Calculus for Biological Scientists

Calculus for Biological Scientists

Jeff Shriner
Colorado State University

May 9, 2024

Acknowledgements

The general style of this text was motivated by the Active Calculus[1] suite by Dr. Matthew Boelkins and contributing authors. Considerable content from Active Calculus - Single Variable[2] was used and modified for the purposes of serving life science majors. Chapter 1, section 3.2, section 3.3, section 3.7, and section 4.1 were created by the current author. All other sections contain content which originated from *Active Calculus - Single Variable*, and was modified to fit the needs of this text. I am grateful to all of the contributors of the Active Calculus suite for allowing me to use the fantastic resources that have been developed through years of hard work.

The current .html version of the text is possible only because of the amazing work of Rob Beezer and his development of the original Mathbook XML, now known as PreTeXt[3].

Finally, I am grateful to the many graduate students, instructors, and students that I have worked with in teaching the course for which this text was developed. The examples, exercises, and overall content layout is better because of their collective engagement and feedback with the material.

I welcome user feedback[4] to correct errors and collect suggestions for improvements.

This text was created with grant support from Colorado State University Libraries and the Colorado Department of Higher Education.

[1] activecalculus.org
[2] activecalculus.org/single/book-1.html
[3] pretextbook.org
[4] docs.google.com/forms

Our Goals

As in *Active Calculus*, our general goal is to provide a free resource that actively engages the reader in learning mathematical concepts in the context of life sciences. In the text we aim to stress the *process* of mathematics by using activities and interactives to encourage deepened understanding over memorization of facts. Towards these general goals, the text was designed to

- Provide a guide for reading a scientific text. We ask motivating, big picture questions before each section, and encourage the reader to answer them on their own at the end of each section before reading a summary.
- Provide opportunities for the reader to actively engage with concepts independently and with others. We present a Warm-Up activity at the beginning of each section for the reader to ponder, which contains previous concepts in a way that will be relevant to the current section they are reading. We also present activities in each section which are meant to be explored in groups, and encourage the reader to deepen their understanding of the concepts being presented, as well as strengthen their oral and written communication in a technical setting.
- Provide tools for using technology to strengthen understanding and build intuition. We embed Desmos interactives that guide the reader in visualizing concepts and computations, and give an illustration of how technology can be used effectively to assist in learning something new.

Students: Read this!

This book is different.

The text is available in HTML and PDF, both of which are free. If you are going to use the book electronically, the best mode is the HTML version. The HTML version looks great in any browser, including on a smartphone, and the links are much easier to navigate in HTML than in PDF. It also gives you access to the interactive elements of the text. Some particular direct suggestions about using the HTML follow among the next few paragraphs. It is also wise to download and save the PDF, since you can use the PDF offline, while the HTML version requires an internet connection.

This book is intended to be read sequentially and engaged with, much more than to be used as a lookup reference. For example, each section begins with a short introduction and a Warm-Up; you should read the short introduction and complete the Warm-Up prior to class, even if your instructor does not require you to do so. Most Warm-Up activities can be completed in 10-20 minutes and are intended to be accessible based on the understanding you have from preceding sections. There are not answers provided to Warm-Up activities, as these are designed simply to get you thinking about ideas that will be helpful in work on upcoming new material. There is great benefit in thinking about these questions before class, even if you are unsure about the correctness of your answers.

As you use the book, think of it as a workbook, not a *worked-book*. There is a great deal of scholarship that shows people learn better when they actively engage and struggle with ideas themselves, rather than passively watch others. Thus, instead of reading worked examples or watching an instructor complete examples, you will engage with Activities that prompt you to grapple with concepts and develop deep understanding. You should expect to spend time in class working with peers on Activities and getting feedback from them and from your instructor. Your goal should be to do all of the activities in the relevant sections of the text and keep a careful record of your work. Answers to the activities are not provided in the text, though you can get answers by asking questions in class or outside of class (such as in office hours).

Each section concludes with a Summary, which re-visits the Motivating Questions from the beginning of the section. After you have read the section and worked the Activities, you should attempt to answer these questions in your own words first. You can then expand answers to check your understanding. This is a good place to find a short list of key ideas that are most essential to take from the section.

At the end of each section, you'll find written Exercises. These are designed to encourage you to connect ideas, investigate new situations, and write about your understanding. You are encouraged to work these exercises with others, focusing on your process rather than a final answer. Answers to selected exercises are provided in Appendix A. Worked solutions are not provided, so you are encouraged to discuss your thought process and reasoning for exercises with others, even if your final answer matches what is listed in the back. Your process and reasoning are the most important things for deepening your understanding and ensuring you can use important concepts on your own in the future, so use the answers as a guide, but not a replacement for engaging in the exercises with others.

The best way to be successful in mathematics generally and calculus specifically is to strive to make sense of the main ideas. We make sense of ideas by asking questions, interacting with others, attempting to solve problems, making mistakes, revising attempts, and writing and speaking about our understanding. This text has been designed to help you make sense of calculus in the context of the life sciences. We wish you the best as you undertake the large and challenging task of doing so!

Instructors: Read this!

This text contains most of the content you would expect from a typical first semester calculus course, though may be lighter in some areas (for example, there are no sections for implicit differentiation or related rates). It also contains content that you would not expect from a typical first semester calculus course. Chapter 1 contains pre-calculus content as well as an introduction to discrete-time dynamical systems, and chapter 3 contains some applications of derivatives in analyzing discrete-time dynamical systems. We approach Chapter 4 from the perspective of analyzing continuous-time dynamical systems (differential equations), and cover most of the standard first semester concepts in integration while doing so.

Among the two formats (HTML, PDF), the HTML is optimal for display in class if you have a suitable projector. The HTML is also best for navigation, as links to internal and external references are much more obvious, and Desmos interactives can be used directly from the HTML version. We recommend saving a downloaded version of the PDF format as a backup in the event you don't have internet access.

The text is written so that each section corresponds to one to two hours of class meeting time. A typical instructional sequence when starting a new section might look like the following:

- Students complete a Warm-Up in advance of class. Class begins with a short debrief among peers followed by all class discussion. (5-10 minutes)
- Brief lecture and discussion to build on the Warm-Up and set the stage for the next activity. (5-10 minutes)
- Students engage with peers to work on and discuss the first activity in the section. (10-15 minutes)
- Brief discussion and possibly lecture to reach closure on the preceding activity, followed by transition to new ideas. (Varies, but 5-15 minutes)
- Possibly repeat with the next activity, and summarize section at the very end.

If the section was not completed, the next hour of class would be similar, but without the Warm-Up at the beginning of class.

There is a suite of Desmos activities[5] that can be used (or copied and modified) to accompany the text. They were created to be used for a final hands-on experience of important topics during a full class period after the relevant content has already been covered in class.

[5] `teacher.desmos.com/collection/5f5050fa1aec2537b2ab3543`

The Exercises in the text are meant to be conceptual problems for students to engage with in writing and orally. We recommend also using an online homework system (such as WeBWorK) to provide students more procedural practice in which they can receive immediate feedback.

The PreTeXt source code for the text is shared at GitHub[6]. If you find errors in the text or have other suggestions, use the feedback link[7] in the HTML version (found at the bottom left), or email the author directly.

[6] github.com
[7] docs.google.com/forms

Contents

Chapter 1

Functions As Models

1.1 Biology and Calculus

Motivating Questions

- What is a mathematical model and why do we need them?
- What are some common mathematical models?
- What are some mathematical processes that can help us become better problem solvers?

A central theme in science is to study systems and how they *change*. Calculus can be viewed broadly as the study of change, and is therefore a useful tool to understand as a scientist. In this section we'll begin to gain an understanding of the relationship between biology and mathematics. We begin with a warm-up describing a simple situation you may need to explore in the future.

Warm-Up 1.1.1 You are in the field measuring the height of a bine each day under certain conditions. You organize your measurements in the following table:

Table 1.1.1

t **(days)**	h **(feet)**
0	1
1	2.90
2	4.61
3	6.15
4	7.53
5	8.78
6	9.90
7	10.91
8	11.82
9	12.64
10	13.38

1. Organize your data as a graph.
2. What do you think the height of the bine will be after 11 days?

1.1.1 Mathematical Models

If you compare your answer to Warm-Up 1.1.1 with other students, you will likely find that your estimation of the bine height after 11 days is different. You may have used slightly different reasoning to come up with your final answer.

Mathematical modeling is a way to convert data that we measure and observe into something that we can analyze and make predictions based on.

Mathematical Model.

A *mathematical model* is a mathematical representation that describes a system we have observed and measured.

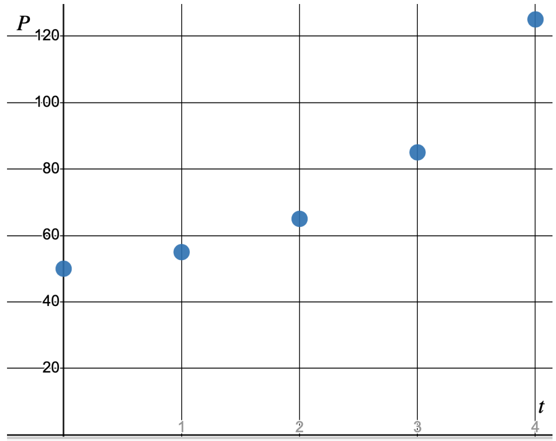
The process of developing and using mathematical models often involves the following steps:

1. *Observe* and *measure* a system you are interested in.
2. *Describe* patterns in your observations and measurements.
3. *Model* the system with a mathematical representation that shares the patterns of the system.
4. *Analyze* the system and *make predictions* about the system using the mathematical model.

The focus of our text will be on analyzing and making predictions about how systems change given a mathematical model. The mathematical models we will use the most and will explore in much more detail are summarized below:

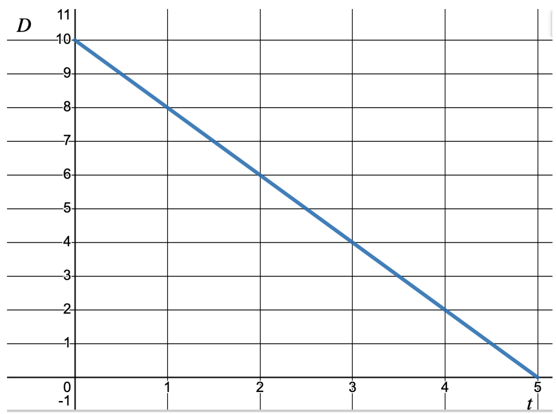
- A **function** describes a specific type of relationships between different quantities. It can be expressed using multiple representations. Most models involve a function relationship.
- A **discrete-time dynamical system (DTDS)** describes a sequence of measurements made at equally spaced intervals. Functions are an important component of these systems.
- A **continuous-time dynamical system (CTDS)** describes measurements taken over an entire time interval. Functions are an important component of these systems.

Example 1.1.2 Discrete-Time Dynamical System. Each year t, a population P of wolves is measured. A model which describes the population at a yearly interval is a discrete-time dynamical system. The graph of the relationship between population and time of such a model might look something like this:



□

Example 1.1.3 Continuous-Time Dynamical System. As time t passes in seconds, my distance D from the door is measured. A model which describes my distance over the entire continuous time interval is a continuous-time dynamical system. The graph of the relationship between distance and time of such a model might look something like this:



□

In Example 1.1.2 and Example 1.1.3, we described the relationships explicitly; that is, we showed what a population value or a distance value would be for a given time value. In practice, this is not typically the representation we begin with for these models. For a DTDS, we give a first example below in Example 1.1.4, and will explore in much more detail in Section 1.7. The common representation and analysis of a CTDS is the main topic of Chapter 4.

Example 1.1.4 Updating Function and Initial Value. A strawberry plant begins as one plant. Each year, the plant produces 2 daughter plants. We can describe the growth of this system using an **initial value** (p_0) and an **updating function**, which uses the current population value (p_t) to describe what the next population value (p_{t+1}) will be:

$$p_0 = 1$$
$$p_{t+1} = p_t + 2$$

Using the initial value and the updating function, we can determine the population of strawberry plants (p_t) at a given year t:

Table 1.1.5

t	p_t
0	$p_0 = 1$
1	$p_1 = p_0 + 2 = 3$
2	$p_2 = p_1 + 2 = 5$
3	$p_3 = p_2 + 2 = 7$

□

1.1.2 Mathematical Processes

Along with mathematical content, our text will also emphasize *mathematical processes* that enhance problem solving skills. The processes we will utilize most are:

- *Creating examples and counterexamples.* We are likely used to watching an example from someone else to try and understand a concept, but the process of *creating* your own example or counterexample can be even more illuminating. In the process of constructing examples and counterexamples, you must think deeply about important aspects of a concept, and typically get repetition in applying relevant procedural tools. This process can be frustrating, as it does not usually happen quickly, but it can be a vital step in learning something well, and is a tools mathematicians use every day.

- *Utilizing technology.* Technology can be an extremely useful tool for gaining intuition about a concept and for verifying conclusions. We'll practice using technology in our study of Calculus to reinforce and enhance our conceptual understanding, not to replace it. The main tool we will use is Desmos[1].

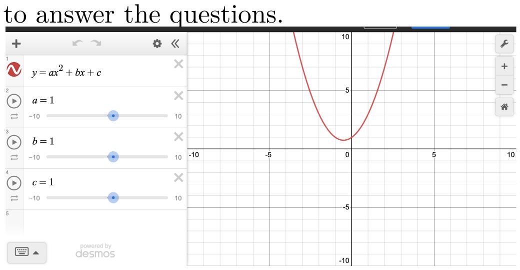
Activity 1.1.2 Creating Examples. Consider a population that changes according to the updating function

$$b_{t+1} = 0.5b_t + 1.$$

1. Find an example of an initial value b_0 for which the population increases over time.

2. Find an example of an initial value b_0 for which the population decreases over time.

Activity 1.1.3 Utilizing Technology. The following activity explores an algebra topic to practice utilizing technology to gain understanding of a topic. The general equation of a quadratic function is $ax^2 + bx + c$. Use the interactive below to answer the questions.



www.desmos.com/calculator/xdluusffkw

1. Which of a, b, or c determines whether the graph opens up or down?

[1]Desmos.com

2. What types of values make the graph open down?

1.1.3 Summary

Below we re-visit the Motivating Questions from the beginning of the section. It is good practice to attempt to summarize in your own words before viewing the answers.

- **Question 1.1.6** What is a mathematical model and why do we need them? □

  A mathematical model is a mathematical representation that describes a system we have observed and measured. We can manipulate mathematical models in order to analyze and predict how a system is changing.

- **Question 1.1.7** What are some common mathematical models? □

  Common mathematical models are functions, discrete-time dynamical systems, and continuous-time dynamical systems.

- **Question 1.1.8** What are some mathematical processes that can help us become better problem solvers? □

  Processes that can help us become better problem solvers are creating examples and counterexamples, and learning how to utilize technology in a way that reinforces and enhances our understanding of a concept.

1.1.4 Exercises

1. Use the Desmos graphing calculator[2] to plot the points given in the table of Warm-Up 1.1.1. Then graph $y = -19 \cdot 0.9^x + 20$. What do you notice? Use this graph to predict the height of the bine after 11 days.

2. The representations of models that we will practice working with in this text are motivated by the models that are used to analyze real systems.

arXiv and bioRxiv. The papers in Exercise 1.1.4.2 come from arXiv[3] and bioRxiv[4], which are examples of open-access archives of scholarly articles in many disciplines.

Articles posted here have not been peer-reviewed, though they are pre-prints of articles that typically get submitted to peer-reviewed journals. These open-access archives are a good way for authors to make their research available to the public quickly, and for the scientific community to remain current on what research is happening in their field.

(a) Browse the following papers which describe research in the field of biology. Focus on the models being used and how they are represented.

- Dynamical behavior of colony migration system[5]
- Discovering dynamical models of human behavior[6]

(b) Reflect on your observations by considering the following questions:

- What types of models do you see?
- What do you notice about the use of numbers in the models?
- What are some questions that you have about how symbols are being used in the models? The way symbols are used to

[2]`Desmos.com/calculator`

create meaning is called **notation**, and we will address many questions regarding notation as we continue our development and analysis of mathematical models.

1.2 Functions

Motivating Questions

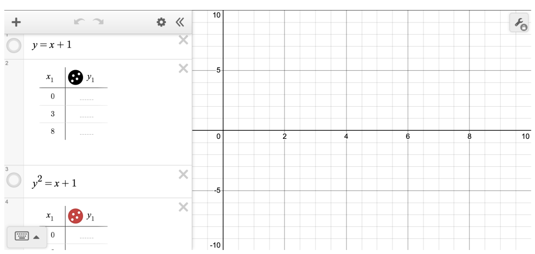
- What makes a relationship a function?
- What are different ways we can describe a function?
- What does it mean when there are letters in a model?
- What are different ways that we can combine functions?
- What does it mean for a function to be invertible?
- How can we describe how a function changes over time?

Functions are a core concept in the study of calculus. In this section we outline common representations of functions and some general properties that we will use throughout future sections.

Warm-Up 1.2.1 Consider the following equations:

a. $y = x + 1$ b. $y^2 = x + 1$

For each equation above, fill in the following tables:



www.desmos.com/calculator/do5rormmpj

1.2.1 Functions and their Representations

Understanding systems is all about being able to describe the *relationship* between the quantities involved. The population of a species is related to time, the energy an animal exerts when running is related to its speed, and the growth of a plant is related to its exposure to sunlight. Equations, tables, and graphs are all mathematical objects that help us represent relationships between quantities, and it is beneficial to be able to represent a single relationship in multiple different ways.

The systems we are typically interested in are a special type of relationship called a *function*:

Definition 1.2.1 Function. Given two related quantities x and y, we say y **is a function of** x if every x value in the relationship is related to exactly one y value. ♢

[3] arxiv.org
[4] biorxiv.org
[5] https://arxiv.org/pdf/2206.07016.pdf
[6] https://www.biorxiv.org/content/10.1101/2022.03.20.484666v1.full

Using this terminology with our previous illustrations, we would say a population is a function of time, energy is a function of speed, and the height of a plant is a function of sunlight exposure. Notice that there is an ordering implied when we talk about function relationships. When we say "y is a function of x", we call x the **input** (or **independent variable**) and y the **output** (or **dependent variable**).

Example 1.2.2 Looking at Warm-Up 1.2.1, one equation represents y as a function of x and one does not.

a. In $y = x + 1$, y is a function of x. Our reasoning will sound different depending on whether you are looking at the equation, table, or graph, but in all cases our reasoning is based on Definition 1.2.1:

   - *Equation:* For an x value, the corresponding y value is obtained by adding one. This gives a unique y value for every x value.
   - *Table:* Every x value in the table is listed exactly once. This means each x value in the table is associated with exactly one y value.

     Warning: When relationships are represented via tables, it is possible you have limited information about the relationship. Unless you are told that the table shows every x, y pair in the relationship, you should assume there may be other x, y pairs in the relationship not listed in the table.
   - *Graph:* If you pick an x value and draw a vertical line at that x value, you will intersect the graph exactly once. This means each x value corresponds to exactly one y value. This is known as the **vertical line test** to determine if the graph of a relationship represents a function relationship.

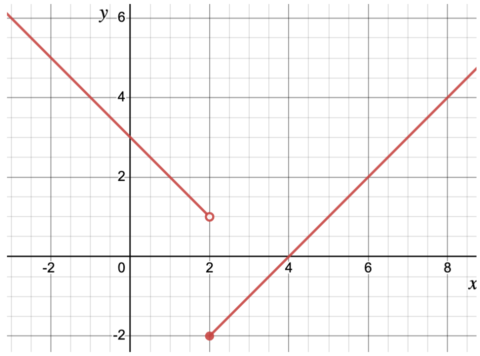
b. In $y^2 = x + 1$, y is *not* a function of x. Our reasoning will sound different depending on whether you are looking at the equation, table, or graph, but in all cases our reasoning is based on Definition 1.2.1:

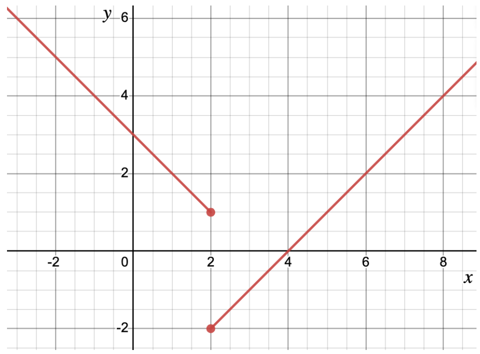
   - *Equation:* For an x value, the corresponding y value is obtained by the rule $\pm\sqrt{x+1}$. This means that when $x = 0$, for example, the associated y values are ± 1. This means there is an x value associated to more than one y value, and so this relationship is not a function.
   - *Table:* There are x values in the table listed more than once, and associated with different y values. This means there are x values in the relationship associated with more than one y value.
   - *Graph:* You can pick an x value and draw a vertical line at that x value such that you intersect the graph more than once. This means you have found an x value which corresponds to more than one y value. This is known as the **vertical line test** to determine if the graph of a relationship represents a function relationship.

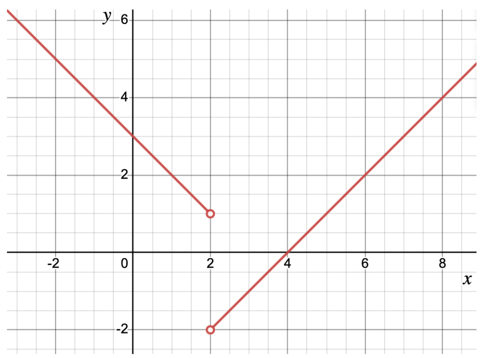
□

Activity 1.2.2 Function Relationships. Practice using Definition 1.2.1 by answering each question below. Focus on verbalizing your thought process. Your reasoning is more important than your final answer.

1. In the equation $y^2 = x + 1$, is x a function of y?
2. Is student height a function of student name?
3. For each graph below, is y a function of x?

- 

- 

- 

The graphs in Activity 1.2.2 are very similar, but they all represent different relationships. Of those that are functions, they differ in their domain and range, which we define below:

Definition 1.2.3 Domain and Range.

- The **domain of a function** is the set of all possible input values.
- The **range of a function** is the set of all possible output values.

◊

Remark 1.2.4 The domain of a function is typically determined by one or more of the following:

- *Operations:* There may be operations used to define a function which are undefined for specific inputs. For example, in $y = \frac{1}{x}$, $x = 0$ is not in the domain because the operation of division by 0 is not defined.
- *Context:* When a function is used as a model, there may be some input values that don't make sense for what the function represents. For example, if height is a function of time given by $h = t^2$, negative t values are not in the domain even though operationally the function is defined for negative t values.

1.2.2 Explicit and Recursive Functions

When given a function relationship, there are two main ways we might express that relationship:

1. **Explicit Function**: An explicit function describes the output (or dependent variable) explicitly in terms of the input (or independent variable). For example, $y = x + 1$ is an explicit function. However, moving forward we will most often write explicit functions using **function notation**, which means we write the output (dependent variable) as $f(x)$ instead of y. The function $y = x + 1$ is written using function notation as $f(x) = x + 1$. Note that we may name the function whatever we'd like (it does not have to be "f"). For example, if the output was "population" and the input was "time", we might call our function $p(t)$.

2. **Recursive Function**: A recursive function is typically used when modeling discrete systems, and describes the next output (or dependent variable) in terms of the current output (or dependent variable). For example, $p_{t+1} = p_t + 2$ is a recursive function. In words, this function says "the next output is equal to the current output plus two." Notice that this relationship does have an independent variable t, it is just not used in describing the relationship like in an explicit function.

1.2.3 Variables and Parameters

As illustrated in Exercise 1.1.4.2, modeling real systems typically involves representing quantities with letters and numbers. It is important that we know how to interpret the letters that we see in function representations.

Definition 1.2.5 Variable and Parameter.

- A **variable** represents a quantity that can change in a fixed model.
- A **parameter** represents a quantity that is constant in a fixed model.

◊

Example 1.2.6 Variables and Parameters. The function $h(t) = mt + b$ describes a relationship whose graph looks like a straight line. One benefit of using function notation is that it clearly indicates to us which letters represent variables. The notation "$h(t)$" tells us that t is the input (or independent variable), and $h(t)$ is the output (or dependent variable). This means that for a fixed line, the input value t can change, and $h(t)$ changes along with it.

Once we identify the independent and dependent variable, every other letter represents a parameter. In this example, m and b are parameters. This means that for a fixed line, m and b are constant values. □

One benefit of using parameters is that it allows us to describe many models at once, and we can then analyze how a specific system behaves based on its parameter values.

1.2.4 Combinations of Functions

Just as there are ways to combine two numbers to get a new number, we can also combine two functions to get a new function. Given two functions $g(x)$ and $k(x)$, we can define the following combinations:

Definition 1.2.7 Combining Functions.

- **Sum**: $(g + k)(x) = g(x) + k(x)$

- **Difference**: $(g - k)(x) = g(x) - k(x)$
- **Product**: $(g \cdot k)(x) = g(x) \cdot k(x)$
- **Quotient**: $\left(\frac{g}{k}\right)(x) = \frac{g(x)}{k(x)}$
- **Composition**: $(g \circ k)(x) = g(k(x))$

$\Diamond$

Example 1.2.8 Combining Functions. Let $f(x) = x^2+1$ and $g(x) = 2x-3$.

- $$\begin{aligned}(f+g)(x) &= f(x) + g(x)\\ &= (x^2+1) + (2x-3)\\ &= x^2+2x-2\end{aligned}$$

- [1]
$$\begin{aligned}(f-g)(x) &= f(x) - g(x)\\ &= (x^2+1) - (2x-3)\\ &= x^2-2x+4\end{aligned}$$

- $$\begin{aligned}(f\cdot g)(x) &= f(x)\cdot g(x)\\ &= (x^2+1)\cdot(2x-3)\\ &= 2x^3-3x^2+2x-3\end{aligned}$$

- $$\begin{aligned}\left(\frac{f}{g}\right)(x) &= \frac{f(x)}{g(x)}\\ &= \frac{(x^2+1)}{(2x-3)}\end{aligned}$$

- $$\begin{aligned}(f\circ g)(x) &= f(g(x))\\ &= g(x)^2+1\\ &= (2x-3)^2+1\\ &= 4x^2-12x+10\end{aligned}$$

- $$\begin{aligned}(g\circ f)(x) &= g(f(x))\\ &= 2f(x)-3\\ &= 2(x^2+1)-3\\ &= 2x^2-1\end{aligned}$$

□

Composition is a way to combine functions that we cannot do with numbers, and allows us to "chain" multiple relationships together.

Activity 1.2.3 Composition. Let $m(w)$ be the number of mosquitos entering a house if w windows are open, and $b(m)$ be the number of mosquito bites someone gets if there are m mosquitos in a house.

1. In this context, which composition makes sense: $b\circ m$ or $m\circ b$?

[2]Note the importance in using parentheses!

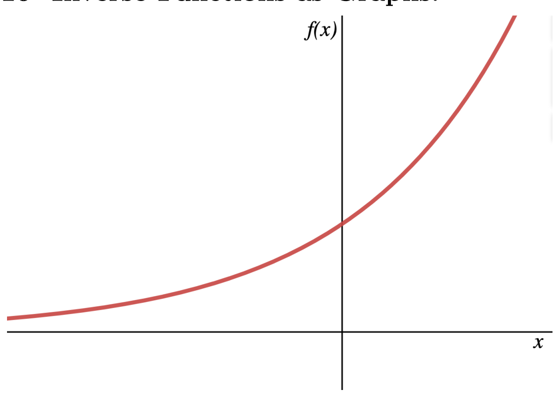
2. In words, what relationship does the composition that makes sense describe?

1.2.5 Inverse Functions

We noticed in Definition 1.2.1 that there is an order implied in a function relationship. Depending on the question you would like to answer, the ordering of the function may or may not be desirable. For example, if you had a model where population was a function of time, and you wanted to know when the population would reach $1,000$, this ordering would be desirable. However, if the population was decreasing and you wanted to know after how many years the population would become extinct, it may be desirable to have a model expressing time as a function of population. The property of being able to switch input and output and still have a function relationship is what we define below:

Definition 1.2.9 Inverse Function. Let $f(x) = y$ be a function of x. If x is also a function of y, we say that f is **invertible**. We denote x as a function of y as f^{-1}, and call it the **inverse function of** f. ♢

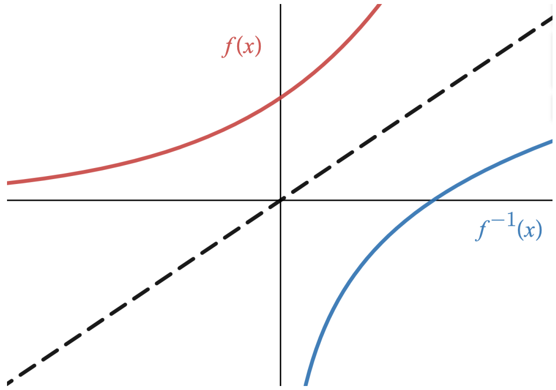
Example 1.2.10 Inverse Functions as Graphs.



To determine if this function has an inverse function, we must determine if the relationship obtained by swapping input and output is also a function. We can test this directly on the graph of f by remembering the vertical line test used in Example 1.2.2, and thinking about how vertical lines change when we swap input and output.

A vertical line in the x, y plane is a line whose points have a constant x value. For example, the points $(1, 0), (1, 1), (1, 2), (1, 3)$ all lie on the vertical line $x = 1$. When we swap input and output, those points become $(0, 1), (1, 1), (2, 1), (3, 1)$, which all lie on the *horizontal* line $y = 1$. So we can determine if the function f is invertible by drawing horizontal lines on the graph of f. If every horizontal line intersects the graph at most once, then f is invertible. Otherwise, the function f is not invertible.

This is known as the **horizontal line test** for testing invertibility. In this example we see the function is invertible, and we can visualize the graph of f^{-1} by reflecting the graph of f over the diagonal line $y = x$:



□

Example 1.2.11 Inverse Functions as Equations.

- Let $f(x) = 3x + 1$. If f has an inverse function, x would be the output and $f(x)$ would be the input. As an equation, this means we will try to re-write the function relationship as "$x =$". For notational purposes, before we start the computation we'll write the function as $y = 3x + 1$ instead of using function notation.

$$\begin{aligned} y &= 3x + 1 \\ y - 1 &= 3x \\ \frac{y-1}{3} &= x \\ x &= \frac{y-1}{3} \end{aligned}$$

  This shows us that f is invertible, and also what the rule is for the inverse function. Instead of using x as the output variable, we switch back to the standard notation and write f^{-1} as a function of x: $f(x) = 3x + 1$ and $f^{-1}(x) = \frac{x-1}{3}$.

- Let $g(x) = x^2 + 1$. We'll try the same process as above to determine if there is an inverse function, g^{-1}.

$$\begin{aligned} y &= x^2 + 1 \\ y - 1 &= x^2 \\ \pm\sqrt{y-1} &= x \\ x &= \pm\sqrt{y-1} \end{aligned}$$

  Note that x is not a function of y, and so our original function $g(x)$ is not invertible.

□

Invertible functions and their inverse functions have a special property under composition. As illustrated in Example 1.2.8, the order of composition matters. That is, for two functions $f(x)$ and $g(x)$, the compositions $f(g(x))$ and $g(f(x))$ are, in general, different. However, composition with an invertible function and its inverse function is an example of a special case when the order does not matter. Even more, the composition will always result in the same function:

Fact 1.2.12 *For an invertible function $f(x)$ and its inverse function $f^{-1}(x)$,*

$$f(f^{-1}(x)) = f^{-1}(f(x)) = x.$$

1.2.6 Average Rate of Change

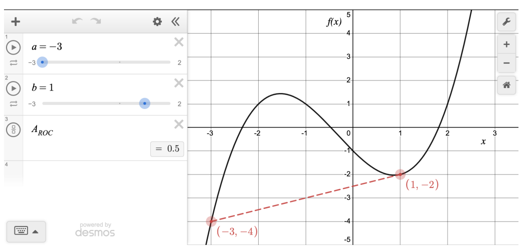
We mentioned in Section 1.1 that a main goal in mathematical modeling is to study how systems *change*. Given a function, one way we can describe how the function changes over time is by computing an *average rate of change*. This is also a foundational concept that will be used to extend our tools for describing change in Chapter 2.

Definition 1.2.13 Average Rate of Change. Given a function $f(x)$ and x values a and b, the **average rate of change between a and b** is the number

$$AROC_{[a,b]} = \frac{f(b) - f(a)}{b - a}.$$

◊

Being able to compute an average rate of change is necessary, but understanding what an average rate of change represents graphically will be even more important for our future developments in describing how functions change. Use the interactive provided to complete the statements below before viewing the answers.



www.desmos.com/calculator/lomycilmj5

1. The value of $AROC_{[-3,2]}$ is 1

2. The line connecting the points $(-3, f(-3))$ and $(2, f(2))$ (*increases/ decreases/stays the same*). increases

3. The value of $AROC_{[-1,1]}$ is -1.5

4. The line connecting the points $(-1, f(-1))$ and $(1, f(1))$ (*increases/ decreases/stays the same*). decreases

5. The value of $AROC_{[-2,2]}$ is 0

6. The line connecting the points $(-2, f(-2))$ and $(2, f(2))$ (*increases/ decreases/stays the same*). stays the same

7. Graphically, the value of $AROC_{[a,b]}$ describes whether the line connecting the two points $(a, f(a))$ and $(b, f(b))$ increases, decreases, or stays the same. Further, the larger the average rate of change is in absolute value, the steeper the line will increase or decrease. The line connecting the two points is called the **secant line**.

Activity 1.2.4 AROC.

1. Sketch the graph of a function f such that $AROC_{[0,2]} > 0$ and f is increasing on the interval $(0, 2)$.

2. Sketch the graph of a function f such that $AROC_{[0,2]} > 0$ and f increases and decreases on the interval $(0, 2)$.

3. Sketch the graph of a function f such that $AROC_{[0,2]} > 0$ and f is decreasing on the interval $(0, 2)$.

1.2.7 Summary

- **Question 1.2.14** What makes a relationship a function? □

  A quantity y is a function of a quantity x when each x value is associated with exactly one y value.

- **Question 1.2.15** What are different ways we can describe a function? □

  We can describe function relationships using equations, tables, and graphs. Function relationships can also be described explicitly or recursively.

- **Question 1.2.16** What does it mean when there are letters in a model? □

  A letter in function may represent a variable (a changing quantity) or a parameter (a constant quantity). Function notation is useful for helping us recognize the variables in a model.

- **Question 1.2.17** What are different ways that we can combine functions? □

  We can combine functions like numbers using addition, subtraction, multiplication, and division. We can also put functions inside of other functions using composition.

- **Question 1.2.18** What does it mean for a function to be invertible? □

  A function is invertible if swapping domain and range results in another function relationship.

- **Question 1.2.19** How can we describe how a function changes over time? □

  We can describe how a function changes over a specific time interval by computing the average rate of change over that time interval. Graphically, this number tells us the steepness of the secant line connecting the two end points, and if the secant line is increasing or decreasing.

1.2.8 Exercises

1. Give an example of a relationship in which y is not a function of x. Represent your example as an equation, table, and graph.

2. Give an example of a relationship in which y is a function of x but is not an invertible function. Represent your example as an equation, table, and graph.

3. Consider the equation $h = kt^2 - m^2t + cm$.

(a) Write the equation using function notation if t is the independent variable and h is the dependent variable. What are the parameters?

(b) Write the equation using function notation if m is the independent variable and h is the dependent variable. What are the parameters?

4.

(a) Sketch the graph of a single function $q(x)$ that meets all of the following criteria:

- The domain of $q(x)$ is $[-1, 0) \cup (0, 4]$.
  Note: This notation means the domain is between -1 and 4, not including 0.
- The range of $q(x)$ is $[0, 3]$.
- $q(-1) > q(4)$.

(b) Is your function from the previous part invertible? Explain why or why not.

5. The functions $g(x)$ and $h(x)$ are defined on the domain $(-\infty, \infty)$. Compute the following values given that

- $g(-1) = 2$ and $h(-1) = -10$, and
- $g(x)$ and $h(x)$ are inverse functions of each other (i.e., $g(x) = h^{-1}(x)$ and $h(x) = g^{-1}(x)$).

