Chapter 4

Nonequilibrium Dynamics: Oscillation

4.1 Oscillations in Nature

We now have to make a detour out of mathematics into science. We have to ask: what are the fundamental kinds of behaviors that can be seen in a scientific system, and what do they look like mathematically?

We have all seen scientific concepts of *equilibrium* playing a fundamental role in many scientific theories.

Chemistry.	We are told that chemical substances placed in a box will quickly go to equilibrium, called "chemical equilibrium."
Thermodynamics.	A hot cup of coffee in a cooler room will quickly go to an equilibrium temperature with the environment, a condition called "thermodynamic equilibrium."
Economics.	We are told that a free market with many small traders will reach an equilibrium price where supply meets demand, called "economic equilibrium."
Physiology.	We are taught the doctrine of homeostasis, which says that the body regulates all physiological variables, such as temperature and hormone levels, to remain in "physiological equilibrium."
Ecology.	Older theories were often phrased in terms of equilibrium concepts such as "carrying capacity" and "climatic climax." The population rises or falls until it reaches the ecosystem's carrying capacity, or the community composition changes until it reaches a state determined by climate and soil, at which point the system is in "ecological equilibrium."

Oscillation in Chemistry and Biology

If "equilibrium" truly described scientific phenomena, we could stop the investigation right here and begin to look for point attractors in all of our models of natural phenomena.

But are systems in nature really governed by equilibrium dynamics? No! The problem is that in every one of the above examples, in every one of these sciences, the doctrine of equilibrium behavior is factually wrong or at least incomplete as a description of the behavior of those systems.

We already saw, in Chapter 1, many types of systems in which the fundamental behavior is oscillation, not equilibrium. Hormones oscillate, and ecosystem populations oscillate. There are

A. Garfinkel et al., *Modeling Life*,
DOI 10.1007/978-3-319-59731-7_4

also thermodynamic oscillations, and oscillations in economic markets. In fact, in each science there has been a battle over the existence of oscillatory phenomena, eventually resulting in the grudging acceptance of oscillation as a fundamental mode of behavior (Garfinkel 1983).

Oscillations in Biochemistry

A typical conflict over the existence of oscillation took place in biochemistry. In 1958, while working in the Soviet Union, chemist B.P. Belousov studied the reduction of bromate by malonic acid, a well-known laboratory model for the Krebs cycle. He saw something remarkable. The colorless liquid turned yellow, then, a minute or so later, turned colorless again, and then a minute or so after that, turned yellow again. It kept up this oscillating behavior for hours. The first reliable oscillatory chemical reaction had been observed (Figure 4.1).

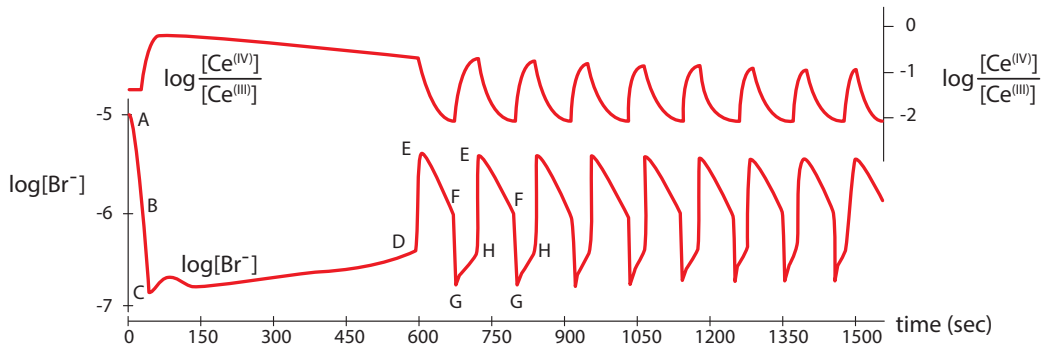


Figure 4.1: Oscillations in reaction products in the Belousov reaction. Redrawn with permission from "Oscillations in chemical systems II: Thorough analysis of temporal oscillation in the bromate–cerium–malonic acid system," by R.J. Field, E. Koros, and R.M. Noyes, (1972), *Journal of the American Chemical Society* 94(25):8649–8664. Copyright 1972 American Chemical Society.

When he tried to publish his results, he met a stone wall of rejection: such a thing as an oscillatory chemical reaction is not even possible, he was told, because it violated the Second Law of thermodynamics, which says that entropy increases with time in every chemical reaction, and therefore perpetual oscillation is impossible. What the critics failed to grasp was that no one was claiming to have found a perpetual oscillator, only one that oscillates *for a long time*. This violates the ideology of "equilibrium," but there is nothing physically wrong with the concept of a process that oscillates for a long time, by importing energy and exporting waste (for example, you). Indeed, the 1977 Nobel Prize in Chemistry was awarded for "contributions to nonequilibrium thermodynamics", including a thermodynamic theory of oscillatory chemical reactions.

Oscillations in Physiology

Body temperature. In all mammals, body temperature shows a clear 24-hour rhythm, whose amplitude can be as much as 1°. This daily rhythm is not the result of simple external cues such as the light–dark cycle, because it persists even in continuous darkness (Figure 4.2).

Hormones. Virtually all mammalian hormones show oscillatory behavior at a number of time scales. This is true of men as well as women. The dynamics of estradiol, the principal estro-

gen, displays oscillations at the 1-to-2-hour scale as well as the 12-hour scale. Note that the oscillations have a much larger amplitude during the daytime (Figure 4.3).

Gene expression. Genes are often under regulation that causes them to express in an oscillatory pattern, with cycles ranging from hours to days (Figure 4.4). Oscillatory gene expression has been detected in many genes, including *Hes1*, which is critical in neural development, and p53, the "guardian angel gene," which is critical in cancer regulation. These oscillations include circadian (24-hour) rhythms, and other higher-frequency rhythms.

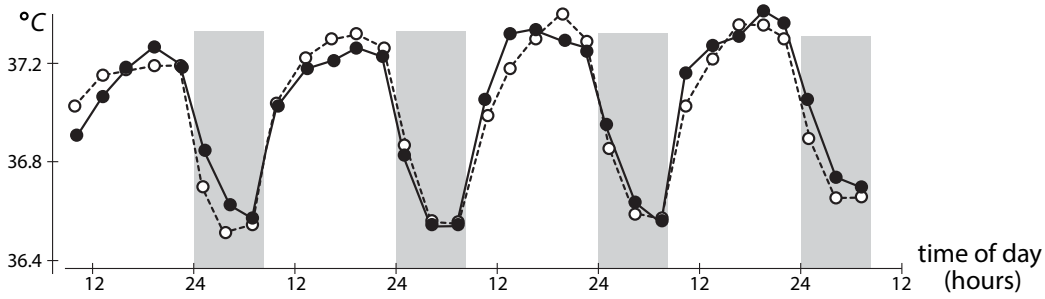


Figure 4.2: Four days of core body temperature (measured rectally) in human subjects. Researchers plotted the average of six human volunteers over four days. Closed circles represent the condition of an artificial light–dark cycle, while the open circles represent the same individuals in continuous darkness. Shaded areas are sleep times. Redrawn from "Human circadian rhythms in continuous darkness: entrainment by social cues," by J. Aschoff, M. Fatranska, H. Giedke, P. Doerr, D. Stamm, and H. Wisser, (1971), *Science* 171(3967):213–15. Reprinted with permission from AAAS.

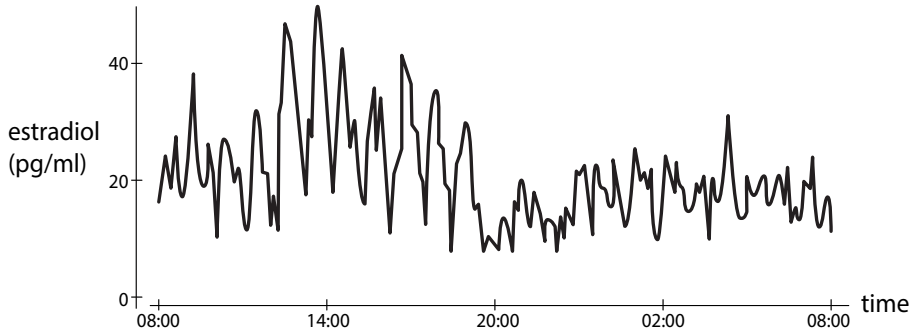


Figure 4.3: Multifrequency oscillations in estradiol in a 25-year-old normal female, mid-to-late follicular phase. Redrawn with permission from "Synchronicity of frequently sampled, 24-h concentrations of circulating leptin, luteinizing hormone, and estradiol in healthy women," by J. Licinio, A.B. Negrão, C. Mantzoros, V. Kaklamani, M.-L. Wong, P.B. Bongiorno, A. Mulla, L. Cearnal, J.D. Veldhuis, and J.S. Flier, (1998), *Proceedings of the National Academy of Sciences* 95(5):2541–2546. Copyright 1998 by National Academy of Sciences, U.S.A.

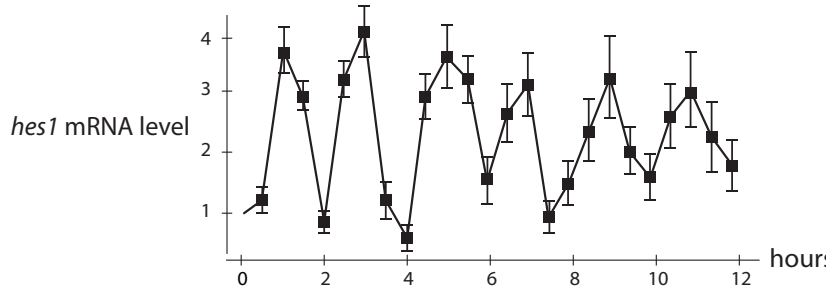


Figure 4.4: Two-hour oscillations in the expression of the gene *Hes1*. Redrawn from "Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop," by H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K. Yoshikawa, and R. Kageyama, (2002), *Science* 298(5594):840–843. Reprinted with permission from AAAS.

Transient Versus Long-Term Behavior

The existence of oscillation must be accepted as a fact. But how are we to understand it and model it mathematically?

We want to say that these systems are in a kind of "dynamic equilibrium," but we don't yet have a way to say this mathematically. We will now develop the mathematical concept corresponding to this oscillatory type of "dynamic equilibrium."

In order to model this concept of equilibrium, we have to make a distinction between *transient behavior* and *long-term behavior*.

When we look at the dynamics of a system, there are two different questions we might be interested in. We can think of them roughly as short-term versus long-term behavior.

Short-term behavior (*transients*). When we start a system with a given initial condition, the system immediately begins to react. This initial short-term response is called *transient*, which can be either an adjective or a noun. For example, if we look at an epidemic population model of susceptible–infected type, we might set S_0 and I_0, the initial numbers of the two populations, and then want to know how the system immediately responds: does the infection get larger or smaller?

Long-term behavior (*asymptotics*). More often, we are interested in the system's long-term behavior pattern, because that is usually what we observe. If we are studying neurons, the heart, metabolic systems, or ecosystems, we are typically looking at a system that has settled into a definite long-term behavior. This behavior "as t approaches infinity" is called the *asymptotic behavior* of the system.

We are therefore led to make a definition, to try to capture the idea of "long-term behavior." If X is the state space of a dynamical system, then we define an *attractor* of the dynamical system as

(1) a set A contained in X such that

(2) there is a neighborhood of initial conditions that all approach A as t approaches infinity.

Let's unpack that. "A set A contained in X," refers to a collection of points in state space. This could be one point, or a curve, or a more complex shape. And "there is a neighborhood of initial conditions that all approach A" just means that if you start close enough to A, you will eventually approach it. We are deliberately not stating just how close "close enough" is, because this can be very different for different attractors.

Exercise 4.1.1 What concept have you previously encountered that describes the neighborhood (to be precise, the largest such neighborhood) in this definition?

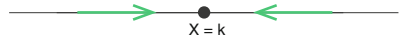
An **attractor** of a dynamical system on the state space X is a set A contained in X such that for a neighborhood of initial conditions X_0, the trajectories going forward from X_0 all approach A, that is,

$$\text{the distance } d\Big(X(t),\ A\Big) \to 0 \text{ as } t \to \infty$$

We have already seen examples of attractors, namely, the stable equilibrium points of Chapter 3. Think about the model of a population with crowding,

$$X' = bX - \frac{b}{k}X^2$$

and recall the behavior at and near $X = k$:



In other words, the point $X = k$ satisfies the definition of an attractor:
(1) it is a set (consisting of one point) in X,
(2) and for all points in a neighborhood of $X = k$, the flow is toward $X = k$.

Exercise 4.1.2 What is the largest neighborhood of $X = k$ for which this is true?

As $t \to \infty$, every initial condition around $X = k$ approaches the point $X = k$. Therefore, $X = k$ is called a *point attractor*. Note that the state point gets closer and closer to $X = k$ without actually ever reaching or touching it. This is called approaching $X = k$ asymptotically.

Exercise 4.1.3 Draw a vector field for a one-dimensional system with three attractors.

Another example is the spring with friction (Figure 4.5). Look at the equilibrium point $(0, 0)$. Note that in a neighborhood around $(0, 0)$, all initial conditions flow to $(0, 0)$ as $t \to \infty$. Thus, in this system, the point $(0, 0)$ is a point attractor.

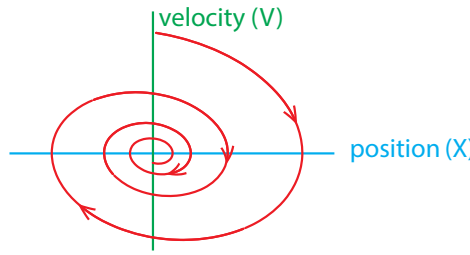


Figure 4.5: Point attractor in the model of a spring with friction.

The simplest attractor is a point. "Point attractor" is another name for "stable equilibrium point," and it is a model for equilibrium control of systems.

Stable Oscillations

We've already seen some models that produce oscillation, including the frictionless spring and the shark–tuna model.

However, these models are not good models for biological oscillations. The biggest problem with them is that they are not *robust*. In both of these models, the behavior depends forever on the initial condition. If you are on a trajectory and are perturbed even slightly, there is no return to the original trajectory. The system "remembers" the perturbation forever.

This is generally undesirable in a biological system. For instance, the body temperature rhythm should be stable to perturbations: if you have a fever one day, you want to be able to return to the normal oscillation.

In order to understand how to model these kinds of "robust" oscillations, we have to think a little bit about dynamical systems. It turns out that dynamics gives us a perfect language to talk about this concept.

First of all, we need to mathematically define the concept of oscillation. There are two ways to look at it: 1) in the time series of a variable, and 2) in the state space trajectory.

1) If X is a state variable, the function $X(t)$ is an oscillation if and only if it is periodic; that is, if there is a constant P (called the *period* of the oscillation) such that for all times t, $X(t+P) = X(t)$. In other words, the function $X(t)$ repeats itself after P time units.
2) In state space, a trajectory represents an oscillation if and only if it is a closed loop, which is often referred to as a closed orbit.

Exercise 4.1.4 Why does the first condition being true mean that the second must be true? Why does the second being true mean that the first must be true?

But is this sufficient to capture the notion of "dynamic equilibrium"? No, there is one more very important piece to the definition. In the shark–tuna system and the frictionless spring, behaviors were indeed represented by closed orbits in state space. However, when perturbed slightly, the behavior goes to a different oscillation from the one that existed before the perturbation. The new oscillation neither approaches the original oscillation nor moves away from it. We say that these oscillations are *neutrally stable*.

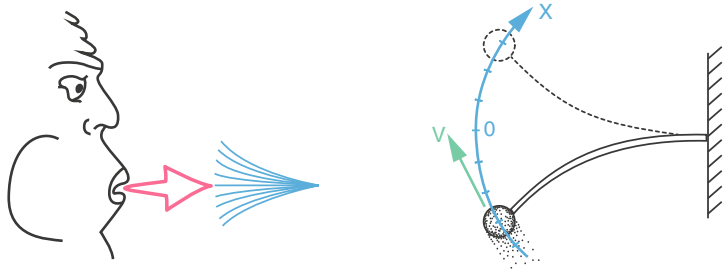
What we need are models for oscillations that are stable. Stable oscillations are better models for biological oscillations than the neutrally stable oscillations in the shark–tuna and frictionless spring models.

The concept of an attractor gives us a perfect definition of a stable oscillation. We can now define a periodic attractor.

A *periodic attractor* is an attractor that is a closed orbit, also called a stable limit cycle, or limit cycle attractor.

Rayleigh's Clarinet: A Stable Oscillation

A beautiful set of examples of stable limit cycles can be found in the pioneering work by Lord Rayleigh (1842–1919) on the physics behind musical instruments. Here we present his analysis of the clarinet reed. Our account closely follows the excellent presentation in Abraham and Shaw's *Dynamics: The Geometry of Behavior* (Abraham and Shaw 1985).



Rayleigh modeled the reed of the clarinet as a thin, flexible wand attached to a solid object, with a mass on its end. The clarinetist supplies energy to the system by blowing along the long axis of the wand.

