Moisés Santillán

Nonlinear Dynamics in the Life Sciences

January 12, 2024

Springer

Chapter 1
Why don't clouds fall?

Abstract In this chapter we introduce differential equations as a tool to study change. We describe a way to classify them, and the nature of its solutions when they exist. We also introduce the concept of initial value problems to narrow down a family of solutions to a particular one, and provide a method to solve a very specific set of easily solvable differential equations. We later study a phenomenon using mathematical modeling, and show step by step the path to do so: beginning with understanding the phenomenon, to writing a mathematical expression to describe it, to solving it using the method introduced earlier in this chapter and finally to discussing the obtained results.

1.1 A Brief introduction to modeling change with differential equations

The natural world around us is constantly changing. To adequately study and understand changing phenomena, appropriate mathematical tools are needed that can describe systems and their evolution over time. Differential equations provide just such a framework, allowing representation of natural processes in terms of rates of change.

Specifically, differential equations relate the derivative of an unknown variable to functions of that variable and/or independent variables. For instance:

$$\frac{dy}{dx} = f(x), \tag{1.1}$$

$$\frac{dy}{dx} = f(x,y), \tag{1.2}$$

$$\frac{\partial y}{\partial x_1} + \frac{\partial y}{\partial x_2} = f(y,x_1,x_2), \tag{1.3}$$

Equations (1.1) and (1.2) involve a single unknown variable and are called ordinary differential equations (ODEs). Equation (1.3) contains partial derivatives with respect to multiple independent variables and is thus a partial differential equation (PDE). Differential equations provide a mathematical language for quantitatively describing natural systems and how they evolve over space and time.

A solution to a differential equation refers to any function that satisfies the equation. Often, differential equations admit multiple solution functions, collectively called a family of solutions. However, real-world systems are uniquely defined by their initial state. By specifying an initial condition, the problem becomes an initial value problem. This additional information allows narrowing down the family of solutions to a single, unique solution corresponding to the system's behavior over time. Some basic differential equations can be solved analytically through explicit mathematical techniques. But for most equations, numerical methods are necessary to approximate solutions since exact closed-form solutions are intractable.

1.2 The method of separation of variables: a technique for solving certain differential equations

Differential equations of the form $dy/dx = F(x,y)$ admit closed-form solutions via the method of separation of variables when $F(x,y)$ can be expressed as the product of a function solely of x, $g(x)$, and a function solely of y, $h(y)$:

$$\frac{dy}{dx} = F(x,y) = g(x)h(y). \tag{1.4}$$

Letting $h(y) = 1/f(y)$, Equation (1.4) can be rearranged such that the variables x and y are isolated on opposite sides:

$$f(y)\frac{dy}{dx} = g(x).$$

This separable equation is then solved by integrating both sides with respect to x:

$$\int f(y)\frac{dy}{dx}dx = \int g(x)dx,$$
$$\int f(y)dy = \int g(x)dx,$$

provided the integrals exist.

1.3 Modeling the fall of cloud droplets

Mathematical modeling utilizing differential equations has increasingly become a valuable complement to experimental research in understanding natural phenomena. Modeling involves more than just solving equations - it is a complete process. The first step is to develop a mathematical model through formulation of governing equations that describe the key aspects of the system or problem. Crafting an appropriate model requires balance, as models need to be sufficiently simple to analyze while still accurately representing the phenomenon. Once established, models are used to gain insights through mathematical and analytical techniques rather than just obtaining numeric solutions. Critically, the results must then be interpreted by relating the mathematical abstractions back to the real system. This contextualization provides enhanced understanding and conclusions. Moreover, answering questions through modeling often leads to new questions, thereby driving development of improved or new models in an iterative process. The overarching goal of mathematical modeling is to complement empirical studies by mechanistically linking mathematics to observed natural phenomena.

As an example of the mathematical modeling process, consider the observable phenomenon of clouds remaining suspended in the sky despite being composed of water/ice; materials denser than air. Why do clouds not simply fall to the earth under the force of gravity given the density of their constituents?

To construct an initial model to investigate this question, we begin by abstracting the intricate, large-scale cloud system into its fundamental building blocks. Clouds are formed from innumerable minuscule droplets and particulates. Therefore, modeling the dynamics of a single representative component isolates the key issue while reducing overall complexity. This abstraction enables formulation of a simplified framework to theoretically explore the forces governing a single constituent's behavior, providing insights into the large-scale behavior of clouds as aggregates of many such parts.

At the level of the individual droplet, the key elements governing its dynamics are the downward pulling force of gravity and the counteracting upward pushing force of air resistance opposing the droplet's movement. Other complex cloud-scale phenomena are negligible for this preliminary model. Figure 1.1 provides a conceptual illustration of this abstracted perspective, visually depicting the isolated droplet subject to only two delineated forces via a diagrammatic force representation.

Consider the positive axis directed along the weight vector. Then by Newton's second law, we can write:

$$ma = mg - F_r, \tag{1.5}$$

where a is droplet acceleration, m its mass, g gravity, and F_r frictional air force. Assuming a spherical droplet, Stokes' law for a sphere in a fluid gives

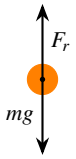


Fig. 1.1 Diagrammatic force representation of a solitary descending water droplet (or ice particle), with the downward pointing weight and upward pointing air resistance delineated.

$F_r = 6\pi\mu r v$, with μ the air viscosity, r the droplet radius, and v velocity. Since $a = \dot{v}$, we can write $F_r = \eta v$, where $\eta = 6\pi\mu r$. Substituting into (1.5) and solving for $\dot{v}$ yields:

$$\dot{v} = \frac{\eta}{m}\left(\frac{mg}{\eta} - v\right). \tag{1.6}$$

Defining $\tau = m/\eta$ and $v_T = mg/\eta$ allows us to rewrite Equation (1.6) as:

$$\dot{v} = \frac{1}{\tau}(v_T - v), \tag{1.7}$$

Applying the method of separation of variables to Equation (1.7) yields:

$$\int \frac{1}{v - v_T} dv = \int \frac{-1}{\tau} dt.$$

Integrating both sides gives:

$$\ln(v - v_T) = \frac{-t}{\tau} + C.$$

Solving this for the velocity v produces:

$$v = v_T + Ce^{\frac{-t}{\tau}}.$$

The initial condition $v(0) = v_0$ gives:

$$C = v_0 - v_T.$$

Substituting this result provides the expression:

$$v = v_T + (v_0 - v_T)e^{\frac{-t}{\tau}}. \tag{1.8}$$

Which describes the time evolution of the droplet's velocity.

To understand the droplet's velocity behavior over very long time periods, we calculate:

$$\lim_{t\to\infty} v = \lim_{t\to\infty} v_T + (v_0 - v_T)\cancelto{0}{e^{\frac{-t}{\tau}}} = v_T$$

This indicates that the droplets will eventually reach a terminal constant velocity of v_T. Figure 1.2 displays plots of Equation (1.8) for differing initial velocities v_0 while holding τ fixed. Notice that all solutions converge to the common value of v_T, regardless of the initial conditions. This illustrates that droplets decelerate over time, eventually matching the characteristic falling speed v_T defined by fluid properties and gravitational acceleration.

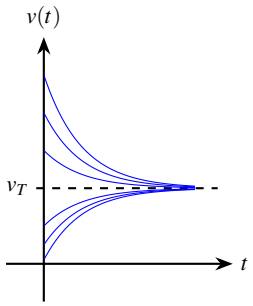


Fig. 1.2 Plots of Equation (1.8) with varying initial velocities v_0 (blue curves), holding τ fixed. The dashed line denotes the terminal velocity v_T. Note that all blue asymptotically approach v_T.

Let us further analyze the terminal velocity. Recall that $v_T = mg/\eta$. However, we can express mass as $m = V\rho$, where V is droplet volume and ρ its density. Since the droplet is assumed spherical, substituting gives:

$$v_T = \frac{V\rho g}{\eta} = \frac{\frac{4}{3}\pi r^3 \rho g}{6\pi\mu r} = \frac{2\rho g}{9\mu} r^2, \tag{1.9}$$

Which reveals the final velocity is proportional to the square of the droplet radius. That is, larger droplets reach a higher terminal speed.

Let us now explore the behavior of Equation (1.8) for varying values of τ. Figure 1.3 displays plots of normalized particle velocity versus time for cases where $\tau/2$, τ, and 2τ are considered, while holding v_0/v_T fixed. Notably, as τ increases, it takes longer for the droplet velocity to approach v_T. This observation leads us to identify τ as the relaxation time of the system, representing the duration required to reach equilibrium.

To investigate the rate at which the droplet velocity approaches v_T, let us examine $\tau = m/\eta$:

$$\tau = \frac{m}{\eta} = \frac{V\rho}{\eta} = \frac{\frac{4}{3}\pi r^3 \rho}{6\pi\mu r} = \frac{2}{9}\frac{\rho}{\mu} r^2, \tag{1.10}$$

Once again, we find that τ is directly proportional to the square of the droplet radius. Therefore, larger droplets have longer relaxation times to attain their terminal velocity.

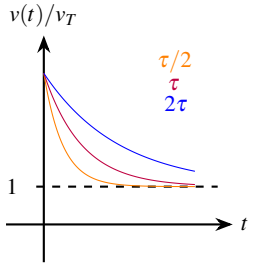


Fig. 1.3 Plots of the normalized particle velocity $v(t)/v_T$, obtained from Equation (1.8), for different values of τ. The orange curve represents $\tau/2$, the purple curve represents τ, and the blue curve represents 2τ. A fixed value of v_0/v_T was considered in all plots.

The above analysis provides insight into why microscopic droplets present in clouds do not readily fall under the force of gravity. Equation (1.10) demonstrates that the relaxation time τ is directly proportional to the square of the droplet radius. Given typical cloud droplet sizes on the order of 10 μm, this results in relaxation times that are exceedingly short; on the order of milliseconds based on standard fluid properties.

According to Equation (1.8), these microscopic droplets would reach 99% of their terminal velocity v_T within a fraction of a second due to the ultrafast relaxation timescale. However, for cloud droplet sizes, the associated terminal velocity v_T itself as given by Equation (1.9) is minuscule, on the order of centimeters per second.

Therefore, over observational timescales relevant to the human eye, micrometer-sized cloud droplets do not perceptibly change in velocity. This is a result of both the very small terminal velocities and exceedingly rapid approach to equilibrium velocity. Their behavior appears static, even though they are slowly descending owing to gravity.

Only larger droplets within clouds have observable terminal velocities and thus visibly fall. Coincidentally, this analysis helps explain the triggering of rain from clouds. Precipitation occurs through coalescence of multiple microscopic droplets into larger falling raindrops, encouraged by processes like droplet collisions, updrafts/downdrafts, cooling temperatures, and presence of foreign particles.

Expanding the model to include ambient air currents

Thus far, our analysis has focused on the simplified scenario where air currents are neglected and droplets fall under the isolated actions of gravity and drag forces. However, clouds exist within a dynamic atmospheric environment

subjected to fluctuating air movements. To gain a more realistic perspective, let us now expand our model to include the effects of ambient wind forces.

Figure 1.4 displays a revised free body diagram depicting a single droplet experiencing an additional time-varying force $f(t)$ imposed by surrounding air motions. Previously, the droplet equilibrium was determined by a balance between the constant downward pull of gravity and upwards drag resistance. With inclusion of $f(t)$, the force budget acting on each droplet becomes more complex as transient lift/drag disturbances alter the net downward acceleration at any given instant. Modeling this third force offers opportunity to deepen understanding of cloud-scale behaviors emerging from interactions between falling constituent droplets and turbulent airflow fluctuations.

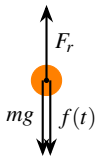


Fig. 1.4 Free body diagram of a single falling droplet of water (or particle of ice). Pointing downwards the weight of the droplet and the force produced by an air current, and pointing upwards the resistance of the air.

Choosing the vertical axis positive in the direction of gravitational acceleration, the equation of motion is:

$$m\dot{v} = mg + f(t) - \eta v, \tag{1.11}$$

where $f(t)$ denotes the time-varying lift or drag imposed by surrounding winds. Let us define the total force acting on the droplet as $F(t) = mg + f(t)$. Isolating the drag term in Equation (1.11) yields:

$$\dot{v} = \frac{\eta}{m}\left(\frac{F(t)}{\eta} - v\right). \tag{1.12}$$

Consistent with our prior analysis, we identify the relaxation time as $\tau = m/\eta$ and the instantaneous terminal velocity as $v_T(t) = F(t)/\eta$. Substituting these into Equation (1.12) gives:

$$\dot{v} = \frac{1}{\tau}\left(v_T(t) - v\right) \tag{1.13}$$

Which has the same form when time-varying wind effects were not included on the droplet motion.

Our previous analysis of the equation of motion without time-varying winds showed that the droplet velocity asymptotically approaches the con-

stant terminal velocity v_T, independent of initial conditions. Additionally, we determined the relaxation time τ governs how rapidly this approach occurs.

When extending our model to include time-varying winds inducing a transient terminal velocity $v_T(t)$, similar behavior is expected for the droplet velocity $v(t)$. Specifically, $v(t)$ will tend towards the instantaneous value of $v_T(t)$ at each moment in time. The rate at which $v(t)$ adjusts to changes in the fluctuating $v_T(t)$ is determined by the relaxation time τ.

For droplets experiencing fluctuations in $v_T(t)$ due to winds, our prior analysis indicates that those with smaller τ will track variations in $v_T(t)$ most closely over time. Within clouds, where droplet sizes yield relaxation times on the order of milliseconds, this property allows for rapid synchronization between $v(t)$ and microscale variations in airflow velocity induced by turbulence.

On observable human timescales, this near-perfect tracking imposed by minute τ obscures individual droplet motions. Instead, clouds appear nebulous and suspended motionless despite droplets smoothly adjusting their velocities downward while also following transient atmospheric currents. Therefore, the small droplet sizes characteristic of clouds effectively couple velocities to instantaneous air movements, contributing significantly to their visual appearance.

Discussion

In this chapter, we introduced differential equations and discussed using them to develop mathematical models of physical phenomena. We applied this approach to investigate the behavior of cloud droplets and why they do not visibly fall under gravity.

Through analytical modeling of a cloud droplet experiencing gravity, drag, and wind forces, we derived equations of motion and solved them to determine droplet velocity over time. This revealed that each droplet reaches a constant terminal settling velocity, v_T, irrespective of initial conditions.

Analysis of the model showed the scaling relationships of both v_T and relaxation time constant τ with droplet radius r. Specifically, we found v_T and τ are directly proportional to r^2. Given cloud droplet radii are on the micron scale, this results in two key properties:

1. Droplets attain v_T extremely rapidly due to exceedingly short τ.
2. The actual magnitude of v_T is minuscule due to the r^2 scaling.

Therefore, while droplets gradually descend at their slow terminal velocities, this imperceptible motion occurs below human observational resolution scales.

Additionally, ambient winds disrupting continuous descent contribute further to clouds' static visual appearance. Overall, the mathematical modeling

elucidated the underlying reasons clouds evade falling visibly despite gravitational forces acting on constituent droplets.

Even simple mathematical models of natural phenomena, like the linear drag model developed here for cloud droplets, can provide deep insights into observable behaviors that may otherwise be taken for granted. By distilling relationships down to their essential mathematical elements, qualitative and quantitative predictions emerge regarding how underlying physical factors control system-level properties.

The rest of this book is dedicated to expanding on such modeling techniques, specifically through nonlinear dynamics approaches. Subsequent chapters will introduce nonlinear methods and apply them to study the time-evolving behavior of various biological systems. The overarching goal is to utilize analytical and computational nonlinear dynamics tools to elucidate the unexpectedly complex and sometimes counterintuitive dynamics that often arise in living systems due to nonlinear interactions between components.

Chapter 2
From Ancient Math to Modern Science: The Fantastic Journey of the Exponential Function

Abstract For centuries, the exponential function has been a cornerstone of mathematics, and its influence has extended far beyond the confines of this discipline. This chapter will delve into the rich history of the exponential function, tracing its origins back to ancient civilizations and exploring its evolution over time. The chapter will also highlight some of the key mathematical breakthroughs that led to the exponential function. Finally, we will present a brief survey of how the exponential function is used in fields such as physics, and engineering.

A story goes that once God Krishna took the form of a wise man and went to the king's court. He challenged the king to a game of chess because he loved playing it. Before the game started, they had to decide what the prize would be if the sage won. The sage said that he only wanted a small amount of rice, but the amount had to be calculated using the chessboard. A grain of rice would be placed in the first square and the number of rice grains would double for each successive square on the board. The king lost the game and had to give the sage the rice. But as he started putting the rice grains on the board, the king realized that he wouldn't have enough rice to pay his debt. Krishna then showed himself in his true form and told the king that he didn't have to pay all the rice right away, but he could do it gradually. Every day, the king would serve free rice pudding to people who came to the temple until the debt was paid off.

This story has different versions. In some versions, God Krishna is replaced by a servant, the inventor of chess, or a craftsman who makes the best chessboards. In others, rice is substituted by wheat. The ending is also different. In some versions, the ruler has the person who should get the reward killed, while in others, the reward is given only if each grain is counted individually. But the message stays the same: the explosive increase in a pattern where

each step is multiplied by the same number (geometric progression) instead of just being added by the same amount (arithmetic progression).

With a history dating back to the Greeks and possibly even the Sumerians, geometric progressions—a discrete form of the exponential function—boasts a rich legacy. Despite this, it wasn't until the 18th century, with the contribution of some of the world's most renowned mathematicians, that the exponential function was finally uncovered. Nevertheless, its influence has been tremendous and spans across numerous scientific domains. A key milestone in this history is the creation of logarithms, which was a seminal event.

Logarithms were independently developed by the mathematicians John Napier, from Scotland, and Jost Bürgi, from Switzerland, to simplify computations in spherical trigonometry, which is used in astronomy and celestial navigation. Despite Bürgi's probable creation of his system around 1600, Napier's discovery was published first in 1614 in the book Mirifici Logarithmorum Canonis Descriptio (Description of the Wonderful Canon of Logarithms), making him widely known as the inventor of logarithms and greatly influencing its subsequent evolution.

Logarithms establish a connection between the operation of multiplication on the positive real numbers and addition on the real number line. Napier, in particular, viewed logarithms as the relationship between two particles moving along a line—one at a constant speed, the other at a speed proportional to its distance from a fixed endpoint. In current terms, Napier's logarithm (NapLog) can be related to the natural logarithm ($\ln$) in this way:

$$\mathsf{NapLog}(x) = -10^7 \ln(x/10^7).$$

English mathematician Henry Briggs made two visits to Edinburgh to collaborate with John Napier in 1616 and 1617. During their conversations, they reached an agreement on Briggs' proposed modification to Napier's logarithms. After his second trip, Briggs released the first table of his improved logarithms, now known as common or base 10 logarithms ($\log_{10}$), in 1617. The widespread use of common logarithms grew rapidly due to their ease in performing complex calculations during a time when calculators were not available. This was mainly due to the fact that our numbering system is built on powers of 10. Nevertheless, the natural logarithm holds a more prominent place in the history of mathematics because of its impact on the discipline's evolution.

