Derivatives and Integrals

2.1   What Is X'?

Let's focus on the change vector X'. Euler's method can give us a very deep insight into what X' is. Recall that a state variable X is actually a function of t (time). If we think of t as representing the "current" time, then the value of X "now" is written as $X(t)$. Since each step of Euler's method moves us forward by Δt time units, the time at the next step will be $t + \Delta t$. So what we previously called "new X" is really $X(t + \Delta t)$. Using this notation, we can rewrite the equation for Euler's method as

$$X(t + \Delta t) \approx X(t) + \Delta t \cdot X'(t)$$

We have written this as an approximation, because Euler's method does not give us the actual value of X at a later time (the red curve) but only an approximation of it (the blue jagged line).

Since we want to gain an understanding of X', let's rearrange this equation to solve for $X'(t)$. First, subtracting $X(t)$ from both sides yields

$$X(t + \Delta t) - X(t) \approx \Delta t \cdot X'(t)$$

We then turn the equation around and divide by Δt to get

$$X'(t) \approx \frac{X(t + \Delta t) - X(t)}{\Delta t}$$

In other words, the rate of change of X is approximately the difference between two values of X at two slightly different times, divided by the difference in those times. We will now turn to explicating precisely what this means and how we can find X' exactly rather than just approximating it. This is the subject generally called "calculus."

2.2   Derivatives: Rates of Change

Instantaneous Rates of Change

In Chapter 1, we emphasized that *quantities are changed by rates*, and that if X represents a quantity, then X', the rate at which X is changing, must be "quantity per unit time."

"Per" always means "divided by," so the rate of change of a quantity should then be a change in the quantity divided by the change in time:

$$\text{rate of change} = \frac{\text{change in quantity}}{\text{change in time}}$$

Suppose you drive from point A to point B. Here, the quantity is "distance," and the rate of change of distance with respect to time is called "velocity" or "speed."[1]

Let's say the total distance from A to B was 10 miles, and your trip took a half hour. Then, following common sense, we can define your average speed over the whole trip as

$$\text{average speed}_{\text{from } A \text{ to } B} = \frac{10 \text{ miles}}{0.5 \text{ hours}} \quad \text{or} \quad 20 \text{ miles per hour}$$

But we can look at your average speed over any time interval. If we let $X(t)$ be your progress, the distance covered from A to B as a function of time t (Figure 2.1), then for any time interval (t_1, t_2) in that half hour, we can define your average speed over that time interval as

$$\text{average speed}_{(t_1, t_2)} = \frac{\text{change in distance}}{\text{change in time}} = \frac{X(t_2) - X(t_1)}{t_2 - t_1}$$
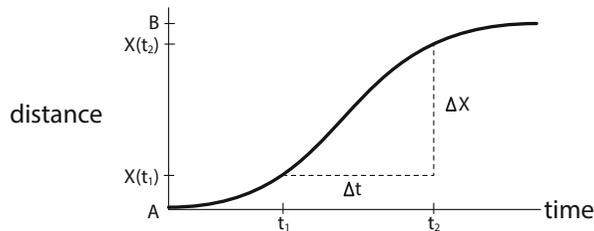

Figure 2.1: An example of distance X covered from A to B as a function of time t.

We will use a standard notation: the change in t, from t_1 to t_2, we will call Δt, So

$$\Delta t = t_2 - t_1$$

and the corresponding change in X, $X(t_2) - X(t_1)$, we will call ΔX:

$$\Delta X = X(t_2) - X(t_1)$$

So

$$\text{average speed}_{(t_1, t_2)} = \frac{X(t_2) - X(t_1)}{t_2 - t_1} = \frac{\Delta X}{\Delta t}$$

Exercise 2.2.1 A bowling lane is 60 feet long. If a bowling ball is released at $t = 0$ and reaches the pins 2.5 seconds later, what is its average speed?

If we now choose a smaller time interval (t_1, t_3), we get a smaller ΔX over the smaller Δt (Figure 2.2), and a new estimate of average speed, over this shorter interval (shown in red). If we then take t_3, t_4, etc., closer and closer to t_1, then we get a succession of average speeds over shorter and shorter intervals.

---

[1] In physics, "velocity" means "speed plus direction," so velocities can be positive or negative. "Speed" is a more colloquial term, and it is generally thought of as only positive. So, if your car was backing up, we would say that your *speed* going backward was 5 miles/hour, and your *velocity* was −5 miles/hour.
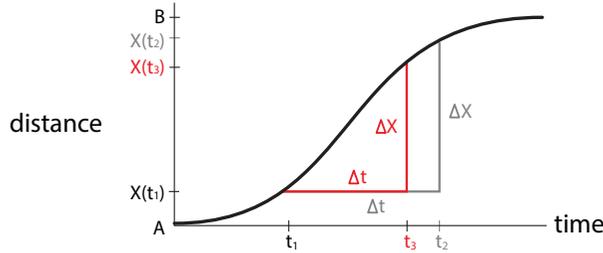
Figure 2.2: Choosing a smaller Δt gives an average velocity over a shorter time interval.

Exercise 2.2.2 Let's say we are given the functional form of the curve in Figure 2.2:

$$f(t) = (B - A)\frac{t^4}{1 + t^4} + A$$

Here we assume $A = 0$ and $B = 1$. Calculate estimates of the average speed over several intervals beginning at $t = 1$, say $\Delta t = 0.5, 0.2$, and 0.1.

Clearly, we can compute average speed over any time interval, no matter how short. Now we want to go further and ask what might seem like an odd question: we talked about your *average* speed over any given *time interval*. Does it make any sense to talk about your speed *at* a *point* in time? Can we make sense of the concept of your *instantaneous* speed at a time t_0?

On the one hand, it makes perfect sense to say "well, at *exactly* 1:15 p.m., when I was partway there, I was definitely going at some speed or other." But on the other hand, if you tried to apply the definition of average speed, you would get

$$\text{instantaneous speed}_{\text{at 1:15 pm}} = \frac{\Delta X}{\Delta t} = \frac{0}{0}$$

which is absurd.

This paradox was known to the ancient Greeks (look up *Zeno's paradoxes*), but it wasn't really answered until the 1600s, with the work of Newton and Leibniz. They realized that the way to approach the idea of instantaneous velocity at t_0 is to look at the average velocity over a *small* interval, from t_0 to $t_0 + \Delta t$, and then let that interval get smaller and smaller, approaching zero; that is, let $\Delta t \to 0$.[2]

If that process produced an actual number as its limiting value (not $\frac{0}{0}$), then we could well call that limiting value the *instantaneous velocity at* t_0.

We can now define the *instantaneous speed* at t_0 to be the value that these successive approximations approach as Δt gets closer and closer to 0, more formally, the *limit* of these values as Δt approaches 0:

$$\text{instantaneous speed}_{t_0} = \lim_{\Delta t \to 0} \frac{\Delta X}{\Delta t} = \lim_{\Delta t \to 0} \frac{X(t_0 + \Delta t) - X(t_0)}{\Delta t}$$

---

[2]Actually, Newton and Leibniz tried to reason using the concept of an "infinitesimally small quantity." It wasn't until the 1800s that the idea of instantaneous velocity was put on a rigorous foundation using the notion of *limits*. In the 1960s, the notion of "infinitesimally small quantity" was made rigorous by UCLA mathematician Abraham Robinson in his nonstandard analysis.

Exercise 2.2.3 If at some instant an object's speed is 30 $\frac{mi}{h}$, will it travel 30 miles in the next hour?

We began this chapter using Euler's method to get an *approximation* for $X'(t)$, and concluded that

$$\underbrace{X'(t)}_{\substack{\text{instantaneous} \\ \text{rate of change}}} \quad \approx \quad \underbrace{\frac{X(t + \Delta t) - X(t)}{\Delta t}}_{\text{average rate of change}}$$

Now we can say that $X'(t)$, the left-hand side of this equation, is the instantaneous rate of change of X at time t, and the right-hand side of this equation is exactly the average rate of change of X from time t to time $t + \Delta t$, as we just defined it. In Euler's method, we learned that we can make the approximation more accurate by making Δt very close to 0.

This connection between average rates of change and instantaneous rates of change is the foundation for the subject that is called "calculus."

Example: A Falling Object

Legend has it that Galileo dropped balls from the Leaning Tower of Pisa and measured their time of flight. This is not true. The time intervals involved are too short for him to measure using then-existing technology. What he actually did was slow the process down by rolling balls down an inclined plane and measuring the time intervals with a water clock.

He then summarized his findings in a law that is applicable to falling objects.

Let $H(t)$ be the height of the ball above the ground t seconds after we let go of it. According to Galileo, if we ignore air resistance slowing the ball down, its height will be

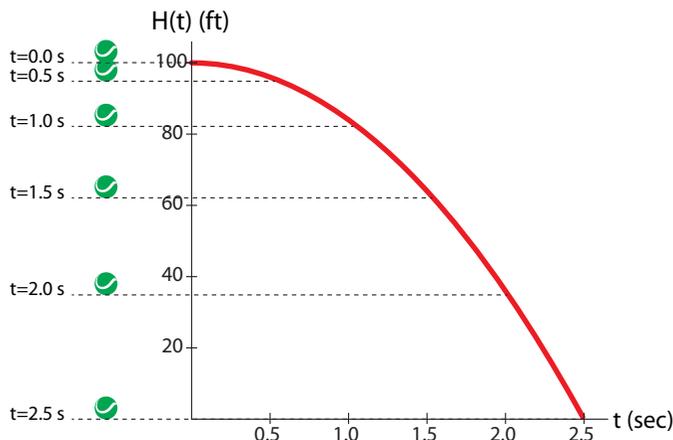$$H(t) = H(0) - 16t^2$$

where $H(0)$ is the initial height.[3]

Figure 2.3: Graph of $H(t) = 100 - 16t^2$, representing the height H of the ball, t seconds after being dropped from an initial height of 100 feet.

---

[3] The value "16" results from the assumption that H is in feet and t is in seconds.

Based on this, can we say how fast the ball is falling *exactly* 1.5 seconds after we release it from an initial height of 100 ft (Figure 2.3)?

Since we are asking for the instantaneous velocity at a time t, which is the instantaneous rate of change of the function H at $t = 1.5$, we are looking for $H'(1.5)$.

How do we compute $H'(1.5)$? By considering the average rate of change of H over various time intervals, and then letting the time intervals get smaller and smaller, that is, making Δt approach 0.

Let's begin by considering a time interval of 0.1 s, from $t = 1.5$ to $t = 1.6$, so

$$\Delta t = 1.6 - 1.5 = 0.1\,\text{s}$$

The average rate of change of H over this time interval is

$$
\begin{aligned}
\text{average rate of change} &= \frac{H(1.5 + \Delta t) - H(1.5)}{\Delta t} \\
&= \frac{H(1.6) - H(1.5)}{0.1} \\
&= \frac{(100 - 16 \cdot 1.6^2) - (100 - 16 \cdot 1.5^2)}{0.1} \\
&= -49.6\,\frac{\text{ft}}{\text{s}}
\end{aligned}
$$

Exercise 2.2.4   Notice that our calculation results in a negative number. Why does this make sense?

The value $\Delta t = 0.1\,\text{s}$ represents a fairly short time interval, so we can consider this to be an approximation of $H'(1.5)$:

$$H'(1.5) \approx \frac{H(1.6) - H(1.5)}{0.1} = -49.6\,\frac{\text{ft}}{\text{s}}$$

As in Euler's method, we can make this approximation better by using a smaller Δt. If we redo the calculation with time interval $\Delta t = 0.01$, we get

$$H'(1.5) \approx \frac{H(1.51) - H(1.5)}{0.01} = \frac{(100 - 16 \cdot 1.51^2) - (100 - 16 \cdot 1.5^2)}{0.01} = -48.16\,\frac{\text{ft}}{\text{s}}$$

We can get sharper estimates of $H'(1.5)$ by using even smaller values of the time interval Δt, for example, $\Delta t = 0.001$.

$$H'(1.5) \approx \frac{H(1.501) - H(1.5)}{0.001} = \frac{(100 - 16 \cdot 1.501^2) - (100 - 16 \cdot 1.5^2)}{0.001} = -48.016\,\frac{\text{ft}}{\text{s}}$$

Exercise 2.2.5   Approximate $H'(1.5)$ using the time interval $\Delta t = 0.0001$.

Estimates with smaller and smaller values of Δt have resulted in a series of estimates of $H'(1.5)$. The actual value of $H'(1.5)$ is the limit of these estimates as Δt approaches 0. But what is that limit?

Exercise 2.2.6   Why do we not allow Δt to reach 0?

Exercise 2.2.7   Use successive approximations to find the object's speed at $t = 1$ second.

Finding $H'(t)$

In our example, the successive estimates of $H'(1.5)$ are -49.6, -48.16, and -48.016. These estimates look like they are getting closer and closer to 48. But how can we be sure that this is the exact value of $H'(1.5)$?

We can answer this mathematically by doing a symbolic calculation. Instead of using specific values of Δt, as above, we will do a symbolic calculation using the symbol Δt.

$$\begin{array}{cc} \text{average rate of change} \\ \text{at t = 1.5} \end{array} = \frac{H(1.5 + \Delta t) - H(1.5)}{\Delta t} \qquad (2.1)$$

We know the function H, and so we can plug it into equation 2.1. Since $H(t) = 100 - 16t^2$, we can compute

$$H(1.5) = 100 - 16 \cdot 1.5^2 = 64$$

The next quantity we need is $H(1.5 + \Delta t)$:

$$\begin{aligned} H(1.5 + \Delta t) &= 100 - 16 \cdot (1.5 + \Delta t)^2 \\ &= 100 - 16 \cdot \left(1.5^2 + 3\Delta t + (\Delta t)^2\right) \\ &= 100 - 36 - 48\Delta t - 16 \cdot (\Delta t)^2 \\ &= 64 - 48\Delta t - 16 \cdot (\Delta t)^2 \end{aligned}$$

Substituting these two expressions into equation 2.1 gives us

$$\begin{array}{cc} \text{average rate of change} \\ \text{at t = 1.5} \end{array} = \frac{\left(64 - 48\Delta t - 16 \cdot (\Delta t)^2\right) - 64}{\Delta t}$$

Notice that the 64's in the denominator cancel each other (not a coincidence).

$$\begin{aligned} \begin{array}{cc} \text{average rate of change} \\ \text{at t = 1.5} \end{array} &= \frac{\left(\cancel{64} - 48\Delta t - 16 \cdot (\Delta t)^2\right) - \cancel{64}}{\Delta t} \\ &= \frac{-48\Delta t - 16 \cdot (\Delta t)^2}{\Delta t} \\ &= \frac{\cancel{\Delta t} \cdot (-48 - 16\Delta t)}{\cancel{\Delta t}} \\ &= -48 - 16\Delta t \end{aligned}$$

We have now found a general expression for the average rate of change at $t = 1.5$ as a function of Δt, and it is obvious what will happen as Δt approaches 0:

$$-48 - 16\Delta t \longrightarrow -48 \quad \text{as} \quad \Delta t \longrightarrow 0$$

which gives us the exact value of $H'(1.5)$ as

$$H'(1.5) = \lim_{\Delta t \to 0} \frac{H(1.5 + \Delta t) - H(1.5)}{\Delta t} = \lim_{\Delta t \to 0} (-48 - 16\Delta t) = -48$$

Exercise 2.2.8   Carry out a similar calculation for $t = 2$.