(a) $(g + h)(-1)$

(b) $(g - h)(-1)$

(c) $(g \cdot h)(-1)$

(d) $(h \circ g)(-10)$

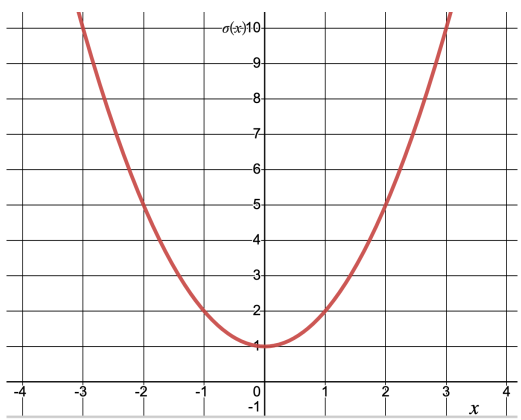
(e) $(g \circ h)(5)$

6. Let $\sigma(x) = x^2 + 1$. Calculate each average rate of change below. Then use the graph provided to illustrate what each calculation represents graphically.

(a) $AROC_{[0,3]}$

(b) $AROC_{[-2,2]}$

(c) $AROC_{[-3,1]}$



1.3 Units and Dimensions of Functions

Motivating Questions

- When using functions as models, how do we ensure our models are meaningful?
- When computing with functions that are being used as models, how do we ensure our computations are meaningful?
- How can we visualize unit conversions graphically?

Suppose you were told that $h(t) = \dfrac{10t+1}{t+2}$ represented the height h of a tree after time t, and were asked to use it to estimate the height of the tree after 5 years. Do you plug in 5 years for t, or 60 months? If you compute the value 8 for h, is that 8 meters, feet, or something else? In order to make functions usable as models, we must understand how variables are being measured, and how those variables are being combined to produce something meaningful.

Warm-Up 1.3.1 Separate the following words into groups, and explain why you separated them the way that you did.

meter, liter, second, cubic feet, mile, year

1.3.1 Units and Dimensions

In Warm-Up 1.3.1, there are multiple ways you may have separated the words into different groups. One way would be to group them based on *what* each is measuring, called a **dimension**. Within a single group (dimension), each word describes *how* we might measure that dimension, called a **unit**.

Example 1.3.1 Units and Dimensions. Volume is a dimension that may be measured in one of many different units, including m^3, ft^3, gallons, and liters.

Mass is a dimension that may be measured in one of many different units, including grams, kilograms, and milligrams. □

1.3.2 Conversion Factors and Fundamental Relations

There are situations when it is necessary to change the units in which a quantity is being measured in order to make a meaningful computation. To convert between different units, we use **conversion factors**, which are ratios that express the same quantity of a dimension in two different units. For example, the ratio $\dfrac{1000\mathrm{g}}{1\mathrm{kg}}$ is a conversion factor because $1000\mathrm{g} = 1\mathrm{kg}$.

We can multiply numerical expressions by conversion factors without changing the quantity being represented since the numerator and denominator are equal, it is the same as multiplying by 1. If we use conversion factors correctly, we can also get unwanted units to cancel and convert to new units.

Example 1.3.2 Conversion Factors. Let $p(t) = 3t$ be the mass of a bacterial colony after t minutes measured in grams, and $q(t) = t + 2$ be the mass of a separate bacterial colony after t minutes measured in kilograms. If we wanted to find the total mass of all bacterial colonies, we cannot just add the functions as they are written, since mass is being measured in different units for each colony. We could either convert $p(t)$ to units of kilograms, or $q(t)$ to units of grams. For this example, we will do the latter:

$$(t+2)\cancel{\text{kg}} \cdot \frac{1000\text{g}}{1\cancel{\text{kg}}} = 1000 \cdot (t+2)\text{g}$$

Notice that we had the choice to write the conversion factor as $\frac{1000\text{g}}{1\text{kg}}$ or $\frac{1\text{kg}}{1000\text{g}}$, and we made our choice so that the unwanted unit (kilograms) would cancel.

Now that $p(t)$ and $q(t)$ are both measured in grams, we may add them together to get the total mass:

$$p(t)\text{g} + q(t)\text{g} = (3t + 1000 \cdot (t+2))\text{g} = (1003t + 2000)\text{g}.$$

□

Activity 1.3.2 Let $V(t) = 2t$ measure the volume (in liters) of a balloon after t seconds, and $m(V) = 6 + V$ measure the mass (in grams) of the balloon when the volume is V milliliters. Use conversion factors to determine a meaningful expression for the function $m(V(t))$. What does this function represent?

Though different dimensions measure different things, it is possible that two different dimensions are related. For example, the radius of a circle is related to its area and the volume of a glass of water is related to its mass. The precise relationship between multiple dimensions is know as a **fundamental relation**, and can help us compute useful quantities. It is good practice to pay attention to units being used within fundamental relations to make sure we understand the units of the quantity being computed, and avoid common errors. Below we list some commonly used fundamental relations:

Fundamental Relations.

- $A = \pi r^2$ (area A of a circle from the radius r)
- $V = \frac{4}{3}\pi r^3$ (volume V of a sphere from the radius r)
- $m = \rho A$ (mass m from the density ρ and area A)
- $m = \rho V$ (mass m from the density ρ and volume V)

Example 1.3.3 Density. A circular bacterial colony with radius 10mm has a density of $0.2\frac{\text{mg}}{\text{mm}^2}$. The units given for the density are a good reminder that we'll be using the fundamental relation $m = \rho A$ to compute the mass, since it gives us mass per unit area. The mass of the colony is then

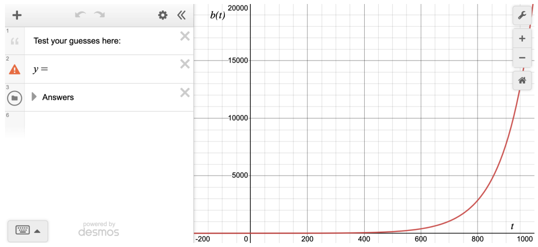
$$\begin{aligned} m &= \rho A \\ &= \rho \pi r^2 \\ &= 0.2\frac{\text{mg}}{\text{mm}^2} \cdot \pi \cdot (10\text{mm})^2 \\ &= 0.2\frac{\text{mg}}{\cancel{\text{mm}^2}} \cdot \pi 100\cancel{\text{mm}^2} \\ &= 20\pi \text{ mg}. \end{aligned}$$

□

Activity 1.3.3 A spherical liquid droplet has density $13\frac{\text{g}}{\text{cm}^3}$. If the radius of the droplet is 2.3mm, what is its mass?

1.3.3 Function Transformations

If we have a function $b(t)$ which measures a population in millions after t hours, how would converting the population unit to thousands change the graph of $b(t)$? How would converting the time unit to days change the graph of $b(t)$? Take a few moments with the interactive below to explore before viewing the answers.



www.desmos.com/calculator/l9rqsarqfw

In both unit conversions above, the resulting graphs are **function transformations** of the original graph $b(t)$. Function transformations are useful outside of unit conversions in recognizing the general shape and other properties of graphs of complicated functions, so we summarize the different types of transformations in the table below:

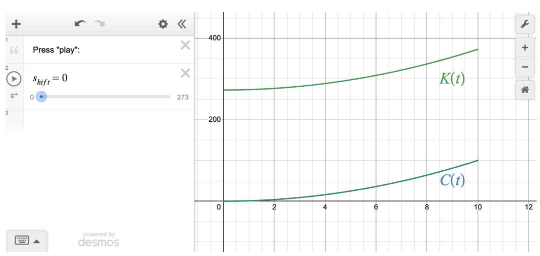
Table 1.3.4 Function transformations of a function $f(x)$

	Vertical	Horizontal
Shift	$f(x)+c$	$f(x+c)$
Scale	$cf(x)$	$f(cx)$
Reflection	$-f(x)$	$f(-x)$

Notice that vertical changes to graphs result from adding or multiplying outside of the function notation (since this changes output values, or y values, of functions), and that horizontal changes to graphs result from adding or multiplying inside of the function notation (since this changes input values, or x values, of functions).

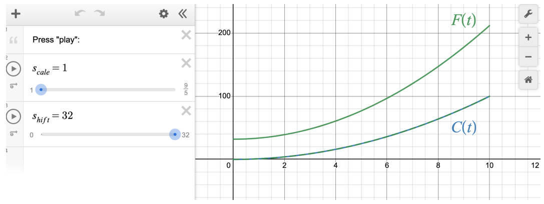
Example 1.3.5 Unit Conversions and Transformtions. A function $C(t)$ gives the temperature in Celsius of an object after t seconds.

The temperature of the object in Kelvin would be given by $K(t) = C(t)+273$, which graphically is a vertical shift up 273 units:



www.desmos.com/calculator/8mntygrvd8

The temperature of the object in Fahrenheit would be given by $F(t) = \frac{9}{5}C(t) + 32$, which graphically is a vertical scale by a factor of $\frac{9}{5}$ followed by a vertical shift up 32 units:



www.desmos.com/calculator/bggzlirwpz

□

1.3.4 Summary

- **Question 1.3.6** When using functions as models, how do we ensure our models are meaningful? □

  To ensure our models are meaningful, we must be sure we are using appropriate fundamental relations that describe the quantities we are measuring, and appropriate units to measure those variable quantities.

- **Question 1.3.7** When computing with functions that are being used as models, how do we ensure our computations are meaningful? □

  We can use the units of variable quantities when making computations to ensure that units combine and cancel appropriately to give a meaningful output.

- **Question 1.3.8** How can we visualize unit conversions graphically? □

  Unit conversions result in a transformation of the graph of the function being used. Function transformations can be useful in recognizing the general shape and other properties of graphs of complicated functions.

1.3.5 Exercises

1. Explain the error in the following argument:

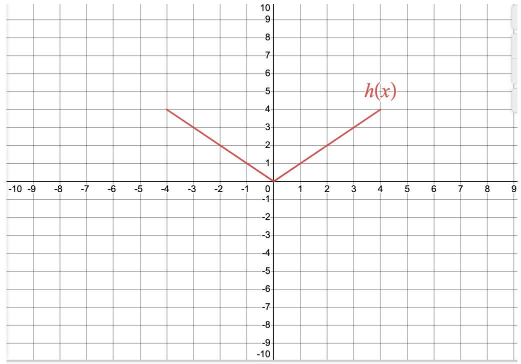
"The area of a square of length 10 cm is 100. If we instead measure the square in meters (so the length is 0.1 m), the area of the same square is 0.01. Since in one case the area gets big and in the other case the area gets small, this means the actual area of a square can change depending on the units used to measure the length."

2. The conversion from degrees Celsius to degrees Fahrenheit is given by $F = \frac{9}{5}C + 32$, where F is measured in degrees Fahrenheit and C is measured in degrees Celsius.

(a) What are the units of 32?

(b) What are the units of $\frac{9}{5}$?

3. The graph of a function $h(x)$ is given below:



On the same coordinate system, sketch the graph of each transformation of $h(x)$.

(a) $h(x) - 2$

(b) $0.5h(x)$

(c) $h(0.5x)$

(d) $-2h(x) + 1$

1.4 Linear Functions

Motivating Questions

- What is the difference between a linear function and a proportional relationship?
- How is the equation of a linear function related to its graph?
- What do solutions to linear equations represent graphically?

There are some families of functions that describe very common relationships. We will study a few of these families in this section, Section 1.5, and Section 1.6.

Warm-Up 1.4.1 The table of a function $g(x)$ is given below. Compute $AROC_{[-1,0]}$, $AROC_{[0,2]}$, and $AROC_{[2,5]}$, recalling Subsection 1.2.6 as needed.

Table 1.4.1

x	$g(x)$
-1	-5
0	-1
2	7
5	19

1.4.1 Linear and Proportional Relationships

You should have found in Warm-Up 1.4.1 that no matter which interval $[a, b]$ you chose, $AROC_{[a,b]}$ was the same number. This property is the defining characteristic of linear functions:

Definition 1.4.2 Linear Function. A **linear function** is a function such that $AROC_{[a,b]}$ is constant for every two x values a and b in the domain of the function. ◊

If a function has a constant average rate of change, we can say something very specific about the equation for the function. For suppose $f(x)$ is a linear function. Then we know that for any x value, the average rate of change between x and 0 is constant (call it m). That is,

$$\begin{aligned} AROC_{[0,x]} &= \frac{f(x)-f(0)}{x-0} \\ &= m \end{aligned}$$

We can re-write $\frac{f(x)-f(0)}{x-0} = m$ as

$$f(x)-f(0) = m(x-0),$$

or

$$f(x) = mx + f(0).$$

Linear Function Equation.

The equation of a linear function in **slope-intercept form** is $f(x) = mx + f(0)$. The value m is called the **slope** of the function, and the value $f(0)$ is called the y**-intercept**.

The equation of a linear function in **point-slope form** is $f(x) = f(a) + m(x-a)$. The value m is the slope of the function, and $(a, f(a))$ is any point on the graph of the function.

A related concept is that of a *proportional relationship*, which we define next.

Definition 1.4.3 Proportional Relationship. A quantity y is **proportional** to a quantity x if they are related by the equation $y = kx$, where k is a parameter. The value k is called the **proportionality constant**. ◊

Example 1.4.4 Linear and Proportional Relationships. The function $\ell(x) = -2x - 4$ is a linear function with slope -2 and y-intercept -4.

The function $p(x) = 8x$ represents a proportional relationship. $p(x)$ is proportional to x with proportionality constant 8. □

Activity 1.4.2 Determine whether each table below represents a linear relationship, proportional relationship, or neither.

x	$f(x)$
-4	-10
-2	-8
0	-6
2	-4
4	-2

Table 1.4.5

x	$g(x)$
-4	16
-2	4
0	0
2	4
4	16

Table 1.4.6

x	$h(x)$
-4	-8
-2	-4
0	0
2	4
4	8

Table 1.4.7

x	$k(x)$
-4	-1
-2	1
0	3
2	5
4	7

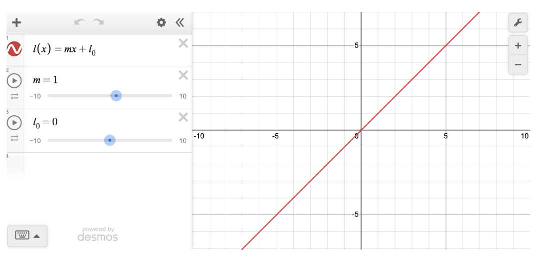
Table 1.4.8

1.4.2 Graphs of Linear Relationships

Notice that in the equation of a linear function $f(x) = mx + f(0)$, the slope m is the constant average rate of change of the function. We can use our previous knowledge about average rates of change from Subsection 1.2.6 and transformations from Subsection 1.3.3 to help us understand the graphical behavior of linear functions.

Activity 1.4.3 Let $l(x) = mx + l(0)$ be a linear function.

1. First, use concepts from Subsection 1.2.6 to complete the sentences below. Then use the interactive provided to support your answers.
    - No matter the value of m, the graph of $l(x)$ is always...
    - If $m > 0$, then the graph of $l(x)$...
    - If $m < 0$, then the graph of $l(x)$...
    - If $m = 0$, then the graph of $l(x)$...
    - If m is very large in absolute value, then the graph of $l(x)$...
    - If m is very small in absolute value, then the graph of $l(x)$...
2. With $y = x$ as your parent function, use concepts from Subsection 1.3.3 to complete the sentences below. Then use the interactive provided to support your answers.
    - If $l(0) > 0$, then the graph of $l(x)$...
    - If $l(0) < 0$, then the graph of $l(x)$...
    - If $l(0) = 0$, then the graph of $l(x)$...



www.desmos.com/calculator/gzznupiaus

1.4.3 Solutions of Linear Equations

A **linear equation** is any equation involving only linear functions. A **solution to a linear equation** is any value that makes the equation true.

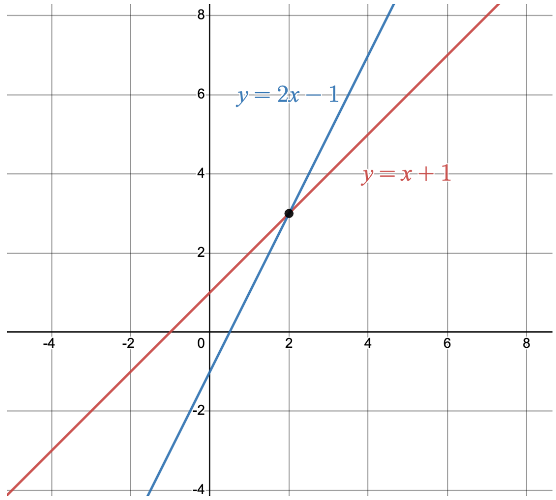
Example 1.4.9 Linear Equations and Solutions. The equation $x + 1 = 2x - 1$ is a linear equation. A solution to this equation is $x = 2$ since $(2) + 1 = 2(2) - 1$. This solution could be found algebraically by solving for "x":

$$\begin{aligned} x + 1 &= 2x - 1 \\ 1 &= x - 1 \\ 2 &= x \end{aligned}$$

□

Given two linear functions $\ell_1(x)$ and $\ell_2(x)$, finding a solution to the linear equation $\ell_1(x) = \ell_2(x)$ means finding an x value when the associated y values on each line are equal. Graphically, this means finding a point of intersection.

Example 1.4.10 Solutions to Linear Equations Graphically. In Example 1.4.9, we saw that a solution to the linear equation $x + 1 = 2x - 1$ is $x = 2$. Graphically, this means if we graphed $y = x + 1$ and $y = 2x - 1$, there would be an intersection point at $x = 2$:



□

Activity 1.4.4 Give an example of a linear equation for each scenario below, or explain why an example does not exist. You are encouraged to think about each scenario graphically.

1. A linear equation which has exactly one solution.
2. A linear equation which has no solution.
3. A linear equation which has two solutions.
4. A linear equation in which every number is a solution.

1.4.4 Summary

- **Question 1.4.11** What is the difference between a linear function and a proportional relationship? □

  A linear function is a function relationship which has a constant average rate of change between any two points. A proportional relationship is a specific type of linear function in which the y- intercept is equal to 0.

- **Question 1.4.12** How is the equation of a linear function related to its graph? □

  The graph of a linear function can be determined by its slope and y-intercept. The slope determines whether the graph increases or decreases, and how quickly it increases or decreases. The y-intercept is a vertical shift of the graph of $y = mx$ which determines where the graph crosses the y-axis.

- **Question 1.4.13** What do solutions to linear equations represent graphically? □

  Solutions to linear equations are points of intersection when graphing the lines on the left and right of the "=" sign. It is possible to have exactly one solution, no solution, or infinitely many solutions.

1.4.5 Exercises

1. Give an example, or explain why no example exists, of the graph of a function which is

(a) linear but not proportional.

(b) proportional but not linear.

(c) linear and proportional.

2. Write the equation of the line that passes through the points $(-2, 10)$ and $(4, -3)$ using

(a) point-slope form.

(b) slope-intercept form.

3. Find the solution(s) to $mx + 1 = 2x - 3$, where m is a parameter (your answer will be $x =$ something in terms of m). For what value(s) of m is there no solution?

4. Let $p(x) = mx + p(0)$ be a linear function representing the price $p(x)$ (in dollars) for riding x miles.

(a) What are the units of $p(0)$?

(b) What are the units of m?

1.5 Exponential and Logarithmic Functions

Motivating Questions

- What is an exponential function?
- What is a logarithmic function?
- How do we solve exponential equations?

Warm-Up 1.5.1 The table of a function $h(x)$ is given below. Compute $AROC_{[-1,0]}$, $AROC_{[0,1]}$, and $AROC_{[1,2]}$, recalling Subsection 1.2.6 as needed.

Table 1.5.1

x	$h(x)$
-1	0.5
0	1
1	2
2	4

Then sketch the visual representation of each average rate of change on a graph to get an idea of how the function changes over time.

1.5.1 Exponential Functions

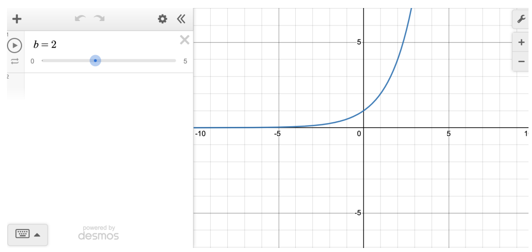
Properties of Exponents. In working with exponential functions, we will often use properties of exponents. We will not review those in full detail in this text, but you may find a review with examples at Paul's Online Notes: Properties of Exponents[1].

[1] tutorial.math.lamar.edu/Classes/Alg/IntegerExponents.aspx#Int_Props

You should have found in Warm-Up 1.5.1 that the value of $AROC_{[a,b]}$ was dependent on the interval $[a, b]$. Since this function does not have a constant average rate of change, we know that it does not represent a linear relationship that we studied in Section 1.4. There is, however, still an important pattern to be discovered in this relationship: each consecutive output value is multiplied by the constant 2 in order to get the next output value. This is the defining characteristic of an **exponential function**:

Definition 1.5.2 Exponential Function. An **exponential function** is a function of the form $f(x) = b^x$, for some constant b where $b > 0$ and $b \neq 1$. The constant b is called the **base** of the exponential. ◊

If $b > 1$ we call the function **exponential growth**. If $0 < b < 1$ we call the function **exponential decay**. Use the interactive to view how the graph of $f(x) = b^x$ changes for different values of b:



www.desmos.com/calculator/h9kzns2y9j

Example 1.5.3 Exponential Models. Exponential functions can be accurate in modeling population growth on a restricted domain. If you have a population that initially has 1 individual, and each individual produces one more individual every year (i.e., the population doubles every year), you could model this with the exponential function $p(t) = 2^t$. □

You may observe a system that behaves exponentially, but does not exactly fit a model of the form $y = b^x$. For example, the output value associated with $x = 0$ may not be always be $b^0 = 1$. This is an instance in which transformations from Subsection 1.3.3 can be useful. A general transformation of an exponential function has the form

$$f(x) = v \cdot b^{h(x-r)} + u.$$

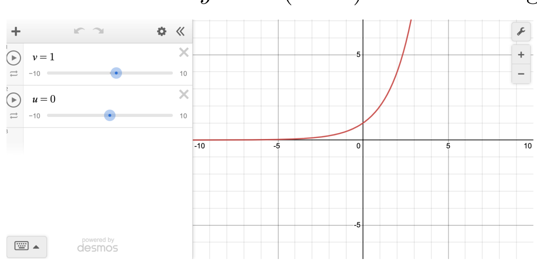
Using rules of exponents this can be written as

$$f(x) = (vb^{-hr}) \cdot (b^h)^x + u.$$

Defining $v_1 = vb^{-hr}$ as a new vertical scale and $b_1 = b^h$ as a new base, we can write the transformation as

$$f(x) = v_1 \cdot (b_1)^x + u,$$

using only a vertical scale and vertical shift on the graph of $y = (b_1)^x$. Therefore, we will focus on transformations of exponential functions in the form $f(x) = v \cdot b^x + u$. Take a moment to use the interactive below to describe how each parameter transforms $y = b^x$ $(b > 1)$ before viewing the answers.



www.desmos.com/calculator/aosgwl700w

- v is a vertical scale by a factor of $|v|$. When a function has the form $f(x) = vb^x$, v is the y-intercept.
- u is a vertical shift. It will shift the graph up or down $|u|$ units.

Though $AROC_{[a,b]}$ is not constant for exponential functions, we can still use these computations to help identify when a relationship fits an exponential model (or a transformation of one) using the following fact:

Fact 1.5.4 *If* $f(x) = v \cdot b^x + u$, *then*

$$\frac{AROC_{[x+1,x+2]}}{AROC_{[x,x+1]}} = b$$

for every x *value.*

Example 1.5.5 Constant AROC Ratio. Looking at the table in Warm-Up 1.5.1, we compute

$$\frac{AROC_{[x+1,x+2]}}{AROC_{[x,x+1]}}$$

for $x = -1$ and $x = 0$:

$$\frac{AROC_{[-1+1,-1+2]}}{AROC_{[-1,-1+1]}} = \frac{AROC_{[0,1]}}{AROC_{[-1,0]}} = \frac{1}{0.5} = 2$$

$$\frac{AROC_{[0+1,0+2]}}{AROC_{[0,0+1]}} = \frac{AROC_{[1,2]}}{AROC_{[0,1]}} = \frac{2}{1} = 2$$

The table lists points of the function $f(x) = 2^x$, which is why we computed the constant $b = 2$ for each ratio. □

Activity 1.5.2 Determine whether the tables below represent an exponential relationship (or a transformation of one). If yes, what is the base?

Table 1.5.6

x	$g(x)$
-2	5
-1	3
0	1
1	-1
2	-3

Table 1.5.7

x	$k(x)$
-2	15
-1	7
0	3
1	1
2	0

It can be useful to write exponential functions using specific bases, depending on the information we are given about the system and what question(s) we are trying to answer. There are two related contexts in which this will be relevant:

Doubling Time and Half Life.

Given a function $p(t) = p_0 \cdot b^t$, $b > 1$, the **doubling time** is the input value t_d such that $p(t_d) = 2p_0$.

Given a function $p(t) = p_0 \cdot b^t$, $0 < b < 1$, the **half life** is the input value t_h such that $p(t_h) = \frac{1}{2}p_0$.

Example 1.5.8 Doubling Time. If we know that the doubling time of a population is 3 years, then we can quickly write down a model for this population using the base 2:

$$p(t) = p_0 \cdot 2^{t/3},$$

where t is measured in years. Note that when we plug in $t = 3$, we get

$$p(3) = p_0 \cdot 2^{3/3} = 2p_0,$$

which shows that the doubling time is $t_d = 3$ as desired. If we'd like to know how much the population increases each *year*, we can use rules of exponents to write

$$p(t) = p_0 \cdot (2^{1/3})^t,$$

to see that the base is $2^{1/3} \approx 1.26$. This means the population increases by approximately 26% each year. □

Another common base for writing exponential functions is the number e. This is an irrational number equal to approximately 2.718, which is special in calculus for reasons we will see in Section 2.5. Writing an exponential function such as 2^x using base e means re-writing 2^x as e^{kx} for some constant k. Using rules of exponents to write

$$2^x = e^{kx} = (e^k)^x,$$

we see this means finding a constant k such that $2 = e^k$. How do we find such a k?

Similarly, in Example 1.5.8, we were given the doubling time and found an equation. What if we were given an equation $p(t) = 10 \cdot (1.5)^t$, and asked to find the doubling time? We would need to solve the equation $20 = 10 \cdot (1.5)^t$, or equivalently $2 = 1.5^t$, for t.

Solving $2 = e^k$ for k and $2 = 1.5^t$ for t are similar in that the variable we are looking to solve for is in the *exponent*. These are called **exponential equations**, and we need tools from Subsection 1.5.2 to solve them.

1.5.2 Logarithmic Functions and Solving Exponential Equations

Properties of Logarithms. In working with logarithmic functions, we will often use properties of logarithms. We will not review those in full detail in this text, but you may find a review with examples at Paul's Online Notes: Properties of Logarithms[2].

Question 1.5.9 Are exponential functions invertible? *Hint*: Picture what the graphs of exponential functions look like. □

Yes, because they pass the horizontal line test (see Example 1.2.10).

[2] `tutorial.math.lamar.edu/Classes/Alg/LogFunctions`

Definition 1.5.10 Logarithmic Function of Base b. Given an exponential function with base b, $y = b^x$, its inverse function is called the **logarithmic function of base** b. It is written $y = \log_b(x)$. ◊

Base 10 and base e are used often and have special notation:

The Common and Natural Log.

Instead of writing $\log_{10}(x)$, we write $\log(x)$. Note that every logarithm has a base, so if you don't see one indicated, it is assumed to be 10! This is called the **common logarithm**.

Instead of writing $\log_e(x)$, we write $\ln(x)$. This is called the **natural logarithm**.

Since logarithmic functions are inverse functions of exponentials, and inverse functions are obtained by swapping the input and output of the original function, we can always interpret logarithmic expressions in terms of exponential expressions:

Interpreting Log Expressions.

The expression

$$y = \log_b(x)$$

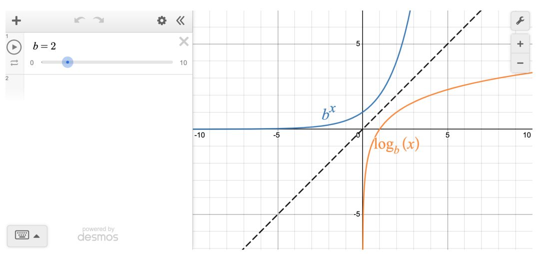
is equivalent to the expression

$$b^y = x.$$

Example 1.5.11 Interpreting Log Expressions.

$\log_2(8) = 3$ since $2^3 = 8$.
$\log(0.01) = -2$ since $10^{-2} = 0.01$.
$\ln(\sqrt{e}) = 0.5$ since $e^{0.5} = \sqrt{e}$. □

Since the graph of an inverse functions is related to the graph of the original function by reflection across the line $y = x$ (Example 1.2.10), we can see what graphs of logarithmic functions look like as the base b changes:



www.desmos.com/calculator/s17e2hlhqk

A property of logarithms that make them a useful tool for solving exponential equations is the following:

Composing Logarithms and Exponentials.

$$\log_b(a^x) = x \cdot \log_b(a)$$

In words, this says that if everything inside of a logarithm is raised to a power, we can bring the power outside of the logarithm using multiplication. Note also that the bases do not need to match to use this property. If they do match, we get a nicer simplification:

$$\log_b(b^x) = x \cdot \log_b(b) = x$$

Example 1.5.12 Solving Exponential Equations Using Logarithms. We will write $y = 2^x$ in the form $y = e^{kx}$. As seen previously, this means we must solve $2 = e^k$ for k:

$$\begin{aligned} 2 &= e^k \\ \ln(2) &= \ln(e^k) \\ \ln(2) &= k \end{aligned}$$

So $k = \ln(2)$, which means $y = 2^x$ can be written as $y = e^{\ln(2)x}$.

We will find the doubling time for the function $p(t) = 10 \cdot (1.5)^t$. Since $p(0) = 10$, we must find the value t such that $p(t) = 2 \cdot p(0) = 20$:

$$\begin{aligned} 20 &= 10 \cdot (1.5)^t \\ 2 &= 1.5^t \\ \log(2) &= \log(1.5^t) \\ \log(2) &= t \cdot \log(1.5) \\ \frac{\log(2)}{\log(1.5)} &= t \end{aligned}$$

So $t_d = \dfrac{\log(2)}{\log(1.5)} \approx 1.71$. □

Activity 1.5.3 A fish population is changing according to the model $p(t) = 500 \cdot e^{-2t}$.

1. Find the doubling time or half life of the population (whichever is appropriate).

2. By what percentage does the population increase or decrease each year?

1.5.3 Scaling and Fitting Data with Logarithms

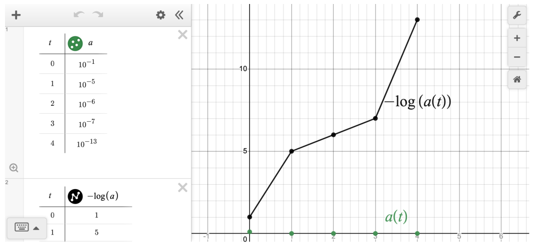
Aside from being a tool for solving exponential equations, logarithms can also be useful for scaling and fitting data, as our next examples illustrate.

Example 1.5.13 Scaling Data. Some systems produce data that are very large or small, and/or cover a very large range. It can be difficult to plot such data in an effective way. Logarithms are useful for taking very large (or small) numbers and making them more manageable, and for compressing a data set that covers a very large range.

For example, you may have a function that measures the acidity of soil over time. Acidity is based on the concentration of hydrogen ions present, and can be measured in moles per liter (M). These numbers are very small, and can vary widely if the acidity of the soil is changing.

The example below shows what the data might look like to plot the molarity a against time t in the first table. In the window given, it's difficult to see any difference between the points. Further, because they cover such a large range of values (.1 to 10^{-13}), it is difficult to change the window to display the relationship in an effective way.

Using the second table, we plot $-\log(a)$ against time t. In fact, this is the definition of **pH** for measuring acidity. As you can see, this makes the numbers more manageable to plot, the relationship of acidity with respect to time easier to describe qualitatively, and changes the range of outputs to values between 1 and 13.



www.desmos.com/calculator/ytbhq6wvz7

□

Example 1.5.14 Fitting Data. After collecting data, we may want to try and fit that data with a specific type of function. For example, we may suspect our data can be modeled by a **power function**: a function of the form $f(x) = kx^p$. We would need to find values for the parameters k and p that best fit our data.

If we take the log of both sides of $f(x) = kx^p$, we get

$$\begin{aligned} \log(f(x)) &= \log(kx^p) \\ \log(f(x)) &= \log(k) + \log(x^p) \\ \underbrace{\log(f(x))}_{y} &= \underbrace{p}_{m} \cdot \underbrace{\log(x)}_{x} + \underbrace{\log(k)}_{y_0} \end{aligned}$$

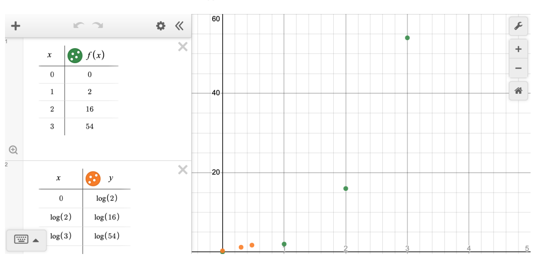
This shows us that if our data fits a power function, then we can plot $\log(f(x))$ as our output and $\log(x)$ as our input and the resulting graph should be linear. Moreover, the slope of the linear function will be the power p and the y-intercept of the linear function will be $\log(k)$.

The example below shows our original data points in the first table that we'd like to fit with a power function $f(x) = kx^p$. The second table shows our data after taking the log of both input and output. We suspect the second table should be a linear function, so we can compute the average rate of change between each interval to confirm and find the slope:

$$\frac{\log(16) - \log(2)}{\log(2) - 0} = 3$$

$$\frac{\log(54) - \log(16)}{\log(3) - \log(2)} = 3$$

So we see that $p = 3$. We can also see the y-intercept from the table is $\log(2)$, so $k = 2$. This means the power function that fits our original data is $y = 2x^3$.



www.desmos.com/calculator/drueualibc

□

1.5.4 Summary

- **Question 1.5.15** What is an exponential function? ☐

  An exponential function is a function that has the form $f(x) = b^x$ for some constant b where $b > 0$ and $b \neq 1$. A special property of these functions is that the ratio of consecutive average rates of change is equal to b. That is,

  $$\frac{AROC_{[x+1,x+2]}}{AROC_{[x,x+1]}} = b$$

  for every x value.

- **Question 1.5.16** What is a logarithmic function? ☐

  A logarithmic function of base b is the inverse function of the exponential function of base b. That is, $\log_b(x) = y$ means $b^y = x$.

- **Question 1.5.17** How do we solve exponential equations? ☐

  To solve equations for a variable that is in the exponent, we can apply logarithms to both sides of the equation and use properties of logarithms. Specifically, we use the property that $\log_b(a^x) = x \cdot \log_b(a)$.

1.5.5 Exercises

1. Let $j(x) = e^{kx}$. For what values of k is $j(x)$ increasing? For what values of k is $j(x)$ decreasing? Justify your answers by using what you know about the graph of $y = b^x$.

2. In the table below, determine the value of c that would make the points fit an exponential relationship with base 0.25.

Table 1.5.18

x	y
0	8
1	c
2	3

3. Use logarithms to find a power function that fits the data in the table below:

Table 1.5.19

x	$h(x)$
0	0
2	24
4	384
6	1944

4. You have data that you suspect fits an exponential model of the form $f(x) = a \cdot b^x$. Show that if you plot $\log(f(x))$ as the output and x as the input, you should get a linear relationship. What is the slope? What is the y-intercept?

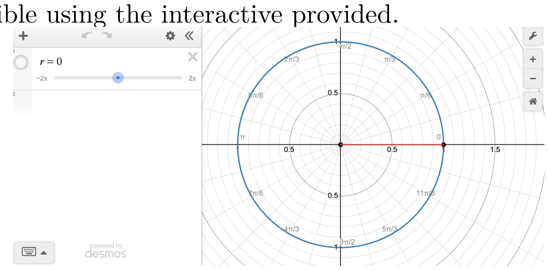
1.6 Trigonometric Functions

Motivating Questions

- How are $\sin(x)$, $\cos(x)$, and $\tan(x)$ defined?
- How can we use trigonometric functions in modeling?