In 1649, Alphonse Antonio de Sarasa, who was previously a student of Grégoire de Saint-Vincent, demonstrated that the area $A(t)$ of the region bounded by the hyperbola $xy = 1$ from $x = 1$ to $x = t$ obeys the following relation common to all logarithmic function:

$$A(t \times u) = A(t) + A(u).$$

It was soon realized that this characteristic could lead to the creation of a new type of logarithm. Mercator, in fact, published the first tables of what

is now referred to as the natural logarithm in his book Logaritmotechnia in 1668.

In 1748, Leonhard Euler's classic book Introductio in analysin infinitorum (Introduction to the Analysis of the Infinite) marked the final step in the development of logarithms, exponential functions, and trigonometric functions. Prior to Euler, these mathematical concepts were usually defined using integral calculus. However, Euler changed this by introducing exponentiation a^x for constant a in the positive real numbers, leading to the creation of the logarithm to base a. He also named the natural logarithm, calling it the "natural or hyperbolic logarithm" due to its connection to the quadrature of the hyperbola. Euler designated number e as the base for the natural logarithm, i.e. the number whose natural logarithm is equal to 1.

Jacob Bernoulli's groundbreaking work in compound interest paved the way for the discovery of the constant later named e by Euler. Bernoulli's contribution can be succinctly described as follows: he examined the growth of capital when invested at an annual interest rate of 100% and is compounded at n intervals. At the end of the year, the capital would have multiplied by a factor of

$$\left(1+\frac{1}{n}\right)^n.$$

Bernoulli showed that as the number of compounding intervals increases, this factor approaches a constant value between 2 and 3.

Euler expanded upon Bernoulli's work by defining the exponential and natural logarithmic functions as follows:

$$\begin{aligned}\exp(x) &= \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n \\ \ln(x) &= \lim_{n\to\infty} n(x^{1/n}-1).\end{aligned}$$

Additionally, Euler established that the exponential function is a exponentiation function with base e ($\exp(x) = e^x$) and that the exponential and logarithmic functions are inverse of each other.

Godfrey H. Hardy wrote in A Mathematician's Apology that "a mathematical idea is significant if it can be connected, in a natural and illuminating way, with a large complex of other mathematical ideas. Thus a serious mathematical theorem, a theorem which connects significant ideas, is likely to lead to important advances in mathematics itself and even in other sciences." Hardy used Pythagoras's proof of the irrationality of $\sqrt{2}$ as an example, highlighting how a simple and elegant theorem can open up new avenues for the development of mathematics.

The exponential function is a prime example of a mathematical idea that has left a lasting impact, not just in mathematics but in other fields as well. It is considered a cornerstone for the advancement of all modern areas of mathematics. Its applications are diverse and can be seen in fields such as physics, where it explains processes ranging from radioactive decay to popu-

lation growth. In finance, it is employed in modeling interest rates and stock prices. The exponential function is also a critical component in engineering, contributing to the development of electrical circuits, control systems, and communication systems, just to mention a few examples.

In Introductio in analysin infinitorum, Euler introduced the equation named after him:

$$e^{ix} = \cos x + i \sin x.$$

Specifically, when $x = \pi$, this formula leads to the well-known Euler identity:

$$e^{i\pi} + 1 = 0.$$

which is considered by many to be one of the most beautiful equations in mathematics, as it combines the three basic mathematical operations (addition, multiplication, and exponentiation) and relates five fundamental mathematical constants (0, the additive identity; 1, the multiplicative identity; the unit of imaginary numbers, i; e and π). The Euler identity is a point of convergence of various mathematical disciplines, such as arithmetic, trigonometry, and complex number theory.

Euler formula has been a vital aspect in physics and engineering, particularly through Fourier analysis. This method allows the transformation of complex functions into sums of simple trigonometric functions. Some of the most significant uses include:

- Signal processing: Fourier analysis is used to analyze and manipulate signals such as audio, images, and voice.
- Image compression: Fourier analysis is used in image compression algorithms to reduce the amount of data needed to represent an image.
- Spectral analysis: Fourier analysis is applied to study the frequency components of signals, such as those produced by vibrating objects or electromagnetic waves.
- Filter design: Fourier analysis is used to create electronic filters that remove unwanted frequencies from signals.
- Heat transfer: Fourier analysis is applied to examine heat flow in solids and fluids.
- Quantum mechanics: Fourier analysis is used to analyze the behavior of particles in a quantum state.
- Seismology: Fourier analysis is employed to examine the propagation of seismic waves.
- Electromagnetic analysis: Fourier analysis is applied to study the behavior of electromagnetic fields.

In conclusion, the exponential function transcends the realm of mathematics theory and has become a vital tool across various fields. Its impact has been revolutionary in shaping our comprehension of both natural and social sciences and has been a driving force behind numerous technological advancements.

Further reading. For those seeking a deeper understanding of the history of logarithmic and exponential functions, the History of the Exponential and Logarithmic Concepts series of articles, published in 1913 in the American Mathematical Monthly, volume 20, numbers 1 through 7, may be of interest. These articles can be accessed for free on Jstor at the following URL: https://www.jstor.org/

Chapter 3
Exponential Decay

Abstract This chapter develops both deterministic and stochastic models to analyze how population size changes over time under the assumption of an age-independent death rate in closed populations. Beginning with a distribution function describing population age structure, a convection equation governs aging in the absence of deaths. Introducing constant mortality leads to exponential decay of total size. A stochastic approach models population size as a random variable and relates transition probabilities, recovering the deterministic solution when averaging expected size. Solving for individual lifetime distributions links dynamics across scales through averaging. Extensions to age-dependent but stationary mortality demonstrate self-organized exponential decay, connecting perspectives from random interactions to emergent collective behaviors. The established foundations connect stochastic and deterministic views on simple populations and lay the groundwork for further complexity.

Population dynamics is the study of how populations change over time and space due to various biological and environmental factors. Central topics in population dynamics include birth, death, immigration and emigration, density dependence, and biotic and abiotic interactions between individuals within a population or between species.

In this chapter, we will begin exploring population dynamics by studying a simple death process in closed populations. Specifically, we will develop both deterministic and stochastic models to analyze how population size changes over time under the assumption of an exponentially distributed age-independent death rate. This forms the basis for understanding more complex population dynamics incorporating additional realistic features like age-structure, density dependence, and other stochastic birth and death mechanisms.

By establishing these foundations, we aim to connect individual-level stochasticity to emergent deterministic behaviors at the population scale.

Dynamics of aging

Let us analyze the dynamics of a closed population with no births or deaths over time. We define the function $\eta(\tau,t)d\tau$ as the fraction of individuals in the population with ages in the interval $[\tau,\tau+d\tau]$ at time t. This function fully describes the population state at any given time and how it evolves. Specifically, the total population at any time t is given by:

$$N(t) = \int_0^\infty \eta(\tau',t)d\tau'.$$

In the absence of births, deaths, and migration, individuals can only age over time. This means the plot of the function $\eta(\tau,t)$ shifts to the right as time passes. We can mathematically express this as:

$$\eta(\tau,t+\Delta t) = \eta(\tau-\Delta t,t).$$

Expanding the left-hand side (LHS) and right-hand side (RHS) terms in a Taylor series about t and τ, respectively, gives:

$$\eta(\tau,t) + \frac{\partial \eta}{\partial t}\Delta t + \cdots = \eta(\tau,t) - \frac{\partial \eta}{\partial \tau}\Delta t + \ldots.$$

Simplifying and taking the limit as $\Delta t \to 0$ yields the governing equation:

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \eta}{\partial \tau}. \tag{3.1}$$

The RHS can be viewed as an age-related flux term. Therefore, this equation expresses conservation of population as the rate of change within an infinitesimal age interval equals the net outflux due to aging.

Since the total population $N(t)$ was assumed to remain constant over time due to no births, deaths, or migration, the last integral must satisfy:

$$\int_0^\infty \frac{\partial \eta}{\partial \tau'}d\tau' = 0.$$

This expresses the fact that aging redistributes individuals among age classes but does not change the total population size.

Age-independent death rate

We can modify Equation (3.1) to account for deaths by including an additional death rate term

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \eta}{\partial \tau} - \mu(\tau)\eta, \tag{3.2}$$

where $\mu(\tau)$ is the death-rate constant for individuals of age τ. It represents the probability per unit time that an individual of age τ dies.

Integrating Equation (3.2) gives:

$$\frac{dN}{dt} = -\int_0^\infty \mu(\tau')\eta(\tau',t)d\tau'.$$

Now assume the death rate is independent of age. Then:

$$\frac{dN}{dt} = -\mu N. \tag{3.3}$$

This differential equation has the well-known solution of exponential decay:

$$N(t) = N_0 e^{-\mu t},$$

where N_0 is the initial population size. This demonstrates that under the assumption of a constant, age-independent death rate, the population will decrease exponentially over time.

An important parameter that arises from this solution is the population half-life ($t_{1/2}$), defined as the time required for the population to decrease to half its initial value. Specifically:

$$\frac{N_0}{2} = N_0 e^{-\mu t_{1/2}}$$

Solving this equation for $t_{1/2}$ gives:

$$t_{1/2} = \frac{\ln 2}{\mu} \tag{3.4}$$

Stochastic description

In the previous section, we modeled population dynamics deterministically and showed the population decays exponentially over time when death rates are independent of age. Here, we develop a stochastic framework to analyze this process more rigorously.

Let n be a random variable representing population size. Define $P(n = N;t)$ as the probability the population has a value of N individuals at time t:

$$P(n = N;t)$$

Additionally, let μ denote the probability per unit time that an individual dies independently of others. Considering these variables, the evolution of $P(n = N;t)$ is described by:

$$P(n = N;t+\Delta t) = P(n = N;t)(1-\mu N\Delta t) + P(n = N+1;t)\mu(N+1)\Delta t \quad (3.5)$$

This equation accounts for two possibilities that could result in a population of size N over a short time interval Δt. 1) the population was already size N at time t and no deaths occurred. 2) The population was size $N+1$ at time t and one death occurred. We assume that Δt is small enough that more than one death is highly unlikely.

We can rearrange Equation (3.5) to obtain:

$$\frac{P(n = N;t+\Delta t) - P(n = N;t)}{\Delta t} = P(n = N+1;t)\mu(N+1) - P(n = N;t)\mu N$$

Taking the limit as the time interval Δt approaches 0 yields the following differential equation governing the time evolution of the probability $P(n = N;t)$:

$$\frac{dP(n = N;t)}{dt} = P(n = N+1;t)\mu(N+1) - P(n = N;t)\mu N \quad (3.6)$$

Equation (3.6) describes the forward Kolmogorov equation for this stochastic death process. It relates the time derivative of the probability of being in state N individuals to the probabilities of transitions between states due to individual death events.

There are several approaches to analyze Equation (3.6). One method is to take the time derivative of the expected population size, represented by:

$$E_n = \sum_{N=0}^{\infty} NP(n = N;t)$$

By multiplying both sides of Equation (3.6) by N and taking the sum from 0 to infinity, we obtain:

$$\frac{d}{dt}\sum_{N=0}^{\infty} NP(n = N;t) = \mu\sum_{N=0}^{\infty} N(N+1)P(n = N+1;t) - \mu\sum_{N=0}^{\infty} N^2P(n = N;t)$$

Through some algebraic steps, this equation can be shown to reduce to:

$$\frac{dE_n}{dt} = -\mu E_n \quad (3.7)$$

Notice that Equation (3.7) is identical to the deterministic Equation (3.3). This implies the deterministic solution corresponds to the average behavior of repeated stochastic experiments, rather than describing any single exper-

iment. It predicts exponential decay when considering expected population sizes over numerous trials.

Another approach is to find an exact solution to Equation (3.6). It can be verified through substitution that:

$$P(n=N;t)=\frac{E_n^N(t)e^{-E_n(t)}}{N!},$$

where $E_n(t)$, satisfies Equation (3.7). This has the form of a Poisson distribution, with mean E_n and standard deviation $\sqrt{E_n}$.

For large values of E_n, the standard deviation becomes negligible compared to the mean. Therefore, even though the deterministic model in Equation (3.3) does not describe individual stochastic trajectories, its predictions are expected to closely match the behavior of single realizations when population sizes are sufficiently large.

While stochastic fluctuations are prominent for small populations, the deterministic exponential decay approximation becomes increasingly accurate as the number of individuals grows. This solution helps connect the statistical properties of the underlying stochastic process to the emergent deterministic behavior predicted by the differential equation model.

Here is one way to rewrite the section on the individual lifetime distribution:

3.1 Individual Lifetime Distribution

As individual death is a stochastic process, the survival time of each individual, denoted by the random variable τ, is also random. Let $P(\tau=T)dT$ represent the probability that an individual survives up to age T and dies between ages T and $T+dT$. We can write an equation for this probability as:

$$P(\tau=T)dT=\left(1-\int_0^T P(\tau=T')dT'\right)\mu dT$$

The term in parentheses is the probability of surviving until age T, while μdT is the probability of dying between T and $T+dT$.

Differentiating this expression gives the differential equation:

$$\frac{dP(\tau=T)}{dT}=-\mu P(\tau=T)$$

Subject to the normalization condition, the solution is:

$$P(\tau=T)=\mu e^{-\mu T}$$

Therefore, individual lifetimes follow an exponential distribution. The mean lifetime is thus:

$$E_\tau = \frac{1}{\mu}$$

Discussion

So far, we have analyzed a simple decaying population model considering both deterministic and stochastic frameworks. The key deterministic result was that when death rates are constant over time, the total population size decays exponentially according to Equation (3.3). By developing a stochastic description, we were able to account for randomness at the individual level. Interestingly, when considering expected population sizes, the deterministic and stochastic descriptions coincide as shown by Equation (3.7).

This emergence of deterministic behavior from underlying stochastic processes is an important phenomenon. While survival of individuals is inherently random, averaging over many trials washes out variability, resulting in smooth exponential decay. This demonstrates how populations self-organize simpler collective dynamics from complex interactions between constituents.

The death rate parameter μ takes on different meanings depending on the description. Deterministically, it characterized the system's exponential decay profile. Stochastically, it represented the probability of individual mortality. Such multi-scale modeling allows μ to provide insight across descriptive levels.

In summary, even simple population models like exponential decay showcase the interplay between stochastic and deterministic perspectives. Randomness at the individual scale shapes probabilistic population fluctuations, yet deterministic laws emerge at larger scales where variation averages out. This theory establishes foundations for more realistic extensions incorporating additional biological complexities.

Epilogue

Let us reexamine the population distribution function $\eta(\tau,t)$. We can define the normalized function:

$$\rho(\tau,t) = \frac{\eta(\tau,t)}{N(t)},$$

where $N(t) = \int_0^\infty \eta(\tau',t)d\tau'$. Function $\rho(\tau,t)$ describes the probability density that a randomly selected individual from the population has age τ at time t.

Rewriting Equation (3.2) in terms of ρ:

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \eta}{\partial \tau} - N(t)\mu(\tau)\rho(\tau,t).$$

Upon integration, this leads to:

$$\frac{dN(t)}{dt} = -\widetilde{\mu}(t)N(t), \tag{3.8}$$

where $\widetilde{\mu}(t) = \int_0^\infty \mu(\tau')\rho(\tau',t)d\tau'$ is the average death rate.

Notably, if the age distribution $\rho(\tau,t)$ remains stationary (independent of t), then the average death rate $\widetilde{\mu}$ is constant. In this case, Equation (3.8) again predicts exponential decay of $N(t)$, even with age-dependent mortality $\mu(\tau)$.

In conclusion, a stationary age structure is another way of producing exponential population decay through a constant average death rate. Therefore, observing exponential decay alone does not fully characterize the underlying stochastic process. Additional information, such as the lifetime distribution of individuals, is needed to distinguish the underlying stochastic process.

Chapter 4
Exponential Growth

Abstract In the previous chapter, we introduced our first population dynamics model, incorporating deaths alone, and demonstrated that this leads to exponential decay. Here, we expand the model to include both births and deaths. Considering these combined processes, we show that populations can either grow or decline exponentially, depending on whether the birth rate exceeds the death rate or vice versa. We delve more deeply into the phenomenon of exponential growth, investigating its remarkable properties. To build intuition, we first consider a thought experiment exploring the explosive potential of exponential growth in a bacterial culture. We then examine some key mathematical underpinnings, showing how exponential functions exhibit geometric behavior when evaluated over arithmetic sequences. Understanding these fundamental characteristics provides critical insight into how even modest growth rates may translate to astonishingly rapid increases over long timescales. In this chapter, our aim is to develop a richer appreciation for both the power and limitations of exponential growth dynamics.

4.1 Exponential Growth

In the previous chapter, we examined how deaths influence population dynamics. Now, we will expand our study to incorporate the impact of births. It is important to consider that, in the absence of migration, new individuals arise in the population solely through the reproduction of existing individuals, as spontaneous generation is not possible. Therefore, we can modify Equation (3.2) to accommodate births as follows:

$$\frac{\partial \eta}{\partial t} = -\frac{\partial \eta}{\partial \tau} + \beta(\tau)\eta - \mu(\tau)\eta. \tag{4.1}$$

In this equation, the term $\partial\eta/\partial t$ represents the rate of change of the population density η with respect to time t. The first term on the right-hand side, $-\partial\eta/\partial\tau$, accounts for the effect of aging on the population. The second term, $\beta(\tau)\eta$, introduces the influence of births, where $\beta(\tau)$ represents the birth rate as a function of age (τ). Finally, the third term, $-\mu(\tau)\eta$, considers the mortality rate $\mu(\tau)$, which represents the rate at which individuals of age τ die.

Based on the findings presented in the previous section of Chapter 3, and assuming that the probability density $\rho(\tau,t) = \nu(\tau,t)/N(t)$ remains stationary, we can integrate Equation (4.1) to obtain the following expression:

$$\frac{dN}{dt} = \widetilde{\beta}N - \widetilde{\mu}N, \tag{4.2}$$

where $\widetilde{\beta}$ and $\widetilde{\mu}$ represent the average birth and death rates, respectively. These parameters are defined as $\widetilde{\beta} = \int_0^\infty \beta(\tau')\rho(\tau')d\tau'$ and $\widetilde{\mu} = \int_0^\infty \mu(\tau')\rho(\tau')d\tau'$. The population size is represented by $N(t) = \int_0^\infty \eta(\tau')d\tau'$. It is worth noting that, if both β and μ are age-independent, we obtain the same result without requiring the stationarity of η.