The procedure that we just applied to find $H'(1.5)$ can be generalized to any t, and we have now developed a general procedure for finding $H'(t)$:

$$H'(t) = \lim_{\Delta t \to 0} \frac{H(t + \Delta t) - H(t)}{\Delta t}$$

Carrying out this calculation, we get

$$H'(t) = \lim_{\Delta t \to 0} \frac{H(t + \Delta t) - H(t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\left(H(0) - 16(t + \Delta t)^2 \right) - H(0) - 16t^2}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\left(\cancel{H(0)} \; \cancel{-16t^2} - 32t \cdot \Delta t - \Delta t^2 \right) - \cancel{H(0)} \; \cancel{-16t^2}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\cancel{\Delta t} \cdot (32t - \Delta t)}{\cancel{\Delta t}}$$

$$= \lim_{\Delta t \to 0} (-32t - \Delta t)$$

$$= -32t$$

So we can now say, for the function $H(t) = H(0) - 16t^2$, that we can calculate $H'(t)$ for any t_0.

Exercise 2.2.9   Use this result to find the object's velocity at $t = 2$.

The derivative of $H(t)$ at the point t_0 is the limit as $\Delta t \to 0$ of the quantity $\frac{\Delta H}{\Delta t}$.

$$\left. H'(t) \right|_{t_0} = \lim_{\Delta t \to 0} \left. \frac{\Delta H}{\Delta t} \right|_{t_0}$$

In fact, this procedure can be carried out for many functions $X(t)$, including most functions that can be expressed as a formula. (For exceptions, see "Do all functions have derivatives?")

If $X(t)$ is any function of t, then we can almost always define $\left. X'(t) \right|_{t_0}$, which is the instantaneous rate of change of X at time t_0. This is called the *derivative* of X at the point t_0.

Variables Other Than Time

We've been talking so far about functions of time, and rates of change with respect to time, like velocity, which is the rate of change of distance with respect to time. But we can also talk about functions of *any* variable, not just time. And then we can ask about how the value of the function changes with respect to changes in the variable.

For example, suppose you are climbing a mountain or ascending in an airplane. You observe that as you go higher, the outside air pressure decreases. So air pressure is a function of altitude.

If we let $H =$ height and $P =$ air pressure, then there is some function

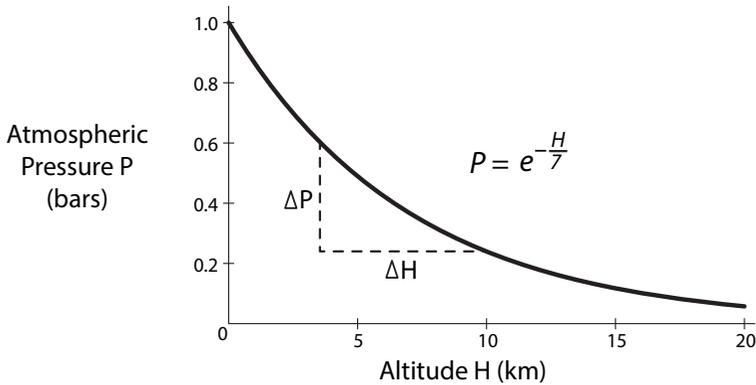$$P = f(H)$$

In fact, it looks like Figure 2.4.

Figure 2.4: Atmospheric pressure P as a function of altitude H.

We can now ask: how much does pressure change with respect to height? We could even ask: "what is the *rate* at which P is changing with respect to H?"

There are many examples in science where we are looking at one variable as a function of another variable. In fact, science is all about looking for relationships that show one variable as a function of another.

In chemistry, we study the properties of gases, such as their pressures and volumes. We know that if we put a gas under pressure, say from a piston, the volume of the gas decreases. So there is a function that gives the volume V for a given value of pressure P, $V = f(P)$. (This is called Boyle's law.) And again, we can talk about the rate at which V is changing with respect to P.

Astrophysics studies the properties of stars, such as their distance from us and their velocities (which can be figured out from the color of the light they emit). There is a famous law, called Hubble's law, that says that the velocity with which a star is receding from us is a function of its distance from us.

In biology, the subject of *allometry* studies the basic physical measurements that can be made on animals, like body length, skull size, heart rate, and metabolic rate. It studies how these characteristics are related to other physical characteristics, such as body mass. It is interesting to look at how these scale with each other over a very wide range of sizes, from ants to whales. For example, let's say we are looking at how heart rate H scales with body mass M. We look at the data from thousands of species and find that they lie on a curved line, giving us $H = f(M)$. Depending on the shape of the curve, we can talk about the rate at which H is changing with respect to M.

So any time one quantity Y can be expressed as a function of some other quantity X, we can ask: if X changes, how much will Y change in response? Let's define this concept of rate precisely.

For example, let's go back to our example of the relation between air pressure P and altitude H (Figure 2.4). We can define the concept of the rate of change of P with respect to H exactly as we did in defining a rate of change with respect to time. First we define an average rate of change over some interval as $\frac{\Delta P}{\Delta H}$.

For example, if we went from an altitude of $H = 2\,\text{km}$ to $H = 5\,\text{km}$, then the *change in altitude* is

$$\Delta H = 5 - 2 = 3\,\text{km}$$

We will see the atmospheric pressure drop, from 0.75 bars at $H = 2\,\text{km}$ down to 0.49 bars at $H = 5\,\text{km}$. Therefore, the *change in pressure* is

$$\Delta P = 0.49 - 0.75 = -0.26\ bars$$

We then define the average rate of change, over the interval from $H = 2\,\text{km}$ to $H = 5\,\text{km}$, of atmospheric pressure (P) with respect to altitude (H), by

$$\text{average rate of change of } P \text{ with respect to } H = \frac{\Delta P}{\Delta H} = \frac{-0.26\,\text{bar}}{3\,\text{km}} = -0.09\,\frac{\text{bar}}{\text{km}}$$

This is the average rate of change over an interval.

It then makes perfect sense to do exactly what we did with respect to time: pick an arbitrary point H_0, and define the *instantaneous* rate of change of P with respect to H at the point H_0 as the limit of the average rate of change as the interval ΔH approaches to zero:

$$\text{instantaneous rate of change of } P \text{ with respect to } H \text{ at } H_0 = \lim_{\Delta H \to 0} \frac{\Delta P}{\Delta H}\bigg|_{H_0}$$

Exercise 2.2.10 The function describing how air pressure varies with elevation is

$$P(H) = 101352e^{-\frac{0.28H}{2396}}$$

where P is measured in pascals and H in meters. Approximate the rate of change of P with respect to H at a height of 2000 meters.

Notation

We have now defined a concept: for any $Y = f(X)$, at any point X_0, the instantaneous rate of change of Y with respect to X is

$$\lim_{\Delta X \to 0} \frac{\Delta Y}{\Delta X}\bigg|_{X_0} = \lim_{X \to X_0} \frac{f(X) - f(X_0)}{X - X_0}$$

The only question is: what to call this? We have been using Y' to mean the derivative with respect to time, and we can extend this to allow Y' to mean the derivative with respect to some other variable, but the only problem is that we have no way of saying what that other variable is. We don't have a terminology yet for the instantaneous rate of change with respect to some arbitrary variable.

We are rescued by Gottfried Leibniz, the co-inventor of calculus in the late 1600s, along with Isaac Newton. Newton had favored a notation something like our X' (actually an X with a dot over it, $\dot{X}$), so X' is called the Newtonian form. Leibniz, on the other hand, wanted to stress that this is a ratio of two quantities, ΔX and Δt, so he adopted somewhat odd notation. He looked at the ratio $\frac{\Delta X}{\Delta t}$ and decided to refer to $\lim_{\Delta t \to 0} \frac{\Delta X}{\Delta t}$ as a ratio he called $\frac{dX}{dt}$.

$$\textbf{Leibniz} \qquad \frac{dX}{dt} = \lim_{\Delta t \to 0} \left.\frac{\Delta X}{\Delta t}\right|_{t_0}$$

$$\textbf{Newton} \qquad X' = \lim_{\Delta t \to 0} \left.\frac{\Delta X}{\Delta t}\right|_{t_0}$$

There is a clear drawback to the Leibniz notation. What is "dX"? What is "dt"? How can we take their ratio if we don't know what the individual terms mean? And why can't we divide top and bottom by d?

The answer to all these questions is that the Leibniz notation can't really be read as the ratio of two anythings, and the terms dX and dt don't really mean anything by themselves.[4]

Rather, the whole expression

$$\text{``} \left.\frac{dX}{dt}\right|_{t_0} \text{''} \qquad \textbf{means} \qquad \text{``} \lim_{\Delta t \to 0} \left.\frac{\Delta X}{\Delta t}\right|_{t_0} \text{''}$$

The big advantage of the Leibniz notation is that we can now state explicitly both of the variables in the limit. This makes it possible to return to our original question what to call $\lim_{\Delta X \to 0} \left.\frac{\Delta Y}{\Delta X}\right|_{X_0}$, where X is arbitrary. We will call it

$$\left.\frac{dY}{dX}\right|_{X_0} \qquad \text{or} \qquad \left.\frac{df}{dX}\right|_{X_0}$$

So everywhere in this text, when we say X', we usually mean $\frac{dX}{dt}$, but we will sometimes use the notation Y' for convenience when the relevant variable is obvious. And when we want to refer to a function of some other variable, such as $P = f(H)$, we will usually call the instantaneous rate of change $\frac{dP}{dH}$.

"Sensitivity"

The quantity we just defined as the "instantaneous rate of change of Y with respect to X" can also be seen as the definition of the concept of "sensitivity." When we are talking about the "sensitivity of Y to X," we are really talking about the quantity $\frac{dY}{dX}$.

Suppose, for example, we are looking at a drug for cancer chemotherapy. We run experiments and determine what percent of cancer cells are still alive when we give D amount of drug. If we call the percentage of cancer cells still alive P, then our experiments give us $P = f(D)$. A typical graph might look like this:

---
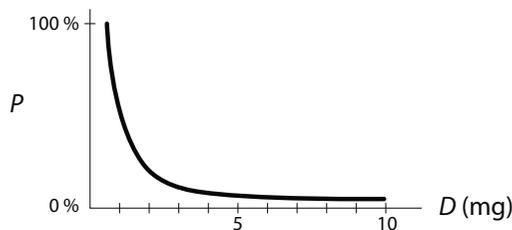
[4]In fact, Leibniz was criticized by many of his fellow mathematicians for this, and for centuries his notion of dX as an "infinitesimal" quantity was frowned upon. The philosopher Bishop Berkeley ridiculed $\frac{dX}{dt}$ as "the ratio of the ghosts of two departed quantities." It was not until the 1960s, nearly 300 years later, that Leibniz was fully vindicated when UCLA math professor Abraham Robinson came up with an idea called nonstandard analysis, which provided a mathematically rigorous foundation for infinitesimals.

We can then talk about the sensitivity of the cancer cells to increasing drug dosages. What we mean is

$$\left.\frac{dP}{dD}\right|_{D_0}$$

So we can say, for example, that for dosages below 2 milligrams, the cancer cells are highly sensitive to the drug, because $\left.\frac{dP}{dD}\right|_{D_0}$ is more negative when $D_0 < 2$.

Further Exercises 2.2

1. The rate of change of the position of a car at some time t_0 is given by $\frac{dX}{dt} = 55$. What does this mean in plain English?

2. You are studying the athletic performance of runners. You have two motion-triggered cameras that produce time-stamped photographs.

   a) The runner reaches the first camera, at the 500 m mark, at 9:03:05 a.m. and the second camera, at the 600 m mark, at 9:03:25 a.m. What is her average speed over that time interval?

   b) When is she running at that speed?

   c) How could you change your measurement setup (without getting new equipment) to better approximate the runner's instantaneous speed at 500 m?

3. You are an ecologist studying bottom-dwelling stream invertebrates. You need to measure the speed at which the water is flowing at a particular point you have chosen to study. You have a stopwatch, a long measuring tape, a supply of Ping-Pong balls (which float and are easy to see), and brightly colored flags that can be used to mark points along the shore or in the water. How would you use this equipment to estimate the instantaneous speed of the water? (You may want to include a diagram with your response.)

4. Use successive approximations to approximate the derivative of the functions below at the points specified.

   a) $f(X) = 6X^5$ at $X = 2$

   b) $f(x) = 7X^3 + 2$ at $X = 3.5$

   c) $f(X) = \sin X$ at $X = -3$ (use radians)

   d) $f(X) = \sin\left(\ln(X^3 + 1)\right)$ at $X = 4$

5. You are studying a new blood-pressure-lowering drug. You find that blood pressure is not very sensitive to the drug at low doses, very sensitive at intermediate doses, and not very sensitive at high doses. Rephrase this statement in terms of $\frac{dP}{dM}$, where P is blood pressure and M is the drug dosage.