There are many relationships in which a pattern repeats itself over a set period of time, such as outside temperatures, hours of sunlight in a day, and tide levels. Trigonometric functions are useful in modeling these relationships.

Warm-Up 1.6.1 When using trigonometric functions in calculus, we use a measurement called **radians**. Address the questions below as completely as possible using the interactive provided.

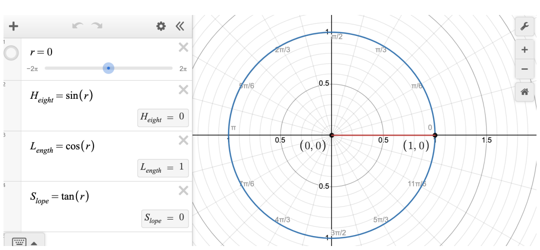


www.desmos.com/calculator/ql8vnbcfpc

1. What do you think "radians" (r) is a measurement of?
2. What is the difference between a positive radian measurement and a negative radian measurement?
3. How many radians trace out the entire circle of radius 1? What is the conversion factor between radians and degrees?
4. What is a radian measure that corresponds to the point $(0, 1)$? $(-1, 0)$? $(0, -1)$? $(1, 0)$?
5. What is the arc length of the segment of the circle traced out when $r = 1$?

1.6.1 The Unit Circle and Trigonometric Functions

The circle in Warm-Up 1.6.1 is called the **unit circle**. As r changes, the line connecting $(0, 0)$ to the unit circle moves around the circle, eventually repeating itself. This repetition is the characteristic that allows us to define functions that repeat a pattern over time. Use the interactive below to determine the definitions of $\sin(r)$, $\cos(r)$, and $\tan(r)$ before viewing the answers.



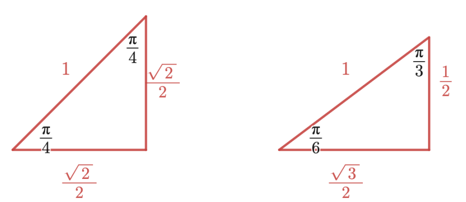
www.desmos.com/calculator/j6vqzcmwlf

- $\sin(r)$ For a value of r, you can draw a right triangle in the unit circle by drawing a vertical line up or down to the x-axis. $\sin(r)$ is the height of that triangle. Equivalently, $\sin(r)$ is the y value of the point on the unit circle associated with r.

- $\cos(r)$ For a value of r, you can draw a right triangle in the unit circle by drawing a vertical line up or down to the x-axis. $\cos(r)$ is the length of that triangle. Equivalently, $\cos(r)$ is the x value of the point on the unit circle associated with r.
- $\tan(r)$ For a value of r, you can draw a right triangle in the unit circle by drawing a vertical line up or down to the x-axis. $\tan(r)$ is the slope of the triangle's hypotenuse. Equivalently, $\tan(r)$ is the slope of the line associated with r connecting $(0, 0)$ to the unit circle.

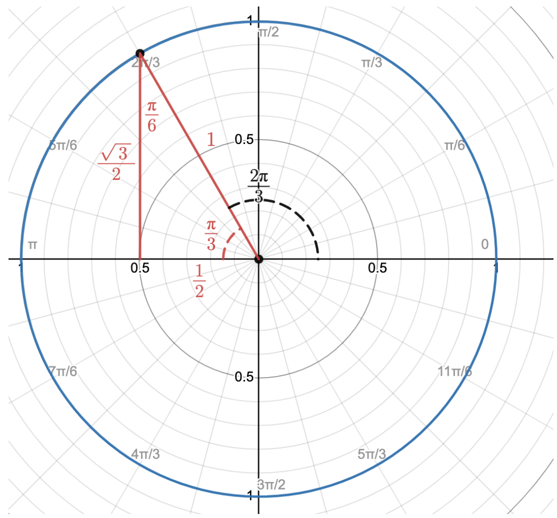
Remark 1.6.1 As with any function, we can use any symbol to represent the input variable inside a trigonometric function. However, it is common to use the greek letter θ ("theta") to represent the input variable in these cases: $\sin(\theta)$, $\cos(\theta)$, and $\tan(\theta)$. The important thing to remember is that the output values of these functions come from the unit circle as described above.
If you are computing the value of trigonometric functions using technology, you'll want to be sure the device you are using is measuring the input in *radians*.
There may be times when it is useful to understand the values of trigonometric functions at special points on the unit circle. You can compute trigonometric values at any multiple of $\frac{\pi}{4}$ or $\frac{\pi}{6}$ just by remembering the lengths of two triangles:

$\frac{\pi}{4}$-$\frac{\pi}{4}$ and $\frac{\pi}{6}$-$\frac{\pi}{3}$ Right Triangles.



As seen in the unit circle, we can draw a right triangle within the unit circle for different values of r. When r is a multiple of $\frac{\pi}{4}$ or $\frac{\pi}{6}$, that triangle will be one of the special triangles , from which we can compute the desired trigonometric value.

Example 1.6.2 Computations in the Unit Circle. We will compute trigonometric values of $\frac{2\pi}{3}$. The first step is to identify the picture of $\frac{2\pi}{3}$ in the unit circle:



As the picture shows, we can then draw a right triangle within the unit circle by drawing a vertical line down to the input axis. We can then determine all of the angles of the triangle, and fill in each length, which tells us the coordinates of the point on the unit circle. Pay attention to whether the coordinates are positive or negative! These are exactly what define, sine, cosine, and tangent of the angle:

$$\cos\left(\frac{2\pi}{3}\right) = -\frac{1}{2}$$
$$\sin\left(\frac{2\pi}{3}\right) = \frac{\sqrt{3}}{2}$$
$$\tan\left(\frac{2\pi}{3}\right) = \frac{\frac{\sqrt{3}}{2} - 0}{-\frac{1}{2} - 0} = -\sqrt{3}$$

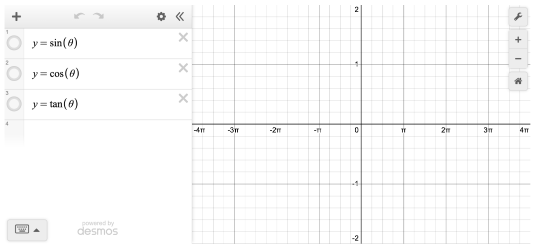
□

Activity 1.6.2 Answer the following questions without using technology:

1. Determine if $\cos\left(\frac{5\pi}{4}\right)$ is positive, negative, or zero. Then compute the exact value.
2. Determine if $\sin\left(\frac{7\pi}{6}\right)$ is positive, negative, or zero. Then compute the exact value.
3. Determine if $\tan\left(-\frac{\pi}{3}\right)$ is positive, negative, or zero. Then compute the exact value.
4. Determine if $\sin(-\pi)$ is positive, negative, or zero. Then compute the exact value.
5. What is the domain of $\tan(\theta)$? How do you know?

1.6.2 Graphs and Transformations of Trigonometric Functions

When we graph trigonometric functions, we are graphing the output value for a given input angle θ. Take a moment to try and sketch the graphs of $\sin(\theta)$, $\cos(\theta)$, and $\tan(\theta)$ based on how these values change as θ changes in the unit circle. Then check your answers with the graph below.

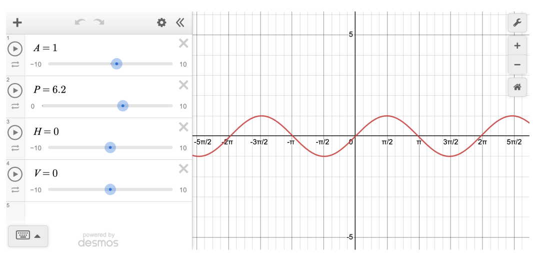


www.desmos.com/calculator/ucda1wv80v

Transformations of trigonometric functions work as described in Subsection 1.3.3, and allow us to change different properties of a repeating pattern based on our observations. However, there is some different vocabulary that we use to describe the transformations in this context based on which graphical properties they change. The general form a transformation of $\sin(x)$ and $\cos(x)$ is

$$A\sin\left(\frac{2\pi}{P}(x-H)\right)+V$$
$$A\cos\left(\frac{2\pi}{P}(x-H)\right)+V$$

Take a moment to describe what each transformation does on the graph of $\sin(x)$ before viewing the answers.



www.desmos.com/calculator/ozeie9oeqv

- A This performs a vertical scale on the graph. If $A < 0$, it vertically reflects the graph. The value of $|A|$ is called the **amplitude** and is half the distance between the maximum (MAX) and minimum (MIN) values of the function. It can be computed by $|A| = \dfrac{MAX - MIN}{2}$.
- P This contributes to the horizontal scale of the graph. It is called the **period** of the function, and determines how long it takes for the function to start repeating itself.
- H This performs a horizontal shift of the graph. It is called a **phase shift** of the function.
- V This performs a vertical shift of the graph. It is the y value over which the graph has vertical symmetry, and is called the **average** of the function. It can be computed by $V = \dfrac{MAX + MIN}{2}$.

Example 1.6.3 Transformations. We would like to model the outside temperature over one day. We will measure this as a function of t hours since midnight. We know the minimum temperature of 43°F occurs at midnight ($t = 0$), and the maximum temperature is 75°F.

Since the minimum temperature occurs at $t = 0$, we will use a transformation of $\cos(x)$. It is possible to use $\sin(x)$, but this would require a phase shift, where as using $\cos(x)$ does not. So we will write a model of the form $f(t) = A\cos\left(\frac{2\pi}{P}(t - H)\right) + V$.

The amplitude is $|A| = \frac{75 - 43}{2} = 16$. Since we'd like the minimum value to occur at $t = 0$, we'll let $A = -16$.

The average is $V = \frac{75 + 43}{2} = 59$.

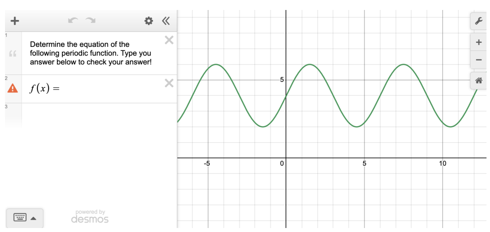
Our model is to repeat after one day (24 hours), so the period should be $P = 24$. This means the horizontal scale is $\frac{2\pi}{24} = \frac{\pi}{12}$. Therefore, our model is

$$f(t) = -16\cos\left(\frac{\pi}{12}t\right) + 59.$$

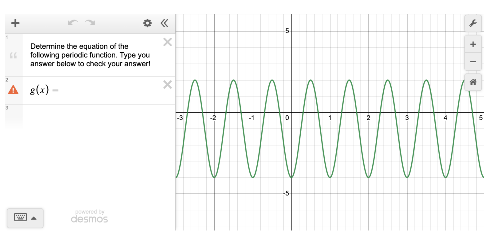
□

Activity 1.6.3

1. Determine an equation for each of the following transformations. You can type your answer into the interactive to check your answer.



www.desmos.com/calculator/jrcijo8m1i



www.desmos.com/calculator/qjnxuv4s83

2. Use transformations to justify each of the following trigonometric identities:

   - $\cos(x) = \sin\left(x + \frac{\pi}{2}\right)$
   - $\cos(-x) = \cos(x)$
   - $\sin(-x) = -\sin(x)$

1.6.3 Summary

- **Question 1.6.4** How are $\sin(\theta)$, $\cos(\theta)$, and $\tan(\theta)$ defined? □

  These functions are defined based on the coordinates of the unit circle associated with the angle θ. $\sin(\theta)$ is the y value, $\cos(\theta)$ is the x value, and $\tan(\theta)$is the slope of the line connecting the point $(0, 0)$ to the point on the unit circle.

- **Question 1.6.5** How can we use trigonometric functions in modeling? □

  Trigonometric functions are useful in modeling relationships that repeat a pattern over a specific period of time. The maximum/minimum values, average value, and period can all be customized using transformations.

1.6.4 Exercises

1. Is there an angle θ such that $\sin(\theta) > 0$ and $\tan(\theta) < 0$? Give an example if one exists, or explain why no such example is possible. If an example exists, illustrate your example with a sketch within the unit circle.

2. Is there an angle θ such that $\sin(\theta) = 0$ and $\tan(\theta) < 0$? Give an example if one exists, or explain why no such example is possible. If an example exists, illustrate your example with a sketch within the unit circle.

3. Sleepiness has two cycles, a circadian rhythm with a period of 24 hours and an ultradian rhythm with period of 4 hours. Both have a phase shift of 0 (in relation to cosine). The average value of the circadian rhythm is 3 sleepiness units and the average value of the ultradian rhythm is 1 sleepiness unit. The amplitude of the circadian rhythm is 2 sleepiness units, and that of the ultradian rhythm is 0.5 sleepiness units.

   (a) Find models for sleepiness over the course of a day due to the circadian rhythm $c(t)$ and ultradian rhythm $u(t)$, where t is measured in hours since midnight.

   (b) Use Desmos[1] to graph your models and verify they have the desired properties.

4. Use Desmos[2] to describe all of the solutions to $\sin(x) = \frac{1}{2}$. How many solutions are there on the domain $[0, \pi]$? How many solutions are there on the domain $[-\pi, 0]$?

1.7 Discrete-Time Dynamical Systems

Motivating Questions

- What is a discrete-time dynamical system, and how can we represent one using functions?
- How can we make predictions about a discrete-time dynamical system?

We were introduced briefly to discrete-time dynamical systems in Section 1.1. In this section and Section 1.8, we will recall the basic vocabulary of these systems, and go into detail regarding how we can analyze these systems.

[1] `Desmos.com`
[2] `Desmos.com`

Warm-Up 1.7.1 A tree is initially 2 meters tall, and grows 150 centimeters each year.

1. Write the relationship of height versus time as a *recursive* function (recall Subsection 1.2.2). Be careful with units! (Section 1.3)

2. Use your recursive relationship to determine the height of the tree after 5 years.

1.7.1 Representing Discrete-Time Dynamical Systems

Recall that a **discrete-time dynamical system** (or **DTDS**) describes a sequence of measurements made at equally spaced intervals. The relationship in Warm-Up 1.7.1 is a DTDS because it describes the height of a tree h (the dependent variable) at discrete times t (the independent variable, defined at $t = 1, 2, 3, \ldots$).

You were asked to describe this relationship recursively, which is a natural way to describe how measurements change from one measurement to the next. This recursive description is called an **updating function**. In order to use the updating function to determine the height of the tree after 5 years, we need to know the initial height of the tree. This is called the **initial value** of the system.

Example 1.7.1 Representing a DTDS. Consider the relationship described in Warm-Up 1.7.1. Let h represent the height of the tree after t years. The initial value is $h_0 = 2$ meters, and the updating function is $h_{t+1} = h_t + 1.5$. Note:

- We use special notation when describing relationships recursively. The independent variable t is given in the subscript, but it is not explicitly used to describe the relationship.

- We had to be careful with units. The initial value was given in meters, while the constant amount of change each year was given in centimeters. We converted to a consistent unit of measure before writing the updating function.

□

In computing with updating functions, it will become useful to identify the function rule used for the recursive relationship. We do this by writing the rule explicitly, treating the current value of the recursive relationship as the input and the next value in the recursive relationship as the output. We call this the **updating function rule**. For example, in Example 1.7.1, the updating function is $h_{t+1} = h_t + 1.5$. Recognizing h_t as the input and h_{t+1} as the output, the updating function rule is $f(x) = x + 1.5$.

Activity 1.7.2 A population of bacteria double each minute. There is initially 3 million bacteria.

1. Write the updating function and initial value for this DTDS. Focus on using correct notation.

2. What is the updating function rule?

3. Use your updating function and initial value to determine the population of bacteria after 4 minutes.

1.7.2 Solutions to Discrete-Time Dynamical Systems

Representing discrete-time dynamical systems using updating functions can be a convenient way to model our observations. However, this can be a difficult representation to analyze and make predictions off of. For example, if you wanted to use the updating function to determine the height of the tree in Example 1.7.1 after 50 years, you would need to iterate the updating function 50 times! In order to analyze a DTDS, it is useful to be able to convert the updating function to other representations of the system.

Example 1.7.2 Creating Tables and Graphs. Using the updating function and initial value, you can create a table of values that describes the relationship between the dependent and independent variables of your system. Looking again at the system in Example 1.7.1, we could iterate the updating function to produce the following table:

Table 1.7.3

t	h_t	h_{t+1}
0	2	3.5
1	3.5	5
2	5	6.5
3	6.5	8

Note that Table 1.7.3 holds information about the relationship of height versus time in two different representations. The right two columns display information about the updating function of the system (h_t versus h_{t+1}), while the left two columns display information about the input and output variables of the system *explicitly* (t versus h_t). This explicit relationship is called the **solution function** of the DTDS.

Using Table 1.7.3, we could also create graphs of this system. Note that we must be careful to label axes appropriately to make it clear which representation of the system we are graphing.

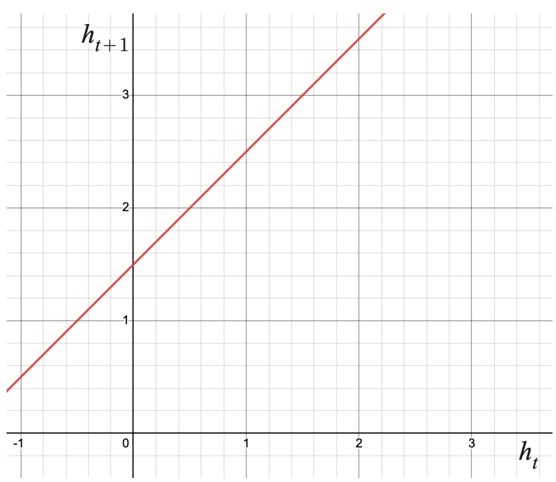


Figure 1.7.4 Graph of the updating function.

Note that the updating function rule is $f(x) = x + 1.5$, which is used to help us graph the updating function.

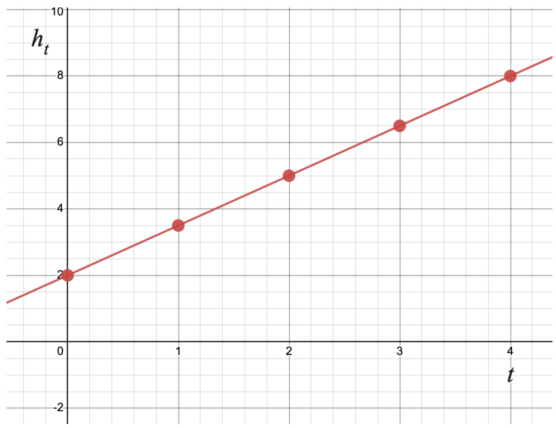


Figure 1.7.5 Graph of the solution function.

Note the discrete dots plotted on the graph of the solution function. These are points from the table and values of the DTDS. They happen to fit the relationship $h_t = 1.5t + 2$, which you see graphed continuously on the graph. □

The explicit form of the solution function is very useful to know if we are able to find it. In Example 1.7.2, this was the function $h_t = 1.5t + 2$. It allows us to answer questions about the system that would be very tedious to do using the updating function. For example, the question regarding the tree height after 50 years is much easier to answer using the solution function:

$$h_{50} = 1.5(50) + 2 = 77 \text{ m}$$

Finding the explicit solution function from a recursive updating function, however, can be a challenging (and sometimes impossible) task. In the following examples, we'll look at a few cases in which we can determine the solution function explicitly.

In order to do so, it is important to recognize the information that an updating function is actually giving us. For example, if you re-write the updating function

$$h_{t+1} = h_t + 1.5$$

as

$$\frac{h_{t+1} - h_t}{1} = 1.5,$$

you may notice that the left side of the equation is the expression for $AROC_{[t,t+1]}$ for the solution function h_t. Recursive equations written this way are called **difference equations**, and can help us determine what type of function the solution function is by analyzing its average rate of change.

Example 1.7.6 A Linear Solution Function. Consider the updating function from Example 1.7.1, $h_{t+1} = h_t + 1.5$. As worked out above, written as a difference equation we see that

$$h_{t+1} - h_t = 1.5.$$

In words, this says the function h_t has a constant average rate of change of 1.5. We know from Section 1.4 that this describes a linear function, and so h_t must be linear with slope 1.5:

$$h_t = 1.5t + h_0$$

Further, we know that $h_0 = 2$, so the solution function is $h_t = 1.5t + 2$. □

Example 1.7.7 An Exponential Solution Function. Consider the updating function $b_{t+1} = 2b_t$ with initial value $b_0 = 3$. Written as a difference equation, this becomes

$$b_{t+1} - b_t = b_t.$$

Since the average rates of change of b_t are not constant, we know that b_t cannot be a linear function. However, if we look at the ratio of consecutive average rates of change (Fact 1.5.4), we compute

$$\begin{aligned}\frac{AROC_{[t+1,t+2]}}{AROC_{[t,t+1]}} &= \frac{b_{t+2} - b_{t+1}}{b_{t+1} - b_t}\\ &= \frac{b_{t+1}}{b_t}\\ &= \frac{2b_t}{b_t}\\ &= 2\end{aligned}$$

This shows that a solution function of the form $b_t = v \cdot 2^t + u$ would fit the discrete system described by the updating function $b_{t+1} = 2b_t$. Create a table similar to that of Table 1.7.3 using the updating function to verify that the solution function is described by $b_t = 3 \cdot 2^t$. □

Example 1.7.8 An Exponential Transformation Solution Function. Consider the updating function $p_{t+1} = 0.5p_t + 1$ with initial value $p_0 = 10$. Written as a difference equation, this becomes

$$p_{t+1} - p_t = -0.5p_t + 1.$$

We can see once again that the solution function cannot be linear, so we will check the ratio of consecutive average rates of change:

$$\begin{aligned}\frac{AROC_{[t+1,t+2]}}{AROC_{[t,t+1]}} &= \frac{p_{t+2} - p_{t+1}}{p_{t+1} - p_t}\\ &= \frac{(0.5p_{t+1} + 1) - (0.5p_t + 1)}{p_{t+1} - p_t}\\ &= \frac{0.5p_{t+1} - 0.5p_t}{p_{t+1} - p_t}\\ &= \frac{0.5(p_{t+1} - p_t)}{p_{t+1} - p_t}\\ &= 0.5\end{aligned}$$

This shows that a solution function of the form $p_t = v \cdot 0.5^t + u$ would fit the discrete system described by the updating function $p_{t+1} = 0.5p_t + 1$. However, unlike in the previous example, creating a table of values will show that the solution function is *not* $p_t = 10 \cdot 0.5^t$:

Table 1.7.9

t	p_t	p_{t+1}	$10 \cdot 0.5^t$
0	10	6	10
1	6	4	5
2	4	3	2.5
3	3	2.5	1.25
4	2.5	2.25	0.625
5	2.25	2.125	0.3125

We can, however, notice from the table that the values of p_t seem to be getting closer to 2 as time goes on. We will develop a tool for verifying this hypothesis in Section 1.8, but creating a table like above is a good way to build intuition for how the solution behaves.

We know that the solution function is a transformation of the exponential function $f(t) = 0.5^t$, and that we'd like there to be a horizontal asymptote at $y = 2$, since this is what the outputs seems to approach as t gets large. Since $f(t) = 0.5^t$ is exponential decay and has a horizontal asymptote at $y = 0$, we will shift the function up 2 using transformations:

$$p_t = v \cdot 0.5^t + 2$$

Now we will find the vertical scale required to satisfy the initial value $p_0 = 10$:

$$\begin{aligned} 10 &= v \cdot 0.5^0 + 2 \\ 10 &= v + 2 \\ 8 &= v \end{aligned}$$

Thus, the solution function is given by $p_t = 8 \cdot 0.5^t + 2$. Create a table of values for $t = 1, 2, 3, 4, 5$ and 6 to verify this solution function fits the points of this DTDS. □

We can summarize our findings from the previous examples with the following:

Solution Functions of $p_{t+1} = bp_t + m,\ b > 0$.

Let $p_{t+1} = bp_t + m$ be an updating function with initial value p_0.

- If $b = 1$, the corresponding solution function is linear with slope m and y-intercept p_0:

$$p_t = mt + p_0$$

- If $m = 0$, the corresponding solution function is exponential with base b and y-intercept p_0:

$$p_t = p_0 \cdot b^t$$

- If $b > 0$, $b \neq 1$, and $m \neq 0$, the corresponding solution function is a transformation of an exponential function with base b.

Activity 1.7.3 Determine the solution function associated with each updating function and initial value.

1. $d_{t+1} = d_t - 4,\ d_0 = 100$
2. $k_{t+1} = 1.8k_t,\ k_0 = 20$
3. $m_{t+1} = 0.8m_t + 5,\ m_0 = 20$

1.7.3 Summary

- **Question 1.7.10** What is a discrete-time dynamical system, and how can we represent one using functions? □

  A discrete-time dynamical system describes a sequence of measurements made at equally spaced intervals. It is common to represent a DTDS using an updating function (a recursive function) and an initial value.

- **Question 1.7.11** How can we make predictions about a discrete-time dynamical system? ☐

  We can make predictions about a DTDS by describing its solution function. We can do this by creating a table of several values, plotting several points on a graph, and in some cases, finding the equation of the solution function explicitly.

1.7.4 Exercises

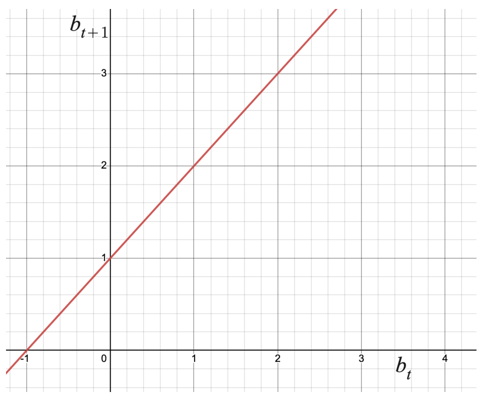
1. Let $f(x)$ be an updating function rule.

(a) If $f(x)$ is invertible, what does $f^{-1}(x)$ represent in the context of the DTDS? Illustrate your answer using the updating function $h_{t+1} = h_t + 1.5$.

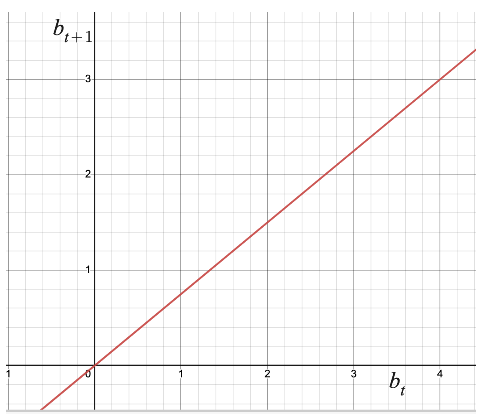
(b) What does $(f \circ f)(x)$ represent in the context of the DTDS? Illustrate your answer using the updating function $h_{t+1} = h_t + 1.5$.

2. Determine the solution function associated with each updating function below. Use an initial value of $b_0 = 5$ for each.

(a)



(b)



3. A patient starts with a concentration of medicine in her bloodstream equal to 10 milligrams per liter. Each day, the patient decomposes 60% of the medication in her bloodstream. However, the doctor gives her enough medication daily to increase the concentration in her bloodstream by 3 milligram per liter.

(a) What is the updating function and initial condition describing this DTDS? Graph the updating function, labeling axes appropriately.

(b) Create a table similar to that of Table 1.7.3 using the updating function to find the concentration of medication on day 7. Then graph the values of the solution function labeling axes appropriately.

(c) Find the equation of the solution function for this DTDS explicitly.

1.8 Analyzing Discrete-Time Dynamical Systems

Motivating Questions

- How can we describe the qualitative behavior of a solution function without finding the equation of the solution function explicitly?
- What are equilibrium values of a DTDS and how can we find and describe them?

We ended Section 1.7 by finding explicit equations of solution functions for certain types of updating functions. In general, this is a very difficult task. It is beneficial to have other tools for describing the *qualitative* behavior of a solution function, especially when we can't find the equation explicitly.

The qualitative behavior of a solution function describes how the function behaves as time goes on. Does the solution function increase or decrease for a particular initial value? Does the solution function approach a specific value? In this section we will learn a graphical tool for describing the behavior of a solution function, and how to identify important values that dictate the behavior of the system.

Warm-Up 1.8.1 Consider the updating function $g_{t+1} = 0.5g_t + 2$ with initial value $g_0 = 20$.

1. Sketch the graph of the updating function. Identify the updating function rule to help you, and clearly label your axes.
2. Complete the table below using the updating function. Plot the relevant points in your table on the graph of the updating function from the previous part.

Table 1.8.1

t	g_t	g_{t+1}
0		
1		
2		
3		

1.8.1 Cob-Webbing

We have seen several examples in which we create tables like Table 1.8.1 using the updating function and initial value. **Cob-webbing** is a method which mimics the iteration process used to create such a table on the graph of the *updating function* of a DTDS. Instead of producing the specific values like a table, cob-webbing is an efficient algorithm that can show us the qualitative behavior of a solution function starting from an initial value of our choosing. The cob-webbing algorithm is described below.

Cob-webbing Algorithm.

Begin with the graph of an updating function $b_{t+1} = f(b_t)$, the diagonal line $b_{t+1} = b_t$, and an initial condition b_0.

1. Identify b_0 on the horizontal axis (b_t) of the graph of the updating function.

2. Travel vertically until you intersect the updating function, and draw a closed circle.

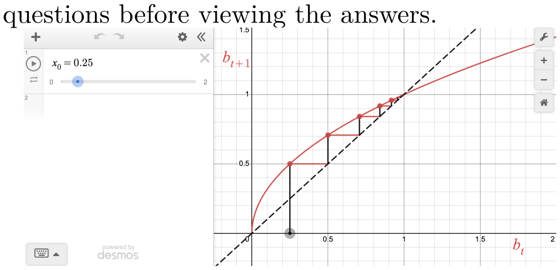
3. From the closed circle, travel horizontally until you intersect the diagonal line $b_{t+1} = b_t$.

4. From this point on the diagonal line, travel vertically until you intersect the updating function, and draw a closed circle.

5. Repeat step 3 through step 4 as many times as desired.

The output value of the closed circles that are generated on the graph of the updating function using the cob-webbing algorithm are consecutive output values of the solution function. By generating these points graphically, we can visually see how the solution function is behaving. Is it increasing or decreasing? Is it approaching a particular value?

Example 1.8.2 Cob-webbing. The interactive below shows the graph of an updating function $x_{t+1} = f(x_t)$ which is more complicated than those that we analyzed previously. Use the interactive to visualize cob-webbing and answer the questions before viewing the answers.



www.desmos.com/calculator/ea7a7h5tw9

1. If $x_0 = 0.25$, is the solution function increasing or deacreasing? The solution function x_t is increasing, since the closed circles on the updating function have output values that are getting larger.

2. If $x_0 = 0.25$, is the solution function increasing/decreasing at an increasing or decreasing rate? The solution function is increasing at a decreasing rate, since it looks like the closed circles are getting bigger by a smaller amount each time.

3. If $x_0 = 0.25$, is the solution function approaching a particular value? The solution function is getting closer and closer to the value $x_t = 1$.

4. Find an example of an initial value for which the solution function will decrease. The solution function will decrease if $x_0 > 1$.

□

1.8.2 Equilibrium Values and Stability

In Example 1.8.2, you'll notice that the solution function approaches the value $x_t = 1$, regardless of what the initial condition is. This is an example of a special value in a DTDS, which we define now.

Definition 1.8.3 Equilibrium Values. Let $b_{t+1} = f(b_t)$ be an updating function for a DTDS. An output value b^* is an **equilibrium value** of the system if $f(b^*) = b^*$. ◊

Remark 1.8.4

1. It is also common to call an equilibrium value b^* an **equilibrium point**, referring to the point (b^*, b^*) which is on the graph of the updating function.

2. Another way to phrase Definition 1.8.3 is that an equilibrium value is an output value that is unchanged by the updating function. In Example 1.8.2, we can see that $x^* = 1$ is an equilibrium value because if we start cob-webbing at $x_0 = 1$, the next output value is $x_1 = 1$.

3. Equilibrium values help to separate the domain of initial values into pieces on which we can describe the general behavior of a solution function. In Example 1.8.2, $x^* = 1$ is the value to compare against to determine if the solution function will increase or decrease. If $x_0 > 1$, the solution function will decrease. If $x_0 < 1$, the solution function will increase.

4. Equilibrium values often have significant meaning in the context of the system being modeled. For example, if the updating function in Example 1.8.2 is modeling a population with respect to time, $x^* = 1$ may represent a carrying capacity of that system.

Given an updating function, we can find equilibrium values *algebraically* or *graphically*, depending on how the updating function is represented.

Example 1.8.5 Finding Equilibria. Let $p_{t+1} = 0.5p_t + 1$.

- *Finding Equilibria Algebraically.* An equilibrium value is a value p^* that is unchanged by the updating function. That is, p^* is an equilibrium value if

  $$p^* = 0.5p^* + 1.$$

  We can subtract $0.5p^*$ from both sides to get

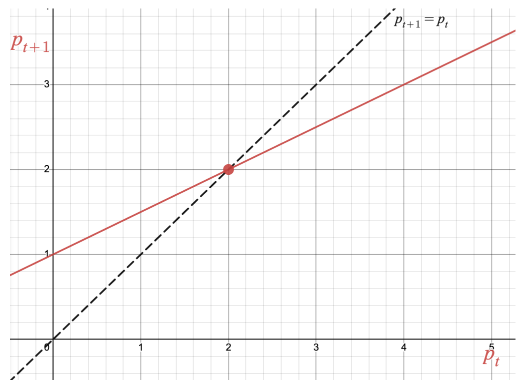
  $$0.5p^* = 1,$$

  and then divide both sides by 0.5 to get

  $$p^* = \frac{1}{0.5} = 2.$$

  This shows that $p^* = 2$ is an equilibrium value of the system, which we can verify by computing $0.5 \cdot \mathbf{2} + 1 = \mathbf{2}$.

- *Finding Equilibria Graphically.* We can graph the updating function $p_{t+1} = 0.5p_t + 1$ by recognizing the updating function rule as $f(x) = 0.5x + 1$:



We graph the updating function along with the diagonal line $p_{t+1} = p_t$ because graphically, a value which is unchanged by the updating function can be identified by finding an intersection point with the line $p_{t+1} = p_t$. We can see on the graph that the diagonal intersects the updating function at $(2, 2)$, which means $p^* = 2$ is an equilibrium value of the system.

□

Remark 1.8.6 Note that the updating function explored in Example 1.8.5 is the same as the one we analyzed in Example 1.7.8 when trying to use transformations to find a solution function equation explicitly. In Example 1.7.8, we used a table of iterated values to make an educated guess at what value the system was approaching in order to determine what a horizontal asymptote should be of the solution function. In Example 1.8.5, we see a more calculated way to find what a horizontal asymptote should be by calculating the equilibrium value of the system.

Activity 1.8.2 Consider a bacterial population which is modeled by the updating function $b_{t+1} = rb_t$, where the parameter r is the per capita reproduction rate.

1. Assuming $r = 2$, find the equilibrium value(s) of this system algebraically.
2. Assuming $r \neq 1$, find the equilibrium value(s) of this system algebraically.
3. Sketch a graph of the updating function if $r = 2$. Cob-web using several different initial values to describe how the solution function behaves with respect to the equilibrium value(s).
4. Sketch a graph of the updating function if $r = 0.5$. Cob-web using several different initial values to describe how the solution function behaves with respect to the equilibrium value(s).

Solution functions need not always approach an equilibrium value as time goes on. We describe how a system behaves with respect to its equilibrium values with the following vocabulary:

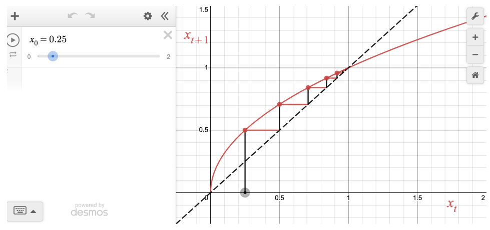
Definition 1.8.7 Stability. Let p^* be an equilibrium value of a DTDS.