By introducing an effective growth rate $\beta = \widetilde{\alpha} - \widetilde{\mu}$, we can reformulate Equation (4.2) as follows:

$$\frac{dN}{dt} = \widetilde{\alpha}N, \tag{4.3}$$

This differential equation has a solution given by:

$$N(t) = N_0 e^{\alpha t}, \tag{4.4}$$

where N_0 represents the initial population size. If the death rate surpasses the birth rate, as studied in the previous chapter, the population would decay exponentially. However, if the opposite scenario occurs, where the birth rate surpasses the death rate, the population would experience exponential growth.

4.2 Explosiveness of the Exponential Growth

To develop an intuitive understanding of exponential growth, let's consider a thought experiment involving an E. coli bacterial culture initiated from a single bacterium, assuming it can grow exponentially indefinitely. In such a scenario, the time evolution of the culture's mass is governed by the following equation:

$$m(t) = m_b e^{\alpha t}, \tag{4.5}$$

where m_b represents the initial mass of an individual bacterium, and α denotes the growth rate.

Now, let's explore how long it would take for the culture's mass to reach the magnitude of planet Earth ($M_E \approx 6 \times 10^{24}$ kg). Given that an individual bacterium weighs approximately 10^{-15} kg and an E. coli culture can double its mass every 30 minutes, we can solve for t in Equation (4.5):

$$t = \frac{1}{\alpha} \ln \frac{m(t)}{m_b}.$$

By substituting $m(t) = M_E$ and $\alpha = \ln(2)/30\,\text{min}$, we can calculate the required time:

$$T_E = \frac{30,\text{min}}{\ln 2} \ln \frac{M_E}{m_b} \approx 4,000\,\text{min} \approx 2.8\,\text{days}.$$

In other words, an exponentially growing bacterial culture starting from a single bacterium would accumulate a mass equal to that of planet Earth in less than 3 days. This thought experiment vividly illustrates the remarkable explosive potential of exponential growth.

4.3 Geometric Sequences and the Exponential Function

To comprehend the explosive nature of exponential growth, it is valuable to explore a fundamental property of the exponential function. Let us consider the function defined as:

$$f(x) = e^{ax}.$$

Now, suppose we have a value x_2 such that:

$$f(x + x_2) = 2f(x).$$

In other words:

$$e^{a(x+x_2)} = 2e^{ax}.$$

By solving for x_2 in the equation above, we find:

$$x_2 = \frac{\ln 2}{\alpha}.$$

Notably, the value of x_2 remains constant regardless of the initial value of x. This implies that, irrespective of the starting point, the value of $f(x)$ doubles when a value of x_2 is added to the argument. Consequently, if we construct an arithmetic sequence of the form:

$$x_{i+1} = x_i + x_2,$$

the sequence:

$$y_i = e^{ax_i},$$

will exhibit geometric behavior ($y_{i+1} = 2y_i$). The remarkable growth potential of geometric sequences has captivated humanity for centuries, giving rise to legends like the one illustrated in Chapter 2.

4.4 Discussion

While the exponential growth model provides valuable insight into population dynamics, it has important limitations when considered over long time periods. As we have seen, even tiny growth rates predict populations will eventually exceed all available resources as their mass approaches infinity. In reality, finite carrying capacities constrain population expansion. Additionally, exponential growth assumes resources are unlimited and external influences negligible, violating the basic principles of resource competition and natural regulation within closed ecological systems. In the next chapter, we will introduce a more realistic population model called the logistic model. The logistic equation incorporates density-dependent effects like resource competition and saturation, allowing populations to stabilize at an environmental carrying capacity. Studying the logistic model will provide a more nuanced understanding of how biotic and abiotic factors interact to govern long-term population persistence in nature. This will lay the foundation for exploring more complex dynamic interactions between species in ecological communities.

Chapter 5
Local Stability Analysis

Abstract This chapter introduces the key concepts and methods involved in performing a local stability analysis of steady states for nonlinear dynamical systems. Steady states, or fixed points, are defined as constant solutions to a differential equation that do not change over time. Their stability determines whether nearby solutions will converge or diverge from the steady state. While graphical analysis can assess stability in 1D, a more rigorous analytical approach is required for higher dimensions. The method developed is to linearize the nonlinear dynamics near a steady state, resulting in a linear approximation governing small perturbations from the fixed point. The stability criterion is then derived from the solution to this linearized system - a steady state will be stable only if perturbations decay over time, indicated by the slope of the original function being negative at that point. This technique allows classification of steady state stability across any system dimensionality using analytical mathematics.

In the preceding chapters, we introduced basic linear population dynamics models and analyzed how their behavior evolves over time. However, to develop more realistic representations, it is necessary to consider nonlinear differential equations. Attaining a rigorous global understanding of these models necessitates a formal stability analysis of their steady states. A steady state corresponds to a constant solution. For a population, this represents an equilibrium point at which birth and death rates precisely balance. Studying the stability of steady states allows for the classification of long-term dynamical behavior and the discernment of global dynamics arising from local perturbations.

Just as a balance at rest indicates equilibrium but imparts no insight into subsequent motion upon disturbance, steady states alone do not reveal dynamical fate. Their stability determines their attracting or repelling character, with implications for qualitative trajectories that may resemble balls

inducted into concavities or convexities. Linearizing functions near steady states permits estimation of the evolution of small deviations, providing local qualitative forecasts. When stability analysis and linearization are combined, they provide qualitative understanding that exceeds model solutions, just as one may predict a ball's motion from terrain alone despite transients obscuring endpoints. This chapter will introduce the tools to perform local stability analysis of 1-dimensional dynamical models.

5.1 Steady States

Consider the univariate ordinary differential equation:

$$\frac{dx}{dt} = f(x), \tag{5.1}$$

wherein $f(x)$ denotes an arbitrary smooth function. Steady states, or fixed points, for this differential equation represent unchanging constant solutions independent of time (t).

Let $x(t) = x^*$ represent a steady state. Its derivative with respect to time is null given its lack of time-dependence. Furthermore, by definition it must satisfy Equation (5.1). Upon substitution, we obtain:

$$f(x) = 0$$

Accordingly, the steady states equate to the roots of function $f(x)$. This provides a method for determining the fixed points of any first-order ordinary differential equation by resolving the relationship $f(x^*) = 0$.

5.2 Assessment of Fixed Point Stability via Graphical Analysis

Consider the function $f(x)$ depicted in Figure 5.1. Two roots, denoted x_1 and x_2, of the equation $f(x) = 0$ are evident. Consequently, these locations define the steady states of the differential equation $dx/dt = f(x)$.

Let us now analyze the stability of the fixed points. Consider x_1 and note the positive slope of $f(x)$ at this location. A positive slope implies that for any initial condition marginally greater than x_1, the value of $f(x)$ will be positive. Accordingly, this yields a positive rate of change for the value of x. Thus, the solution will move in a progressively greater direction, distancing itself from x_1. Similarly, for any initial value slightly less than x_1, the slope is such that the rate of change of x is negative. Consequently, the solution will decrease over time, drifting ever farther from x_1. From this analysis, we can see that x_1 is unstable as it repels nearby solutions over time.

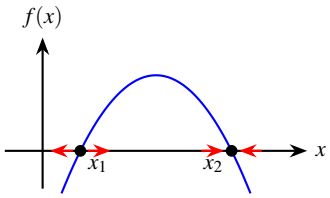


Fig. 5.1 Graphical depiction of the function $f(x)$ with two roots. This figure illustrates the function $f(x)$ which possesses two roots, denoted by x_1 and x_2. The slope of $f(x)$ is positive in the vicinity of x_1, as depicted by the upward sloping tangent. In contrast, the slope is negative near x_2, as shown by the downward sloping tangent. These slope characteristics determine the qualitative behavior of solutions initialized proximate to the steady states. Specifically, red arrows indicate that perturbations tend to diverge from x_1 while converging towards x_2. Therefore, x_1 represents an unstable fixed point whereas x_2 denotes a stable steady state. This graphical analysis conveys how the slope's sign at steady state locations qualifies their attracting or repelling nature.

In contrast, let us examine the stability of the other fixed point, x_2. At this location, the slope of $f(x)$ is negative. Therefore, for any initial condition marginally greater than x_2, the value of $f(x)$ will be negative. This implies the rate of change of x will also be negative, causing the solution to decrease in value over time and move closer to x_2. Similarly, for initial values slightly less than x_2, the negative slope indicates the rate of change of x is positive. Hence, the solution will increment in value, again tending towards x_2. We can thus deduce that perturbations from x_2 will be attracted back to this fixed point location over time. Consequently, x_2 is stable as it attracts solutions.

In summary, this graphical analysis demonstrates that the stability of a fixed point for the differential equation $dx/dt = f(x)$ is determined solely by the sign of the slope of the function $f(x)$ evaluated at that fixed point. A positive slope implies perturbations will move further away from the fixed point over time, characterizing it as unstable. In contrast, a negative slope signifies that perturbations will be attracted back to the fixed point location, marking it as stable.

5.3 Local stability analysis

While the preceding graphical approach provides useful qualitative insights into fixed point stability for univariate systems, it is not readily generalized to higher dimensions. The graphic analysis relies on visualizing the slope of a scalar-valued function, which is only possible for one-dimensional models. However, real-world phenomena often evolve according to multidimensional dynamical systems that cannot be embedded in two-dimensional space for intuitive slope evaluation. Therefore, a more systematic technique is required that permits stability classification of fixed points for differential equations

inhabiting abstract dimensionalities. In this section, we develop the method of local stability analysis through linearization. By approximating nonlinear dynamics near steady states via linear algebra, this approach furnishes a quantitative framework to assess stability in any dimensional system. These analytical tools supersede qualitative geometric intuitions, instead providing mathematics directly applicable to complexity beyond what may be visualized.

Consider once more the differential equation:

$$\frac{dx}{dt} = f(x). \tag{5.2}$$

We want to analyze the stability of a steady state value x^*. To do this, we look at an arbitrary solution $x(t)$ that starts out close to x^*. For the steady state to be locally stable, we need to show that $x(t)$ converges to x^* as time goes to infinity:

$$\lim_{t\to\infty} x(t) = x^* \tag{5.3}$$

It's helpful to define the "distance function" $\delta x(t)$:

$$\delta x(t) = x(t) - x^*. \tag{5.4}$$

This tells us how far $x(t)$ is from the steady state x^* at any time t. For our analysis to work, $\delta x(t)$ needs to stay much smaller than x^* as time changes.

We can write the stability condition in terms of $\delta x(t)$ as well:

$$\lim_{t\to\infty} \delta x(t) = 0. \tag{5.5}$$

In other words, for x^* to be stable, any small perturbations need to disappear over long times.

We seek to derive the governing differential equation for the dynamics of $\delta x(t)$. It can be readily shown from its definition in Equation (5.4) that:

$$\frac{d\delta x(t)}{dt} = \frac{dx(t)}{dt}.$$

Substituting this expression and Equation (5.4) into the original system dynamics in Equation (5.2) yields:

$$\frac{d\delta}{dt} = f(x^* + \delta x).$$

Leveraging the assumption that δx is small relative to x^*, we can approximate $f(x^* + \delta x)$ via a Taylor series expansion truncated after the first order term:

$$f(x^* + \delta x) \approx f(x^*) + f'(x^*)\delta x,$$

where $f'(x^*) = df/dx|_{x=x^*}$. Furthermore, since x^* defines a steady state, $f(x^*) = 0$. Thus, the differential equation for δx reduces to the following linear form:

$$\frac{d\delta x}{dt} = f'(x^*)\delta x.$$

The solution to this equation takes the form

$$\delta x(t) = \delta x_0 e^{f'(x^*)t},$$

where δx_0 denotes the initial perturbation from the steady state x^*. It is evident from this last equation that a necessary condition for $\delta x(t)$ to asymptotically approach zero as $t \to \infty$ is that $f'(x^*) < 0$. In other words, the steady state x^* will only exhibit stable behavior if the slope of the function $f(x)$ is negative when evaluated at $x = x^*$. In summary, local stability is assured only if perturbations decay over time, which occurs solely when the linearized dynamics represented by $f'(x^*)$ are negative.

5.4 Discussion

In this chapter, we introduced the method of local stability analysis and applied it to determine the stability of steady states for a univariate nonlinear dynamical system. Through linearization near steady states and solution of the resulting linear differential equation governing perturbations, we recovered the same criterion for stability established by the graphical analysis in the previous chapter - namely, that stability is ensured only when the slope of the governing function is negative, as evaluated at the steady state. Interestingly, this analysis arrived at the same qualitative conclusion through an analytical approach, rather than a geometric visualization. This mathematical technique has the advantage of being readily generalized to higher-dimensional systems that cannot be depicted graphically. The method of local stability analysis through linearization provides a systematic framework to assess fixed point stability for nonlinear dynamical models of any dimensionality. In this way, insights into long-term dynamical behavior near steady states can be captured quantitatively rather than relying on low-dimensional intuitions. This analytical approach will prove invaluable for nonlinear multi-species population and ecosystem models in subsequent chapters.

Chapter 6
On the interconnected origins of statistics, probability theory and population dynamics

Abstract This chapter traces the interconnected conceptual origins of statistics, probability theory, and population dynamics that emerged in seminal developments of the 17th century. It examines foundational contributions from figures such as Cardano, Pascal, Graunt, Huygens, and Bernoulli who established the theoretical and empirical bases for analyzing natural and social phenomena using mathematical and data-driven methods. Key advances discussed include the early application of quantitative methods to games of chance by Cardano, the founding of probability theory by Pascal and Fermat, Graunt's establishment of empirical demography, and the development of probabilistic population modeling through life tables. The paper argues this transformative period saw the beginning of quantitative, evidence-based approaches across scientific domains through linked conceptual roots in analyzing randomness, uncertainty, and human populations.

The fields of statistics, probability theory, and population dynamics share interconnected conceptual origins that emerged in the 17th century. During this period, mathematicians, philosophers, and early scientists began developing new empirical and mathematical approaches to understand natural and social phenomena. Where previously such domains had been studied through qualitative reasoning or abstract models, quantitative analysis utilizing data and probability concepts now took root.

Three pivotal developments laid the foundation for these disciplines as we understand them today. The first was the establishment of probability theory as a mathematical subject area through the seminal correspondence between Blaise Pascal and Pierre de Fermat in 1654. In establishing clear definitions and analytical techniques for calculating probabilities, their work launched probability as a rigorous field of study.

Concurrently, John Graunt published his study of London mortality records titled Observations in 1662. This work is considered the origin of

modern demography and statistics, as Graunt systematically compiled and analyzed parish death data using crude early statistical techniques. His work demonstrated the value of a quantitative, empirical approach and helped establish population modeling and life tables.

Connecting these two threads was the development of probabilistic concepts of population dynamics. Christiaan Huygens and later Nicholas Bernoulli built on Graunt's work by interpreting his partial life table probabilistically. This work in the late 17th century established the theoretical framework for life tables based on probabilistic models of mortality rates.

Together, these developments tracing back to figures like Cardano, Pascal, Fermat, Graunt, Huygens and Bernoulli represented the interconnected foundational concepts of statistics, probability theory and population modeling that are still core to these fields today. Through increasingly mathematical and data-driven methods, scientists began developing powerful new analytical tools for understanding the natural and social world in quantitative terms.

6.1 Background

The intellectual context of the 17th century set the stage for the formative developments in probability, statistics and population modeling. Advances in mathematics, along with changing philosophies of science, fostered quantitative and empirical approaches to understanding the natural world.

Mathematicians of the early 1600s had formalized algebra, introducing symbolic notation that allowed solving generalized problems rather than specific numerical examples. This mathematical maturity enabled tackling more complex real-world problems. Concurrently, Francis Bacon and others advocated inductive, experiment-based scientific methods rather than overtly theoretical deduction alone.

Early proto-statistical techniques also emerged. Blood counts recorded by Belgian Lambert van Heemskerck in 1590 showed histograms emerging as a tool to visualize data distributions. Galileo's 1611 displacement experiments laid foundations for experimental design and estimation of measurement error. Dice problems studied mathematically by Cardano in his 1556 work De Ludo Aleae implicitly demonstrated probabilistic concepts, though a rigorous theory was lacking.

In natural philosophy, Aristotelian teleological thought remained dominant yet faced challenges from the mechanistic views of Descartes. John Wallis' De Algebra of 1693 signaled a shift towards mathematical representations. Investigation of celestial phenomena like comets, sunspots and nova by Kepler, Brahe and Galileo revealed complexity rather than the astronomically "perfect" former worldview.

This context set the stage for progress. Developing numerical techniques and willingness to quantify real-world phenomena opened the doors for ap-

plying mathematics to problems like games, mortality and astronomy. Empirical natural knowledge supplanted overtly theoretical tradition, recognizing phenomena like errors and variation fundamental to scientific understanding. Challenges to established worldviews enabled new probabilistic and quantifying perspectives on inherently complex, divers phenomena. This lay groundwork for the probability, statistics and life tables emerging in the mid-1600s traced in subsequent sections.

6.2 Cardano's work

One of the earliest works to apply mathematical thinking to games of chance was Girolamo Cardano's 1526 treatise De Ludo Aleae (On Games of Chance). As a physician and polymath deeply engaged in mathematics, Cardano sought to determine equitable rules and stake structures for various gambling activities through probabilistic reasoning. Though he lacked a theoretical framework, Cardano's work marked a pioneering effort in what would later become probability theory.

In De Ludo Aleae, Cardano analyzed dice, cards and various board games that involved elements of chance. For each, he aimed to calculate the mathematical expectation of each potential outcome in order to establish fair stakes. For example, in analyzing common dice games, Cardano outlined the relative probabilities of rolling different numbers and their associated payouts. However, his arithmetic calculations contained errors and he did not consistently apply proportional reasoning to determine equitable outcomes.

A problematic endeavor Cardano explored was the "problem of points;" how winnings should be divided between two players if an incomplete game of chance was stopped before a winner was determined.. While proposing solutions based on proportional reasoning, Cardano's logic contained logical gaps. For a game half-completed, he suggested stakes be split evenly rather than considering the changing probabilities as the game progressed. This flawed proportional method would continue to plague proposed solutions for some time.

Despite analytical limitations from lacking a rigorous probability system, De Ludo Aleae marked an intellectual milestone as one of the earliest works to systematically apply mathematical and quantitative thinking to games involving chance. It helped establish chance and probability as worthwhile subjects of mathematical inquiry. While Cardano's specific solutions were imperfect, his work helped catalyze further development and paved the way for the eventual theoretical foundations later established by Pascal, Fermat, and other 17th century thinkers. De Ludo Aleae highlighted the value of quantitative analysis and paved the way for probability's maturation.

6.3 Pascal and Fermat

The birth of probability theory as a rigorous mathematical discipline owes to the seminal correspondence between the French mathematician Blaise Pascal and French lawyer Pierre de Fermat in 1654. Seeking to solve the unresolved "problem of points," their letters established fundamental concepts and analytical techniques that formed the basis of probability as a field of study.