2.3   Derivatives: A Geometric Interpretation

From Secant to Tangent

Let's now look at the concept "$\frac{dY}{dX}$" geometrically. First, let's make a geometric picture of the *average* rate of change of a function at a point. Suppose Y is a function of X. At the point X_1, the average rate of change of Y with respect to X over the interval (X_1, X_2) is

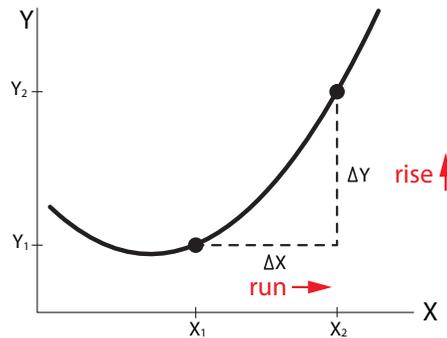$$\left.\frac{\Delta Y}{\Delta X}\right|_{X_1} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

Figure 2.5: An example of ΔX (run) and its corresponding ΔY (rise).

Looking at this geometrically, we see that ΔY is the change in the vertical direction, and ΔX is the change in the horizontal direction (Figure 2.5). (These are sometimes called "rise" and "run.")

What we want is the average rate of change, that is, the quantity $\frac{\Delta Y}{\Delta X}$. We can visualize this quantity by drawing the blue straight line directly connecting the two points (X_1, Y_1) and (X_2, Y_2) (Figure 2.6). The slope of this line is

$$\frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\Delta Y}{\Delta X}$$
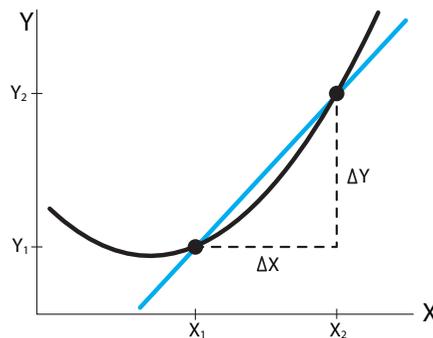
Figure 2.6: Secant line connecting the point (X_1, Y_1) to the point (X_2, Y_2), where $(X_2, Y_2) = (X_1 + \Delta X, Y_1 + \Delta Y)$.

This line is called the *secant*[5] to the curve through these two points. We can say a lot about this blue secant line. The crucial concept here is the notion of *slope*. The slope of a straight line is defined as $\frac{\Delta Y}{\Delta X}$ taken over any two points on the line. This is exactly the concept we need:

$$\left.\frac{\Delta Y}{\Delta X}\right|_{X_1} = \text{slope of the secant line connecting } (X_1, Y_1) \text{ and } (X_2, Y_2)$$

To summarize,

$$\text{average rate of change} = \text{slope of secant} = \frac{\Delta Y}{\Delta X}$$

Exercise 2.3.1   Calculate the slope of the secant line to the graph of $Y = \frac{X}{1+X}$ from $X = 1$ to $X = 3$.

Now that we have defined the average rate of change $\frac{\Delta Y}{\Delta X}$, we want to let ΔX get smaller and smaller, in order to get a geometric picture of $\frac{dY}{dX}$, which is the limit of $\frac{\Delta Y}{\Delta X}$ as ΔX approaches 0.

As ΔX gets smaller and smaller, the blue secant lines cut through smaller and smaller portions of the curve near X_1 (Figure 2.7).
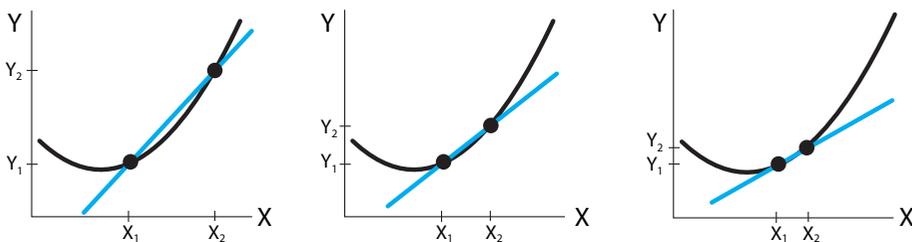

Figure 2.7: The slope of secant lines gradually changes as X_2 approaches X_1.

As we do this, the blue secant line gets closer and closer to the curve, until finally it approaches a line that "just touches" the curve at the point (X_1, Y_1).[6] This is the line shown in red (Figure 2.8). This limiting red line is called the *tangent line*[7] to the curve $Y = f(X)$ at the point (X_1, Y_1).

---

[5]From the Latin word "secare," meaning "to cut."

[6]The notion of "just touches" is being left slightly vague here. And the concept "just touches" doesn't even work for certain examples, like $f(X) = X^3$ at $X = 0$, where the tangent line is a horizontal line cutting through the curve. In fact, the true definition of "tangent" *requires* the concept of derivative. The tangent is the line whose slope is equal to the derivative of the function at that point.

[7]The word "tangent" comes from the Latin *tangere*, meaning "to touch."
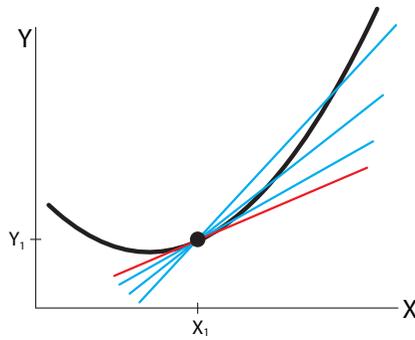
Figure 2.8: The limit of the secant process, as X_2 approaches X_1, is the red line, called the tangent to the black curve at the point (X_1, Y_1).

In summary,

as Δx $\longrightarrow$ 0
secant lines $\longrightarrow$ tangent line
slope of secant lines $\longrightarrow$ slope of tangent line
$\|$ $\|$
average rates of change $\longrightarrow$ instantaneous rate of change

" $\longrightarrow$ " means "converges to"

> If $Y = f(X)$ is the graph of Y as a function of X, then
> (1) the slope of the line tangent to the curve $Y = f(X)$ at the point X_0
> $=$ (2) the derivative of f at the point X_0, $\frac{df}{dX}\big|_{X_0}$ (or equivalently, the derivative of Y with respect to X, $\frac{dY}{dX}\big|_{X_0}$.)
> $=$ (3) the instantaneous rate of change of Y with respect to X (or f with respect to X) at the point X_0.

Exercise 2.3.2   Find the slope of the secant line crossing the graph of $f(t) = 200 - 16t^2$ at the following values of t. What value is the slope approaching?

a) $t = 2$, $t = 2.5$ b) $t = 2$, $t = 2.1$ c) $t = 2$, $t = 2.05$

The Equation of the Tangent Line

We now know that the quantity $\frac{dY}{dX}\big|_{X_0}$ is the slope of the tangent line to $Y = f(X)$ at the point X_0 (Figure 2.8).

We can use that fact to derive the actual equation for the tangent line. The best-known form of the equation for a line is the slope–intercept form (Figure 2.9),

$$Y = mX + b \quad \text{where } m = \text{slope, } b = \text{Y-intercept}$$

There is, however, a different way of writing an equation for a line that will be more useful to us. To develop it, we start with the slope–intercept form. We know the slope m. It's $\frac{dY}{dX}\big|_{X_0}$.
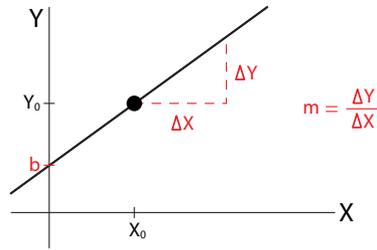
Figure 2.9: The line $Y = mX + b$ has slope m and intercepts the Y-axis at b.

But what is b? We find b by realizing that (X_0, Y_0) is a point on this line, and therefore

$$Y_0 = mX_0 + b$$

which implies

$$b = Y_0 - mX_0$$

If we substitute that back into the equation for the line, we get

$$Y = mX + b = mX + (Y_0 - mX_0)$$

Rearranging yields,

$$Y = m(X - X_0) + Y_0$$

which yields

$$(Y - Y_0) = m(X - X_0)$$

This is called the "point–slope" form of the equation for a line, since it explicitly involves the slope and a reference point on the line. Now, we can put everything together. The equation of the tangent line to $f(X)$ at $X = X_0$ is

$$(Y - Y_0) = \left.\frac{dY}{dX}\right|_{X_0} (X - X_0)$$

It is especially significant for us, since it gives us $(Y - Y_0)$ as a *linear function* of $(X - X_0)$.

Exercise 2.3.3 Find the equation of the tangent line to $f(t) = 200 - 16t^2$ at $t = 2$.

Exercise 2.3.4 Write equations for the following lines in both slope–intercept and point–slope form.
a) The line that has a slope of 2 and a Y-intercept of -54.
b) The line that has a slope of -3 and passes through the point $(2, 6)$.
c) The line that passes through the points $(1, 7)$ and $(3, 5)$.

Further Exercises 2.3

1. If for some function f, $f(2) = 5$ and $f'(2) = -3$, what is the tangent line to f at $X = 2$?

2. If some function f has the tangent line $y - 2 = 4(t - 16)$ at the point implied by the equation, what are $f(16)$ and $f'(16)$?

3. Find the tangents to the following functions at the points given. Then, graph the function and the tangent in Sage. *(Hint: You found these slopes in Further Exercise 2.2.4 on page 73.)*

   a) $f(X) = 6X^5$ at $X = 2$

   b) $f(x) = 7X^3 + 2$ at $X = 3.5$

   c) $f(X) = \sin X$ at $X = -3$ (use radians)

   d) $f(X) = \sin\left(\ln(X^3 + 1)\right)$ at $X = 4$

2.4   Derivatives: Linear approximation

Linear Functions

Throughout this book, we will often use the method of approximation by a very special class of functions, called linear functions. The equation for the tangent line is an important example of this.

Here, we will discuss the idea of linear functions in one variable. Later, we will see that all of Chapter 6 is devoted to the subject of linear functions in many variables.

In one variable, a function $Y = f(X)$ is said to be linear if it meets two conditions:

(1) $f(X_1 + X_2) = f(X_1) + f(X_2)$ for all X_1 and X_2 and
(2) $f(aX) = af(X)$ for every real number a

These are extremely strong requirements, and few functions can meet them. For example, the function $f(X) = X^2$ can't meet either of them.

Exercise 2.4.1   Verify that $f(X) = X^2$ is not a linear function. *(Hint: Apply the definition.)*

Exercise 2.4.2   Check whether $f(X) = X + 1$ is a linear function.

It turns out that the only functions that can meet the requirements for linearity are those in the family of functions

$$f(X) = kX \quad \text{where } k \text{ is a real number}$$

All linear functions of one variable have this form, and all functions having this form are linear. Notice that the relation $Y = mX + b$ is not a linear function, unless $b = 0$. It's the equation for a straight line, but it is not a linear function. The terminology is unfortunate, but at this point we have no choice but to keep this slightly confusing fact in mind.

This is why we prefer to write the equation for the tangent line in the linear point–slope form

$$\Delta Y = m \cdot \Delta X \quad \text{or} \quad \Delta Y = \left.\frac{dY}{dX}\right|_{X_0} \cdot \Delta X$$

Exercise 2.4.3   What is the complete equation for the tangent line to $Y = f(X)$ at the point $(X_0, f(X_0))$?

Zooming In on Curves

Let's expand on the theme of the derivative as a linear approximation to a function at a point. Look at the graph of Y as a function of X and its tangent line at the point X_0 in Figure 2.10. As we zoom in on that point, the curve looks more and more like the tangent line.
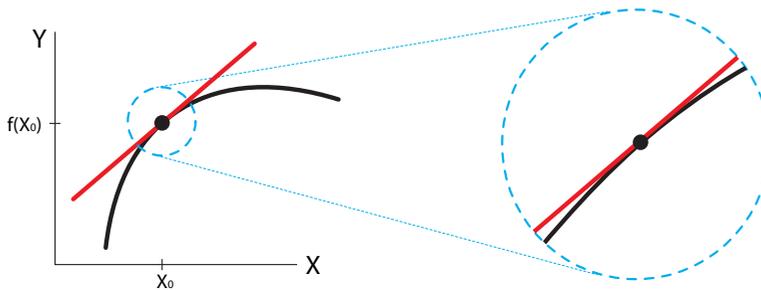

Figure 2.10: A tangent line (red) to a curve (black) at a point (black dot). Zooming in at the black dot, the curve begins to resemble the tangent line.

We can make this intuitive idea precise by realizing that near X_0, the line is an *approximation* to the curve.

$$\textbf{line} \qquad Y - Y_0 = \frac{df}{dX}\Big|_{X_0} \cdot (X - X_0)$$

$$\textbf{curve} \qquad Y - Y_0 \approx \frac{df}{dX}\Big|_{X_0} \cdot (X - X_0)$$

To put it another way, we know that the average rate of change $\frac{\Delta Y}{\Delta X}\big|_{X_0}$ is an approximation to $\frac{dY}{dX}\big|_{X_0}$. In symbols,

$$\frac{\Delta Y}{\Delta X}\Big|_{X_0} \quad \approx \quad \frac{dY}{dX}\Big|_{X_0}$$

This approximation gets better and better as ΔX approaches 0.

So as ΔX approaches 0, the line $\Delta Y = \frac{df}{dX}\big|_{X_0} \cdot \Delta X$ is a better and better approximation to the curve f at the point X_0

$$\Delta f \approx \frac{df}{dX}\Big|_{X_0} \cdot \Delta X$$

Exercise 2.4.4 In SageMath, pick a function and a point on the function. Plot the function at several magnification levels. Describe what you see.

Linear Approximation

Since the tangent line is an approximation to a function at a point, we can use it to find approximate values of the function near the point. In particular, the $\Delta Y = \frac{dY}{dX}|_{X_0} \cdot \Delta X$ form of the equation for the tangent line makes it natural to calculate the change in Y produced by a change in X.

