Chapter 2: An Introduction to Calculus

Matthew Eichhorn

Contents

1 Introduction

At the end of Chapter 1, we discussed **Euler's Method**, a numerical method for approximating the path of a trajectory via a sequence of small steps. We saw that our choice of step size, Δt, affected the quality of our approximation; a smaller Δt resulted in an approximate trajectory that more closely aligned with the actual curve. This brings up the question:

> How small of a Δt is small enough?

To state this more exactly, is there some sufficiently small step size that guarantees that Euler's Method will produce a decent approximation? Again, the answer to this question depends on our notion of "decent", but, as is shown by the following example, we can in one sense answer this question "no."

We return our attention to the mass-spring system without friction, also known as the **simple harmonic oscillator**. Recall that in this system, we are concerned with the position (X) and velocity (V) of a cart attached to the end of a spring. Setting all of the physical parameters of the system to 1, we obtain change equations

$$\begin{aligned} X' &= V \\ V' &= -X. \end{aligned}$$

We saw that trajectories for this system appear as concentric circles in the (X, V)-plane. From a physical standpoint, these periodic trajectories make sense. Without the action of an external force, such as friction, we would expect the cart to oscillate back and forth forever. Moreover, overlaying one of these trajectories (in our case, the trajectory through $(1, 0)$) on the vector field, we see that the direction of travel along the trajectory agrees with the change vectors in the vector field.

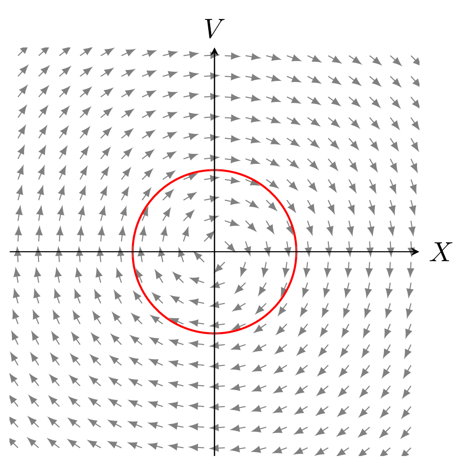


Figure 1: Trajectory of Mass-Spring System with initial condition $(X, V) = (1, 0)$

Now, we consider the approximate trajectories found by Euler's method. These are given for 3 different step sizes in the figure below.

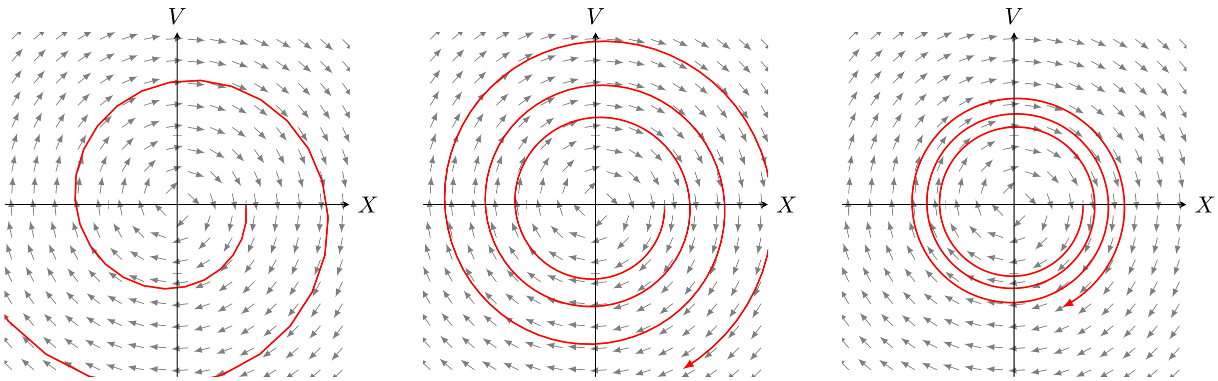


Figure 2: Approximate Trajectories from Euler's Method with $\Delta t = 0.25$ (left), 0.1 (middle), and 0.05 (right).

All three of these approximate trajectories is an outward spiral. Moreover, *any* positive Δt, no matter how small, will result in an outward spiral [1]. Not only is any trajectory predicted by Euler's Method qualitatively different from the true trajectory (they do not have the same shape), but it is also qualitatively different. These approximate trajectories predict a gradual increase in speed and distance traveled by the spring, defying physical reality.

It is here that we seem to have reached a conundrum. In the last chapter, we learned that choosing a smaller step size always results in a more accurate approximation; however, we have just shown that for the mass-spring system, *any* step size leads to an approximation that is qualitatively different from the true trajectory. Therefore, how can we be assured that our circular trajectory is actually correct? It is with calculus that we are able to tackle these questions.

In calculus, we are not so much concerned as to a specific value of Δt that we should pick, but rather with the question

What is happening to the approximate trajectory as we make Δt smaller and smaller?

Return our attention to Figure 2, we see that as Δt gets smaller, our spiral becomes tighter and tighter. Calculus pushes this idea one step further, asking about where we would eventually end up if we continued this process, if we were able to make Δt "infinitely-small". In this case, we know that Euler's method gives us a better and better approximation, so this end point should be the exact circular trajectory in Figure 1. Thinking a little more abstractly, if we continue to tighten the spiral more and more, eventually all of the layers of the spiral would appear to "squeeze" on top of each other, giving us this same circular trajectory.

[1] To see this, observe that the change vector at a point in the vector field is perpendicular to the line segment connecting that point to the origin. Therefore, moving any positive amount along the change vector increases the distance from the origin (by the Pythagorean theorem), resulting in the outward spiral (try drawing a picture of this).

In this way, calculus provides us with the vantage point to resolve our supposed conundrum, serving as a bridge between a calculation that was computationally feasible (Euler's Method) and our physical reality. This shift in vantage point may seem strange at first, but it is the central idea underlining calculus. It is likely the most abstract mathematics that most people have seen, and probably will ever see; however, in this abstraction is some beautiful, "oddly-satisfying" math.

Here, we only scratch the surface of calculus. We begin by introducing the **limit**, the basic building block of calculus. From here, we use limits to define the **derivative**, allowing us to study how functions change. Next, we introduce the **integral** and the notion of accumulation of a function. The **Fundamental Theorem of Calculus** gives us a beautiful connection between these seemingly-disparate notions, unifying calculus into one cohesive subject. Finally, we dip our toes into **differential equations**, looking at how we can quantitatively approach some of the same systems that we have already qualitatively studied, including logistic population growth and the Lotka-Volterra predation model.

2 Limits and Continuity

It can be said that calculus is really just the study of **limits**. In fact, the derivative and the integral, traditionally lauded as the fundamental building blocks of calculus are both defined in terms of limits, as we will see in the coming chapters. Therefore, limits serve as a good starting point for our discussion of calculus. Beyond this, limits are crucial to developing an understanding of the behavior of dynamical systems, such as the biological systems that we study in this course. For example, we saw in the previous chapter that the population model with crowding, or the **logistic model**:

$$X' = rX\left(1 - \frac{X}{k}\right)$$

tends toward a population carrying capacity, k. However, through simulation, we can see that an initially-small population will never, in any finite amount of time, reach this carrying capacity. We are in a similar conundrum to the Euler's Method example from above, and once again, we see that limits will provide us with the mathematical language needed to express this phenomenon.

2.1 What is a Limit?

Previously, we have viewed functions procedurally, perhaps as a "machine" that takes in X-values from the domain and returns y-values from the range. From this viewpoint, the only way for us to understand the function's behavior at some point $X = a$ is to evaluate $f(a)$. Pictorially, we can represent this as taking the graph of $f(X)$, and using our hands to cover most of the graph, only leaving a narrow sliver at $X = a$. This can pose a problem, for example when the function is not defined at a. Plugging into the function is impossible at a; however, limits can often provide us with a way to still glean some information about the function's behavior here.

As our first example, we consider the function $f(X) = \dfrac{X^2 - 4}{X - 2}$. We would like to know about this function's behavior around $X = 2$. Plugging into the function will not work because the function is not defined at $X = 2$; the denominator of the function is 0 here. Factoring the numerator, we see that $f(X) = \dfrac{(X+2)(X-2)}{(X-2)}$. Therefore, if we plug in any other number than 2, the denominator is non-zero and cancels with the $(X-2)$ factor in the numerator. Hence, we can similarly represent $f(X)$ as $X+2$ when $X \neq 2$, and note that it is undefined at $X = 2$. Using mathematical symbols, this gives

$$f(X) = \begin{cases} X+2 & X \neq 2 \\ \text{undefined} & = 2 \end{cases}$$

Graphically, we have

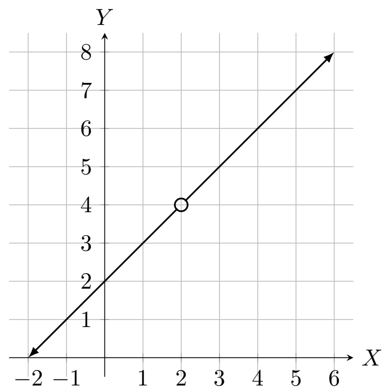


Figure 3: Graph of $Y = f(X)$

The function $f(X)$ has a **hole** at $X = 2$. However, looking at the graph, we see that around $X = 2$, the values of the function are very close to 4. We express this idea as a limit,

$$\lim_{X \to 2} f(X) = 4,$$

which we read as "the limit, as X approaches 2, of $f(X)$ is 4". Returning to our analogy from above, the limit of a function is our best guess[2] at a function's value at $X = a$ if lay down a thin rod covering $X = a$ (see Figure 4).

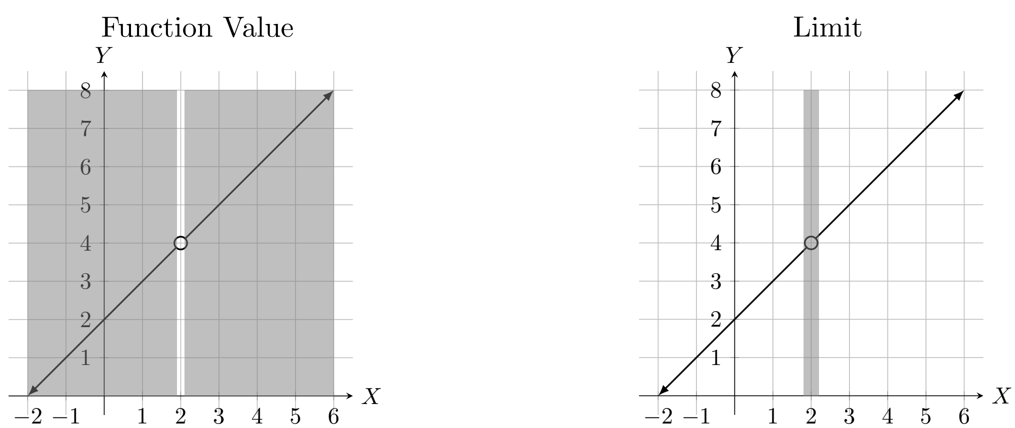


Figure 4: By "greying-out" parts of this graph, we see that the limit $\lim_{X \to 2} f(X) = 4$ exists even though the function $f(X)$ is undefined at $X = 2$.

[2]Of course, nothing precludes us from defining this covered value to be whatever we'd like.

With this intuition in mind, we can define the limit as follows [3].

Definition 2.1 (Limit)**.** Given a function $f(X)$, we say that the limit of $f(X)$ as X approaches a is L, or using mathematical symbols,

$$\lim_{X \to a} f(X) = L$$

for some X-value a and some Y-value L, if as we evaluate $f(X)$ at X-values closer to a (but not exactly $X = a$), the function values get closer and closer to L.

As we see from above, this agrees with our example $f(X)$. As we trace the curve from either direction toward $X = 2$, the value of the function approaches 4. Now that we have introduced limits, a next natural question would be how to evaluate them. We answer this question below.

2.2 Calculating Limits

The definition of the limit is based on how the function behaves as we evaluate closer and closer to a specific value $X = a$. Therefore, the most naïve approach to computing a limit is to perform such calculations. For example, with our function $f(X) = \dfrac{X^2 - 4}{X - 2}$, we can construct the table of values:

X	1	1.9	1.99	1.999	1.9999	2.0001	2.001	2.01	2.1	3
$f(X)$	3	3.9	3.99	3.999	3.9999	4.0001	4.001	4.01	4.1	5

From this tables, we can see that the value of $f(X)$ approaches 4 as X approaches 2, so we conclude that $\lim_{X \to 2} f(X) = 4$. While this method is straightforward, and seems to follow from the definition of the limit, it is not ideal for two reasons. First, it is a lot of work to have to construct these tables of values every time we need to take a limit. Second, for some functions, we can construct a table of values that misrepresents the value of the limit, so this procedure is by no means robust. Instead, we present some rules that will allow us to take limits of more and more complicated functions. The first rule deals with the limit of a constant function.