Without the clarinetist, the system is simply a spring with friction (from air resistance), and it produces a spiraling in trajectory (Figure 4.6, left). If we bend the reed up or down, it will oscillate in a damped manner and eventually return to the equilibrium position. This behavior can be modeled using Hooke's law ($F_s = -k_1 X$, with $k_1 = 1$) with simple linear friction ($F_f = k_2 V$, with $k_2 = 1$). Assuming the mass $m = 1$, we get

$$X' = V$$
$$V' = F_s - F_f = -X - V$$

This gives us exactly the behavior of a spring with friction, namely, a spiraling in to a stable equilibrium point.

When the clarinetist blows on the reed, the situation is changed. Rayleigh reasoned that blowing supplies energy to the system and therefore acts like the opposite of friction, or in other words, like "negative friction." Thus, for this system, the function that relates "friction" to velocity has a negative slope, which results in a spiraling out of the trajectory (Figure 4.6, right).

Exercise 4.1.5 Write the equations for a spring with "negative friction."

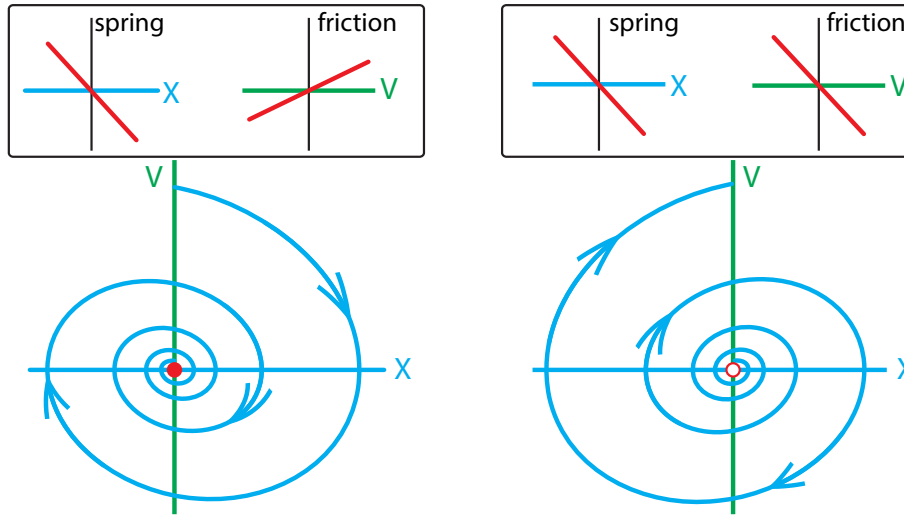


Figure 4.6: Left: When the friction force is positive, the system has a point attractor of spiral type at (0, 0). Right: When the friction is negative, the origin becomes a spiral type unstable equilibrium.

Of course, a trajectory that spirals out forever isn't realistic. What actually happens is that if the wand is moving slowly (V is small), then blowing on it will actually accelerate it, so the force of the breath is in the same direction as the motion and adds energy to the system. But if the velocity of the wand is high, the blowing produces conventional friction (due to air resistance), which retards the motion. So how do we model this? Rayleigh needed a function of V that had a negative slope (negative friction) for small values of V and a positive slope (positive friction) for large values of V. The simplest way to do this is with a cubic function like

$$F_f = (V^3 - V)$$

(Figure 4.7).

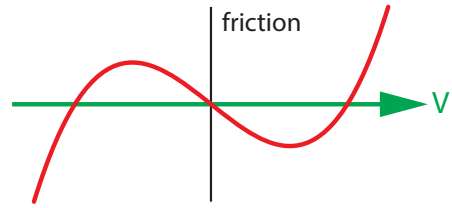


Figure 4.7: A hypothetical nonlinear friction force.

What behavior results from this nonlinear friction? Rayleigh reasoned in state space. He argued that since the small-V behavior produces a spiraling out, and the large-V behavior produces a spiraling in, between these two there must be a single closed orbit trapped between the other two kinds of trajectories. (This was not proved until 50 years later, by Poincaré, using his new invention, topology.)

This new kind of friction then gives us a new differential equation:

$$X' = V$$
$$V' = -X - (V^3 - V)$$

A simulation of this equation results in the trajectories shown in Figure 4.8, right.

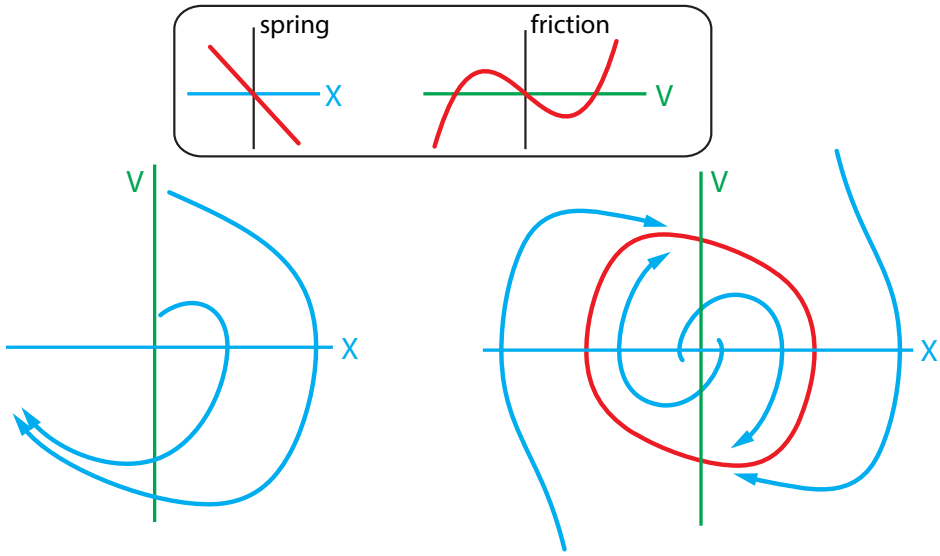


Figure 4.8: Upper: spring force and friction force for the Rayleigh clarinet model. Lower Left: Two representative trajectories for this model. Lower Right: All trajectories, from any initial condition except (0, 0), approach the red loop asymptotically.

Consider the closed orbit shown in red. Note an interesting fact about it, which we have not seen before: if you choose an initial condition that is not on the red loop, **the ensuing trajectory will get closer and closer to the red loop, and will approach it as $t \to \infty$**. This is true whether you are inside the red loop or outside it; all trajectories, with the exception of the one point at $(0, 0)$, approach the red loop arbitrarily closely.

In other words, **the red loop fits the definition of an attractor**. It is our first example of a closed orbit attractor, or periodic attractor. A third name for these is based on the idea that just as an equilibrium point is a limit point, the red loop is a *limit cycle*, and so these are called *limit cycle attractors*.[1] Note that another name for the red loop is a *stable limit cycle*. It is stable in exactly the same sense as a stable equilibrium point: if you perturb the system off the cycle, the behavior returns to the cycle. So it really is an attractor.

Exercise 4.1.6 Sketch a phase portrait that shows an *unstable* limit cycle.

We said that closed orbit attractors are better models for biological oscillations. They are also better models for musical instruments: we want the character of the musical note to be stable

[1] Some sources refer to all closed trajectories as "limit cycles." On the other hand, a few reserve the term for stable closed trajectories.

under small changes. For example, when we blow harder, we want the quality of the note and its frequency to be stable, and only its amplitude to change. Now, the *quality* of the note, what musicians call the timbre, is what makes a trumpet playing a note sound different from a guitar playing the same note. What gives a note its quality is the overtones, or higher harmonics of the fundamental frequency. These harmonics show up in the trajectory by giving the oscillation a noncircular shape.

Let's model "blowing harder." Rayleigh suggested that it can be modeled by changing the "friction" term, so that the negative friction region is broader. For example, if we take $F_f = 0.5V^3 - V$, then we get the solid limit cycle shown in Figure 4.9.

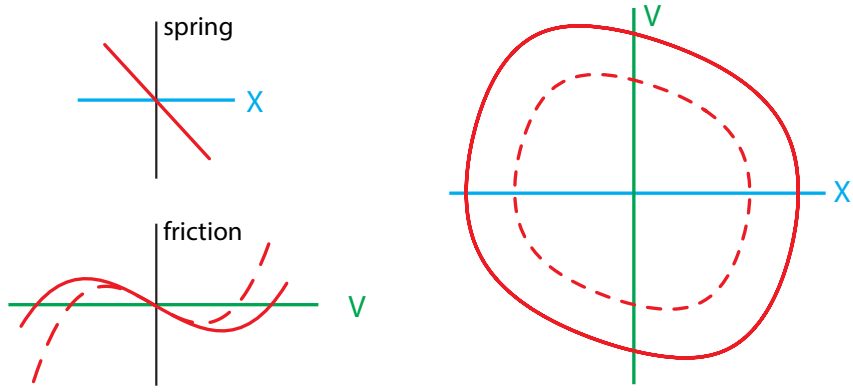


Figure 4.9: Blowing harder. Left: The solid lines show the forces in the Rayleigh clarinet model, under the "blowing harder" condition. The dotted line represents the model without blowing harder. Right: Limit cycle attractors for the two models. Note similarity of shape.

Note that it has the same shape as the smaller one. This is important, because it is the shape of the trajectory that gives the instrument its characteristic sound. Additionally, the fundamental frequency is unchanged, which is also critical in a musical instrument.

Exercise 4.1.7 Simulate the model

$$X' = V$$
$$V' = -X - (aV^3 - V)$$

for three different positive values of a and compare the trajectories and time series.

A little research will turn up fascinating mathematical models of other musical instruments. Abraham and Shaw's beautiful book has several of them (Abraham and Shaw 1985). Models of the clarinet reed have been improved since Rayleigh's time, and many other musical instruments have been mathematically modeled.

Further Exercises 4.1

1. Is a saddle point an attractor? Justify your answer.
2. Does a trajectory that approaches a limit cycle attractor ever reach the attractor? Explain.
3. Give an example of an equilibrium concept from science or everyday life (other than those described in the text) and describe what aspects of system behavior it captures and what it fails to capture.
4. Describe jet lag and recovery from it in dynamical terms.
5. Sketch an *unstable* limit cycle. If the limit cycle has a single equilibrium point (and no other limit cycles) inside it, what kind of equilibrium must the point be?
6. Suppose a 2D system has a stable equilibrium point that is located somewhere outside a limit cycle. Can a trajectory starting inside the limit cycle reach this point? Justify your answer. *(Hint: It may help to draw the situation.)*
7. Suppose you are studying a system of differential equations, and you find an unstable spiral equilibrium point. You also find a trajectory that makes a complete loop around that equilibrium point. In a 2D state space, these conditions usually cause that "loop trajectory" to be a limit cycle attractor.
   a) If the state space is three-dimensional, does the loop have to be a limit cycle attractor? Explain.
   b) Can you think of a way that these conditions could occur in a 2D state space so that the loop is not a limit cycle attractor? Explain. A picture is a good idea. *(Hint: It can happen, but it's extremely unlikely.)*

4.2 Mechanisms of Oscillation

As we begin to model oscillatory phenomena in nature, we will see some common themes across all of our models. In particular, there are typical causes or mechanisms for stable oscillatory behavior. The two most important are *steep negative feedback* and *time delays*.

The Hypothalamic/Pituitary/Gonadal Hormonal Axis

Let's start by examining hormone oscillations (Figure 4.3). An elementary model of an endocrine control system was first proposed by W. Smith (Smith 1983).

The gonads (ovaries in females, testes in males) secrete hormones, called estradiol and progesterone in females and testosterone in males. For simplicity here, we will assume that it is one hormone, which we will call G (for gonad). What makes the gonads secrete their output? They are under the control of two hormones made by the pituitary, luteinizing hormone (LH) and follicle-stimulating hormone (FSH). These hormones stimulate the gonads: the more LH and FSH the pituitary makes, the more G the gonads make. As another simplifying assumption, we'll model a single generic pituitary hormone, which we'll call P.

If the pituitary gland controls the gonads, what controls the pituitary gland? In the 1970s, it was discovered that the pituitary (which is in the head but not technically in the brain) is actually under the control of the brain. The hypothalamus, a part of the brain located a millimeter away from the pituitary, secretes releasing factors that cause the pituitary to secrete its hormones. The hypothalamic factor relevant to the system we are studying is gonadatropin-releasing hormone, which we'll call H (for "hypothalamus"). The more H is secreted by the hypothalamus, the more P is secreted by the pituitary.

Where is this chain of glands driving glands going to end? It ends by closing the loop. The hypothalamus senses the circulating levels of G and responds to high levels of G by down-regulating its output of H. Figure 4.10 summarizes the situation.

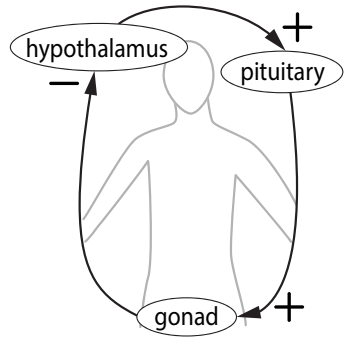


Figure 4.10: In mammals, the Hypothalamic-Pituitary-Gonad system forms a negative feedback loop.

We can now specify a few dynamical assumptions and start writing the differential equations for this system. Earlier, we said that H stimulates the production of P, and P stimulates the production of G. We will assume that this stimulation is directly proportional to the concentration of the stimulating hormone, with proportionality constant 1. Furthermore, we'll assume that the decrease in hormone concentration caused by that hormone is proportional to the concentration of that hormone. The equations we now have are

$$H' = \text{☁} - k_1 H$$
$$P' = H - k_2 P$$
$$G' = P - k_3 G$$

The cloud symbol ☁ in the equation for H' represents an unknown function of G that decreases as G increases but never goes negative. One possibility for such a function is the family of *decreasing sigmoids*

$$\text{☁} = \frac{1}{1+G^n}$$

shown in Figure 4.11.

Notice that for our negative feedback function, we have chosen a function that is never negative! The term "negative feedback" actually encompasses two somewhat different types of behavior. In the more straightforward case, an increase in some quantity leads to an actual decrease in that quantity. The examples we have seen so far fall into this class. The second kind of negative feedback is a bit more subtle. It occurs when the feedback loop cannot actually take away from the quantity in question but can decrease its growth rate. An example of this

is seeing your bank account balance get low and curtailing your spending in response. Even if you reduced spending all the way to zero, this could not actually increase the amount of money in your account. Spending reductions do, however, slow down the decline of your bank balance. Here, we see a biological example of this kind of negative feedback. It is a biological fact that the hypothalamus can secrete only H. It can't suck H back up! So the form of the negative feedback has to be the second kind; it has to be modeled by a function that is declining but never negative.

The shape of this function depends on n, as shown in Figure 4.11. Notice that the middle portion gets steeper; that is, it is more sensitive to changes in G as n increases. Here we will choose a relatively steep value, let's say $n = 9$. Thus, the overall equations are

$$H' = \frac{1}{1+G^n} - k_1 H$$
$$P' = H - k_2 P$$
$$G' = P - k_3 G$$

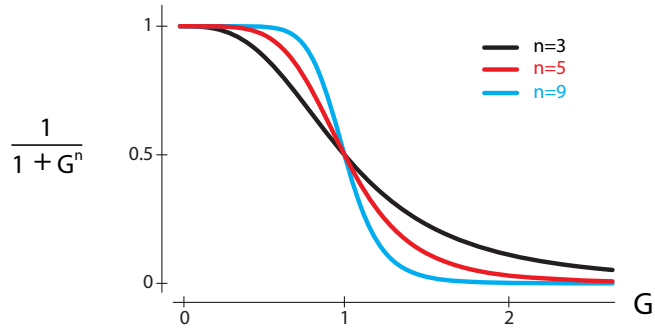


Figure 4.11: Negative feedback functions, with varying steepness.

A simulation of this model, using $k_1 = k_2 = k_3 = 0.2$, and $n = 9$, shows clear oscillations; Figure 4.12.

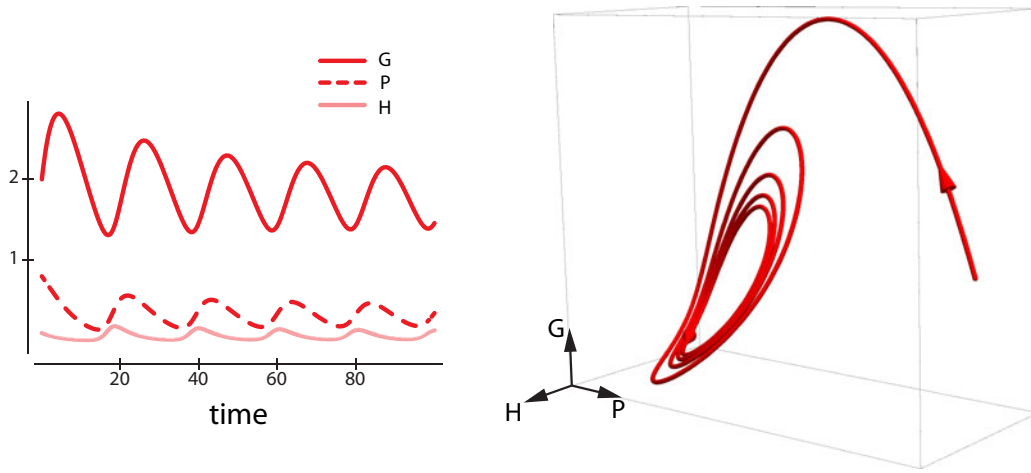


Figure 4.12: Limit cycle attractor in the H/P/G model.