Let's look at our example of atmospheric pressure P as a function of height H above sea level. In this case, we can say,

$$\Delta P \approx \left.\frac{dP}{dH}\right|_{H_0} \cdot \Delta H \quad \text{when } \Delta H \text{ is small}$$

Suppose that at some H_0, the rate of change of P with respect to H is $-0.1 \frac{\text{bars}}{\text{km}}$. Then we can say that if the airplane goes a little bit higher, say $\Delta H = 0.01$ km, then the atmospheric pressure will have changed by approximately

$$\Delta P \approx \left(-0.1 \frac{\text{bars}}{\text{km}} \right) \cdot \left(0.01 \, \text{km} \right) = -0.001 \, \text{bars}$$

Note that we are estimating the effect of a small change ΔH in the nonlinear function $P(H)$ at a point H_0 using the linear approximation to the function H_0. This will result in a small error in the estimate of ΔP, an error that will get smaller and smaller as ΔH approaches 0 (Figure 2.11). It is in this sense that the line

$$\Delta Y = \left.\frac{dY}{dX}\right|_{X_0} \cdot \Delta X$$
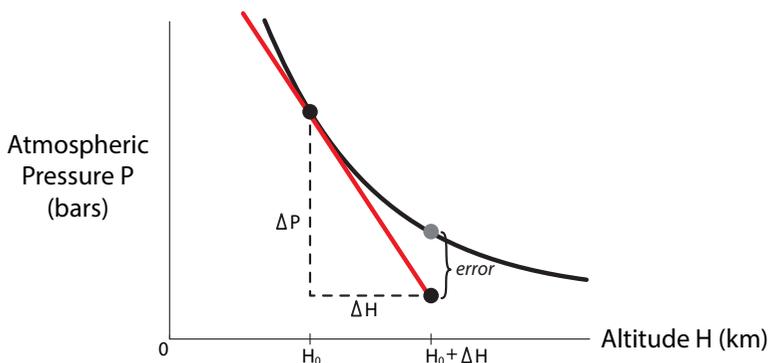
is a linear approximation to f at the point X_0.

Figure 2.11: The tangent line (red) to the curve (black) of pressure P as a function of altitude H is an approximation to the curve. The error gets smaller as ΔH decreases.

Exercise 2.4.5   In the example of the falling object, we calculated its velocity $H'(1.5)$, the rate of change of height with respect to time, at 1.5 seconds after it was released. We got the answer $-48\,\frac{\text{ft}}{\text{s}}$. Now estimate how far the ball will drop in the next 0.01 seconds. In other words, let $\Delta t = 0.01$ seconds, and calculate an approximate value for ΔH.

Exercise 2.4.6   The equation for the height of the falling ball is

$$H(t) = H(0) - 16t^2$$

Use this equation to calculate the actual change in H from $t = 1.5\,\text{s}$ to $t = 1.51$ s. How close is this actual ΔH to the ΔH you calculated in Exercise 2.4.5?

Summary

We have now seen three concepts of the derivative $\frac{dY}{dX}\big|_{X_0}$.
   (1) as the rate of change of Y with respect to X at the point X_0
   (2) as the slope of the tangent line to $Y = f(X)$ at the point X_0
   (3) as the linear approximation to $Y = f(X)$ at the point X_0
   Of these, the last is the most important: it is the idea of the derivative as a linear approximation that generalizes very naturally to n dimensions. This will be our focus in Chapter 6 and Chapter 7.

All Functions Differentiable?

If a function has a derivative at a point, we say it is differentiable at that point. If a function is differentiable at some point, it has a unique tangent at that point. What conditions does the function have to meet to have a unique tangent at a given point?
   First of all, it must be *continuous* at that point. A function is continuous at a point if the curve through the point can be drawn without lifting the pen from the paper. For example, the function

$$f(X) = \begin{cases} X^2 & 0 \le X \le 2 \\ X^3 & 2 < X \le 3 \end{cases}$$

is not continuous at the point $X = 2$ (Figure 2.12). We can't even discuss the derivative at the point $X = 2$ because there is no linear approximation to the right of $X = 2$ in the function X^2, and no linear approximation to the left of $X = 2$ in the function X^3. No matter how much we zoom in on $X = 2$, the function never looks like a straight line through the point.
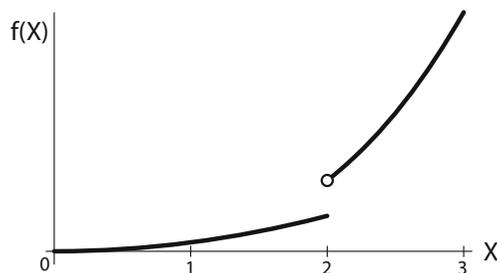

Figure 2.12: This function $f(X)$ is discontinuous at $X = 2$, therefore the derivative at $X = 2$ does not exist.

But even when the function is continuous, it still may not be differentiable. Consider

$$g(X) = \begin{cases} 0 & X \le 0 \\ X & 0 \le X \end{cases}$$

and look at the point $X = 0$. The function g cannot have a derivative at $X = 0$, because to the left of $X = 0$ it has slope 0, and to the right of $X = 0$ it has slope 1 (Figure 2.13).
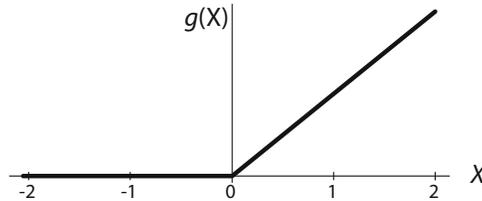

Figure 2.13: This function $g(X)$ is continuous at $X = 0$, but the kink means that the slope to the left of $X = 0$ is 0, and the slope to the right of $X = 0$ is a positive number, so there is no derivative at $X = 0$.

The lack of a derivative at $X = 0$ is also clear when we look closely at the function $g(X)$ near $X = 0$.

When we first defined the concept of derivative, we said that the derivative is the slope of the tangent to the curve, and the tangent to the curve can be visualized by zooming in closer and closer until the curved function resembles a straight line.

But when we zoom in on the function $g(X)$ near $X = 0$, we see the problem: *the function never resembles a straight line, no matter how much we zoom in* (Figure 2.14).
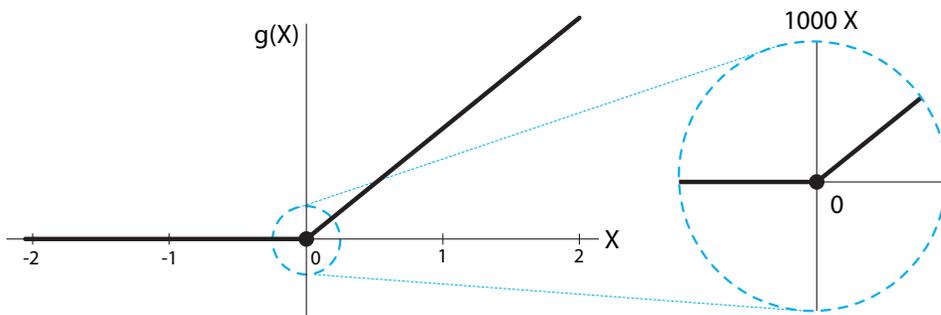

Figure 2.14: As we zoom in on the point $X = 0$, the function does not get flatter and flatter, because there is a corner there.

So not all functions have a linear approximation that becomes better and better as we zoom in. In particular, a cusp or corner will always look like a cusp or corner, regardless of the scale at which we view the function. Therefore, the function does not have a linear approximation at the cusp or corner and is not differentiable there (Figure 2.15).
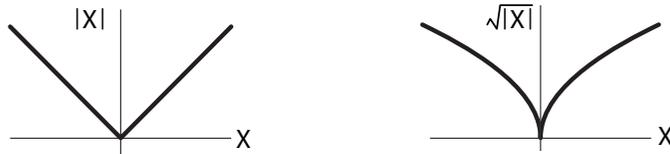
Figure 2.15: The functions $|X|$ and $\sqrt{|X|}$ have corners or cusps at $X = 0$ and are not differentiable there.

For the sake of completeness, we will mention a way for a continuous function without cusps or corners to have a point where it is not differentiable. This happens when the function has a vertical tangent at some point. Since the derivative is the slope of the tangent and the slope of a vertical line is undefined (it's infinite, and infinity is not a number), a function is not differentiable where it has a vertical tangent (Figure 2.16).
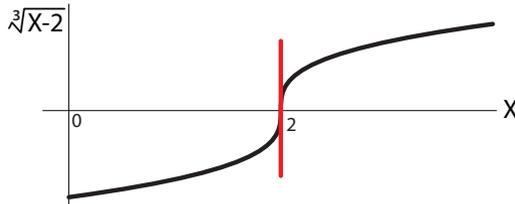

Figure 2.16: The function $\sqrt[3]{X - 2}$ has infinite slope at $X = 2$ and so is not differentiable there.

Exercise 2.4.7 View $f(X) = |X|$ at several zoom levels and show that the corner at $X = 0$ remains a sharp corner no matter how closely you zoom in. Briefly explain why this means that it does not have a derivative at $X = 0$.

Exercise 2.4.8 Is the function in Figure 2.15 differentiable at all points shown other than $X = 0$?

Further Exercises 2.4

1. You are studying a new blood pressure drug. At a dose of 5 mg, the slope of the dose–response curve is $-2\ \frac{\text{mmHg}}{\text{mg}}$. Approximately how much would a patient's blood pressure change if the drug dose was increased to 5.1 mg?

2. Suppose $g(N)$ measures the size of tomatoes produced by a tomato plant as a function of the amount N of nitrogen that is available to the plant.

    a) Explain in plain English (without using the word "derivative") what the quantity $\frac{dg}{dN}$ means.

    b) If at some instant $\frac{dg}{dN}$ was equal to 5, and N was then increased by 0.04, what would you expect to happen to g? Be as specific as possible.

3. You have developed a robotic ant to help you study insect behavior. As the ant travels, it keeps track of its position and the slope of the surface it's on and mathematically models its local environment.

    a) The ant has traveled 10 cm horizontally and 6 cm vertically from its starting point on a twig with a slope of 0.5. If this is the only information the ant has, what function best approximates the geometry of the twig at the point the ant is on?

    b) The ant can use its model of the environment to plan its movements. In particular, it wants its next step to take it no higher than 0.1 cm above its current location. How far can the ant travel horizontally and still accomplish this?

4. Sketch graphs of functions that match the following descriptions:

    a) The function is discontinuous at $X = 2$ but continuous everywhere else.

    b) The function is continuous at $X = 5$ but has no tangent line there.

    c) The function is not differentiable at $X = 1$ but has a tangent line there.

2.5 The Derivative of a Function

Given a function $Y = f(X)$, we now understand the concept of the derivative of f at a point X_0.

$$\left.\frac{df}{dX}\right|_{X_0} = \lim_{\Delta X \to 0} \frac{f(X_0 + \Delta X) - f(X_0)}{\Delta X}$$

Using this definition, given any point X_0, we can assign a number to that point: the value of $\frac{df}{dX}$ at X_0.

This means that we have defined a new function from $\mathbb{R}$ to $\mathbb{R}$: the function that assigns to a point X the value of $\frac{df}{dX}$ at that point X. We call this new function the *derivative* of f, and we write it as $\frac{df}{dX}$. The process of finding $\frac{df}{dX}$ given f is called differentiating f or "taking the derivative of f."

For example, consider the upper graph in Figure 2.17. It is the graph of some function $f(X)$. At every point, f has a tangent and that tangent has a slope. On the left-hand side, the slopes are positive; in the middle region, they are negative; and in the right-hand region, they become positive again. The graph that records the slope of f at every point is the blue curve shown immediately below the graph of f. The blue curve is the graph of the function $\frac{df}{dX}$.

In general, if f is any function

$$f : \mathbb{R} \to \mathbb{R}$$

and f has a derivative everywhere, then there is another function

$$\frac{df}{dX} : \mathbb{R} \longrightarrow \mathbb{R}$$

called the derivative of f. For example, we worked out earlier that if $H(t) = H(0) - 16t^2$, then $H'(t) = -32t$.
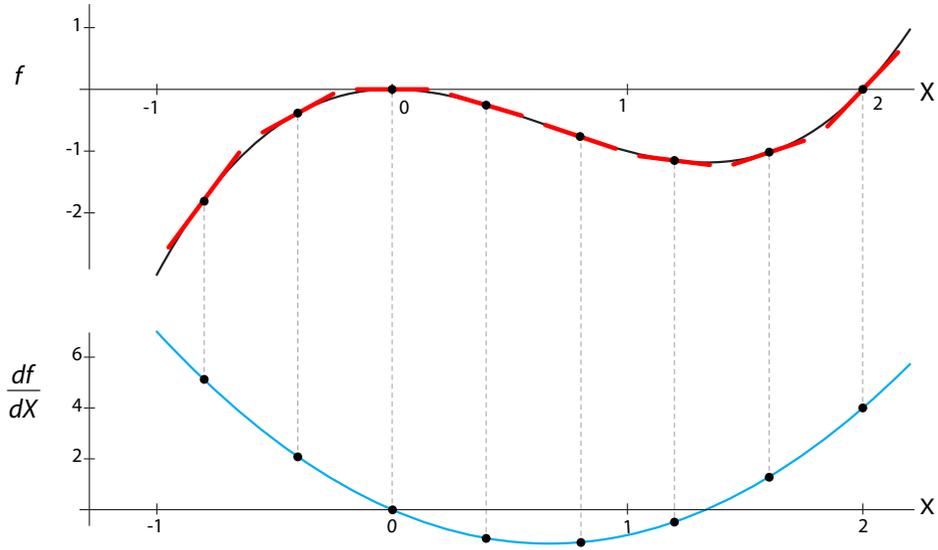
Figure 2.17: Upper: a function $f(X)$, with tangents shown (red lines) at representative points. Lower: the slopes of the tangents are plotted to form the function $\frac{df}{dX}$ (blue curve).

Exercise 2.5.1 Match each function f in the top row to its derivative f' in the bottom row. We have done the first one for you. Make sure you understand this, and then match the others.

We will now take a big step in abstraction, going meta on the whole idea of functions. You might remember that we defined the function concept quite generally, using such examples as coffee shop menus and Martian DNA. However, the functions we've actually worked with have acted on nothing more exotic than numbers and points. So what was the purpose of all that abstraction?

We have now come to a place where it is very helpful to think about functions that act on other functions. Leibniz notation suggests that we can think of "$\frac{d}{dX}$" as a function of functions, a function that takes as its input a function f and returns another function $\frac{df}{dX}$ (Figure 2.18).

Let's work an example. Let's take our falling ball, whose height at time t after release from initial height $H(0)$ is given by
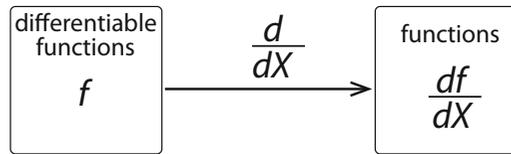
function H $\qquad H(t) = H(0) - 16t^2$

Figure 2.18: Differentiation is a function from differentiable functions to their derivatives.