If the output values of solution functions get as close to p^* as we'd like when initial values are close to p^*, we say that p^* is a **stable equilibrium value**.

Otherwise, we say that p^* is an **unstable equilibrium value**. ◊

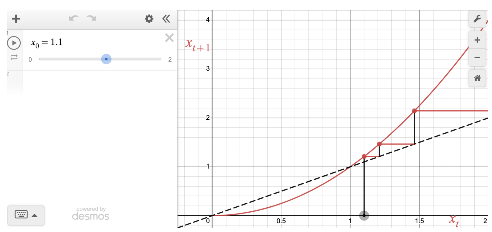
Example 1.8.8 Stability.

- *Stable Equilibrium Value.* For the updating function $x_{t+1} = f(x_t)$ below, the value $x^* = 1$ is a stable equilibrium value. Change the initial value to different numbers above and below the equilibrium value, and notice that in all cases, the solution function tends towards the equilibrium value as time goes on.



www.desmos.com/calculator/wxqars743x

- *Unstable Equilibrium Value.* For the updating function $x_{t+1} = f(x_t)$ below, the value $x^* = 1$ is an unstable equilibrium value. Change the initial value to different numbers above and below the equilibrium value, and notice the solution function does not always tend towards the equilibrium value as time goes on.



www.desmos.com/calculator/pmijlwm6az

□

1.8.3 Summary

- **Question 1.8.9** How can we describe the qualitative behavior of a solution function without finding the equation of the solution function explicitly? □

  We can describe the qualitative behavior of a solution function by cob-webbing on the graph of the updating function. The series of points generated by cob-webbing show us whether a solution function is increasing or decreasing, how it is increasing or decreasing, and if the solution function is approaching a particular value as time goes on.

- **Question 1.8.10** What are equilibrium values of a DTDS and how can we find and describe them? □

  Equilibrium values are values that are unchanged by the updating function. We can find equilibrium values algebraically using the equation of the updating function or graphically using the graph of the updating function. We can categorize equilibrium values as stable or unstable based on how solution functions behave with initial values that are close to the equilibrium value.

1.8.4 Exercises

1. A **linear updating function** is an updating function whose updating function rule is a linear function. Is it possible for a DTDS with a linear updating function to have exactly two equilibrium values? Explain how you know.

2. Give an example of the graph of an updating function that has 2 stable equilibrium values and 1 unstable equilibrium value. Use cob-webbing to illustrate why your example works.

1.9 Applications: The Lung Model and Competing Species

Motivating Questions

- How can we model gas exchange in the lung using a discrete-time dynamical system?
- What is a weighted average?
- How can we model a system which involves two distinct populations that change over time?

In our final section of the chapter we will survey two discrete-time dynamical systems, using the tools we've developed so far to analyze the behavior of each system over time.

Warm-Up 1.9.1 Consider the updating function $p_{t+1} = 0.7p_t + 0.6$.

1. Determine the equilibrium value(s) of this system algebraically.
2. Create tables similar to that of Table 1.7.3 using the updating function and appropriate initial values to guess whether the equilibrium value(s) from the previous part are stable or unstable.
3. Graph the updating function and use cob-webbing to verify your conclusion regarding the stability of the equilibrium value(s).

1.9.1 The Lung Model

We will develop a basic DTDS measuring the concentration c_t of a gas in the lung after t breaths (exhale and inhale). As a starting place for developing our model, we will make several assumptions:

- There is a fixed maximum volume V of air in the lung.
- The volume of air breathed out during exhale W is the same as the volume of air breathed in during inhale.
- The concentration β of the gas being measured is constant in the outside air.

This is a simplified model, but it is important to note that models are never perfect. It is important to analyze a model for its limitations, and adjust your model accordingly if there is a factor important to your study that is not being considered. An example of how one might adjust the lung model presented here is given in Exercise 1.9.4.1.

We will measure volume in liters (L) and the amount of gas particles in millimoles (mmol). Then the concentration of the gas will be measured in millimoles per liter (mmol/L). In order to develop our model, we will look at the concentration of gas at different stages of the breathing process, using unit conversions (Section 1.3) to guide our computations. The current concentration in the lung is represented by c_t, and we are looking for the concentration after one more breath, c_{t+1}.

Table 1.9.1

Stage	**Volume (L)**	**Concentration (mmol/L)**	**Amount (mmol)**
Air in Lung Before Exhale	V L	c_t mmol/L	$V\cancel{L} \cdot c_t \frac{\text{mmol}}{\cancel{L}} = Vc_t$ mmol
Exhaled Air	W L	c_t mmol/L	$W\cancel{L} \cdot c_t \frac{\text{mmol}}{\cancel{L}} = Wc_t$ mmol
Air in Lung After Exhale	$(V-W)$ L	c_t mmol/L	$(V-W)\cancel{L} \cdot c_t \frac{\text{mmol}}{\cancel{L}} = (V-W$
Inhaled Air	W L	β mmol/L	$W\cancel{L} \cdot \beta \frac{\text{mmol}}{\cancel{L}} = W\beta$ mmol

The final stage of a single breath is to analyze the air in the lung after inhale. During this stage, however, we are looking to compute the concentration instead of being given it as in previous stages. Concentration is computed as "amount" divided "volume". The volume of air after inhale is the full volume of the lung, V L. The amount of gas particles in the lung after inhale is the sum of of the amount left in the lung after exhale, plus the amount breathed in during inhale:

$$(V-W)c_t \text{ mmol} + W\beta \text{ mmol}$$

Thus, the new concentration of gas in the lung after one breath is

$$\begin{aligned} c_{t+1} &= \frac{(V-W)c_t + W\beta}{V} \\ &= \frac{(V-W)}{V}c_t + \frac{W}{V}\beta \\ &= \left(1 - \frac{W}{V}\right)c_t + \frac{W}{V}\beta. \end{aligned}$$

We summarize the development of the lung model below:

The Lung Model.

A DTDS which measures the concentration c_t (in mmol/L) of a gas in the lungs after t breaths is given by

$$c_{t+1} = (1-p)c_t + p\beta,$$

where β is the outside concentration of the gas being measured and $p = \frac{W}{V}$ is the percentage of the lung's volume that is being filled with outside air during each inhale.

Activity 1.9.2 Consider a lung model with $V = 6$ L, $W = 2$ L, and $\beta = 3$ mmol/L.

1. Compute the equilibrium value(s) of this system algebraically.
2. Graph the updating function and use cob-webbing to classify the equilibrium value(s) as stable or unstable.
3. Interpret your previous answers in the context of the lung. Do your answers make sense?

The updating function for the lung model is an example of a *weighted average*, which we define now.

Definition 1.9.2 Weighted Average. A **weighted average** of two numbers a and b is a sum of the form

$$(1-p)a+pb,$$

where p is any number such that $0 \le p \le 1$. The numbers p and $1-p$ are called **weights** of the weighted average. ♢

Remark 1.9.3

- When $p=0.5$, this is just the average of two numbers:

$$(1-0.5)a+0.5b=0.5a+0.5b=\frac{a+b}{2}$$

- The weights p and $1-p$ represent a percentage, allowing us to place more importance (or weight) on one number over another.
- The weights p and $1-p$ always sum to 1 (100%).

Example 1.9.4 Weighted Average.

1. The sum

$$0.8(2)+0.2(10)$$

is a weighted average of the numbers 2 and 10. The number 2 is weighted as 80% while the number 10 is weighted as 20%. The sum is equal to 3.6, which is much closer to 2 than 10 since 2 is being given more importance.

2. The sum

$$0.2(2)+0.8(10)$$

is a weighted average of the numbers 2 and 10. The number 2 is weighted as 20% while the number 10 is weighted as 80%. The sum is equal to 8.4, which is much closer to 10 than 2 since 10 is being given more importance.

□

Activity 1.9.3 The lung model $c_{t+1}=(1-p)c_t+p\beta$ is a weighted average.

1. What are the two numbers being averaged?
2. What are the weights of the weighted average?
3. Interpret the weighted average in the context of what is being modeled. What is being given more importance and why?

1.9.2 A Model for Competing Species

Consider a situation in which there are two different bacterial strains growing in the same environment. We'd like to describe the make-up of these two populations in a single discrete-time dynamical system. We'll start by modeling each strain separately assuming an exponential growth model. Let the first strain's population after t minutes be represented by x_t and the second strain's population be represented by y_t. Then each population can be modeled separately as

$$x_{t+1}=sx_t$$
$$y_{t+1}=ry_t,$$

where s and r are the per capita growth rates of the first and second strain, respectively.

In order to create a single model that describes both of these populations, we will measure the **population proportion** of one of the strains. A population proportion is the fraction that a sub-population represents with respect to the total population. Let p_t be the population proportion of the first strain after t minutes. Then

$$p_t = \frac{\text{The population of strain 1 after t minutes}}{\text{The total population after t minutes}} = \frac{x_t}{x_t + y_t}.$$

Note that even though p_t is being measured with respect to the first strain's population, we could also compute the population proportion of the second strain, $\frac{y_t}{x_t + y_t}$, from this expression by computing $1 - p_t$, since the sum of the two population proportions should equal 1 (100%).

We would like to have an updating function that describes the relationship p_t. That is, we want an equation of the form $p_{t+1} = f(p_t)$. Towards that end, we compute

$$\begin{aligned}
p_{t+1} &= \frac{x_{t+1}}{x_{t+1} + y_{t+1}} \\
&= \frac{sx_t}{sx_t + ry_t} \\
&= \frac{sx_t}{sx_t + ry_t} \cdot \frac{\frac{1}{x_t+y_t}}{\frac{1}{x_t+y_t}} \\
&= \frac{s\frac{x_t}{x_t+y_t}}{s\frac{x_t}{x_t+y_t} + r\frac{y_t}{x_t+y_t}} \\
&= \frac{sp_t}{sp_t + r(1 - p_t)} \\
&= \frac{sp_t}{r + (s - r)p_t}
\end{aligned}$$

We summarize the development of this competing species model below:

Competing Species Model.

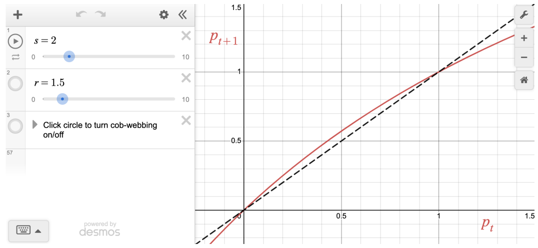
If species 1 grows with per capita growth rate s and species 2 grows with per capita growth rate r, the population proportion of species 1 is modeled by the updating function

$$p_{t+1} = \frac{sp_t}{r + (s - r)p_t}.$$

We note here that the competing species model is an example of a **non-linear updating function**. That is, the updating function rule $f(x) = \frac{sx}{r + (s - r)x}$ is not a linear function. This can make finding equilibrium values algebraically more complex, and also allow us to have multiple different equilibrium values. The type of equations that are most relevant for us to be able to solve in this section are **quadratic equations**. You can find a nice review with solutions of solving quadratic equations at Paul's Online Notes, Part 1[1] and Part 2[2].

Activity 1.9.4 Use the interactive below to answer questions regarding the equilibrium values of the competing species model.

[1] tutorial.math.lamar.edu/Classes/Alg/SolveQuadraticEqnsI.aspx
[2] tutorial.math.lamar.edu/Classes/Alg/SolveQuadraticEqnsII.aspx



www.desmos.com/calculator/jmrevtw05y

1. Do the equilibrium values of this model depend on the values of s and r? Explain.

2. Compute the equilibrium values of $p_{t+1} = \dfrac{3p_t}{2 + (3-2)p_t}$ algebraically, then use the interactive to verify your answers.

3. Compute the equilibrium values of $p_{t+1} = \dfrac{sp_t}{r + (s-r)p_t}$ algebraically.

4. Does the stability of the equilibrium values depend on the values of s and r? Use cob-webbing to determine when each equilibrium value is stable/unstable. Try to do the cob-webbing on your own before using the button provided that illustrates the cob-webbing.

5. What do the equilibrium values represent in the context of the competing species model? How can you interpret the stability of each equilibrium value in the context of the competing species model?

1.9.3 Summary

- **Question 1.9.5** How can we model gas exchange in the lung using a discrete-time dynamical system? □

  We can model gas exchange in the lung using the lung model.

- **Question 1.9.6** What is a weighted average? □

  A weighted average between two numbers a and b is a sum of the form $(1-p)a + pb$, where $0 \le p \le 1$. Weighted averages give us a way to place more importance on one number over another when computing the average value.

- **Question 1.9.7** How can we model a system which involves two distinct populations that change over time? □

  We can form a DTDS that measures a population proportion. One example of this is when two distinct population exhibit exponential growth, we get the competing species model. One more example of this in the context of migration models is outlined in Exercise 1.9.4.5 and Exercise 1.9.4.4.

1.9.4 Exercises

1. Our current model for gas exchange in the lung does not account for absorption of the chemical in the body, which can occur for chemicals such as oxygen. If a fraction α of the chemical is absorbed before breathing out, the new DTDS becomes

$$c_{t+1} = (1-p)(1-\alpha)c_t + p\beta,$$

where $p = \frac{W}{V}$ and β is the outside concentration of the chemical.

(a) Explain in words why multiplying c_t by $(1-\alpha)$ makes sense.

(b) Find the equilibrium value(s) c^* algebraically. Your answer should be in terms of p, α, and β.

(c) Let $p = 0.5$ and $\beta = 0.2$. Use the table to test several values of α to see how absorption effects equilibria. As absorption increases, what happens to the equilibrium concentration?

Table 1.9.8

α	c^*
0	
0.25	
0.5	
0.75	
1	

2. Consider the lung model given by $c_{t+1} = 0.4c_t + 1.2$, $c_0 = 1$.

(a) Find the solution function for this DTDS.

Hint: Example 1.7.8 and Remark 1.8.6

(b) Use the solution function to determine the approximate number of breaths it will take for the concentration to reach 1.99 mmol/L.

3. Explain why the sum $0.2(10) + 0.6(3)$ is not a weighted average.

4. Suppose two nearby islands have populations of butterflies, with x_t on the first island and y_t on the second. Each year, 20% of the butterflies from the first island fly to the second and 30% of the butterflies from the second fly to the first.

Also suppose each butterfly that begins the year on the first island produces one offspring after migration (whether they find themselves on the first or the second island). Those that begin the year on the second island do not reproduce. Assume that no butterflies die.

(a) As an initial example, suppose we start with 100 butterflies on each island. Find the number after migration and after reproduction on both islands.

Hint: The population of island 1 is 190 and island 2 is 110, but it is *how* you calculate these numbers that you want to pay attention to.

(b) Find equations for x_{t+1} and y_{t+1}, paying attention to how you calculated the previous part to identify patterns.

(c) Let p_t represent the proportion of butterflies on the first island after t years. Find the updating function for p_t.

(d) Find the equilibrium value of the system p_t. There should be only one that makes biological sense. Then classify the equilibrium value as stable or unstable. You may use any method of your choice.

(e) Interpret the meaning of the equilibrium value and its stability in the context of the model.

5. A population of salmon live in one of two territories in the Atlantic ocean. There are k_t fish in territory 1 and h_t fish in territory 2. Every year, 80% of the fish from territory 1 migrate to territory 2, and 75% of the fish from territory 2 migrate to territory 1.

(a) Find an equation for k_{t+1} in terms of k_t and h_t.

(b) Find an equation for h_{t+1} in terms of k_t and h_t.

(c) Let p_t represent the proportion of fish living in territory 1 after t years. Find the updating function for p_t.

(d) Find all equilibrium values of the system p_t and classify each as stable or unstable. You may use any method of your choice.

(e) Interpret the meaning of each equilibrium value and its stability in the context of the model.

Chapter 2

The Derivative

2.1 Limits of Functions

Motivating Questions

- What is the mathematical notion of *limit* and what role do limits play in the study of functions?
- What is the meaning of the notation $\lim_{x \to a} f(x) = L$?
- How do we go about determining the value of the limit of a function at a point?
- What is instantaneous velocity and how do we compute it?

We've seen in Chapter 1 that functions can model many interesting phenomena, such as population growth and temperature patterns over time. We can use calculus to study how a function value changes in response to changes in the input variable.

Think about a falling ball whose position function is given by $s(t) = 64-16t^2$. The average rate of change (also called *average velocity* in this context) on the interval $[1, x]$ is given by

$$AROC_{[1,x]} = \frac{s(x) - s(1)}{x - 1} = \frac{(64 - 16x^2) - (64 - 16)}{x - 1} = \frac{16 - 16x^2}{x - 1}.$$

Note that the average velocity is a function of x. That is, the function $g(x) = \frac{16-16x^2}{x-1}$ tells us the average velocity of the ball on the interval from $t = 1$ to $t = x$. Calculus provides tools to describe how a function changes at a given instant; that is, at a single point in time. This is called the **instantaneous rate of change**, or in the context of a position function, the **instantaneous velocity**. To find the instantaneous velocity of the ball when $t = 1$, we need to know what happens to $g(x)$ as x gets closer and closer to 1. But also notice that $g(1)$ is not defined, because it leads to the quotient $0/0$.

This is where the notion of a *limit* comes in. By using a limit, we can investigate the behavior of $g(x)$ as x gets arbitrarily close, but not equal, to 1. We first use the graph of a function to explore points where interesting behavior occurs.

Warm-Up 2.1.1 Suppose that g is the function given by the graph below. Use the graph in Figure 2.1.1 to answer each of the following questions.

a. Determine the values $g(-2)$, $g(-1)$, $g(0)$, $g(1)$, and $g(2)$, if defined. If the function value is not defined, explain what feature of the graph tells you this.

b. For each of the values $a = -1$, $a = 0$, and $a = 2$, complete the following sentence: "As x gets closer and closer (but not equal) to a, $g(x)$ gets as close as we want to ______."

c. What happens as x gets closer and closer (but not equal) to $a = 1$? Does the function $g(x)$ get as close as we would like to a single value?

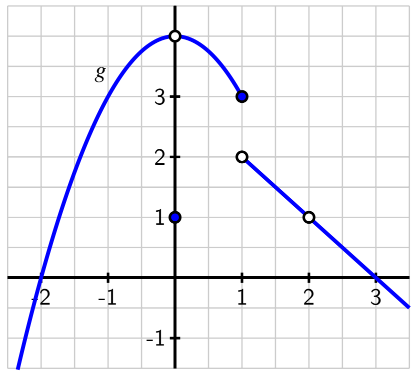


Figure 2.1.1 Graph of $y = g(x)$ for Warm-Up 2.1.1.

2.1.1 The Notion of Limit

Limits give us a way to identify a trend in the values of a function as its input variable approaches a particular value of interest. We need a precise understanding of what it means to say "a function f has limit L as x approaches a." To begin, think about a recent example.

In Warm-Up 2.1.1, we saw that as x gets closer and closer (but not equal) to 0, $g(x)$ gets as close as we want to the value 4. At first, this may feel counterintuitive, because the value of $g(0)$ is 1, not 4. But limits describe the behavior of a function *arbitrarily close to* a fixed input, and the value of the function *at* the fixed input does not matter. More formally,[1] we say the following.

Definition 2.1.2 Given a function f, a fixed input $x = a$, and a real number L, we say that f **has limit** L **as** x **approaches** a, and write

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

provided that we can make $f(x)$ as close to L as we like by taking x sufficiently close (but not equal) to a. If we cannot make $f(x)$ as close to a single value as we would like as x approaches a, then we say that f **does not have a limit as** x **approaches** a. ◊

Example 2.1.3 For the function g pictured in Figure 2.1.1, we make the

[1] What follows here is not what mathematicians consider the formal definition of a limit. To be completely precise, it is necessary to quantify both what it means to say "as close to L as we like" and "sufficiently close to a." That can be accomplished through what is traditionally called the epsilon-delta definition of limits. The definition presented here is sufficient for the purposes of this text.

following observations:

$$\lim_{x\to -1} g(x) = 3, \ \lim_{x\to 0} g(x) = 4, \text{ and } \lim_{x\to 2} g(x) = 1.$$

When working from a graph, it suffices to ask if the function *approaches* a single value from each side of the fixed input. The function value *at* the fixed input is irrelevant. This reasoning explains the values of the three limits stated above.

However, g does not have a limit as $x \to 1$. There is a jump in the graph at $x = 1$. If we approach $x = 1$ from the left, the function values tend to get close to 3, but if we approach $x = 1$ from the right, the function values get close to 2. There is no single number that all of these function values approach. This is why the limit of g does not exist at $x = 1$. □

Example 2.1.3 shows that when considering the limit of a function at a point, we must consider how the function behaves from the different sides of that point. Therefore, we introduce the notion of *left* and *right* (or *one-sided*) limits.

Definition 2.1.4 We say that f *has limit* L_1 *as* x *approaches* a *from the left* and write

$$\lim_{x\to a^-} f(x) = L_1$$

provided that we can make the value of $f(x)$ as close to L_1 as we like by taking x sufficiently close to a while always having $x < a$. We call L_1 the left-hand limit of f as x approaches a. Similarly, we say L_2 *is the right-hand limit of* f *as* x *approaches* a and write

$$\lim_{x\to a^+} f(x) = L_2$$

provided that we can make the value of $f(x)$ as close to L_2 as we like by taking x sufficiently close to a while always having $x > a$. ◊

In the graph of the function f in Figure 2.1.5, we see that

$$\lim_{x\to 1^-} f(x) = 2 \text{ and } \lim_{x\to 1^+} f(x) = 3.$$

Precisely because the left and right limits are not equal, the overall limit of f as $x \to 1$ fails to exist.

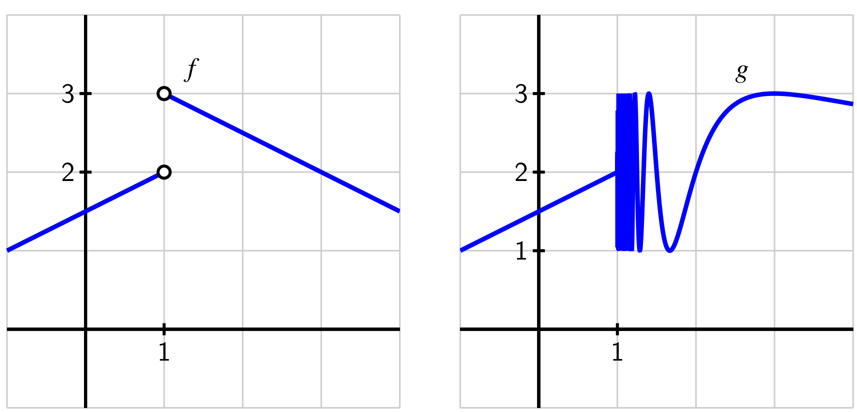


Figure 2.1.5 Functions f and g that each fail to have a limit at $a = 1$.

For the function g pictured at right in Figure 2.1.5, the function fails to have a limit at $a = 1$ for a different reason. While the function does not have a jump in its graph at $a = 1$, it is still not the case that g approaches a single value as x approaches 1. In particular, due to the infinitely oscillating behavior of g to the right of $a = 1$, we say that the right-hand limit of g as $x \to 1^+$ does not exist, and thus $\lim_{x\to 1} g(x)$ does not exist.

To summarize, if either a left- or right-hand limit fails to exist or if the left- and right-hand limits are not equal to each other, the overall limit does not exist.

> A function f has limit L as $x \to a$ if and only if
>
> $$\lim_{x\to a^-} f(x) = L = \lim_{x\to a^+} f(x).$$

For any function f, there are typically three ways to answer the question "does f have a limit at $x = a$, and if so, what is the limit?" The first is to reason graphically as we have just done with the example from Warm-Up 2.1.1. If we have a formula for $f(x)$, there are two additional possibilities:

1 Evaluate the function at a sequence of inputs that approach a on either side (typically using some sort of computing technology), and ask if the sequence of outputs seems to approach a single value.

2 Use the algebraic form of the function to understand the trend in its output values as the input values approach a.

The first approach produces only an approximation of the value of the limit, while the latter can often be used to determine the limit exactly.

Example 2.1.6 Limits of Two Functions. For each of the following functions, we'd like to know whether or not the function has a limit at the stated a-values. Use both numerical and algebraic approaches to investigate and, if possible, estimate or determine the value of the limit. Compare the results with a careful graph of the function on an interval containing the points of interest.

a. $f(x) = \frac{4-x^2}{x+2}$; $a = -1$, $a = -2$ b. $g(x) = \sin\left(\frac{\pi}{x}\right)$; $a = 3$, $a = 0$

Solution. a. We first construct a graph of f along with tables of values near $a = -1$ and $a = -2$.

x	$f(x)$
-0.9	2.9
-0.99	2.99
-0.999	2.999
-0.9999	2.9999
-1.1	3.1
-1.01	3.01
-1.001	3.001
-1.0001	3.0001

Table 2.1.7 Table of f values near $x = -1$.

x	$f(x)$
-1.9	3.9
-1.99	3.99
-1.999	3.999
-1.9999	3.9999
-2.1	4.1
-2.01	4.01
-2.001	4.001
-2.0001	4.0001

Table 2.1.8 Table of f values near $x = -2$.

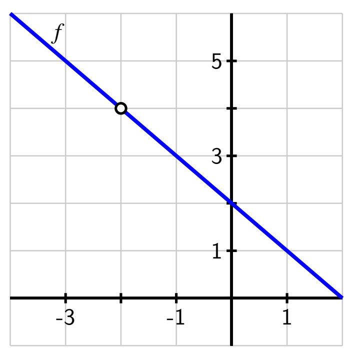


Figure 2.1.9 Plot of $f(x)$ on $[-4, 2]$.

From Table 2.1.7, it appears that we can make f as close as we want to 3 by taking x sufficiently close to -1, which suggests that $\lim_{x\to -1} f(x) = 3$. This is also consistent with the graph of f. To see this a bit more rigorously and from an algebraic point of view, consider the formula for f: $f(x) = \frac{4-x^2}{x+2}$. As $x \to -1$, $(4 - x^2) \to (4 - (-1)^2) = 3$, and $(x + 2) \to (-1 + 2) = 1$, so as $x \to -1$, the numerator of f tends to 3 and the denominator tends to 1, hence $\lim_{x \to -1} f(x) = \frac{3}{1} = 3$.

The situation is more complicated when $x \to -2$, because $f(-2)$ is not defined. If we try to use a similar algebraic argument regarding the numerator and denominator, we observe that as $x \to -2$, $(4 - x^2) \to (4 - (-2)^2) = 0$, and $(x + 2) \to (-2 + 2) = 0$, so as $x \to -2$, the numerator and denominator of f both tend to 0. We call $0/0$ an *indeterminate form*. This tells us that there is somehow more work to do. From Table 2.1.8 and Figure 2.1.9, it appears that f should have a limit of 4 at $x = -2$.

To see algebraically why this is the case, observe that

$$\begin{aligned}\lim_{x \to -2} f(x) &= \lim_{x \to -2} \frac{4 - x^2}{x + 2} \\ &= \lim_{x \to -2} \frac{(2 - x)(2 + x)}{x + 2}.\end{aligned}$$

It is important to observe that, since we are taking the limit as $x \to -2$, we are considering x values that are close, but not equal, to -2. Because we never actually allow x to equal -2, the quotient $\frac{2+x}{x+2}$ has value 1 for every possible value of x. Thus, we can simplify the most recent expression above, and find that

$$\lim_{x \to -2} f(x) = \lim_{x \to -2} 2 - x.$$

This limit is now easy to determine, and its value clearly is 4. Thus, from several points of view we've seen that $\lim_{x \to -2} f(x) = 4$.

b. Next we turn to the function g, and construct two tables and a graph.

Table 2.1.11

$gx = 3$

x	$g(x)$
2.9	0.84864
2.99	0.86428
2.999	0.86585
2.9999	0.86601
3.1	0.88351
3.01	0.86777
3.001	0.86620
3.0001	0.86604

Table 2.1.11

x	$g(x)$
-0.1	0
-0.01	0
-0.001	0
-0.0001	0
0.1	0
0.01	0
0.001	0
0.0001	0

Table 2.1.12 Table of g values near $x = 0$.

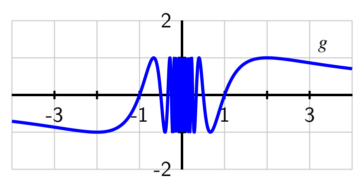


Figure 2.1.13 Plot of $g(x)$ on $[-4, 4]$.

First, as $x \to 3$, it appears from the table values that the function is approaching a number between 0.86601 and 0.86604. From the graph it appears that $g(x) \to g(3)$ as $x \to 3$. The exact value of $g(3) = \sin(\frac{\pi}{3})$ is $\frac{\sqrt{3}}{2}$, which is approximately 0.8660254038. This is convincing evidence that

$$\lim_{x \to 3} g(x) = \frac{\sqrt{3}}{2}.$$

As $x \to 0$, we observe that $\frac{\pi}{x}$ does not behave in an elementary way. When x is positive and approaching zero, we are dividing by smaller and smaller positive values, and $\frac{\pi}{x}$ increases without bound. When x is negative and approaching zero, $\frac{\pi}{x}$ decreases without bound. In this sense, as we get close to $x = 0$, the inputs to the sine function are growing rapidly, and this leads to increasingly rapid oscillations in the graph of g betweem 1 and -1. If we plot the function

$g(x) = \sin\left(\frac{\pi}{x}\right)$ with a graphing utility and then zoom in on $x = 0$, we see that the function never settles down to a single value near the origin, which suggests that g does not have a limit at $x = 0$.

How do we reconcile the graph with the righthand table above, which seems to suggest that the limit of g as x approaches 0 may in fact be 0? The data misleads us because of the special nature of the sequence of input values $\{0.1, 0.01, 0.001, \ldots\}$. When we evaluate $g(10^{-k})$, we get $g(10^{-k}) = \sin\left(\frac{\pi}{10^{-k}}\right) = \sin(10^k\pi) = 0$ for each positive integer value of k. But if we take a different sequence of values approaching zero, say $\{0.3, 0.03, 0.003, \ldots\}$, then we find that

$$g(3 \cdot 10^{-k}) = \sin\left(\frac{\pi}{3 \cdot 10^{-k}}\right) = \sin\left(\frac{10^k\pi}{3}\right) = \frac{\sqrt{3}}{2} \approx 0.866025.$$

That sequence of function values suggests that the value of the limit is $\frac{\sqrt{3}}{2}$. Clearly the function cannot have two different values for the limit, so g has no limit as $x \to 0$. □

An important lesson to take from Example 2.1.6 is that tables can be misleading when determining the value of a limit. While a table of values is useful for investigating the possible value of a limit, we should also use other tools to confirm the value.

Activity 2.1.2 Estimate the value of each of the following limits by constructing appropriate tables of values. Then determine the exact value of the limit by using algebra to simplify the function. Finally, plot each function on an appropriate interval to check your result visually.

a. $\displaystyle\lim_{x\to 1} \frac{x^2-1}{x-1}$ b. $\displaystyle\lim_{x\to 0} \frac{(2+x)^2-4}{x}$ c. $\displaystyle\lim_{x\to 0} \frac{\sqrt{x+1}-1}{x}$

Recall that our primary motivation for considering limits of functions comes from our interest in studying the rate of change of a function. To that end, we close this section by revisiting our previous work with average and instantaneous velocity and highlighting the role that limits play.

2.1.2 Instantaneous Velocity

Suppose that we have a moving object whose position at time t is given by a function s. We know that the average velocity of the object on the time interval $[a, b]$ is $AROC_{[a,b]} = \frac{s(b)-s(a)}{b-a}$. We define the *instantaneous velocity* at a to be the limit of average velocity as b approaches a. Note particularly that as $b \to a$, the length of the time interval gets shorter and shorter (while always including a). We will write $IROC_{t=a}$ for the instantaneous velocity at $t = a$, and thus

$$IROC_{t=a} = \lim_{b\to a} AROC_{[a,b]} = \lim_{b\to a} \frac{s(b)-s(a)}{b-a}.$$

Equivalently, if we think of the changing value b as being of the form $b = a + h$, where h is some small number, then we may instead write

$$IROC_{t=a} = \lim_{h\to 0} AROC_{[a,a+h]} = \lim_{h\to 0} \frac{s(a+h)-s(a)}{h}.$$

Again, the most important idea here is that to compute instantaneous velocity, we take a limit of average velocities as the time interval shrinks.

Activity 2.1.3 Consider a moving object whose position function is given by $s(t) = t^2$, where s is measured in meters and t is measured in minutes.

a. Determine the most simplified expression for the average velocity of the object on the interval $[3, 3 + h]$, where $h > 0$.

b. Determine the average velocity of the object on the interval $[3, 3.2]$. Include units on your answer.

c. Determine the instantaneous velocity of the object when $t = 3$. Include units on your answer.

The closing activity of this section asks you to make some connections among average velocity, instantaneous velocity, and slopes of certain lines.

Activity 2.1.4 For the moving object whose position s at time t is given by the graph in Figure 2.1.14, answer each of the following questions. Assume that s is measured in feet and t is measured in seconds.

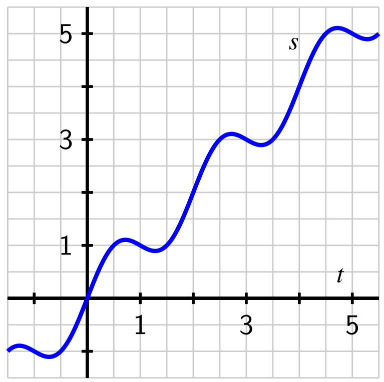


Figure 2.1.14 Plot of the position function $y = s(t)$ in Activity 2.1.4.

a. Use the graph to estimate the average velocity of the object on each of the following intervals: $[0.5, 1]$, $[1.5, 2.5]$, $[0, 5]$. Draw each line whose slope represents the average velocity you seek.

b. How could you use average velocities or slopes of lines to estimate the instantaneous velocity of the object at a fixed time?

c. Use the graph to estimate the instantaneous velocity of the object when $t = 2$. Should this instantaneous velocity at $t = 2$ be greater or less than the average velocity on $[1.5, 2.5]$ that you computed in (a)? Why?

2.1.3 Summary

- **Question 2.1.15** What is the mathematical notion of *limit* and what role do limits play in the study of functions? □

  Limits enable us to examine trends in function behavior near a specific point. In particular, taking a limit at a given point asks if the function values nearby tend to approach a particular fixed value.

- **Question 2.1.16** What is the meaning of the notation $\lim_{x \to a} f(x) = L$? □

  We read $\lim_{x \to a} f(x) = L$, as "the limit of f as x approaches a is L," which means that we can make the value of $f(x)$ as close to L as we want by taking x sufficiently close (but not equal) to a.

- **Question 2.1.17** How do we go about determining the value of the limit of a function at a point? □

To find $\lim_{x\to a} f(x)$ for a given value of a and a known function f, we can estimate this value from the graph of f, or we can make a table of function values for x-values that are closer and closer to a. If we want the exact value of the limit, we can work with the function algebraically to understand how different parts of the formula for f change as $x \to a$.

- **Question 2.1.18** What is instantaneous velocity and how do we compute it? □

  We find the instantaneous velocity of a moving object at a fixed time by taking the limit of average velocities of the object over shorter and shorter time intervals containing the time of interest.

2.1.4 Exercises

1. Consider the function whose formula is $f(x) = \frac{16-x^4}{x^2-4}$.

a. What is the domain of f?

b. Use a sequence of values of x near $a = 2$ to estimate the value of $\lim_{x\to 2} f(x)$, if you think the limit exists. If you think the limit doesn't exist, explain why.

c. Use algebra to simplify the expression $\frac{16-x^4}{x^2-4}$ and hence work to evaluate $\lim_{x\to 2} f(x)$ exactly, if it exists, or to explain how your work shows the limit fails to exist. Discuss how your findings compare to your results in (b).

d. True or false: $f(2) = -8$. Why?

e. True or false: $\frac{16-x^4}{x^2-4} = -4-x^2$. Why? How is this equality connected to your work above with the function f?

f. Based on all of your work above, construct an accurate, labeled graph of $y = f(x)$ on the interval $[1, 3]$, and write a sentence that explains what you now know about $\lim_{x\to 2} \frac{16-x^4}{x^2-4}$.