Rule 1: $\lim_{X \to a} c = c$

Since the value of the constant function c is always c, we can be assured that any sequence of X-values approaching a all evaluate to c. Therefore, by our definition, the limit is c.

If this rule seems self-evident or obvious, that is because it is. However, it is worth stating because it will be useful for our later computations.

Our second rule deals with the simplest linear function $f(X) = X$.

Rule 2: $\lim_{X \to a} X = a$

[3] For those of you with a calculus background, you may notice that this definition is not the precise $\epsilon - \delta$ definition of the limit. Here, we do not concern ourselves with this mathematical rigor, but instead focus on gaining understanding of the main ideas.

Again, the reasoning for this rule is essentially tautological. Since the output of this function is exactly its input (we call this the **identity** function for this reason), evaluating this function at X-values closer and closer to $X = a$ results in outputs that also get closer and closer to a.

The next rules deal with how we can combine functions. Specifically, they show that limits preserve addition, subtraction, scalar multiplication, multiplication, and division. We present these rules without justification [4].

Rules 3-7: Suppose that

$$\lim_{X \to a} f(X) = L \qquad \text{and} \qquad \lim_{X \to a} g(X) = M,$$

for constants a, L, M, and suppose that c is any constant. Then,

3. $\displaystyle \lim_{X \to a} \Big[f(X) + g(X)\Big] = L + M$
4. $\displaystyle \lim_{X \to a} \Big[f(X) - g(X)\Big] = L - M$
5. $\displaystyle \lim_{X \to a} \Big[c \cdot f(X)\Big] = cL$
6. $\displaystyle \lim_{X \to a} \Big[f(X) \cdot g(X)\Big] = LM$
7. $\displaystyle \lim_{X \to a} \left[\frac{f(X)}{g(X)}\right] = \frac{L}{M}, \qquad \text{as long as } M \neq 0$

While in isolation, these rules don't seem very powerful, they can allow us to compute limits of more complicated functions, as shown in the following examples.

Example 2.1. For any natural number n, $\lim\limits_{X \to a} X^n = a^n$. This is an application of Rules 2 and 6, because

$$\lim_{X \to a} X^n = \lim_{X \to a} \underbrace{\Big[X \cdot X \cdot \ldots \cdot X\Big]}_{n \text{ times}} = \underbrace{\left(\lim_{X \to a} X\right) \cdot \left(\lim_{X \to a} X\right) \cdot \ldots \cdot \left(\lim_{X \to a} X\right)}_{n \text{ times}} = \underbrace{a \cdot a \cdot \ldots \cdot a}_{n \text{ times}} = a^n$$

Example 2.2. By using Rules 3, 4, and 5, we can find a rule for determining the limit of any polynomial $p(X)$, namely $\lim\limits_{X \to a} p(X) = p(a)$ for all a[5]. As an example,

$$\lim_{X \to 3} \big[X^2 + 2X - 3\big] = \lim_{X \to 3} X^2 + 2 \cdot \lim_{X \to 3} X - \lim_{X \to 3} 3 = 3^2 + 2 \cdot 3 - 3 = 12$$

We again return to our example $f(X) = \frac{X^2-4}{X-2} = \begin{cases} X+2 & X \neq 2 \\ \text{undefined} & X = 2 \end{cases}$. Note that the definition of the limit does not consider the value of the function at $X = a$. Therefore, $\lim\limits_{X \to 2} f(X) = \lim\limits_{X \to 2} \Big[X + 2\Big] = 4$. In general, we are always able to cancel these **removable discontinuities** when calculating limits. This is one main reason that limits are so powerful: they allow us to circumvent problem values where a function is not defined. We give another example of limit calculations with removable discontinuities.

[4] Justification is given in more advanced calculus or **analysis** classes, where the concept of limits is defined precisely.
[5] As we will see later, this means that polynomials are **continuous** functions.

Example 2.3. We calculate

$$\lim_{X\to 3}\frac{X^2+2X-15}{X^2-10X+21}.$$

Plugging in $X=3$ to this rational expression [6], we see that the denominator is 0, so the expression is undefined at $X=3$. Therefore, we are going to need to be more careful. Factoring the numerator and denominator, we see that

$$\frac{X^2+2X-15}{X^2-10X+21}=\frac{(X-3)(X+5)}{(X-3)(X-7)}.$$

Since we are taking the limit as $X\to 3$, we need not consider the value of the rational expression at $X=3$. For any value other than 3, the factors of $(X-3)$ in the numerator and denominator will cancel, so

$$\lim_{X\to 3}\frac{X^2+2X-15}{X^2-10X+21}=\lim_{X\to 3}\frac{(X-3)(X+5)}{(X-3)(X-7)}=\lim_{X\to 3}\frac{(X+5)}{(X-7)}=\frac{8}{-4}=-2.$$

Consulting the graph of this function,

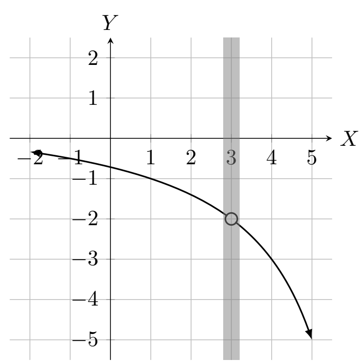


Figure 5: $Y=\frac{X^2+2X-15}{X^2-10X+21}$

confirms our calculation.

2.3 Continuity

Here, we introduce a property of functions, **continuity**, that is extremely important in calculus. The consequences of continuity are far reaching, and we will only scratch the surface here. We begin with the definition.

Definition 2.2 (Continuity)**.** A function $f(X)$ is continuous at $X=a$ if

$$\lim_{X\to a} f(X)=f(a),$$

that is, if the limit of $f(X)$ as X approaches a is equal to the function value. A function that is continuous at every point in its domain is called a **continuous function**.

[6] As a reminder, a **rational expression** is a fraction with polynomials in the numerator and denominator.

Note that Example 2.2 tells us that polynomials are continuous. In fact, a much more powerful statement, that we present without proof, is also true.

Theorem 2.1. All of the familiar functions that we work with, including polynomials, rational functions, exponential functions, logarithmic functions, and trig functions are continuous everywhere that they are defined.

This result is what will allow us to do calculus on these familiar functions. Most of the time, this will be a result of the the following, and final, limit rule.

Rule 8: Suppose that $f(X)$ is continuous at a and suppose that $g(X)$ is a function with $\lim_{X\to a} g(X) = L$. Then,

$$\lim_{X\to a} f(g(X)) = f(L).$$

While this rule looks complicated, it essentially allows us to pull limits inside of continuous functions. This will become important when we introduce differentiation rules.

One question that you may have is "Theorem 2.1 seems to suggest that every function is continuous. When are functions *not* continuous?" A simple answer to this question is that a function is not continuous wherever it is not defined. Additionally, a function is also not continuous wherever the limit does not exist.

This shifts our question to "When can a limit fail to exist?" The main answer to this question are functions that approach different values as we near a value $X = a$ from the left and from the right. This phenomenon mainly occurs in piecewise-defined functions, such as the following example.

Example 2.4. The signum function $\operatorname{sgn}(X)$ is defined to be be 1 when X is positive, -1 when X is negative, and 0 when X is 0:

$$\operatorname{sgn}(X) = \begin{cases} -1 & X < 0 \\ 0 & X = 0 \\ 1 & X > 0. \end{cases}$$

Graphically,

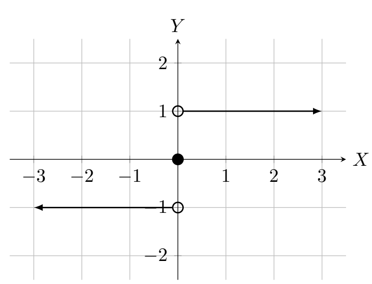


Figure 6: $Y = \operatorname{sgn}(X)$

As we approach 0 from the right, the y-values are consistently 1, suggesting we are approaching 1. We notate this as $\lim_{X\to 0^+} \operatorname{sgn}(X) = 1$. Similarly, as we approach 0 from the left, the y-values are consistently -1, suggesting we are approaching -1. We notate this as $\lim_{X\to 0^-} \operatorname{sgn}(X) = -1$. Since we approach different values depending on the approach direction, the limit $\lim_{X\to 0} \operatorname{sgn}(X)$ does not exist.

2.4 Infinite Limits and Limits at Infinity

We conclude our discussion of limits with a discussion of infinite limits and limits at infinity. We start with infinite limits.

Definition 2.3. We say that $\lim_{X \to a} f(X) = \infty$ (the limit of $f(X)$ as X approaches a is infinity), if as we consider X-values approaching a, the value of the $f(X)$ increases without bound.

We say that $\lim_{X \to a} f(X) = -\infty$, if as we consider X-values approaching a, the value of the $f(X)$ decreases without bound.

Therefore, infinite limits occur at **vertical asymptotes** of a function.

Example 2.5. Consider the function $f(X) = \frac{1}{X^2}$, shown below.

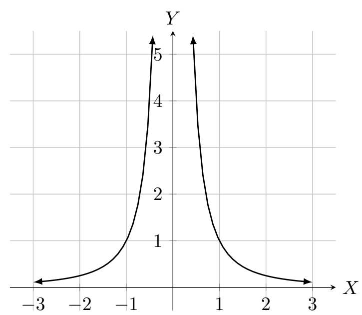


Figure 7: $Y = \frac{1}{X^2}$

Since $f(X)$ increases without bound as X approaches 0 from either direction, $\lim_{X \to 0} f(X) = \infty$.

Note that not all vertical asymptotes coincide with infinite limits, as is illustrated by the following example.

Example 2.6. Consider the function $g(X) = \tan(X)$, shown below.

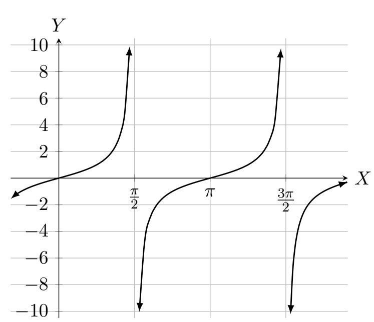


Figure 8: $Y = \tan(X)$

Since $f(X)$ increases without bound as X approaches $\frac{\pi}{2}$ from the left, but $f(X)$ decreases without bound as X approaches $\frac{\pi}{2}$ from the right, $\lim_{X \to \frac{\pi}{2}}$ does not exist.

Finally, we consider limits at infinity.

Definition 2.4. We say that $\lim_{X\to\infty} f(X) = L$ when the value of $f(X)$ gets closer to L as X gets bigger and bigger.

We say that $\lim_{X\to-\infty} f(X) = M$ when the value of $f(X)$ gets closer to M as X gets smaller and smaller.

As an interesting example of such limits, we consider the case of **rational functions**. Recall that a rational function $r(X)$ is defined as the quotient of 2 polynomials $r(X) = \dfrac{p(X)}{q(X)}$. Also, recall that the **leading term** of a polynomial is the one that includes the highest power of the variable. The **degree** is the power of this term, and the **leading coefficient** is its coefficient. For simplicity, we assume that all leading coefficients are positive, but these results can be easily extended using Rule 5 for limit calculations.

We summarize the results in the following theorem [7].

Theorem 2.2. Suppose that $r(X) = \dfrac{p(X)}{q(X)}$ is a rational function. Then,

$$\lim_{X\to\infty} r(X) = \begin{cases} 0 & \deg(p) < \deg(q) \\ \frac{\text{leading coefficient of } p}{\text{leading coefficient of } q} & \deg(p) = \deg(q) \\ \infty & \deg(p) > \deg(q) \end{cases}$$

In the first two cases, we say that $r(X)$ has a horizontal or end behavior asymptote. We conclude the chapter with a couple of examples using the above Theorem.

Example 2.7. Consider the function

$$r(X) = \frac{X^3 - 3X^2}{7 - X^4}.$$

Since the degree of the numerator (3) is less than the degree of the denominator (4), $\lim_{X\to\infty} r(X) = 0$.

Example 2.8. Consider the function

$$s(X) = \frac{X^2 - 1000}{2X}.$$

Since the degree of the numerator (2) is greater than the degree of the denominator (1), $\lim_{X\to\infty} s(X) = \infty$.

Example 2.9. Consider the function

$$t(X) = \frac{18X^2 - 13X + 7}{3X - 12X^2 + 4}.$$

Since both the numerator and denominator have degree 2,

$$\lim_{X\to\infty} t(X) = \frac{18}{-12} = -\frac{3}{2}.$$

[7] We do not include the proof here. It uses the fact that $\lim_{X\to\infty} \frac{1}{X} = 0$, along with the limit rules presented above.