Notice that all three hormones oscillate. The trajectory approaches a closed loop attractor, which is the steady state for the system. If we performed the experiment of starting at a variety of initial conditions, we would see a remarkable fact: all trajectories approach the same closed loop attractor. And if we perturbed the system off the closed loop attractor, it would quickly return to it. Thus, this is a stable oscillation in the endocrine system.

Exercise 4.2.1 Verify that for values of n less than 8, the system goes to a stable equilibrium, but as n passes 8, the equilibrium point becomes unstable, and a stable oscillation is created.

Exercise 4.2.2 Verify that a variety of initial conditions all approach the same limit cycle attractor in the H/P/G system.

Highly sensitive negative feedback loops are one of the major causes of oscillations in biological systems. To see why steep negative feedback results in oscillatory behavior, imagine a parent teaching a teenager to drive. The teen is trying to keep the car in the center of the lane, and the parent tells them to correct right or correct left, as appropriate. This is an example of a negative feedback loop. If the parent's sensitivity to the car's position is reasonable, the car will travel in a fairly straight line down the center of the lane. But what happens if the parent yells, "go right" when the car drifts a little bit to the left? The startled teenager will overcorrect, taking the car too far to the right. The parent will then start yelling, "go left," the teen will overcorrect again, and the car will oscillate back and forth, as illustrated in Figure 4.13.

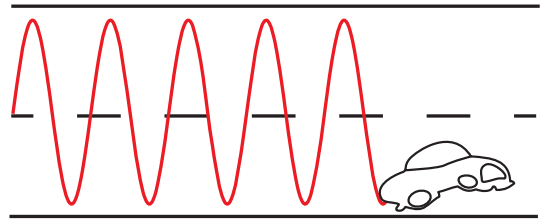


Figure 4.13: Schematic of the behavior of a car whose driver is under very steep feedback control. The driver overcorrects in each direction.

While it is clear that steep negative feedback is a cause of these oscillations, it is important to understand that it is not sufficient by itself to produce these oscillations. To see why, consider an even simpler negative feedback model. Let's eliminate the middleman between H and G, and assume that the hypothalamic feedback could somehow be applied instantaneously to the gonad. In other words, let H control G directly, resulting in a new model:

$$H' = \frac{1}{1+G^n} - k_1 H$$
$$G' = H - k_3 G$$

This negative feedback model will not oscillate, no matter how steep the feedback.

Exercise 4.2.3 Verify this assertion.

The reason is that eliminating the middleman eliminated a key *time delay* in the process that was necessary to generate oscillation. In this case, the time delay is created by the fact that the hypothalamus must change the pituitary, and then the pituitary changes the gonad.

While steep negative feedback is an important cause of oscillation in this system, it is also important to remember that *time delays* also play a role.

Respiratory Control of CO_2

This endocrine time delay is modeled by having intermediate steps in the process. There is another way to model time delays—explicitly.

The explicit approach involves writing differential equations in which the rate of change of the state variable is a function of the value of that variable some time ago. For example, we might have $X'(t) = 2X(t-5)$, where $X(t-5)$ is the value of X at a time 5 time units before the present time. Such equations, which explicitly include time delays, are called *delay differential equations*. The value of the delay is commonly written τ (the Greek letter tau), so it's common to see expressions such as $X(t-\tau)$.

Exercise 4.2.4 In the delay differential equation $Y'(t) = 16Y(t-2) + 8Y(t)$, what does $Y(t-2)$ refer to? What does $Y(t)$ refer to?

One important delay differential equation in biology is the Mackey–Glass model of respiratory control of CO_2 (Mackey and Glass 1977). One function of breathing is to control the concentration of carbon dioxide in the blood, a quantity we will represent with the variable X. This is carried out by increasing the breathing rate when CO_2 is high, thereby shoveling out more CO_2. The control of the breathing rate (also called the ventilation rate) is carried out by chemoreceptors in the brain, which send instructions to the nerves controlling the lung.

Now let's make a model of this process, which is essentially going to be

$$\begin{aligned} X' &= \text{things that increase } CO_2 - \text{things that decrease } CO_2 \\ &= \text{body metabolism} - \text{ventilation} \end{aligned}$$

Let's assume that the body's rate of metabolic production of CO_2 is a constant, which we'll call L.

Now we need to model the effect of ventilation. Carbon dioxide is excreted by the lungs; each breath has a volume of CO_2 that depends on the current CO_2 concentration in the blood in the lung, which is the variable X. So then the rate of excretion of CO_2 is equal to

$$CO_2/\text{breath} \quad \times \quad \text{breaths/minute}$$

The term "breaths/minute" in the excretion of CO_2 from the lungs is the ventilation rate V, which is controlled by CO_2 concentration in the blood. When the CO_2 concentration is low, the ventilation rate is low, but when CO_2 is high, the ventilation rate is close to the maximum. We need a function that summarizes this. A.V. Hill, the physiologist who first studied this, used a function that has become so popular that in physiology it is now called a "Hill function."[2] It is

[2] In ecology, the same function is sometimes called the "Holling Type III function" and is used to model the feeding behavior of vertebrates.

the family of *increasing* sigmoid functions

$$Y = \frac{X^n}{1 + X^n}$$

For increasing values of n, the function gets steeper and steeper, as shown in Figure 4.14. We would therefore like to write the model as

$$\begin{aligned} X' &= L - V \cdot X \\ &= L - \frac{V_{max} \cdot X^n}{1 + X^n} \cdot X \end{aligned}$$

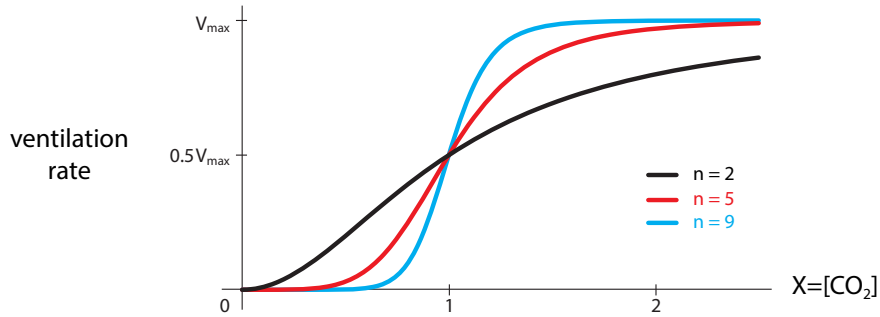


Figure 4.14: Three examples of the Hill function for ventilation, $\frac{V_{max} \cdot X^n}{1+X^n}$.

We use V_{max} to scale the sigmoid function so that its maximum value is the maximum ventilatory rate, called V_{max}.

Exercise 4.2.5 What aspect of the function does V_{max} control?

There is one problem with this, however. There is an X in the ventilation rate Hill function, and there is an X that it is multiplying, but they are not the same X! There is a **delay** between gas exchange in the lungs and the effect on CO_2-monitoring neurons in the brain. In simple terms, it takes time for blood to get from the lungs to the brain. Therefore, the brain is responding not to the current CO_2 concentration in the lung but to the concentration some time ago. (In the body, this delay is on the order of 0.2 minutes.) Thus, the ventilation rate function really needs to be

$$V = V_{max} \cdot \frac{{X_\tau}^n}{1 + {X_\tau}^n}$$

where X_τ is the time-delayed value $X(t-\tau)$, the value of X at time τ time units ago. With this addition, the Mackey–Glass equation becomes

$$X' = L - \frac{V_{max} \cdot {X_\tau}^n}{1 + {X_\tau}^n} \cdot X$$

The state variable is X, but we are most interested in the quantity V, the ventilation rate. For low values of n and τ, the system goes to a stable equilibrium. When X is in equilibrium, so is V, and the result is a steady breathing rate. But if we increase n or τ (or both), the

model starts to oscillate (Figure 4.15), with the breathing rate waxing and waning over 30 seconds. These oscillations in breathing rate, called *Cheyne–Stokes breathing*, are observed in heart failure patients as well as those with stroke or other neurologic conditions (Figure 4.16). Heart failure patients have longer circulation times, due to low pumping efficiency, and so have higher values of τ, while stroke patients often suffer from "hyperreflexia," in which reflex reactions are exaggerated, and therefore can be modeled as having an increased n.

Exercise 4.2.6 Let

$$X' = 6 - \frac{16 \cdot X(t-0.2)^5}{1 + X(t-0.5)^5} \cdot X$$

Assume that for all $t \leq 0$, $X(t) = 0.5$. Use Euler's method with a step size of 0.1 to approximate $X(0.3)$.

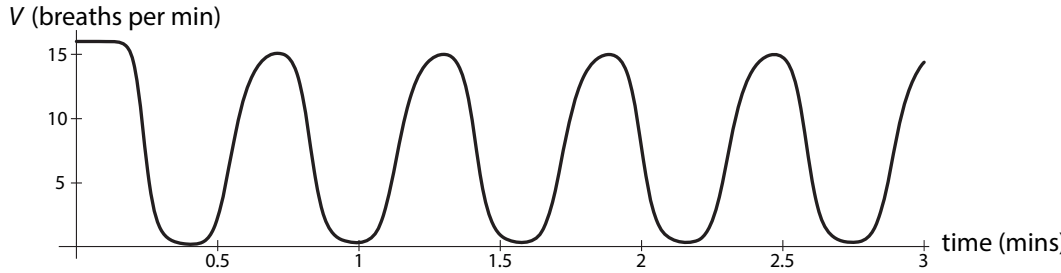


Figure 4.15: A simulation of the Mackey–Glass respiration model developed in the text, with $L = 6$, $V_{max} = 16$, $n = 5$, and $\tau = 0.2$.

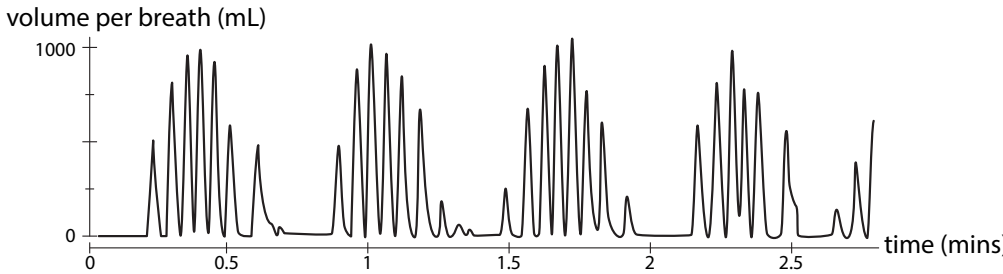


Figure 4.16: Cheyne–Stokes breathing in a spinal cord injury patient. Redrawn with permission from "Sleep disordered breathing in chronic spinal cord injury," by A. Sankari, A. Bascom, S. Oomman, and M.S. Badr, (2014), *Journal of Clinical Sleep Medicine* 10(1):65–72. Copyright 2014 American Academy of Sleep Medicine.

You should be surprised to see oscillations coming from a single-variable model. (Why?) The reason this is possible is that the state of a delay differential equation is not just the current value of the variable. Proceeding from one integration step to the next in a delay differential equation requires information about the value of the variable τ time units ago. Consequently, delay differential equations are actually infinite-dimensional, since we need to know the whole history of values, information about an infinite number of points, to simulate them. This allows delay differential equations to display behaviors that are otherwise possible only in two or three dimensions.

The kinds of delays modeled by delay differential equations are what we might call "transfer delays." For example, the Mackey–Glass model contains a delay because it takes time for blood to get from the lungs to the brain. However, delays in negative feedback loops can cause oscillations even without an explicit delay in the equations. The HPG model contains such a "process delay."

Exercise 4.2.7 Verify that both n and τ must be sufficiently large for oscillation to happen in this system.

Muscle Tremor

The same dynamics are at work in many cases in which oscillation is a pathology. Consider the simplest type of control system in skeletal muscle: the monosynaptic stretch reflex. Muscles contract because they are given an electrical signal from the controlling neurons, called motor neurons. There is a negative feedback loop that regulates muscle position and helps the muscle maintain a constant position in space: when a skeletal muscle is stretched by external forces, I_a sensory neurons register this stretch and increase their signaling to the primary α-motor neuron (in the spinal cord) governing that muscle. This results in the motor neuron increasing its firing, which results in the muscle contracting. Thus there is a negative feedback loop (Figure 4.17).

increase in L $\longrightarrow$ increased stretch reflex firing $\longrightarrow$ increased motor neuron firing $\longrightarrow$ decreased L

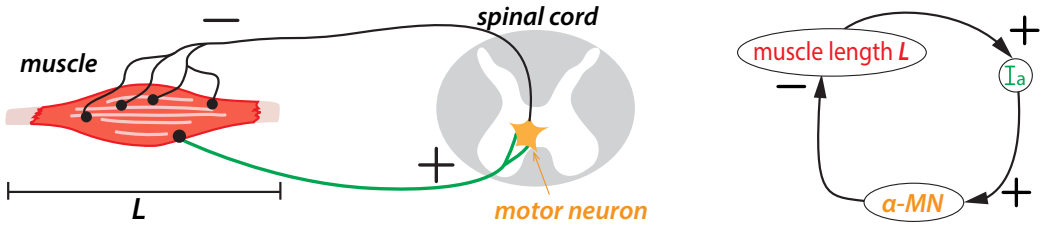


Figure 4.17: Left: There is a simple stretch reflex arc that runs from a muscle to the motor neurons that control it. Right: Schematic of the arc shows that it is a negative feedback loop.

Under normal conditions, this negative feedback loop maintains a fairly steady muscle position. But in many pathological conditions, the steady state of the limb is lost, and pathological oscillations result, called tremor (Figure 4.18, Figure 4.19).

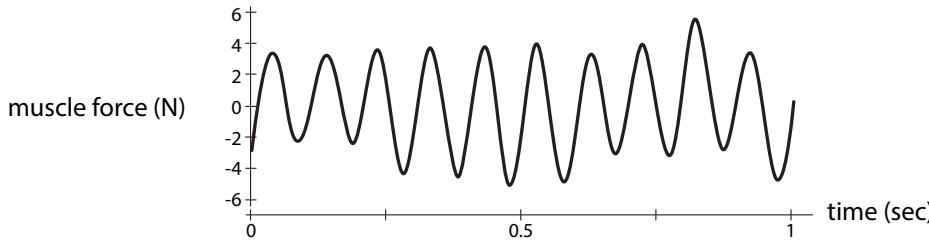


Figure 4.18: Stretch reflex-induced oscillation in the force at the elbow joint of a normal human subject. The reflex has been enhanced by a spring load. Redrawn from "Alpha band cortico-muscular coherence occurs in healthy individuals during mechanically-induced tremor," by F. Budini, L.M. McManus, M. Berchicci, F. Menotti, A. Macaluso, F. Di Russo, M.M. Lowery, and G. De Vito, (2014), *PloS one* 9(12):e115012. Copyright 2014 Budini et al.

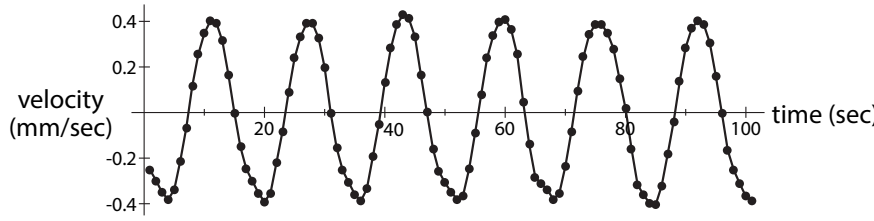


Figure 4.19: Parkinsonian tremor in the index finger of a patient (subject v4) off medication. Drawn from data provided in supplement to Beuter et al. (2001) https://www.physionet.org/physiobank/database/tremordb/.

Dynamical systems theory can give us an insight into the mechanisms behind tremors. As we have seen, there are two kinds of factors that can cause a negative feedback system to go into oscillation: steep slopes and increased time delays.

Both of these occur in various pathologies that exhibit tremor. For example, multiple sclerosis (MS) is a disease in which the insulation of the neuron becomes damaged, leading to slower conduction and hence increased time delay in the system. MS patients suffer from muscle tremor, and it is very tempting to speculate that this might be the mechanism.

Another group of patients that exhibit muscle tremor are stroke patients. Here, the mechanism is different: one of the roles the brain plays when healthy is to suppress the sensitivity of peripheral reflexes. But in stroke, which is caused by a burst or clogged artery in the brain, that suppression is lost, and there is a resulting "hyperreflexia" (similar to that in respiration) in the stretch reflex, resulting in stroke-related tremor.

Oscillations in Insulin and Glucose

Insulin is a hormone that is released by the pancreas in response to a rise in blood glucose, for example after a meal. The insulin then facilitates the entry of glucose into muscle cells, where it is metabolized. The dynamics of "glucose makes insulin go up, insulin makes glucose go down" is then a classic negative feedback loop.