2. Let $g(x) = -\frac{|x+3|}{x+3}$.

a. What is the domain of g?

b. Use a sequence of values near $a = -3$ to estimate the value of $\lim_{x\to -3} g(x)$, if you think the limit exists. If you think the limit doesn't exist, explain why.

c. Use algebra to simplify the expression $\frac{|x+3|}{x+3}$ and hence work to evaluate $\lim_{x\to -3} g(x)$ exactly, if it exists, or to explain how your work shows the limit fails to exist. Discuss how your findings compare to your results in (b). (Hint: $|a| = a$ whenever $a \geq 0$, but $|a| = -a$ whenever $a < 0$.)

d. True or false: $g(-3) = -1$. Why?

e. True or false: $-\frac{|x+3|}{x+3} = -1$. Why? How is this equality connected to your work above with the function g?

f. Based on all of your work above, construct an accurate, labeled graph of $y = g(x)$ on the interval $[-4, -2]$, and write a sentence that explains what you now know about $\lim_{x\to -3} g(x)$.

3. For each of the following prompts, sketch a graph on the provided axes of a function that has the stated properties.

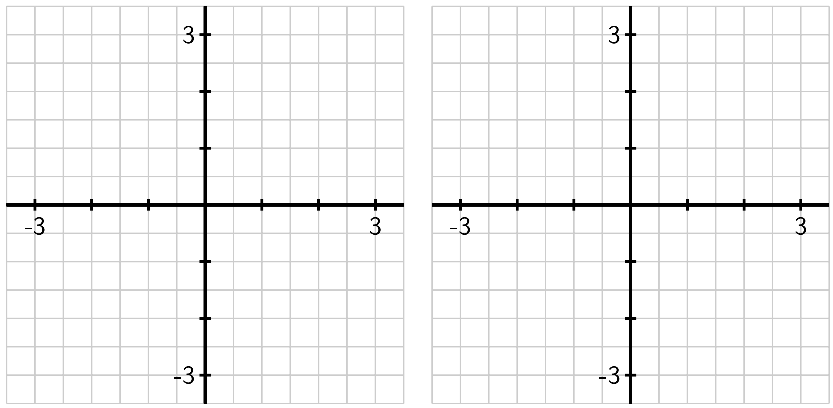


Figure 2.1.19 Axes for plotting $y = f(x)$ in (a) and $y = g(x)$ in (b).

a. $y = f(x)$ such that

- $f(-2) = 2$ and $\lim_{x\to-2} f(x) = 1$
- $f(-1) = 3$ and $\lim_{x\to-1} f(x) = 3$
- $f(1)$ is not defined and $\lim_{x\to1} f(x) = 0$
- $f(2) = 1$ and $\lim_{x\to2} f(x)$ does not exist.

b. $y = g(x)$ such that

- $g(-2) = 3$, $g(-1) = -1$, $g(1) = -2$, and $g(2) = 3$
- At $x = -2, -1, 1$ and 2, g has a limit, and its limit equals the value of the function at that point.
- $g(0)$ is not defined and $\lim_{x\to0} g(x)$ does not exist.

4. A bungee jumper dives from a tower at time $t = 0$. Her height s in feet at time t in seconds is given by $s(t) = 100\cos(0.75t) \cdot e^{-0.2t} + 100$.

a. Write an expression for the average velocity of the bungee jumper on the interval $[1, 1 + h]$.

b. Use computing technology to estimate the value of the limit as $h \to 0$ of the quantity you found in (a).

c. What is the meaning of the value of the limit in (b)? What are its units?

2.2 The Derivative of a Function at a Point

Motivating Questions

- How is the average rate of change of a function on a given interval defined, and what does this quantity measure?
- How is the instantaneous rate of change of a function at a particular point defined? How is the instantaneous rate of change linked to average rate of change?

- What is the derivative of a function at a given point? What does this derivative value measure? How do we interpret the derivative value graphically?
- How are limits used formally in the computation of derivatives?

The *instantaneous rate of change* of a function is an idea that sits at the foundation of calculus. It is a generalization of the notion of instantaneous velocity and measures how fast a particular function is changing at a given point. If the original function represents the position of a moving object, this instantaneous rate of change is precisely the velocity of the object. In other contexts, instantaneous rate of change could measure the number of cells added to a bacteria culture per day, the number of additional gallons of gasoline consumed by increasing a car's velocity one mile per hour, or the number of dollars added to a mortgage payment for each percentage point increase in interest rate. The instantaneous rate of change can also be interpreted geometrically on the function's graph, and this connection is fundamental to many of the main ideas in calculus.

Recall that for a function $f(x)$ on a domain $[a, b]$, the **average rate of change between a and b** is given by

$$AROC_{[a,b]} = \frac{f(b) - f(a)}{b - a}.$$

In a similar way as in Subsection 2.1.2, if we write the second x value b as $a + h$ (where $|h|$ can be thought of as the distance of b from a), we get the following form of the average rate of change of a function between two points:

Definition 2.2.1 For a function f, the **average rate of change** of f on the interval $[a, a + h]$ is given by the value

$$AROC_{[a,a+h]} = \frac{f(a + h) - f(a)}{h}.$$

This form of the average rate of change of a function is called the **difference quotient**. ♢

A benefit of using the difference quotient to write an average rate of change is that it allows us to place focus on a single x (the x value a) instead of two x values. Our ultimate goal, as we'll explore below, is to describe how a function is changing at a single x value.

It is essential that you understand how the average rate of change of f on an interval is connected to its graph.

Warm-Up 2.2.1 Suppose that f is the function given by the graph below and that a and $a + h$ are the input values as labeled on the x-axis. Use the graph in Figure 2.2.2 to answer the following questions.

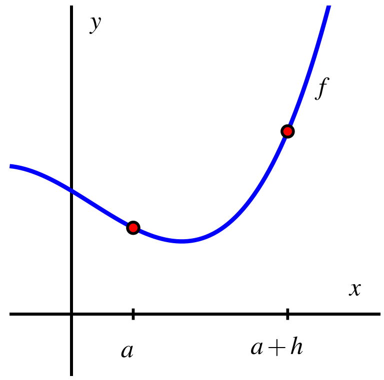


Figure 2.2.2 Plot of $y = f(x)$ for Warm-Up 2.2.1.

a. Locate and label the points $(a, f(a))$ and $(a+h, f(a+h))$ on the graph.

b. Construct a right triangle whose hypotenuse is the line segment from $(a, f(a))$ to $(a+h, f(a+h))$. What are the lengths of the respective legs of this triangle?

c. What is the slope of the line that connects the points $(a, f(a))$ and $(a+h, f(a+h))$?

d. Write a meaningful sentence that explains how the average rate of change of the function on a given interval and the slope of a related line are connected.

2.2.1 The Derivative of a Function at a Point

Just as we defined instantaneous velocity in terms of average velocity in Subsection 2.1.2, we now define the instantaneous rate of change of a function at a point in terms of the average rate of change of the function f over related intervals. This instantaneous rate of change of f at a is called "the *derivative* of f at a," and is denoted by $f'(a)$.

Definition 2.2.3 Let f be a function and $x = a$ a value in the function's domain. We define the **derivative of f with respect to x evaluated at** $x = a$, denoted $f'(a)$, by the formula

$$f'(a) = \lim_{h \to 0} AROC_{[a,a+h]} = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

provided this limit exists. ◊

Aloud, we read the symbol $f'(a)$ as either "f-prime at a" or "the derivative of f evaluated at $x = a$." Using notation from Subsection 2.1.2, the derivative at a could also be written as $IROC_{t=a}$. However, moving forward we will default to using the prime notation for derivative.

Much of the next several chapters will be devoted to understanding, computing, applying, and interpreting derivatives. For now, we observe the following important things.

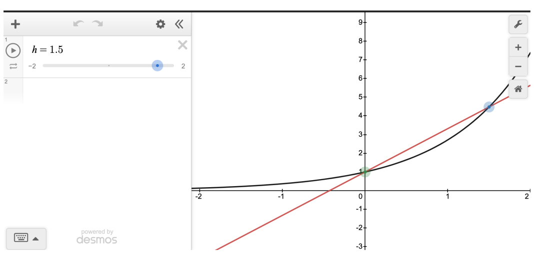
Note 2.2.4

- The derivative of f at the value $x = a$ is defined as the limit of the average rate of change of f on the interval $[a, a+h]$ as $h \to 0$. This limit may not exist, so not every function has a derivative at every point.

- We say that a function is *differentiable* at $x = a$ if it has a derivative at $x = a$.
- The derivative is a generalization of the instantaneous velocity of a position function: if $y = s(t)$ is a position function of a moving body, $s'(a)$ tells us the instantaneous velocity of the body at time $t = a$.
- Because the units on $\frac{f(a+h)-f(a)}{h}$ are "units of $f(x)$ per unit of x," the derivative has these very same units. For instance, if s measures position in feet and t measures time in seconds, the units on $s'(a)$ are feet per second.
- Because the quantity $\frac{f(a+h)-f(a)}{h}$ represents the slope of the line through $(a, f(a))$ and $(a+h, f(a+h))$, when we compute the derivative we are taking the limit of a collection of slopes of lines. Thus, the derivative itself represents the slope of a particularly important line.

We first consider the derivative at a given value as the slope of a certain line.

When we compute an instantaneous rate of change, we allow the interval $[a, a+h]$ to shrink as $h \to 0$. We can think of one endpoint of the interval as "sliding towards" the other. In particular, provided that f has a derivative at $(a, f(a))$, the point $(a+h, f(a+h))$ will approach $(a, f(a))$ as $h \to 0$. Because the process of taking a limit is a dynamic one, it can be helpful to use computing technology to visualize it. Use the interactive below to change the distance h from the x value $a = 0$. Try to verbalize what the line connecting the two points starts to look like as h gets very close (but not equal to) 0.



www.desmos.com/calculator/eb9qn2xyge

Figure 2.2.5 shows a sequence of figures with several different lines through the points $(a, f(a))$ and $(a+h, f(a+h))$, generated by different values of h. These lines (shown in the first three figures in magenta), are often called **secant lines** to the curve $y = f(x)$. A secant line to a curve is simply a line that passes through two points on the curve. For each such line, the slope of the secant line is $m = \frac{f(a+h)-f(a)}{h}$, where the value of h depends on the location of the point we choose. We can see in the diagram how, as $h \to 0$, the secant lines start to approach a single line that passes through the point $(a, f(a))$. If the limit of the slopes of the secant lines exists, we say that the resulting value is the slope of the **tangent line** to the curve. This tangent line (shown in the right-most figure in green) to the graph of $y = f(x)$ at the point $(a, f(a))$ has slope $m = f'(a)$.

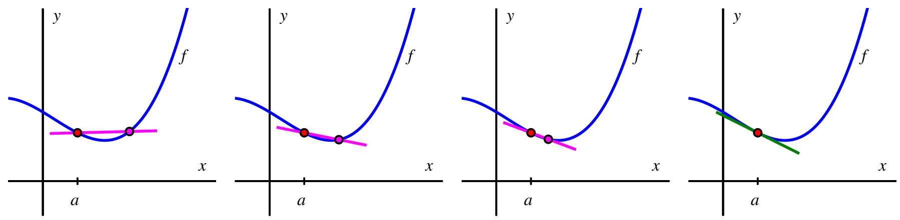


Figure 2.2.5 A sequence of secant lines approaching the tangent line to f at $(a, f(a))$.

If the tangent line at $x = a$ exists, the graph of f looks like a straight line when viewed up close at $(a, f(a))$. In Figure 2.2.6 we combine the four graphs in Figure 2.2.5 into the single one on the left, and zoom in on the box centered at $(a, f(a))$ on the right. Note how the tangent line sits relative to the curve $y = f(x)$ at $(a, f(a))$ and how closely it resembles the curve near $x = a$.

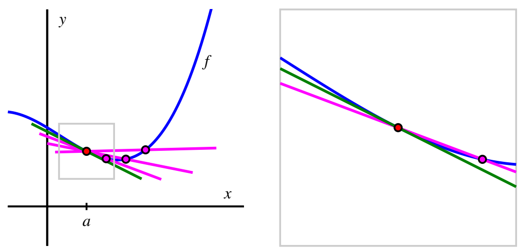


Figure 2.2.6 A sequence of secant lines approaching the tangent line to f at $(a, f(a))$. At right, we zoom in on the point $(a, f(a))$. The slope of the tangent line (in green) to f at $(a, f(a))$ is given by $f'(a)$.

Note 2.2.7 The instantaneous rate of change of f with respect to x at $x = a$, $f'(a)$, also measures the slope of the tangent line to the curve $y = f(x)$ at $(a, f(a))$.

The following example demonstrates several key ideas involving the derivative of a function.

Example 2.2.8 Using the limit definition of the derivative. For the function $f(x) = x - x^2$, use the limit definition of the derivative to compute $f'(2)$. In addition, discuss the meaning of this value and draw a labeled graph that supports your explanation.

Solution. From the limit definition, we know that

$$f'(2) = \lim_{h \to 0} \frac{f(2+h) - f(2)}{h}.$$

Now we use the rule for f, and observe that $f(2) = 2 - 2^2 = -2$ and $f(2+h) = (2+h) - (2+h)^2$. Substituting these values into the limit definition, we have that

$$f'(2) = \lim_{h \to 0} \frac{(2+h) - (2+h)^2 - (-2)}{h}.$$

In order to let $h \to 0$, we must simplify the quotient. Expanding and distributing in the numerator,

$$f'(2) = \lim_{h \to 0} \frac{2 + h - 4 - 4h - h^2 + 2}{h}.$$

Combining like terms, we have

$$f'(2) = \lim_{h \to 0} \frac{-3h - h^2}{h}.$$

Next, we remove a common factor of h in both the numerator and denominator and find that

$$f'(2) = \lim_{h \to 0} (-3 - h).$$

Finally, we are able to take the limit as $h \to 0$, and thus conclude that $f'(2) = -3$. We note that $f'(2)$ is the instantaneous rate of change of f at the point $(2, -2)$. It is also the slope of the tangent line to the graph of $y = x - x^2$ at the point $(2, -2)$. Figure 2.2.9 shows both the function and the line through $(2, -2)$ with slope $m = f'(2) = -3$.

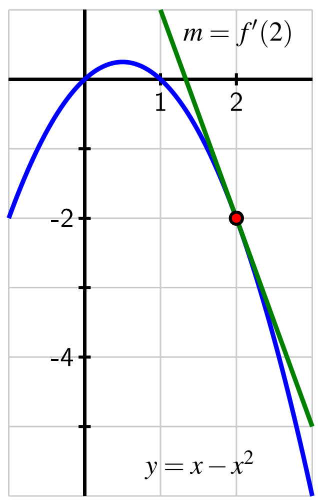


Figure 2.2.9 The tangent line to $y = x - x^2$ at the point $(2, -2)$.

□

The following activities will help you explore a variety of key ideas related to derivatives.

Activity 2.2.2 Consider the function f whose formula is $f(x) = 3 - 2x$.

a. What familiar type of function is f? What can you say about the slope of f at every value of x?

b. Compute the average rate of change of f on the intervals $[1, 4]$, $[3, 7]$, and $[5, 5 + h]$; simplify each result as much as possible. What do you notice about these quantities?

c. Use the limit definition of the derivative to compute the exact instantaneous rate of change of f with respect to x at the value $a = 1$. That is, compute $f'(1)$ using the limit definition. Show your work. Is your result surprising?

d. Without doing any additional computations, what are the values of $f'(2)$, $f'(\pi)$, and $f'(-\sqrt{2})$? Why?

Activity 2.2.3 A water balloon is tossed vertically in the air from a window.

The balloon's height in feet at time t in seconds after being launched is given by

$$\begin{aligned} s(t) &= -16t^2 + 16t + 32 \\ &= -16(t - 0.5)^2 + 36. \end{aligned}$$

Use this function to respond to each of the following questions.

a. Sketch an accurate, labeled graph of s on the axes provided in Figure 2.2.10. You should be able to do this without using computing technology.

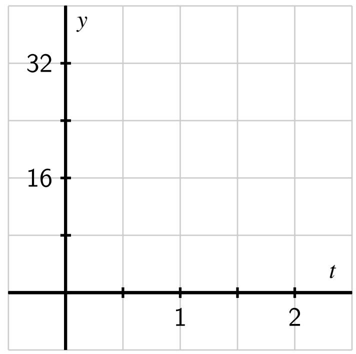


Figure 2.2.10 Axes for plotting $y = s(t)$ in Activity 2.2.3.

b. Compute the average rate of change of s on the time interval $[1, 2]$. Include units on your answer and write one sentence to explain the meaning of the value you found.

c. Use the limit definition to compute the instantaneous rate of change of s with respect to time, t, at the instant $a = 1$. Show your work using proper notation, include units on your answer, and write one sentence to explain the meaning of the value you found.

d. On your graph in (a), sketch two lines: one whose slope represents the average rate of change of s on $[1, 2]$, the other whose slope represents the instantaneous rate of change of s at the instant $a = 1$. Label each line clearly.

e. For what values of a do you expect $s'(a)$ to be positive? Why? Answer the same questions when "positive" is replaced by "negative" and "zero."

Activity 2.2.4 A rapidly growing city in Arizona has its population P at time t, where t is the number of decades after the year 2010, modeled by the formula $P(t) = 25000e^{t/5}$. Use this function to respond to the following questions.

a. Sketch a rough graph of P for $t = 0$ to $t = 5$ on the axes provided in Figure 2.2.11.

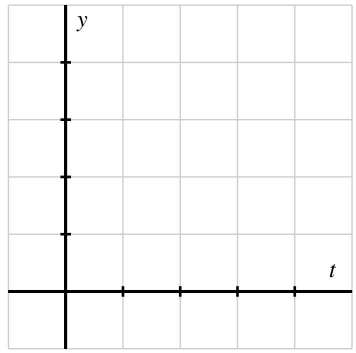


Figure 2.2.11 Axes for plotting $y = P(t)$ in Activity 2.2.4.

b. Compute the average rate of change of P between 2030 and 2050. Include units on your answer and write one sentence to explain the meaning (in everyday language) of the value you found.

c. Use the limit definition to write an expression for the instantaneous rate of change of P with respect to time, t, at the instant $a = 2$. Explain why this limit is difficult to evaluate exactly.

d. Estimate the limit in (c) for the instantaneous rate of change of P at the instant $a = 2$ by using several small h values. Once you have determined an accurate estimate of $P'(2)$, include units on your answer, and write one sentence (using everyday language) to explain the meaning of the value you found.

e. On your graph above, sketch two lines: one whose slope represents the average rate of change of P on $[2, 4]$, the other whose slope represents the instantaneous rate of change of P at the instant $a = 2$.

f. In a carefully-worded sentence, describe the behavior of $P'(a)$ as a increases in value. What does this reflect about the behavior of the given function P?

2.2.2 Summary

- **Question 2.2.12** How is the average rate of change of a function on a given interval defined, and what does this quantity measure? □

  The average rate of change of a function f on the interval $[a, b]$ is $\frac{f(b)-f(a)}{b-a}$. The units on the average rate of change are units of $f(x)$ per unit of x, and the numerical value of the average rate of change represents the slope of the secant line between the points $(a, f(a))$ and $(b, f(b))$ on the graph of $y = f(x)$. If we view the interval as being $[a, a+h]$ instead of $[a, b]$, the meaning is still the same, but the average rate of change is now computed using the difference quotient, $\frac{f(a+h)-f(a)}{h}$.

- **Question 2.2.13** How is the instantaneous rate of change of a function at a particular point defined? How is the instantaneous rate of change linked to average rate of change? □

  The instantaneous rate of change with respect to x of a function f at a value $x = a$ is denoted $f'(a)$ (read "the derivative of f evaluated at a" or

"f-prime at a") and is defined by the formula

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

provided the limit exists. Note particularly that the instantaneous rate of change at $x = a$ is the limit of the average rate of change on $[a, a+h]$ as $h \to 0$.

- **Question 2.2.14** What is the derivative of a function at a given point? What does this derivative value measure? How do we interpret the derivative value graphically? □

  Provided the derivative $f'(a)$ exists, its value tells us the instantaneous rate of change of f with respect to x at $x = a$, which geometrically is the slope of the tangent line to the curve $y = f(x)$ at the point $(a, f(a))$. We even say that $f'(a)$ is the "slope of the curve" at the point $(a, f(a))$.

- **Question 2.2.15** How are limits used formally in the computation of derivatives? □

  Limits allow us to move from the rate of change over an interval to the rate of change at a single point.

2.2.3 Exercises

1. Consider the graph of $y = f(x)$ provided in Figure 2.2.16.

a. On the graph of $y = f(x)$, sketch and label the following quantities:

- the secant line to $y = f(x)$ on the interval $[-3, -1]$ and the secant line to $y = f(x)$ on the interval $[0, 2]$.
- the tangent line to $y = f(x)$ at $x = -3$ and the tangent line to $y = f(x)$ at $x = 0$.

b. What is the approximate value of the average rate of change of f on $[-3, -1]$? On $[0, 2]$? How are these values related to your work in (a)?

c. What is the approximate value of the instantaneous rate of change of f at $x = -3$? At $x = 0$? How are these values related to your work in (a)?

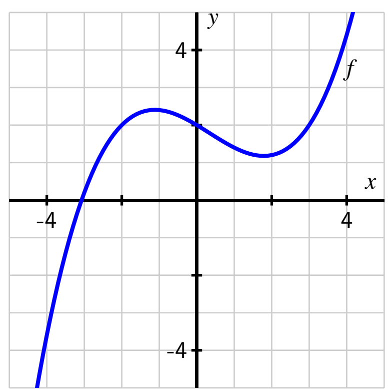


Figure 2.2.16 Plot of $y = f(x)$.

2. For each of the following prompts, sketch a graph on the provided axes in Figure 2.2.17 of a function that has the stated properties.

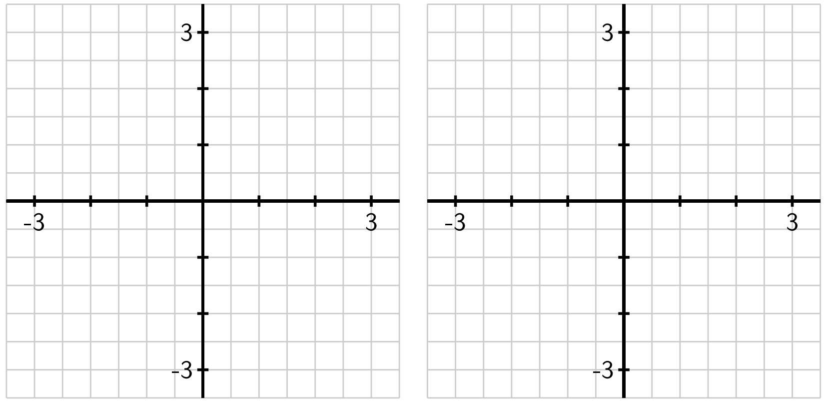


Figure 2.2.17 Axes for plotting $y = f(x)$ in (a) and $y = g(x)$ in (b).

a. $y = f(x)$ such that

- the average rate of change of f on $[-3, 0]$ is -2 and the average rate of change of f on $[1, 3]$ is 0.5, and
- the instantaneous rate of change of f at $x = -1$ is -1 and the instantaneous rate of change of f at $x = 2$ is 1.

b. $y = g(x)$ such that

- $\frac{g(3)-g(-2)}{5} = 0$ and $\frac{g(1)-g(-1)}{2} = -1$, and
- $g'(2) = 1$ and $g'(-1) = 0$

3. Suppose that the population, P, of China (in billions) can be approximated by the function $P(t) = 1.15(1.014)^t$ where t is the number of years since the start of 1993.

a. According to the model, what was the total change in the population of China between January 1, 1993 and January 1, 2000? What will be the average rate of change of the population over this time period? Is this average rate of change greater or less than the instantaneous rate of change of the population on January 1, 2000? Explain and justify, being sure to include proper units on all your answers.

b. According to the model, what is the average rate of change of the population of China in the ten-year period starting on January 1, 2012?

c. Write an expression involving limits that, if evaluated, would give the exact instantaneous rate of change of the population on today's date. Then estimate the value of this limit (discuss how you chose to do so) and explain the meaning (including units) of the value you have found.

d. Find an equation for the tangent line to the function $y = P(t)$ at the point where the t-value is given by today's date.

4. The goal of this problem is to compute the value of the derivative at a point for several different functions, where for each one we do so in three different ways, and then to compare the results to see that each produces the same value.

For each of the following functions, use the limit definition of the derivative to compute the value of $f'(a)$ using three different approaches: strive to use the algebraic approach first (to compute the limit exactly), then test your result using numerical evidence (with small values of h), and finally plot the graph of $y = f(x)$ near $(a, f(a))$ along with the appropriate tangent line to estimate the value of $f'(a)$ visually. Compare your findings among all three approaches; if you are unable to complete the algebraic approach, still work numerically and graphically.

a. $f(x) = x^2 - 3x$, $a = 2$

b. $f(x) = \frac{1}{x}$, $a = 1$

c. $f(x) = \sqrt{x}$, $a = 1$

d. $f(x) = 2 - |x - 1|$, $a = 1$

e. $f(x) = \sin(x)$, $a = \frac{\pi}{2}$

2.3 The Derivative Function

Motivating Questions

- How does the limit definition of the derivative of a function f lead to an entirely new (but related) function f'?
- What is the difference between writing $f'(a)$ and $f'(x)$?
- How is the graph of the derivative function $f'(x)$ related to the graph of $f(x)$?
- What are some examples of functions f for which f' is not defined at one or more points?

We now know that the instantaneous rate of change of a function $f(x)$ at $x = a$, or equivalently the slope of the tangent line to the graph of $y = f(x)$ at $x = a$, is given by the value $f'(a)$. In all of our examples so far, we have identified a particular value of a as our point of interest: $a = 1$, $a = 3$, etc. But it is not hard to imagine that we will often be interested in the derivative value for more than just one a-value, and possibly for many of them. In this section, we explore how we can move from computing the derivative at a single point to computing a formula for $f'(a)$ at any point a. Indeed, the process of "taking the derivative" generates a new function, denoted by $f'(x)$, derived from the original function $f(x)$.

Warm-Up 2.3.1 Consider the function $f(x) = 4 - x^2$.

a. Use the limit definition to compute the derivative values: $f'(0)$, $f'(1)$, $f'(2)$, and $f'(3)$.

b. Observe that the work to find $f'(a)$ is the same, regardless of the value of a. Based on your work in (a), what do you conjecture is the value of $f'(4)$? How about $f'(5)$? (Note: you should *not* use the limit definition of the derivative to find either value.)

c. Conjecture a formula for $f'(a)$ that depends only on the value a. That is, in the same way that we have a formula for $f(x)$ (recall $f(x) = 4 - x^2$), see if you can use your work above to guess a formula for $f'(a)$ in terms of a.

2.3.1 How the derivative is itself a function

In your work in Warm-Up 2.3.1 with $f(x) = 4x - x^2$, you may have found several patterns. One comes from observing that $f'(0) = 4$, $f'(1) = 2$, $f'(2) = 0$, and $f'(3) = -2$. That sequence of values leads us naturally to conjecture that $f'(4) = -4$ and $f'(5) = -6$. We also observe that the particular value of a has very little effect on the process of computing the value of the derivative through the limit definition. To see this more clearly, we compute $f'(a)$, where a represents a number to be named later. Following the now standard process of using the limit definition of the derivative,

$$\begin{aligned} f'(a) &= \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} = \lim_{h\to 0} \frac{4(a+h)-(a+h)^2-(4a-a^2)}{h} \\ &= \lim_{h\to 0} \frac{4a+4h-a^2-2ha-h^2-4a+a^2}{h} = \lim_{h\to 0} \frac{4h-2ha-h^2}{h} \\ &= \lim_{h\to 0} \frac{h(4-2a-h)}{h} = \lim_{h\to 0} (4-2a-h). \end{aligned}$$

Here we observe that neither 4 nor $2a$ depend on the value of h, so as $h \to 0$, $(4-2a-h) \to (4-2a)$. Thus, $f'(a) = 4-2a$.

This result is consistent with the specific values we found above: e.g., $f'(3) = 4-2(3) = -2$. And indeed, our work confirms that the value of a has almost no bearing on the process of computing the derivative. We note further that the letter being used is immaterial: whether we call it a, x, or anything else, the derivative at a given value is simply given by "4 minus 2 times the value." We choose to use x for consistency with the original function given by $y = f(x)$, as well as for the purpose of graphing the derivative function. For the function $f(x) = 4x - x^2$, it follows that $f'(x) = 4 - 2x$.

Because the value of the derivative function is linked to the graph of the original function, it makes sense to look at both of these functions plotted on the same domain.

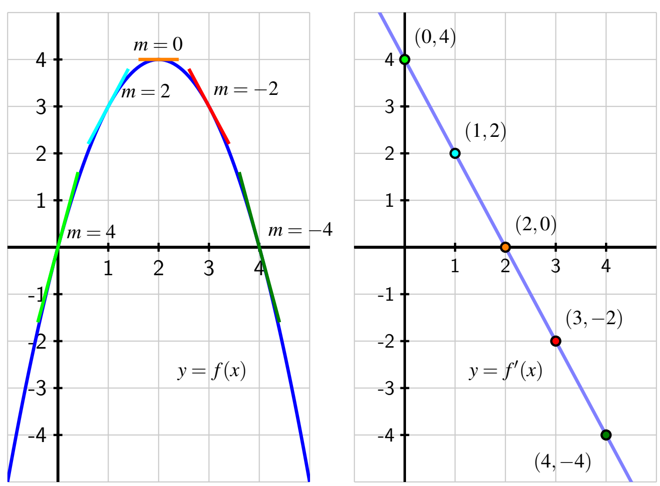
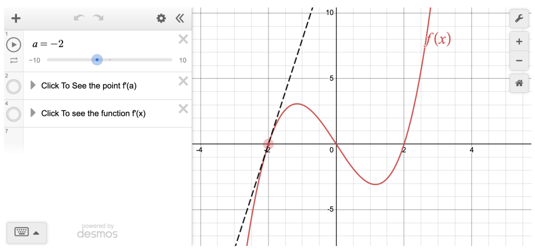


Figure 2.3.1 The graphs of $f(x) = 4x - x^2$ (at left) and $f'(x) = 4 - 2x$ (at right). Slopes on the graph of f correspond to heights on the graph of f'.

In Figure 2.3.1, on the left we show a plot of $f(x) = 4x - x^2$ together with a selection of tangent lines at the points we've considered above. On the right, we show a plot of $f'(x) = 4 - 2x$ with emphasis on the heights of the derivative graph at the same selection of points. Notice the connection between colors in the left and right graphs: the green tangent line on the original graph is tied to the green point on the right graph in the following way: *the slope of the tangent line* at a point on the lefthand graph is the same as the *height* at the corresponding point on the righthand graph. That is, at each respective value of x, the slope of the tangent line to the original function is the same as the height of the derivative function. Do note, however, that the units on the vertical axes are different: in the left graph, the vertical units are simply the output units of f. On the righthand graph of $y = f'(x)$, the units on the vertical axis are units of f per unit of x.

An excellent way to explore how the graph of $f(x)$ generates the graph of $f'(x)$ is through an interactive like the one below:



www.desmos.com/calculator/uvurolvyev

- Change the input value a, and describe how slopes of tangent lines to the graph of $f(x)$ change. Are they positive or negative? Are they big or small?

- Activate the folder to see the point $f'(a)$ plotted. These are points on the graph of $f'(x)$, and should match your descriptions from the previous part.

- Activate the folder to see the graph of $f'(x)$, and visualize how the slopes of tangent lines to the graph of $f(x)$ determine the output values of the graph of $f'(x)$.

In Section 2.2 when we first defined the derivative, we wrote the definition in terms of a value a to find $f'(a)$. As we have seen above, the letter a is merely a placeholder, and it often makes more sense to use x instead. For the record, here we restate the definition of the derivative.

Definition 2.3.2 Let f be a function and x a value in the function's domain. We define the **derivative of** f, a new function called f', by the formula $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$, provided this limit exists. ◊

We now have two different ways of thinking about the derivative function:

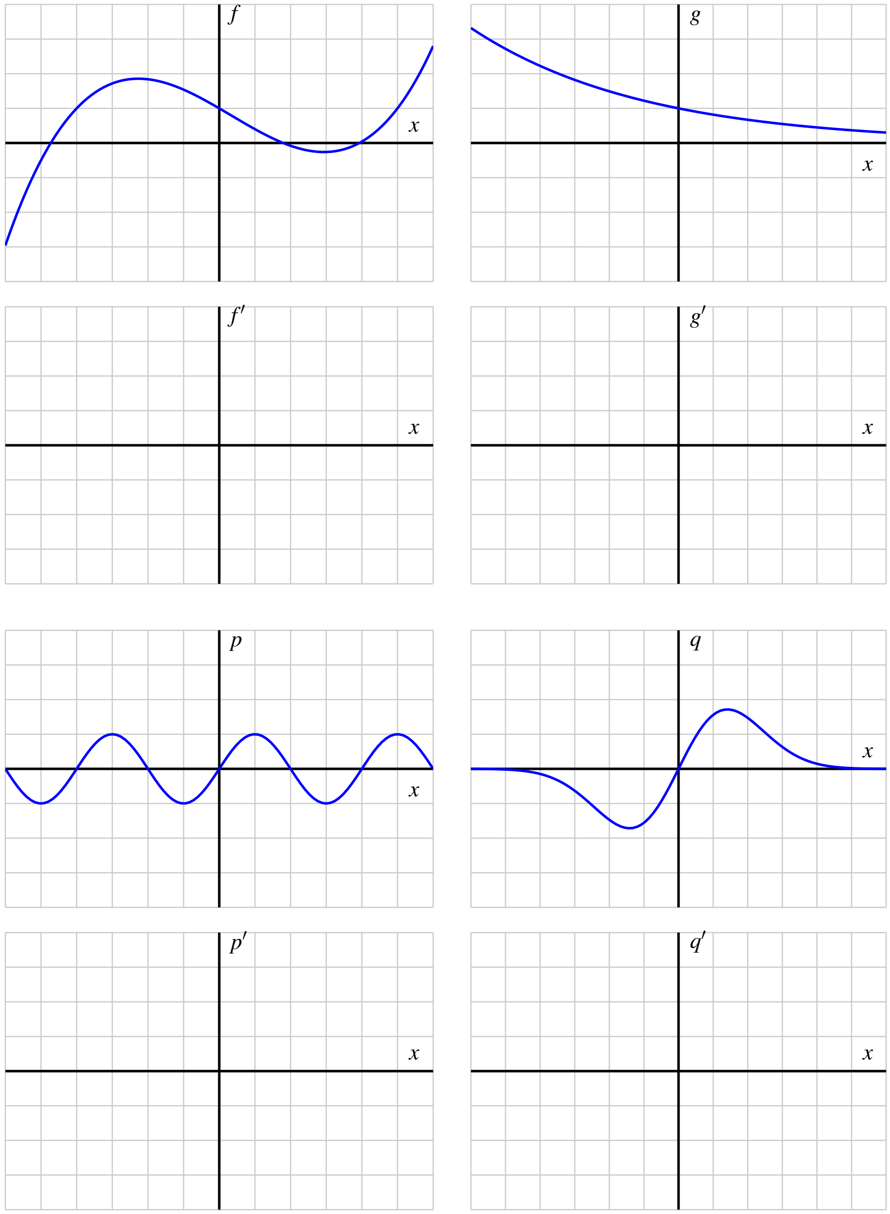
1 given a graph of $y = f(x)$, how does this graph lead to the graph of the derivative function $y = f'(x)$? and

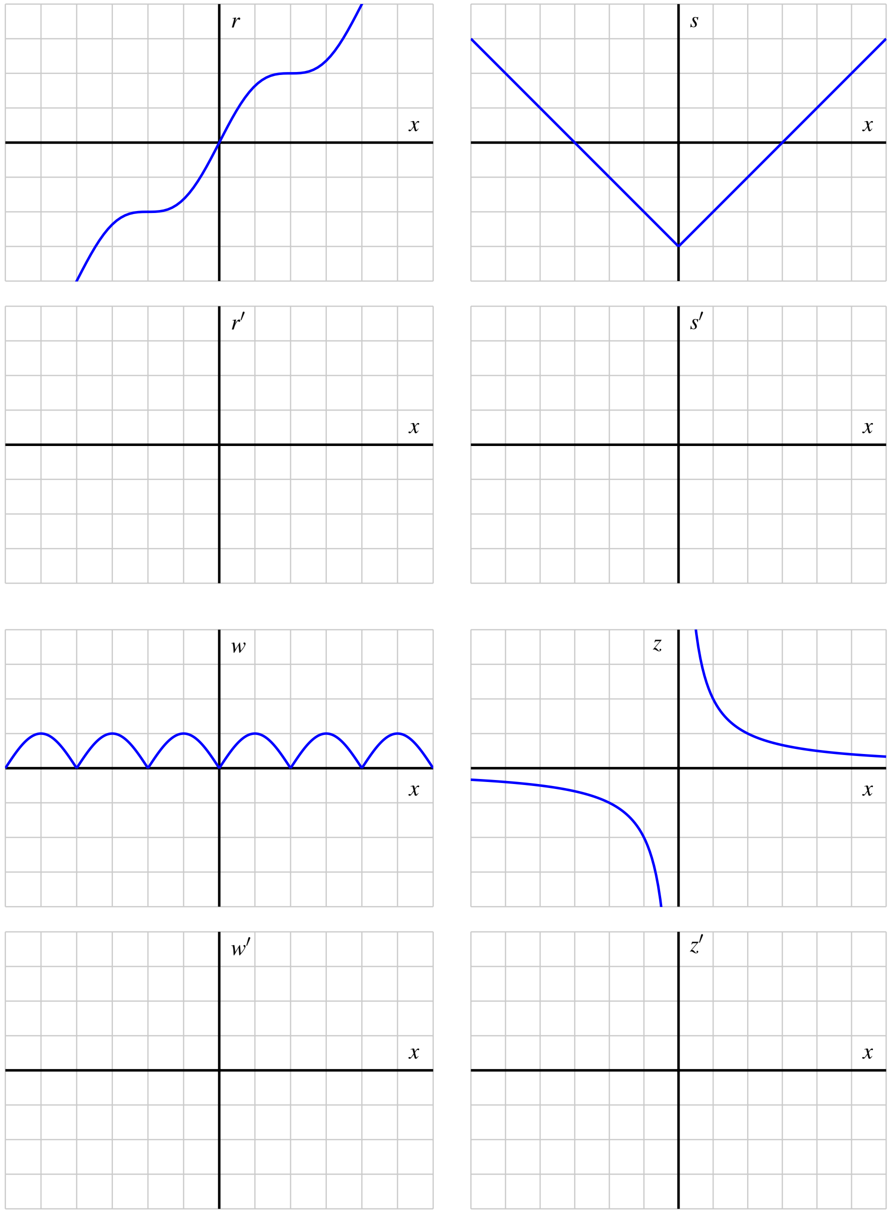
2 given a formula for $y = f(x)$, how does the limit definition of derivative generate a formula for $y = f'(x)$?