3 Defining the Derivative

As we have seen, calculus provides us with a new perspective to better describe the behavior of functions. Limits give us the ability to express the value that a function is tending toward, even when the function is not defined there (in the case of removable discontinuities) or can never get there (in the case of limits at infinity). In this section, we will motivate and define another concept, the **derivative**, a measure of a function's sensitivity.

More concretely, the derivative measures the **instantaneous rate of change** of a function, how much we expect the value of the function to change relative to a small change, or **perturbation**, in the input. As is the case with many topics in calculus, we can look to physics for a motivating example.

Example 3.1. Consider a rock being dropped from a 144 foot high bridge. The projectile motion equations tell us that we can approximate the height of the rock by the function

$$Y(t) = 144 - 16t^2,$$

where Y is the height of the rock above the ground and t is the time since the rock was dropped. By plugging into this function, we can determine the height of the rock at any time, but what if we want to know about its **velocity**? Are we able to use this function to, for example, determine the velocity of the falling rock 1 second after being dropped? To answer this question, we can think about how velocity is defined. Since velocity is the signed version of speed, we have

$$\text{velocity} = \frac{\text{change in position}}{\text{change in time}},$$

or the amount of distance that the object travels in one unit of time (in our case, seconds). It is this form of the velocity that a speed camera uses to determine the speed of a car. The camera takes many pictures of an intersection in quick succession. Then, by comparing the position of cars in successive pictures using road markings the camera can calculate the speed of the car and notify the police of a speeder.

We can mimic this idea in our bridge example. Suppose that we have a camera that can take pictures in quick succession, say every 0.1 seconds. Suppose we take pictures 1 second and 1.1 seconds after the rock is dropped. The first picture will show the rock at height $Y(1) = 128$ feet, and the second picture will show the rock at height $Y(1.1) = 124.64$. Therefore, the velocity of the rock over this 0.1 second interval (we call this an **average rate of change**) is

$$\frac{124.64 - 128}{1.1 - 1} = -33.6\text{ft/sec}.$$

However, this is not quite what we were looking for. We wanted the velocity *exactly* 1 second after the rock was dropped, not the average velocity over an interval of time. To get our exact answer, we will need to revisit some ideas from our discussion of limits.

3.1 The Limit Definition of the Derivative

In a sense, this question we are asking seems somewhat paradoxical. How can we calculate the velocity, a quantity that describes the movement of an object, at a specific instant, during which the object is not moving? Hopefully by now, this type of question will seem familiar. Just as before, the answer lies in limits. If we cannot reason about the movement of an object at one instant in time, we can instead think about the movement in smaller and smaller time intervals. Suppose that we do this for our falling rock example. We already calculated the average velocity over a 0.1 second interval as -33.6 ft/sec. If we instead look over a 0.01 second time interval, we can compute the average velocity as

$$\frac{Y(1.01) - Y(1)}{1.01 - 1} = \frac{127.6784 - 128}{0.01} = -32.16\text{ft/sec.}$$

We summarize similar calculations in the following table:

time interval	0.1	0.01	0.001	0.0001	0.00001
average velocity (ft/sec)	-33.6	-32.16	-32.016	-32.0016	-32.00016

Note that here we are looking at time intervals starting at 1 second (for example, between 1 and 1.01 seconds). We could have also looked at small time intervals ending at 1 second (for example, between 0.99 and 1 seconds).

These average velocities are tending toward -32 ft/sec as we make the time interval smaller and smaller. This seems to suggest that the instantaneous velocity is -32 ft/sec, which is in fact correct. Even more important than this answer was the technique that we used to arrive at it. The table that we computed looks similar to the one that we used when calculating limits. In fact, we can express our instantaneous velocity calculation as a limit, the limit of the average velocity as the time interval tends toward 0 seconds.

We will introduce the variable h to denote the "width" of the time interval. In our example, taking pictures after 1 second and 1.1 seconds corresponds to considering the time interval $h = 0.1$. Then, the instantaneous speed after 1 second is

$$\lim_{h \to 0} \frac{Y(1+h) - Y(1)}{h},$$

where the numerator of this fraction is the change in the rock's height during the time interval and the denominator is the amount of time that passed. Using the fact that

$$Y(1+h) = 144 - 16(1+h)^2,$$

we can expand this calculation to fund that

$$\lim_{h \to 0} \frac{Y(1+h) - Y(1)}{h} = \lim_{h \to 0} \frac{(144 - 16(1+h)^2) - 128}{h} = \lim_{h \to 0} \frac{-32h - 16h^2}{h}.$$

Note here, that we have a removable discontinuity at $h = 0$ because we are able to factor h our of the numerator. Cancelling these factors of h in the numerator and denominator, our limit simplifies to

$$\lim_{h \to 0} -32 - 16h = -32\text{ft/sec.}$$

We can generalize this idea to any function, defining the notion of the derivative of a function.

Definition 3.1. Given a function $f(X)$, its derivative, which we notate as either [8] $f'(X)$ or $\frac{df}{dX}$ is given by the function

$$\lim_{h\to 0} \frac{f(X+h)-f(X)}{h},$$

defined wherever this limit exists.

Notice that this definition is more general than our falling rock example. Notice that instead of focusing on the rate of change at one specific point ($t = 1$ in our example), we have defined the derivative to be a function, which takes in an X-value and returns the instantaneous rate of change of $f(X)$ at that X-value.

Returning to our rock example, we can more generally calculate the derivative of $Y(t)$. We have,

$$\begin{aligned}
Y'(t) &= \lim_{h\to 0} \frac{Y(t+h)-Y(t)}{h} \\
&= \lim_{h\to 0} \frac{(144-16(t+h)^2)-(144-16t^2)}{h} \\
&= \lim_{h\to 0} \frac{144-16t^2-32th-16h^2-144+16t^2}{h} \\
&= \lim_{h\to 0} \frac{-32th-16h^2}{h} \\
&= \lim_{h\to 0} -32t-16h \\
&= -32t
\end{aligned}$$

Using the derivative, we can see that after 2 seconds, the rock will have velocity $Y'(2) = -64$ ft/sec, and when the rock hits the ground after 3 seconds[9], the rock will have velocity $Y'(3) = -96$ ft/sec.

[8] The prime notation is due to Isaac Newton, and the differential notation is due to Gottfried Leibniz, who independently discovered calculus, and thus codified different notational conventions.

[9] note that $Y(3) = 0$

3.2 Thinking about the Derivative Geometrically

Another, perhaps even more beautiful, way that we can look at the derivative is geometrically. We focus our attention on fraction inside of the limit definition of the derivative,

$$\frac{f(X+h)-f(X)}{h},$$

which is called a **difference quotient**. Drawing a picture of this, we have

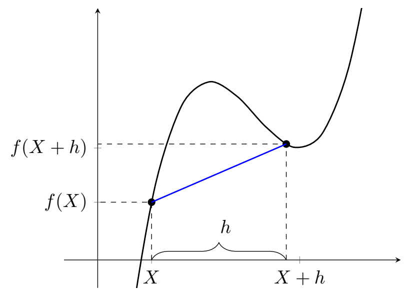


Figure 9: Visualizing the Difference Quotient

The numerator of the difference quotient is the difference in the heights of the black points, and the denominator is their horizontal spacing. Therefore, the value of the difference quotient is the familiar "rise-over-run" slope of the blue line segment connecting these points. The line containing this segment is a **secant line** of f at X (it "cuts" the function at X).

As we make h smaller and smaller, these points move closer and closer together on the curve, and the slope of this segment changes. The limit of these secant lines (shown in red below) is called the **tangent line** to f at X (it only "touches" the function at X).

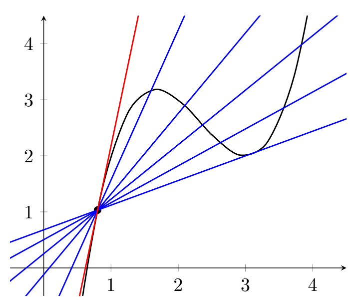


Figure 10: Secant and Tangent Lines

The slope of this red tangent line is the derivative, $f'(X)$. Moreover, the tangent line sits very close to the function at this point of tangency [10]. For this reason, we can think of the derivative $f'(X)$ as a function that takes in X-values and returns the slope of the $f(X)$ at those X-values.

[10] We will come back to this fact in a later when we discuss using the derivatives to calculate **linear approximations** to a function.

3.3 Calculations Using the Limit Definition

To conclude this section, we provide some examples of using the limit definition to calculate derivatives. As we will see in the next section, the use of more general **differentiation rules** will provide us with more efficient ways to compute complicated derivatives.

Example 3.2. As our first example, we consider a line $\ell(X) = mX + b$ (as is typical, we use m to represent the slope and b to represent the Y-intercept). Using the limit definition, we find that

$$\ell'(X) = \lim_{h\to 0} \frac{\ell(X+h) - \ell(X)}{h} = \lim_{h\to 0} \frac{(m(X+h)+b) - (mX+b)}{h} = \lim_{h\to 0} \frac{mh}{h} = m.$$

The derivative of $\ell(X)$ is m at every X-value. This agrees with our geometric interpretation of the derivative as the slope of a function, because m is the slope of $\ell(X)$.

Example 3.3. Polynomials are some of the easier functions to take derivatives using the limit definition. For example, the derivative of $g(X) = X^3$ is

$$g'(X) = \lim_{h\to 0} \frac{g(X+h) - g(X)}{h} = \lim_{h\to 0} \frac{(X+h)^3 - X^3}{h} = \lim_{h\to 0} \frac{\left(X^3 + 3X^2h + 3Xh^2 + h^3\right) - X^3}{h}$$

$$= \lim_{h\to 0} \frac{3X^2h + 3Xh^2 + h^3}{h} = \lim_{h\to 0} 3X^2 + 3Xh + h^2 = 3x^2.$$

We will see in the next section that this is an immediate consequence of the **power rule** for differentiation.

Example 3.4. As a final example, we consider the function $k(X) = \frac{1}{X}$. By the limit definition,

$$k'(X) = \lim_{h\to 0} \frac{k(X+h) - k(X)}{h} = \lim_{h\to 0} \frac{\frac{1}{X+h} - \frac{1}{X}}{h} = \lim_{h\to 0} \frac{\frac{X-(X+h)}{X(X+h)}}{h} = \lim_{h\to 0} \frac{\frac{-h}{X(X+h)}}{h} = \lim_{h\to 0} \frac{-1}{X(X+h)}$$

Whenever $X \neq 0$, this limit exists, and is equal to $\frac{-1}{X^2}$. However, notice that $k(X)$ is not defined at 0, so we need not worry about this case. Hence, $k'(X) = \frac{-1}{X^2}$. This is, in fact, another consequence of the power rule.

4 Differentiation Rules

In this section, we introduce many rules for differentiation which allow us to calculate derivatives without appealing to the cumbersome limit definition of the derivative introduced in the previous section.

4.1 Derivatives of Some Simple Functions

To begin, we will explore the derivatives of some basic functions that we will use as building blocks in more complicated derivatives.

The Power Rule:

The power rule allows us to take derivatives of powers of X. Formally, it says that for any constant n,

$$\frac{d}{dX}\Big[X^n\Big] = nX^{n-1}.$$

Note here, that we have written the rule in Leibniz's notation because it allows for a cleaner representation. This rule should be read "the derivative (with respect to X) of X^n is nX^{n-1}."

Idea of the Proof:

We can provide a proof of the power rule in the case that n is a natural number using the Binomial Theorem[11]:

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i$$

The cases $n = 0, 1$ are handled by Example 3.2 above. For $n \geq 2$, we plug in X for a and h for b in the Binomial Theorem to find that

$$\begin{aligned}(X+h)^n &= X^n + \binom{n}{1} X^{n-1}h + \sum_{i=2}^{n} \binom{n}{i} X^{n-i}h^i \\ &= X^n + nX^{n-1}h + h^2 f(X,h),\end{aligned}$$

where $f(X,h)$ is a polynomial expression in X and h.

Plugging into the limit definition for the derivative, we have

$$\begin{aligned}\frac{d}{dX}\Big[X^n\Big] &= \lim_{h\to 0} \frac{(X+h)^n - X^n}{h} \\ &= \lim_{h\to 0} \frac{X^n + nX^{n-1}h + h^2 f(X,h) - X^n}{h} \\ &= \lim_{h\to 0} \frac{nX^{n-1}h + h^2 f(X,h)}{h} \\ &= \lim_{h\to 0} nX^{n-1} + hf(X,h) \\ &= nX^{n-1},\end{aligned}$$

[11] A proof for general n requires the technique of **logarithmic differentiation**, which we will not cover in this course.

where second term cancels in the last step by limit Rule 6.

Following are some example applications of the power rule.

Example 4.1. Let $f(X) = X^5$. Then, $f'(X) = 5X^{(5-1)} = 5X^4$.