The dynamics of glucose and insulin were first studied in a mathematical model by Sturis et al. Their paper, called "Computer model for mechanisms underlying ultradian oscillations of

glucose and insulin," was the first to explain insulin–glucose oscillations as emerging from the feedback dynamics of the insulin–glucose system itself (Sturis et al. 1991b). Following the logic of their analysis, insulin (I) is increased by glucose in a saturating manner, and is decreased by the usual degradation, giving us

$$I' = \underbrace{\frac{k_1 \cdot G^4}{1+G^4}}_{\text{glucose spurs insulin production by the pancreas}} - \underbrace{k_2 \cdot I}_{\text{degradation of insulin}}$$

Glucose (G) is changed by four factors:

- G is increased by external sources (such as meals).
- G is also increased by glucose production by the liver. This production is inhibited by insulin (I).
- G is degraded at a rate k_4.
- G combines with I in the muscle to metabolize G.

This gives us the G' equation as

$$G' = \underbrace{\frac{k_3}{1+I^2}}_{\text{Insulin inhibits glucose production in the liver}} + \underbrace{Ext}_{\text{external glucose (meals)}} - \underbrace{k_4 \cdot G}_{\text{degradation of glucose}} - \underbrace{G \cdot I}_{\text{insulin facilitates glucose utilization by muscle}}$$

Parameters : $k_1 = 1, k_2 = 0.1, k_3 = 1, k_4 = 0.1, Ext = \begin{cases} 5, & \text{if } 1 < t < 2 \\ 0, & \text{otherwise} \end{cases}$.

In this model, the transient intake of glucose results in a spike of insulin and then a return of both quantities to equilibrium values (Figure 4.20).

However, as Sturis et al. observe, this model is not physiologically realistic, because it assumes that the response of the insulin system to the rise in glucose is instantaneous. In fact, it takes time for the pancreas to respond to the rise in glucose. When we amend the model to include this time delay, we get a new I' expression:

$$I' = \frac{k_1 \cdot {G_\tau}^4}{1+{G_\tau}^4} - k_2 \cdot I$$

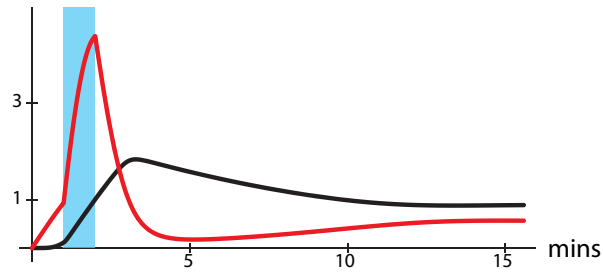


Figure 4.20: Glucose (red) and insulin (black) in response to an external dose of glucose (blue rectangle).

where τ is the time delay in the response. If we let $\tau = 15$ minutes, then the system goes into oscillation, even with a constant glucose infusion ($Ext = 1$), Figure 4.21.

And of course, insulin and glucose in the body actually do oscillate, as seen in a figure we saw in Chapter 1 (Figure 1.5 on page 5) and reprint here (Figure 4.22).

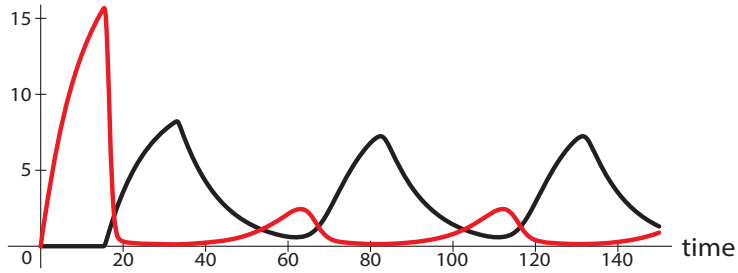


Figure 4.21: Glucose (red) and insulin (black) in response to a constant dose of glucose.

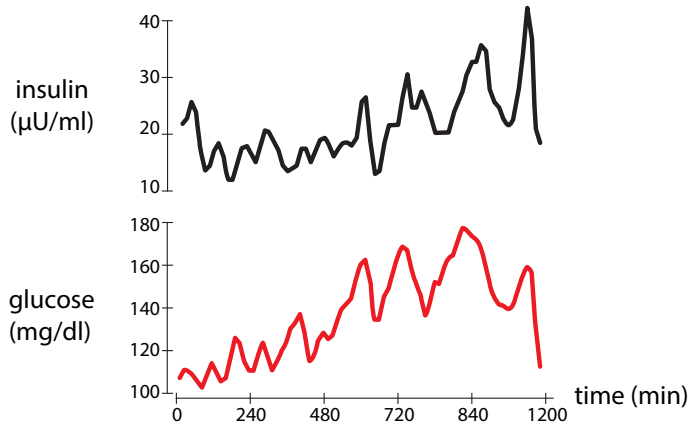


Figure 4.22: Insulin and glucose oscillations in a human volunteer under constant glucose infusion, (Sturis et al., 1991a). Redrawn from "Aspects of oscillatory insulin secretion," by J. Sturis, K.S. Polonsky, J.D. Blackman, C. Knudsen, E. Mosekilde, and E. Van Cauter, *In Complexity, Chaos, and Biological Evolution?* by E. Mosekilde and L. Mosekilde, eds., (1991), volume 270, pp. 75–93. New York: Plenum Press. Copyright 1991 by Plenum Press. With permission of Springer.

In these systems, the principal cause of oscillation is the introduction of time delays into the negative feedback system. We already spoke of steep negative feedback as a cause of oscillation (in the presence of some time delay). Our cartoon example of steep negative feedback was the hyperactive parent teaching a child to drive and causing constant overreaction that resulted in oscillation. We can make another cartoon example to illustrate the role of time delays (Figure 4.23).

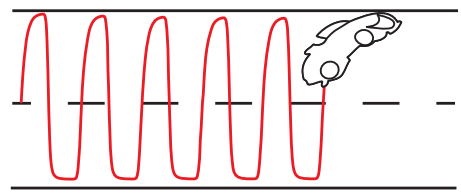


Figure 4.23: Schematic of the behavior of a car whose driver is under negative feedback control with a time delay.

Imagine another parent teaching a child to drive, only now the parent is inattentive; maybe the parent is texting. There is therefore a short delay before the parent responds to the car's drift.

Now, the car will also oscillate, because the driver will have drifted well to the left by the time the parent's corrective is issued. (Indeed, police officers look for drivers who are "weaving" down the road, because oscillations in the vehicle's path could well be a sign of the slower reflexes caused by alcohol consumption.)

Oscillatory Gene Expression

With so much physiology operating in an oscillatory manner, it should not be surprising to learn that in many critical physiological systems, gene expression operates in an oscillatory manner, because rhythmic gene expression has to be coordinated to, and in some cases actually drive, these rhythmic processes.

Therefore, cells have evolved mechanisms to produce oscillatory gene expression. Most of these mechanisms depend on some kind of negative feedback, where the gene produces a product that inhibits that very gene.

A good example is the tumor suppressor gene called p53. It has been called "the guardian of the genome," "the guardian angel gene," and the "master watchman," referring to its role in conserving stability by preventing genome mutation. It is known, for example, that after damage to DNA (by radiation, in this case), p53 levels rise.

Scientists knew that p53 induces the production of another protein called Mdm2, and that Mdm2 actually inhibits p53 and increases p53 degradation (Figure 4.24) (Lahav et al. 2004).

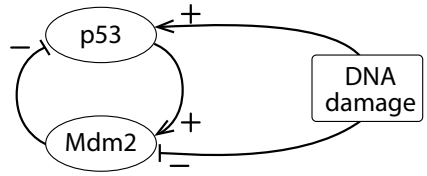


Figure 4.24: Negative feedback in the p53-mdm2 system.

This is obviously a negative feedback loop. However, the function of this negative feedback loop was not immediately clear. Some speculated that its function was to ensure "stability" of this critical protein, by providing a kind of thermostat-like control of its level.

Then, one group actually followed the expression of the two genes over time. They found that "p53 was expressed in a series of discrete pulses after DNA damage." The two genes were expressed in an oscillatory manner, with p53 expression always leading that of Mdm2 (Figure 4.25).

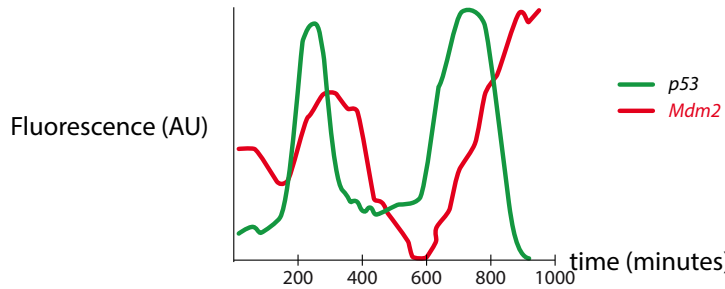


Figure 4.25: Redrawn by permission from Macmillan Publishers Ltd: Nature Genetics ("Dynamics of the p53-mdm2 feedback loop in individual cells," by G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A.J. Levine, M.B. Elowitz, and U. Alon (2004), *Nature Genetics* 36(2):147–150), copyright 2004.

The group developed a series of models reflecting various hypotheses about the mechanism producing the oscillatory gene expression. The models all produce oscillations, but each has a characteristic frequency, amplitude, and waveform, which can be used to choose one model over another.

For example, one model postulates an upstream activator of p53, which they call S, and could therefore be a protein that is produced by damaged DNA.

- S then activates p53 ($= X$), which then activates Mdm2 ($= Y$) after a time delay τ.
- Mdm2 then combines with p53 to degrade it, resulting in a $-XY$ term in the X' equation.
- The S protein is assumed to be produced at a constant rate β_S, and then Mdm2 combines with S to degrade it, producing the $-SY$ term in the S' equation.
- S then activates p53 ($= X$) in a sigmoidal manner, after a time delay τ. This is the primary event post-DNA damage.
- p53 is inhibited by Mdm2 by a mechanism in which Mdm2 binds to p53 and inactivates it (the $-XY$ term in the X' equation).

This results in a set of differential equations

$$\begin{aligned} \textbf{p53} \qquad & X' = \beta_X \frac{S^n}{1+S^n} - \alpha_{XY} XY \\ \textbf{Mdm2} \qquad & Y' = \beta_Y X(t-\tau) - \alpha_Y Y \\ \textbf{DNA damage molecule} \qquad & S' = \beta_S - \alpha_S Y S \end{aligned}$$

A simulation of these equations confirms the existence of oscillations in gene expression. Note that the period is ≈ 6 hours, which agrees with the data (Figure 4.26).

Other models are based on alternative mechanisms, and the outputs of the models can be compared to data in order to rule out one mechanism or another.

The authors present a very interesting interpretation of the oscillations. After noting that the response of the cell to DNA damage is an oscillatory series of pulses, they call this a "digital" response, because the cell's response to larger DNA damage is to emit a larger *number* of identical pulses, as opposed to just producing a higher constant output, which they call an "analogue" response. They suggest that the digital response is more effective, since higher-amplitude pulses or higher constant levels of p53 can easily be toxic. This same reasoning has been used to explain oscillations in hormone levels.

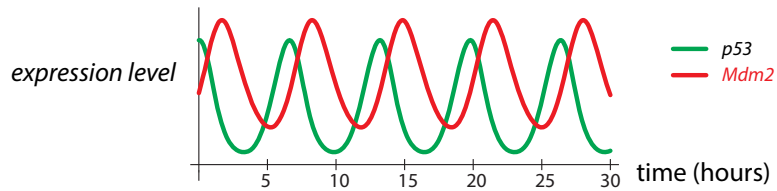


Figure 4.26: Simulation of Model VI of (Lahav et al. 2004). Green (lower) tracing is p53, red (upper) is Mdm2. Parameter values are $\beta_X = 0.9, \alpha_{XY} = 1.4, \beta_Y = 1.2, \alpha_Y = 0.8, \beta_S = 0.9, \alpha_S = 2.7, \tau = 0.9, X_0 = 0, Y_0 = 0.9, S_0 = 0$.

A second example of oscillation in gene expression is in the Hes1 system, which we have already seen (Figure 4.4 on page 174). Hirata et al. developed a model to explain these Hes1 oscillations. In their model, the messenger RNA (mRNA) for Hes1 ($= Y$) is converted into Hes1 protein ($= X$) at a rate B. They postulate an "interaction factor" ($= Z$), which would combine with Hes1 protein to degrade it. Thus, there are $-XZ$ terms in both the X' and Z'

equations. The Hes1 protein is assumed to inhibit its own transcription, that is, Hes1 mRNA. This inhibition is modeled by the decreasing sigmoid function $\frac{E}{1+X^2}$. Note that there is another decreasing sigmoid term in the Z' equation, $\frac{F}{1+X^2}$, implying that Hes1 protein also inhibits the production of the Hes1 interaction factor (Figure 4.27).

The overall model is

$$\textbf{Hes1 protein} \quad X' = -AXZ + BY - CX$$

$$\textbf{Hes1 mRNA} \quad Y' = \frac{E}{1+X^2} - DY$$

$$\textbf{Hes1 Interaction factor} \quad Z' = -AXZ + \frac{F}{1+X^2} - GZ$$

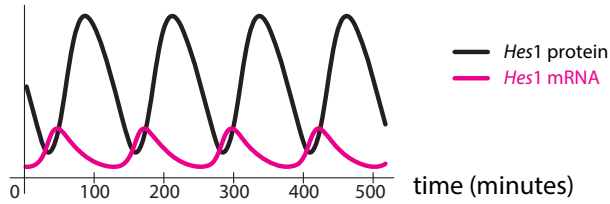


Figure 4.27: Simulation of the Hirata et al. model of Hes1 oscillations. Parameter values are $A = 0.022, B = 0.3, C = 0.031, D = 0.028, E = 0.5, F = 20, G = 0.3$.

Finally, a third kind of model for oscillatory gene expression has been developed by a group of researchers at the University of Texas Medical School in Houston. They focused on the role of *transcription factors*, which are all-important regulators of gene expression.

Genes have subsections that are called *response elements*. These are parts of the gene that easily bind to different kinds of signaling molecules, called transcription factors, and respond by increasing or decreasing transcription, which is the process by which DNA is converted into mRNA.

In many cases, the gene for a transcription factor can be inhibited by the transcription factor protein, generating a powerful negative feedback mechanism that can generate oscillations in gene expression.

Smolen et al. propose a model in which the transcription factor A induces its own transcription as well as the transcription of a second transcription factor R, which then inhibits both A's transcription and its own. The structure of their model is shown in Figure 4.28.

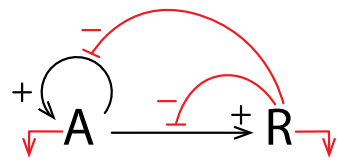


Figure 4.28: Positive and negative feedback loops in the model of an inhibitory transcription factor R, which is activated by transcription factor A.

The differential equations are

$$A' = \frac{k_1 A^2}{A^2 + k_2\left(1 + \frac{R}{k_3}\right)} - k_4 A + r_{bas}$$

$$R' = \frac{k_5 A^2}{A^2 + k_6\left(1 + \frac{R}{k_7}\right)} - k_8 R$$

Note the general form of the model. The large terms in the A' and R' equations have the same form. They are both increasing sigmoids in A, which means that A spurs the production of both itself and R. The presence of R in the denominator of the increasing sigmoids means that greater amounts of R will decrease the production of A and itself.

Simulating their model confirms the existence of oscillations, with a period of one to two hours (Figure 4.29).

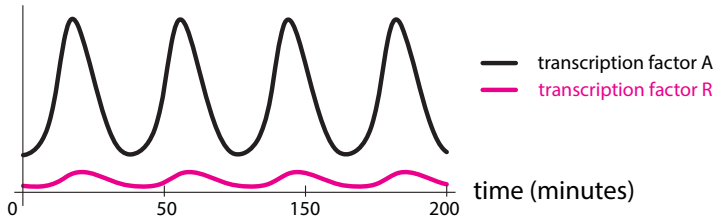


Figure 4.29: Simulation of the Smolen et al. model of gene transcription factor oscillations. Parameter values are $k_1 = 10.5, k_2 = 10, k_3 = 0.2, k_4 = 1, k_5, = 0.3, k_6 = 10, k_7 = 0.2, k_8 = 0.2, r_{bas} = 0.4$.

Further Exercises 4.2

1. Some people have difficulty maintaining a stable weight. Instead, they gain a lot of weight, go on a diet, lose the weight, but then gain it back. This pattern is sometimes referred to as yo-yo dieting.
   a) What kind of feedback loop is involved in this situation?
   b) Use your understanding of feedback loops and oscillations to suggest what might help such a person to stabilize their weight.
2. While traveling, you find yourself in a hotel room in which using the thermostat leads to large oscillations in the room's temperature. The thermostat responds to the room's air temperature by turning on an air conditioner on the other side of the room if the temperature near the thermostat gets too warm. Similarly, when the temperature near the thermostat gets cold, the air conditioner switches off. What could the builder of the hotel have done to prevent the oscillations you are experiencing?
3. Give an example (other than those in the text) of an oscillation caused primarily by a highly sensitive negative feedback loop and another one caused by time delays.