Both of these perspectives are explored in the following activities.

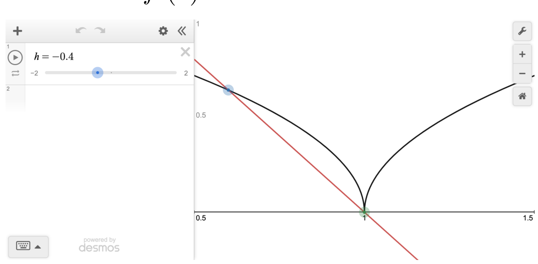
Activity 2.3.2 For each given graph of $y = f(x)$, sketch an approximate graph of its derivative function, $y = f'(x)$, on the axes immediately below. The scale of the vertical axis for the graph of f' does not need to be accurate.

When you are finished with all 8 graphs, write several sentences that describe your overall process for sketching the graph of the derivative function, given the graph of the original function. What are the values of the derivative function that you tend to identify first? What do you do thereafter? How do key traits of the graph of the derivative function exemplify properties of the graph of the original function?





Activity 2.3.2 included some graphs for which the associated derivative function was not defined for certain x values. The interactive below contains another example of a function $f(x)$ for which $f'(1)$ does not exist. Use the interactive to see how secant lines behave as the distance h from $x = 1$ gets closer to zero from both the left and the right. Compare this behavior to the interactive from Section 2.2, and write a thoughtful sentence explaining why it makes sense that $f'(1)$ does not exist.



www.desmos.com/calculator/vokjwfntsl

Now, recall the opening example of this section: we began with the function $y = f(x) = 4x - x^2$ and used the limit definition of the derivative to show that $f'(a) = 4 - 2a$, or equivalently that $f'(x) = 4 - 2x$. We subsequently graphed the functions f and f' as shown in Figure 2.3.1. Following Activity 2.3.2, we now understand that we could have constructed a fairly accurate graph of $f'(x)$ *without* knowing a formula for either f or f'. At the same time, it is useful to know a formula for the derivative function whenever it is possible to find one.

In the next activity, we further explore the more algebraic approach to finding $f'(x)$: given a formula for $y = f(x)$, the limit definition of the derivative will be used to develop a formula for $f'(x)$.

Activity 2.3.3 For each of the listed functions, determine a formula for the derivative function. For the first two, determine the formula for the derivative by thinking about the nature of the given function and its slope at various points; do not use the limit definition. For the last three, use the limit definition. Pay careful attention to the function names and independent variables. It is important to be comfortable with using letters other than f and x. For example, given a function $p(z)$, we call its derivative $p'(z)$.

a. $f(x) = 1$

b. $g(t) = t$

c. $p(z) = z^2$

d. $F(t) = \dfrac{1}{t}$

e. $G(y) = \sqrt{y}$

2.3.2 Summary

- **Question 2.3.3** How does the limit definition of the derivative of a function f lead to an entirely new (but related) function f'? □

  The limit definition of the derivative, $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$, produces a value for each x at which the derivative is defined, and this leads to a new function $y = f'(x)$. It is especially important to note that taking the derivative is a process that starts with a given function (f) and produces a new, related function (f').

- **Question 2.3.4** What is the difference between writing $f'(a)$ and $f'(x)$? □

  There is essentially no difference between writing $f'(a)$ (as we did regularly in Section 2.2) and writing $f'(x)$. In either case, the variable is just a placeholder that is used to define the rule for the derivative function.

- **Question 2.3.5** How is the graph of the derivative function $f'(x)$ related to the graph of $f(x)$? □

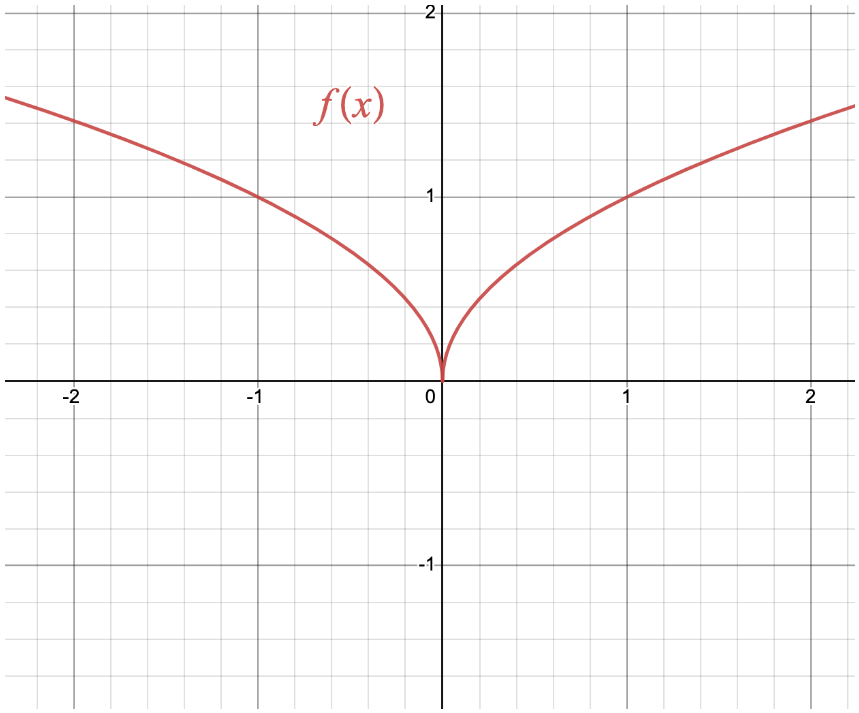
  Given the graph of a function $y = f(x)$, we can sketch an approximate graph of its derivative $y = f'(x)$ by observing that *heights* on the derivative's graph correspond to *slopes* on the original function's graph.

- **Question 2.3.6** What are some examples of functions f for which f' is not defined at one or more points? □

  In Activity 2.3.2, we encountered some functions that had sharp corners on their graphs, such as the shifted absolute value function. At such points, the derivative fails to exist, and we say that f is not differentiable there. It suffices to understand this as a consequence of the jump that must occur in the derivative function at a sharp corner on the graph of the original function.

2.3.3 Exercises

1. On the graph of the function $f(x)$ below, sketch a rough graph of its derivative function $f'(x)$. Use it to explain why $f'(0)$ does not exist.



Note: The point $(0, f(0))$ is called a **cusp**.

2. Consider the function $g(x) = x^2 - x + 3$.

a. Use the limit definition of the derivative to determine a formula for $g'(x)$.

b. Use a graphing utility to plot both $y = g(x)$ and your result for $y = g'(x)$; does your formula for $g'(x)$ generate the graph you expected?

c. Use the limit definition of the derivative to find a formula for $p'(x)$ where $p(x) = 5x^2 - 4x + 12$.

d. Compare and contrast the formulas for $g'(x)$ and $p'(x)$ you have found. How do the constants 5, 4, 12, and 3 affect the results?

3. Let g be a continuous function (that is, one with no jumps or holes in the graph) and suppose that a graph of $y = g'(x)$ is given by the graph on the right in Figure 2.3.7.

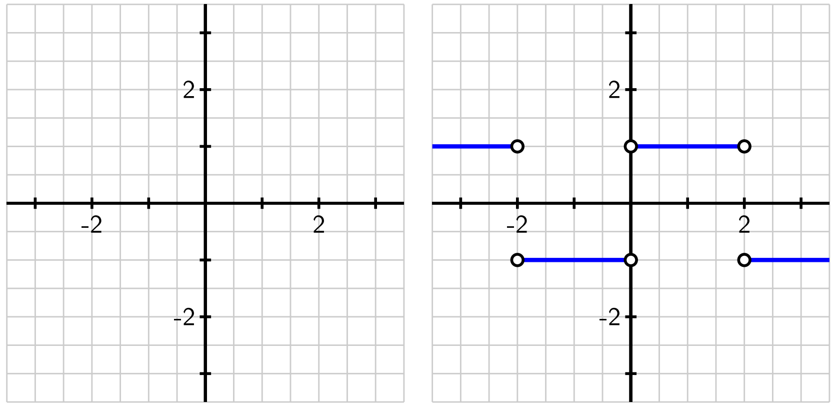


Figure 2.3.7 Axes for plotting $y = g(x)$ and, at right, the graph of $y = g'(x)$.

a. Observe that for every value of x that satisfies $0 < x < 2$, the value of $g'(x)$ is constant. What does this tell you about the behavior of the graph of $y = g(x)$ on this interval?

b. On what intervals other than $0 < x < 2$ do you expect $y = g(x)$ to be a linear function? Why?

c. At which values of x is $g'(x)$ not defined? What behavior does this lead you to expect to see in the graph of $y = g(x)$?

d. Suppose that $g(0) = 1$. On the axes provided at left in Figure 2.3.7, sketch an accurate graph of $y = g(x)$.

4. For each graph that provides an original function $y = f(x)$ in Figure 2.3.8, your task is to sketch an approximate graph of its derivative function, $y = f'(x)$, on the axes immediately below. The scale of the vertical axis for the graph of f' does not need to be accurate.

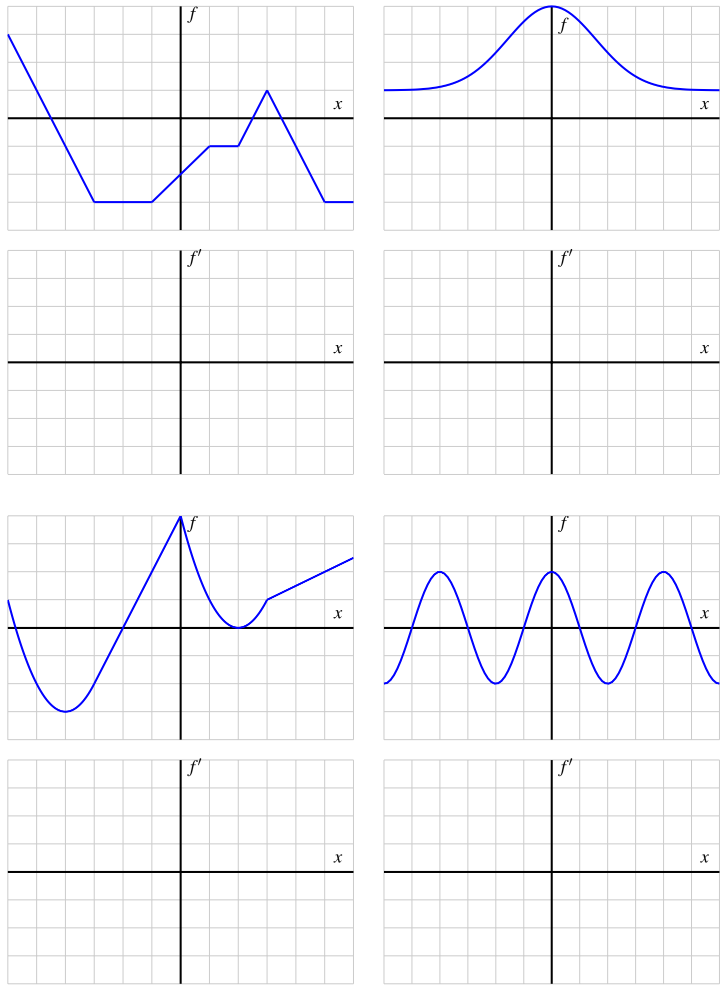


Figure 2.3.8 Graphs of $y = f(x)$ and grids for plotting the corresponding graph of $y = f'(x)$.

2.4 The Second Derivative

Motivating Questions

- How does the derivative of a function tell us whether the function is increasing or decreasing on an interval?
- What can we learn by taking the derivative of the derivative (the *second* derivative) of a function f?
- What does it mean to say that a function is concave up or concave down? How are these characteristics connected to certain properties of the derivative of the function?
- What are the units of the second derivative? How do they help us understand the rate of change of the rate of change?

Given a differentiable function $y = f(x)$, we know that its derivative, $y = f'(x)$, is a related function whose output at $x = a$ tells us the slope of the tangent line to $y = f(x)$ at the point $(a, f(a))$. That is, heights on the derivative graph tell us the values of slopes on the original function's graph.

At a point where $f'(x)$ is positive, the slope of the tangent line to f is positive. Therefore, on an interval where $f'(x)$ is positive, the function f is increasing (or rising). Similarly, if $f'(x)$ is negative on an interval, the graph of f is decreasing (or falling).

The derivative of f tells us not only *whether* the function f is increasing or decreasing on an interval, but also *how* the function f is increasing or decreasing. Look at the two tangent lines shown in Figure 2.4.1. We see that near point A the value of $f'(x)$ is positive and relatively close to zero, and near that point the graph is rising slowly. By contrast, near point B, the derivative is negative and relatively large in absolute value, and f is decreasing rapidly near B.

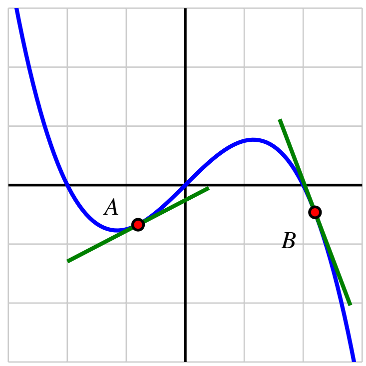


Figure 2.4.1 Two tangent lines on a graph.

Besides asking whether the value of the derivative function is positive or negative and whether it is large or small, we can also ask "how is the derivative changing?"

Because the derivative, $y = f'(x)$, is itself a function, we can consider taking its derivative — the derivative of the derivative — and ask "what does the derivative of the derivative tell us about how the original function behaves?" We start with an investigation of a moving object.

Warm-Up 2.4.1 The position of a car driving along a straight road at time t in minutes is given by the function $y = s(t)$ that is pictured in Figure 2.4.2. The car's position function has units measured in thousands of feet. For instance, the point $(2, 4)$ on the graph indicates that after 2 minutes, the car has traveled 4000 feet.

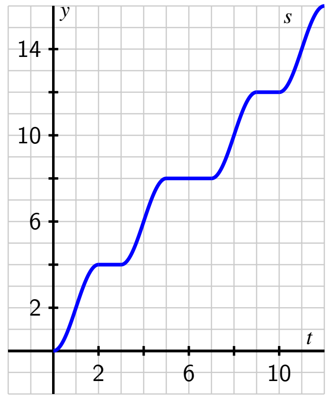


Figure 2.4.2 The graph of $y = s(t)$, the position of the car (measured in thousands of feet from its starting location) at time t in minutes.

a. In everyday language, describe the behavior of the car over the provided time interval. In particular, you should carefully discuss what is happening on each of the time intervals $[0, 1]$, $[1, 2]$, $[2, 3]$, $[3, 4]$, and $[4, 5]$, plus provide commentary overall on what the car is doing on the interval $[0, 12]$.

b. On the lefthand axes provided in Figure 2.4.3, sketch a careful, accurate graph of $y = s'(t)$.

c. What is the meaning of the function $y = s'(t)$ in the context of the given problem? What can we say about the car's behavior when $s'(t)$ is positive? when $s'(t)$ is zero? when $s'(t)$ is negative?

d. Rename the function you graphed in (b) to be called $y = v(t)$. Describe the behavior of v in words, using phrases like "v is increasing on the interval ..." and "v is constant on the interval ...."

e. Sketch a graph of the function $y = v'(t)$ on the righthand axes provide in Figure 2.4.3. Write at least one sentence to explain how the behavior of $v'(t)$ is connected to the graph of $y = v(t)$.

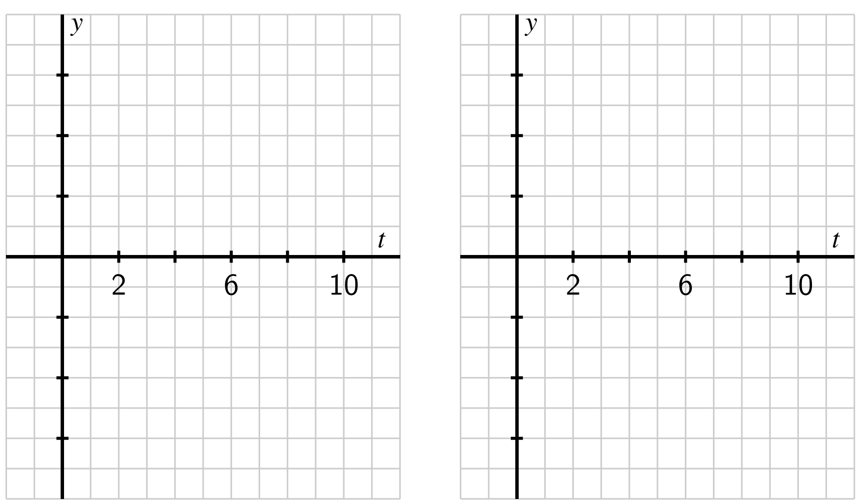


Figure 2.4.3 Axes for plotting $y = v(t) = s'(t)$ and $y = v'(t)$.

2.4.1 Increasing or decreasing

So far, we have used the words *increasing* and *decreasing* intuitively to describe a function's graph. Here we define these terms more formally.

Definition 2.4.4 Given a function $f(x)$ defined on the interval (a, b), we say that f **is increasing on** (a, b) provided that for all x, y in the interval (a, b), if $x < y$, then $f(x) < f(y)$. Similarly, we say that f **is decreasing on** (a, b) provided that for all x, y in the interval (a, b), if $x < y$, then $f(x) > f(y)$. ◊

Simply put, an increasing function is one that is rising as we move from left to right along the graph, and a decreasing function is one that falls as the value of the input increases. If the function has a derivative, the sign of the derivative tells us whether the function is increasing or decreasing.

Let f be a function that is differentiable on an interval (a, b). It is possible to show that that if $f'(x) > 0$ for every x such that $a < x < b$, then f is increasing on (a, b); similarly, if $f'(x) < 0$ on (a, b), then f is decreasing on (a, b).

For example, the function pictured in Figure 2.4.5 is increasing on the entire interval $-2 < x < 0$, and decreasing on the interval $0 < x < 2$. Note that the value $x = 0$ is not included in either interval since at this location, the function is changing from increasing to decreasing.

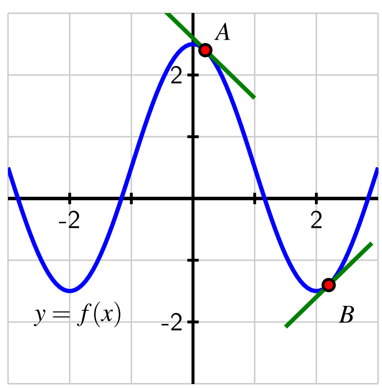


Figure 2.4.5 A function that is decreasing on the intervals $-3 < x < -2$ and $0 < x < 2$ and increasing on $-2 < x < 0$ and $2 < x < 3$.

2.4.2 The Second Derivative

We are now accustomed to investigating the behavior of a function by examining its derivative. The derivative of a function f is a new function given by the rule

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.$$

Because f' is itself a function, it is perfectly feasible for us to consider the derivative of the derivative, which is the new function $y = [f'(x)]'$. We call this resulting function *the second derivative* of $y = f(x)$, and denote the second derivative by $y = f''(x)$. Consequently, we will sometimes call f' "the first derivative" of f, rather than simply "the derivative" of f.

Definition 2.4.6 The **second derivative** is defined by the limit definition of the derivative of the first derivative. That is,

$$f''(x) = \lim_{h\to 0} \frac{f'(x+h) - f'(x)}{h}.$$

◊

The meaning of the derivative function still holds, so when we compute $y = f''(x)$, this new function measures slopes of tangent lines to the curve $y = f'(x)$, as well as the instantaneous rate of change of $y = f'(x)$. In other words, just as the first derivative measures the rate at which the original function changes, the second derivative measures the rate at which the first derivative changes. The second derivative will help us understand how the rate of change of the original function is itself changing.

2.4.3 Concavity

In addition to asking *whether* a function is increasing or decreasing, it is also natural to inquire *how* a function is increasing or decreasing. There are three basic behaviors that an increasing function can demonstrate on an interval, as pictured in Figure 2.4.7: the function can increase more and more rapidly, it can increase at the same rate, or it can increase in a way that is slowing down. Fundamentally, we are beginning to think about how a particular curve bends, with the natural comparison being made to lines, which don't bend at all. More than this, we want to understand how the bend in a function's graph is tied to behavior characterized by the first derivative of the function.

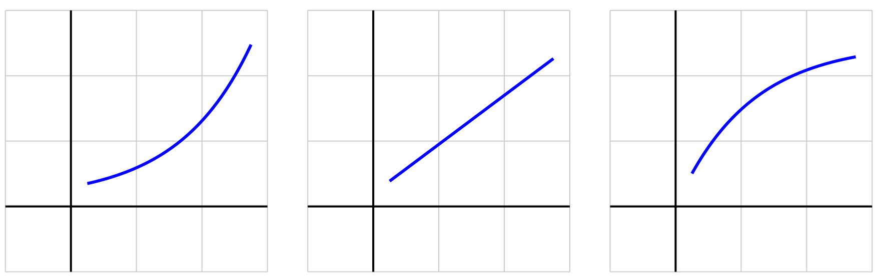


Figure 2.4.7 Three functions that are all increasing, but doing so at an increasing rate, at a constant rate, and at a decreasing rate, respectively.

On the leftmost curve in Figure 2.4.7, draw a sequence of tangent lines to the curve. As we move from left to right, the slopes of those tangent lines will increase. Therefore, the rate of change of the pictured function is increasing, and this explains why we say this function is *increasing at an increasing rate*. For the rightmost graph in Figure 2.4.7, observe that as x increases, the function increases, but the slopes of the tangent lines decrease. This function is *increasing at a decreasing rate*.

Similar options hold for how a function can decrease. Here we must be extra careful with our language, because decreasing functions involve negative slopes. Negative numbers present an interesting tension between common language and mathematical language. For example, it can be tempting to say that "-100 is bigger than -2." But we must remember that "greater than" describes how numbers lie on a number line: $x > y$ provided that x lies to the right of y. So of course, -100 is less than -2. Informally, it might be helpful to say that "-100 is more negative than -2." When a function's values are negative, and those values get more negative as the input increases, the function must be decreasing.

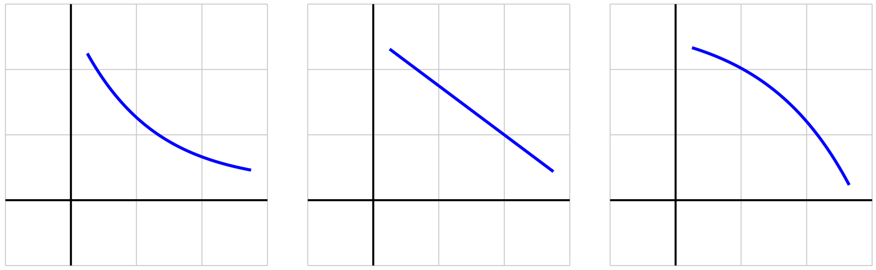


Figure 2.4.8 From left to right, three functions that are all decreasing, but doing so in different ways.

Now consider the three graphs shown in Figure 2.4.8. Clearly the middle graph depicts a function decreasing at a constant rate. Now, on the first curve, draw a sequence of tangent lines. We see that the slopes of these lines get less and less negative as we move from left to right. That means that the values of the first derivative, while all negative, are increasing, and thus we say that the leftmost curve is *decreasing at an increasing rate*.

This leaves only the rightmost curve in Figure 2.4.8 to consider. For that function, the slopes of the tangent lines are negative throughout the pictured interval, but as we move from left to right, the slopes get more and more negative. Hence the slope of the curve is decreasing, and we say that the function is *decreasing at a decreasing rate*.

We now introduce the notion of *concavity* which provides simpler language to describe these behaviors. When a curve opens upward on a given interval, like the parabola $y = x^2$ or the exponential growth function $y = e^x$, we say that the curve is *concave up* on that interval. Likewise, when a curve opens down, like the parabola $y = -x^2$ or the opposite of the exponential function $y = -e^x$, we say that the function is *concave down*. Concavity is linked to both the first and second derivatives of the function.

In Figure 2.4.9, we see two functions and a sequence of tangent lines to each. On the lefthand plot, where the function is concave up, observe that the tangent lines always lie below the curve itself, and the slopes of the tangent lines are increasing as we move from left to right. In other words, the function f is concave up on the interval shown because its derivative, f', is increasing on that interval. Similarly, on the righthand plot in Figure 2.4.9, where the function shown is concave down, we see that the tangent lines alway lie above the curve, and the slopes of the tangent lines are decreasing as we move from left to right. The fact that its derivative, f', is decreasing makes f concave down on the interval.

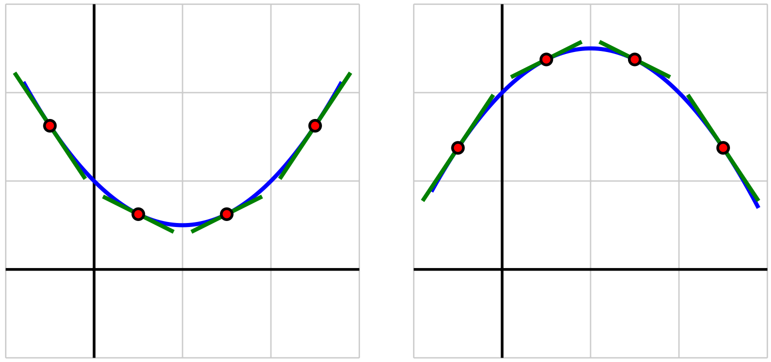


Figure 2.4.9 At left, a function that is concave up; at right, one that is concave down.

We state these most recent observations formally as the definitions of the terms *concave up* and *concave down*.

Definition 2.4.10 Let f be a differentiable function on an interval (a, b). Then f is **concave up** on (a, b) if and only if f' is increasing on (a, b); f is **concave down** on (a, b) if and only if f' is decreasing on (a, b). ◊

Activity 2.4.2 The position of a car driving along a straight road at time t in minutes is given by the function $y = s(t)$ that is pictured in Figure 2.4.11. The car's position function has units measured in thousands of feet. Remember that you worked with this function and sketched graphs of $y = v(t) = s'(t)$ and $y = v'(t)$ in Warm-Up 2.4.1.

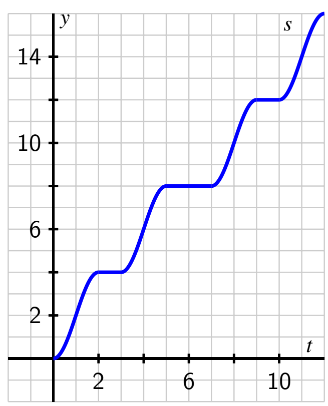


Figure 2.4.11 The graph of $y = s(t)$, the position of the car (measured in thousands of feet from its starting location) at time t in minutes.

a. On what intervals is the position function $y = s(t)$ increasing? decreasing? Why?

b. On which intervals is the velocity function $y = v(t) = s'(t)$ increasing? decreasing? neither? Why?

c. **Acceleration** is defined to be the instantaneous rate of change of velocity, as the acceleration of an object measures the rate at which the velocity of the object is changing. Say that the car's acceleration function is named

$a(t)$. How is $a(t)$ computed from $v(t)$? How is $a(t)$ computed from $s(t)$? Explain.

d. What can you say about s'' whenever s' is increasing? Why?

e. Using only the words *increasing*, *decreasing*, *constant*, *concave up*, *concave down*, and *linear*, complete the following sentences. For the position function s with velocity v and acceleration a,

- on an interval where v is positive, s is ___________.
- on an interval where v is negative, s is ___________.
- on an interval where v is zero, s is ___________.
- on an interval where a is positive, v is ___________.
- on an interval where a is negative, v is ___________.
- on an interval where a is zero, v is ___________.
- on an interval where a is positive, s is ___________.
- on an interval where a is negative, s is ___________.
- on an interval where a is zero, s is ___________.

Exploring the context of position, velocity, and acceleration is an excellent way to understand how a function, its first derivative, and its second derivative are related to one another. In Activity 2.4.2, we can replace s, v, and a with an arbitrary function f and its derivatives f' and f'', and essentially all the same observations hold. In particular, note that the following implications hold: f'' being positive on an interval implies f' is increasing on the interval implies f is concave up on the interval. Likewise, f'' being negative on an interval implies f' is decreasing on the interval implies f is concave down on the interval.

Activity 2.4.3 A potato is placed in an oven, and the potato's temperature F (in degrees Fahrenheit) at various points in time is taken and recorded in the following table. Time t is measured in minutes.

t	$F(t)$
0	70
15	180.5
30	251
45	296
60	324.5
75	342.8
90	354.5

Table 2.4.12 Select values of $F(t)$.

a. Though it's not possible to say for sure, do you think that $F'(30) > 0$ or $F'(30) < 0$? Explain your reasoning.

b. Though it's not possible to say for sure, do you think that $F''(30) > 0$ or $F''(30) < 0$? Explain your reasoning.

Activity 2.4.4 This activity builds on our experience and understanding of how to sketch the graph of f' given the graph of f.

In Figure 2.4.13, given the respective graphs of two different functions f, sketch the corresponding graph of f' on the first axes below, and then sketch f'' on the second set of axes. In addition, for each, write several careful sentences in the spirit of those in Activity 2.4.2 that connect the behaviors of f, f', and

f''. For instance, write something such as

> f' is __________ on the interval __________, which is connected to the fact that f is __________ on the same interval __________, and f'' is __________ on the interval.

but of course with the blanks filled in. The scale of the vertical axis for the graph of f' does not need to be accurate.

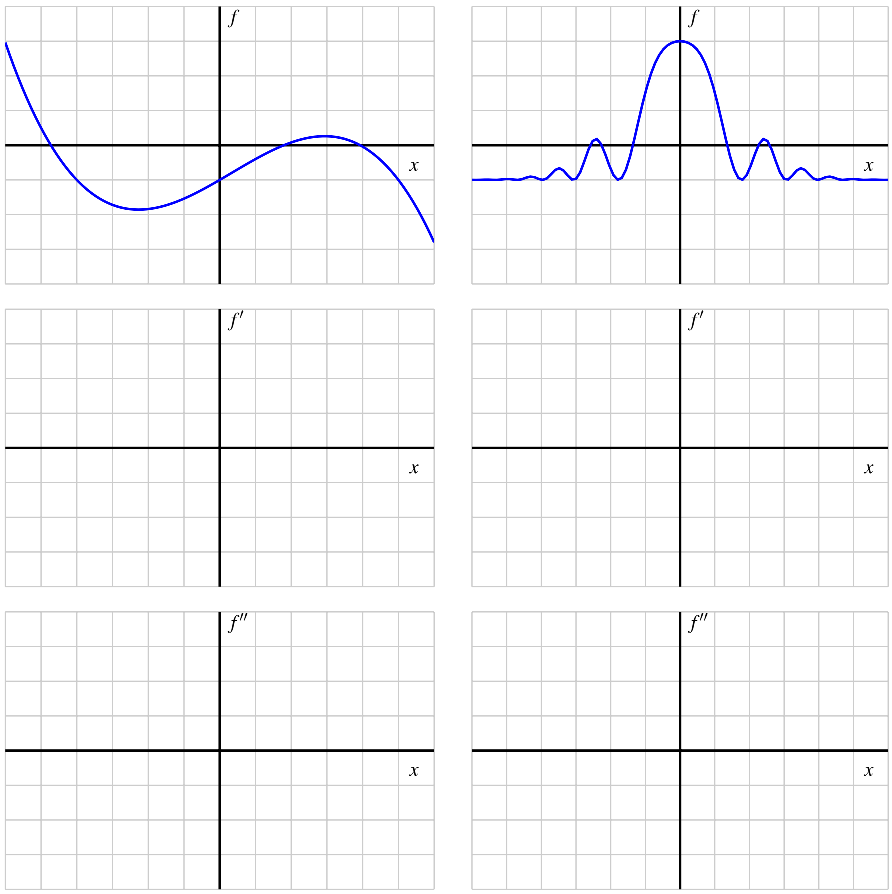


Figure 2.4.13 Two given functions f, with axes provided for plotting f' and f'' below.

2.4.4 Summary

- **Question 2.4.14** How does the derivative of a function tell us whether the function is increasing or decreasing on an interval? ☐

  A differentiable function f is increasing on an interval whenever its first derivative is positive, and decreasing whenever its first derivative is negative.

- **Question 2.4.15** What can we learn by taking the derivative of the derivative (the *second* derivative) of a function f? ☐

  By taking the derivative of the derivative of a function f, we arrive at the second derivative, f''. The second derivative measures the instantaneous rate of change of the first derivative. The sign of the second derivative tells us whether the slope of the tangent line to f is increasing or decreasing.

- **Question 2.4.16** What does it mean to say that a function is concave up or concave down? How are these characteristics connected to certain properties of the derivative of the function? ☐

  A differentiable function is concave up whenever its first derivative is increasing (which is true when its second derivative is positive), and concave down whenever its first derivative is decreasing (which is true when its second derivative is negative). Examples of functions that are everywhere concave up are $y = x^2$ and $y = e^x$; examples of functions that are everywhere concave down are $y = -x^2$ and $y = -e^x$.

- **Question 2.4.17** What are the units of the second derivative? How do they help us understand the rate of change of the rate of change? ☐

  The units on the second derivative are "units of output per unit of input per unit of input." They tell us how the value of the derivative function is changing in response to changes in the input. In other words, the second derivative tells us the rate of change of the rate of change of the original function.

2.4.5 Exercises

1. Suppose that $y = f(x)$ is a twice-differentiable function such that f'' is continuous for which the following information is known: $f(2) = -3$, $f'(2) = 1.5$, $f''(2) = -0.25$.

a. Is f increasing or decreasing near $x = 2$? Is f concave up or concave down near $x = 2$?

b. Do you expect $f(2.1)$ to be greater than -3, equal to -3, or less than -3? Why?

c. Do you expect $f'(2.1)$ to be greater than 1.5, equal to 1.5, or less than 1.5? Why?

d. Sketch a graph of $y = f(x)$ near $(2, f(2))$ and include a graph of the tangent line.