Pascal introduced the groundbreaking idea of mathematical expectation, which allowed calculating the average probability-weighted payoff of each possible outcome. Through examining games with varying numbers of trials like rolling dice, Pascal discovered patterns in the binomial coefficients that became known as Pascal's triangle. This arrangement reveals the coefficients of binomial expansions and allowed calculating probabilities and expectations through what is known as the binomial distribution.

Fermat independently devised an analytical approach using proportional reasoning about remaining possibilities. Through comparing solutions, Pascal and Fermat established that their methods agreed, validating probability could be understood mathematically. They then applied these to systematically solve variations of the problem of points.

Significantly, Pascal provided clear definitions of probability, expectation, and equity in games of chance. This marked the establishment of probability theory as an organized field based on precise definitions and analytical calculations rather than intuition. Their breakthrough correspondence founded the modern mathematical understanding of probability that remains fundamental to statistics, risk analysis and related disciplines. While further refined by later thinkers, Pascal and Fermat created the indispensable theoretical underpinnings.

6.4 Huygens' treatise

Building upon the foundational work of Pascal and Fermat, Christiaan Huygens' 1657 treatise De Ratiociniis in Ludo Aleae marked an important step forward in developing probability theory. Through elegantly solving problems involving games of chance, Huygens refined key concepts and helped disseminate probabilistic reasoning to a broader intellectual audience.

Huygens presented five problems to demonstrate logical solutions using the mathematical expectation and probability techniques established earlier. His treatment expanded the scope beyond Pascal's original problem of points by considering scenarios like dividing stakes between players leaving ongoing games and determining equitable initial and ongoing stakes.

What distinguished Huygens' approach was his employment of geometric representations to arrive at solutions. By visualizing probabilities and expected values geometrically, he provided intuitive demonstrations accessible

to contemporary mathematicians and scholars. Problems involving compound probability or independent events were depicted especially clearly through his innovative geometric approach.

Beyond pragmatic applications, Huygens' treatise introduced theoretical advances. He properly considered cases involving more complex stochastic processes with multiple dependent variables. The concept of independence between events was also more rigorously defined.

Huygens helped cement probability theory as a legitimate field meriting dedicated study. By refining concepts through elegant visual proof, he disseminated probabilistic reasoning to a broader intellectual community. His treatise engaged fellow mathematical thinkers while communicating foundations developed by Pascal and Fermat to a general educated readership. It served to further legitimize and strengthen the young discipline of probability.

6.5 Graunt's Bills of Mortality

While Continental mathematicians formalized probability theory, in England John Graunt published his seminal work Observations Made upon the Bills of Mortality in 1662. By systematically compiling and analyzing parish mortality records from London, Graunt is considered the founder of modern empirical demography and statistics.

Graunt studied weekly Bills of Mortality listing burials in each parish from 1603-1660. He assessed data quality, discussed trends in mortality from different causes, and noted seasonal patterns. By comparing burial totals to estimated births, Graunt proposed London's total population was around 460,000.

He constructed what is considered the first life table by calculating survivorship ratios at each age based on proportional mortality. While crude, this provided insights into longevity and helped estimate life expectancies. Graunt also tested hypotheses, finding for example that mortality was lower when more births occurred.

Through his careful collection and assessment of real-world evidence, Graunt established demography as an empirical discipline. Crude analytical techniques like comparing proportions laid foundations for later statistical methods. Unlike earlier works grounded in theory, Graunt emphasized pragmatic, data-driven analysis and population modeling.

Observations represented an intellectual departure by quantitatively interrogating social phenomena. It influenced later demographers and materially aided statecraft through providing population estimates. Most significantly, Graunt established mortality records and life tables as foundations for future probabilistic population modeling developed by Huygens, Bernoulli and later thinkers. His work originated empirical social science inquiry through numbers.

6.6 Probabilistic life tables and early insurance mathematics

Building upon Graunt's empirical establishment of life tables, later 17th century mathematicians developed probabilistic interpretations that represented important conceptual advances. Correspondence between Constantijn and Christiaan Huygens in 1669 began interpreting life tables through expected value calculations, laying early foundations.

However, the first systematic probabilistic construction and interpretation of Graunt's partial life table emerged in Nicholas Bernoulli's 1709 dissertation for the University of Basel. Seeing Graunt's observations as samples from an imagined infinite population, Bernoulli established the probability of survival and death at each age based on frequencies in the dataset.

This conceptualization described what is now known as a stationary life table based on regular and consistent mortality probabilities at each age level. Bernoulli fitted Graunt's London figures to his theoretical model to derive survival probabilities and life expectancies, demonstrating how pragmatic demographic records could inform probabilistic population modeling.

Concurrently, nascent life insurance and annuity practices drove interest in mathematically calculating fair premiums and payout values. Jan de Witt evaluated annuity pricing in 1671 based on Dutch life tables, though not fully probabilistically. Edmond Halley's 1693 paper introduced what became known as Halley's life table and outlined uses including insurance pricing, suggesting age-specific rates.

While emerging pragmatically, these developments established the analytical theoretical framework treating populations and longevity in probabilistic terms. Refined life tables coupled with probability concepts formed the mathematical underpinnings for later insurance and social policy applications informed by demographic patterns. They launched systematic population modeling central to modern actuarial science and epidemiology.

6.7 Discussion

By the late 17th century, the conceptual foundations for statistics, probability theory, and population modeling as quantitative fields of study had begun to solidify. Individuals including Cardano, Pascal, Graunt, Huygens, Bernoulli and others made seminal contributions developing mathematical and empirical approaches for analyzing phenomena involving chance, data, and human populations.

Pascal and Fermat established probability theory on a rigorous theoretical basis through analytical techniques like mathematical expectation. Graunt originated modern empirical demography through systematically compiling and analyzing real-world mortality data. Connecting these lines of inquiry,

later mathematicians developed probabilistic interpretations of demographic phenomena like life tables.

Throughout this period, there was a discernible shift towards more quantitative, data-driven methods of understanding the natural and social worlds. Inspired by early empirical pioneers, scientists increasingly used numerical evidence and mathematical representations to model irregular and complex systems. They developed analytical tools to estimate parameters, assess uncertainties, and test hypotheses rigorously against observations.

By the late 1600s, the theoretical and empirical foundations were largely in place for statistics, probability, and population modeling as quantitative fields. Figures like Pascal, Graunt, and Bernoulli established key concepts involving mathematical modeling of variation, uncertainty, and human demography using data-driven methods. While continuing to evolve significantly, the origins of these disciplines can be traced to this era when early thinkers began placing such topics on a rigorous analytical basis.

Even today, statistics, probability, and population modeling remain interconnected and influence each other's ongoing progress. Advancements in one field often spur developments in related areas. For example, new statistical techniques enable analysts to more accurately model population dynamics when considering demographic uncertainties or changes over time. Similarly, probabilistic methods allow epidemiologists and other scientists to better estimate public health risks and design evidence-based policies. The significant conceptual links established in their 17th century beginnings continue shaping the nature of inquiry across these quantitative fields.

Chapter 7
Logistic Model

Abstract This chapter presents the logistic growth model, a seminal population dynamics model that incorporates density-dependent negative feedback on growth rates. After historically introducing population regulation concepts from Malthus and Verhulst's development of the first quantitative model, we derive the logistic differential equation and its closed-form solution. The model exhibits long-term convergence to a stable carrying capacity regardless of initial conditions. Through local stability analysis of steady states, analytical and graphical approaches confirm the model predicts sigmoidal growth trajectories for starts below half carrying capacity, and monotonic convergence above it. While density-dependence captures universal population constraints, the simplicity of the logistic equation demonstrates how fundamental processes can form the basis for mathematical descriptions reflecting real ecological dynamics over long timescales, making it one of the most widely applied population models.

While the exponential growth model provides a simple and mathematically tractable framework for studying populations over short time periods, it has significant limitations when extrapolated over long timescales. Most notably, the model predicts unconstrained population expansion leading to infinite size, which is not feasible in reality. All natural populations are bound by finite resources in their environment. The exponential model also ignores factors like resource competition and carrying capacity effects that act to slow growth as the population approaches the limits of its habitat. As a result, the model cannot realistically depict long-term population behavior or stability. For populations existing in a closed ecological system with density-dependent influences, another type of population model is needed that incorporates limitations on growth from resource limitations and crowding effects. This will allow an investigation of population regulation and equilibrium conditions. The development of such a model, which involves including density-dependent

feedback in the growth term, is necessary to more accurately represent population dynamics over long periods aligned with biological timescales.

7.1 A brief historical introduction

Thomas Malthus was one of the earliest scholars to study population dynamics. In his 1798 treatise "An Essay on the Principle of Population", Malthus hypothesized that human populations have an inherent capacity to grow exponentially if unchecked by outside forces. This is because populations can increase geometrically through reproduction. In contrast, he argued that the food resources required to support growing populations only increase in an arithmetic progression over time. As populations grow exponentially yet resources increase linearly, Malthus recognized that individuals would inevitably come into competition for limiting resources like food, space, and other necessities of life. This competition, he proposed, serves to slow population growth from its unchecked exponential rate down to a level sustainable given the available resources. Malthus' ideas introduced the concept that populations are regulated by external factors related to resource availability and by the scarcity and struggle for existence that arises due to overpopulation. His insights formed the basis for later formal mathematical models of population dynamics.

While Malthus recognized that populations cannot grow exponentially indefinitely, he did not provide a mathematical model to describe how growth becomes limited over time. It was the Belgian mathematician Pierre Verhulst who first developed an explicit equation incorporating this regulation mechanism. In 1838, Verhulst modified the simple exponential growth model based on Malthus' arguments about competition for resources. He proposed adding a negative feedback term that increases in proportion to the population size, representing the growing scarcity and resultant effect on per capita growth rate as the population approaches the environment's carrying capacity.

7.2 The logistic model

We begin by recalling the differential equation that governs exponential population growth:

$$\frac{dN}{dt} = \alpha N,$$

where α is a constant growth rate parameter. Verhulst proposed modifying this model by substituting parameter α with a decreasing function of the population size N, denoted $\alpha(N)$. Specifically, he proposed a linear decay function of the form:

$$\alpha(N) = r\left(1 - \frac{N}{K}\right),$$

where r represents the intrinsic per capita growth rate in the absence of density-dependent effects. Meanwhile, K denotes the environment's biological carrying capacity; the maximum sustainable population size given resource constraints. This negative feedback term grows proportionally as N approaches K, introducing density-dependent regulation into the exponential growth formulation. Verhulst's ingenious substitution of a variable growth rate function was a landmark development that produced the first quantitative model embodying Malthus' concept of self-limitation through resource competition.

Through this modification, Verhulst produced the following nonlinear ordinary differential equation to govern population dynamics:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right). \tag{7.1}$$

Verhulst named the solution to this differential equation the "logistic function". By extension, the overall model is commonly referred to as the "logistic model" of population growth. However, it is also appropriately termed the "Verhulst model

We can solve the logistic differential equation (Eq. 7.1) using variable transformations. First, introduce the change of variable $x = 1/N$. This transforms Equation 7.1 to:

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

Next, define the variable $y = x - 1/K$, yielding:

$$\frac{dy}{dt} = -ry.$$

This has the solution $y(t) = y_0 e^{-rt}$, where y_0 is the initial value. Transforming back through the steps, we arrive at the solution for the original variable N:

$$N(t) = \left[\frac{1}{K} + \left(\frac{1}{N_0} - \frac{1}{K}\right)e^{-rt}\right]^{-1}. \tag{7.2}$$

Through judicious variable substitutions, we have derived an explicit closed-form solution to the logistic differential equation governing population dynamics.

Equation (7.2) is the well known logistic function. Consider the characteristics of this solution. The first term inside the brackets, $1/K$, represents the asymptotic carrying capacity. As time increases, the exponential term e^{-rt} monotonically decays to zero. Therefore, in the limit of large time, the population N will approach the value:

$$\lim_{t\to\infty} N = K.$$

This shows that regardless of the initial population N_0, the solution will always converge to the environmental carrying capacity K. Physically, this makes sense; as time progresses, density-dependent effects will gradually dampen the growth rate until it balances perfectly with mortality to establish an equilibrium at the maximum sustainable size. The timescale to reach this steady state depends on the intrinsic growth rate r, but the ultimate attractor is always K. Therefore, the logistic model comprehensively captures the self-regulation of populations towards their local resource limits envisioned by Malthus.

7.3 Local stability of steady states

While we were able to derive an analytical solution to the logistic equation in the previous section, this is not always possible for nonlinear dynamical systems. Even when a closed-form solution cannot be obtained, important qualitative insights into dynamic behavior can still be gained by analyzing the system's steady states and their local stability properties. Though analytical solutions offer complete descriptions of trajectories over time, local stability analysis offers a general, mathematically rigorous technique to characterize global qualitative behaviors. In this section, we will use such an approach by finding steady states of the logistic model and applying linear stability theory to assess their stability.

To perform a local stability analysis, we first rewrite the logistic differential equation (Eq. 7.1) in the general form:

$$\frac{dN}{dt} = f(N),$$

where $f(N) = rN(1-N/K)$. The steady state solutions, also known as fixed points, correspond to values of N where $f(N) = 0$. One can readily show that the fixed points are $N_1^* = 0$ and $N_2^* = K$.

We next calculate the derivative of $f(N)$ to determine the slope at each fixed point:

$$f'(N) = r - \frac{2rN}{K}.$$

Evaluating f' gives: $f'(N_1^*) = r > 0$ and $f'(N_2^*) = -r < 0$. This indicates that perturbations will grow near $N_1^* = 0$, making it unstable. Meanwhile, near the carrying capacity $N_2^* = K$, perturbations will decay over time, signifying stability.

The previous analytic stability analysis can be visualized graphically. Figure (7.1) depicts the growth rate function $f(N) = rN(1-N/K)$. This illustrates

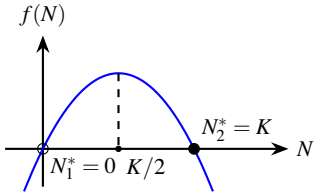


Fig. 7.1 Graphical analysis of fixed points and stability in the logistic growth model. This figure shows the growth rate function $f(N) = rN(1-N/K)$. The two roots of this function, where it intersects the horizontal axis, are the fixed points (steady state solutions): $N_1^* = 0$ and $N_2^* = K$. Local stability is determined by the slope of $f(N)$ at each fixed point. At $N_1^* = 0$, the slope is positive indicating nearby trajectories will diverge over time. Therefore, N_1^* denotes an unstable fixed point, represented by an open circle. In contrast, the slope of $f(N)$ is negative at the carrying capacity $N_2^* = K$. Thus, small perturbations will contract back towards the fixed point. This characterizes N_2^* as a stable steady state solution, shown as a filled circle.

the two fixed points where $f(N)$ intersects the horizontal axis, representing steady state solutions.

The fixed points occur at population values of $N_1^* = 0$ and $N_2^* = K$, coinciding with the steady state solutions found earlier. To assess local stability, we evaluate the slope of $f(N)$ at each fixed point. At $N_1^* = 0$, the positive slope means nearby trajectories will diverge over time. Therefore, N_1^* denotes an unstable node, shown as an open circle. In contrast, the negative slope at the carrying capacity $N_2^* = K$ implies small perturbations will decay back to this fixed point. This characterizes N_2^* as locally stable, represented by a filled circle.

The graphical analysis in Figure 7.1 provides additional qualitative insights into solution trajectories. Since the function $f(N)$ is increasing over the range $0 < N < K/2$, populations initialized below half the carrying capacity ($N_0 < K/2$) will experience accelerated growth as they approach this midpoint. However, as $f(N)$ is decreasing for $K/2 < N < \infty$, these solutions will then decelerate as diminishing resources induce density-dependent regulation toward the asymptote. Thus, trajectories beginning below $K/2$ exhibit the characteristic S-shaped sigmoid curve.

On the other hand, solutions starting above half the carrying capacity ($N_0 > K/2$) will converge directly to K in a qualitatively different manner, as $f(N)$ is negative over this interval. Their approach will be steadily dampened regardless of whether their initial value is below or above the stable steady state $N_2^* = K$. Therefore, by examining the increasing and decreasing intervals of $f(N)$, we gain insight into qualitatively distinct population growth dynamics depending on initial population size relative to the system's threshold midpoint value.

The qualitative insights gained from analyzing the growth rate function $f(N)$ can be validated by plotting the solutions of the logistic differential equation over time. Figure 7.2 plots the logistic function given by Equation

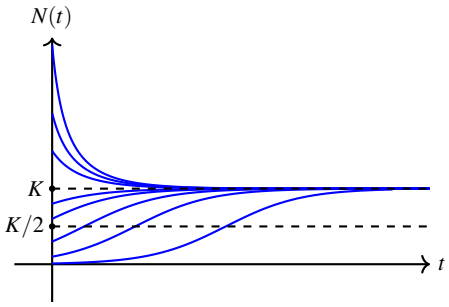


Fig. 7.2 Qualitative trajectories of the logistic solution for varying initial conditions. This figure graphically depicts the logistic function given by Equation (7.2) over time for different initial population sizes N_0. Solutions initialized below half the carrying capacity $K/2$ exhibit characteristic sigmoidal growth curves, accelerating as they approach the midpoint before slowing down due to density dependence. In contrast, populations starting above $K/2$ demonstrate qualitatively different monotonic convergence (without an inflection point) to the stable steady state at carrying capacity K, regardless of whether they start above or below the fixed point.

(7.2) for a range of initial population sizes N_0. This illustrates how initial conditions smaller than half the carrying capacity $K/2$ exhibit sigmoidal growth curves, consistent with the posited trajectory shapes based on the increasing and decreasing intervals of $f(N)$. Solutions initialized above $K/2$ also match expectations, showing direct monotonic convergence.

By comparing the predicted dynamics from analytic and graphical stability analysis (Figure 7.1) against analytic solutions in Figure 7.2, we find strong agreement between the local mathematical characterization near fixed points and the emergent global qualitative behaviors of the overall system across a spectrum of initial conditions. This corroborates that the linearization techniques employed to analyze steady states provide insights extending well beyond the local neighborhood of equilibria to inform our broader understanding of logistic population growth models.

7.4 Discussion

The logistic equation, despite its simple mathematical form, captures essential qualitative features of real-world population dynamics. The inclusion of nonlinear density-dependent feedback simulates the influence of finite resources and competition on growth. This confers carrying capacity behavior observed universally in populations, from microbial colonies to animal populations. Thus the logistic model has broad relevance in ecology, providing a starting point to understand outbreaks, booms, and stabilization in a variety of species. Its success stems from distilling dynamics to fundamental

processes without species-specific complexity. As one of the earliest and most widely employed models in mathematical biology, it demonstrates how even minimal representations can reflect key characteristics of living populations, especially those of unicellular organisms like bacteria exhibiting exponential growth followed by limited expansion.