Example 4.2. Let $g(X) = \sqrt{X} = X^{\frac{1}{2}}$. Then, $g'(X) = \frac{1}{2}X^{\left(\frac{1}{2}-1\right)} = \frac{1}{2}X^{-\frac{1}{2}} = \frac{1}{2\sqrt{X}}$.

Example 4.3. Let $h(X) = 1 = X^0$. Then, $h'(X) = 0 \cdot X^{-1} = 0$.

Example 4.4. Let $j(X) = \frac{1}{X^2} = X^{-2}$. Then, $j'(X) = -2X^{-3} = \frac{-2}{X^3}$.

Derivatives of $\sin(X)$ and $\cos(X)$:

Next, we consider the derivatives of some of the elementary trig functions, sine and cosine. We start by looking at the graphs of $\sin(X)$ and $\cos(X)$.

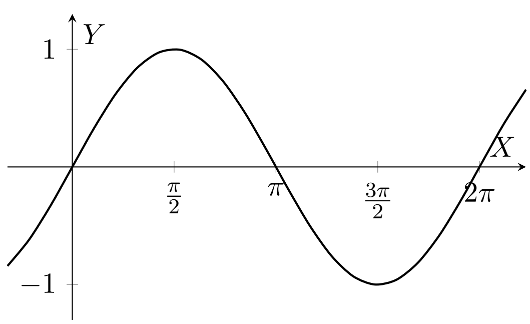


Figure 11: $Y = \sin(X)$

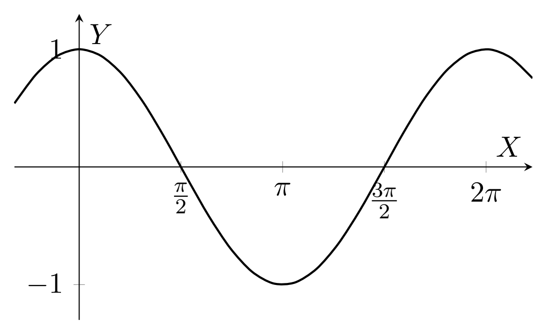


Figure 12: $Y = \cos(X)$

Notice that $\sin(X)$ has horizontal tangent lines at its peak at $\frac{\pi}{2}$ and its trough at $\frac{3\pi}{2}$. Therefore, its derivative should be 0 at $\frac{\pi}{2}$ and $\frac{3\pi}{2}$. Additionally, $\sin(X)$ has its steepest positive slope at 0 (and again at 2π), and its steepest negative slope at π. Therefore, the maximum values of its derivative should occur at $X = 0, 2\pi$, and the minimum value at $X = \pi$. Notice that $\cos(X)$ has all of these properties. In fact,

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

In order to show this from the limit definition, we will need to use two special limits:

$$\lim_{X\to 0} \frac{\sin(X)}{X} = 1 \qquad\qquad \lim_{X\to 0} \frac{\cos(X)-1}{X} = 0.$$

The first of these limits requires the application of the **Squeeze Theorem**, something that we will not cover in this course. The second limit follows from the first one, which you can verify as an exercise.

With these two limits in hand, we can differentiate $\sin(X)$ using the limit definition.

$$\frac{d}{dX}\Big[\sin(X)\Big] = \lim_{h\to 0} \frac{\sin(X+h)-\sin(X)}{h}.$$

Here, we use the trig identity

$$\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$$

to continue our simplification.

$$\begin{aligned}
&= \lim_{h\to 0} \frac{\sin(X)\cos(h) + \cos(X)\sin(h) - \sin(X)}{h} \\
&= \lim_{h\to 0} \frac{\sin(X)\left(\cos(h) - 1\right) + \cos(X)\sin(h)}{h} \\
&= \sin(X) \cdot \lim_{h\to 0} \left[\frac{\cos(h)-1}{h}\right] + \cos(X) \cdot \lim_{h\to 0} \left[\frac{\sin(h)}{h}\right] \\
&= \sin(X) \cdot 0 + \cos(X) \cdot 1 \\
&= \cos(X).
\end{aligned}$$

By similar reasoning, both geometric and algebraic, we can show that

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

We will give a different proof of this fact using the **chain rule** in discussion.

Derivatives of e^X and $\ln(X$:

Finally, we give the differentiation rule for e^X and $\ln(X)$:

$$\frac{d}{dX}\Big[e^X\Big] = e^X.$$

$$\frac{d}{dX}\Big[\ln(X)\Big] = \frac{1}{X}.$$

We present these rules without proof, as the proofs require more advanced techniques than we will cover in the course.

4.2 Derivative Rules for Combining Functions

Now that we have established differentiation rules for some common base functions, we will look at rules for combining functions. The first rules that we will look at allow us to take derivatives of **linear combinations** of functions. That is, functions that can be expressed as a sum or difference of simpler functions.

Sum Rule:

Just as with limits (Rule 3), the derivative of a sum is the sum of the the corresponding derivatives:

$$\frac{d}{dX}\Big[f(X) + g(X)\Big] = \frac{df}{dX} + \frac{dg}{dX}$$

To prove this, let the function $s(X)$ be defined as the sum $s(X) := f(X) + g(X)$. Then, using the limit definition

$$\begin{aligned}
\frac{ds}{dX} &= \lim_{h\to 0} \frac{s(X+h) - s(X)}{h} \\
&= \lim_{h\to 0} \frac{(f(X+h) + g(X+h)) - (f(X) + g(X))}{h} \\
&= \lim_{h\to 0} \left[\frac{f(X+h) - f(X)}{h} + \frac{g(X+h) - g(X)}{h}\right] \\
&= \lim_{h\to 0} \left[\frac{f(X+h) - f(X)}{h}\right] + \lim_{h\to 0} \left[\frac{g(X+h) - g(X)}{h}\right] \\
&= \frac{df}{dX} + \frac{dg}{dX}
\end{aligned}$$

Example 4.5. Consider the function

$$f(X) = X^2 + \sin(X).$$

This function is the sum of the functions X^2 and $\sin(X)$. By the Sum Rule, its derivative will be the sum of the functions $\frac{d}{dX}\left[X^2\right] = 2X$ and $\frac{d}{dX}\left[\sin(X)\right] = \cos(X)$, so

$$f'(X) = 2X + \cos(X).$$

Constant Multiple Rule:

Similarly, we have a constant multiple rule for derivatives that inherits from the constant multiple rule for limits (Rule 5). For any constant c:

$$\frac{d}{dX}\left[c \cdot f(X)\right] = c \cdot \frac{df}{dX}$$

To prove this, let the function $m(X)$ be defined as the constant multiple $m(X) := c \cdot f(X)$. Then, using the limit definition

$$\begin{aligned}
\frac{dm}{dX} &= \lim_{h\to 0} \frac{m(X+h) - m(X)}{h} \\
&= \lim_{h\to 0} \frac{c \cdot f(X+h) - c \cdot f(X)}{h} \\
&= \lim_{h\to 0} \left[c \cdot \frac{f(X+h) - f(X)}{h}\right] \\
&= c \cdot \lim_{h\to 0} \frac{f(X+h) - f(X)}{h} \\
&= c \cdot \frac{df}{dX}
\end{aligned}$$

By combining these rules with the power rule, we can take derivatives of all polynomials, as is shown in the following example.

Example 4.6. Suppose that $f(X) = 2X^3 - 7X + 4$. Then,

$$\begin{aligned}\frac{df}{dX} &= 2 \cdot \frac{d}{dX}\Big[X^3\Big] - 7 \cdot \frac{d}{dX}\Big[X^1\Big] + 4 \cdot \frac{d}{dX}\Big[X^0\Big] \\ &= 2 \cdot 3X^2 - 7 \cdot 1 + 4 \cdot 0 \\ &= 6X^2 - 7\end{aligned}$$

<u>Product Rule:</u>

Next, we look at the rule for differentiating a product of functions, $p(X) = f(X) \cdot g(X)$. Unfortunately, the product rule is not as simple as the sum rule; the derivative of a product is *not* the product of the derivatives. In other words,

$$\frac{d}{dX}\Big[f(X) \cdot g(X)\Big] \neq \frac{df}{dX} \cdot \frac{dg}{dX}$$

For some intuition about the product rule (but not a rigorous proof), we can draw a picture. We can think of the function $p(X)$ as representing the area of a rectangle with width $f(X)$ and height $g(X)$. Then, $\frac{dp}{dX}$ represents the rate that the area of this rectangle is changing when we perturb X. Suppose that we increase X by h. Then, the width of the rectangle changes by $\frac{df}{dX} \cdot h$ and the height of the rectangle changes by $\frac{dg}{dX} \cdot h$.

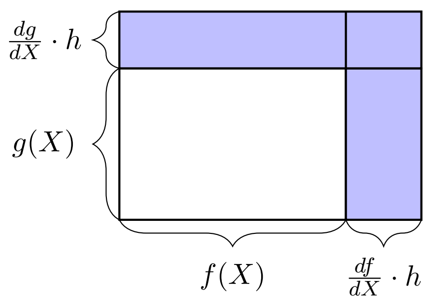


Figure 13: Visual representation of $p(X) = f(X) \cdot g(X)$

The change in the area is shaded in blue. From the picture, this shaded region has area

$$f(X) \cdot \frac{dg}{dX} \cdot h + g(X) \cdot \frac{df}{dX} \cdot h + \frac{df}{dX} \cdot \frac{dg}{dX} \cdot h^2.$$

Therefore, the instantaneous rate of change of this area is

$$\begin{aligned}\frac{dp}{dX} &= \lim_{h \to 0} \frac{f(X) \cdot \frac{dg}{dX} \cdot h + g(X) \cdot \frac{df}{dX} \cdot h + \frac{df}{dX} \cdot \frac{dg}{dX} \cdot h^2}{h} \\ &= \lim_{h \to 0} \Big[f(X) \cdot \frac{dg}{dX} + g(X) \cdot \frac{df}{dX} + h \cdot \frac{df}{dX} \cdot \frac{dg}{dX}\Big] \\ &= f(X) \cdot \frac{dg}{dX} + g(X) \cdot \frac{df}{dX}.\end{aligned}$$

This is the product rule:

$$\frac{d}{dX}\Big[f(X) \cdot g(X)\Big] = f(X) \cdot \frac{dg}{dX} + g(X) \cdot \frac{df}{dX}.$$

The mnemonic you can use to remember this is "first times derivative of the second plus second times derivative of the first." Following are a couple of examples where the product rule is used.

Example 4.7. Consider the function $f(X) = X^2 \sin(X)$. This is a product of two functions (X^2 and $\sin(X)$), so to take the derivative, we need to use the product rule. Computing the derivative, we have

$$\frac{d}{dX}\Big[X^2 \sin(X)\Big] = X^2 \cdot \frac{d}{dX}\Big[\sin(X)\Big] + \sin(X) \cdot \frac{d}{dX}\Big[X^2\Big] = X^2 \cos(X) + 2X \sin(X).$$

Example 4.8. Consider the function $g(X) = X \ln(X) - X$. We will need to use the product rule to compute the derivative of this first term. We have

$$\frac{dg}{dX} = X \cdot \frac{d}{dX}\Big[\ln(X)\Big] + \ln(X) \cdot \frac{d}{dX}\Big[X\Big] - 1 = X \cdot \frac{1}{X} + \ln(X) - 1 = \ln(X).$$

Quotient Rule:

The quotient rule handles differentiation of a quotient of two functions. Just like with the product rule, the derivative of a quotient is *not* the quotient of the derivatives. Rather, we have

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

A common mnemonic for the quotient rule is "low d-high minus high d-low over low squared." We omit a proof of the quotient rule here, but you will verify the quotient rule using the product and chain rules in discussion. Again, we conclude with some examples of using the quotient rule.

Example 4.9. Consider the function $f(X) = \frac{e^X}{X}$. We compute its derivative using the quotient rule:

$$\frac{df}{dX} = \frac{X \cdot \frac{d}{dX}\left[e^X\right] - e^X \cdot \frac{d}{dX}\left[X\right]}{X^2} = \frac{Xe^X - e^X}{X^2}.$$

Example 4.10. Consider the function $g(X) = \tan(X) = \frac{\sin(X)}{\cos(X)}$. By the quotient rule,

$$\frac{dg}{dX} = \frac{\cos(X) \cdot \frac{d}{dX}\left[\sin(X)\right] - \sin(X) \cdot \frac{d}{dX}\left[\cos(X)\right]}{\cos^2(X)} = \frac{\cos^2(X) + \sin^2(X)}{\cos^2(X)} = \frac{1}{\cos^2(X)} = \sec^2(X).$$

Note that the third equality follows from the Pythagorean identity $\sin^2(X) + \cos^2(X) = 1$.

Chain Rule:

Our last rule, the chain rule, allows us to take derivatives of **compositions** of functions. Many people find this to be the most difficult rule, mainly because it can be difficult to spot and understand the composition.