We already showed that for every time t,

derivative function of H $\quad\quad \dfrac{dH}{dt} = H'(t) = -32t$

Thus we can say that "the derivative of $H(t)$ is $-32t$," understanding that what we mean is that at every point t,

$$\Delta H = (-32t) \cdot \Delta t$$

is the linear approximation to H.

Exercise 2.5.2   What does "the derivative of $f(x)$ is $7x + 4.5$" mean? Give two answers.

Given a function $Y = f(X)$, we calculate the derivative function

$$\frac{dY}{dX} \quad \left(\text{or } \frac{df}{dX} \right)$$

by finding

$$\frac{f(X + \Delta X) - f(X)}{\Delta X}$$

and letting ΔX approach 0.

Let's try an example, the function

$$Y = f(X) = X^2 + X$$

To calculate its derivative function, we plug the definition of f into the expression for the average change of Y with respect to X

$$\frac{f(X + \Delta X) - f(X)}{\Delta X} = \frac{\left((X + \Delta X)^2 + (X + \Delta X) \right) - (X^2 + X)}{\Delta X}$$

$$= \frac{\cancel{X^2} + 2X \cdot \Delta X + (\Delta X)^2 + \cancel{X} + \Delta X - \cancel{X^2} - \cancel{X}}{\Delta X}$$

$$= \frac{\cancel{\Delta X} \cdot (2X + \Delta X + 1)}{\cancel{\Delta X}}$$

$$= 2X + \Delta X + 1$$

Letting ΔX approach 0, this expression becomes

$$\lim_{\Delta X \to 0} \frac{f(X + \Delta X) - f(X)}{\Delta X} = \lim_{\Delta X \to 0} (2X + \Delta X + 1)$$

$$= 2X + 1$$

So we can say that "the derivative of $X^2 + X$ is $2X + 1$." What we mean is that at every point X_0, the slope of the tangent line to $Y = X^2 + X$ is $2X_0 + 1$.

Exercise 2.5.3 Find the derivative of the function $f(X) = X^3$ as in the above example. (Recall from algebra that $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$.)

Exercise 2.5.4 Calculate the slope of the tangent line to the graph of $Y = X^3$ at $X = 1$.

Higher Order Derivatives

Once we have the idea that the derivative is a function that takes a function $f(X)$ and assigns to it the function $\frac{df}{dX}$, we can ask: what if we applied this function twice? That is, if the derivative $\frac{df}{dX}$ is a function from $\mathbb{R} \to \mathbb{R}$ then does it have a derivative itself? The answer is yes, and the derivative of the derivative is called the *second derivative* of f with respect to X, and is generally written as

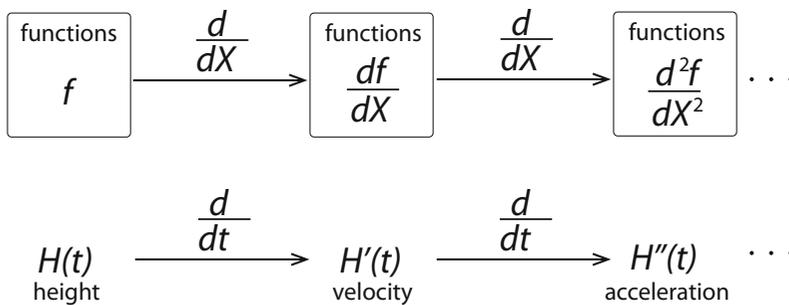$$\text{second derivative of } f \text{ with respect to } X = \frac{d^2 f}{dX^2}$$

Figure 2.19: If the function $\frac{df}{dX}$ is itself differentiable, then we can take its derivative to get a third function, called the second derivative of f. For the falling body, the derivative of height is velocity, and the derivative of velocity is acceleration.

The best-known example of a second derivative is in the motion of an object in space. If $H(t)$ is an object's position at time t, then the first derivative of H with respect to time, $H'(t)$, is called the *velocity* of the object at time t. The derivative of velocity, $H'(t)$, with respect to time is the second derivative of H with respect to time, and is called *acceleration* (Figure 2.19).

Exercise 2.5.5 Find the second derivative of $Y(X) = X^3 + 15X^2 + 3$.

Exercise 2.5.6 The growth of cells in a petri dish slows down over time. Is the second derivative of the function giving the number of cells positive or negative?

Derivatives of Famous Functions

The same method we just used, plugging "$t + \Delta t$" into f, subtracting $f(t)$, dividing by Δt, then letting $\Delta t \to 0$, works generally to find the derivatives of many well-known functions, though in

many cases, special technical tricks have to be used. However, this is a tedious process involving much algebra, so it's useful to know the derivatives of common functions.

Traditional calculus courses take great care in deriving these derivative functions for a large class of functions. Here, we will simply present these rules and functions. For those who are curious, try a quick Internet search. We list the most important here.

The Derivative of a Constant:
For any number c,

$$\frac{d}{dX}(c) = 0$$

Exercise 2.5.7   Why does this make sense?

Power Functions:
For any constant $n \neq 0$,

$$\frac{d}{dX}(X^n) = nX^{n-1}$$

That is, if $f(X) = X^n$, then its derivative is $f'(X) = nX^{n-1}$. This holds even for values of n that are not integers, such as fractions.

A good way to visualize the process is this: to find the derivative of something raised to a power n, first bring the exponent down in front of the expression, and then decrease the exponent by 1.

For example, to find the derivative of X^8, first bring the 8 down in front of the expression, and then decrease the exponent to 7, to end up with $8X^7$. To find the derivative of $X^{\frac{1}{3}}$, first bring the $\frac{1}{3}$ down in front of the expression, and then decrease the exponent to $\frac{1}{3} - 1 = -\frac{2}{3}$, to end up with $\frac{1}{3}X^{-\frac{2}{3}}$.

Exercise 2.5.8   Differentiate:

  a) $f(X) = X^5$ b) $f(X) = X^{-3}$ c) $f(X) = X^{17.2}$

Exercise 2.5.9   The maximum life-span, L, of a mammalian species increases with average body mass B as roughly $L(B) = B^{0.25}$. What is the rate of increase of life-span with body mass?

Exponential Functions:

$$\frac{d}{dX}(e^{kX}) = ke^{kX}$$

Logarithmic Functions:

$$\frac{d}{dX}(\ln X) = \frac{1}{X}$$

Trigonometric Functions:

$$\frac{d}{dX}\Big(\sin(X)\Big) = \cos(X)$$

$$\frac{d}{dX}\Big(\cos(X)\Big) = -\sin(X)$$

Putting Functions Together

We often want to combine simple functions into more complex ones. There are several rules for how to find the derivatives of these complex functions in terms of the derivatives of their components.

Here we present the necessary rules.

The Constant Multiple Rule:

If c is a constant and $f(X)$ is a function of X, and we let

$$h(X) = c \cdot f(X) \quad \text{or simply} \quad h = c \cdot f$$

then

$$\frac{dh}{dX} = \frac{d(c \cdot f)}{dX} = c \cdot \frac{df}{dX}$$

In other words, a constant multiple just passes through the derivative unchanged.

For example,

$$\frac{d}{dX}3X^2 = 3\frac{d}{dX}X^2 = 3 \cdot 2X = 6X$$

Exercise 2.5.10   Differentiate:

  a) $f(X) = 4X^8$ b) $f(X) = 3.5X^{-2}$ c) $f(X) = \pi X^{4.3}$

The Addition Rule:

If $f(X)$ and $g(X)$ are two functions of X, and we let

$$h(X) = f(X) + g(X) \quad \text{or simply} \quad h = f + g$$

then

$$\frac{dh}{dX} = \frac{d(f + g)}{dX} = \frac{df}{dX} + \frac{dg}{dX}$$

In other words, the derivative of the sum of two functions is the sum of their derivatives.

Exercise 2.5.11   A similar rule holds for subtraction. Why?

Exercise 2.5.12   Apply the addition and subtraction rules to calculate the derivative of the function $f(X) = X + X^2 - 2X^3 + 2X^4$.

Exercise 2.5.13 What is the rule for differentiating a function of the form $h(X) = f(X) + c$, where c is a constant? Justify your answer in terms of the rules we already know.

Be careful not to confuse the rule you just developed with the constant multiple rule!

- The derivative of 5 *times* something is 5 times the derivative of the something. In that case, the constant 5 stays in place, unchanged.

- The derivative of something *plus* 5 (or *minus* 5) is just the derivative of the something. In this case, the constant 5 vanishes when you take the derivative.

The Product Rule:
For two functions $f(X)$ and $g(X)$, if we let $h(X)$ be their product,

$$h(X) = f(X) \cdot g(X) \quad \text{or simply} \quad h = f \cdot g$$

then

$$\frac{dh}{dX} = \frac{d(f \cdot g)}{dX} = \frac{df}{dX} \cdot g + f \cdot \frac{dg}{dX}$$

The Quotient Rule:
If $f(X)$ and $g(X)$ are functions of X, and we let

$$h(X) = \frac{f(X)}{g(X)} \quad \text{or simply} \quad h = \frac{f}{g}$$

then

$$\frac{dh}{dX} = \frac{d\left(\frac{f}{g}\right)}{dX} = \frac{\frac{df}{dX} \cdot g - \frac{dg}{dX} \cdot f}{g^2}$$

Exercise 2.5.14 Differentiate the following functions:

a) $f(t) = \sin(t)\cos(t)$

b) $h(X) = \dfrac{X^2}{3X + 5}$

c) $f(X) = \dfrac{4X}{\sqrt{X} + 2}$

d) $g(Y) = (3Y^6)\ln Y$

Often, we have to deal with functions that are embedded in other functions, for example, the function $h(X) = \sqrt{X^2 + 1}$. This is a composite function: there is an inner function $X^2 + 1$ and an outer function $\sqrt{}$, and we first apply the inner function to X and then apply the outer function to the result.

If we call the inner function $g(X)$ and the outer function $f(X)$, then $h(X)$ can be written as $h(X) = f(g(X)) = (f \circ g)(X)$. (See Section 1.2 if you want to review composition of functions.)

The rule for differentiating a composite function is called the *chain rule*.

The Chain Rule:

If $f(X)$ and $g(X)$ are functions of X, and we let

$$h(X) = f\Big(g(X)\Big) \quad \text{or simply} \quad h = f \circ g$$

then

$$\frac{dh}{dX} = \frac{d(f \circ g)}{dX} = \frac{df}{dg} \cdot \frac{dg}{dX}$$

The expression $\dfrac{df}{dg}$ needs clarification. By $\dfrac{df}{dg}$, we mean the derivative of f, treating the whole expression $g(X)$ as if it were a variable. It's equivalent to setting $Y = g(X)$; then $\dfrac{df}{dX} = \dfrac{df}{dY} \cdot \dfrac{dY}{dX}$.

Let's work out an example of the chain rule. We know that an object dropped from a height $H(0)$ will, after t seconds, be at the height

$$\textbf{height equation} \qquad H(t) = H(0) - 16t^2$$

As the altitude of the object decreases, the atmospheric pressure on it will increase by the relationship $P(H) = e^{-\frac{H}{7}}$.

We want a function that will give us the atmospheric pressure as a function of time. To do that, we need to make a composite of these two functions. However, we have a slight problem, which is that the "H" in the falling object equation is in feet, and the "H" in the pressure equation is measured in kilometers. Therefore, we need to be explicit about this. Since 1 kilometer = 3281 feet, we have

$$\textbf{pressure equation in ft} \qquad P(H) = e^{-\frac{1}{3281} \cdot \frac{H}{7}}$$

To find the rate of change of pressure P with respect to time, on a falling object dropped from an initial height of $H(0)$ km, we use the chain rule:

$$\frac{dP}{dt} = \frac{dP}{dH} \cdot \frac{dH}{dt}$$

$$= \frac{d\Big(e^{-\frac{1}{3281} \cdot \frac{H}{7}}\Big)}{dH} \cdot \frac{d\Big(H(0) - 16t^2\Big)}{dt}$$

$$= \left(-\frac{1}{3281} \cdot \frac{1}{7} \cdot e^{-\frac{1}{3281} \cdot \frac{H}{7}} \right) \cdot \left(-32t \right)$$

$$= \left(\frac{1}{3281} \cdot \frac{1}{7} \cdot e^{-\frac{1}{3281} \cdot \frac{H}{7}} \right) \cdot \left(32t \right)$$

$$= \left(\frac{1}{3281} \cdot \frac{1}{7} \cdot e^{-\frac{1}{3281} \cdot \frac{H(0)-16t^2}{7}} \right) \cdot \left(32t \right)$$

If the object is dropped from $H(0) = 10{,}000$ ft, then after $t = 10$ s, the rate of change of pressure is

$$\left. \frac{dP}{dt} \right|_{t=5\,\text{s}} = \left(\frac{1}{3281} \cdot \frac{1}{7} \cdot e^{-\frac{1}{3281} \cdot \frac{10000 - 16 \cdot (10)^2}{7}} \right) \cdot \left(32 \cdot 10 \right) \approx 0.01 \frac{\text{bars}}{\text{s}}$$

Exercise 2.5.15   Write the following expressions of $h(X)$ as a composition of two functions, one outer function $f(Y)$ and one inner function $g(X)$, so that $f(g(X)) = h(X)$. Then, find the derivative of each.

a) $h(X) = (X^3 + 1)^2$ b) $h(X) = \sqrt{X^5}$ c) $h(X) = e^{X^2+1}$

Further Exercises 2.5

1. Differentiate the following functions:
   a) $f(X) = 2.5X$
   b) $g(X) = 8X + 4$
   c) $f(X) = 3X^4 - 6X^2 + 5X + 10$
   d) $\tan X = \dfrac{\sin X}{\cos X}$
   e) $y(X) = e^X \sin X$
   f) $f(t) = 2.5\cos(t + \pi) + 10$ (Functions like this are often used to model seasonally varying parameters.)
   g) $w(t) = (t^6 + 26t^4 - t^3 + 179)^{73}$
   h) $f(X) = e^{\sqrt{X}}$
   i) $f(t) = 3t^7 + 4t^5 - \sqrt{t}$
   j) $f(X) = \dfrac{1}{1 + X}$ (You will see this function and the two that follow in more advanced models later in this book.)
   k) $f(X) = \dfrac{X}{1 + X}$
   l) $f(X) = \dfrac{X^2}{1 + X^2}$

2. What is the slope of the tangent line to the graph of $Y = e^{X^2}$ at $X = 1$?

3. Find the linear approximation to the function
$$f(X) = (X + 2)^3 - e^{3X}$$
   at $X_0 = 0$.
   a) First, give your answer in the form $\Delta f \approx m \cdot \Delta X$.
   b) Expand your answer from part (a) by rewriting Δf as $f(X) - f(X_0)$ and ΔX as $X - X_0$, and solving for $f(X)$. (Note: What is $f(X_0)$?)
   c) What is $f(0.2)$, approximately?
   d) Use your answer from part (b) to write down the equation for the tangent line to $f(X)$ at $X_0 = 0$.