4. Meerkats are highly social small carnivores that live in southern Africa. They rely on each other to raise their young. Use the following assumptions to model the number of *adult* meerkats, M, in a population. You can invent parameters as necessary.
   - The per capita rate at which meerkats give birth to babies who survive to adulthood is a steep sigmoid function of the adult population, with higher reproductive success at higher populations.
   - Meerkats die of natural causes at a constant per capita rate d.
   - Meerkats are preyed upon by eagles and jackals. These predators have many other prey, so their population does not depend on the meerkat population.
   - The rate at which jackals prey on meerkats is a nonsigmoid saturating function of the meerkat population.
   - The rate at which eagles prey on meerkats is a sigmoid function of the meerkat population. The sigmoid is not very steep.

5. The garibaldi is a large orange fish that lives off the coast of California and Baja California. Use the assumptions below to write a differential equation for the size of an *adult* garibaldi population.
   - The number of adults joining a population is the number of eggs laid times the fraction that hatch times the fraction that survive to adulthood.
   - Garibaldis lay eggs at a constant per capita rate, b.
   - Because garibaldis sometimes eat their own eggs, the fraction of eggs that hatch is a declining sigmoid function of the adult population.
   - Larval garibaldis float as plankton before becoming adults and joining a population. Thus, the number of individuals joining a population is proportional to the number that hatched six years earlier, with proportionality constant r.
   - Adult garibaldis die at a constant per capita rate d.

6. At a picnic, you drop a cookie, which promptly attracts the attention of a nearby ant colony. Let A be the number of ants *on the cookie*.
   a) When ants find food, they secrete a pheromone as they return to the anthill that causes other ants to follow their path. The greater the number of ants that do this, the more pheromone there is, and the greater the number of ants that go to the cookie. However, when there are many ants on the cookie, some go home empty-mandibled. Seeing these unsuccessful ants discourages new ants from going to the cookie. Sketch the graph of a function that fits this description (the number of ants going to the cookie as a function of the number of ants on the cookie) and write an equation for it. Briefly explain why you chose the shape that you did.

   b) Write a differential equation for A based on your answer to the previous part and the following assumptions. Feel free to create parameters as necessary.
      - Ants decide whether or not to go to the cookie as soon as they leave the anthill and do not change their minds once the decision has been made.
      - It takes ten minutes for an ant to travel between the anthill and the cookie.
      - Ants on the cookie leave at a constant per capita rate k.

7. The logistic equation predicts that when a small population is introduced to a new habitat, it will smoothly grow until reaching carrying capacity and then level off. However, what we often observe in such cases is an overshoot and collapse pattern, in which the population grows to a high density and then crashes.
   a) Let N be the *adult* population. Use the following assumptions to model this system.
      – The total birth rate is a logistic function of the adult population.
      – After being born, individuals take τ time units to mature into adults.
      – Adults have a constant per capita death rate d.
   b) Simulate the model for $r = 1.2$, $K = 50$, $d = 0.1$, and $\tau = 2.8$. Describe your observations.
   c) What happens if you change r? What about τ?
8. Recall the hypothalamus-pituitary-gonad (H/P/G) model:

$$\begin{aligned} H' &= \frac{1}{1+G^n} - k_1 H \\ P' &= H - k_2 P \\ G' &= P - k_3 G \end{aligned}$$

   a) Find the equilibrium points of this system when $n = 1$. How many are there that are biologically meaningful?
   b) For values of n other than 1, it is difficult/impossible to find the equilibrium points by hand. Use a graphing calculator or the `find_root` command in Sage to find a biologically meaningful equilibrium point of this system for $n = 2$, and for $n = 7$, $n = 8$, and $n = 9$. *(Hint: There is a clever way to find/approximate this equilibrium point graphically.)* See whether you can find it.

4.3 Bifurcation and the Onset of Oscillation

Glycolysis

Earlier in this chapter, we discussed oscillatory chemical reactions. You might think that such reactions are merely laboratory curiosities, useful for amusing students but not very important practically. You would be badly mistaken, because *glycolysis*, one of the fundamental sources of energy in living systems, typically operates in an oscillatory manner.

Glycolysis is one of the body's fundamental metabolic processes, producing the energy molecules that cells can consume. It is perhaps the most ancient metabolic pathway, and it can proceed without oxygen. High-intensity/short-duration activities like sprinting are fueled by glycolysis.

Glycolysis also fuels the yeast cells that are used to brew alcohol. When these yeast cells are grown in a high sugar medium, their outputs become oscillatory.

The earliest observations of glycolytic oscillations were in these yeast cells. They do not require the structure of the cell, and can even be seen in cell-free suspensions (Ghosh and Chance 1964). When cells are suspended in a medium containing glucose, the individual cells synchronize to produce macroscopic oscillations. (Figure 4.30).

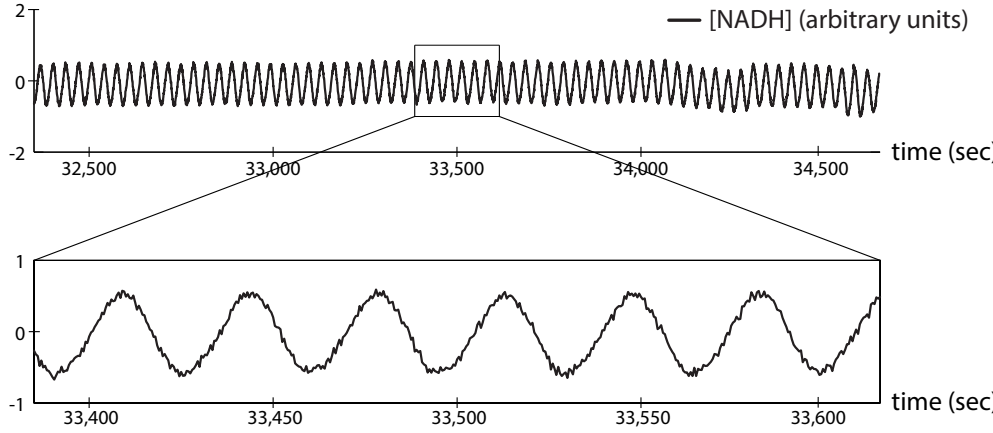


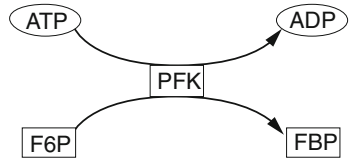
Figure 4.30: Glycolytic oscillations in a suspension of yeast cells. The vertical axis is the metabolic intermediate NADH. Redrawn by permission from Macmillan Publishers Ltd: Nature "Sustained oscillations in living cells," by S. Danø, P.G. Sørensen, and F. Hynne, (1999), *Nature* 402(6759):320–322, copyright 1999.

Research (Chou et al. 1992; Luciani et al. 2006) has suggested that these glycolytic oscillations may be physiologically functional, since they are coupled to oscillations in insulin-producing pancreatic β cells (Figure 4.31, top).

Interventions that disrupt intracellular Ca^{2+} oscillations also abolish glycolytic oscillations, which are essential for insulin secretion and are impaired in diabetes (Figure 4.31, bottom).

The simplest mechanism for glycolysis focuses on the reaction governed by the enzyme phosphofructokinase (PFK), the so-called Higgins–Selkov model. When glucose is processed by the metabolic system, the first part is the two-step conversion of glucose to fructose-6-phosphate (F6P). Then the enzyme PFK governs the key step in glycolysis, which is the conversion of F6P into fructose 1,6-biphosphate (FBP). FBP then is an energy molecule that fuels cellular metabolism and produces large quantities of ATP (adenosine triphosphate) molecules downstream; ATP is the form of energy actually used by the cell.

The PFK reaction itself requires one molecule of ATP, which is converted to the less-useful ADP (adenosine diphosphate).



PFK is an enzyme, and it requires for its activation to be bound with two molecules of ADP. As is typical for a catalyst, the molecules are not consumed by the catalytic reaction, so the

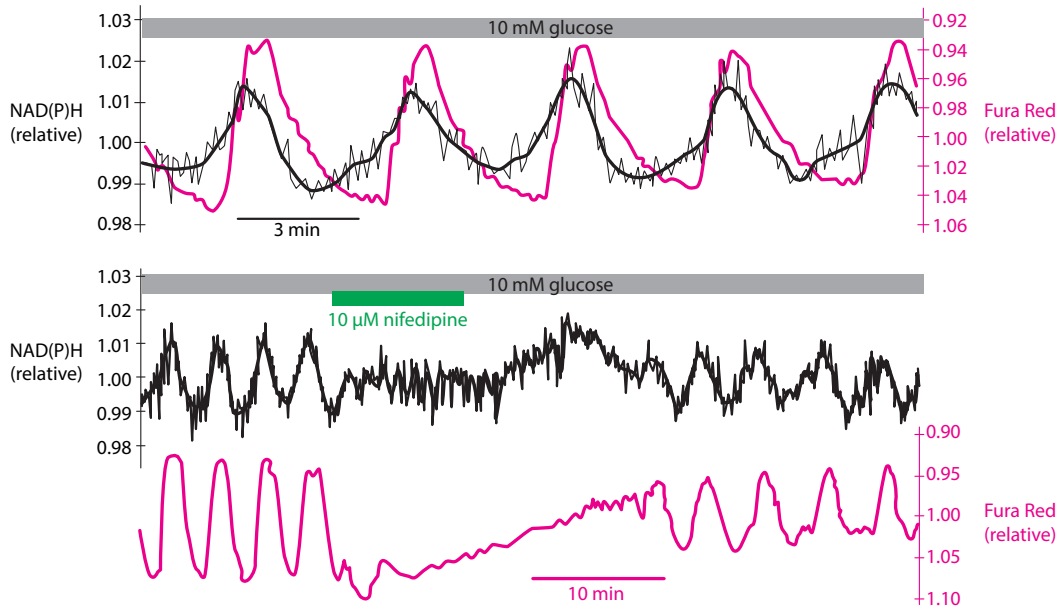


Figure 4.31: Top: synchronization of glycolysis and intracellular Ca^{2+} oscillations in mouse pancreatic islet β cells. The black tracing is the glycolytic intermediate NAD(P)H. The red tracing is a fluorescent indicator of intracellular Ca^{2+}. Bottom: intracellular Ca^{2+} oscillations (Fura Red) and glycolytic oscillations in mouse pancreatic islet β cells. When the intracellular Ca^{2+} oscillations are disrupted by nifedipine, a calcium-channel blocker, both oscillations are inhibited. Redrawn from "Ca2+ controls slow NAD(P)H oscillations in glucose-stimulated mouse pancreatic islets," by D.S. Luciani, S. Misler, and K.S. Polonsky, (2006), *Journal of Physiology* 572(2):379–392. Copyright 2006 John Wiley & Sons. Reprinted with permission from John Wiley & Sons.

overall reaction scheme is

$$\text{☁} \xrightarrow{V_0} \text{F6P}$$
$$\text{F6P} + 2\,\text{ADP} \xrightarrow[\text{PFK}]{c} 3\,\text{ADP}$$
$$\text{ADP} \xrightarrow{k} \text{☁}$$

where the clouds ☁ mean "the environment."

So, from these reaction schemes, we follow the approach of Chapter 1 on how to write differential equations from chemical laws (page 34), and write the differential equation, letting S = [F6P], and P = [ADP]:

$$S' = V_0 - cSP^2$$
$$P' = cSP^2 - kP$$

Exercise 4.3.1 Explain what each term in this model means and why it has the algebraic form (for example, SP^2) that it does.

A simulation of this model shows that with a small change in k_1, the system changes from equilibrium behavior (Figure 4.32, left), to a stable limit cycle oscillation (Figure 4.32, right).

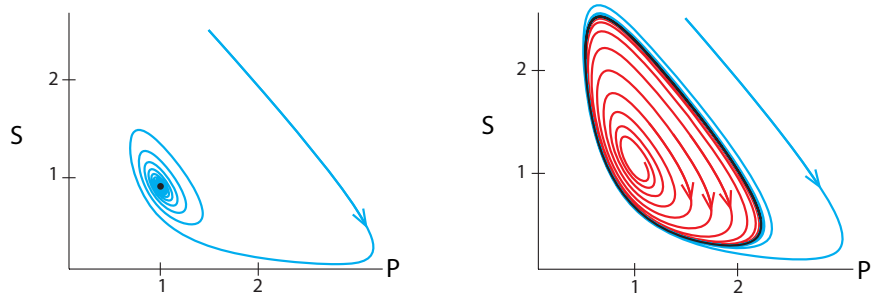


Figure 4.32: $V_0 = 1, k = 1, c = 1.1$ (left), and $c = 0.9$ (right).

Since this early model, there have been several more-sophisticated models of oscillation in glycolysis. See, for example, the paper of Boiteux et al. (1975).

Stable Oscillations in an Ecological Model

We will now consider an ecological model that is more realistic than the Lotka–Volterra (shark–tuna) model in Chapter 1. It's called the Holling–Tanner model (Tanner 1975).

Let us call our prey population N and our predator population P. The Lotka–Volterra model assumed that in the absence of predators, the prey population would grow exponentially. This is clearly unrealistic, since prey population growth must be constrained by resources. (If nothing else, the population will eventually run out of space!) Thus, we will assume that in the absence of herbivores, the prey would grow logistically. The expression for this is the familiar $rN(1 - \frac{N}{K})$.

Another problem with the Lotka–Volterra model is more subtle. The predation term in that model has the form aNP, where N is the prey population and P is the predator population. This means that at every value of P, the amount of prey consumed by the predators is simply proportional to the amount of prey available. No matter what, the predators never get full. This might be an acceptable model if the prey population is small compared to what the predators are capable of consuming, but we can't guarantee that this will always be the case.

This problem can be resolved by making each individual predator's rate of consuming prey level off as prey density increases. The expression for predation becomes $f(N)P$, where $f(N)$ is the function describing how an individual predator's consumption rate changes with prey abundance. (In the Lotka–Volterra model, $f(N) = aN$.) If $f(N)$ plateaus as N increases, there is a limit to how much predators can eat, which makes biological sense. One common choice for $f(N)$ is

$$f(N) = \frac{C_{max} \cdot N}{N + h}$$

where C_{max} is a predator's maximum consumption rate and h is the half-saturation density, the prey density at which consumption is half the maximum rate (Figure 4.33).[3] We will use this function in our model.

[3]Mathematically, this function is called a rectangular hyperbola, but it goes by several other names in biology, including the "Holling Type II functional response" in ecology and "Michaelis–Menten kinetics" in biochemistry.

Exercise 4.3.2 Why can h act as a half-saturation density? In other words, what is the consumption rate when $N = h$, and what does this mean biologically?

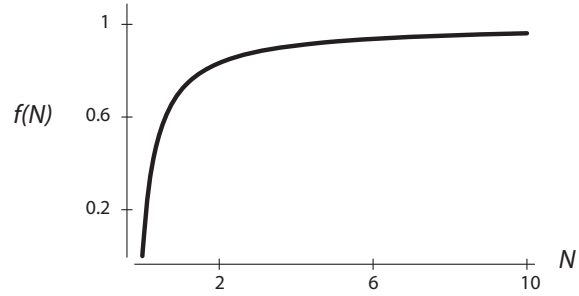


Figure 4.33: The function $f(N) = \frac{C_{max} \cdot N}{N+h}$, when $C_{max} = 1$ and $h = 0.4$.

Putting these assumptions together gives us the following differential equation for the prey population:

$$N' = r_1 N(1 - \frac{N}{k}) - \frac{wN}{d+N}P$$

In this equation, w is the maximum consumption rate, and d is the half-saturation density.

We will assume that the predator population also grows logistically. However, its carrying capacity is set not by an unmodeled environment but by the prey population. More specifically, if j is the number of prey needed to support one predator, then jP is the number of prey necessary to support a population of P predators. If jP is less than the actual prey population, N, the predator population can grow. However, if jP is greater than N, the predator population has exceeded its carrying capacity and must decline. These assumptions translate into the equation

$$P' = r_2 P(1 - \frac{jP}{N})$$

This model undergoes a dramatic change in behavior as w, the maximum consumption rate, increases. When w is low, the system has a stable equilibrium, as shown in Figure 4.34.

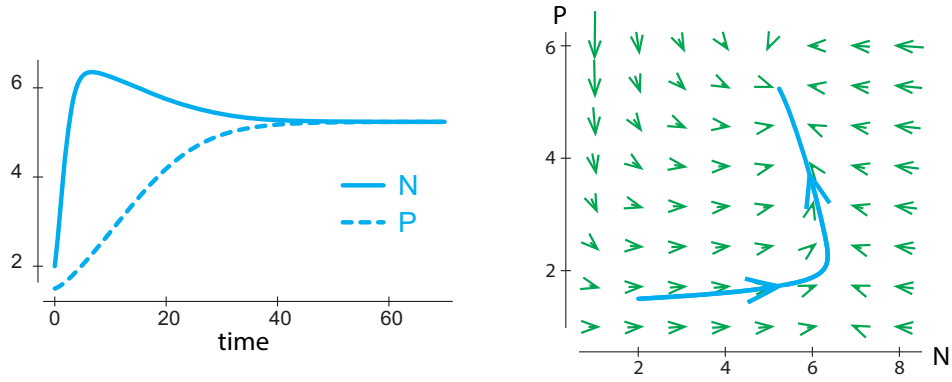


Figure 4.34: A simulation of the Holling–Tanner model, with $r_1 = 1$, $r_2 = 0.1$, $k = 7$, $d = 1$, $j = 1$, and $w = 0.3$.