To compose functions is to make the input of one function another function, instead of just the variable, X. For example, we can compose $f(X) = \sin(X)$ and $g(X) = X^2$ in two different ways:

1. $f(g(X))$, read "f of g of X" is the result of plugging $g(X) = X^2$ in for X in $f(X) = \sin(X)$, yielding $\sin(X^2)$.

2. $g(f(X))$, read "g of f of X" is the result of plugging $f(X) = \sin(X)$ in for X in $g(X) = X^2$, yielding $(\sin(X))^2 = \sin^2(X)$.

The chain rule, expressed in Newton's notation is

$$[f(g(X))]' = f'(g(X)) \cdot g'(X).$$

The derivative is the derivative of the outside function, composed with the inside function times the derivative of the inside function. To represent the chain rule in Leibniz's notation, we use the variable Y to represent $g(X)$, so $f(g(X)) = f(Y)$. Then,

$$\frac{df}{dX} = \frac{df}{dY} \cdot \frac{dY}{dX}.$$

For some intuition for the correctness of the chain rule (but not a rigorous proof), we will again turn to a picture.

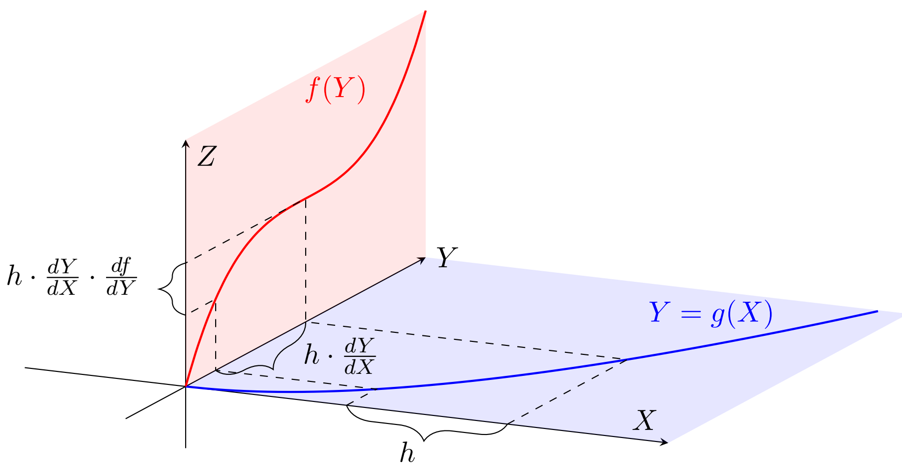


Figure 14: Visualizing the Chain Rule

In this picture, the input to the function $f(g(X))$ is a value on the X-axis, and its output is on the Z-axis. We split the evaluation of $f(g(X))$ into 2 parts. First, we plug X into g (the blue function), to compute $Y = g(X)$. From this value on the Y-axis, we plug into f (the red function), to compute $Z = f(Y) = f(g(X))$.

To compute the derivative of $f(g(X))$ is to determine the difference in height of the outputs for two inputs separated by h (thinking of h as getting very small). When h is small, two inputs h apart will be sent by g to two outputs $h \cdot \frac{dY}{dX}$ apart on the Y-axis (this is the meaning of the derivative of g). But these outputs are inputs to f, which will in turn map them to points $h \cdot \frac{dY}{dX} \cdot \frac{df}{dY}$ apart. Thus, for small h, a change of h in the input to $f(g(X))$ results in a change of $h \cdot \frac{dY}{dX} \cdot \frac{df}{dY}$ to the output, so the derivative

$$\frac{df}{dX} = \lim_{h \to 0} \frac{h \cdot \frac{dY}{dX} \cdot \frac{df}{dY}}{h} = \frac{dY}{dX} \cdot \frac{df}{dY}.$$

Here, we give some examples of the chain rule.

Example 4.11. Consider the function $f(X) = (3X^2 - 4X)^7$. This is a composition of functions, with inner function $Y = 3X^2 - 4X$ and outer function $f(Y) = Y^7$. Using the chain rule, we compute

$$\begin{aligned}
\frac{df}{dX} &= \frac{d}{dY}\Big[Y^7\Big] \cdot \frac{d}{dX}\Big[3X^2 - 4X\Big] \\
&= 7Y^6 \cdot (6X - 4) \\
&= 7(3X^2 - 4X)^6 \cdot (6X - 4).
\end{aligned}$$

Notice that, in the last step, we substitute into Y so that our final expression of the derivative is in terms of X only.

Example 4.12. Consider the function $g(X) = \sin^2(X) + \cos^2(X)$. Both terms of the function are compositions, where the inner functions are the respective trig functions (sine and cosine), and the outer functions are both squaring. By the chain rule, we have

$$\begin{aligned}
\frac{dg}{dX} &= 2\sin(X) \cdot \frac{d}{dX}\Big[\sin(X)\Big] + 2\cos(X) \cdot \frac{d}{dX}\Big[\cos(X)\Big] \\
&= 2\sin(X) \cdot \cos(X) + 2\cos(X) \cdot (-\sin(X)) \\
&= 2\sin(X)\cos(X) - 2\sin(X)\cos(X) \\
&= 0.
\end{aligned}$$

Since the derivative of g is 0 for all X, g has slope 0 everywhere. Thus, g is a horizontal line, or a constant function. To figure out the value of this constant, we can plug in any value of X. Choosing $X = 0$, we have $g(0) = \sin^2(0) + \cos^2(0) = 0^2 + 1^2 = 1$.

We have proven the Pythagorean identity $\sin^2(X) + \cos^2(X) = 1$.

4.3 Examples of More Complicated Derivatives

We finish this section with some examples of more complicated derivatives. Such examples usually require the application of multiple rules. We separate each of the rules onto its own line to make the calculations easier to follow.

Example 4.13. We calculate the derivative of $f(X) = 3X^2 e^X \cos(X)$.

$$\begin{aligned}
\frac{df}{dX} &= 3X^2 \cdot \frac{d}{dX}\Big[e^X \cos(X)\Big] + \big(e^X \cos(X)\big) \cdot \frac{d}{dX}\Big[3X^2\Big] && \text{(product rule)} \\
&= 3X^2 \cdot \left(e^X \cdot \frac{d}{dX}\Big[\cos(X)\Big] + \cos(X) \cdot \frac{d}{dX}\Big[e^X\Big]\right) + \big(e^X \cos(X)\big) \cdot 6X && \text{(product rule)} \\
&= 3X^2 \cdot \big(-e^X \sin(X) + e^X \cos(X)\big) + 6Xe^X \cos(X)
\end{aligned}$$

Example 4.14. We calculate the derivative of $g(X) = \frac{X\ln(X)}{e^X}$.

$$\begin{aligned}
\frac{dg}{dX} &= \frac{e^X \cdot \frac{d}{dX}\Big[X\ln(X)\Big] - X\ln(X) \cdot \frac{d}{dX}\Big[e^X\Big]}{(e^X)^2} && \text{(quotient rule)} \\
&= \frac{e^X \cdot \Big(X \cdot \frac{d}{dX}\Big[\ln(X)\Big] + \ln(X) \cdot \frac{d}{dX}\Big[X\Big]\Big) - Xe^X\ln(X)}{e^{2X}} && \text{(product rule)} \\
&= \frac{e^X \cdot \big(X \cdot \frac{1}{X} + \ln(X) \cdot 1\big) - Xe^X\ln(X)}{e^{2X}} \\
&= \frac{e^X \cdot (1 + \ln(X)) - Xe^X\ln(X)}{e^{2X}} \\
&= \frac{1 + \ln(X) - X\ln(X)}{e^X}
\end{aligned}$$

Example 4.15. We calculate the derivative of $h(X) = \sin\left(e^{2X}\right)$. This is a (nested) composition of functions with innermost function $Y = 2X$, middle function $Z = e^Y$, and outer function $h(Z) = \sin(Z)$. Therefore,

$$\begin{aligned}
\frac{dh}{dX} &= \frac{d}{dZ}\Big[\sin(Z)\Big] \cdot \frac{d}{dX}\Big[e^{2X}\Big] && \text{(chain rule)} \\
&= \frac{d}{dZ}\Big[\sin(Z)\Big] \cdot \frac{d}{dY}\Big[e^Y\Big] \cdot \frac{d}{dX}\Big[2X\Big] && \text{(chain rule)} \\
&= \cos(Z) \cdot e^Y \cdot 2 \\
&= 2\cos(e^Y) \cdot e^Y \\
&= 2\cos(e^{2X}) \cdot e^{2X}
\end{aligned}$$

Example 4.16. We calculate the derivative of $k(X) = \frac{X\sin(\sqrt{X})}{\cos(X)}$. In the numerator, we have a composition of functions, with inner function $Y = \sqrt{Z}$ and outer function $\sin(Y)$.

$$\begin{aligned}
\frac{dk}{dX} &= \frac{\cos(X) \cdot \frac{d}{dX}\Big[X\sin(\sqrt{X})\Big] - X\sin(\sqrt{X}) \cdot \frac{d}{dX}\Big[\cos(X)\Big]}{\cos^2(X)} && \text{(quotient rule)} \\
&= \frac{\cos(X) \cdot \Big(X \cdot \frac{d}{dX}\Big[\sin(\sqrt{X})\Big] + \sin(\sqrt{X}) \cdot \frac{d}{dX}\Big[X\Big]\Big) + X\sin(\sqrt{X})\sin(X)}{\cos^2(X)} && \text{(product rule)} \\
&= \frac{\cos(X) \cdot \Big(X \cdot \frac{d}{dY}\Big[\sin(Y)\Big] \cdot \frac{d}{dX}\Big[\sqrt{X}\Big] + \sin(\sqrt{X})\Big) + X\sin(\sqrt{X})\sin(X)}{\cos^2(X)} && \text{(chain rule)} \\
&= \frac{\cos(X) \cdot \Big(X \cdot \cos(Y) \cdot \frac{1}{2\sqrt{X}} + \sin(\sqrt{X})\Big) + X\sin(\sqrt{X})\sin(X)}{\cos^2(X)} \\
&= \frac{\cos(X) \cdot \Big(\sqrt{X} \cdot \cos(\sqrt{X}) + 2\sin(\sqrt{X})\Big) + 2X\sin(\sqrt{X})\sin(X)}{2\cos^2(X)}
\end{aligned}$$

5 Linear Approximation

To finish our introduction of derivatives, we will introduce the concept of linear approximation. Previously, we discussed that the derivative can be viewed geometrically as a slope of the **tangent line** to our function at a specified X-value. Here, we expand upon this idea of the tangent line; we will discuss how to compute its equation and see that it can, in some cases, serve as a good approximation to our function. To motivate this, we turn to an example.

Example 5.1. Consider the function $f(X) = X^2$, with derivative $\frac{df}{dX} = 2X$. At $X = 1$, the function has value 1 and derivative 2. Therefore, the tangent line to $f(X)$ at $X = 1$ has slope 2 and passes through the point $(1, 1)$. Thus, the tangent line has equation $L(X) = 2X - 1$. We plot both $f(X)$ and $L(X)$ below.

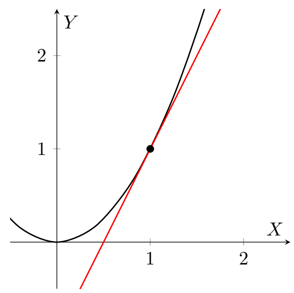


Figure 15: $Y = f(X)$ and $Y = L(X)$

Notice that near the point of tangency, $(1, 1)$, the tangent line $L(X)$ is almost indistinguishable from $f(X)$.Therefore, as long as we are close to $X = 1$, we can use $L(X)$, the equation of a line, to accurately approximate the values of $f(X)$. For this reason, we call $L(X)$ a **linear approximation** to $f(X)$ at $X = 1$.

5.1 Calculating the Linear Approximation

In order to use the linear approximation, we will need a way to find it. We are looking for the equation of a line. To find such an equation we will need to know its slope, m, and one point that it passes through (X_0, Y_0). With this information, we can write its equation using point-slope form:

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

Adding Y_0 to both sides, we can rearrange to be:

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

To find the slope and the point, we can turn directly to our intuition about the tangent line. We know that the slope of the tangent line at X_0 is the derivative $f'(X_0)$. But we also know that the tangent line passes through the point of tangency, which lies on our original function. That is, it passes through the point $(X_0, f(X_0))$. Substituting this information into the equation above, we obtain the linear approximation equation to $f(X)$ at $X = X_0$:

$$L(X) = f'(X_0)(X - X_0) + f(X_0).$$

Verify that $L(X)$ from Example 5.1 adheres to this equation. Below are some examples of calculating the tangent line.

Example 5.2. We consider calculating the tangent line to $f(X) = \sin(X)$ at $X = 0$. Here, $f(0) = 0$, so the tangent line passes through the origin. Its slope is $f'(0) = \cos(0) = 1$, so its equation is $L(X) = X$. We can verify this by looking at a graph.