4. In mammals, resting metabolic rate M is related to body mass B as approximately

$$M = 0.8B^{3/4}$$

a) Find the linear approximation to this function for a body mass of 100 grams.

b) An animal species that currently averages 100 grams in mass evolves to have an average mass of 110 grams. Use the linear approximation to estimate how much its metabolic rate would change.

5. The number of species, S, living on an island or habitat fragment of area A can be modeled as $S = cA^z$, where $z > 0$. We can measure area in units that make $c = 1$, simplifying the equation into $S = A^z$. A common value for z is approximately 0.45. Find $\frac{dS}{dA}$, and explain the meaning of this function and its significance for biodiversity conservation.

6. The optimal cruising speed V for a bird in flight is given by the allometric equation

$$V = 12M^{1/6}$$

where M is the mass of the bird in kilograms, and V is in meters per second. The average female bald eagle weighs around 5.6 kg. Find the linear approximation to the equation above at $M = 5.6$. If a female American golden eagle is about 0.7 kg lighter than a bald eagle, how much faster or slower would you expect its optimal cruising speed to be? (Note: You can check that your answer is in the ballpark by just plugging numbers into the equation above, but the point of this problem is to use a linear approximation to arrive at your answer.)

7. Find the instantaneous rate of change of the function $f(X) = 3\sin X + \ln X$ at $X = \pi$.

8. Find the tangent to the function $f(X) = 5X^4 + 3X^2 - 9$ at $X = 2$.

9. In this section, we saw that it makes sense to think of differentiation as a function. Is this function linear? Justify your answer.

2.6 Integration

So far we have focused on differentiation: if we know $f(X)$, can we find $f'(X)$ (also called $\frac{df}{dX}$)?

As we saw in the previous section, the answer to this question is frequently yes. Most of the famous formulas, like X^2, X^n, e^X, or $sin(X)$, can be symbolically differentiated, and the product rule, quotient rule, and chain rule give us a way to get derivatives of compound expressions made up of these functions.

But what about the reverse process? If we are given $f'(X)$, can we recover $f(X)$? This reverse process is called *integration*. Of course, when we are given a function that is obviously the derivative of another function, this process is easy. For example, if we are given $f'(t) = 2t$ then we know that any function of the form $t^2 + c$, where c is a constant, has as its derivative $2t$. We say that $f(t) = t^2 + c$ is the *antiderivative* of $f'(t) = 2t$.

**Exercise 2.6.1** Why is the $+c$ necessary? *Find $\frac{d}{dt}(t^2 + 5)$ and $\frac{d}{dt}(t^2 - 1)$.*

Exercise 2.6.2 Find the antiderivative of $f'(X) = 3X^2$.

But these are very special cases, and in most cases, given the functional form of $f'(t)$, it is impossible to state the antiderivative $f(t)$, and we have to rely on approximations. So, just as in differential equations in Chapter 1, symbolic differentiation is usually easy, and symbolic integration is usually hard.
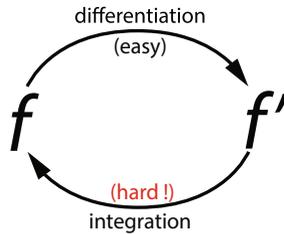

Before we go into how to get f back from f', let's ask why: *why would we want to do this?* One important class of cases involves a given function $f'(t)$ as a rate of something, and we want to figure out the total amount of the something deposited by that rate.

For example, suppose we are given the speed $V(t)$ of a car from a recording of the speedometer. So we have $V(t)$. But $V(t)$ is just $X'(t)$, where X is the car's position at time t, assuming it starts at position 0 (which makes the constant c in the antiderivative to be 0). Can we recover $X(t)$ from $V(t)$?
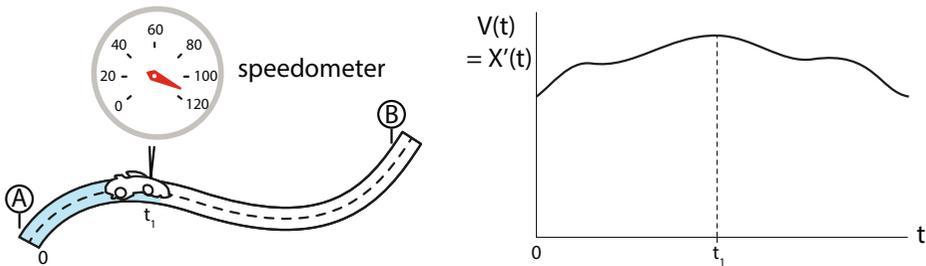
For another example, we might know $D'(t)$, the rate of drug delivery to a patient. This is the readout of the flow meter attached to the intravenous drip. But what we want to know is not $D'(t)$, but rather $D(t)$, the cumulative amount of drug that was delivered to the patient up to time t.
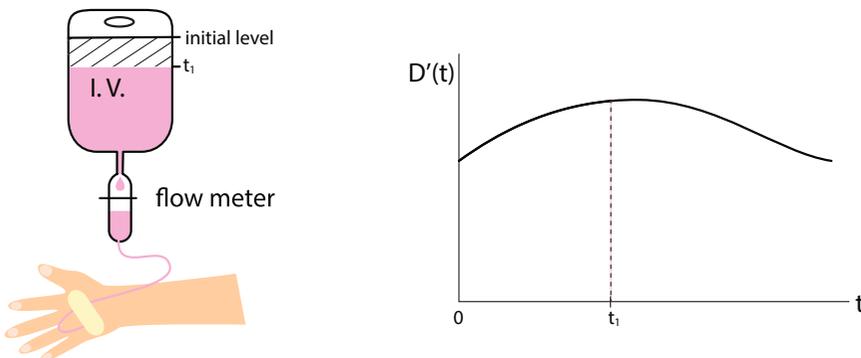
Euler and Riemann: Adding Up Little Rectangles

Let's look at the case of recovering distance traveled, $X(t)$, given the speed $V(t)$. Suppose we are given $V(t)$ as a function, $V(t) = 3t^2$. We know from the power rule that if $X(t) = t^3 + c$, then $X'(t) = 3t^2$. So we can say immediately that in t seconds, the car has traveled a total of $t^3 + c$ miles. Since at $t = 0$, it was at c miles, the distance it has covered is t^3 miles. But what about the case in which X' is not obviously the derivative of some function?

One way to do this is to use a version of Euler's method. Suppose $V(t) = X'(t)$ is the velocity of the car. Let's write down the equation for Euler's method:

$$\text{new } X = \text{old } X + X'(\text{old } X) \cdot \Delta t$$

If we make the table to calculate Euler's method, it looks like

t	old X	X'(old X)	X'(old X) $\cdot \Delta t$	new $X = $ old $X + X'$(old X) $\cdot \Delta t$
0	X_0	$X'(X_0)$	$X'(X_0) \cdot \Delta t$	$X_{\Delta t} = X_0 + X'(X_0) \cdot \Delta t$
Δt	$X_{\Delta t}$	$X'(X_{\Delta t})$	$X'(X_{\Delta t}) \cdot \Delta t$	$X_{2\Delta t} = X_0 + X'(X_0) \cdot \Delta t + X'(X_{\Delta t}) \cdot \Delta t$
$2\Delta t$	$X_{2\Delta t}$	$X'(X_{2\Delta t})$	$X'(X_{2\Delta t}) \cdot \Delta t$	$X_{3\Delta t} = X_0 + X'(X_0) \cdot \Delta t + X'(X_{\Delta t}) \cdot \Delta t + X'(X_{2\Delta t}) \cdot \Delta t$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Exercise 2.6.3 In this table, what are the red entries in the last column? The black entries?

In the differential equation case, the critical step in Euler's method is the calculation, at each Δt, of the new value of X'. Since the differential equation gives us X' as a function of X, $X' = f(X)$, we used that function f to calculate the new X'.

Here we don't have to do that, because we are given $X'(t)$ explicitly as a function of t, and so we don't have to recalculate it from the previous X value.

This makes the procedure of approximating $X(t)$ much easier: we can input $X'(t)$ directly from the formula and do not have to recalculate it using new values of X for every Δt. Euler's method then takes a particularly simple form:

t	$V(t)$	$V(t) \cdot \Delta t$	$X(t)$
0	$V(0)$	$V(0) \cdot \Delta t$	$X(0) = X_0$
Δt	$V(\Delta t)$	$V(\Delta t) \cdot \Delta t$	$X_0 + V(0) \cdot \Delta t$
$2\Delta t$	$V(2\Delta t)$	$V(2\Delta t) \cdot \Delta t$	$X_0 + V(0) \cdot \Delta t + V(\Delta t) \cdot \Delta t$
$3\Delta t$	$V(3\Delta t)$	$V(3\Delta t) \cdot \Delta t$	$X_0 + V(0) \cdot \Delta t + V(\Delta t) \cdot \Delta t + V(2\Delta t) \cdot \Delta t$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

If we summarize this simplified table, we see that

$$X(t) = \underbrace{X(0)}_{\substack{\text{location at} \\ \text{time 0}}} + \underbrace{V(0 \cdot \Delta t) \cdot \Delta t}_{\substack{\text{distance covered in first} \\ \Delta t}} + \underbrace{V(1 \cdot \Delta t) \cdot \Delta t}_{\substack{\text{distance covered in second} \\ \Delta t}} + \underbrace{V(2 \cdot \Delta t) \cdot \Delta t}_{\substack{\text{distance covered in third} \\ \Delta t}} + \cdots$$

Notice what this is saying: for each time interval, we are adding a little increment. If we represent time as $t = 0, \Delta t, 2\Delta t, 3\Delta t, \cdots, k\Delta t, \cdots$, then the little increments are each

$$V(k \cdot \Delta t) \cdot \Delta t$$

as k ranges from 0 to n, where $n \cdot \Delta t$ is the stopping time. (In other words, n is the number of Δt's necessary to get to the stopping time.) This has a geometric interpretation that we will discuss a little later.

Each little increment $V(k \cdot \Delta t) \cdot \Delta t$ represents the distance the car travels in that Δt. We get this by assuming that the velocity V is constant over the short interval Δt, which enables us to use the formula

$$\text{distance} = \text{velocity} \times \text{time}$$

to calculate the distance traveled.

This sum of little increments is called a *Riemann sum.*

We can summarize the Riemann sum as

$$X(t) \approx X(0) + \overset{k=n}{\underset{k=0}{\mathbf{Sum}}}\ V(k \cdot \Delta t) \cdot \Delta t$$

where the expression

$$\overset{k=n}{\underset{k=0}{\mathbf{Sum}}}\ f(k)$$

means "the sum of all terms $f(k)$, where k takes on every integer value from 0 to n."

Finally, we use the Greek uppercase sigma (Σ) to stand for "sum," and we have

$$X(t) \approx X(0) + \sum_{k=0}^{k=n} V(k \cdot \Delta t) \cdot \Delta t$$

Exercise 2.6.4   Compute:

a) $\displaystyle\sum_{k=0}^{k=3} 2k$

b) $\displaystyle\sum_{k=0}^{k=4} k^3$

c) $\displaystyle\sum_{k=0}^{k=3} 6k + 2$

Procedure for the Riemann Sum

(1) Break down the total elapsed time into many small Δt's.
(2) Assume that $V(t)$ is constant over each small interval Δt.
(3) Use the equation "distance $=$ velocity $\times$ time" to calculate the distance traveled in that Δt.
(4) Add up the little distances.

Exercise 2.6.5   Find the Riemann sum for $f(X) = X^2 + 5$ between $X = 0$ and $X = 2$ using a step size of 0.5.

In this way, we can approximate the function $X(t)$ by a finite sum of little increments that depend on Δt. This approximation gets better and better as Δt gets smaller. So we take the final step of letting Δt approach 0. We can then replace the "approximately equals" sign by "exactly equals":

$$X(t) = X(0) + \lim_{\Delta t \to 0} \sum_{k=0}^{k=n} V(k \cdot \Delta t) \cdot \Delta t$$

We need a new symbol for this infinite limit of the sum Σ as Δt approaches 0. The standard symbol for this is a big script S shape called the "integral sign"

$$X(t) = X(0) + \int_0^t X' \cdot dt \qquad (2.2)$$

The expression $\int_0^t X' \cdot dt$ is called *the definite integral of $X'(t)$ from 0 to t*. Equation (2.2) is called the *f*undamental theorem of calculus.

To get some intuition for the fundamental theorem of calculus, let's rewrite equation (2.2) as

$$X(t) - X(0) = \int_0^t X' \cdot dt \qquad (2.3)$$

Now, let's think about the IV drip example. To find out how much fluid the patient has received, you could follow the procedure described here: add up the little rectangles. Or you could record how much fluid is in the IV bag at the start of treatment, how much is left at the end, and subtract. The fundamental theorem of calculus tells us that the results will be the same.

An Example: A Speeding Car

Let's use a Riemann sum approximation to find an integral that is not explicitly known. Suppose the speed of a car, in feet per second, is

$$V(t) = 2\sqrt{1000 - t^3}$$

This corresponds to a car starting at around 45 miles per hour ($\approx$ 65 feet per second) and slowing down to a complete stop over an interval of 10 seconds.
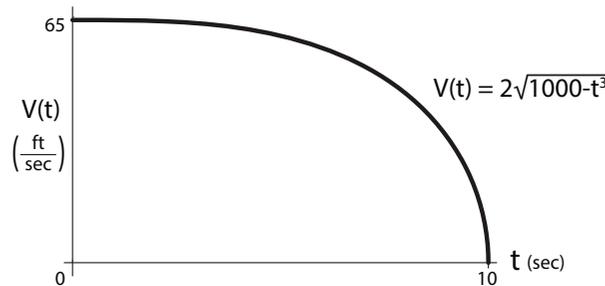
If we want to figure out how far the car has traveled at a particular time t, we need to find $X(t)$.