2. For a certain function $y = g(x)$, its derivative is given by the function pictured in Figure 2.4.18.

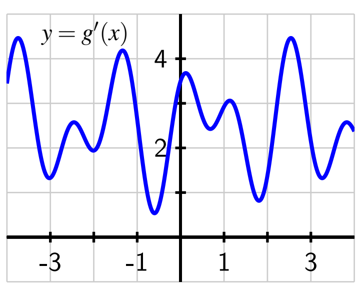


Figure 2.4.18 The graph of $y = g'(x)$.

a. What is the approximate slope of the tangent line to $y = g(x)$ at the point $(2, g(2))$?

b. How many real number solutions can there be to the equation $g(x) =$

0? Justify your conclusion fully and carefully by explaining what you know about how the graph of g must behave based on the given graph of g'.

c. On the interval $-3 < x < 3$, how many times does the concavity of g change? Why?

d. Use the provided graph to estimate the value of $g''(2)$.

3. For each prompt that follows, sketch a possible graph of a function on the interval $-3 < x < 3$ that satisfies the stated properties.

a. $y = f(x)$ such that f is increasing on $-3 < x < 3$, concave up on $-3 < x < 0$, and concave down on $0 < x < 3$.

b. $y = g(x)$ such that g is increasing on $-3 < x < 3$, concave down on $-3 < x < 0$, and concave up on $0 < x < 3$.

c. $y = h(x)$ such that h is decreasing on $-3 < x < 3$, concave up on $-3 < x < -1$, neither concave up nor concave down on $-1 < x < 1$, and concave down on $1 < x < 3$.

d. $y = p(x)$ such that p is decreasing and concave down on $-3 < x < 0$ and is increasing and concave down on $0 < x < 3$.

2.5 Elementary Derivative Rules

Motivating Questions

- What are alternate notations for the derivative?
- How can we use the algebraic structure of a function $f(x)$ to compute a formula for $f'(x)$?
- What is the derivative of a power function of the form $f(x) = x^n$? What is the derivative of an exponential function of the form $f(x) = b^x$?
- If we know the derivative of $y = f(x)$, what is the derivative of $y = kf(x)$, where k is a constant?
- If we know the derivatives of $y = f(x)$ and $y = g(x)$, how do we compute the derivative of $y = f(x) + g(x)$?

In Section 2.3, we developed the concept of the derivative of a function. We now know that the derivative f' of a function f measures the instantaneous rate of change of f with respect to x. The derivative also tells us the slope of the tangent line to $y = f(x)$ at any given value of x. So far, we have focused on interpreting the derivative graphically or, in the context of a physical setting, as a meaningful rate of change. To calculate the value of the derivative at a specific point, we have relied on the limit definition of the derivative,

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

In the remainder of this chapter, we investigate how the limit definition of the derivative leads to interesting patterns and rules that enable us to find a formula for $f'(x)$ quickly, *without* using the limit definition directly. For example, we would like to apply shortcuts to differentiate a function such as $g(x) = 4x^7 - \sin(x) + 3e^x$

Warm-Up 2.5.1 Functions of the form $f(x) = x^n$, where $n = 1, 2, 3, \ldots$, are often called **power functions**. The first question below revisits work we did earlier in Section 2.3, and the following questions extend those ideas to higher powers of x.

a. Use the limit definition of the derivative to find $f'(x)$ for $f(x) = x^2$.

b. Use the limit definition of the derivative to find $f'(x)$ for $f(x) = x^3$.

c. Use the limit definition of the derivative to find $f'(x)$ for $f(x) = x^4$. (Hint: $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$. Apply this rule to $(x+h)^4$ within the limit definition.)

d. Based on your work in (a), (b), and (c), what do you conjecture is the derivative of $f(x) = x^5$? Of $f(x) = x^{13}$?

e. Conjecture a formula for the derivative of $f(x) = x^n$ that holds for any positive integer n. That is, given $f(x) = x^n$ where n is a positive integer, what do you think is the formula for $f'(x)$?

2.5.1 Some Key Notation

In addition to our usual f' notation, there are other ways to denote the derivative of a function, as well as the instruction to take the derivative. If we are thinking about the relationship between y and x, we sometimes denote the derivative of y with respect to x by the symbol

$$\frac{dy}{dx}$$

which we read "dee-y dee-x." For example, if $y = x^2$, we'll write that the derivative is $\frac{dy}{dx} = 2x$. This notation comes from the fact that the derivative is related to the slope of a line, and slope is measured by $\frac{\Delta y}{\Delta x}$. Note that while we read $\frac{\Delta y}{\Delta x}$ as "change in y over change in x," we view $\frac{dy}{dx}$ as a single symbol, not a quotient of two quantities.

We use a variant of this notation as the instruction to take the derivative. In particular,

$$\frac{d}{dx}[\square]$$

means "take the derivative of the quantity in $\square$ with respect to x." For example, we may write $\frac{d}{dx}[x^2] = 2x$.

It is important to note that the independent variable can be different from x. If we have $f(z) = z^2$, we then write $f'(z) = 2z$. Similarly, if $y = t^2$, we say $\frac{dy}{dt} = 2t$. And it is also true that $\frac{d}{dq}[q^2] = 2q$. This notation may also be used for second derivatives: $f''(z) = \frac{d}{dz}\left[\frac{df}{dz}\right] = \frac{d^2 f}{dz^2}$.

In what follows, we'll build a repertoire of functions for which we can quickly compute the derivative.

2.5.2 Constant, Power, and Exponential Functions

So far, we know the derivative formula for two important classes of functions: constant functions and power functions. If $f(x) = c$ is a constant function, its graph is a horizontal line with slope zero at every point. Thus, $\frac{d}{dx}[c] = 0$. We summarize this with the following rule.

Constant Functions.

For any real number c, if $f(x) = c$, then $f'(x) = 0$.

Example 2.5.1 If $f(x) = 7$, then $f'(x) = 0$. Similarly, $\frac{d}{dx}[\sqrt{3}] = 0$. □

In your work in Warm-Up 2.5.1, you conjectured that for any positive integer n, if $f(x) = x^n$, then $f'(x) = nx^{n-1}$. This rule can be formally proved for any positive integer n, and also for any nonzero real number (positive or negative).

Power Functions.

For any nonzero real number n, if $f(x) = x^n$, then $f'(x) = nx^{n-1}$.

Example 2.5.2 Using the rule for power functions, we can compute the following derivatives. If $g(z) = z^{-3}$, then $g'(z) = -3z^{-4}$. Similarly, if $h(t) = t^{7/5}$, then $\frac{dh}{dt} = \frac{7}{5}t^{2/5}$, and $\frac{d}{dq}[q^{\pi}] = \pi q^{\pi-1}$. □

It will be instructive to have a derivative formula for one more type of basic function. For now, we simply state this rule without explanation or justification; we will explore why this rule is true in one of the exercises. And we will encounter graphical reasoning for why the rule is plausible in Warm-Up 2.6.1.

Exponential Functions.

For any positive real number b, if $f(x) = b^x$, then $f'(x) = b^x \ln(b)$.

Example 2.5.3 If $f(x) = 2^x$, then $f'(x) = 2^x \ln(2)$. Similarly, for $p(t) = 10^t$, $p'(t) = 10^t \ln(10)$. It is especially important to note that when $a = e$, where e is the base of the natural logarithm function, we have that

$$\frac{d}{dx}[e^x] = e^x \ln(e) = e^x$$

since $\ln(e) = 1$. This is an extremely important property of the function e^x: its derivative function is itself! □

Note carefully the distinction between power functions and exponential functions: in power functions, the variable is in the base, as in x^2, while in exponential functions, the variable is in the power, as in 2^x. As we can see from the rules, this makes a big difference in the form of the derivative.

Activity 2.5.2 Use the three rules above to determine the derivative of each of the following functions. For each, state your answer using full and proper notation, labeling the derivative with its name. For example, if you are given a function $h(z)$, you should write "$h'(z) =$" or "$\frac{dh}{dz} =$" as part of your response.

a. $f(t) = \pi$

b. $g(z) = 7^z$

c. $h(w) = w^{3/4}$

d. $p(x) = 3^{1/2}$

e. $r(t) = (\sqrt{2})^t$

f. $s(q) = q^{-1}$

g. $m(t) = \dfrac{1}{t^3}$

h. $f(x) = \dfrac{1}{r} + r^2$

2.5.3 Constant Multiples and Sums of Functions

Next we will learn how to compute the derivative of a function constructed as an algebraic combination of basic functions. For instance, we'd like to be able to take the derivative of a polynomial function such as

$$p(t) = 3t^5 - 7t^4 + t^2 - 9,$$

which is a sum of constant multiples of powers of t. To that end, we develop two new rules: the Constant Multiple Rule and the Sum Rule.

How is the derivative of $y = kf(x)$ related to the derivative of $y = f(x)$? Recall that when we multiply a function by a constant k, we vertically stretch the graph by a factor of $|k|$ (and reflect the graph across $y = 0$ if $k < 0$). This vertical stretch affects the slope of the graph, so the slope of the function $y = kf(x)$ is k times as steep as the slope of $y = f(x)$. Thus, when we multiply a function by a factor of k, we change the value of its derivative by a factor of k as well.[1]

The Constant Multiple Rule.

For any real number k, if $f(x)$ is a differentiable function with derivative $f'(x)$, then $\frac{d}{dx}[kf(x)] = kf'(x)$.

In words, this rule says that "the derivative of a constant times a function is the constant times the derivative of the function."

Example 2.5.4 If $g(t) = 3 \cdot 5^t$, we have $g'(t) = 3 \cdot 5^t \ln(5)$. Similarly, $\frac{d}{dz}[5z^{-2}] = 5(-2z^{-3}) = -10z^{-3}$. □

Next we examine a sum of two functions. If we have $y = f(x)$ and $y = g(x)$, we can compute a new function $y = (f + g)(x)$ by adding the outputs of the two functions: $(f + g)(x) = f(x) + g(x)$. Not only is the value of the new function the sum of the values of the two known functions, but the slope of the new function is the sum of the slopes of the known functions. Therefore[2], we arrive at the following Sum Rule for derivatives:

The Sum Rule.

If $f(x)$ and $g(x)$ are differentiable functions with derivatives $f'(x)$ and $g'(x)$ respectively, then $\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$.

In words, the Sum Rule tells us that "the derivative of a sum is the sum of the derivatives." It also tells us that a sum of two differentiable functions is also differentiable. Furthermore, because we can view the difference function $y = (f-g)(x) = f(x)-g(x)$ as $y = f(x)+(-1\cdot g(x))$, the Sum Rule and Constant Multiple Rules together tell us that $\frac{d}{dx}[f(x) + (-1 \cdot g(x))] = f'(x) - g'(x)$, or that "the derivative of a difference is the difference of the derivatives." We can now compute derivatives of sums and differences of elementary functions.

Example 2.5.5 Using the sum rule, $\frac{d}{dw}(2^w + w^2) = 2^w \ln(2) + 2w$. Using both the sum and constant multiple rules, if $h(q) = 3q^6 - 4q^{-3}$, then $h'(q) = 3(6q^5) - 4(-3q^{-4}) = 18q^5 + 12q^{-4}$. □

Activity 2.5.3 Use only the rules for constant, power, and exponential functions, together with the Constant Multiple and Sum Rules, to compute the derivative of each function below with respect to the given independent variable. Note well that we do not yet know any rules for how to differentiate the product or quotient of functions. This means that you may have to do some algebra first on the functions below before you can actually use existing rules to compute the desired derivative formula. In each case, label the derivative you calculate

[1]The Constant Multiple Rule can be formally proved as a consequence of properties of limits, using the limit definition of the derivative.

[2]Like the Constant Multiple Rule, the Sum Rule can be formally proved as a consequence of properties of limits, using the limit definition of the derivative.

with its name using proper notation such as $f'(x)$, $h'(z)$, dr/dt, etc.

a. $f(x) = x^{5/3} - x^4 + 2^x$

b. $g(x) = 14e^x + 3x^5 - x$

c. $h(z) = \sqrt{z} + \frac{1}{z^4} + 5^z$

d. $r(t) = \sqrt{53}\,t^7 - \pi e^t + e^4$

e. $s(y) = (y^2 + 1)(y^2 - 1)$

f. $q(x) = \dfrac{x^3 - x + 2}{x}$

g. $p(a) = 3a^4 - 2a^3 + 7a^2 - a + 12$

In the same way that we have shortcut rules to help us find derivatives, we introduce some language that is simpler and shorter. Often, rather than say "take the derivative of f," we'll instead say simply "differentiate f." Similarly, if the derivative exists at a point, we say "f is differentiable at that point," or that f can be differentiated.

As we work with the algebraic structure of functions, it is important to develop a big picture view of what we are doing. Here, we make several general observations based on the rules we have so far.

- The derivative of any polynomial function will be another polynomial function, and that the degree of the derivative is one less than the degree of the original function. For instance, if $p(t) = 7t^5 - 4t^3 + 8t$, p is a degree 5 polynomial, and its derivative, $p'(t) = 35t^4 - 12t^2 + 8$, is a degree 4 polynomial.

- The derivative of any exponential function is another exponential function: for example, if $g(z) = 7 \cdot 2^z$, then $g'(z) = 7 \cdot 2^z \ln(2)$, which is also exponential.

- We should not lose sight of the fact that all of the meaning of the derivative that we developed in Section 2.3 still holds. The derivative measures the instantaneous rate of change of the original function, as well as the slope of the tangent line at any selected point on the curve.

Activity 2.5.4 Each of the following questions asks you to use derivatives to answer key questions about functions. Be sure to think carefully about each question and to use proper notation in your responses.

a. Find the slope of the tangent line to $h(z) = \sqrt{z} + \frac{1}{z}$ at the point where $z = 4$.

b. A population of cells is growing in such a way that its total number in millions is given by the function $P(t) = 2(1.37)^t + 32$, where t is measured in days.

   i. Determine the instantaneous rate at which the population is growing on day 4, and include units on your answer.

   ii. Is the population growing at an increasing rate or growing at a decreasing rate on day 4? Explain.

c. Find an equation for the tangent line to the curve $p(a) = 3a^4 - 2a^3 + 7a^2 - a + 12$ at the point where $a = -1$.

d. What is the difference between being asked to find the *slope* of the tangent line (asked in (a)) and the *equation* of the tangent line (asked in (c))?

When functions are used as models, they are not always represented explicitly as in the examples we've seen so far in this section. See Exercise 1.1.4.2 for a reminder of the types of function representations we may need to analyze when modeling real-world systems. When a function is defined in terms of other unknown functions, we can still compute a derivative using derivative rules.

We will start practicing this skill in this section, and continue throughout this chapter as we develop more derivative rules.

Example 2.5.6 Symbolic Derivative Computations. Suppose $f(x)$ and $g(x)$ are differentiable functions.

Let $\ell(x) = 2f(x) + 6g(x)$. Then using the constant multiple rule and the sum rule for derivatives, we compute that $\ell'(x) = 2f'(x) + 6g'(x)$.

Let $k(x) = -3f(x) + 6x^5$. Then using the constant multiple rule and the sum rule for derivatives, we compute that $k'(x) = -3f'(x) + 30x^4$.

Let $t(x) = kx^2 - 4g(x)$. Then using the constant multiple rule and the sum rule for derivatives (and noting that k is a parameter), we compute that $t'(x) = 2kx - 4g'(x)$. □

2.5.4 Summary

- **Question 2.5.7** What are alternate notations for the derivative? □

  Given a differentiable function $y = f(x)$, we can express the derivative of f in several different notations: $f'(x)$, $\frac{df}{dx}$, $\frac{dy}{dx}$, and $\frac{d}{dx}[f(x)]$.

- **Question 2.5.8** How can we use the algebraic structure of a function $f(x)$ to compute a formula for $f'(x)$? □

  The limit definition of the derivative leads to patterns among certain families of functions that enable us to compute derivative formulas without resorting directly to the limit definition. For example, if f is a power function of the form $f(x) = x^n$, then $f'(x) = nx^{n-1}$ for any real number n other than 0. This is called the Rule for Power Functions.

- **Question 2.5.9** What is the derivative of a power function of the form $f(x) = x^n$? What is the derivative of an exponential function of the form $f(x) = b^x$? □

  We have stated a rule for derivatives of exponential functions in the same spirit as the rule for power functions: for any positive real number b, if $f(x) = b^x$, then $f'(x) = b^x \ln(b)$.

- **Question 2.5.10** If we know the derivative of $y = f(x)$, what is the derivative of $y = kf(x)$, where k is a constant? If we know the derivatives of $y = f(x)$ and $y = g(x)$, how do we compute the derivative of $y = f(x) + g(x)$? □

  If we are given a constant multiple of a function whose derivative we know, or a sum of functions whose derivatives we know, the Constant Multiple and Sum Rules make it straightforward to compute the derivative of the overall function. More formally, if $f(x)$ and $g(x)$ are differentiable with derivatives $f'(x)$ and $g'(x)$ and a and b are constants, then

  $$\frac{d}{dx}\left[af(x) + bg(x)\right] = af'(x) + bg'(x).$$

2.5.5 Exercises

1. Let f and g be differentiable functions for which the following information is known: $f(2) = 5$, $g(2) = -3$, $f'(2) = -1/2$, $g'(2) = 2$.

a. Let h be the new function defined by the rule $h(x) = 3f(x) - 4g(x)$. Determine $h(2)$ and $h'(2)$.

b. Find an equation for the tangent line to $y = h(x)$ at the point $(2, h(2))$.

c. Let p be the function defined by the rule $p(x) = -2f(x) + \frac{1}{2}g(x)$. Is p increasing, decreasing, or neither at $a = 2$? Why?

2. Let functions p and q be the piecewise linear functions given by their respective graphs in Figure 2.5.11. Use the graphs to answer the following questions.

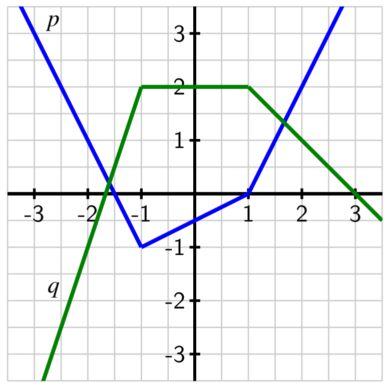


Figure 2.5.11 The graphs of p (in blue) and q (in green).

a. At what values of x is p not differentiable? At what values of x is q not differentiable? Why?

b. Let $r(x) = p(x) + 2q(x)$. Determine $r'(-2)$ and $r'(0)$.

c. Find an equation for the tangent line to $y = r(x)$ at the point $(2, r(2))$.

3. Let $f(x) = a^x$. The goal of this problem is to explore how the value of a affects the derivative of $f(x)$, without assuming we know the rule for $\frac{d}{dx}[a^x]$ that we have stated and used in earlier work in this section.

a. Use the limit definition of the derivative to show that

$$f'(x) = \lim_{h\to 0} \frac{a^x \cdot a^h - a^x}{h}.$$

b. Explain why it is also true that

$$f'(x) = a^x \cdot \lim_{h\to 0} \frac{a^h - 1}{h}.$$

c. Use computing technology and small values of h to estimate the value of

$$L = \lim_{h\to 0} \frac{a^h - 1}{h}$$

when $a = 2$. Do likewise when $a = 3$.

d. Note that it would be ideal if the value of the limit L was 1, for then f would be a particularly special function: its derivative would be simply a^x, which would mean that its derivative is itself. By experimenting with different values of a between 2 and 3, try to find a value for a for which

$$L = \lim_{h\to 0} \frac{a^h - 1}{h} = 1.$$

e. Compute $\ln(2)$ and $\ln(3)$. What does your work in (b) and (c) suggest is true about $\frac{d}{dx}[2^x]$ and $\frac{d}{dx}[3^x]$?

f. How do your investigations in (d) lead to a particularly important fact about the function $f(x) = e^x$?

2.6 Derivatives of the Sine and Cosine Functions

Motivating Questions

- What is a graphical justification for why $\frac{d}{dx}[b^x] = b^x \ln(b)$?
- What do the graphs of $y = \sin(x)$ and $y = \cos(x)$ suggest as formulas for their respective derivatives?
- Once we know the derivatives of $\sin(x)$ and $\cos(x)$, how do previous derivative rules work when these functions are involved?

Throughout Chapter 2, we will develop shortcut derivative rules to help us bypass the limit definition and quickly compute $f'(x)$ from a formula for $f(x)$. In Section 2.5, we stated the rule for power functions,

$$\text{if } f(x) = x^n, \text{ then } f'(x) = nx^{n-1},$$

and the rule for exponential functions,

$$\text{if } b \text{ is a positive real number and } f(x) = b^x, \text{ then } f'(x) = b^x \ln(b).$$

Later in this section, we will use a graphical argument to conjecture derivative formulas for the sine and cosine functions.

Warm-Up 2.6.1 Consider the function $g(x) = 2^x$, which is graphed in Figure 2.6.1.

a. At each of $x = -2, -1, 0, 1, 2$, use a straightedge to sketch an accurate tangent line to $y = g(x)$.

b. Use the provided grid to estimate the slope of the tangent line you drew at each point in (a).

c. Use the limit definition of the derivative to estimate $g'(0)$ by using small values of h, and compare the result to your visual estimate for the slope of the tangent line to $y = g(x)$ at $x = 0$ in (b).

d. Based on your work in (a), (b), and (c), sketch an accurate graph of $y = g'(x)$ on the axes adjacent to the graph of $y = g(x)$.

e. Write at least one sentence that explains why it is reasonable to think that $g'(x) = cg(x)$, where c is a constant. In addition, calculate $\ln(2)$, and then discuss how this value, combined with your work above, reasonably suggests that $g'(x) = 2^x \ln(2)$.

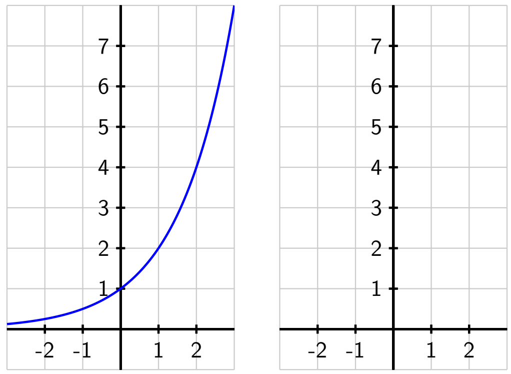


Figure 2.6.1 At left, the graph of $y = g(x) = 2^x$. At right, axes for plotting $y = g'(x)$.

2.6.1 The sine and cosine functions

The sine and cosine functions are among the most important functions in all of mathematics. Sometimes called the *circular* functions due to their definition on the unit circle (Section 1.6), these periodic functions play a key role in modeling repeating phenomena such as tidal elevations, the behavior of an oscillating mass attached to a spring, or weather patterns. Like polynomial and exponential functions, the sine and cosine functions are considered basic functions, ones that are often used in building more complicated functions. As such, we would like to know formulas for $\frac{d}{dx}[\sin(x)]$ and $\frac{d}{dx}[\cos(x)]$, and the next two activities lead us to that end.

Activity 2.6.2 Consider the function $f(x) = \sin(x)$, which is graphed in Figure 2.6.2 below. Note carefully that the grid in the diagram does not have boxes that are 1×1, but rather approximately 1.57×1, as the horizontal scale of the grid is $\pi/2$ units per box.

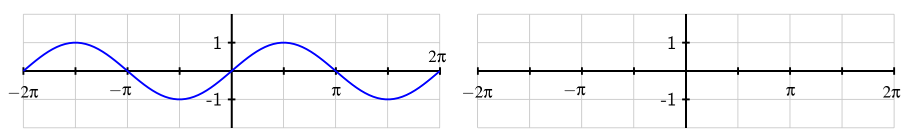


Figure 2.6.2 At left, the graph of $y = f(x) = \sin(x)$.

a. At each of $x = -2\pi, -\frac{3\pi}{2}, -\pi, -\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$, use a straightedge to sketch an accurate tangent line to $y = f(x)$.

b. Use the provided grid to estimate the slope of the tangent line you drew at each point. Pay careful attention to the scale of the grid.

c. Use the limit definition of the derivative to estimate $f'(0)$ by using small values of h, and compare the result to your visual estimate for the slope of the tangent line to $y = f(x)$ at $x = 0$ in (b). Using periodicity, what does this result suggest about $f'(2\pi)$? about $f'(-2\pi)$?

d. Based on your work in (a), (b), and (c), sketch an accurate graph of $y = f'(x)$ on the axes adjacent to the graph of $y = f(x)$.

e. What familiar function do you think is the derivative of $f(x) = \sin(x)$?

Activity 2.6.3 Consider the function $g(x) = \cos(x)$, which is graphed in Figure 2.6.3 below. Note carefully that the grid in the diagram does not have boxes that are 1×1, but rather approximately 1.57×1, as the horizontal scale of the grid is $\pi/2$ units per box.

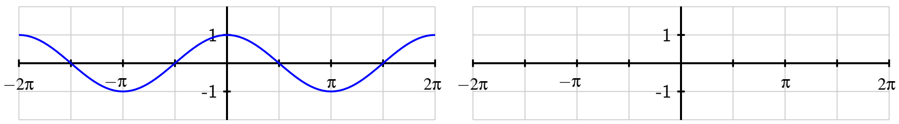


Figure 2.6.3 At left, the graph of $y = g(x) = \cos(x)$.

a. At each of $x = -2\pi, -\frac{3\pi}{2}, -\pi, -\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$, use a straightedge to sketch an accurate tangent line to $y = g(x)$.

b. Use the provided grid to estimate the slope of the tangent line you drew at each point. Again, note the scale of the axes and grid.

c. Use the limit definition of the derivative to estimate $g'(\frac{\pi}{2})$ by using small values of h, and compare the result to your visual estimate for the slope of the tangent line to $y = g(x)$ at $x = \frac{\pi}{2}$ in (b). Using periodicity, what does this result suggest about $g'(-\frac{3\pi}{2})$? can symmetry on the graph help you estimate other slopes easily?

d. Based on your work in (a), (b), and (c), sketch an accurate graph of $y = g'(x)$ on the axes adjacent to the graph of $y = g(x)$.

e. What familiar function do you think is the derivative of $g(x) = \cos(x)$?

The results of the two preceding activities suggest that the sine and cosine functions not only have beautiful connections such as the identities $\sin^2(x) + \cos^2(x) = 1$ and $\cos(x - \frac{\pi}{2}) = \sin(x)$, but that they are even further linked through calculus, as the derivative of each involves the other. The following rules summarize the results of the activities[1].

Sine and Cosine Functions.

For all real numbers x,

$$\frac{d}{dx}[\sin(x)] = \cos(x) \quad \text{and} \quad \frac{d}{dx}[\cos(x)] = -\sin(x).$$

We have now added the sine and cosine functions to our library of basic functions whose derivatives we know. The constant multiple and sum rules still hold, of course, as well as all of the inherent meaning of the derivative.

Activity 2.6.4 Answer each of the following questions. Where a derivative is requested, be sure to label the derivative function with its name using proper notation.

a. Determine the derivative of $h(t) = 3\cos(t) - 4\sin(t)$.

b. Find the exact slope of the tangent line to $y = f(x) = 2x + \frac{\sin(x)}{2}$ at the point where $x = \frac{\pi}{6}$.

c. Find the equation of the tangent line to $y = g(x) = x^2 + 2\cos(x)$ at the point where $x = \frac{\pi}{2}$.

[1]These two rules may be formally proved using the limit definition of the derivative and the expansion identities for $\sin(x + h)$ and $\cos(x + h)$.

d. Determine the derivative of $p(z) = z^4 + 4^z + 4\cos(z) - \sin(\frac{\pi}{2})$.

e. The function $P(t) = 24 + 8\sin(t)$ represents a population of a particular kind of animal that lives on a small island, where P is measured in hundreds and t is measured in decades since January 1, 2010. What is the instantaneous rate of change of P on January 1, 2030? What are the units of this quantity? Write a sentence in everyday language that explains how the population is behaving at this point in time.

2.6.2 Summary

- **Question 2.6.4** What is a graphical justification for why $\frac{d}{dx}[b^x] = b^x \ln(b)$? □

  For an exponential function $f(x) = b^x (b > 1)$, the graph of $f'(x)$ appears to be a scaled version of the original function. In particular, careful analysis of the graph of$f(x) = 2^x$, suggests that $\frac{d}{dx}[2^x] = 2^x \ln(2)$, which is a special case of the rule we stated in Section 2.5.

- **Question 2.6.5** What do the graphs of $y = \sin(x)$ and $y = \cos(x)$ suggest as formulas for their respective derivatives? □

  By carefully analyzing the graphs of $y = \sin(x)$ and $y = \cos(x)$, and by using the limit definition of the derivative at select points, we found that $\frac{d}{dx}[\sin(x)] = \cos(x)$ and $\frac{d}{dx}[\cos(x)] = -\sin(x)$.

- **Question 2.6.6** Once we know the derivatives of $\sin(x)$ and $\cos(x)$, how do previous derivative rules work when these functions are involved? □

  We note that all previously encountered derivative rules still hold, but now may also be applied to functions involving the sine and cosine. All of the established meaning of the derivative applies to these trigonometric functions as well.

2.6.3 Exercises

1. Suppose that $V(t) = 24 \cdot 1.07^t + 6\sin(t)$ represents the value of a person's investment portfolio in thousands of dollars in year t, where $t = 0$ corresponds to January 1, 2010.

a. At what instantaneous rate is the portfolio's value changing on January 1, 2012? Include units on your answer.

b. Determine the value of $V''(2)$. What are the units on this quantity and what does it tell you about how the portfolio's value is changing?

c. On the interval $0 \le t \le 20$, use technology to graph the function $V(t) = 24 \cdot 1.07^t + 6\sin(t)$ and describe its behavior in the context of the problem. Then, compare the graphs of the functions $A(t) = 24 \cdot 1.07^t$ and $V(t) = 24 \cdot 1.07^t + 6\sin(t)$, as well as the graphs of their derivatives $A'(t)$ and $V'(t)$. What is the impact of the term $6\sin(t)$ on the behavior of the function $V(t)$?

2. Let $f(x) = 3\cos(x) - 2\sin(x) + 6$.

a. Determine the exact slope of the tangent line to $y = f(x)$ at the point where $a = \frac{\pi}{4}$.

b. Determine the equation of the tangent line to $y = f(x)$ at the point where $a = \pi$.

c. At the point where $a = \frac{\pi}{2}$, is f increasing, decreasing, or neither?

d. At the point where $a = \frac{3\pi}{2}$, does the tangent line to $y = f(x)$ lie above the curve, below the curve, or neither? How can you answer this question without even graphing the function or the tangent line?

3. Let $s(\theta) = \sin\left(\theta + \dfrac{\pi}{2}\right)$.

(a) Explain why none of the derivative rules we have developed up to this point tell us how to compute $s'(\theta)$ (HINT: We will develop a rule to help us differentiate functions in this form in Section 2.8). Even so, give your best guess as to what you think $s'(\theta)$ equals and explain your thought process.

(b) We can use what we know about function transformations to help us compute $s'(\theta)$.

i. Re-write $s(\theta)$ as an elementary trig function using your knowledge of function transformations.

ii. Take the derivative of $s(\theta)$ in this new form using one of the derivative rules developed in this section.

iii. Use function transformations again to re-write $s'(\theta)$ in a form that resembles the form of your guess from the first part of this problem.

2.7 Derivatives of Products and Quotients

Motivating Questions

- How does the algebraic structure of a function guide us in computing its derivative using shortcut rules?
- How do we compute the derivative of a product of two basic functions in terms of the derivatives of the basic functions?
- How do we compute the derivative of a quotient of two basic functions in terms of the derivatives of the basic functions?
- How do the product and quotient rules combine with the sum and constant multiple rules to expand the library of functions we can differentiate quickly?

So far, we can differentiate power functions (x^n), exponential functions (b^x), and the two fundamental trigonometric functions ($\sin(x)$ and $\cos(x)$). With the sum rule and constant multiple rules, we can also compute the derivative of combined functions.

Example 2.7.1 Differentiate

$$f(x) = 7x^{11} - 4 \cdot 9^x + \pi \sin(x) - \sqrt{3}\cos(x)$$

Because f is a sum of basic functions, we can now quickly say that $f'(x) = 77x^{10} - 4 \cdot 9^x \ln(9) + \pi \cos(x) + \sqrt{3}\sin(x)$. □

What about a product or quotient of two basic functions, such as

$$p(z) = z^3 \cos(z),$$

or

$$q(t) = \frac{\sin(t)}{2^t}?$$

While the derivative of a sum is the sum of the derivatives, it turns out that the rules for computing derivatives of products and quotients are more complicated.

Warm-Up 2.7.1 Let f and g be the functions defined by $f(t) = 2t^2$ and $g(t) = t^3 + 4t$.

a. Determine $f'(t)$ and $g'(t)$.

b. Let $p(t) = 2t^2(t^3 + 4t)$ and observe that $p(t) = f(t) \cdot g(t)$. Rewrite the formula for p by distributing the $2t^2$ term. Then, compute $p'(t)$ using the sum and constant multiple rules.

c. True or false: $p'(t) = f'(t) \cdot g'(t)$.

d. Let $q(t) = \frac{t^3+4t}{2t^2}$ and observe that $q(t) = \frac{g(t)}{f(t)}$. Rewrite the formula for q by dividing each term in the numerator by the denominator and simplify to write q as a sum of constant multiples of powers of t. Then, compute $q'(t)$ using the sum and constant multiple rules.

e. True or false: $q'(t) = \frac{g'(t)}{f'(t)}$.

2.7.1 The product rule

As part (b) of Warm-Up 2.7.1 shows, it is not true in general that the derivative of a product of two functions is the product of the derivatives of those functions. To see why this is the case, we consider an example involving meaningful functions.

Say that we are tracking carbon emissions for a particular population over time. Let $N(t)$ represent the number of individuals on day t, where $t = 0$ represents the first day on which we begin tracking. Let $S(t)$ give the amount of carbon emissions from one individual on day t; the units on $S(t)$ are pounds per person (the rate of carbon emissions). To compute the total carbon emissions for the whole population on day t, we take the product

$$V(t) = N(t)\,\text{individuals} \cdot S(t)\,\text{pounds per individual}.$$

Observe that over time, both the number of individuals and the rate of carbon emissions will vary. The derivative $N'(t)$ measures the rate at which the number of individuals is changing, while $S'(t)$ measures the rate at which the carbon emission per individual is changing. How do these respective rates of change affect the rate of change of the total carbon emission function?

To help us understand the relationship among changes in N, S, and V, let's consider some specific data.

- Suppose that on day 100, there are 520 individuals and the rate of carbon emission is 2.75 pounds per individual. This tells us that $N(100) = 520$ and $S(100) = 2.75$.

- On day 100, the population grows by an additional 12 individuals (so the number of individuals is rising at a rate of 12 individuals per day).

- On that same day the rate of carbon emissions is rising at a rate of 0.75 pounds per individual per day.

In calculus notation, the latter two facts tell us that $N'(100) = 12$ (individuals per day) and $S'(100) = 0.75$ (pounds per individual per day). At what rate is the total carbon emission changing on day 100?

Observe that the increase in total value comes from two sources: the growing number of individuals, and the rising rate of carbon emissions per individual. If only the number of individuals is increasing (and the value of the rate of carbon emission is constant), the rate at which total carbon emissions would rise is the product of the current rate of carbon emission and the rate at which the number of individuals is changing. That is, the rate at which total carbon emissions would change is given by

$$S(100) \cdot N'(100) = 2.75 \, \frac{\text{pounds}}{\text{individual}} \cdot 12 \, \frac{\text{individuals}}{\text{day}} = 33 \, \frac{\text{pounds}}{\text{day}}.$$

Note particularly how the units make sense and show the rate at which the total value V is changing, measured in pounds per day.