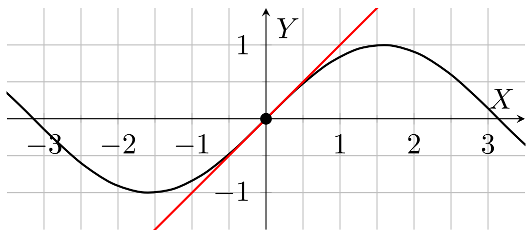


Figure 16: Tangent Line to $Y = \sin(X)$ at $X = 0$

Example 5.3. We consider calculating the tangent line to $g(X) = X^3 - X^2$ at $X = 1$. To calculate the slope of the tangent line we plug $X = 1$ into the derivative $g'(X) = 3X^2 - 2X$, to get $g'(1) = 1$. The point of tangency is $(1, g(1)) = (1, 0)$. Therefore, the tangent line has equation,

$$L(X) = 1(X - 1) + 0 = X - 1.$$

We can verify this by looking at a graph.

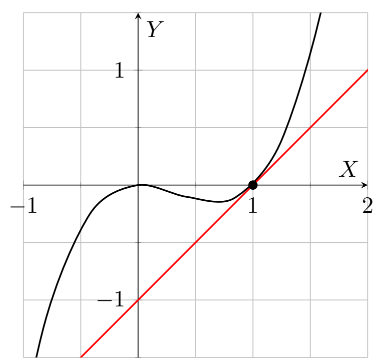


Figure 17: Tangent Line to $Y = X^3 - X^2$ at $X = 1$

5.2 Tangent Lines as Linear Approximations

From the above examples, we see that the tangent line sits very close the the function nearby the point of tangency. This is made concrete by the following theorem.

Theorem 5.1. The tangent line is the best linear approximation to a function nearby the point of tangency.

What this theorem means is that one wants to use a line to approximate a function around a point, then the tangent line at that point is the best choice. We omit the proof of this theorem[12], and instead discuss its application.

Often times, the functions that arise from a real-world application, such as a biological model, can be difficult to work with, even just to compute the values at some points. However, as long as there are some "easy" evaluation points (and provided that we are able to differentiate the function), we can use linear approximations to get a good estimate of the functions behavior.

As a more mathematical example of such a phenomenon, we consider the calculation of a cube root.

Example 5.4. Suppose that we would like to estimate $\sqrt[3]{67}$ without a calculator. We notice that $\sqrt[3]{64} = 4$ is easy to calculate, so we can approximate $\sqrt[3]{67}$ using a linear approximation to $f(X) = \sqrt[3]{X}$ at $X = 64$.

The slope of this approximation is $f'(64) = \frac{1}{3}(64)^{-\frac{2}{3}} = \frac{1}{3} \cdot \frac{1}{8} = \frac{1}{24}$, and the point of tangency is $(64, 4)$. Therefore, the linear approximation equation is

$$L(X) = \frac{1}{24}(x - 64) + 4 = \frac{X + 32}{24}.$$

We compute $L(67) = \frac{99}{24} = \frac{33}{8} = 4.125$. The actual value of $\sqrt[3]{67} \approx 4.062$, so our approximation is fairly accurate.

The other benefit of a linear approximation is that it allows us to interpret the derivative as a **sensitivity** of the dependent variable of our system to changes in the independent variable (this idea was used in our pictorial reasoning about the product and chain rules in Section 4). We conclude this section with an example of a biological system where this idea can be applied.

Example 5.5. Suppose that we can model the population of fish, F, in a pond as a function of the water temperature T (so $F(T) = \ldots$). Further, suppose that the current water temperature is 18 degrees Celsius, and our $F'(18) = -200$. Then, as long as the change in water temperature is not too great, we can expect the fish population to decrease by 200 for every degree increase in temperature in the pond. Thus, if the water temperature increases by 0.5 degrees, then we can expect the fish population to change by $0.5 \cdot (-200) = -100$, that is to decrease by 100 fish.

As we will see in Chapter 3, when we study **bifurcations**, understanding what parameters of a system we have control over, and how our modification of these parameters effects the system is of pivotal importance in environmental management. For example, a government deciding on different water treatment practices, which can change the temperature of the pond, can use the above information to account for the affects of their decision on the fish population.

[12]This is actually a special case of **Taylor's Theorem**, which tells us how we can use higher-degree polynomials to obtain better approximate to function at a point. Taylor's Theorem and its proof are often topics in a second-semester calculus course.

6 Defining the Integral

In the previous sections, we explored the derivative. We learned that the derivative of a function $f(X)$ is another function $f'(X)$ that tells us the instantaneous slope of $f(X)$ at any X-value. For physical intuition, we noticed that taking the derivative of a function $Y(t)$, which measured the height of a falling rock, we could determine its speed. Is it possible to perform this process in reverse? If we had a function that gave us the speed of the rock at any time, and we knew how long it has been since the rock was dropped, is it possible to figure out how far the rock has fallen?

As we will see in the next sections, the answer to this question is yes; the technique that allows it is called **integration**. In this section, we will introduce the definite integral and give a definition of integration in terms of limits. In the next section, we present the Fundamental Theorem of Calculus, one of the most beautiful results in mathematics that masterfully links together the concepts of integration and differentiation. Finally, we will conclude the chapter with a brief foray into the area of differential equations where we will use the calculus techniques that we have developed to quantitatively study the behaviors of change equations.

6.1 Functions and Area

To begin our discussion of integration, we will consider the following problem:

> Given a function, how can I calculate the area between it and the X-axis?

Pictorially, if I look at the graph of a function $f(X)$ and choose X-values a and b, I'd like to calculate the area of the shaded region enclosed by $Y = f(X)$, $Y = 0$, $X = a$, and $X = b$ (see the figure below).

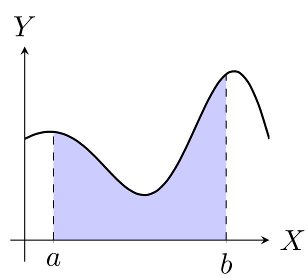


Figure 18: Visualizing the area under $Y = f(X)$

We colloquially refer to this as the "area under the curve", even though this area will sit above the function when the function is below the X-axis. We introduce a new piece of notation, called the integral, to represent this quantity. We'd notate the blue shaded region from Figure 6.1 as

$$\int_a^b f(X)dX,$$

where a and b are the lower and upper bounds of the integral (respectively), $f(X)$ is the function that we are integrating, and dX tells us that our bounds are X-values.

One small idiosyncrasy of the integral is that it assigns a sign (positive or negative) for the area. When the shaded area sits above the X-axis (so $f(X)$ is positive), the integral evaluates to a positive area. However, when the shaded area sits below the X-axis (so $f(X)$ is negative), the integral evaluates to a negative area. Although this may seem arbitrary now, the reason for it will become apparent later.

Example 6.1. We consider the integral $\int_0^{2\pi} \sin(X)\ dX$. This represents the area between the curve $Y = \sin(X)$ and the X-axis, as shown in the following figure.

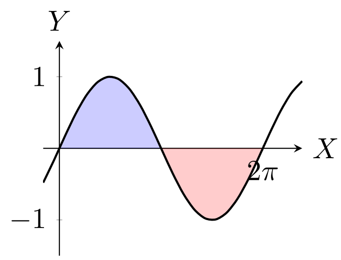


Figure 19: $\int_0^{2\pi} \sin(X)\ dX$

The integral would compute the areas of the blue and red shaded regions. The area of the blue region would be positive because it sits above the X-axis, and the area of the red region would be negative because it sits below the X-axis. In this case, we can see that the regions have the same shape (only flipped), so they have the same area. Hence, the (blue) positive area cancels with the (red) negative area, so the integral evaluates to 0.

For some functions, it is easy to calculate the integral geometrically using area formulas that you are already used to. Such is the case in the following example.

Example 6.2. Consider the integral $\int_2^6 \frac{1}{2}X\ dX$. Visualizing this integral, we have:

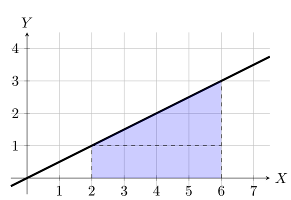


Figure 20: $\int_2^6 \frac{1}{2}X\ dX$

The shaded area consists of a 4×1 rectangle (with area 4) and a 4×2 right triangle (with area 4), so the value of the integral is 8.

However, when the functions that we are integrating are more complicated, we cannot turn to the area formulas that we know. Instead, we will need to be a little more clever, and return to ideas from calculus.

6.2 Riemann Sums

Here, we consider computing $\int_0^2 X^2 \, dX$, shown below.

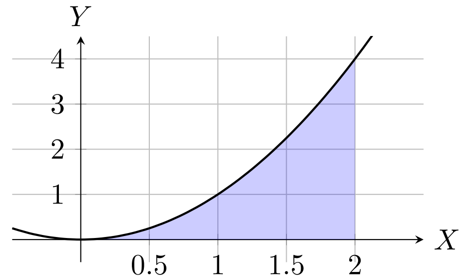


Figure 21: $\int_0^2 X^2 \, dX$

We don't have an easy area formula for this shaded region. Instead, we will come up with a process for estimating the area. The simplest method is to build up the blue area using side-by side rectangles with varying heights, like a shelf of books. As a rule, we will always choose the height of the rectangle so that its top-left corner touches the function. This is called a **left Riemann sum**[13]. Suppose that we choose rectangles with width $h = 0.5$ [14]. Then, we can visualize the left Riemann sum as:

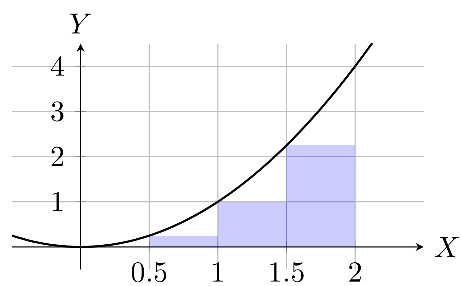


Figure 22: Left Riemann sum with $h = 0.5$

The area approximated by this Riemann sum is simply the sum of the areas of these rectangles, in this case $0.25(0.5) + 1(0.5) + 2.25(0.5) = 1.75$. Note that this is an under-approximation of the area because our rectangles fail to include all of the area under the parabola. We missed the white "triangles" between the tops of our rectangles and the parabola.

How could we make our approximation better?

[13]We can similarly define a right or midpoint Riemann sum, depending on where on the top of the rectangle we want the function to touch.

[14]Although using h for width seems weird, we choose this notation because it is similar to that from derivatives.

If we used more, thinner rectangles, we could fill in some of this missed area. For example, a left Riemann sum with rectangles of width 0.25 looks like:

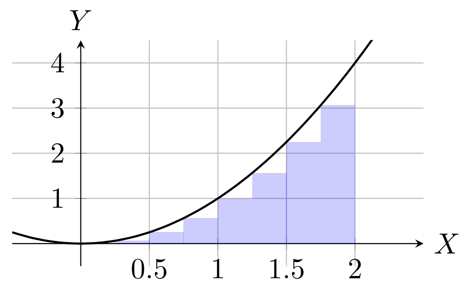


Figure 23: Left Riemann sum with $h = 0.25$

This time, the area computed by the Riemann sum is 2.1875, a slightly larger area because we have filled in some of the gap. At this point, your calculus gears should be turning. By making h smaller, we were able to move closer to the the actual value of the area. Therefore, if we take the limit (assuming it exists) as h approaches 0, we should be able to calculate the exact area, that is the integral.

To write out a limit expression for the integral[15], we first need to write an expression for the area computed by a Riemann sum. Let n be the number of width-h rectangles that we can completely fit between a and b (in mathematical symbols, we write this as $n := \lfloor \frac{b-a}{h} \rfloor$). From our pictures, we see that the height of each rectangles is the value of the function at the left edge of the rectangle. Hence, we can express the area as

$$\underbrace{h \cdot f(a)}_{\text{first rectangle}} + \underbrace{h \cdot f(a+h)}_{\text{second rectangle}} + \cdots + \underbrace{h \cdot f(a+(n-1)h)}_{n\text{'th rectangle}}.$$

In summation notation, we can express this sum more compactly as

$$\sum_{i=0}^{n-1} h \cdot f(a+ih).$$

Here, the upper bound of the sum is calculating how many rectangles of width h fit between a and b. Therefore, from above, we see that taking the limit of this sum gives us the Riemann definition of the integral:

$$\int_a^b f(X)\, dX = \lim_{h \to 0} \left[\sum_{i=0}^{n-1} h \cdot f(a+ih) \right],$$

when this limit exists. In this case, we say that $f(X)$ is **integrable** on $[a, b]$, or simply just integrable when this holds for all a, b. For our purposes, we will use the following fact [16].

Theorem 6.1. Continuous functions are integrable.

[15] There are many different ways to express an integral as a limit, which all evaluate to the correct area. We focus on only one here.