While analytic solutions offer complete trajectories when attainable, local stability analysis provides a generalizable technique for characterizing nonlinear dynamical systems. By linearizing near fixed points, it allows classifying long-term qualitative fates of solutions based on mathematical stability criteria. As demonstrated here for the logistic equation, this approach correctly predicts global dynamic attributes from investigation of a system's local equilibrium properties. Even without closed-form solutions, local analysis thus remains a powerful approach for divining qualitative insights into complexity arising from nonlinear feedbacks governing real natural and engineered systems. Where full descriptions remain elusive, studying stability properties of steady states provides an alternative lens into qualitative behaviors that may emerge from complexity near points of balance.

Chapter 8
Extending the logistic model

In the previous chapter, we introduced the basic logistic population growth model and analyzed its dynamic behavior. This model assumes that density-dependent competition for resources limits exponential growth. However, more complex interactions can also shape population dynamics. Here we extend the logistic equation to incorporate two additional realistic effects: increased mortality and the Allee effect.

8.1 Increased mortality

Up to this point, our logistic population model has assumed a constant per capita death rate. However, environmental factors unrelated to population density, such as stressors or resource limitation, can elevate mortality. To incorporate this effect, we modify the logistic differential equation. Specifically, we add a term representing density-independent background mortality:

$$\frac{dN}{dt} = rN(1-N/K) - aN. \tag{8.1}$$

Here, the parameter $a \geq 0$ controls the strength of increased background mortality. When $a = 0$, this model reduces to the basic logistic equation by recovering the original constant per capita death rate assumption.

To perform a local stability analysis, we first rewrite the logistic differential equation (Eq. 8.1) in the general form:

$$\frac{dN}{dt} = f(N),$$

where $f(N) = rN(1-N/K) - aN$. The steady state solutions, also known as fixed points, are given by values of N that satisfy $f(N) = 0$. This yields two

fixed points: the extinction state $N_1^* = 0$, and the non-zero equilibrium $N_2^* = K(r-a)/r$.

To determine stability, we calculate the derivative of the function $f(N)$ to obtain the slope at each fixed point. The derivative is:

$$f'(N) = (r-a) - \frac{2rN}{K}.$$

Evaluating f' at the two fixed points gives: $f'(N_1^*) = r-a$ and $f'(N_2^*) = a-r$. This reveals that when $a < r$, the extinction state N_1^* is unstable while the positive equilibrium N_2^* is stable. However, when the mortality rate exceeds the intrinsic growth rate ($a > r$), the stability of the fixed points switches; now the extinction state is stable, whereas the non-zero state N_2^* becomes unstable.

Geometrically, we can gain insight into this transition by visualizing how the fixed points vary with the mortality rate parameter a, as illustrated in Figure 8.1. As a increases through the critical value of r, the two fixed points collide at the bifurcation point and trade stability. Topologically, this behavior corresponds to a transcritical bifurcation. Physically, it indicates that beyond a threshold level of background mortality, high death rates outweigh new births even at very low population sizes, thereby preventing the population from persisting over long-term. The bifurcation analysis thus provides a framework for understanding how environmental stresses influence population viability through changes in demographic rates.

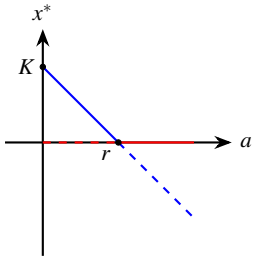


Fig. 8.1 Bifurcation diagram showing how the two steady state solutions, population extinction (N_1^*) and the positive equilibrium population size (N_2^*), vary dynamically with changes to the background mortality rate parameter (a). The extinction fixed point is depicted with red lines, while the positive equilibrium is shown in blue. Stable steady states are depicted with solid lines, and unstable states with dashed lines. As a increases beyond the critical threshold value r, equivalent to the intrinsic growth rate, the fixed points intersect at the transcritical bifurcation point. This bifurcation signifies the exchange of stability between the two solutions, where population persistence loses stability and extinction gains stability with further increases in environmental mortality.

This model exhibits the first bifurcation encountered in this textbook. It is therefore worthwhile to discuss bifurcations in more detail. By definition,

a bifurcation refers to a qualitative change in dynamical behavior that can emerge from small perturbations to parameter values. Specifically, at a bifurcation point the number and/or stability properties of fixed point solutions may abruptly shift. Some common types of bifurcations include:

- Saddle-node bifurcation: this occurs when two fixed points coalesce and annihilate one another.
- Transcritical bifurcation: a stable and unstable fixed point exchange stability at the critical parameter value.
- Pitchfork bifurcation: a single fixed point loses stability, splitting into three solutions of which two are stable.

Bifurcations are significant because they delineate boundaries between distinct qualitative solution behaviors attainable within a model. Identifying these transition points provides insight into how sensitive model predictions are to perturbations of the governing equations.

In the modified logistic population model presented here, the bifurcation that occurs with rising background mortality is of the transcritical type. This bifurcation demarcates how increasing non-density dependent death rates can irreversibly drive the population to extinction once they surpass the capacity for growth. Specifically, as mortality rises past the critical threshold equal to the intrinsic per capita growth rate, the transcritical bifurcation causes the stable fixed point solution representing long-term population persistence to exchange stability with the unstable extinction point. Consequently, above this bifurcation value, mortality dominates reproduction even at very low densities throughout the entire parameter space. This example demonstrates that environmental stressors can precipitate a population tipping point between ongoing viability versus extinction by tilting the demographic balance through small alterations to basic rates of birth and death. Identifying such bifurcation-defined thresholds provides insight into how susceptible population persistence is to fluctuations in underlying biological parameters.

8.2 Allee Effect

Small populations face additional hurdles that threaten their long-term sustainability. This phenomenon, known as the Allee effect, arises due to difficulties finding mates and engaging in beneficial forms of cooperation at very low densities, resulting in decreased per capita growth. To represent this intrinsic difficulty in the logistic model, we introduce an Allee term as follows:

$$\frac{dN}{dt} = rN(1-N/K)(\frac{N}{A}-1). \tag{8.2}$$

In this formulation, parameter $0 < A < K/2$ represents a population threshold below which the effective per capita growth rate becomes negative, meaning that deaths exceed births.

Let us analyze Equation 8.2 in its general functional form, $\dot{N} = f(N)$, where $f(N) = rN(1 - N/K)(N/A - 1)$. The form of this function $f(N)$ is depicted graphically in Figure 8.2. As seen in the figure, $f(N)$ intersects the horizontal axis at three distinct points, indicating the presence of three steady state solutions: population extinction at $N_1^* = 0$, an interior equilibrium at $N_2^* = A$, and a second fixed point at the carrying capacity $N_3^* = K$. Furthermore, the derivatives $f'(N_1^*) < 0$, $f'(N_2^*) > 0$, and $f'(N_3^*) < 0$. This confirms that N_1^* and N_3^* are stable nodes, while N_2^* is an unstable saddle point.

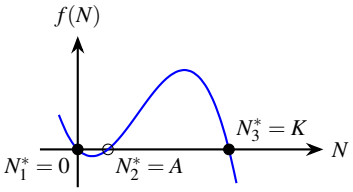


Fig. 8.2 Graphical depiction of the function $f(N) = rN(1-N/K)(N/A-1)$ describing per capita growth rate in the Allee effect model (Equation 8.2). The function intersects the horizontal axis at three fixed points, indicating three steady state solutions: population extinction at $N_1^= 0$, an interior unstable equilibrium at $N_2^= A$, and the second interior stable fixed point at carrying capacity $N_3^= K$. The first and third fixed points are stable nodes whereas the interior point is an unstable saddle, as shown by the signs of the derivatives $f'(N_1^) < 0$, $f'(N_2^) > 0$, and $f'(N_3^) < 0$.

The existence of two locally stable fixed points separated by an unstable steady state in this modified logistic model, namely the extinction state N_1^* and the positive carrying capacity N_3^*, gives rise to bistability in the system's dynamics. Biologically, bistability implies that the population can persist indefinitely at either stable state of extinction or carrying capacity, with the final outcome dependent on initial conditions. Stochastic fluctuations may cause the population to randomly switch between attractors.

The logistic equation exhibiting an Allee effect is the first example in this textbook to demonstrate bistability. As a key feature of nonlinear systems, bistability will emerge again in other examples throughout this textbook, continually deepening our understanding of its consequences for real-world ecological and biological systems.

8.3 Discussion

In this chapter, we explored two extensions to the basic logistic population growth model: increased background mortality and an Allee effect. By incorporating density-independent mortality through an additional death term, the model exhibited a transcritical bifurcation whereby changes to the mortality rate could precipitate a population tipping point between ongoing viability and extinction. This demonstrated how environmental stresses may trigger abrupt, irreversible transitions in population state through subtle parameter changes.

Incorporating an Allee effect produced bistability, in which the population could persist indefinitely at either the extinction or carrying capacity stable states depending on initial conditions. Bistability highlights the sensitivity of ecological systems to random perturbations, with the potential for sudden shifts between alternative community states. It is a nonlinear phenomenon with significant implications for population persistence and community resilience.

More broadly, this chapter demonstrated the utility of dynamical systems theory and tools like bifurcation analysis for gaining mechanistic understanding of population viability. Such approaches aim to tease apart how multiplicative interactions between intrinsic biological processes and density-dependent feedbacks shape complex, often counterintuitive population dynamics over time.

Chapter 9
Local Stability Analysis in 2D Systems

Abstract This chapter develops the technique of local stability analysis for two-dimensional nonlinear dynamical systems modeled by sets of coupled ordinary differential equations. Steady states, or fixed points, where both equations are simultaneously satisfied, are classified into nodes, saddles, and spirals by analyzing the eigenvalue structure of the 2×2 Jacobian matrix computed from the system's partial derivatives near each fixed point. Analytical criteria are derived requiring both eigenvalues to have negative real parts for stability, reducing determination to the Jacobian's trace and determinant. Qualitative analysis of eigenvalue configurations provides insight into global trajectory behavior near different steady state types, equipping nonlinear dynamical modeling with a rigorous yet accessible mathematical approach for stability-based analysis. The methodology introduced generalizes directly to higher dimensions and serves as a foundation for stability-based modeling across sciences involving ordinary differential equations.

In Chapter 5, we developed the method of local stability analysis to characterize the behavior of steady states in one-dimensional nonlinear dynamical systems. This approach reduces the stability criterion to checking the sign of the slope of the governing function at the steady state. However, many natural and engineered systems are described by models with two or more dimensions that cannot be intuitively visualized or analyzed using one-dimensional techniques. It is therefore necessary to generalize the technique of local stability analysis to higher-dimensional systems described by sets of coupled ordinary differential equations.

In this chapter, we extend the methodology of local stability analysis to two-dimensional nonlinear dynamical systems represented by a pair of coupled ordinary differential equations. While graphical analysis is no longer suitable, the algebraic framework of linearization remains valid. We will see that steady states are now defined as simultaneous solutions to two equations.

Linearizing the system dynamics locally yields a 2×2 Jacobian matrix governing small perturbations near the steady state. The matrix's eigenstructure determines stability and allows a qualitative classification.

By deriving analytical stability criteria from the Jacobian eigenvalues, this approach provides insights into even relatively simple 2D models that are inaccessible through low-dimensional intuition alone. The methodology generalizes directly to higher dimensions through matrix analysis. Local stability analysis thus equips the study of multidimensional dynamical models with a systematic and mathematically rigorous technique.

9.1 Steady States in 2D Systems

We begin by considering a general two-dimensional dynamical system governed by a pair of first-order ordinary differential equations:

$$\frac{dx}{dt} = f(x,y), \tag{9.1}$$

$$\frac{dy}{dt} = g(x,y), \tag{9.2}$$

where $f(x,y)$ and $g(x,y)$ are smooth functions determining the rate of change of x and y with respect to time t.

In the one-dimensional case studied previously, a steady state or fixed point was defined as a location x^* where the system's behavior does not change over time. To generalize this concept to two dimensions, we require the system to be stationary at a particular point (x^*,y^*) in the x-y plane. Mathematically, a fixed point (x^*y^*) satisfies::

$$\left.\frac{dx}{dt}\right|_{(x^*,y^*)} = 0, \quad \left.\frac{dy}{dt}\right|_{(x^*,y^*)} = 0.$$

Substituting the definitions of dx/dt and dy/dt yields:

$$f(x^*,y^*) = 0, \quad g(x^*,y^*) = 0.$$

Therefore, steady states or fixed points in two-dimensional systems are points (x^*,y^*) that are simultaneous solutions to the equations $f(x,y) = 0$ and $g(x,y) = 0$.

9.2 Local Linearization and Stability Analysis

To analyze the stability of a fixed point (x^*, y^*), we study how small perturbations from this point evolve over time. Let $\delta x = x - x^*$ and $\delta y = y - y^*$ represent deviations of an arbitrary trajectory from the fixed point coordinates. By definition, these perturbations satisfy:

$$\frac{d\delta x}{dt} = \frac{dx}{dt}, \quad \frac{d\delta y}{dt} = \frac{dy}{dt}.$$

Substituting the ODE definitions yields:

$$\begin{aligned}\frac{d\delta x}{dt} &= f(x^* + \delta x, y^* + \delta y),\\ \frac{d\delta y}{dt} &= g(x^* + \delta x, y^* + \delta y).\end{aligned}$$

Performing a Taylor series approximation truncated to the linear term renders:

$$\begin{aligned}\frac{d\delta x}{dt} &= f_x \delta x + f_y \delta y,\\ \frac{d\delta y}{dt} &= g_x \delta x + g_y \delta y\end{aligned}$$

where the partial derivatives are evaluated at the fixed point:

$$f_x = \left.\frac{\partial f}{\partial x}\right|_{(x^*,y^*)}, \quad f_y = \left.\frac{\partial f}{\partial y}\right|_{(x^*,y^*)}, g_x = \left.\frac{\partial g}{\partial x}\right|_{(x^*,y^*)}, \quad g_y = \left.\frac{\partial g}{\partial y}\right|_{(x^*,y^*)}.$$

These linearized equations can be written compactly in matrix form as:

$$\frac{d\delta \mathbf{r}}{dt} = \mathbf{J}\delta \mathbf{r}, \tag{9.3}$$

in which the Jacobian matrix $\mathbf{J}$ contains the partial derivatives, and $\delta\mathbf{r}$ is the perturbation vector.

$$\mathbf{J} = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}, \quad \delta\mathbf{r} = \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}.$$

Based on the one-dimensional analysis, we expect perturbations to evolve exponentially near the fixed point. Specifically, we make an ansatz that the solution takes the form:

$$\delta \mathbf{r} = \mathbf{v} e^{\lambda t}, \tag{9.4}$$

where the vector $\mathbf{v}$ and scalar parameter λ remain to be determined.

Substitution of Equation (9.4) into Equation (9.3) yields

$$\lambda \mathbf{v} = \mathbf{J}\mathbf{v}. \tag{9.5}$$

This reveals that for the ansatz to be a valid solution, the parameters λ and $\mathbf{v}$ must satisfy the eigenstructure of the Jacobian matrix $\mathbf{J}$. Specifically, λ must be an eigenvalue of $\mathbf{J}$ and $\mathbf{v}$ the corresponding eigenvector.

Since a 2x2 matrix like the Jacobian will generally have two eigenpairs, the complete solution is a superposition of the exponential modes:

$$\delta\mathbf{r} = C_1\mathbf{v_1}e^{\lambda t_1} + C_2\mathbf{v_2}e^{\lambda t_2}, \tag{9.6}$$

with $(\lambda_1, \mathbf{v_1})$ and $(\lambda_2, \mathbf{v_2})$ the eigenpairs of $\mathbf{J}$, while constants C_1 and C_2 are determined by the initial conditions. We can deduce the asymptotic behavior from Equation (9.6): $\delta\mathbf{r}$ decays to zero as $t \to \infty$ if both eigenvalues have negative real parts. In this case, the fixed point is declared locally stable.

9.3 Steady-state classification

The eigenstructure relation in Equation (9.5) can be expressed as:

$$(\mathbf{J} - \lambda\mathbf{I})\mathbf{v} = 0,$$

where $\mathbf{I}$ is the 2D identity matrix. This represents a system of two homogeneous linear equations with the components of $\mathbf{v}$ as unknowns. Generally, the only solution to such a system is the trivial solution $\mathbf{v} = \mathbf{0}$. However, this trivial solution implies $\delta\mathbf{r}(t) = \mathbf{0}$, meaning the considered solution is already at the fixed point. This prevents analysis of the stability. A non-trivial solution exists only when the determinant of the coefficient matrix vanishes, i.e. when:

$$\det(\mathbf{J} - \lambda\mathbf{I}) = 0.$$

Expanding this determinant condition gives:

$$(f_x - \lambda)(g_y - \lambda) - f_y g_x = 0,$$

Which can be written compactly in terms of the trace $\tau = f_x + g_y$ and determinant $\Delta = f_x g_y - f_y g_x$ of the Jacobian matrix $\mathbf{J}$ as:

$$\lambda^2 - \tau\lambda + \Delta = 0. \tag{9.7}$$

Equation (9.7) takes the form of a characteristic quadratic polynomial, whose roots correspond to the eigenvalues of the Jacobian matrix $\mathbf{J}$. The eigenvalues, $\lambda_{1,2}$, can be explicitly computed by solving the characteristic polynomial:

$$\lambda_{1,2} = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2}.$$

An analysis of this equation reveals key properties of the eigenvalues based on the values of the trace and determinant. Specifically:

- If the determinant $\Delta < 0$ (region I in Figure 9.1), then one eigenvalue will be positive while the other is negative, regardless of the sign of the trace τ.
- If $\Delta > 0$ and the discriminant $\tau^2 > 4\Delta$ (regions II and V in Figure 9.1), then both eigenvalues will be real and have the same sign as the trace τ.
- If $\Delta > 0$ but the discriminant $\tau^2 < 4\Delta$, then the eigenvalues form a complex conjugate pair in which τ is the real part (regions III and IV in Figure 9.1). Notably, when the trace $\tau = 0$, the solutions eigenvalues are purely imaginary numbers.

The above analysis has delineated how the values of coefficients Δ and τ determine the possible configurations of the eigenvalues. We now aim to connect these eigenvalue configurations to criteria for classifying the stability of the fixed point.