As w increases, the equilibrium point moves but remains stable. However, as w passes a critical value, the equilibrium becomes unstable, as shown in Figure 4.35. When this happens, a limit cycle attractor appears.

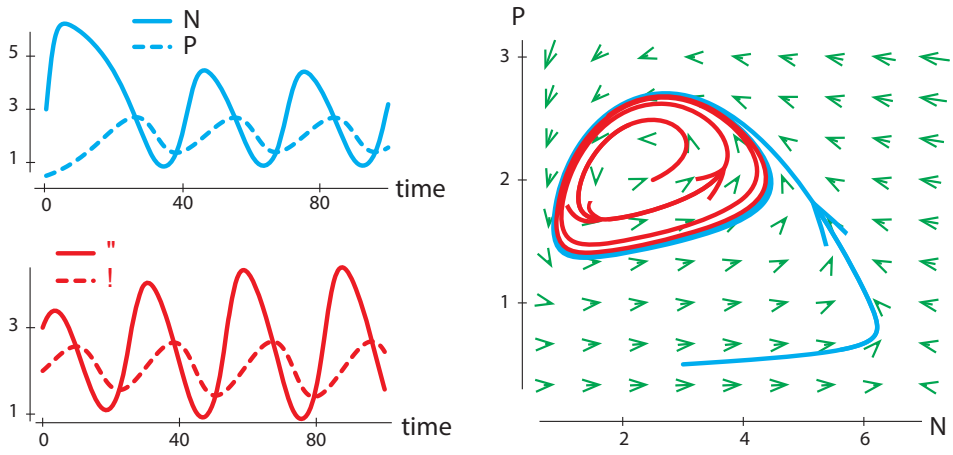


Figure 4.35: Two simulations of the Holling–Tanner model with $w = 1$ starting from different initial conditions and all other parameters as in Figure 4.34.

Exercise 4.3.3 Find the equilibria for this model using the parameter values in Figure 4.34. *(Hint: Work with the second equation first.)*

Exercise 4.3.4 Try intervening in the Holling–Tanner system by introducing predator-removal policies at various phases of the cycle with varying magnitudes. In Chapter 1, we performed shark-removal interventions in the shark–tuna model Figure (1.9 on page 7). How do the results of your interventions compare to those in the shark–tuna (Lotka–Volterra) system?

Hopf Bifurcations

Consider what has happened in each of these models: there is a parameter in the system that creates a change from "stable equilibrium point" to "unstable equilibrium point plus stable limit cycle."

- In Rayleigh's clarinet reed model, it was the slope of the friction term at $V = 0$. When it was positive, the equilibrium point was stable, but when it became negative, the equilibrium point became unstable, and a stable limit cycle was born.
- In the hypothalamic/pituitary/gonadal axis, the critical parameter was n, which reflected the steepness of the negative feedback. When n passed a critical value, the equilibrium point became unstable, and a stable limit cycle was born.
- In the respiratory control model, there were two parameters that produced oscillation: n, which measured the steepness of the negative feedback, and τ, which reflected the time delay in the system.
- In the Selkov glycolysis model, the critical parameter was c, the reaction rate of the catalytic step.

- In the Holling–Tanner model, there are several critical parameters: w, the maximum consumption rate of the predators, as well as r, d, and k, for each of which there are similar critical values.

We have now seen a new example of a "change in the attractors of a differential equation as a parameter passes a critical point," which extends the notion of bifurcation from Chapter 3. So this change is a bifurcation.

This combination of an equilibrium point losing stability and a limit cycle appearing is called a *Hopf bifurcation*. (Its full name is "Poincaré–Andronov–Hopf bifurcation," but it is usually just called a Hopf bifurcation.) It is the first bifurcation we've seen that involves oscillations and therefore cannot occur in one dimension.

The destruction of a stable equilibrium point and its replacement by an unstable equilibrium point and a stable limit cycle attractor is called Hopf bifurcation.

Hopf Bifurcations and the Causes of Oscillation

The theory of Hopf bifurcation gives us unique insights into the mechanisms responsible for oscillatory behavior. It is also a great example of the program of Poincaré, which we mentioned at the end of Chapter 3: explain forms of motion, and changes of forms of motion, by finding bifurcations.

The respiratory control model is an especially good example, because it explicitly depends on two parameters:

(1) n, which controls the steepness of the feedback,

(2) τ, which controls the time delay.

We can then make a two-parameter bifurcation diagram, which is generic for systems with time delay and negative feedback; see Figure 4.36:

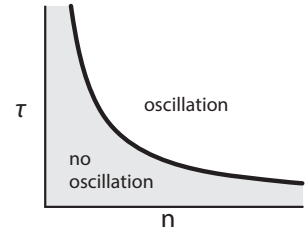


Figure 4.36: A typical bifurcation diagram for a negative feedback system, where n represents steepness of feedback and τ represents time delay.

From this diagram, we can see that:

(1) Oscillation requires at least some steepness of feedback *and* some time delay.

(2) It also requires that at least one of these factors be significantly large.

The Hopf bifurcation diagram enables us to make statements like the following: "the cause of these oscillations is . . ." where the ". . ." will be factors involved in the slope of feedback and/or

time delay in the system. It also enables us to intervene in these systems to enhance or prevent these oscillations.

The chief causes of oscillation in feedback systems are steep negative feedback and time delays.

Further Exercises 4.3

1. Briefly explain the statement due to W. Smith, "Puberty is a Hopf bifurcation." What does this mean?

2. Create a slider-based interactive that allows you to alter C_{max} and h in $\frac{C_{max}N}{N+h}$. Describe how changing these variables affects the shape of the plot and the biological meaning of these changes.

3. Recall the Holling–Tanner predator-prey model:

$$N' = r_1 N(1 - \frac{N}{k}) - \frac{wN}{d+N}P$$
$$P' = r_2 P(1 - \frac{jP}{N})$$

   a) This system is difficult to work with because it has six different parameters, all of which affect the behavior of the system. However, each of them has a biological meaning. Write a brief explanation of what each parameter ($r1, r2, d, j, w, h, k$) means and specify the appropriate units for each one. (Assume that time is measured in years, so that for example, the units of N' are "prey per year" and the units of P are "predators per year.")

   b) What is the state space for which these differential equations are defined? *(Hint: Be careful! There is something here that is slightly different from the usual.)*

   c) Use a graphical analysis (nullclines) to determine how many equilibrium points this system has and say as much as you can about where they occur in the state space. What can you say about the stability of each equilibrium point? *(Hint: It is possible to do this without having to plug in any numbers for the parameters, assuming only that the parameters are all positive numbers. However, you may plug in reasonable numbers for them if you wish. The nullclines should look roughly the same regardless of what numbers you use.)* Also, all but one of the equilibrium points are easy to compute algebraically by hand, but unfortunately this "hard" one is the most interesting.

   d) Suppose $r_1 = 0.4$, $r_2 = 0.03$, $d = 1$, $j = 150$, $w = 300$, $h = 1000$, and $k = 3000$. Find the equilibrium points of this system. You may do this with just algebra, or use a graphical method (nullclines), or use SageMath or a graphing calculator. Note: There is one "interesting" equilibrium point, which is not on either axis, i.e., for which N and P are both nonzero.

e) With the parameters as in part (c), the trajectories approach a limit cycle attractor. Based on this, what can you say about the equilibrium point at which both N and P are nonzero?

f) Now using the same parameters as in part (c), but with $r_1 = 0.2$, find the equilibrium points of the system again. By plotting a trajectory or some time series in SageMath, what can you say this time about the equilibrium point at which both N and P are nonzero? What phenomenon has occurred between $r_1 = 0.2$ and $r_1 = 0.4$?

4. We can also study the Holling–Tanner model using vector fields and simulation. In this problem, we will use the parameter values $r_1 = 1$, $r_2 = 0.1$ $k = 7$, $d = 1$, $j = 1$, and $w = 0.3$.

a) Plot the vector field for this system. Allow both N and P to range between 0 and 10.

b) Simulate and plot the time series for this system for at least two initial conditions, running each simulation for 100 time units. Be sure to keep your simulation results for future use.

c) Plot trajectories for the simulations from the previous exercise and overlay them on the vector field. (All the trajectories should be on one plot.) If necessary, change the plotting range for the vector field so it is big enough for the whole trajectory.

d) Set w to 1 and simulate the model for three different initial conditions, plotting the time series for each. Describe what happens.

e) Plot the vector field for the model with $w = 1$. Then, overlay trajectories from your simulations on the vector field.

f) What is the term for a change in behavior resulting from a change in a parameter, like what you observe here?

5. You also observed oscillations in the Lotka–Volterra predation model, but that model's behavior was different in an important way.

a) Repeat the first three parts of Further Exercise 4.3.4 for the Lotka–Volterra model

$$N' = 0.5N - 0.01NP$$
$$P' = (0.5)(0.01)NP - 0.2P$$

b) How is the behavior of the Holling–Tanner model similar to that of the Lotka–Volterra model? How is it different?

6. Recall the Higgins–Selkov model of glycolysis,

$$S' = V_0 - cSP^2$$
$$P' = cSP^2 - kP$$

a) Simulate this model with $V_0 = 0.5$, $c = 0.23$, and $k = 0.4$ for three different initial conditions. How does the system behave?

b) In real life, for these parameter values, V_0 can range from 0.48 to 0.6. Using any method you choose, approximate the value of V_0 at which the system begins to have persistent oscillations. (You may want to use more than one method.)

4.4 The Neuron: Excitable and Oscillatory Systems

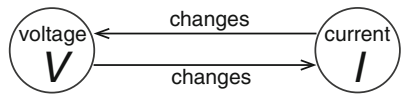
Virtually all the cells in our body have some electrical activity that is essential for their regulation and function. Understanding how this electrical activity is generated and spreads is the subject called "electrophysiology." To grasp it, we need first to understand something about electricity, and second, something about physiology.

A Trip to the Electronics Store

First we will review the necessary facts about electrical circuits. We will pay a visit to the electronics store, but we will be taking home just the differential equations (see, for example, Hirsch et al. (2012)).

In electrical circuit theory, differential equations take a special form. We saw that in mechanics, the fundamental variables are of two kinds: positions and velocities. In electrical circuit theory, the fundamental variables are **voltages** and **currents**, generally denoted by V and I. *Current* is the flow of electric charge, or more concretely, of charged particles (electrons, protons, or ions). *Voltage* is simply a difference in charge between two places. Both voltage and current can be either positive or negative, depending on the direction of the flow (for current) or which location has more charge (for voltage).

In the world of electricity, the form of the differential equations is given by the fact that voltages change currents, and currents change voltages.



The first item we pick up is a **capacitor**. A capacitor is a device that stores electric charge inside an outer shell and releases it when connected to another electric device. The physics behind this storage can vary: the charge can be stored as an electrical field, or it can be stored chemically. When it is stored chemically, this constitutes a *battery*. What matters to us is the charging and discharging of the capacitor/battery, which is described by a simple differential equation:

$$\frac{dV_C}{dt} = \frac{1}{C} \cdot I_C$$

where V_C and I_C are the voltage and current across the capacitor, and C is a constant called the capacitance (here $C = 1$).

This differential equation governs the charging and discharging of the capacitor/battery. It says, for example, that when the capacitor is discharging, the current depletes the stored voltage. And when the capacitor is being charged, the larger the applied current, the faster it will charge.

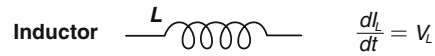
Exercise 4.4.1 What kind of behavior does this differential equation describe?

Exercise 4.4.2 If the capacitor is charging, what is the sign of I_C? If it's discharging?

The second item we find is a little more mysterious: an **inductor** L. The physics behind an inductor is complicated, but it doesn't really concern us here. All that matters to us is that an inductor satisfies a differential equation called *Faraday's law*,

$$\frac{dI_L}{dt} = \frac{1}{L} \cdot V_L$$

where L is a constant called the inductance (here $L = 1$). For us, as mathematical modelers, an inductor is anything that satisfies this differential equation. (In the neuron and cardiac cell, this differential equation describes the opening and closing of ion channels embedded in the cell membrane.)



Exercise 4.4.3 How does a change in the sign of voltage across an inductor affect current? *(Hint: Be careful!)*

The third element is a **resistor** R. Resistors don't have differential equations; instead, there is an algebraic equation that governs their current–voltage relation. It's called *Ohm's law*. You may have learned something by that name in high school or an introductory physics course that was stated as

$$I_R = \frac{1}{R} \cdot V_R \qquad (or\ \text{“}V = RI\text{”})$$

where R is a constant called the "resistance." However, it is not true in general that the voltage across a resistor is equal to some constant R times the current. That's what we would call a *linear* resistor, and not all resistors are linear. Instead, we will talk about a generalized Ohm's law

$$I_R = f(V_R)$$

where f, called the *resistor characteristic*, can take a number of different shapes.

These are the three major types of electric components.

Now let's hook them up into an electrical circuit. The simplest way is to hook up the resistor R, the inductor L, and the capacitor C in parallel, with a voltage source (Figure 4.37):

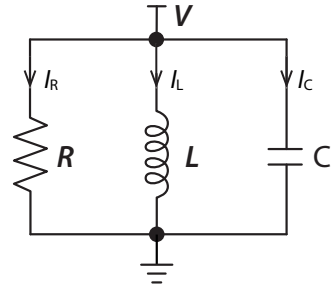


Figure 4.37: An RLC electric circuit.

In order to form the differential equation for this circuit, we need to account for six state variables: the inductor has a voltage V_L and a current I_L, the capacitor has a voltage V_C and a current I_C, and the resistor has a voltage V_R and a current I_R.

At first, it looks like we have six state variables and only two differential equations, plus one algebraic equation (Ohm's law) to account for the six. But once they are hooked up into a circuit, they are no longer independent. Two powerful circuit laws come into play.

Kirchhoff's voltage law (KVL) says that the sum of the voltages around a closed loop must equal 0. Therefore, for the closed loop of the battery and the resistor, we have $V_R - V_0 = 0$, so $V_R = V_0$. Similarly, considering the loops containing the inductor and the capacitor, we can say that $V_L = V_C = V_0$, so all three voltages must be equal.

Kirchhoff's voltage law $\qquad V_R = V_L = V_C$

Kirchhoff's current law (KCL) says that the sum of the currents in and out of a node (circuit component) must equal 0.

Kirchhoff's current law $\qquad I_R + I_L + I_C = 0$

so $I_C = -I_R - I_L$.

Now we can write

$$\begin{aligned} I_L' &= V_L && \text{(Faraday's law)} \\ &= V_C && \text{(by KVL)} \\ V_C' &= I_C && \text{(capacitor law)} \\ &= -I_R - I_L && \text{(by KCL)} \\ I_R &= f(V_R) && \text{(generalized Ohm's law)} \\ &= f(V_C) && \text{(by KVL)} \end{aligned}$$

Collecting these terms and letting $I = I_L$ and $V = V_C$, we get

$$\begin{aligned} I' &= V \\ V' &= -I - f(V) \end{aligned}$$

Now we have a two-variable differential equation. In order to study its behavior, of course, we have to specify the resistor characteristic $f(V)$: as we mentioned, it can take on many different shapes.

We have certainly seen this equation before: it is the equation for a linear spring with friction, with a change of variable names:

electrical	mechanical
$I' = V$	$X' = V$
$V' = -I - f(V)$	$V' = -X - f(V)$

keeping in mind, of course, that the V on the left means "voltage" and the V on the right means "velocity."

Comparing these two equations, we see that the resistor characteristic $f(V)$ in the electrical equation plays the same role as the friction term $f(V)$ in the mechanical equation. This analogy suggests that **resistance is electrical "friction."**

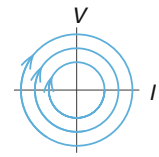
By varying the resistor characteristic $f(V)$, we can produce a variety of behaviors in the electrical circuit.

Case 0: zero resistance. If we could somehow take the resistor out of the circuit, the remaining LC circuit would have zero resistance. Since no current would flow through the resistor, we would have $I_R = f(V_R) = f(V) = 0$. This makes our equation become

$$I' = V$$
$$V' = -I$$

We have seen this equation before: it's just the frictionless spring! In our analogy, we then have

electrical	mechanical
$I' = V$	$X' = V$
$V' = -I$	$V' = -X$



How will this electrical system behave? Just as the frictionless spring oscillates forever, so does the zero-resistance electrical circuit. This continues the analogy of resistance as electrical friction: when it is removed, the system will oscillate in a closed loop forever.

Case 1: linear resistance. Now let's assume a classic linear resistor, in which the resistance is a constant R, and the current is therefore a linear function of voltage:

$$I = \frac{1}{R} \cdot V$$

The constant $\frac{1}{R}$ is often written as g, called the *conductance*.

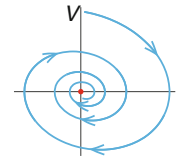
$I = g \cdot V$

This gives us the differential equation

$$I' = V$$
$$V' = -I - gV$$

Pursuing our analogy, we see that this is identical to the spring with simple linear friction. We can therefore say that its behavior will be to spiral inward to the stable $(0, 0)$ equilibrium point. In a time series plot, both variables would exhibit damped oscillations:

electrical	mechanical
$I' = V$	$X' = V$
$V' = -I - gV$	$V' = -X - kV$

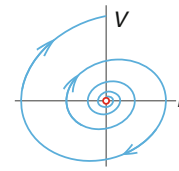


Case 2: "negative resistance." In our discussion of the mechanical system in Rayleigh's clarinet model, we considered the concept of "negative friction." Whatever that might be, we saw that it would be modeled by a friction function that had a negative slope, $f(V) = -kV$.