[16] Proving this fact is not that simple, and is typically left to an **analysis** (rigorous calculus) course.

Therefore, all of the integrals that we will care about taking exist. This still leaves us with the question of how to actually compute integrals. Unlike with the derivative, this limit definition is very messy and almost impossible to use directly. Therefore, we will need to find another way, a back door, to calculate these integrals. We will do this in the following section when we introduce the Fundamental Theorem of Calculus.

6.3 Integral Bound Rules

We conclude this section by introducing some rules involving integral bounds. These rules make use of the geometric view of the integral as an area under a curve.

Rule 1: Given any constant a and any function $f(X)$ defined at a,

$$\int_a^a f(X)\,dX = 0.$$

This rule is fairly straightforward. The area that integral represents has width 0 (it is the integral from a to a), and therefore must have area 0.

Rule 2: Given any constants $a \le b \le c$ and any function $f(X)$ defined on $[a, c]$,

$$\int_a^b f(X)\,dX + \int_b^c f(X)\,dX = \int_a^c f(X)\,dX.$$

To see this, it is easiest to look at a picture.

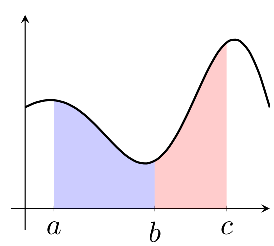


Figure 24: Visualizing Rule 2

Here, $\int_a^b f(X)\,dX$ represents the blue region, $\int_b^c f(X)\,dX$ represents the red region, and $\int_a^c f(X)\,dX$ is the combined regions.

So far, our definition of the integral only allows for integrals where the upper bound is greater than or equal to the lower bound, that is $\int_a^b f(X)\,dX$ where $b \ge a$. To handle the case where $b < a$, we expand our definition so that

$$\int_a^b f(X)\,dX = -\int_b^a f(X)\,dX.$$

That is, flipping the bounds of integration negates the value of the integral. This expanded definition makes Rule 2 hold for arbitrary a, b, c (they no longer need to be in order). This generalized Rule 2 will be used in our proof of the second Fundamental Theorem of Calculus.

7 The Fundamental Theorem of Calculus

At long last, we have built up to the Fundamental Theorem of Calculus, a beautiful result that unifies the two seemingly disparate ideas that we have explored thus far: the derivative and the integral. As a result of the theorem, we will see that these concepts are just two different perspectives of the same idea. Beyond this, the theorem provides us the ability to compute the value of integrals, which seemed intractable from the integral definition alone.

7.1 The First Fundamental Theorem of Calculus

When we studied differentiation, we first discussed calculating derivatives at a specific X-value, using limits to compute the slope of a tangent line to a function, $f(X)$. Then, we expanded this idea by viewing the derivative as a function $f'(X)$ which can tell us this slope at any X-value. Here, we do a similar thing for integrals.

Given a function $f(X)$, we define the function $F(X)$ [17] as

$$F(X) := \int_a^X f(t)\, dt,$$

for some constant a (called the base-point). Take a second to appreciate this. Here, our variable X changes the upper bound of the integral, so the output of our function is an area. Below, we explore an example of such a function.

Example 7.1. Consider the linear function $f(X) = 2X$. Following our notation from above, we define the function

$$F(X) := \int_0^X f(t)\, dt.$$

In this case, we have chosen 0 as our base-point. To simplify our explanation, we focus on positive values of X. Then, for any t-value X, $F(X)$ is the area under $f(t)$ between 0 and X.

We can visualize this as follows.

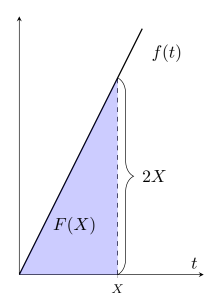


Figure 25: Visualizing $F(X)$

Here, we see that the shaded area representing the integral expression is a triangle with base X and height $f(X) = 2X$. Thus, it has area $\frac{1}{2} \cdot X \cdot 2X = X^2$, so $F(X) = X^2$.

[17] It is standard practice to use the upper case counterpart for this function.

Observe that in our example, $f(X) = 2X$ is the derivative of $F(X) = X^2$. The First Fundamental Theorem of Calculus tells us that this was not a coincidence.

Theorem 7.1 (First Fundamental Theorem of Calculus)**.** Given a continuous function $f(X)$, define the function $F(X)$ by

$$F(X) = \int_a^X f(t)\ dt$$

for some constant a. Then, $F'(X) = f(X)$.

For this reason, we refer to $F(X)$ as the **antiderivative** of $f(X)$.

In order to prove the theorem, we return to the limit definition of the derivative. We have

$$F'(X) = \lim_{h\to 0} \frac{F(X+h) - F(X)}{h} = \lim_{h\to 0} \frac{\int_a^{X+h} f(t)\ dt - \int_a^X f(t)\ dt}{h}$$

Here, we apply the second integral bound rule from Section 6.3 to rewrite the numerator of this last fraction as a single integral, simplifying the expression to

$$= \lim_{h\to 0} \frac{1}{h} \int_X^{X+h} f(t)\ dt.$$

We can represent this integral with a picture.

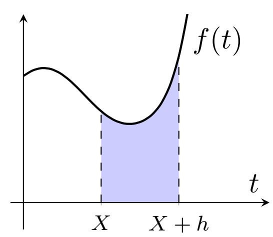


Figure 26: $\int_X^{X+h} f(t)\ dt$

We are dividing this integral, the area of the shaded region by h, its width. Therefore, we are left with a notion of the "average height" of the shaded region, or the average value of the function between X and $X+h$.

To complete our reasoning, we use the fact that $f(t)$ is continuous. This means that as t approaches X, $f(t)$ gets closer and closer to $f(X)$. Therefore, as h shrinks to 0, the function's values on all of the interval $[X, X+h]$ must get closer and closer to $f(X)$, meaning this average value does as well. In short $\lim_{h\to 0} \frac{1}{h} \int_X^{X+h} f(t)\ dt = f(X)$. We have shown that $F'(X) = f(X)$, verifying the theorem.

As noted above, this first Fundamental Theorem motivates the importance of antiderivatives. We also refer to antiderivatives as **indefinite integrals**, and express them using the notation

$$\int f(X)\ dX,$$

an integral without bounds.

In order to compute these indefinite integrals, we use the derivative rules that we have already learned, but in reverse [18]. As an example, we consider the "power rule for integration".

Example 7.2. Consider $\int X^n \, dX$ for some constant n. To compute this antiderivative, we ask ourselves, "What function has X^n as its derivative?" We know that the power rule for differentiation subtracts 1 from the exponent, so we should add 1 when taking the antiderivative. Also, differentiation multiplies by the higher exponent as a coefficient, so we should divide by it for anti-differentiation. Putting this together, we find that

$$\int X^n \, dX = \frac{1}{n+1} X^{n+1}.$$

You can verify that this is an antiderivative by taking its derivative and ensuring we get back to X^n.

For other functions, we perform a similar "reverse the rules" process.

Example 7.3. We have

$$\int \cos(X) \, dX = \sin(X)$$

because the derivative of $\sin(X)$ is $\cos(X)$. An antiderivative of e^X is

$$\int e^X \, dX = e^X$$

because the derivative of e^X is e^X. An antiderivative of $\frac{1}{X}$ is

$$\int \frac{1}{X} \, dX = \ln |X|.$$

Note that we introduce the absolute value in order to handle the case where $X < 0$ (since the log of a negative number is undefined).

Note that since addition and scalar multiplication were preserved by differentiation, they will also be preserved by anti-differentiation, so:

$$\int f(X) + g(X) \, dX = \int f(X) \, dX + \int g(X) \, d(X)$$
$$\int c \cdot f(X) \, dX = c \cdot \int f(X) \, dX.$$

Example 7.4. We compute an antiderivative

$$\int 8x^3 - 9x^2 \, dX.$$

Since anti-differentiation preserves addition and scalar multiplication, we can rewrite this as

$$\begin{aligned}\int 8x^3 - 9x^2 \, dX &= 8 \cdot \int x^3 \, dX - 9 \int x^2 \, dX \\ &= 8 \cdot \frac{1}{4} X^4 - 9 \cdot \frac{1}{3} X^3 \\ &= 2x^4 - 3x^3.\end{aligned}$$

[18] There are more advanced methods for computing complicated antiderivatives such as **u-substitutions** and **integration by parts**, but we will not consider these.

Finally, we note that a function $f(X)$ does not have a unique antiderivative, but rather has infinitely many antiderivatives $F(X)$. This infinite collection results from our different choices of base-points a, which result in functions $F(X)$ that differ by an additive constant (equal to the area under the curve between these different base-points). If $F(X)$ is an antiderivative of $f(X)$ (that is $F'(X) = f(X)$, note that $F(X) + C$ (for any constant C) is also an antiderivative of $f(X)$ because

$$\frac{d}{dX}\Big[F(X) + C\Big] = F'(X) + 0 = f(x).$$

We call this C a **constant of integration**, and we include it in our antiderivatives as a way of symbolizing the presence of this unknown additive constant, depending on our choice of base-point. Note that this constant of integration was omitted from the above examples to allow for a cleaner presentation of the rules. In actuality, this constant should always be included for indefinite integrals, such as in the example that follows.

Example 7.5.

$$\int 4X + \sin(X)\, dX = 4 \cdot \frac{1}{2}X^2 - \cos(X) + C = 2X^2 - \cos(X) + C.$$

Again, we can verify that this is correct by taking of the derivative of $2X^2 - \cos(X) + C$, which is $4X - (-\sin(X)) = 4X + \sin(X)$.

7.2 The Second Fundamental Theorem of Calculus

The second Fundamental Theorem is really a corollary (an immediate consequence) of the first Fundamental Theorem. It gives us the ability to exactly calculate the areas represented as integrals.

Theorem 7.2. Consider a function $f(X)$ and constant bounds $b \leq c$. Then,

$$\int_b^c f(X)\, dX = F(c) - F(b),$$

where $F(X)$ is the antiderivative of $f(X)$.

The easiest verification of this theorem is done with a picture. Suppose that a is the base-point used in defining $F(X)$, chosen so that $a < b$ [19]. Then, we have the following diagram.

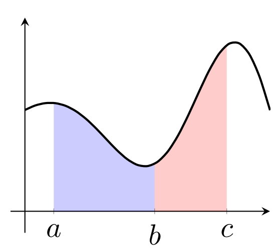


Figure 27: Visualizing the second Fundamental Theorem

[19] The generalized second integral bound rule tells us that this fact is true regardless of the location of a, but this choice produces a nicer picture.

The integral we are concerned with computes the area of the shaded red region. The value $F(c)$ is the area of the combined red and blue region and $F(b)$ is the area of blue region alone. The formula above says that in order to compute the area between b and c, one computes the area between a and c and then subtracts the area between a and b, which exactly agrees with our picture. Note that the second Fundamental Theorem is really just a combination of the first Fundamental Theorem and our second integral bound rule.

As a piece of notation, we often write $F(c) - F(b)$ using the vertical bar notation

$$F(X)\Big|_{X=b}^{X=c}.$$

The second Fundamental Theorem provides us with a process to calculate **definite integrals** (integrals with bounds, whose value is a number, representing area under a curve), namely

1. Compute the antiderivative, $F(X)$, of the function, $f(X)$, inside of the integral.
2. Plug the lower and upper integral bounds into $F(X)$.
3. Subtract the lower bound calculation from the upper bound calculation.

Note that since we are subtracting two values of the antiderivative function, we can disregard the additive constant of integration, since it will cancel out. We illustrate this process in the following examples.