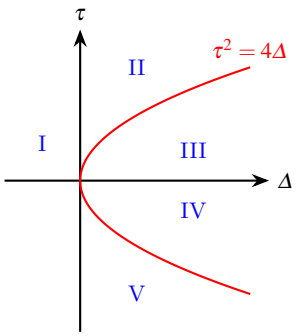
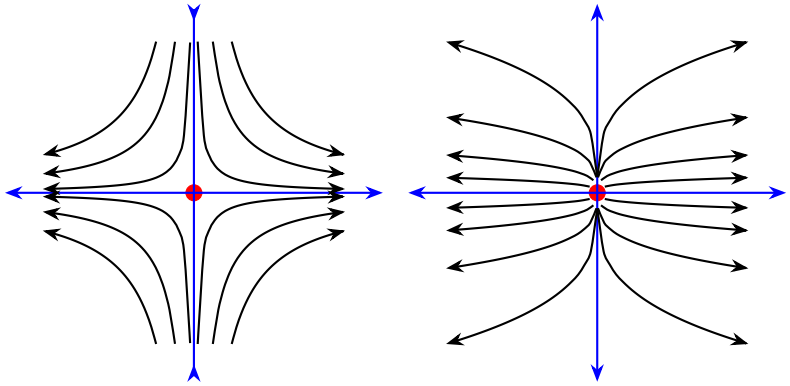


Fig. 9.1 Caption

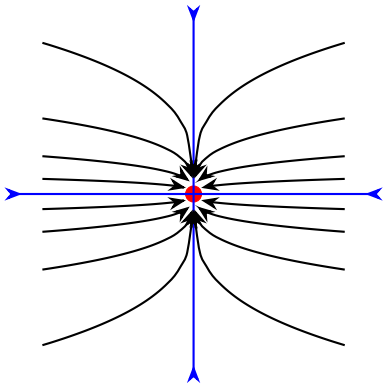
Recall that the general solution of the linearized system is given by Equation (9.6). When the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are unitary and linearly independent, they form a basis of the δx-δy space, and the terms $C_1 e^{\lambda_1 t}$ and $C_2 e^{\lambda_2 t}$ can be interpreted as the coordinates of the perturbation $\delta\mathbf{r}(t)$ at time t in this eigenbasis.

Considering the full solution as $\mathbf{r}(t) = \mathbf{r}^* + \delta\mathbf{r}(t)$, we see that $C_1 e^{\lambda_1 t}$ and $C_2 e^{\lambda_2 t}$ determine how the linearized solution $\mathbf{r}(t)$ evolves in a reference frame defined by the Jacobian eigenvectors and anchored at the fixed point $\mathbf{r}^*$.

When the eigenvalues are real, the eigenbasis coordinates $C_1 e^{\lambda_1 t}$ and $C_2 e^{\lambda_2 t}$ will exponentially decay or grow over time depending on the sign of the respective eigenvalues. A positive eigenvalue leads to exponential growth of the corresponding coordinate, while a negative eigenvalue implies exponential decay. This characterization of the perturbation growth informs the trajectory sketches shown in Figure 9.2 for regions I, II and V of Figure 9.1.



(a) Region I (b) Region II



(c) Region V

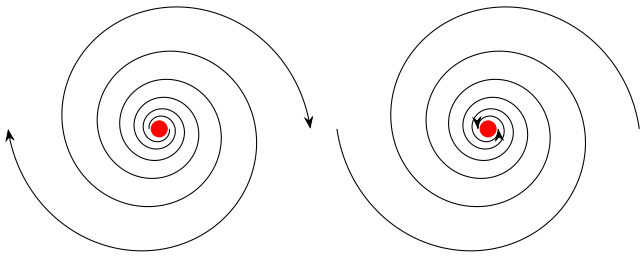
Fig. 9.2 Illustration of phase portraits for different eigenvalue configurations near a fixed point. (a) Region I of Figure 9.1 exhibits saddle node behavior, with trajectories flowing away along one direction and toward the fixed point along the orthogonal direction. (b) Region II displays an unstable node, where all trajectories exponentially diverge from the fixed point. (c) Region V corresponds to a stable node, with converging trajectories.

The steady state in region I, where one eigenvalue is positive and one is negative, corresponds to a saddle node. Near a saddle node, trajectories will flow away in one direction and towards the fixed point in the orthogonal direction, as illustrated in 9.2a. A steady state in region II, where both eigenvalues are positive, is referred to as an unstable node. In this case, small perturbations will always grow exponentially, leading trajectories to flow outward from the fixed point in all directions, as shown in Figure 9.2b. In region V, where both eigenvalues are negative, the steady state is said to be a stable node. Here, perturbations decay exponentially fast, causing trajectories to converge onto the fixed point from all surrounding points in phase space, as depicted in Figure 9.2c.

When the eigenvalues are complex, the linearized system possesses oscillatory behavior near the fixed point. For regions III and IV in Figure 9.1, where the trace τ and determinant Δ satisfy $\tau^2 < 4\Delta$, the eigenvalues form a complex conjugate pair $\lambda_{1,2} = \alpha \pm i\beta$. Substituting this into the general solution of the linearized system in Equation (9.6) yields:

$$\delta\mathbf{r}(t) = C_1\mathbf{v}_1 e^{\alpha t}\cos(\beta t) + C_2\mathbf{v}_2 e^{\alpha t}\sin(\beta t)$$

Observe that the perturbation undergoes rotational motion, with an exponential decay or growth modulated by the real part α of the complex eigenvalues. Since $\alpha = \tau/2$, where τ is the trace of the Jacobian matrix, the sign of τ dictates whether the oscillations decay or grow exponentially over time. Specifically, when $\tau > 0$ the real part is positive ($\alpha > 0$), corresponding to region III of Figure 9.1 where perturbations spiral outward from the fixed point with exponential grow. Conversely, for $\tau < 0$ the real part is negative ($\alpha < 0$), describing region IV where oscillations spiral inward to the fixed point in a decaying manner. These behaviors are illustrated in Figure 9.3.



(a) Region III (b) Region IV

Fig. 9.3 Phase portraits for regions III and IV displaying spiral node behavior near the fixed point. (a) In region III of Figure 9.1, where the trace $\tau < 0$, perturbations decay along a clockwise spiral trajectory towards the fixed point. (b) For region IV with $\tau > 0$, perturbations grow along a counterclockwise spiral, diverging from the fixed point. The real and imaginary parts of the complex eigenvalues determine the exponential decay/growth and rotational motion of perturbations in these cases.

9.4 Discussion

In this chapter, we have presented a framework for classifying the behavior near steady state solutions of 2D dynamical systems using linear stability analysis. By analyzing the eigenvalues of the Jacobian matrix ($\mathbf{J}$) evaluated at a fixed point ($\mathbf{r}^*$), we can gain qualitative insight into how perturbations will evolve in the local vicinity of that steady state.

There are three main categories of steady state behavior that emerge based on the signs and types of eigenvalues:

1. Node: All real eigenvalues lead to exponential decay or growth along principal directions, yielding stable or unstable equilibria.
2. Saddle: Mixed stability with one growing and one decaying eigenvalue mode. Trajectories flow away along one direction and towards the fixed point along the other.
3. Spiral: A complex conjugate eigenvalue pair yields stable or unstable rotational motion around the fixed point, with exponential modulation by the real part of the pair.

The stability of nodes and spirals is determined by the sign of the real part of the eigenvalues.

The classifications provided by this local linear stability analysis framework offer powerful insights into the full nonlinear dynamics of the system. Specifically, the local dynamic behavior near steady states allows for a qualitative description of the global behavior of trajectories throughout the phase space. When combined with numerical techniques, this framework enables comprehensive characterization of system dynamics. In the following chapters, we will explore detailed examples that demonstrate these concepts.

Chapter 10
Competitive Lotka-Volterra model

Abstract This chapter presents and analyzes the Lotka-Volterra competitive modeling framework, which describes the population dynamics of two competing species through coupled logistic growth differential equations that represent per capita growth rates being reduced by interspecific competition. The chapter examines equilibrium states, performs phase plane analysis to determine long-term trajectories defined by different competition coefficient regimes, and analyzes stability of equilibria to relate competitive hierarchies to community outcomes. Despite simplifying assumptions, the model provides explanatory insights into patterns conforming to Gause's competitive exclusion principle by conceptually formulating interspecific interactions through basic growth and competition functions. These foundational results demonstrate how even simple conceptual theory can generate ecological understanding of community assembly when aimed at elucidating underlying mechanisms rather than precise prediction.

10.1 Introduction

Intraspecific competition describes interactions between individuals of the same species competing for limited resources. In contrast, interspecific competition characterizes interactions between distinct but ecologically similar species that utilize overlapping resource pools. Past chapters have addressed intraspecific competition mathematically through the logistic model of population growth regulation.

This chapter focuses on interspecific competition, which occurs when multiple species utilize the same limited resources in a single area. As populations of coexisting species continuously interact through competition for necessities

like food and shelter, their relative abundances shift in complex ways over time. Describing these population dynamics mathematically provides insights into how competition shapes community structure and long-term species coexistence.

We explore the competitive Lotka-Volterra model, one of the earliest and most influential models developed to study interspecific competition. First introduced independently by Lotka in 1925 and Volterra in 1926, this seminal model formulated competition between species using coupled logistic growth functions. Over a century later, the Lotka-Volterra model remains fundamental to understanding phenomena like competitive exclusion and the conditions permitting diverse communities to coexist simultaneously.

10.2 The Competitive Lotka-Volterra Model

Lotka and Volterra developed one of the first mathematical formulations of interspecific competition by extending the logistic growth model to multiple interacting species. Their competitive Lotka-Volterra model describes the population dynamics of two species, which we will refer to as species 1 and 2, competing for a shared limiting resource.

The model assumes:

- Each species exhibits logistic population growth in the absence of the other.
- Competition arises through each species reducing the per capita growth rate of the other.

Based on these assumptions, the governing equations are:

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - a_{12}\frac{N_2}{K_2}\right), \tag{10.1}$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2}{K_2} - a_{21}\frac{N_1}{K_1}\right), \tag{10.2}$$

where:

- N_1 and N_2 are the population sizes of species 1 and 2.
- r_1 and r_2 are the intrinsic growth rates.
- K_1 and K_2 are the carrying capacities.
- a_{12} and a_{21} are the interspecific competition coefficients.

This pair of coupled nonlinear ordinary differential equations represent the simplest mathematical model capturing the dynamics of competing populations based on limiting resources. We will analyze its properties and implications in subsequent sections.

It is convenient to nondimensionalize the model by introducing the following scaling:

$$n_1 = N_1/K_1,$$
$$n_2 = N_2/K_2,$$
$$\tau = r_1 t.$$

This scales population sizes relative to the corresponding carrying capacities and time relative to the intrinsic growth rate of species 1. Substituting these into the original model equations gives:

$$\frac{dn_1}{d\tau} = n_1(1 - n_1 - a_{12}n_2), \tag{10.3}$$
$$\frac{dn_2}{d\tau} = \rho n_2(1 - n_2 - a_{21}n_1), \tag{10.4}$$

with $\rho = r_2/r_1$.

10.3 Analysis of the nondimensional model

We now analyze the dynamics and behavior of the nondimensional competitive Lotka-Volterra model derived in the previous section. We first examine the equilibrium points of the system, found by setting (10.1) and (10.2) equal to zero and solving the resulting equations. This gives four possible equilibrium points:

$$E_1 = (0,0), \qquad E_2 = (1,0),$$
$$E_3 = (0,1), \qquad E_4 = \left(\frac{1-a_{12}}{1-a_{12}a_{21}}, \frac{1-a_{21}}{1+a_{12}a_{21}}\right)$$

The equilibrium points E_1, E_2, and E_3 correspond to boundary equilibria representing the extinction of one or both species. Specifically, E_1 is the point of double extinction, E_2 is the extinction of species 2, and E_3 is the extinction of species 1.

The point E_4 is an interior equilibrium where both species can potentially coexist. Notably, E_4 exists in the positive quadrant where population sizes are greater than zero only when the interspecific competition coefficients satisfy either $a_{12}, a_{21} < 1$ or $a_{12}, a_{21} > 1$.

To analyze the stability of the equilibrium points, we compute the Jacobian matrix of the system defined by Eqs. 10.3-10.4. Taking the partial derivatives, the Jacobian is:

$$J = \begin{bmatrix} 1 - 2n_1 - a_{12}n_2 & -a_{12}n_1 \\ -\rho a_{21}n_2 & \rho(1 - 2n_2 - a_{21}n_1) \end{bmatrix}. \tag{10.5}$$

Evaluated at each equilibrium point E_i, the Jacobian matrix $J(E_i)$ characterizes the linearized behavior near that point. The stability type is determined by the eigenvalues of each $J(E_i)$ matrix.

Evaluating the Jacobian at the equilibrium point $E_1 = (0,0)$ gives:

$$J(E_1) = \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}.$$

As this is a diagonal matrix, its eigenvectors are aligned with the n_1 and n_2 axes. The corresponding eigenvalues are 1 and ρ. Since both eigenvalues are positive, the Jacobian evaluation indicates that E_1 is an unstable node, where small perturbations will drive the populations away from the double extinction equilibrium.

The Jacobian evaluated at E_2 gives

$$J(E_2) = \begin{bmatrix} -1 & -a_{12} \\ 0 & \rho(1-a_{21}) \end{bmatrix}.$$

One of its eigenvectors is aligned with the n_1 axis and its corresponding eigenvalue is -1. The other eigenvector is

$$\left(\frac{a_{12}}{\rho(a_{21}-1)-1}, 1 \right),$$

while its associated eigenvalue is $\rho(1-a_{21})$. Therefore, when $a_{21} > 1$ the eigenvalue is negative and the fixed point is a stable node. Contrarily, when $a_{21} < 1$ the eivengalue is positive and the fixed point is a stable node. Notably, the stability of the fixed point where species n_1 is the sole survivor is determined by its aggressiveness.

The Jacobian evaluated at E_2 gives

$$J(E_2) = \begin{bmatrix} -1 & -a_{12} \\ 0 & \rho(1-a_{21}) \end{bmatrix}.$$

One of its eigenvectors is aligned with the n_1 axis and its corresponding eigenvalue is -1. The other eigenvector is

$$\left(\frac{a_{12}}{\rho(a_{21}-1)-1}, 1 \right),$$

while its associated eigenvalue is $\rho(1-a_{21})$. Therefore, when $a_{21} > 1$, this eigenvalue is negative and the fixed point E_2 is a stable node, representing the stable extinction of species 2. Contrarily, when $a_{21} < 1$ the eigenvalue is positive and E_2 is an unstable node, meaning small perturbations will drive the system away from the extinction of species 2. In summary, the stability of the fixed point where only species 1 survives depends on the competitiveness of species 2, represented by the parameter a_{21}.

The Jacobian evaluated at E_3 is:

$$J(E_3) = \begin{bmatrix} \rho(1-a_{12}) & 0 \\ -a_{21} & -1 \end{bmatrix}.$$

The eigenvector aligned with the n_2 axis has an eigenvalue of -1. The other eigenvector is

$$\left(1, \frac{a_{21}}{\rho(a_{12}-1)-1}\right)$$

with associated eigenvalue $\rho(1-a_{12})$. Therefore, when $a_{12} > 1$ this eigenvalue is negative and E_3 is a stable node, representing the stable extinction of species 1. However, when $a_{12} < 1$ the eigenvalue is positive and E_3 is an unstable node, meaning small perturbations will drive the system away from the extinction of species 1. Analogous to E_2, the stability of the fixed point where only species 2 survives depends on the competitiveness of species 1, given by the parameter a_{12}.

The Jacobian at the interior coexistence equilibrium E_4 is:

$$J(E_4) = \frac{1}{1-a_{12}a_{21}} \begin{bmatrix} a_{12}-1 & a_{12}(a_{12}-1) \\ \rho a_{21}(a_{21}-1) & \rho(a_{21}-1) \end{bmatrix}$$

The stability of E_4 depends on the eigenvalues of $J(E_4)$, which are given by:

$$\lambda_{1,2} = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2},$$

where the trace τ and determinant Δ are:

$$\tau = \frac{(a_{12}-1)+\rho(a_{21}-1)}{1-a_{12}a_{21}}, \quad \Delta = \frac{\rho(a_{12}-1)(a_{21}-1)}{1-a_{12}a_{21}}$$

It can be shown that:

$$\tau^2 - 4\Delta = \frac{[(a_{12}-1)+\rho(a_{21}-1)]^2 + 4\rho a_{12}a_{21}(a_{12}-1)(a_{21}-1)}{(1-a_{12}a_{21})^2}$$

Since E_4 only exists in the positive quadrant when $a_{12}, a_{21} < 1$ or $a_{12}, a_{21} > 1$, this further implies that the eigenvalues are real. If $\Delta > 0$, the eigenvalues have the same sign as τ, while one is positive and one negative otherwise. Finally, E_4 is a saddle when $a_{12}, a_{21} > 1$ because $\Delta < 1$, but a stable node when $a_{12}, a_{21} < 1$ since $\Delta > 0$ and $\tau < 0$.

The previous analysis reveals four possible long-term scenarios based on the competition coefficients a_{12} and a_{21}, as illustrated in Figure 10.3:

A) When $a_{12} > 1$ and $a_{21} < 1$, the extinction equilibrium E_2 is the sole steady state. In this case, species 1 dominates competition and excludes species 2 in the long run for nearly all starting population levels.

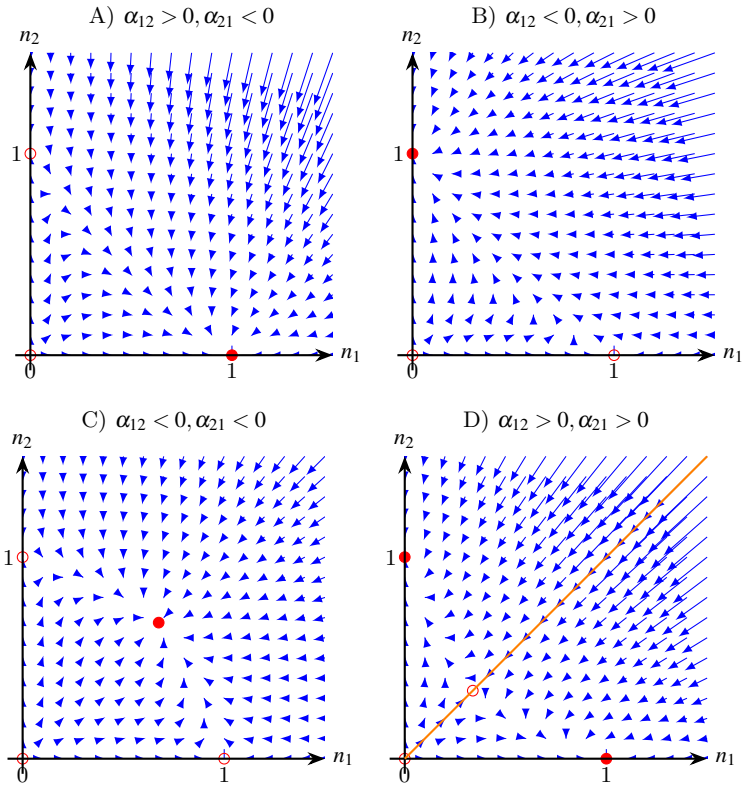


Fig. 10.1 Qualitative behavior of the competitive Lotka-Volterra model in the four regimes defined by the competition coefficients a_{12} and a_{21}. Arrows indicate the direction of flow and stable steady states are shown as filled circles. (A) When $a_{12} > 1$ and $a_{21} < 1$, species 1 competitively excludes species 2 and the system flows to the E_2 equilibrium. (B) For $a_{12} < 1$ and $a_{21} > 1$, species 2 outcompetes species 1 and the long-term outcome is the E_3 equilibrium. (C) When mutualism or weak competition occurs (a_{12}, $a_{21} < 1$), both populations can stably coexist at E_4. (D) Under strong reciprocal competition (a_{12}, $a_{21} > 1$), the system exhibits bistability between the exclusion of species 1 or 2, separated by a separatrix that passes through E_1 andE_4.