In the electrical case, the analogy would be to a system with "negative resistance."

The effect of this function, in both the mechanical and the electrical cases, would then be to produce an unstable equilibrium point, spiraling outward from the origin and producing a time series whose amplitude grows with time:

electrical	mechanical
$I' = V$	$X' = V$
$V' = -I + gV$	$V' = -X + kV$



Just as friction robs energy from a mechanical system, so negative friction would have to *supply* energy to the system. In the case of Rayleigh's clarinet, the energy was being supplied by the clarinetist blowing.

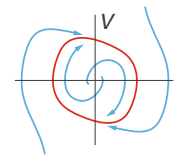
In the case of electrical systems, a similar "negative resistance" would also have to supply energy to the system. This could be a plug in the wall for an electrical circuit. Later, in the case of biological electricity, we will see that the energy supplied is from metabolism.

Case 3: "N"-shaped resistance. In our discussion of Rayleigh's clarinet model, we ended up combining the negative friction produced by the clarinetist with the positive friction inherent in the system to produce an "N"-shaped function, for example, the cubic $V^3 - V$.

If we were to imagine an electrical resistor that had this cubic resistor characteristic,

then our analogy would be complete:

electrical	mechanical
$I' = V$	$X' = V$
$V' = -I - (V^3 - V)$	$V' = -X - (V^3 - V)$



This would result in an electrical system with a limit cycle attractor. The system would go to this attractor and maintain it. At the electronics store, we can buy such devices for 50 cents; they are called tunnel diodes.

In biological systems, as we shall see in the following sections, neurons have regions of negative resistance. When a neuron's resistance characteristic looks like $V^3 - V$, the neuron will exhibit limit cycle behavior and continue oscillating. These neurons are called *pacemaker neurons*.

This concludes our visit to the electronics store. Let's now go on to talk about the physiology behind electrophysiology.

Exercise 4.4.4 Sketch time series for each of the four cases discussed. For each, briefly explain why it makes sense that the time series displays the behavior that it does.

The Electrical Cell

Biological cells create an internal environment that is very different from their external environment (Figure 4.38). In the external environment, which was originally seawater, sodium ions (Na^+) are present in high concentration, around 115 mM, and potassium ions (K^+) are present in relatively low concentration, around 15 mM.

But inside the cell, the situation is reversed: Na^+ concentration is low, while K^+ concentration is relatively high.

This state of ionic disequilibrium is actively maintained by molecular pumps that continually pump Na^+ out of the cell and K^+ in. The pumps require energy to work, and that energy comes from the basic metabolic processes of the body, which convert the food we eat into the molecules that fuel the pumps.

The biochemist Oscar Hechter once began a lecture to a large audience by asking, "What is life?" He paused, and then said, "Ladies and gentlemen, life is the battle against sodium." People laughed, but he was making an excellent point: a large fraction of your lunch goes to generating the energy that fuels the pumps that keep sodium out of our cells.

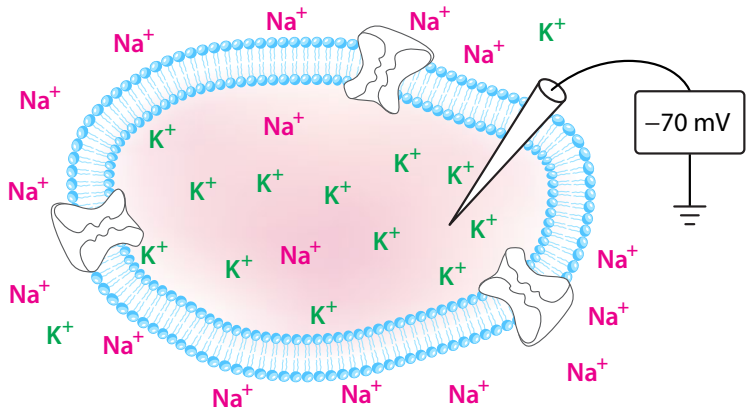


Figure 4.38: The neuron, like many cells, has a high K^+ concentration and a low Na^+ concentration inside the cell. Outside the cell, on the contrary, K^+ concentrations are low, while Na^+ concentrations are high.

The overall effect of this ionic imbalance is that there is a net voltage difference between the inside and outside of a cell, which is typically around −70 mV. That is, there are more + charges outside the cell than there are inside, and this produces the voltage difference across the cell membrane. In the late 1940s, with the development of microelectrode technology, physiologists, including Hodgkin and Huxley, were able to actually measure this voltage difference.

When left undisturbed, a cell remains stable at −70 mV. But the experimenters could administer small *stimulating* currents, again through microelectrodes. What Hodgkin and Huxley saw surprised them (Figure 4.39): when they give the cell a small electric stimulus, it responded with a much larger action and then a return to the resting state.

They realized that this rise and fall of voltage, called the *action potential*, was the key signaling act of the neuron (Hodgkin and Huxley 1939). The small stimulus modeled the receipt of a pulse from another neuron, and the large response was the outgoing signal. They reasoned that this was the basis of neuronal communication.

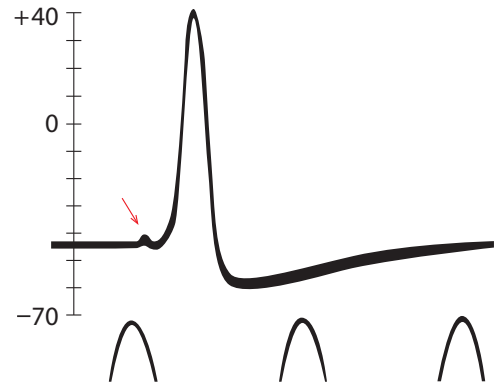


Figure 4.39: First recording of intracellular voltage in a neuron, by Hodgkin and Huxley in 1939. Oscillations at the bottom are time markers that occur every two milliseconds. Note the tiny blip of the stimulus immediately before the onset of the action potential. Redrawn by permission from Macmillan Publishers Ltd: Nature "Action potentials recorded from inside a nerve fibre," by A.L. Hodgkin and A.F. Huxley, (1939), *Nature* 144(3651):710–711, copyright 1939.

The Mechanism of the Action Potential

Hodgkin and Huxley developed a set of hypotheses about how the action potential is generated.

They understood that the rapid increase in voltage had to be produced by a current flowing into the cell, and that the subsequent decrease had to be produced by a current flowing out of the cell. They suspected that these currents were in fact the flow of ions like Na^+ and K^+. After all, ions are charged particles, and the flow of charged particles is a current. But how can this happen? How can sodium ions suddenly start rushing into the cell? How can potassium ions suddenly start flowing out? Hodgkin and Huxley hypothesized that there must be special "particles" that conduct the sodium and potassium ions through the cell membrane. The activity of these carrier particles would then be dependent on the voltages and currents in the system at a given time.

Nowadays, with the advent of molecular biology in the 1970s and 1980s, we know what the "carrier particles" actually are. They are *ion channels*, and their structure and voltage-dependence

are well known. It is remarkable that Hodgkin and Huxley knew none of this but were able to infer the existence of ion channels from macroscopic data and their differential equations.

They went on to develop a four-variable differential equation that described these processes in detail (Hodgkin and Huxley 1952). They were able to produce a simulation of this four-variable equation *by hand calculation* using a mechanical calculator, since electronic computers were new and extremely rare in 1952. Their numerical integration produced a voltage output that closely resembled the actual voltage tracing, and their differential equation was given the Nobel Prize in Physiology in 1963. Good discussions of the Hodgkin–Huxley equations can be found in Keener and Sneyd (2009) and Izhikevich (2007).

Here, we will develop a two-variable simplification of the Hodgkin–Huxley model that captures the essential dynamics, called the *FitzHugh–Nagumo* (FHN) model.

Hodgkin and Huxley stylized the action potential into three stages:

(1) Voltage is elevated by the inrush of Na^+ ions.
(2) Voltage returns to the resting state by the outflow of K^+ ions.
(3) Pumps restore the ion imbalances.

Fast inward process. Hodgkin and Huxley had shown by experiment that the fast inward process was sodium-dependent: removing sodium from the bath water abolished the action potential. So they hypothesized that the voltage elevation was created by the inrush of Na^+ ions. Therefore, the $f(V)$ term in the V' equation must be describing a feature of the sodium conductance.

They also knew that it has a very important feature: if they gave a very tiny stimulus current to the cell, they did not get an action potential. Only a stimulus that was sufficiently strong would elicit the much larger response of the action potential. Therefore, *the equilibrium point of this system must be stable.*

Exercise 4.4.5 How would the cell respond if the equilibrium were unstable?

But then, once the action potential gets underway, there is a positive feedback mechanism at work whereby Na^+ entry into the cell elevates V, which further increases Na^+ entry, etc. This dynamic, in which increases in V cause further increases in V, is a clear example of *negative resistance*.

So they reasoned that the current–voltage curve for the Na^+ resistance had to have a region of negative resistance to account for the explosive increase in voltage. But resistance is the slope of the I/V curve, so this meant that the I/V curve had to have a region with negative slope. However, unlike the examples in the previous section, *the negative resistance region must not include the equilibrium voltage*, or else the equilibrium point would be unstable. So the negative resistance region must lie a small but finite distance away from the equilibrium voltage.

Since

$$V' = -I - f(V)$$
$$I' = V$$

is the master model for the electrical cell, we can model stage 1, the fast inward process, as

$$V' = -I - f(V)$$

Then $f(V)$ has to have certain properties: it has to have a region of negative slope near but not at the equilibrium point and positive slope elsewhere.

A simple function that has those properties is (Figure 4.40)

$$f(V) = V(V-1)(V-a) \qquad \text{with } 0 < a < 1$$

Exercise 4.4.6 Make an interactive that explores the effect of changing parameter a on the shape of the $f(V)$ curve.

If we plot $f(V)$, it is exactly like the friction in the Rayleigh oscillator, except that it is shifted to the right (Figure 4.40). The effect of this shift is to change the stability of the $(0,0)$ equilibrium point. It used to be in the negative friction region in the Rayleigh oscillator model, but now it is in the positively sloped region. Thus, the equilibrium point $(0,0)$ becomes stable.

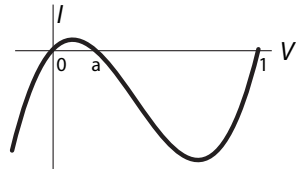
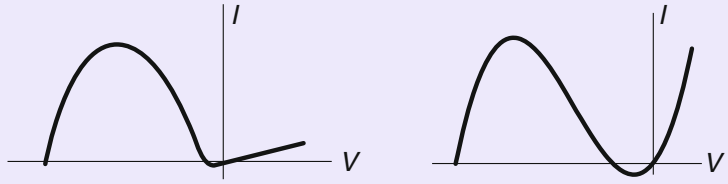


Figure 4.40: Shifted "N"-shaped resistor characteristic function $f(V)$. Here $a = 0.1$.

I/V Curve of the Neuron

Hodgkin and Huxley experimentally recorded the I/V curve of the squid neuron, and found that it had exactly such a negatively sloped region.



On the left is the I/V curve of the squid axon, recorded by Hodgkin and Huxley. On the right is the function $f(V)$ we use to model this process. Here we have plotted $f(-V)$, since in their day, what was meant by V is now what we call $-V$.

It is conventional in the literature to use the function $-f(V)$ and then write the equation for the fast inward process as

$$V' = -I + f(V)$$

where

$$f(V) = V(1-V)(V-a)$$

To reflect the speed of the fast inward process, we will multiply the whole right-hand side of the fast inward equation by $1/\epsilon$, where ϵ is a small number such as $\epsilon = 0.01$. Thus the equation for the fast inward process is now

$$\textbf{fast inward} \qquad V' = \frac{1}{\epsilon}\Big(- I + f(V)\Big)$$

Note the very interesting dynamics that are already contained in this equation. If we consider it a one-variable differential equation $V' = f(V)$, it is exactly the system studied in Chapter 3, called the logistic equation with an Allee effect. It has three equilibrium points, $V = 0$, $V = a$, and $V = 1$. The two equilibrium points at 0 and 1 are stable, and $V = a$ is the unstable threshold. If V is less than a, then V' is negative, and the system goes to the stable equilibrium point at $V = 0$, but if V is greater than a, V increases to the stable equilibrium at $V = 1$. The fast inward dynamics inherits this threshold behavior from the Allee-like character of the resistance curve.

Exercise 4.4.7 Simulate $V' = \frac{1}{\epsilon}\Big(-I+f(V)\Big)$ with $\epsilon = 0.01$ for each form of $f(V)$ discussed in this section. Describe how the system behaves in each case. *(Hint: Try several initial conditions.)*

Recovery process The recovery process is dominated by the flow of K^+ ions. Following Hodgkin and Huxley, we model this as a resistor in series with an inductor. (Why an inductor? Because the current flow through an ion channel changes as a function of voltage, whence $I' = f(V)$, which is the equation for an inductor.)

The recovery phase is therefore represented by the equation for the $[K^+]$ current,

$$\textbf{recovery} \qquad I' = V - \gamma I$$

Combining these insights, we get a model of the electrical cell (Figure 4.41).

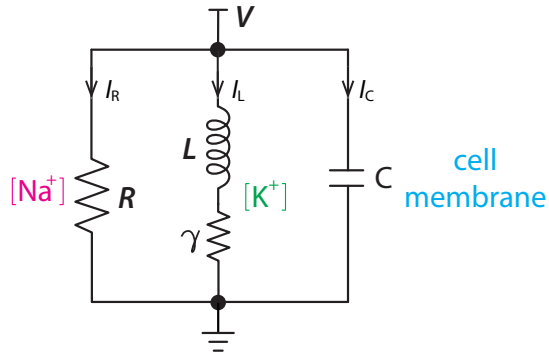


Figure 4.41: Electrical circuit model of a neuron.

Combining the two processes. However, instead of writing $I' = V - gI$, most writers on the subject prefer to create a new variable w for the current, called "recovery," which is identical to our I. The overall equations are then

$$V' = \frac{1}{\epsilon}\Big(- w + f(V) + I_{ext}\Big)$$

$$w' = V - \gamma w$$

where I_{ext} is an external stimulus.

These are called the *FitzHugh–Nagumo equations*, and they are a simple model of the neuronal action potential. Let's study them, both numerically and analytically. We will use as our external stimulus I_{ext}, a square current pulse of duration 0.1 and varying amplitude.

Experiments with the FitzHugh–Nagumo Model

First let's do some experiments with the FitzHugh–Nagumo (FHN) model. We will begin by replicating the experiment of Hodgkin and Huxley. We deliver an extremely small stimulus current I_{ext} to the cell, and the result is a very small deflection of the voltage followed by a quick return to equilibrium (Figure 4.42, left).

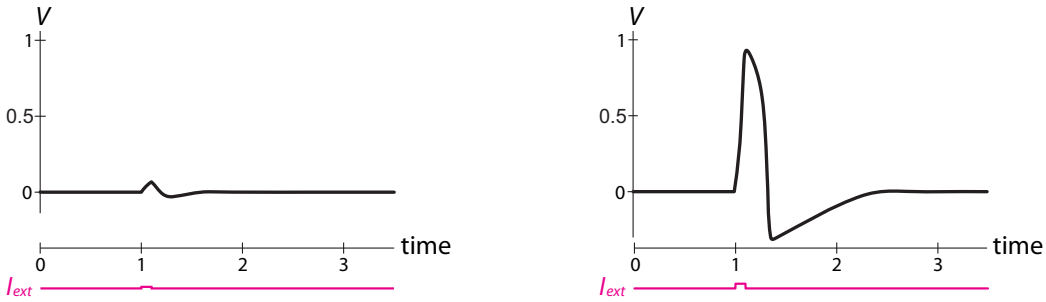


Figure 4.42: Response of the FHN model to a current stimulus pulse delivered at $t = 1$ with duration 0.1. Left: stimulus pulse amplitude = 0.01. Right: stimulus pulse amplitude = 0.03.

But when we increase the amplitude of the stimulus by just a little bit, we get a large action in response, a substantial deflection in voltage, followed by a return to the same equilibrium. This is the action potential (Figure 4.42, right).

For another experiment, let's use as our stimulus not the brief pulse we have been using so far, but a constant input of current. Here, we observe another interesting phenomenon: if the constant current is at a low amplitude, the neuron is quiescent (Figure 4.43, left). But when the constant stimulus has a slightly larger value, the system goes into a permanent oscillation, with a repetitive train of spikes issuing from the neuron (Figure 4.43, right).