Example 7.6. We calculate $\int_2^4 X^3 - 3X^2 + 2\ dX$. The antiderivative of the function $f(X) = X^3 - 3X^2 + 2$ inside of this integral is $F(X) = \frac{1}{4}X^4 - X^3 + 2X$. Thus,

$$\begin{aligned}\int_2^4 X^3 - 3X^2 + 2\ dX &= \Big(\frac{1}{4}X^4 - X^3 + 2X\Big|_{X=2}^{X=4} \\ &= \Big(\frac{4^4}{4} - 4^3 + 2(4)\Big) - \Big(\frac{2^4}{4} - 2^3 + 2(2)\Big) \\ &= (64 - 64 + 8) - (4 - 8 + 4) \\ &= 8\end{aligned}$$

Example 7.7. Just as in Example 6.1, we consider $\int_0^{2\pi} \sin(X)\ dX$, quantitatively this time. The antiderivative of $\sin(X)$ is $-\cos(X)$, so

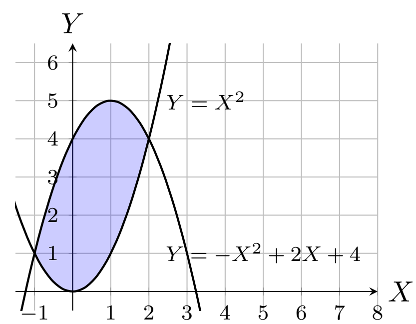
$$\int_0^{2\pi} \sin(X)\ dX = \Big(-\cos(X)\Big|_{X=0}^{X=2\pi} = -\cos(2\pi) - (-\cos(0)) = -1 + 1 = 0,$$

confirming that the positive area under $Y = \sin(X)$ between 0 and π cancels with the negative area over $Y = \sin(X)$ between π and 2π. The area of each of these pieces is

$$\int_0^{\pi} \sin(X)\ dX = \Big(-\cos(X)\Big|_{X=0}^{X=\pi} = -\cos(\pi) + \cos(0) = 1 + 1 = 2.$$

Example 7.8. As a final example, we calculate the area of the shaded region below. The farthest left that the region gets is $X = -1$, and the farthest right is $X = 2$. In this interval, the top of the shaded region is the function $u(X) = -X^2 + 2X + 4$ and the bottom is the function $\ell(X) = X^2$. Therefore, we can express the area as $\int_{-1}^{2} u(X) - \ell(X)\, dX = \int_{-1}^{2} -2X^2 + 2X + 4\, dX$. Evaluating this integral, we have,

$$
\begin{aligned}
&\int_{-1}^{2} -2X^2 + 2X + 4\, dX \\
&= \left(\frac{-2}{3}X^3 + X^2 + 4X\right|_{X=-1}^{X=2} \\
&= \left(\frac{-2}{3}(2)^3 + 2^2 + 4(2)\right) - \left(\frac{-2}{3}(-1)^3 + (-1)^2 + 4(-1)\right) \\
&= \left(\frac{-16}{3} + 4 + 8\right) - \left(\frac{2}{3} + 1 - 4\right) \\
&= 9.
\end{aligned}
$$



8 Separable Differential Equations

We wrap up this chapter with a brief introduction to **differential equations**. We are already familiar with differential equations; these are the change equations that we have studied since the start of the semester. While previously we studied these equations qualitatively, looking at the general shape of their trajectories to describe their long term behavior, here we take a more quantitative approach. We will use the calculus techniques that we have learned in order to solve explicitly for the equations of system trajectories.

We start by introducing some basic terminology and a technique for solving a specific subclass of differential equations. Then, we use this technique to examine exponential growth and decay equations. Finally, we solve for a closed-form representation of the trajectories in the Lotka-Volterra system.

8.1 What is a Differential Equation?

As we mentioned above, differential equations are the familiar change equations that we have studied since the start of the semester. More formally, a differential equation is an equation involving a function, and its derivatives [20]. The **general solution** of a differential equation is the collection of all functions that satisfy the equation. More concretely, this is the set of possible trajectories of the system. An **initial condition**, a known point through which the trajectory must pass, allows us to specify the general solution to a **particular solution** [21].

Example 8.1. Consider the differential equation

$$X \cdot \frac{dY}{dX} = 2Y.$$

A solution to this differential equation is a function $Y = f(X)$ satisfying the above equation. We can verify that the general solution to this equation has the form

$$Y = AX^2,$$

for some constant A. That is, setting A to any constant gives a valid solution to the equation. To verify this, note that $\frac{dY}{dX} = 2AX$, so

$$X \cdot \frac{dY}{dX} = X \cdot 2AX = 2AX^2 = 2Y.$$

Suppose that we care about the particular solution passing through $(1, 4)$. Plugging $X = 1$ and $Y = 4$ into $Y = AX^2$, we find that $A = 4$, so the particular solution is $Y = 4X^2$.

Here, we "solved" this differential equation using a guess-and-check strategy. We guessed that $Y = AX^2$ was the general solution, and plugged it into the differential equation to verify that it is correct. Next, we introduce a technique to solve for such a general solution without guessing.

[20] Here, we use the word "derivatives" because the differential equation may involve higher-order derivatives (e.g., the second derivative, the derivative of the derivative) or partial derivatives (derivatives of multi-variate functions). We do not concern ourselves with these more general cases here, and instead focus only on **ordinary first-order differential equations**.

[21] Here, we use the fact that each point in the state space has a unique trajectory passing through it.

8.2 Solving Separable Differential Equations

While there is a vast array of different techniques for solving differential equation of various forms, we focus our attention on one specific class, the **separable differential equations**. A differential equation is separable if it can be rearranged so that each side of the equation contains only one variable. The differential equation from the example above is separable. This is because we can separate the differential, moving the "dX" to the right side and dividing both sides by XY to get

$$X \cdot \frac{dY}{dX} = 2Y \quad \Longrightarrow \quad X\ dY = 2Y\ dX \quad \Longrightarrow \quad \frac{1}{Y}\ dY = \frac{2}{X}\ dX.$$

Next, we integrate both sides, making sure to add a constant of integration on one side [22]. We get

$$\int \frac{1}{Y}\ dY = \int \frac{2}{X}\ dX \quad \Longrightarrow \quad \ln|Y| = 2\ln|X| + C.$$

From here (when possible), we rearrange the equation to solve for Y as a function of X. In this case,

$$\begin{aligned} \ln|Y| &= 2\ln|X| + C \\ \ln|Y| &= \ln(X^2) + C && \text{(log properties)} \\ Y &= e^C \cdot X^2 && \text{(exponentiate both sides)} \\ Y &= AX^2 && (A := e^C) \end{aligned}$$

Note here, that the additive constant of integration became a multiplicative constant after exponentiation. For this reason, it is important to introduce the constant during the integration step, and not at the end. Note also that we were able to drop the absolute values in this case because X^2 and e^C are both guaranteed to be non-negative.

From this point, we solve for any particular solution by plugging in the initial equation to the general solution and solving for the constant, just as we did in Example 8.1.

To summarize the "separate and integrate" process for determining the particular solution for a separable differential equation:

1. Manipulate the equation so that each side of the equation includes only one variable. This includes separating the derivatives into their constituent **differentials**.
2. Integrate each side with respect to its variable of integration. Introduce a constant of integration on one side of the equation.
3. When possible, manipulate the equation to solve explicitly for one variable as a function of the other.
4. Plug the initial condition into the general solution to determine the value of the arbitrary constant. Substitute this value into the general solution to find the particular solution.

Sometimes, it may be easier to solve for the constant of integration before any manipulations. In these cases, steps 3 and 4 can be interchanged.

[22] It is unnecessary to have a constant of integration on both sides, because we could subtract one of them to the other side and combine it with the other constant.

8.3 Exponential Growth and Decay

One specific class of separable differential equations are exponential growth and decay equations. These equations are of the form

$$X' = rX$$

for some constant r. This is the "simple population model" that we saw at the beginning of the semester. We assume that X represents a positive quantity (the number of individuals in a population), so that when $r > 0$, X' is positive, signifying a growth in the quantity, and when $r < 0$, X' is negative, signifying a decay in the quantity. Notice that this equation captures the *per capita* rates of change that we are familiar with. We know that a population with a *per capita* birth rate will grow exponentially. With our new technique for solving differential equations, we can verify this quantitatively.

Example 8.2. We model the growth of an insect population, X, by the differential equation $X' = 0.05X$, with time measured in days. If the initial insect population was 20, we use the technique for solving separable differential equations to solve for the insect population after 1 week.

First, we separate and integrate the differential equation.

$$\begin{aligned}
\frac{dX}{dt} &= 0.05X \\
\int \frac{dX}{X} &= \int 0.05dt \\
\ln|X| &= 0.05t + C \\
X &= e^{0.05t+C} && \text{(exponentiate both sides)} \\
X &= Ae^{0.05t} && (A := e^C)
\end{aligned}$$

Now, we plug in the initial condition $X(0) = 20$.

$$20 = Ae^{0.05(0)} = Ae^0 = A,$$

so the general solution is $X = 20e^{0.05t}$. Assuming t is measured in days, after 1 week, we expect the insect population to be $20e^{0.35} \approx 28.38$.

While the model gives a reasonable answer in this case, we know that exponential growth is unreasonable because it induces a rapidly increasing positive feedback loop on the system. In fact, this exponential model predicts that after 3 years, the insect population will have grown to $20e^{0.05\cdot 1095} = 20e^{54.75} \approx 6 \cdot 10^{23}$, approximately the same number as there are stars in the universe!

Following the example, we find that the solution to the differential equation $X' = rX$ with initial condition $X(0) = a$ is

$$X = ae^{rt}.$$

In practice, exponential decay is more realistic. In fact, it occurs in systems such as radioactive decay. Radioisotopes have a half-life, the expected amount of time that it will take for half of the mass of the isotope to break down. Since the amount of isotope is decreasing by a multiplicative factor (one half) in a fixed amount of time, we can express this as a *per capita* decay equation, with an exponentially-decaying solution. We will explore the phenomenon of exponential decay more in discussion.

8.4 Differential Equations and the Lotka-Volterra Model

To conclude this section, we take another look at the Lotka-Volterra equations from a quantitative perspective. Recall that these equations are

$$\begin{aligned} X' &= rX - \beta XY \\ Y' &= m\beta XY - sY, \end{aligned}$$

where X represent the population of the prey species and Y represents the population of the predator species. In order to use our "separate and integrate" procedure, we will need to have 1 equation in 2 variables. Here, we have 2 equations in 3 variables (X, Y and time t). To eliminate the time variable, and to combine the change equations, we turn to the chain rule. Recall that the Leibniz representation of the chain rule is

$$\frac{dY}{dt} = \frac{dY}{dX} \cdot \frac{dX}{dt}.$$

Rearranging this by moving $\frac{dX}{dt}$ to the left side, we get

$$\frac{dY}{dX} = \frac{\frac{dY}{dt}}{\frac{dX}{dt}} = \frac{Y'}{X'}.$$

Hence, we can divide the right hand sides of the Lotka-Volterra equations to obtain the differential equation

$$\frac{dY}{dX} = \frac{Y'}{X'} = \frac{m\beta XY - sY}{rX - \beta XY}.$$

Amazingly, this is a separable equation. Cross-multiplying this equation, we find that

$$\begin{aligned} rX - \beta XY \; dY &= m\beta XY - sY \; dX, \\ X(r - \beta Y) \; dY &= Y(m\beta X - s) \; dX, \\ \frac{r - \beta Y}{Y} \; dY &= \frac{m\beta X - s}{X} \; dX, \\ \frac{r}{Y} - \beta \; dY &= m\beta - \frac{s}{X} \; dX. \end{aligned}$$

Integrating this, we have

$$\begin{aligned} \int \frac{r}{Y} - \beta \; dY &= \int m\beta - \frac{s}{X} \; dX, \\ r\ln(Y) - \beta Y &= m\beta X - s\ln(X) + C. \end{aligned}$$

Note that we do not write the absolute values in these logarithms, because X and Y represent species populations and must be non-negative. In this case, we are unable to manipulate this equation to solve for Y in terms of X (or vice versa), so we must settle for this **implicit** representation of the solution.

For a specific example, we consider the Lotka-Volterra system with $r = \frac{2}{3}$, $\beta = \frac{4}{3}$, $m = \frac{3}{4}$, and $s = 1$. In this case, our general solution has the form

$$\frac{2}{3}\ln(Y) - \frac{4}{3}Y = X - \ln(X) + C,$$

which we can rewrite as

$$C = 3\ln(X) + 2\ln(Y) - 3X - 4Y.$$

With the help of software, we can produce plots of these implicit solution curves by determining a collection of (X, Y) points that satisfy this equation for various values of C.

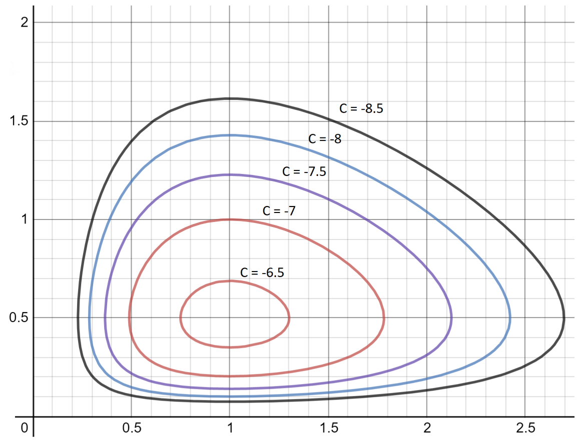


Figure 28: Trajectories of Lotka-Volterra System for Various Values of C

As seen in the figure above, the implicit solution curves to this differential equation form concentric closed loops, matching the qualitative behavior that we observed in the previous chapter. Similar, but slightly more advanced techniques can be used to find explicit solutions to the differential equations describing other systems that we have studied. The mass spring system can be reduced to a second-order differential equation, and can be shown to have a solution expressed in terms of sine and cosine. The logistic population model is also a separable differential equation, but the integration step requires a partial fraction decomposition and a u-substitution, two more advanced integration techniques.