We mentioned earlier that most of the time, it is impossible to find an actual equation for the solution to an integration problem. This is one of those cases. There is no closed-form expression for the antiderivative of $2\sqrt{1000 - t^3}$.[8]

So how do we find the distance we've traveled in the car at some time t? We can use the Riemann sum method to approximate it. So let's suppose $X(0) = 0$, and let's use a step size of $\Delta t = 0.1$ to approximate X at time $t = 10$ seconds, the moment at which the speed is 0 (i.e., when the car comes to a complete stop). We'll be able to use this to find the distance required to stop the car, which can be an important safety consideration.

---

[8] To state this precisely, the antiderivative of $2\sqrt{1000 - t^3}$ is not an *elementary function*. Elementary functions are those made up of a finite number of power functions, trig functions, exponential functions, and their inverses, combined using addition, subtraction, multiplication, division, and composition—in short, anything for which you can write down a simple formula

The Riemann sum is

t	$V(t)$	$V(t) \cdot \Delta t$	$X(t) = \sum V(t) \cdot \Delta t$
0.	63.25	6.325	6.325
0.1	63.25	6.325	12.65
0.2	63.25	6.325	18.97
0.3	63.24	6.324	25.30
0.4	63.24	6.324	31.62
$\vdots$	$\vdots$	$\vdots$	$\vdots$
9.9	10.90	1.090	535.0

Therefore, the total distance traveled in 10 seconds is 535.0 feet.

Exercise 2.6.6 Find the distance traveled by the car in 8 seconds if $V(t) = 4\sqrt{500 - t^3}$. Use a step size of 0.5. You may want to use SageMath or a spreadsheet to help with the calculation.

The Geometry of the Riemann Sum

There is a geometric visualization of the Riemann sum that gives a lot of significant insight into the concept. When we calculated the Riemann sum, we followed the process below:

(1) We broke down the total elapsed time into many Δt's.
(2) We assumed that $V(t)$ was constant over each interval Δt.
(3) Then we used the equation "distance = velocity × time" to calculate the distance traveled in that Δt.
(4) Then we added up the little distances.

This process can be viewed geometrically (Figure 2.20).
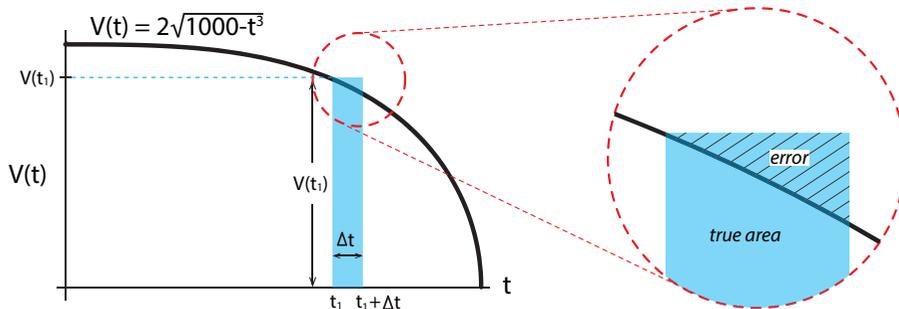

Figure 2.20: The blue rectangle is an approximation to the area under $V(t)$ in a small region of width Δt. The inset illustrates the error in this approximation.

During the brief interval from $t = t_1$ to $t = t_1 + \Delta t$, how far did the car move? We approximated its speed over that short interval using the speed at the beginning of the interval, which is $V(t_1)$, and assumed that the car was constant in velocity over that interval. Based on this assumption, we then used

$$\text{distance} = \text{velocity} \times \text{time}$$

to calculate the distance that the car covered during the time interval t_1 to $t_1 + \Delta t$ as $V(t_1) \cdot \Delta t$.

Now consider the blue rectangle. It has a base that is Δt wide, and its height is $V(t_1)$. So the area of the blue rectangle is given by

$$\text{area} = \text{height} \times \text{width}$$

or in this case, area $= V(t_1) \cdot \Delta t$. But this is the same calculation that we just did for the distance traveled:

$$
\begin{array}{ccccc}
\text{distance} & = & \text{velocity} & \times & \text{time} \\
\updownarrow & & \updownarrow & & \updownarrow \\
V(t_1) \cdot \Delta t & = & V(t_1) & \times & \Delta t \\
\updownarrow & & \updownarrow & & \updownarrow \\
\text{area} & = & \text{height} & \times & \text{width}
\end{array}
$$

In other words, the area of the little blue rectangle is the distance traveled during that Δt.

Then what is the total distance traveled by the car? It must be approximately the sum of the little distances. But then the total distance that the car taveled from 0 to t is approximately equal to the sum of these little rectangles, or $\sum V(t) \cdot \Delta t$ (Figure 2.21).
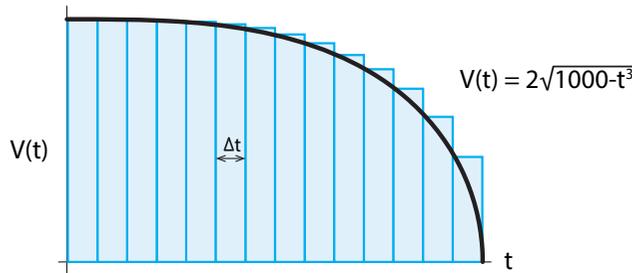

Figure 2.21: We can estimate the area under the curve $V(t)$ by adding up the areas of all the blue rectangles of height $V(t)$ and width Δt.

The sum of the little rectangles is approximately the area under the curve, and the sum of the little rectangles is approximately equal to the total distance the car has driven. In the limit as Δt approaches 0, the two are the same: *distance is the area under the curve of velocity as a function of time.*

Thus, the distance the car traveled from point A at $t = 0$ to $t = t_1$ is the area under the velocity curve between $t = 0$ and $t = t_1$ (Figure 2.22).
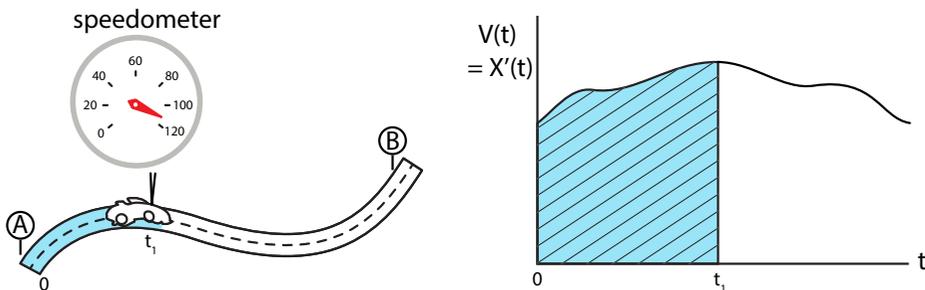

Figure 2.22: Left: The car's speedometer tells us the car's velocity V at any time t. Right: We can graph this Velocity data as $V(t)$. The shaded blue area is equal to the distance the car has traveled from $t = 0$ to $t = t_1$.

And similarly, in the drug drip problem, the cumulative amount of drug delivered from $t = 0$ to $t = t_1$ is equal to the area under the flow rate curve $D'(t)$ (Figure 2.23).
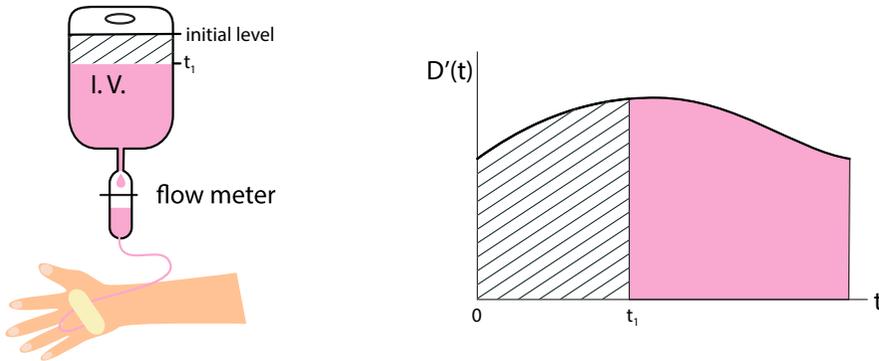

Figure 2.23: Left: The intravenous drug delivery system has a flow meter that tells us, at any time, how fast the drug is flowing into the patient, that is, $D'(t)$, where $D(t)$ is the total amount of drug delivered. Right: The shaded area shows the total amount of drug that has been delivered by time t_1, while the pink region shows the amount of drug left in the bag.

In fact, we can generalize the car velocity $V(t)$ to be any function $f(t)$. In general, let f be any reasonably well behaved function of t.

Let's define a new function of t, called $F(t)$,

$$F(t) = \text{the area under } f(t) \text{ from 0 to } t$$

So, for example, the hatched area is the value of $F(t + \Delta t)$, and the pink area is the value of $F(t)$ (Figure 2.24). Now consider the green rectangle. It has height $f(t)$ and width Δt, so its area is $f(t) \cdot \Delta t$. But now let's look at the green rectangle from the point of view of the area-under-f function F. It is clear that the green rectangle is approximately equal to the area under f from 0 to $t + \Delta t$ (the hatched area) minus the area under f from 0 to t (the pink area). This approximation gets better and better as $\Delta t \to 0$.

$$\underbrace{F(t + \Delta t)}_{\text{hatched area}} - \underbrace{F(t)}_{\text{pink area}} \approx \underbrace{f(t) \cdot \Delta t}_{\text{green rectangle}}$$
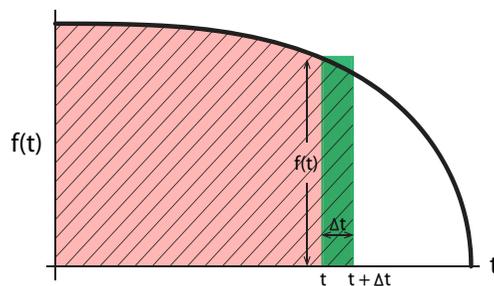

Figure 2.24: The green rectangle is approximately equal to the area under $f(t)$ from t to $t + \Delta t$, which is the shaded area, minus the area under $f(t)$ from 0 to t, which is the pink area.

If we divide both sides by Δt,

$$\frac{F(t + \Delta t) - F(t)}{\Delta t} \approx f(t)$$

then as $\Delta t \rightarrow 0$, the left-hand side is just the definition of F', so

$$F' = f$$

We can now say that the area under f from a to b is just the area under f from 0 to b minus the area under f from 0 to a.

$$F(b) - F(a) = \int_a^b f(t) \cdot dt$$

This is another version of the fundamental theorem of calculus.

Exercise 2.6.7 According to CDC data, the average American six-year-old girl weighs 42.5 pounds, and the average ten-year-old girl weighs 75 pounds. What is the area under the growth (rate of change of weight) function between $t = 6$ and $t = 10$?

Example: A Drug Drip

We can illustrate the general principle of integration by Riemann sums using the example of an IV drip delivering drug to a patient.
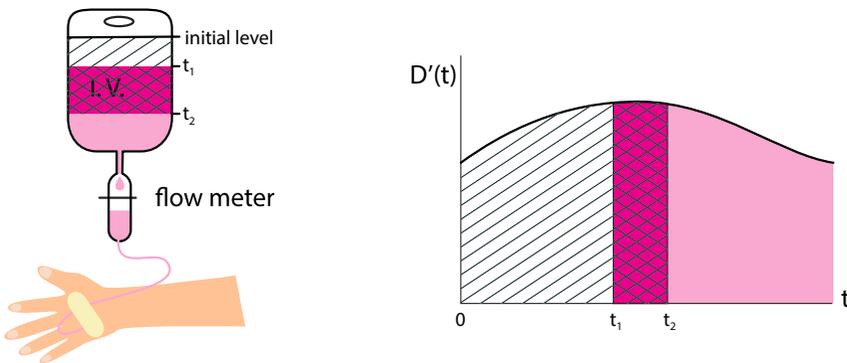

Figure 2.25: The cross-hatched area is the amount of drug that was delivered from t_1 to t_2. It is equal to the total amount delivered from $t = 0$ to $t = t_2$ minus the amount delivered from $t = 0$ to $t = t_1$ (white shaded area).

The flow meter on the drip gives us $D'(t)$, the flow rate as a function of time. But what we need to know is how much drug was delivered to the patient between t_1 and t_2 (Figure 2.25). Consulting the figure, we see that this is the cross-hatched part of the area under the curve $D'(t)$.

Then the cross-hatched area can be found by realizing that

| cross-hatched area $=$	area under $D'(t)$ from 0 to t_2	$-$	area under $D'(t)$ from 0 to t_1
$\updownarrow$	$\updownarrow$		$\updownarrow$
$\int_{t_1}^{t_2} D'(t) \cdot dt \quad =$	$\int_0^{t_2} D'(t) \cdot dt$	$-$	$\int_0^{t_1} D'(t) \cdot dt$
$\updownarrow$	$\updownarrow$		$\updownarrow$
$\int_{t_1}^{t_2} D'(t) \cdot dt \quad =$	$D(t_2)$	$-$	$D(t_1)$

This is an application of the fundamental theorem of calculus.

But it is important to remember that in the typical case, we will be given $D'(t)$ as *data*, the readout of the flow meter, that has been stored as a digital record. We have no idea what mathematical function of time that is, and we can be pretty sure that it isn't some simple mathematical function that happens to have a handy antiderivative function.

Therefore, the expression $\int_{t_1}^{t_2} D'(t) \cdot dt$ is not going to be very helpful in these real-world cases, and the only technique open to us is to add up the little rectangles.

Further Exercises 2.6

1. What is the area under the graph of the function $f(x) = \cos X - X$ between $X = 0$ and $X = \pi$?

2. a) Approximate the area under the graph of $f(X) = X^2$ between $X = 2$ and $X = 4$ using a ΔX of 0.5.

    b) Find the exact area under $f(X) = X^2$ between $X = 2$ and $X = 4$.