B) Conversely, when $a_{12} < 1$ and $a_{21} > 1$, E_3 is the only equilibrium. Now species 2 is competitively superior and drives species 1 extinct in the long run.

C) When a_{12} and a_{21} are both less than 1, E_4 represents the sole steady state. Here, interspecific competition is weak enough to allow the long-term coexistence of both species for most initial conditions.

D) Finally, if a_{12} and a_{21} both exceed 1, the extinction equilibria E_2 and E_3 become stable, creating a bistable system. In this case, the phase plane separates into two basins of attraction defined by a separatrix running through

E_1 and E_4. Populations beginning in either basin asymptotically approach either the E_2 or E_3 equilibrium exclusively.

The results of this model analysis have important implications for understanding competitive interactions and community assembly in ecological systems. When interspecific competition is strong, meaning one species is highly dominant over the other, the model predicts that stable coexistence is not possible long-term. The inferior competitor will be excluded from the community. This upholds Gause's principle of competitive exclusion, which states that no two species can occupy the same niche indefinitely if competing for the same limiting resources.

However, when interspecific competition is weak or mutualistic, the model indicates stable coexistence may occur. Only under conditions of weak resource overlap or facilitation between species can a diverse competitive community persist over generations. This provides insight into how biodiversity is maintained; strong niche differentiation or weak competitive effects allow multiple similar species to stably co-occupy an area together. Overall, the results tie competitive hierarchies directly to potential long-term diversity outcomes, deepening our comprehension of community structure and dynamics.

10.4 Discussion

While simple and stylized, the competitive Lotka-Volterra model yields important insights into interspecific interactions and their implications for community organization. By capturing the essential dynamics of competing populations through logistic growth and competitive reductions to per capita growth rates, the model provides a mathematical framework for exploring how competition shapes long-term community outcomes.

Despite many unrealistic assumptions, such as constant competition coefficients and symmetrical interspecific effects, the Lotka-Volterra model successfully describes the four qualitative behaviors that can emerge from competitive interactions. These regimes, distinguished by the relative competitive abilities of each species, predict different long-term fates ranging from coexistence to exclusion. In this way, the model functions well as an abstract, conceptual explanatory model rather than a quantitatively predictive one.

Explanatory models like Lotka-Volterra aim to provide mechanistic understanding of natural phenomena, rather than detailed forecasts of real populations. They distill the essence of a process into mathematical terms to investigate general principles and properties. While simplifying assumptions neglect ecological complexities, this allows revealing underlying dynamics that may otherwise remain obscure. The insights generated can then guide more realistic, data-driven modeling as our comprehension progresses.

For example, the Lotka-Volterra model demonstrated how competitive hierarchies and resource partitioning influence species diversity through niche

differentiation and exclusion. This basic theoretical relationship between interspecific interactions and community structure informed decades of subsequent empirical research. While empirical studies have revealed scenarios more complex than the idealized model, its explanatory value endures over a century later.

Naturally, limitations arise from unrealistic assumptions. Constant parameters ignore environmental fluctuations or evolutionary changes to competitive abilities over time. Symmetrical effects do not reflect specializations that could permit stable coexistence. Despite this, the Lotka-Volterra model fulfills its purpose of building foundational understanding rather than accurate prediction for any specific system. Its enduring usefulness demonstrates the power of conceptual models to advance ecology conceptually. Overall, the Lotka-Volterra model exemplifies how simple theory can provide explanatory insights that stimulate new hypotheses and guide more sophisticated modeling in future.

Chapter 11
SIR Model

Abstract This chapter introduces the basic SIR modeling framework for studying infectious disease transmission dynamics. The SIR model divides a population into susceptible, infected, and recovered compartments, and represents changes in each compartment through a system of differential equations formulated based on assumptions of homogeneous mixing and exponential recovery rates. The model equations are non-dimensionalized and analyzed to identify fixation points and determine stability through Jacobian analysis, revealing a critical transition at the basic reproduction number R0=1 that governs disease invasion and establishment. Various extensions and applications of the SIR model are also discussed to demonstrate how it provides qualitative and quantitative insights into real disease systems through both analytic solutions and simulation-based analyses, establishing its value as a foundational approach for epidemiological modeling.

11.1 Introduction

Mathematical modeling of infectious disease transmission has been crucial for understanding and controlling epidemics. One of the simplest yet most influential epidemiological models is the SIR model. Proposed in the early 20th century, the SIR model divides a population into three categories - Susceptible, Infected, and Recovered - and represents transmission through a system of differential equations.

Despite its assumptions of homogeneous mixing and exponential recovery rates, the SIR model has provided fundamental insights into disease dynamics. Threshold quantities like the basic reproduction number R0 characterize when an outbreak will occur based on transmission rates. Analytic solutions

also demonstrate a range of epidemiological behaviors from disease invasion and persistence to endemic equilibrium.

The basic SIR framework has served as a starting point for extending models to consider additional biological and external influence realistically. Yet even in its simple form, the model illustrates how self-organizing transmission processes alone can lead to complex episodic outbreak patterns on the population scale. SIR analyses and simulations provide an elegant, versatile tool to study disease control strategies from targeted quarantines to vaccination programs.

In this chapter, we will introduce the SIR modeling framework in detail. We will formulate the compartmental differential equations, obtain analytic solutions, and explore simulation-based analyses. Examples will demonstrate how this basic setup can be applied to real disease systems while also motivating extensions. The goal is to establish the conceptual and technical foundations of SIR modeling as a foundation for more complex and realistic epidemiological representations.

11.2 Model Formulation

The SIR model divides population into three mutually exclusive compartments based on disease status: susceptible (S), infected (I), and recovered (R). The following assumptions are made:

- Individuals mix homogeneously; the probability of contact between any two individuals is equal.
- Upon infection, individuals immediately become infectious.
- Transmission occurs through contact between susceptible and infected individuals. The rate of new infections is thus proportional to both the number of susceptible (S) and infected (I) individuals, with proportionality constant β known as the transmission rate.
- Infected individuals recover at a constant rate γ and acquire lifelong immunity.
- The population birth rate is constant and denoted by α.
- Individuals in all compartments die due to causes other than disease at constant rate μ, and infected individuals also die due to causes directly associated to the disease at a constant rate δ.

These mechanisms are schematically represented in Figure 11.2 which pictures infection dynamics as flux-balance processes. Taking this into consideration, the population evolution in all compartments can then be expressed as a system of ordinary differential equations:

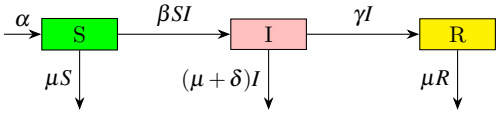


Fig. 11.1 Schematic representation of the SIR model compartmental structure and transitions. Susceptible individuals (compartment S) become infected at a rate proportional to contacts with infectious individuals in compartment I. Infected individuals recover at rate γ and enter the recovered compartment R with lasting immunity. All compartments face a baseline death rate μ. Additionally, infectious individuals face an extra disease-induced death rate δ owing to pathogenic effects.

$$\begin{aligned}
\frac{dS}{dt} &= \alpha - \beta SI - \mu S, \\
\frac{dI}{dt} &= \beta SI - (\gamma + \mu + \delta)I, \\
\frac{dR}{dt} &= \gamma I - \mu R.
\end{aligned}$$

11.3 Normalizing the Model

We can non-dimensionalize the system of differential equations by introducing the following normalized variables:

$$\begin{aligned}
s &= \frac{S}{\alpha/\mu}, \\
i &= \frac{I}{\alpha/(\gamma+\mu+\delta)}, \\
r &= \frac{R}{\alpha/\mu}, \\
t' &= \mu t
\end{aligned}$$

Let us further introduce a non-dimensional time variable $t' = \mu t$ such that With this, the ODE system becomes:

$$\begin{aligned}
\frac{ds}{dt'} &= 1 - R_0 si - s, \\
\frac{di}{dt'} &= \rho(R_0 si - i), \\
\frac{dr}{dt'} &= \varepsilon i - r,
\end{aligned}$$

where

$$R_0 = \frac{\alpha\beta}{\mu(\gamma+\mu+\delta)},$$
$$\rho = \frac{\gamma+\mu+\delta}{\mu},$$
$$\varepsilon = \frac{\gamma}{\gamma+\mu+\delta}.$$

The normalization process yields several advantages. First, it reduces the original model with 5 parameters to an equivalent model with only 3 non-dimensional lumped parameters. This simplifies the analysis. Second, some of these lumped parameters have special epidemiological meaning. In particular, R_0 is known as the basic reproductive number. To understand its significance, rewrite R_0 as:

$$R_0 = \beta \frac{\alpha}{\mu} \frac{1}{\mu+\gamma+\delta}.$$

β represents the number of new infections generated by each infectious individual per unit time. α/μ is the disease-free population size. Thus, $\beta\alpha/\mu$ gives the number of infections generated per unit time by one infectious individual introduced into a completely susceptible population. Meanwhile, $1/(\mu+\gamma+\delta)$ is the average time an individual spends in the infectious compartment before either recovering or dying. Multiplying these terms gives R_0, which can thus be interpreted as the expected number of secondary infections produced by a single infectious individual introduced into a wholly susceptible population. The dynamical significance of this threshold will become clear in subsequent analysis.

11.4 Fixed points and stability analysis

The first two differential equations of the normalized SIR model, governing the dynamics of s and i, form an autonomous subsystem that is decoupled from the equation for r. This subsystem can be analyzed independently of r, then the solutions for i can be substituted into the third equation to solve for r.

Focusing first on the independent (s,i) subsystem will allow us to characterize the overall infection dynamics without consideration of the recovered population. Solving this core transmission model will provide insight into whether the disease can invade and establish infection within the population.

Rewriting this subsystem

$$\frac{ds}{dt'} = 1 - R_0 si - s,$$
$$\frac{di}{dt} = \rho(R_0 si - i).$$

It is straightforward to prove that it has two different fixed points:

$$s_1^* = 1, \qquad s_2^* = \frac{1}{R_0},$$
$$i_1^* = 0, \qquad i_2^* = 1 - \frac{1}{R_0}.$$

To analyze the stability of the disease-free equilibrium, we compute the Jacobian matrix of the reduced subsystem:

$$J = \begin{bmatrix} -1 - R_0 i & -R_0 s \\ \rho R_0 i & \rho(R_0 s - 1) \end{bmatrix}.$$

Evaluating this at the disease-free fixed point $(s_1^*, i_1^*) = (1,0)$ gives:

$$J(1,0) = \begin{bmatrix} -1 & -R_0 \\ 0 & \rho(R_0 - 1) \end{bmatrix}.$$

The eigenvectors of the Jacobian matrix are:

$$v_1 = (1,0),$$
$$v_2 = \left(-\frac{R_0}{1 + \rho(R_0 - 1)}, 1 \right),$$

with corresponding eigenvalues:

$$\lambda_1 = -1, \quad \lambda_2 = \rho(R_0 - 1).$$

These results indicate that the disease-free fixed point is locally asymptotically stable when $R_0 < 1$, and unstable otherwise, transitioning to a saddle point structure at the critical value of $R_0 = 1$. This stability switches at the epidemic threshold of $R_0 = 1$.

We can investigate the stability of the endemic equilibrium by evaluating the Jacobian matrix at (s_2^*, i_2^*):

$$J(s_2^*, i_2^*) = \begin{bmatrix} -R_0 & -1 \\ \rho(R_0 - 1) & 0 \end{bmatrix}.$$

Rather than computing the eigenvalues directly, which could yield cumbersome expressions, we analyze the Jacobian trace (τ) and determinant (Δ) as explained in Chapter 9 and summarized in Figure 9.1:

$$\tau = -R_0,$$
$$\Delta = \rho(R_0 - 1).$$

From these results, we see that $\Delta < 0$ when $R_0 < 1$, indicating the endemic equilibrium is a saddle node in this case. Additionally, for $R_0 > 1$, $\tau < 0$, im-

plying the endemic equilibrium is locally asymptotically stable. In summary, the trace and determinant criteria reveal that the endemic fixed point undergoes a transcritical bifurcation, transitioning from a saddle to a stable node as R_0 passes through the threshold value of 1.

Figure 11.4 summarizes the findings of the local stability analysis. Observe that i_2^* takes negative values when $R_0 < 1$. This means that, in practice, only the disease-free steady state exists when the average number of secondary infections produced by a single infectious individual introduced into a wholly susceptible population is less than one.

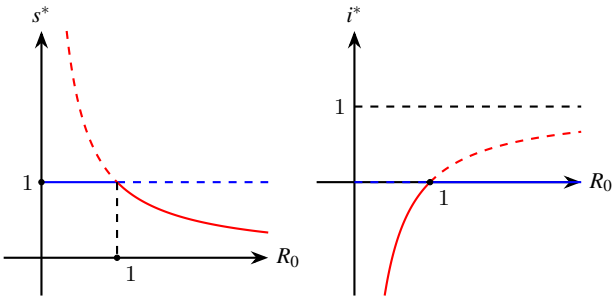


Fig. 11.2 Steady-state values and stability domains as a function of the basic reproduction number, R_0. (Left) In the disease-free state, the fraction of susceptible individuals, s^*, remains at 1, represented by the blue curve. However, in the endemic-disease state s^* decreases inversely with R_0, shown by the red curve. (Right) The normalized infected fraction at equilibrium, i^*, is zero in the disease-free state (blue curve) and increases as $1 - 1/R_0$ in the endemic-disease state (red curve). A transcritical bifurcation occurs at R0=1. For R0 < 1, the disease-free equilibrium is stable (solid lines) while the endemic steady state is unstable (dashed lines). At the bifurcation point, the two equilibria collide and their stability reverses for R0 > 1.

In summary, the SIR model predicts two possible long-term outcomes depending on the value of R_0. Specifically, the model has two steady states: a disease-free equilibrium where the infection is eliminated, and an endemic equilibrium where the disease persists at a constant level in the population.

The stability analysis revealed that a critical transition occurs at $R_0 = 1$. For $R_0 < 1$, the disease-free state is locally asymptotically stable, meaning any small perturbation will fade out as the infection dies from the population. However, when $R_0 > 1$, the disease-free equilibrium loses stability. This same threshold governs the stability of the endemic equilibrium, which undergoes a bifurcation at $R_0 = 1$. Below the threshold, the endemic state is a saddle node and unstable. But for $R_0 > 1$, it transitions to a stable node.

Biologically, R_0 represents the average number of secondary cases produced by one infected individual in a fully susceptible population. Hence, by determining whether R_0 is above or below one, we can predict if a disease invasion is likely to succeed or fail at establishing long-term transmission within the

community. The stability analysis therefore provides key insights into disease dynamics based on the epidemiological threshold defined by R_0.

Based on the definition of R_0, there are a few measures that can be taken to reduce its value:

- Reduce transmission rate (β). Public health interventions that decrease the risk of transmission during contact, such as hand hygiene, masks, distancing, and ventilation, will lower β. Non-pharmaceutical interventions like lockdowns and restrictions on gatherings that reduce social contact frequency also impact β.
- Increase recovery rate (γ). Providing supportive medical care and targeted treatment to infected individuals can help speed their recovery and shorten the duration that they remain infectious, effectively increasing parameter γ.
- Increase removal rate (δ). In agricultural or wildlife settings, humane culling of visibly infected individuals can effectively increase the disease-induced mortality rate parameter δ. However, in human populations a more ethical approach is to quickly identify infectious individuals through widespread testing and rigorous contact tracing and place them in isolation.
- Reduce influx into susceptible class (α). Rather than interpretation as birth rate, α is better viewed as the rate at which new individuals become susceptible, such as through loss of immunity. Vaccination upon birth or entrance into the susceptible population lowers this parameter without unwanted demographic effects.
- Combination of approaches: The most effective strategies often combine multiple types of measures simultaneously. This creates additive reductions in R_0 through independent pathways compared to single measures.

11.5 Discussion

While the basic SIR model makes several restrictive assumptions, it has proven remarkably capable of capturing fundamental mechanisms of disease transmission dynamics. The model's enduring value stems from its ability to isolate a minimal set of elements necessary to mathematically represent infectious propagation through a population based on transmission between individuals. Though heterogeneous mixing, temporary immunity, and other real-world complexities are ignored, the model framework provides conceptual and predictive insights that have remained highly relevant.

Despite its limitations, the SIR model elucidated critical but unintuitive properties like the existence of epidemic thresholds defined by epidemic quantities like R0. Threshold effects and bifurcation behaviors revealed through model analysis explain why certain interventions can entirely curb outbreaks

while others have negligible impact. Such lessons challenge naïve assessments and guide rational intervention design.

Perhaps most remarkably, the simple SIR rules give rise to complex and varied overall infection patterns at the system scale depending on parameter conditions. Self-organized criticality and phase transition phenomena emerge from local inter-individual interactions alone. These phenomena shed light on how diseases can transition between disparate outbreak modes from fleeting invasion to endemic stability.

The model's intrinsic value also lies in its explanatory character; by distilling phenomena to essential mathematical elements, relationships between underlying factors and observable behaviors are uncovered. This enhances conceptual understanding beyond empirical observations alone. The model serves as a framework upon which more realistic representations can be built through principled expansions, while preserving core transmission dynamics.

In summary, despite being formulated over a century ago with limited data and computational power, the SIR model has proven to transcend its restrictive assumptions through theoretical and practical insights that remain highly relevant today. Its minimal yet fundamental representation of transmission acts as the cornerstone upon which our quantitative understanding of infectious disease dynamics continues to be built.

Chapter 12
The interdisciplinary journey of oscillators: from transatlantic navigation to neurophysiology

Abstract Oscillators have proven profoundly versatile, weaving through numerous disciplines to unlock innovations. In this chapter, we explore how oscillators have had a significant impact on the advancement of science and technology throughout history. Both mechanical and electromagnetic oscillators, along with the mathematical models that describe them, have played a crucial role in various fields such as navigation, communications, and the study of neurophysiology. Key contributions in the science of oscillators that have left a lasting influence on other areas are highlighted, and the influence of figures like John Harrison, James C. Maxwell, Heinrich Hertz, Balthasar van der Pol, Alan Hodgkin, and Andrew Huxley is emphasized. We analyze how oscillators have bridged disciplines as diverse as telecommunications and neurophysiology, and discuss how they continue to be essential in current research and the development of advanced interdisciplinary approaches.