And as a final experiment, let's hook up *two* neurons. The coupling between them will be a flow of current between neuron #1 and neuron #2, as actually happens when the neurons are coupled by what are called *gap junctions*. In this case, the coupling is a simple resistor (so $I = \frac{V}{R}$), and the current flow to neuron #1 from neuron #2 is equal to

$$I_{coupling\ 2\to1} = \frac{(V_2 - V_1)}{R}$$

And the flow to neuron #2 from neuron #1 is equal to

$$I_{coupling\ 1\to2} = \frac{(V_1 - V_2)}{R}$$

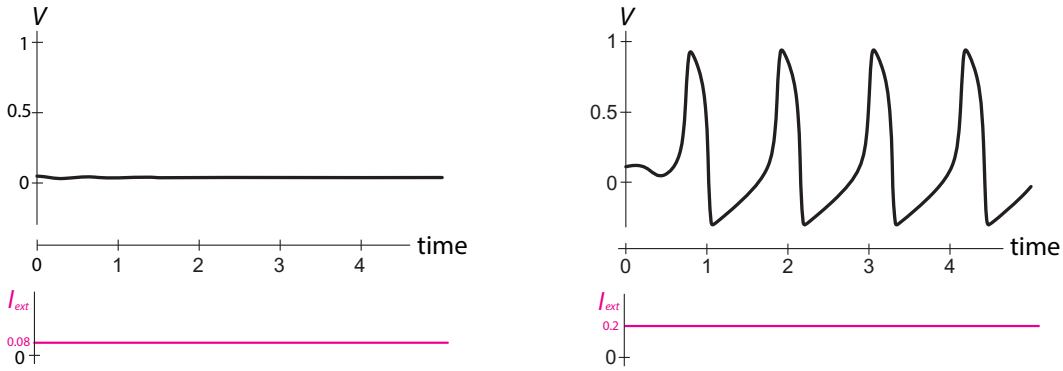


Figure 4.43: Left: response of the FHN system to a small constant current stimulus. Right: response to a slightly larger constant current stimulus.

The overall equation for the two-neuron coupling is

$$
\begin{aligned}
V_1' &= \frac{1}{\epsilon}\Big(- w_1 + f(V_1) + I_{coupling\ 2\to1} \ + I_{ext}\Big) \\
w_1' &= V_1 - \gamma w_1 \\
V_2' &= \frac{1}{\epsilon}\Big(- w_2 + f(V_2) + I_{coupling\ 1\to2}\Big) \\
w_2' &= V_2 - \gamma w_2
\end{aligned}
$$

In this case, we see that neuron #1 passes its excitation to neuron #2, which responds with an action potential after a short delay (Figure 4.44).

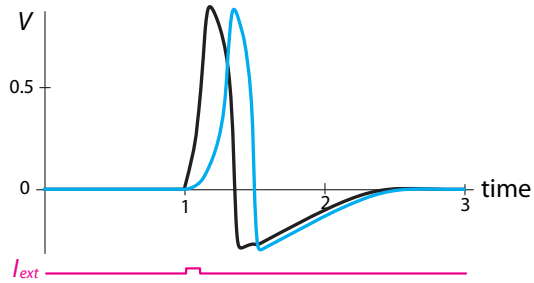


Figure 4.44: Stimulus pulse I_{ext} amplitude $= 0.025$, $R = 45$, $\epsilon = 0.008$, $\gamma = 0.5$, $a = 0.1$.

Hodgkin and Huxley performed the same experiment with their model and realized that this was the key to neuronal communication. They were able to show that if they coupled many of these models in series and used a realistic value for the coupling resistance, the resulting wave of excitation passed down the chain at a speed very close to the measured value of neuronal conduction velocity!

The wave that passes from neuron to neuron, or from heart cell to heart cell, is very similar to the "wave" that is spontaneously formed by crowds at sports stadiums. In both cases, the elements are what are called *excitable elements*. An excitable element is one that has

(1) a stable equilibrium point as its only attractor,

(2) a region of stored energy a small but finite distance away from the equilibrium point.

Such elements will respond to a sufficient stimulus by releasing an excitation of their own, followed by a return to the stable equilibrium point.

Is the Neuron like a Toilet?

There is a good example of an excitable element in the home. It's the flush toilet. The ordinary household toilet satisfies the axioms of an excitable element: very small pushes on the flush handle will produce only a very small response, and a rapid return to the resting state. But if the handle is pushed far enough, the system will spontaneously release a large amount of stored energy. This is the water reservoir in the tank; emptying it produces the large action phase. Then, of course, pumps must go to work, consuming energy, that will pump water back into the tank, to return it to equilibrium.

When excitable elements are hooked up by simple resistive coupling, the result is called an *excitable medium*. One example of a phenomenon that has been modeled as an excitable medium is the occurrence of stadium waves.[4] Similar models have been used to model the spread of forest fires, cardiac electrical conduction, and neural systems.

Dynamics of the FitzHugh–Nagumo Model

All of these phenomena that the neuron displays in reality and in our computer simulations can be explained by careful reference to the phase plane of the model.

Let's first draw the nullclines. To find the V-nullcline, we set $V' = 0$,

$$V' = 0 = \frac{1}{\epsilon}\Big(- w + f(V)\Big)$$

and get

$$w = V(1 - V)(V - a)$$

When we plot this in (V, w) state space, we get the blue curve in Figure 4.45. To find the w-nullcline, we set $w' = 0$ to get

$$w = \frac{1}{\gamma}V$$

which is the red line in Figure 4.45.

[4] Farkas et al. (2002) refers to a stadium wave as "La Ola," Spanish for "wave." They report that the first recorded stadium wave was at Azteca stadium in Mexico City during the 1986 World Cup. Their paper uses an excitable medium model of the stadium wave.

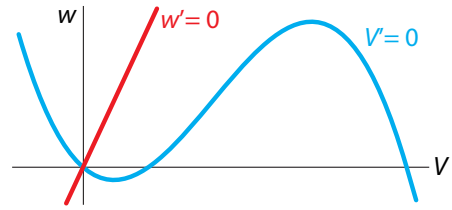


Figure 4.45: Nullclines for the FHN model.

First we find the equilibrium points. Here there is obviously only one, at $(0, 0)$. It is stable, because the slope of the resistor characteristic is negative at this point. (This may sound like the opposite of what we said in the discussion of the Rayleigh and electrical circuit oscillators, where the "negative resistance" region was negatively sloped. But there is no conflict, and both are saying the same thing, because in the FHN model, the resistance term is $+f(V)$, whereas in the Rayleigh model the friction term is $-f(V)$.)

We can then use the nullclines to determine the system's behavior, just as we did in Chapter 3. On the V-nullcline, the change vector (V', w') is $(0, w')$, so there is no horizontal component, and the change vector is purely vertical. On the w-nullcline, the change vector (V', w') becomes $(V', 0)$, so there is no vertical component, and the change vector is purely horizontal (Figure 4.46).

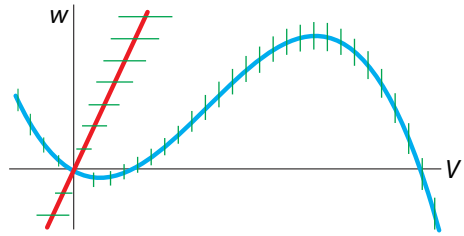


Figure 4.46: The direction of change vectors along the nullclines in the FHN model.

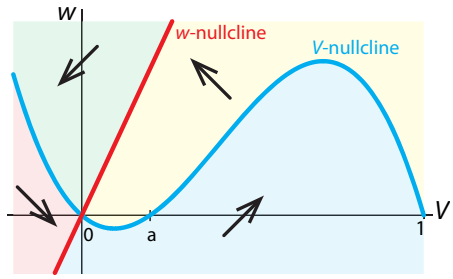
First, let's look at the V-nullcline. The V-nullcline divides state space into a region in which V is growing and a region in which V is decreasing. The only question is which is which, and that is easily answered by looking at the V' equation and realizing that above the blue curve, $w > f(V)$, so V' must be negative; below the blue curve, $w < f(V)$, and therefore V' must be positive.

Similarly, the w-nullcline separates state space into two regions. Since the w' equation is $w' = V - \gamma w$, above the red line $\gamma w > V$, so w' must be negative above the red line, and positive below it.

Together, the two nullclines divide state space into four regions (Figure 4.47).

Exercise 4.4.8 Sketch the nullclines in Figure 4.47 and use test points to confirm that the change vectors are drawn correctly.

The nullcline analysis already gives us a sense of the movement, which can be further confirmed by plotting the vector field superimposed on the nullclines (Figure 4.48).



blue region	$V' > 0$	$w' > 0$
yellow region	$V' < 0$	$w' > 0$
green region	$V' < 0$	$w' < 0$
pink region	$V' > 0$	$w' < 0$

Figure 4.47: The nullclines for the FHN model divide state space into four regimes with distinct behaviors.

We can now plot our first experiment, with the subthreshold and suprathreshold stimuli, on this state space picture. If we plot a trajectory resulting from a low-amplitude stimulus pulse, we see a small counterclockwise orbit, which returns quickly to the stable equilibrium point at $(0, 0)$ (Figure 4.49).

Exercise 4.4.9 In the experiment with the small-amplitude stimulus pulse in Figure 4.49, the stimulus pushed the state point across the blue V-nullcline into "increasing V" territory. Nevertheless, the system returns quickly to the equilibrium point. Why is this so?

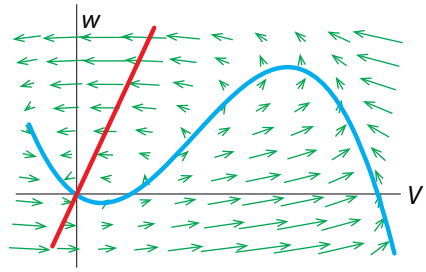


Figure 4.48: Vector field and nullclines for the FHN model. Note the sense of movement.

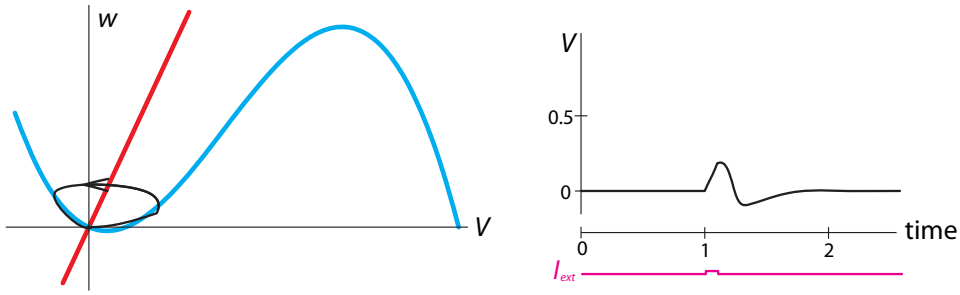


Figure 4.49: Left: One trajectory in state space (black curve) resulting from a low-amplitude stimulus pulse. Right: corresponding time series.

If we increase the amplitude of the stimulus pulse by a little, we get a completely different kind of trajectory, corresponding to an action potential (Figure 4.50). Now the stimulus pulse has pushed the state point well over the blue V-nullcline (phase 1), and now V begins to increase (phase 2). It continues to increase in both V and w, until it crosses the V-nullcline again, and V begins to decline, while w is still increasing (phase 3). In phase 4, the state point has crossed the w-nullcline, and w begins to decrease, while V is still decreasing. And finally, in phase 5, the state point has passed the V-nullcline again, and V decreases along with w until the system relaxes back to the equilibrium point. Note that in this phase, the state point hugs the V-nullcline, meaning that V' is nearly 0 during this phase.

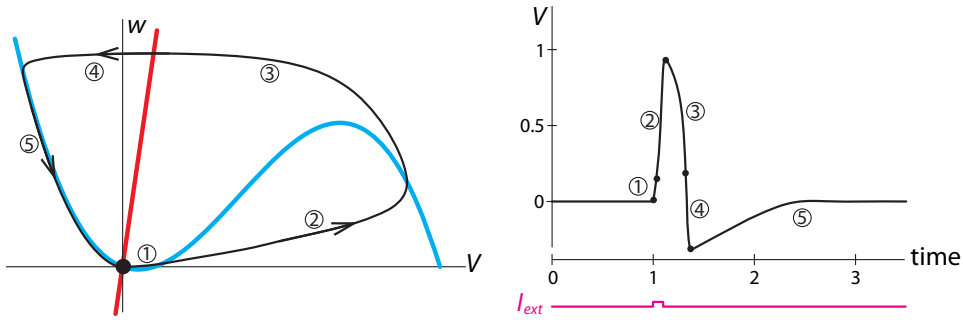


Figure 4.50: Left: State space trajectory (black curve) of the response to a slightly larger stimulus. Right: corresponding time series.

Finally, let's consider the effect of adding a constant stimulus current I_{ext}. Note that the addition of the constant term to the V' equation has the effect of shifting the V-nullcline upward. Now the equilibrium point is no longer at $(0, 0)$.

If the amplitude of the stimulus is small, the new equilibrium point is moved closer to the positively sloped region, but it does not quite reach it (Figure 4.51). As a consequence, the equilibrium point is still stable, although it is so close to the unstable region that even a small perturbation will elicit an action potential.

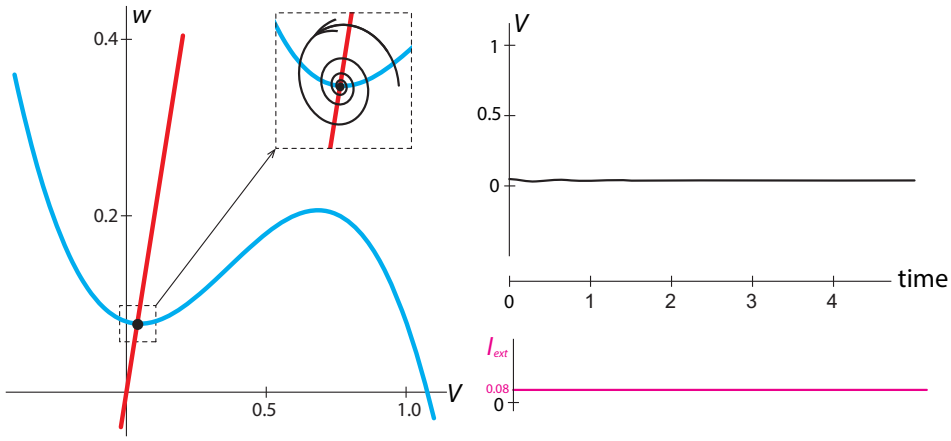


Figure 4.51: Left: Adding a small constant external stimulus, rather than a pulse, moves the blue nullcline upward, but does not essentially change the dynamics of the system. Right: time series of the system's response to a small perturbation.

However, when we increase the amplitude of the stimulus current, we see a different phenomenon: now the equilibrium point has been shifted into the positive-slope region of the V-nullcline, and the system now has an unstable equilibrium point and a stable limit cycle attractor (Figure 4.52). This neuron will fire repetitively. Such neurons are called "pacemaker neurons," and our analysis suggests that there is a deep analogy between these neurons and Rayleigh's model of the clarinet!

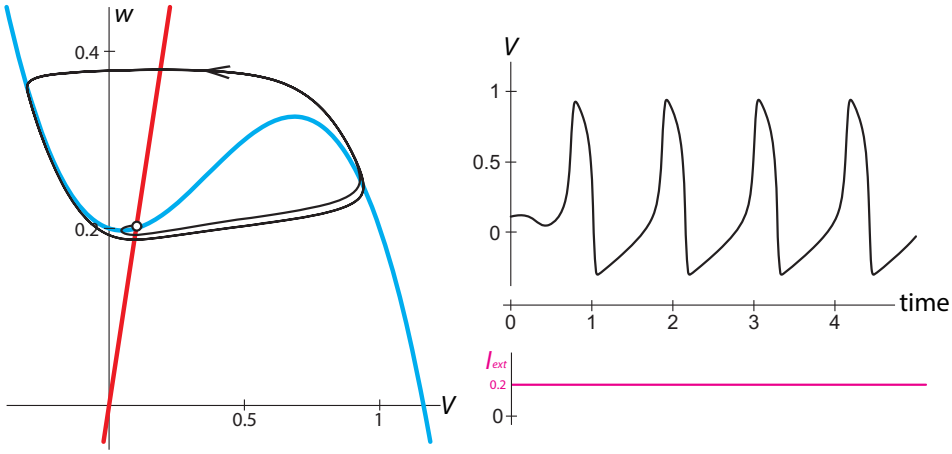


Figure 4.52: Left: Adding a larger constant external current changes the dynamics of the model. Now the red line crosses the blue line in a positively-sloped region, and the equilibrium point becomes unstable. Any small perturbation off the unstable equilibrium point will result in a permanent oscillation (black curve). Right: time series of the behavior.

Further Exercises 4.4

1. A common differential equation, used, for example, to represent ion channel kinetics, is

$$X' = \frac{a_0 - X}{t_0}$$

where a_0 and t_0 are constants.

a) What dynamics follow from this equation?

b) How do a_0 and t_0 affect these dynamics?

2. For the following system,

$$V' = \frac{1}{\epsilon}\Big(- w + f(V) + I_{ext}\Big)$$
$$w' = V - \gamma w$$

where $I_{ext} = 0.08$, $\epsilon = 0.01$, $f(V) = V(1-V)(V-a)$, $a = 0.1$, $g = 0.5$:

a) Calculate the equilibrium points by setting $V' = w' = 0$.

b) Write down the V-nullcline function.

c) Calculate the slope of the V-nullcline at the equilibrium point.