3. a) Approximate the area under the graph of $f(X) = x^4 + 1$ between $X = 1$ and $X = 3$ using $\Delta t = 0.5$.

    b) Find the exact area under the graph of $f(X) = X^4 + 1$ between $X = 1$ and $X = 3$.

    c) How could you make the answer to part (a) closer to the exact value you found in part (b)?

4. You are studying a plant population whose age distribution is given by $X(a) = \frac{10}{9}\frac{1}{a^2}$, where a is age in years. The smallest individuals you can reliably identify are one year old, so the age distribution starts at 1, and the plant can live no longer than ten years. What fraction of the population is between 3 and 6 years old?

5. A building has solar panels on the roof. The graph below shows the amount of power generated by the solar panels. Assume you have the data used to generate the graph, sampled so frequently that it may be regarded as continuous. Describe how you would compute the total amount of electricity generated between 9 a.m. and 11 a.m.
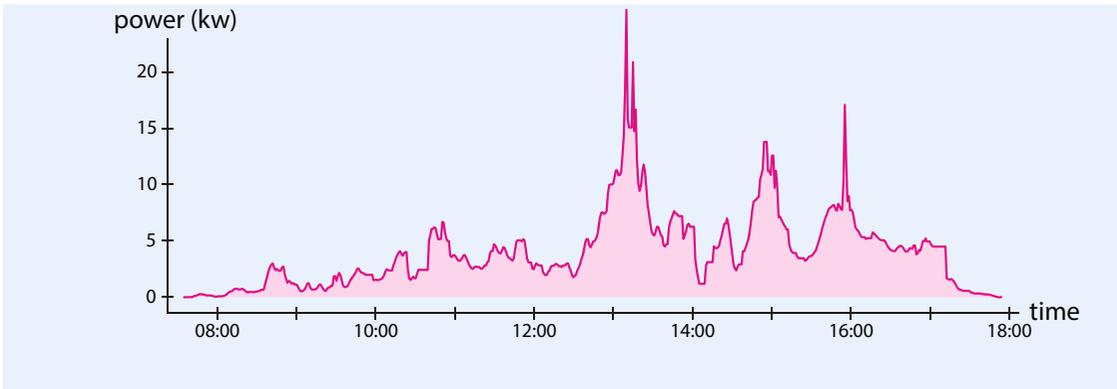
2.7 Explicit Solutions to Differential Equations

In the very rare case in which an antiderivative function can be found, we can use the antiderivative function to create explicit solutions to some simple differential equations.

In Chapter 1, we said that for every well-behaved vector field, the integral curve exists. This is the red curve, and its existence is guaranteed by the fundamental theorem on the existence and uniqueness of solutions to ordinary differential equations.

But we also said that while the red curve is known to exist, the *equation* for the red curve is generally unknown and unknowable, in the sense that most differential equations do not have solutions in terms of elementary functions. We will now deal with one of the few cases in which the equation for the red curve *is* known.

This is called an *explicit solution* to the differential equation $X' = f(X)$, and we can actually write out the function $X(t)$, and then show that

$$\frac{d}{dt}X(t) = f\left(X(t)\right)$$

We will study two of the simplest differential equations, $X' = kX$ and $X' = -kX$. These equations have explicit solutions.

Suppose an individual in a population of size X gives birth, on average, to b offspring per unit time (i.e., the population has per capita birth rate b). The population has per capita death rate d, per capita immigration rate i, and per capita emigration rate e, all assumed to be constants.

In this case, the per capita population growth rate $r = b + i - d - e$ is constant, and we can write the differential equation

$$X' = rX \tag{2.4}$$

What behavior follows from this differential equation? Of course, we can integrate it numerically, using SageMath. But in this case, there is an explicit solution to the differential equation. We saw earlier that

$$\frac{d}{dt}e^{kt} = ke^{kt}$$

so if $X(t) = e^{kt}$, then

$$X'(t) = \frac{d}{dt}X(t) = \frac{d}{dt}e^{kt} = ke^{kt} = kX$$

Therefore, the function $X(t) = e^{kt}$ *solves* the differential equation $X' = kX$.

If we plot both the numerical integration of $X' = kX$ and the explicit solution $X(t) = e^{kt}$, we see that they agree (Figure 2.26). Indeed, e^{kt} is the equation for the true red curve that solves the differential equation. The discrete points are the numerical approximation (blue line) to the red curve.
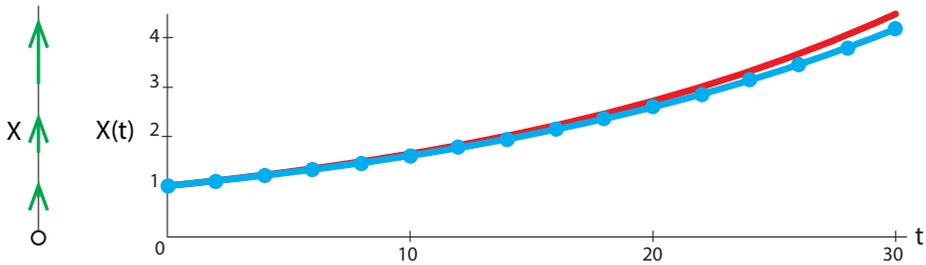

Figure 2.26: Left: state space and vector field for $X' = 0.05X$. Right: time series plots showing $e^{0.05t}$ (red) and Euler integration of $X' = 0.05X$ (blue). Initial condition $X(0) = 1$.

Exercise 2.7.1 Use SageMath to plot e^{kt} for three different values of k, say $k = 0.1$, $k = 1$, and $k = 5$.

The type of growth that corresponds to these equations is called *exponential growth*. How fast is it?

The Rate of Exponential Growth

In the year 1256, the Arab scholar ibn Khallikan wrote down the story of the inventor of chess and the Indian king who wished to reward him for his invention. The inventor asked that the king place one grain of wheat (or in other versions of the legend, rice) on the first square of the chessboard, two on the second, four on the third, and so on, doubling the number of grains with each succeeding square, until the 64 squares were filled. The king thought this a very meager reward, but the inventor insisted. To the king's shock, it turned out that there was no way he could give the inventor that much grain, even if he bankrupted the kingdom.

Let's plot $X(t) = $ # of grains of rice at time t. If we plot the number of grains, starting at $t = 0$ with 1 grain on the first square, then the resulting first few iterations look like Figure 2.27. The red points lie exactly on an exponential curve (black), namely
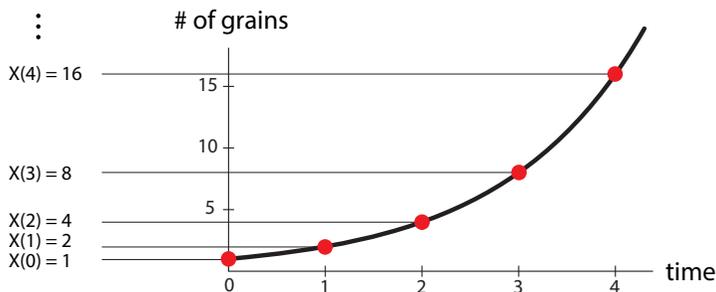

Figure 2.27: The red dots are calculations of $X(t) = 2^t$ for $t = 0, 1, 2, 3$ and 4. The black curve is the continuous function $X(t) = 2^t$, which exhibits exponential growth.

$$X(t) = e^{(\ln 2) \cdot t}$$

because $e^{(\ln 2) \cdot t} = 2^t$.

You can think of the smooth curve $e^{(\ln 2) \cdot t}$ as representing a process in which growth occurs smoothly all the time, whereas the points represent a process in which growth happens and/or is measured only at the time points $t = 0, 1, 2, 3, \ldots$ This latter process is called a discrete-time process and it will be studied in Chapters 5 and 6.

Exercise 2.7.2   How many grains of wheat would there be on the last square of the chessboard?

Exercise 2.7.3   Your employer offers you a choice: be paid \$1 million for thirty days of work or receive \$0.01 on the first day and double your earnings each day. Which do you pick and why?

Exercise 2.7.4   Use SageMath's built-in differential equation solver or your own implementation of Euler's method to simulate equation (2.4) on page 109 for at least three different values of r.

Mind-Blowing Fact of the Day

A sheet of paper is about 0.1 mm thick. If you folded such a sheet of paper in half 50 times, the resulting stack would reach 3/4 of the way to the sun! It would take light 6.3 minutes to travel this distance.

The differential equation $X' = rX$ has the solution $X = X_0 \cdot e^{rt}$.

Exponential Decay

The differential equation $X' = -kX$ models a process in which a constant fraction k of X is removed or dies at any given time, for example, a liquid that evaporates at a constant rate k, or a chemical species that decays into another at a constant rate

$$X \xrightarrow{k} A$$
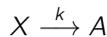
In population dynamics, in a population whose the per capita death rate d exceeded the per capita birth rate b, growth would be governed by $X' = -kX$ where $k = d - b$.

What is the behavior predicted by this model? The vector field is always negative, and the arrows get smaller and smaller as we get closer to zero (Figure 2.28 left).

The differential equation $X' = -kX$ has an explicit solution:

$$X(t) = X_0 \cdot e^{-kt}$$

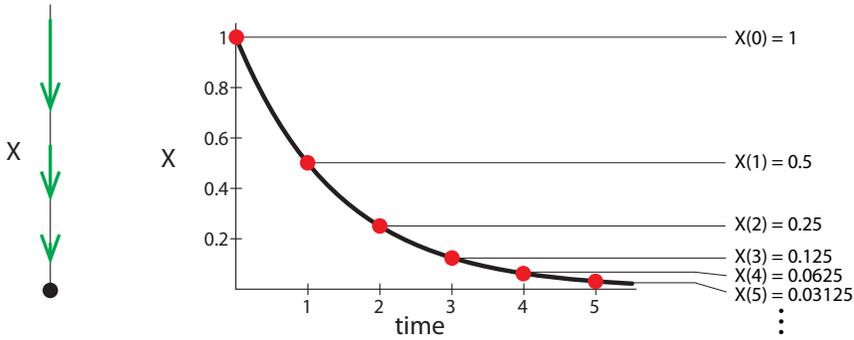**Exercise 2.7.5**   Verify that $X(t) = X_0 \cdot e^{-kt}$ solves the differential equation $X' = -kX$.

Figure 2.28: Vector field and time series for $X' = (\ln 0.5) \cdot X$

The graph of $X(t)$ depits what is called *exponential decay* (Figure 2.28 right). In this case, the differential equation is $X' = (\ln 0.5) \cdot X$, and the explicit solution is $X(t) = e^{(\ln 0.5) \cdot t}$. Just as in exponential growth, we can imagine exponential decay it as a continuous process governed by a differential equation, or as a discrete process in which, say, half of what is left is removed every night at midnight.

When Models Break Down

Over short periods of time, populations can be modeled as growing or decaying exponentially. However, if a population truly underwent exponential growth, it would never stop growing. An exponentially growing population of aardvarks, elephants, or naked mole rats would reach the mass of the Earth, then the solar system and, eventually, the universe. Clearly, something must prevent real populations from growing exponentially, and if we want a population growth model to be useful over more than just the short term, the exponential model must be modified to incorporate processes that stop the population from growing.

Similarly, exponential decay can't be pushed too far. If every night, your roommate removes half of the chocolate left in the fridge, then after a month of this, the fraction of chocolate left would be $\frac{1}{2^{30}} \approx \frac{1}{10^{10}}$, which is smaller than a molecule of chocolate.

The point here is not to bash exponential growth or decay as a model. It fails in a particularly spectacular way, but all models fail at some point. A perfect model would be as complex as the system being modeled and therefore useless. Rather, the lesson is to always do a sanity check when working with models. When using mathematical models in biology, keep an open mind— some strange phenomena first discovered in models have been found in the real world—but use your biological knowledge to judge when a model's predictions are sufficiently wrong to require the model to be modified in order to be useful for the task at hand. There is no universal recipe for doing this, which is why modeling is often called an art.

Exercise 2.7.6   Name two biological factors or processes that might stop a population from growing.

Exercise 2.7.7   A population of 20 million bacteria is growing continuously at a rate of 5% an hour. How many bacteria will there be in 24 hours?

Exercise 2.7.8 A pollutant breaks down at the rate of 2% a year. What fraction of the current amount will be present in 20 years? *Let $X(0) = 1$.*

Further Exercises 2.7

1. The population of an endangered species is declining at a rate of 2.5% a year. If there are currently 4000 individuals of this species, how many will there be in 20 years?

2. If money in a bank account earns an interest rate of 1.5%, compounded continuously, and the initial balance is $1000, how much money will be in the account in ten years?

3. Radioactive iodine, used to treat thyroid cancer, has a half-life of eight days. Find its decay constant, r. *(Hint: You'll need to use natural logarithms.)*

4. Mary is going to have an outdoor party in 10 days. She wants to have her backyard pond covered in water lilies before the party, so she goes to the nursery to buy some water lilies. Mary gives the clerk the dimensions of her pond, and the clerk, knowing the growth rate of the water lilies that he stocks, calculates that if she purchases a single water lily, it will produce a population of ten thousand lilies that will completely cover the surface of the pond in 20 days. Mary reasons that if she buys two water lilies instead of one, she can meet her goal of having the pond surface covered in 10 days. Is there anything wrong with Mary's reasoning? How many water lilies will Mary need to buy to meet her goal?

5. You have $10,000 and can put it either in an account bearing 3.9% interest compounded monthly or one bearing 4% interest compounded annually. If the money will be in the account for five years, which one should you choose?

6. The economic activity of a country is often quantified as the gross domestic product (GDP), which is the sum of private and government consumption, investments, and net exports (the value of exports minus the value of imports). For a developed country such as the United States, economists might see a GDP growth rate of 3% a year as reasonable. However, production and consumption create some pollution. By how much would pollution per dollar of GDP have to decline for pollution levels 50 years from now to be the same as current levels, assuming a 3% annual growth rate of GDP? In 75 years? In 100 years? *(Hint: Let the current pollution level be 1 and find out what the future pollution level would be).*

7. If you want to approximate the time it takes an exponentially growing quantity to double, you can divide 70 by the percentage growth rate. For example, a population growing at 2% a year has a doubling time of about 35 years. Find an exact equation for doubling time and explain why the rule of 70 works as an approximation. *(Hint: Start by trying a few concrete examples.)*