Physics encompasses a number of notable concepts that have shaped the course of scientific and technological progress. The ideal gas and the blackbody radiator, for example, have had a significant impact on various branches of physics and inspired the development of novel technologies. However, the harmonic oscillator is arguably the most profound and versatile concept of all.

The study of oscillators has a long and rich history, weaving together ideas and discoveries from diverse disciplines including physics, mathematics, engineering, biology, and medicine. Research on oscillators has unlocked fundamental insights about the natural world while also enabling transformative technologies.

This chapter will trace the winding journey of oscillator research over centuries, highlighting interdisciplinary collaborations that produced groundbreaking innovations. We will explore how the quest for accurate timekeeping at sea drove innovations in clock engineering, and how this converged with

discoveries in electromagnetism to enable radio technology. Electrical oscillator models would come full circle, providing key insights into biological oscillators underlying heart rhythms and nerve impulses. By tracing this vivid narrative, we will reveal how the incessant human drive to understand and manipulate oscillatory systems has profoundly reshaped civilization.

12.1 The Longitude Problem and the Development of Clocks

In the 2nd century BCE, the Silk Road was established as a network of trade routes linking China with the Mediterranean region, fostering the exchange of goods, ideas, and technologies between Asia and Europe. However, after the fall of Constantinople in 1453, the Silk Road was disrupted, making trade routes to Asia and the Middle East more cumbersome and expensive for Europeans. To circumvent this challenge, explorers set out to discover direct sea routes to Asia, leading to seminal voyages such as Christopher Columbus' discovery of the Americas in 1492, Vasco da Gama's expedition that reached India by sailing around Africa in 1498, and the voyage of circumnavigation initially commanded by Ferdinand Magellan and completed in 1522 (Parry, 1992). These journeys were soon followed by others from various European powers, marking the beginning of the Age of Exploration, which spanned from the 15th to the 18th century and produced significant changes in global trade, culture, and politics.

The contributions of Islamic scholars during the Islamic Golden Age, which spanned from the 8th to the 14th century, in the fields of mathematics, astronomy, medicine, and geography are partly responsible for the accomplishments of the Age of Exploration (Hodgson, 1977). They contributed significantly to the translation and preservation of works from the Greek and Roman eras, which had a major impact on philosophers of the European Renaissance. Since navigation relied heavily on the observation of celestial bodies prior to the 18th century, the contributions in astronomy and mathematics were particularly notable. However, astronomical methods only allowed sailors to determine their latitude, or their north-south position, while determining longitude was a more challenging task. This challenge came to be known as the longitude problem.

Accurate determination of longitude was vital for ensuring safe navigation, particularly during lengthy voyages. Vessels capable of precisely calculating their position faced reduced risks of grounding, collision, or becoming lost at sea. This not only protected the lives of sailors but also safeguarded valuable cargo, minimizing losses for ship owners and merchants. Additionally, efficient navigation reduced travel time, enabling ships to undertake more voyages and enhance profitability. To promote advancements in this crucial area, the British government introduced the Longitude Prize, also referred to as the Longitude Act, in 1714. This competition aimed to encourage the creation of

a dependable and feasible method for determining longitude at sea (Sobel, 2007).

John Harrison, a British clockmaker, successfully tackled the problem of determining longitude. Harrison recognized the importance of accurate timekeeping in solving this problem. Rather than relying on celestial observations, Harrison proposed that a meticulously crafted timepiece could allow sailors to keep track of the time at a reference location and compare it to the local time. By comparing the local and reference times, it was possible to convert the difference in hours, minutes, and seconds into a reliable indicator of longitude.

John Harrison was born in 1693 in Yorkshire, England. From a young age, he displayed exceptional skill as a carpenter and clockmaker. In the 1720s, Harrison moved to London and began designing innovative clocks, including longcase clocks with novel mechanisms. Interested in the Longitude Prize, he set out to design a marine chronometer that could maintain accuracy under the harsh conditions encountered on transoceanic voyages.

Following on the steps of scientists and clock-makers like Christiaan Huygens, Harrison crafted a series of marine chronometers. The first, known as H1, was finished in 1735. It pioneered innovations such as the use of 'grasshopper escapement' and temperature compensation through a bimetallic strip. However, issues with operation at sea meant further refinements were needed.

Over the next few decades, Harrison went on to develop the H2, H3 and finally the H4 models, incrementally improving the timekeeping accuracy and durability. The H4 chronometer, completed in 1759, was a compact, high-performance device that fully met the demands of marine navigation. In rigorous sea trials, it kept time to within a few seconds per day, allowing longitude to be calculated to within half a degree.

The Board of Longitude was reluctant to fully award Harrison the prize money, leading to a prolonged dispute. However, his chronometers were widely recognized as a monumental achievement. Their unprecedented accuracy revolutionized navigation and enabled the great voyages of exploration that expanded global knowledge.

At its core, Harrison's ingenious solution to the longitude problem required developing a sustained mechanical oscillator that could remain perpetual and constant for months under the harsh conditions of ocean voyages. The marine chronometer's clockwork mechanism is precisely an oscillator in which energy dissipated through friction is continually replenished. This allows the periodic motion to be maintained indefinitely, enabling accurate timekeeping over long periods.

The marine chronometer was a keystone technology that helped inaugurate the modern scientific era. By solving the centuries-old quest for accurate timekeeping, Harrison's ingenious clocks allowed great leaps forward in astronomy, surveying, navigation and mapping. His contributions epitomize the enlightenment values of rationality, progress and the mastery of nature through science and technology.

12.2 Harnessing Oscillators for Wireless Communication

The chronometer's sustained mechanical oscillations enabled transformative advances in global navigation and mapping. However, the nineteenth century ushered in a new paradigm as scientists sought to harness the power of electromagnetic waves. Just as the perpetual motion of Harrison's clocks arose from carefully sustaining a harmonic system, steady electrical oscillations would prove vital for powering modern wireless communication.

The pioneering experiments on electricity and magnetism conducted in the eighteenth and early nineteenth centuries laid the foundation for understanding oscillatory electrical phenomena (Baigrie, 2006). Notable contributions came from Hans Christian Ørsted's discovery of the connection between electricity and magnetism in 1820, demonstrating that an electric current produces a magnetic field, and the efforts of Michael Faraday to quantitatively relate electric and magnetic forces. However, it was James Clerk Maxwell who, through his seminal treatise published in 1873, unified the previously fragmented theories into a comprehensive framework (Maxwell, 2011a,b).

Maxwell synthesized prior knowledge within a system of equations describing the interrelation between electric and magnetic fields. A key insight was that these fields could propagate through space as waves, spreading energy outward from a source. One remarkable conclusion was that light itself was simply a high-frequency electromagnetic wave. Although controversial at the time, this revelation that light was electromagnetic rippled across physics and enabled huge leaps in scientific progress.

Maxwell's equations predicted that, in principle, electromagnetic waves of any frequency could propagate through empty space. However, generating such waves to study their properties proved challenging. It was not until 1887 that Heinrich Hertz conclusively demonstrated the existence of electromagnetic waves in a pioneering experiment (Shamos, 1987).

In his laboratory, Hertz sought to generate and detect the elusive electromagnetic waves predicted by Maxwell's equations. To achieve this, he devised an electrical oscillator circuit connected to a dipole antenna. When operated at high voltages, oscillations in the circuit induced corresponding oscillations in the antenna, launching electromagnetic radiation. To detect the waves, Hertz used a separate loop antenna connected to another resonant circuit. When placed in the path of the radiation, oscillations were induced in this receiver loop, providing experimental evidence of propagating electromagnetic waves.

The apparatus engineered by Hertz definitively confirmed Maxwell's theories and opened up a new paradigm in physics. Additionally, Hertz measured properties of the waves including polarization, reflection, refraction and interference, verifying they obeyed the same rules as visible light. This cemented light's electromagnetic basis.

Hertz's experiments proved that coordinated electrical oscillations could be harnessed to generate and receive wireless signals. However, the capabili-

ties of Hertz's apparatus were limited. While suitable for laboratory studies, more powerful and efficient oscillators would be needed to enable practical applications like radio communication.

Early radio transmitters used crude ore detectors and inefficient spark gap generators to produce electromagnetic radiation. Generating steady, high-frequency electrical oscillations was key to enabling continuous wave radio broadcasts. This relied on certain analogies with mechanical oscillators like those used in chronometers. A resonant LC circuit can exhibit oscillations, but losses due to resistance will dampen the oscillations over time. To sustain continuous oscillations, energy must be continually fed into the system, analogous to rewinding the chronometer's drive spring. In an electronic oscillator, amplifying elements like vacuum tubes are used to replenish the energy dissipated in the resonant tank circuit. This negative resistance precisely counteracts losses, allowing persistent high-frequency oscillations. By maintaining energy balance in an electrical harmonic system, electronic oscillators enabled modern wireless communication.

The parallels between mechanical and electrical oscillators highlight the continuity in physics concepts across disciplines. Whether weighing a clock's escapement or balancing amplifiers and dissipation in a radio transmitter, the goal of sustaining an oscillator despite losses united these efforts. Powered by electronic oscillators, radio technology would fundamentally reshape society in the 20th century (Clarke, 2011).

12.3 Modelling Neural Excitation with Oscillators

The transmission of signals in nerves and muscles relies on the propagation of electrical impulses known as action potentials. Understanding the nonlinear dynamics underlying excitation in neurons and cardiac cells would require integrating concepts from physics, mathematics and biology. Once again, oscillator models would provide vital insights, this time by mimicking the spikes and rhythms produced by biological cells.

The Dutch physicist Balthasar van der Pol joined Philips Research Laboratories after receiving his doctorate in 1913. During his time there, he made significant contributions to the field of electronics, especially in radio technology. Van der Pol was a key figure in the development of the Philips radio receiver, which was a huge success at the time. According to one of his biographers, "Radio might have remained a field of haphazard empiricism along with wild commercial ventures, but for the influence of men like Van der Pol who stressed the need for a more scientific approach" (Bremmer et al., 1960, Introduction).

In 1926, van der Pol derived a nonlinear differential equation to describe the behavior of vacuum tube circuits used in early radios. While analyzing his model equations, van der Pol made an intriguing discovery. When the value

of a parameter known as the damping coefficient was small, the system exhibited the familiar traits of a harmonic oscillator. However, as the damping coefficient increased to larger values, the model solutions diverged from the conventional behavior of harmonic oscillators. Instead of following smooth periodic sinusoidal oscillations, the model solutions displayed alternating periods of rapid and slow changes. Van der Pol coined the term "relaxation oscillations" to describe this behavior (van der Pol, 1926).

The spiking behavior of Van der Pol oscilator was analogous to the action potentials transmitted along neurons and the cells of the heart pacemaker. Building on this, van der Por created a circuit with three oscillators to reproduce the electrical behavior of the heart. Using this method, he successfully replicated specific cardiac arrhythmias, demonstrating the model's ability to capture the complex dynamics of cardiac electrical activity (van der Pol and van der Mark, 1928).

In parallel to the oscillator modeling work, new technologies enabled direct measurements of neural electrical signals. The development of operational amplifiers (op-amps) in the early 1940s was a key advance. Op-amps can perform high-gain voltage amplification and mathematical operations like addition, subtraction, integration and differentiation. This made signal processing and analysis much more feasible. In 1949, Kenneth Cole and Howard Curtis invented the voltage clamp technique using an op-amp based feedback circuit to control the voltage across a cell membrane (Brown, 2020). This allowed the underlying ionic currents to be precisely measured for the first time. The voltage clamp provided pivotal insights into the ionic basis of neural electrical signaling.

Fundamental discoveries in thermodynamics and physical chemistry by pioneers like Max Planck and Walther Nernst also contributed to understanding the electrical behavior of cell membranes. Nernst's work led to his eponymous equation in 1889, relating an ion's equilibrium potential to its concentration gradient. The Nernst-Planck equations described the motion of charged particles, extending diffusion to include electrostatic effects.

Another significant milestone was Cole's adoption of the squid giant axon as a model for studying membrane electrical properties in 1936, following J. Z. Young's suggestion. The size and spacious lumen of the squid axon made it more amenable to experimental techniques such as intracellular recording, which were previously impractical with smaller axons.

While collaborating with Cole in the 1930s, Hodgkin realized the potential of using the squid giant axon to record intracellular action potentials. Returning to Cambridge in 1938, he recruited the undergraduate Andrew Huxley. Huxley's outstanding physics and mathematics background paired with Hodgkin's neurophysiology expertise. Their collaborative project resulted in a series of papers providing quantitative insights into the biophysical basis of the action potential (Brown, 2020).

The culminating paper in the Hodgkin and Huxley series established the field of quantitative membrane biophysics, integrating mathematical model-

ing with empirical measurements. Hodgkin and Huxley developed a model of interconnected differential equations and performed painstaking calculations to match the model output to their voltage clamp data. Lacking access to early computers, Huxley relied on a desktop calculator, achieving remarkable precision through rigor. This yielded fundamental insights into the ionic mechanisms mediating the initiation and conduction of the action potential.

The Hodgkin-Huxley model marked a pivotal breakthrough in neuroscience, offering the first quantitative, mechanistic picture of neural excitability. Their integrative approach combining math and biology yielded insights into ion channel function while spurring new fields like computational neuroscience. This achievement emerged from interdisciplinary collaboration, as experts in physics, mathematics and physiology united around the common goal of demystifying the action potential. The Hodgkin-Huxley collaboration highlights the power of bridging disciplinary divides to propel discovery. Their synthesis of diverse perspectives unveiled fundamental insights previously out of reach.

Though groundbreaking, the complexity of the Hodgkin-Huxley model posed challenges for broader usage in an era before digital computers. In the 1960s, Richard FitzHugh at the NIH sought to investigate the model's mathematical properties using nonlinear dynamics techniques. To solve the equations across parameters, FitzHugh collaborated with John Moore to build an analog computer from op-amps, multipliers and plotters (Izhikevich and FitzHugh, 2006). This allowed them to graphically visualize solutions, though operating the intricate analog simulator required sophisticated engineering and math skills. The model complexity and lack of digital computing motivated efforts to derive simplified models of neural excitation.

Guided by Cole's insights, FitzHugh modified van der Pol's relaxation oscillator equations to distinguish key features of the Hodgkin-Huxley model. The aim was to separate the dynamics of sodium and potassium ion flow across the membrane from the regenerative excitation process. Originally called the Bonhoeffer-van der Pol equations, these were later renamed the FitzHugh-Nagumo equations because, around the same time, Japanese engineer Jin-Ichi Nagumo invented an electronic circuit using tunnel diodes that reproduced the key cubic nonlinearity. The simplified FitzHugh-Nagumo model provided an accessible approximation to the complex Hodgkin-Huxley system. Reconfiguring the analog computer to solve these reduced equations enabled more extensive mathematical analyses of neural excitation dynamics.

The quest to understand the electrical basis of neural signaling exemplifies the potency of cross-disciplinary pollination. Integrating oscillator models from physics and engineering with emerging techniques to probe nerve impulses yielded pivotal insights into the mechanisms of neural excitation. Constructing fruitful analogies between biological and man-made oscillators led to accessible mathematical representations that captured the essence of spiking dynamics. By breaking down barriers between physics, mathematics and physiology, pioneers transformed understanding of the fundamental pro-

cesses underlying thought itself. The intertwined narrative of biological and technological oscillators underscores how synthesizing diverse perspectives propels discovery.

12.4 Conclusion

The history of oscillators unveils the intricate relationship between science and technology. Driven by practical needs, advancements in timekeeping brought about significant changes in navigation. Not only this pursuit had profound political and economical effects but also led to fundamental scientific discoveries.

Likewise, electronic oscillators played a crucial role in powering wireless communication systems that transformed society. Yet, when examined through mathematical models, they also shed light on the electrical rhythms of the nervous system. Time and again, the manipulation of oscillatory systems for engineering purposes unintentionally deepened our scientific understanding, crossing disciplinary boundaries.

The enduring significance of oscillators lies in their versatility as model systems. The concept of the harmonic oscillator provides an approximate description for a wide range of systems, including physical, biological, and engineering domains. This commonality across different areas facilitates analogies that enhance the process of discovery.

Whether it's adjusting escapement mechanisms, optimizing amplifiers, or simulating action potentials, researchers have consistently identified common patterns that can be expressed mathematically. The science of oscillators has thrived through the integration of ideas from various fields. Pioneers from diverse backgrounds collaborated to convey concepts between disciplines, building bridges that connected unexpected yet fertile intellectual territories.

By tracing the history of oscillations through the centuries, we gain insight into the very evolution of science itself. While curiosity guides exploration in specialized fields, it's the process of synthesis that breathes life into the deepest ideas. The journey of oscillators highlights how collaborative human effort, aimed at understanding natural rhythms, regardless of their origin, propels us toward a deeper understanding of nature and contributes to the shaping of our civilization.

Finally, oscillators still play a crucial part in today's science and technology. High-frequency electronic oscillators act like the heartbeat of modern computers, making sure that different parts work together and share information at the right time. In computer science, they facilitate complex calculations and data processing by synchronizing clock cycles and enabling high-speed data transfer within computer systems. In the internet, they play a central role in precise data transmission, ensuring the uninterrupted flow of information during digital interactions. These unsung heroes quietly con-

tribute to the seamless operation of computer technology and the interconnected world we rely on daily.

References

Baigrie B (2006) Electricity and magnetism. Greenwood Guides to Great Ideas in Science, Greenwood Press, Westport, CT

Bremmer H, Bouwkamp C, Pol, van der B (eds) (1960) Selected scientific papers Balthasar van der Pol. North-Holland Publishing Company, Netherlands

Brown A (2020) A companion guide to the Hodgkin-Huxley papers A companion guide to the Hodgkin-Huxley papers. Physiological Society

Clarke AC (2011) How the world was one. Hachette UK

Hodgson MGS (1977) The venture of Islam: The classical age of Islam v.1. University of Chicago Press, Chicago, IL

Izhikevich E, FitzHugh R (2006) FitzHugh-nagumo model. Scholarpedia 1(9):1349, DOI 10.4249/scholarpedia.1349, URL https://doi.org/10.4249/scholarpedia.1349

Maxwell J (2011a) A Cambridge library collection - physical sciences A treatise on electricity and magnetism: Volume 1. Cambridge University Press, Cambridge, England

Maxwell J (2011b) A Cambridge library collection - physical sciences A treatise on electricity and magnetism: Volume 2. Cambridge University Press, Cambridge, England

Parry JH (1992) The discovery of the sea. University of California Press, Berkeley, CA

van der Pol B (1926) LXXXVIII. on "relaxation-oscillations". The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2(11):978–992, DOI 10.1080/14786442608564127

van der Pol B, van der Mark J (1928) LXXII. the heartbeat considered as a relaxation oscillation, and an electrical model of the heart. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 6(38):763–775, DOI 10.1080/14786441108564652

Shamos MH (1987) Great Experiments in Physics. Dover Publications, Mineola, NY

Sobel D (2007) Longitude. Walker, New York, NY