If instead the number of individuals is constant, but the rate of carbon emissions is rising, the rate at which the total carbon emission would rise is the product of the number of individuals and the rate of change of the carbon emission rate. The total carbon emission is rising at a rate of

$$N(100) \cdot S'(100) = 520 \, \text{individuals} \cdot 0.75 \, \frac{\text{pounds per individual}}{\text{day}} = 390 \, \frac{\text{pounds}}{\text{day}}.$$

Of course, when both the number of individuals and the rate of carbon emissions are changing, we have to include both of these sources. In that case the rate at which the total carbon emission is rising is

$$V'(100) = S(100) \cdot N'(100) + N(100) \cdot S'(100) = 33 + 390 = 423 \, \frac{\text{pounds}}{\text{day}}.$$

We expect the total carbon emission from the population to rise by about 423 pounds on the 100th day.[1]

Next, we expand our perspective from the specific example above to the more general and abstract setting of a product p of two differentiable functions, f and g. If $P(x) = f(x) \cdot g(x)$, our work above suggests that $P'(x) = f'(x)g(x) + f(x)g'(x)$. Indeed, a formal proof using the limit definition of the derivative can be given to show that the following rule, called the *product rule*, holds in general.

Product Rule.

If f and g are differentiable functions, then their product $P(x) = f(x) \cdot g(x)$ is also a differentiable function, and

$$P'(x) = f'(x)g(x) + f(x)g'(x).$$

[1] While this example highlights why the product rule is true, there are some subtle issues to recognize. For one, if the rate of carbon emissions really does rise exactly 0.75 pounds per individual on day 100, and the number of individuals really rises by 12 on day 100, then we'd expect that $V(101) = N(101) \cdot S(101) = 532 \cdot 3.5 = 1862$. If, as noted above, we expect the total carbon emission to rise by 423 pounds, then with $V(100) = N(100) \cdot S(100) = 520 \cdot 2.75 = 1430$, then it seems we should find that $V(101) = V(100) + 423 = 1853$. Why do the two results differ by 9? One way to understand why this difference occurs is to recognize that $N'(100) = 12$ represents an *instantaneous* rate of change, while our (informal) discussion has also thought of this number as the total change in the number of individuals over the course of a single day. The formal proof of the product rule reconciles this issue by taking the limit as the change in the input tends to zero.

In light of the earlier example involving carbon emissions, the product rule also makes sense intuitively: the rate of change of P should take into account both how fast f and g are changing, as well as how large f and g are at the point of interest. In words the product rule says: if P is the product of two functions f (the first function) and g (the second), then "the derivative of P is the derivative of the first times the second, plus the first times the derivative of the second." It is often a helpful mental exercise to say this phrasing aloud when executing the product rule.

Example 2.7.2 If $P(z) = z^3 \cdot \cos(z)$, we can use the product rule to differentiate P. The first function is z^3 and the second function is $\cos(z)$. By the product rule, P' will be given by the derivative of the first, $3z^2$, times the second, $\cos(z)$, plus the first, z^3, times the derivative of the second, $-\sin(z)$. That is,

$$P'(z) = 3z^2 \cos(z) + z^3(-\sin(z)) = 3z^2 \cos(z) - z^3 \sin(z).$$

□

Activity 2.7.2 Use the product rule to answer each of the questions below. Throughout, be sure to carefully label any derivative you find by name. It is not necessary to algebraically simplify any of the derivatives you compute.

a. Let $m(w) = 3w^{17}4^w$. Find $m'(w)$.

b. Let $h(t) = (\sin(t) + \cos(t))t^4$. Find $h'(t)$.

c. Determine the slope of the tangent line to the curve $y = f(x)$ at the point where $a = 1$ if f is given by the rule $f(x) = e^x \sin(x)$.

d. Find the equation of the tangent line to the function $y = g(x)$ at the point where $a = -1$ if g is given by the rule $g(x) = (x^2 + x)2^x$.

2.7.2 The quotient rule

Because quotients and products are closely linked, we can use the product rule to understand how to take the derivative of a quotient. Let $Q(x)$ be defined by $Q(x) = f(x)/g(x)$, where f and g are both differentiable functions. It turns out that Q is differentiable everywhere that $g(x) \neq 0$. We would like a formula for Q' in terms of f, g, f', and g'. Multiplying both sides of the formula $Q = f/g$ by g, we observe that

$$f(x) = Q(x) \cdot g(x).$$

Now we can use the product rule to differentiate f.

$$f'(x) = Q'(x)g(x) + Q(x)g'(x).$$

We want to know a formula for Q', so we solve this equation for $Q'(x)$.

$$Q'(x)g(x) = f'(x) - Q(x)g'(x)$$

and dividing both sides by $g(x)$, we have

$$Q'(x) = \frac{f'(x) - Q(x)g'(x)}{g(x)}.$$

Finally, we recall that $Q(x) = \frac{f(x)}{g(x)}$. Substituting this expression in the preceding equation, we have

$$Q'(x) = \frac{f'(x) - \frac{f(x)}{g(x)}g'(x)}{g(x)}$$

$$= \frac{f'(x) - \frac{f(x)}{g(x)}g'(x)}{g(x)} \cdot \frac{g(x)}{g(x)}$$
$$= \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

This calculation gives us the *quotient rule*.

Quotient Rule.

If f and g are differentiable functions, then their quotient $Q(x) = \frac{f(x)}{g(x)}$ is also a differentiable function for all x where $g(x) \neq 0$ and

$$Q'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

As with the product rule, it can be helpful to think of the quotient rule verbally. If a function Q is the quotient of a top function f and a bottom function g, then Q' is given by "the derivative of the top times the bottom, minus the top times the derivative of the bottom, all over the bottom squared." Note that remembering the order of the product rule given in the previous subsection can help in remembering the correct order for the numerator of the quotient rule.

Example 2.7.3 If $Q(t) = \sin(t)/2^t$, we call $\sin(t)$ the top function and 2^t the bottom function. By the quotient rule, Q' is given by the derivative of the top, $\cos(t)$ times the bottom, 2^t, minus the top, $\sin(t)$, times the derivative of the bottom, $2^t \ln(2)$, all over the bottom squared, $(2^t)^2$. That is,

$$Q'(t) = \frac{\cos(t)2^t - \sin(t)2^t \ln(2)}{(2^t)^2}.$$

In this particular example, it is possible to simplify $Q'(t)$ by removing a factor of 2^t from both the numerator and denominator, so that

$$Q'(t) = \frac{\cos(t) - \sin(t)\ln(2)}{2^t}.$$

□

In general, we must be careful in doing any such simplification, as we don't want to execute the quotient rule correctly but then make an algebra error.

Activity 2.7.3 Use the quotient rule to answer each of the questions below. Throughout, be sure to carefully label any derivative you find by name. That is, if you're given a formula for $f(x)$, clearly label the formula you find for $f'(x)$. It is not necessary to algebraically simplify any of the derivatives you compute.

a. Let $r(z) = \frac{3^z}{z^4+1}$. Find $r'(z)$.

b. Let $v(t) = \frac{\sin(t)}{\cos(t)+t^2}$. Find $v'(t)$.

c. Determine the slope of the tangent line to the curve $R(x) = \dfrac{x^2 - 2x - 8}{x^2 - 9}$ at the point where $x = 0$.

d. When a camera flashes, the intensity I of light seen by the eye is given by the function

$$I(t) = \frac{100t}{e^t},$$

where I is measured in candles and t is measured in milliseconds. Compute $I'(0.5)$, $I'(2)$, and $I'(5)$; include appropriate units on each value; and discuss the meaning of each.

2.7.3 Combining rules

In order to apply the derivative shortcut rules correctly we must recognize the fundamental structure of a function.

Example 2.7.4 Determine the derivative of the function

$$f(x) = x\sin(x) + \frac{x^2}{\cos(x)+2}.$$

How do we decide which rules to apply? Our first task is to recognize the structure of the function. This function f is a sum of two slightly less complicated functions, so we can apply the sum rule[2] to get

$$\begin{aligned} f'(x) &= \frac{d}{dx}\left[x\sin(x) + \frac{x^2}{\cos(x)+2}\right] \\ &= \frac{d}{dx}\left[x\sin(x)\right] + \frac{d}{dx}\left[\frac{x^2}{\cos(x)+2}\right] \end{aligned}$$

Now, the left-hand term above is a product, so the product rule is needed there, while the right-hand term is a quotient, so the quotient rule is required. Applying these rules respectively, we find that

$$\begin{aligned} f'(x) &= \ (x\cos(x)+\sin(x)) + \frac{(\cos(x)+2)2x - x^2(-\sin(x))}{(\cos(x)+2)^2} \\ &= x\cos(x)+\sin(x) + \frac{2x\cos(x)+4x+x^2\sin(x)}{(\cos(x)+2)^2}. \end{aligned}$$

□

Example 2.7.5 Differentiate

$$s(y) = \frac{y\cdot 7^y}{y^2+1}.$$

The function s is a quotient of two simpler functions, so the quotient rule will be needed. To begin, we set up the quotient rule and use the notation $\frac{d}{dy}$ to indicate the derivatives of the numerator and denominator. Thus,

$$s'(y) = \frac{\frac{d}{dy}\left[y\cdot 7^y\right]\cdot(y^2+1) - y\cdot 7^y\cdot\frac{d}{dy}\left[y^2+1\right]}{(y^2+1)^2}.$$

Now, there remain two derivatives to calculate. The first one, $\frac{d}{dy}\left[y\cdot 7^y\right]$ calls for use of the product rule, while the second, $\frac{d}{dy}\left[y^2+1\right]$ needs only the sum rule. Applying these rules, we now have

$$s'(y) = \frac{[1\cdot 7^y + y\cdot 7^y\ln(7)](y^2+1) - y\cdot 7^y[2y]}{(y^2+1)^2}.$$

While some simplification is possible, we are content to leave $s'(y)$ in its current form. □

[2]When taking a derivative that involves the use of multiple derivative rules, it is often helpful to use the notation $\frac{d}{dx}\left[\ \right]$ to wait to apply subsequent rules. This is demonstrated in each of the two examples presented here.

Success in applying derivative rules begins with recognizing the structure of the function, followed by the careful and diligent application of the relevant derivative rules. The best way to become proficient at this process is to do a large number of examples.

Activity 2.7.4 Use relevant derivative rules to answer each of the questions below. Throughout, be sure to use proper notation and carefully label any derivative you find by name.

a. Let $f(r) = (5r^3 + \sin(r))(4^r - 2\cos(r))$. Find $f'(r)$.

b. Let $p(t) = \dfrac{\cos(t)}{t^6 \cdot 6^t}$. Find $p'(t)$.

c. Let $g(z) = 3z^7 e^z - 2z^2 \sin(z) + \frac{z}{z^2+1}$. Find $g'(z)$.

d. A moving particle has its position in feet at time t in seconds given by the function $s(t) = \frac{3\cos(t) - \sin(t)}{e^t}$. Find the particle's instantaneous velocity at the moment $t = 1$.

e. Suppose that $f(x)$ and $g(x)$ are differentiable functions and it is known that $f(3) = -2$, $f'(3) = 7$, $g(3) = 4$, and $g'(3) = -1$. If $p(x) = f(x) \cdot g(x)$ and $q(x) = \dfrac{f(x)}{g(x)}$, calculate $p'(3)$ and $q'(3)$.

As the algebraic complexity of the functions we are able to differentiate continues to increase, it is important to remember that all of the derivative's meaning continues to hold. Regardless of the structure of the function f, the value of $f'(a)$ tells us the instantaneous rate of change of f with respect to x at the moment $x = a$, as well as the slope of the tangent line to $y = f(x)$ at the point $(a, f(a))$.

2.7.4 Summary

- **Question 2.7.6** How does the algebraic structure of a function guide us in computing its derivative using shortcut rules? □

  If a function is a sum, product, or quotient of simpler functions, then we can use the sum, product, or quotient rules to differentiate it in terms of the simpler functions and their derivatives.

- **Question 2.7.7** How do we compute the derivative of a product of two basic functions in terms of the derivatives of the basic functions? □

  The product rule tells us that if P is a product of differentiable functions f and g according to the rule $P(x) = f(x)g(x)$, then

  $$P'(x) = f'(x)g(x) + f(x)g'(x).$$

- **Question 2.7.8** How do we compute the derivative of a quotient of two basic functions in terms of the derivatives of the basic functions? □

  The quotient rule tells us that if Q is a quotient of differentiable functions f and g according to the rule $Q(x) = \frac{f(x)}{g(x)}$, then

  $$Q'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

- **Question 2.7.9** How do the product and quotient rules combine with the sum and constant multiple rules to expand the library of functions we can differentiate quickly? □

Along with the constant multiple and sum rules, the product and quotient rules enable us to compute the derivative of any function that consists of sums, constant multiples, products, and quotients of basic functions. For instance, if F has the form

$$F(x) = \frac{2a(x) - 5b(x)}{c(x) \cdot d(x)},$$

then F is a quotient, in which the numerator is a sum of constant multiples and the denominator is a product. Thus, the derivative of F can be found by applying the quotient rule and then using the sum and constant multiple rules to differentiate the numerator and the product rule to differentiate the denominator.

2.7.5 Exercises

1. Show that $\frac{d}{dx}[\tan(x)] = \sec^2(x)$.

HINT: Recall that $\tan(x) = \frac{\sin(x)}{\cos(x)}$ and $\sec(x) = \frac{1}{\cos(x)}$.

2. Let f and g be differentiable functions for which the following information is known: $f(2) = 5$, $g(2) = -3$, $f'(2) = -1/2$, $g'(2) = 2$.

a. Let h be the new function defined by the rule $h(x) = g(x) \cdot f(x)$. Determine $h(2)$ and $h'(2)$.

b. Find an equation for the tangent line to $y = h(x)$ at the point $(2, h(2))$ (where h is the function defined in (a)).

c. Let r be the function defined by the rule $r(x) = \frac{g(x)}{f(x)}$. Is r increasing, decreasing, or neither at $a = 2$? Why?

d. Find an equation for the tangent line to $y = r(x)$ at the point $(2, r(2))$ (where r is the function defined in (c)).

3. Let functions p and q be the piecewise linear functions given by their respective graphs in Figure 2.7.10. Use the graphs to answer the following questions.

a. Let $r(x) = p(x) \cdot q(x)$. Determine $r'(-2)$ and $r'(0)$.

b. Find an equation for the tangent line to $y = r(x)$ at the point $(2, r(2))$.

c. Let $z(x) = \frac{q(x)}{p(x)}$. Determine $z'(0)$ and $z'(2)$.

d. Let $\ell(x) = p(x) \cdot e^x$. Determine $\ell'(0)$.

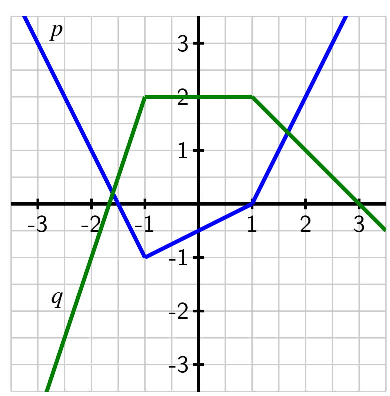


Figure 2.7.10 The graphs of p (in blue) and q (in green).

4. A farmer with large land holdings has historically grown a wide variety of crops. With the price of ethanol fuel rising, he decides that it would be prudent to devote more and more of his acreage to producing corn. As he grows more and more corn, he learns efficiencies that increase his yield per acre. In the present year, he used 7000 acres of his land to grow corn, and that land had an average yield of 170 bushels per acre. At the current time, he plans to increase his number of acres devoted to growing corn at a rate of 600 acres/year, and he expects that right now his average yield is increasing at a rate of 8 bushels per acre per year. Use this information to answer the following questions.

a. Say that the present year is $t = 0$, that $A(t)$ denotes the number of acres the farmer devotes to growing corn in year t, $Y(t)$ represents the average yield in year t (measured in bushels per acre), and $C(t)$ is the total number of bushels of corn the farmer produces. What is the formula for $C(t)$ in terms of $A(t)$ and $Y(t)$? Why?

b. What is the value of $C(0)$? What does it measure?

c. Write an expression for $C'(t)$ in terms of $A(t)$, $A'(t)$, $Y(t)$, and $Y'(t)$. Explain your thinking.

d. What is the value of $C'(0)$? What does it measure?

e. Based on the given information and your work above, write the equation of the tangent line to $C(t)$ at the point $(0, C(0))$.

5. You will find unit conversions (Section 1.3) to be especially helpful in answering the following questions.

Let $f(v)$ be the gas consumption (in liters/km) of a car going at velocity v (in km/hour). In other words, $f(v)$ tells you how many liters of gas the car uses to go one kilometer if it is traveling at v kilometers per hour. In addition, suppose that $f(80) = 0.05$ and $f'(80) = 0.0004$.

a. Let $g(v)$ be the distance the same car goes on one liter of gas at velocity v. What is the relationship between $f(v)$ and $g(v)$? Hence find $g(80)$ and $g'(80)$.

b. Let $h(v)$ be the gas consumption in liters per hour of a car going at velocity v. In other words, $h(v)$ tells you how many liters of gas the car uses in one hour if it is going at velocity v. What is the algebraic relationship between $h(v)$ and $f(v)$? Hence find $h(80)$ and $h'(80)$.

c. How would you explain the practical meaning of these function and derivative values to a driver who knows no calculus? Include units on each of the function and derivative values you discuss in your response.

2.8 Derivatives of Compositions

Motivating Questions

- What is a composite function and how do we recognize its structure algebraically?
- Given a composite function $C(x) = f(g(x))$ that is built from differentiable functions f and g, how do we compute $C'(x)$ in terms of f, g, f', and g'? What is the statement of the Chain Rule?

In addition to learning how to differentiate a variety of basic functions, we have also been developing our ability to use rules to differentiate certain algebraic combinations of them.

Example 2.8.1 State the rule(s) to find the derivative of each of the following combinations of $f(x) = \sin(x)$ and $g(x) = x^2$:

$$s(x) = 3x^2 - 5\sin(x),$$

$$p(x) = x^2 \sin(x), \text{and}$$

$$q(x) = \frac{\sin(x)}{x^2}.$$

Solution. Finding s' uses the sum and constant multiple rules, because $s(x) = 3g(x) - 5f(x)$. Determining p' requires the product rule, because $p(x) = g(x) \cdot f(x)$. To calculate q' we use the quotient rule, because $q(x) = \frac{f(x)}{g(x)}$. □

There is one more natural way to combine basic functions algebraically, and that is by *composing* them. For instance, let's consider the function

$$C(x) = \sin(x^2),$$

and observe that any input x passes through a *chain* of functions. In the process that defines the function $C(x)$, x is first squared, and then the sine of the result is taken. Using an arrow diagram,

$$x \longrightarrow x^2 \longrightarrow \sin(x^2).$$

In terms of the elementary functions f and g, we observe that x is the input for the function g, and the result is used as the input for f. We write

$$C(x) = f(g(x)) = \sin(x^2)$$

and say that C is the **composition** of f and g (see Subsection 1.2.4). We will refer to g, the function that is first applied to x, as the *inner* function, while f, the function that is applied to the result, is the *outer* function.

Given a composite function $C(x) = f(g(x))$ that is built from differentiable functions f and g, how do we compute $C'(x)$ in terms of f, g, f', and g'? In the same way that the rate of change of a product of two functions, $p(x) = f(x) \cdot g(x)$, depends on the behavior of both f and g, it makes sense intuitively that the rate of change of a composite function $C(x) = f(g(x))$ will also depend on some combination of f and g and their derivatives. The rule that describes how to compute C' in terms of f and g and their derivatives is called the *chain rule*.

But before we can learn what the chain rule says and why it works, we first need to be comfortable decomposing composite functions so that we can correctly identify the inner and outer functions, as we did in the example above with $C(x) = \sin(x^2)$.

Warm-Up 2.8.1 For each function given below, identify its fundamental algebraic structure. In particular, is the given function a sum, product, quotient, or composition of basic functions? If the function is a composition of basic functions, state a formula for the inner function g and the outer function f so that the overall composite function can be written in the form $f(g(x))$. If the function is a sum, product, or quotient of basic functions, use the appropriate rule to determine its derivative.

a. $h(x) = \tan(2^x)$

b. $p(x) = 2^x \tan(x)$

c. $r(x) = (\tan(x))^2$

d. $m(x) = e^{\tan(x)}$

e. $w(x) = \sqrt{x} + \tan(x)$

f. $z(x) = \sqrt{\tan(x)}$

2.8.1 The chain rule

Often a composite function cannot be written in an alternate algebraic form. For instance, the function $C(x) = \sin(x^2)$ cannot be expanded or otherwise rewritten, so it presents no alternate approaches to taking the derivative. But some composite functions can be expanded or simplified, and these provide a way to explore how the chain rule works.

Example 2.8.2 Let $f(x) = -4x + 7$ and $g(x) = 3x - 5$. Determine a formula for $C(x) = f(g(x))$ and compute $C'(x)$. How is C' related to f and g and their derivatives?

Solution. By the rules given for f and g,

$$\begin{aligned} C(x) &= f(g(x)) \\ &= f(3x-5) \\ &= -4(3x-5)+7 \\ &= -12x+20+7 \\ &= -12x+27. \end{aligned}$$

Thus, $C'(x) = -12$. Noting that $f'(x) = -4$ and $g'(x) = 3$, we observe that C' appears to be the product of f' and g'. □

It may seem that Example 2.8.2 is too elementary to illustrate how to differentiate a composite fuction. Linear functions are the simplest of all functions, and composing linear functions yields another linear function. While this example does not illustrate the full complexity of a composition of nonlinear functions, at the same time we will see in Section 3.1 that any differentiable function is *locally* linear, which essentially means that any function with a derivative behaves like a line when viewed up close. The fact that the derivatives of the linear functions f and g are multiplied to find the derivative of their composition turns out to be a key insight.

We now consider a composition involving a nonlinear function.

Example 2.8.3 Let $C(x) = \sin(2x)$. Use the double angle identity to rewrite C as a product of basic functions, and use the product rule to find C'. Rewrite C' in the simplest form possible.

Solution. Using the double angle identity for the sine function, we write

$$C(x) = \sin(2x) = 2\sin(x)\cos(x).$$

Applying the product rule and simplifying, we find

$$C'(x) = 2\sin(x)(-\sin(x)) + \cos(x)(2\cos(x)) = 2(\cos^2(x) - \sin^2(x)).$$

Next, we recall that the double angle identity for the cosine,

$$\cos(2x) = \cos^2(x) - \sin^2(x).$$

Substituting this result into our expression for $C'(x)$, we now have that

$$C'(x) = 2\cos(2x).$$

□

In Example 2.8.3, if we let $g(x) = 2x$ and $f(x) = \sin(x)$, we observe that $C(x) = f(g(x))$. Now, $g'(x) = 2$ and $f'(x) = \cos(x)$, so we can view the structure of $C'(x)$ as

$$C'(x) = 2\cos(2x) = g'(x)f'(g(x)).$$

In this example, as in the example involving linear functions, we see that the derivative of the composite function $C(x) = f(g(x))$ is found by multiplying the derivatives of f and g, but with f' evaluated at $g(x)$.

It makes sense intuitively that these two quantities are involved in the rate of change of a composite function: if we ask how fast C is changing at a given x value, it clearly matters how fast g is changing at x, as well as how fast f is changing at the value of $g(x)$. It turns out that this structure holds for all differentiable functions[1] as is stated in the Chain Rule.

Chain Rule.

If g is differentiable at x and f is differentiable at $g(x)$, then the composite function C defined by $C(x) = f(g(x))$ is differentiable at x and

$$C'(x) = f'(g(x))g'(x).$$

As with the product and quotient rules, it is often helpful to think verbally about what the chain rule says: "If C is a composite function defined by an outer function f and an inner function g, then C' is given by the derivative of the outer function evaluated at the inner function, times the derivative of the inner function."

It is helpful to identify clearly the inner function g and outer function f, compute their derivatives individually, and then put all of the pieces together by the chain rule.

Example 2.8.4 Determine the derivative of the function

$$r(x) = (\tan(x))^2.$$

Solution. The function r is composite, with inner function $g(x) = \tan(x)$ and outer function $f(x) = x^2$. Organizing the key information involving f, g, and their derivatives, we have

$f(x) = x^2$	$g(x) = \tan(x)$
$f'(x) = 2x$	$g'(x) = \sec^2(x)$
$f'(g(x)) = 2\tan(x)$	

Applying the chain rule, we find that

$$r'(x) = f'(g(x))g'(x) = 2\tan(x)\sec^2(x).$$

□

As a side note, we remark that $r(x)$ is usually written as $\tan^2(x)$. This is common notation for powers of trigonometric functions: $\cos^4(x)$, $\sin^5(x)$, and $\sec^2(x)$ are all composite functions, with the outer function a power function and the inner function a trigonometric one.

Activity 2.8.2 For each function given below, identify an inner function g and outer function f to write the function in the form $f(g(x))$. Determine $f'(x)$, $g'(x)$, and $f'(g(x))$, and then apply the chain rule to determine the derivative

[1] Like other differentiation rules, the Chain Rule can be proved formally using the limit definition of the derivative.

of the given function.

a. $h(x) = \cos(x^4)$

b. $p(x) = \sqrt{\tan(x)}$

c. $s(x) = 2^{\sin(x)}$

d. $z(x) = \cos^4(x)$

e. $m(x) = (\sqrt{x} + e^x)^9$

2.8.2 Using multiple rules simultaneously

The chain rule now joins the sum, constant multiple, product, and quotient rules in our collection of techniques for finding the derivative of a function through understanding its algebraic structure and the basic functions that constitute it. It takes practice to get comfortable applying multiple rules to differentiate a single function, but using proper notation and taking a few extra steps will help.

Example 2.8.5 Find a formula for the derivative of $h(t) = 3^{t^2+2t}\sin^4(t)$.

Solution. We first observe that h is the product of two functions: $h(t) = a(t) \cdot b(t)$, where $a(t) = 3^{t^2+2t}$ and $b(t) = \sin^4(t)$. We will need to use the product rule to differentiate h. And because a and b are composite functions, we will need the chain rule. We therefore begin by computing $a'(t)$ and $b'(t)$.

Writing $a(t) = f(g(t)) = 3^{t^2+2t}$, and finding the derivatives of f and g, we have

$$\begin{aligned} f(t) &= 3^t & g(t) &= t^2 + 2t \\ f'(t) &= 3^t \ln(3) & g'(t) &= 2t + 2 \\ f'(g(t)) &= 3^{t^2+2t}\ln(3) \end{aligned}$$

Thus, by the chain rule, it follows that $a'(t) = f'(g(t))g'(t) = 3^{t^2+2t}\ln(3)(2t+2)$.

Turning next to b, we write $b(t) = r(s(t)) = \sin^4(t)$ and find the derivatives of r and s.

$$\begin{aligned} r(t) &= t^4 & s(t) &= \sin(t) \\ r'(t) &= 4t^3 & s'(t) &= \cos(t) \\ r'(s(t)) &= 4\sin^3(t) \end{aligned}$$

By the chain rule,

$$b'(t) = r'(s(t))s'(t) = 4\sin^3(t)\cos(t).$$

Now we are finally ready to compute the derivative of the function h. Recalling that $h(t) = 3^{t^2+2t}\sin^4(t)$, by the product rule we have

$$h'(t) = \frac{d}{dt}[3^{t^2+2t}]\sin^4(t) + 3^{t^2+2t}\frac{d}{dt}[\sin^4(t)].$$

From our work above with a and b, we know the derivatives of 3^{t^2+2t} and $\sin^4(t)$, and therefore

$$h'(t) = 3^{t^2+2t}\ln(3)(2t+2)\sin^4(t) + 3^{t^2+2t}4\sin^3(t)\cos(t).$$

□

Activity 2.8.3 For each of the following functions, find the function's derivative. State the rule(s) you use, label relevant derivatives appropriately, and be sure to clearly identify your overall answer.

a. $p(r) = 4\sqrt{r^6 + 2e^r}$

b. $m(v) = \sin(v^2)\cos(v^3)$

c. $h(y) = \dfrac{\cos(10y)}{e^{4y} + 1}$

d. $s(z) = 2^{z^2 \tan(z)}$

e. $c(x) = \sin(e^{x^2})$

The chain rule now adds substantially to our ability to compute derivatives. Whether we are finding the equation of the tangent line to a curve, the instantaneous velocity of a moving particle, or the instantaneous rate of change of a certain quantity, if the function under consideration is a composition, the chain rule is indispensable.

Activity 2.8.4 Use known derivative rules, including the chain rule, as needed to answer each of the following questions.

a. Find an equation for the tangent line to the curve $y = \sqrt{e^x + 3}$ at the point where $x = 0$.

b. If $s(t) = \dfrac{1}{(t^2 + 1)^3}$ represents the position function of a particle moving horizontally along an axis at time t (where s is measured in inches and t in seconds), find the particle's instantaneous velocity at $t = 1$. Is the particle moving to the left or right at that instant?

c. At sea level, air pressure is 30 inches of mercury. At an altitude of h feet above sea level, the air pressure, P, in inches of mercury, is given by the function $P = 30e^{-0.0000323h}$. Compute dP/dh and explain what this derivative function tells you about air pressure, including a discussion of the units on dP/dh. In addition, determine how fast the air pressure is changing for a pilot of a small plane passing through an altitude of 1000 feet.

d. Suppose that $f(x)$ and $g(x)$ are differentiable functions and that the following information about them is known:

Table 2.8.6 Data for functions f and g.

x	$f(x)$	$f'(x)$	$g(x)$	$g'(x)$
-1	2	-5	-3	4
2	-3	4	-1	2

If $C(x)$ is a function given by the formula $f(g(x))$, determine $C'(2)$. In addition, if $D(x)$ is the function $f(f(x))$, find $D'(-1)$.

2.8.3 The composite version of basic function rules

As we gain more experience with differention, we will become more comfortable in simply writing down the derivative without taking multiple steps. This is particularly simple when the inner function is linear, since the derivative of a linear function is a constant.

Example 2.8.7 Use the chain rule to differentiate each of the following composite functions whose inside function is linear:

$$\frac{d}{dx}\left[(5x+7)^{10}\right] = 10(5x+7)^9 \cdot 5,$$

$$\frac{d}{dx}\left[\tan(17x)\right] = 17\sec^2(17x), \text{ and}$$

$$\frac{d}{dx}\left[e^{-3x}\right] = -3e^{-3x}.$$

□

More generally, following is an excellent exercise for getting comfortable with the derivative rules. Write down a list of all the basic functions whose derivatives we know, and list the derivatives. Then write a composite function with the inner function being an unknown function $u(x)$ and the outer function being a basic function. Finally, write the chain rule for the composite function. The following example illustrates this for two different functions.

Example 2.8.8 To determine

$$\frac{d}{dx}[\sin(u(x))],$$

where u is a differentiable function of x, we use the chain rule with the sine function as the outer function. Applying the chain rule, we find that

$$\frac{d}{dx}[\sin(u(x))] = \cos(u(x)) \cdot u'(x).$$

This rule is analogous to the basic derivative rule that $\frac{d}{dx}[\sin(x)] = \cos(x)$, and gives us a quick justification of the fact you showed using transformations in Exercise 2.6.3.3 that $\frac{d}{d\theta}[\sin\left(\theta + \frac{\pi}{2}\right)] = \cos\left(\theta + \frac{\pi}{2}\right)$.

Similarly, since $\frac{d}{dx}[a^x] = a^x \ln(a)$, it follows by the chain rule that

$$\frac{d}{dx}[a^{u(x)}] = a^{u(x)} \ln(a) \cdot u'(x).$$

This rule is analogous to the basic derivative rule that $\frac{d}{dx}[a^x] = a^x \ln(a)$. □

2.8.4 Summary

- **Question 2.8.9** What is a composite function and how do we recognize its structure algebraically? □

  A composite function is one where the input variable x first passes through one function, and then the resulting output passes through another. For example, the function $h(x) = 2^{\sin(x)}$ is composite since $x \longrightarrow \sin(x) \longrightarrow 2^{\sin(x)}$.

- **Question 2.8.10** Given a composite function $C(x) = f(g(x))$ that is built from differentiable functions f and g, how do we compute $C'(x)$ in terms of f, g, f', and g'? What is the statement of the Chain Rule? □

  Given a composite function $C(x) = f(g(x))$ where f and g are differentiable functions, the chain rule tells us that

  $$C'(x) = f'(g(x))g'(x).$$

2.8.5 Exercises

1. Consider the basic functions $f(x) = x^3$ and $g(x) = \sin(x)$.

a. Let $h(x) = f(g(x))$. Find the exact instantaneous rate of change of h at the point where $x = \frac{\pi}{4}$.

b. Which function is changing most rapidly at $x = 0.25$: $h(x) = f(g(x))$ or $r(x) = g(f(x))$? Why?

2. Let $u(x)$ be a differentiable function. For each of the following functions, determine the derivative. Each response will involve u and/or u'.

a. $p(x) = e^{u(x)}$

b. $q(x) = u(e^x)$

c. $r(x) = \cos(u(x))$

d. $s(x) = u(\cos(x))$

e. $a(x) = u(x^4)$

f. $b(x) = (u(x))^4$

3. Let functions p and q be the piecewise linear functions given by their respective graphs in Figure 2.8.11. Use the graphs to answer the following questions.

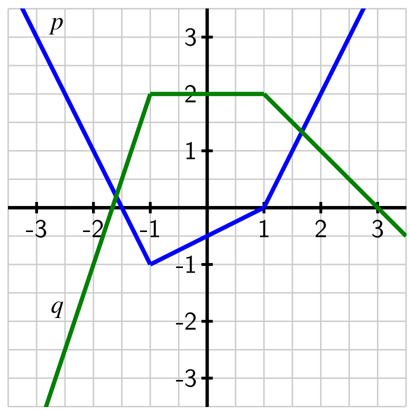


Figure 2.8.11 The graphs of p (in blue) and q (in green).

a. Let $C(x) = p(q(x))$. Determine $C'(0)$ and $C'(3)$.

b. Let $Y(x) = q(q(x))$ and $Z(x) = q(p(x))$. Determine $Y'(3)$ and $Z'(0)$.

4. If a spherical tank of radius 4 feet has h feet of water present in the tank, then the volume of water in the tank is given by the formula

$$V = \frac{\pi}{3}h^2(12 - h).$$

a. At what instantaneous rate is the volume of water in the tank changing with respect to the *height* of the water at the instant $h = 1$? What are the units on this quantity?

b. Now suppose that the height of water in the tank is being regulated by an inflow and outflow (e.g., a faucet and a drain) so that the height of the water at time t is given by the rule $h(t) = \sin(\pi t) + 1$, where t is measured in hours (and h is still measured in feet). At what rate is the height of the water changing with respect to time at the instant $t = 2$?

c. Continuing under the assumptions in (b), at what instantaneous rate is the volume of water in the tank changing with respect to *time* at the instant $t = 2$?

d. What are the main differences between the rates found in (a) and (c)? Include a discussion of the relevant units.

2.9 Derivatives of Inverse Functions

Motivating Questions

- What is the derivative of the natural logarithm function?
- If g is the inverse of a differentiable function f, how is g' computed in terms of f, f', and g?

Much of mathematics centers on the notion of a function. Indeed, throughout our study of calculus, we are investigating the behavior of functions, with particular emphasis on how fast the output of the function changes in response to changes in the input. Because each function represents a process, a natural question to ask is whether or not the particular process can be reversed. That is, if we know the output that results from the function, can we determine the input that led to it? And if we know how fast a particular process is changing, can we determine how fast the inverse process is changing?

One of the most important functions in all of mathematics is the natural exponential function $f(x) = e^x$. Its inverse, the natural logarithm, $g(x) = \ln(x)$, is similarly important. One of our goals in this section is to learn how to differentiate the logarithm function. First, we review some of the basic concepts surrounding functions and their inverses (see Subsection 1.2.5).

Warm-Up 2.9.1 The equation $y = \frac{5}{9}(x - 32)$ relates a temperature given in x degrees Fahrenheit to the corresponding temperature y measured in degrees Celsius.

a. Solve the equation $y = \frac{5}{9}(x-32)$ for x to write x (Fahrenheit temperature) in terms of y (Celsius temperature).

b. Let $C(x) = \frac{5}{9}(x-32)$ be the function that takes a Fahrenheit temperature as input and produces the Celsius temperature as output. In addition, let $F(y)$ be the function that converts a temperature given in y degrees Celsius to the temperature $F(y)$ measured in degrees Fahrenheit. Use your work in (a) to write a formula for $F(y)$.

c. Next consider the new function defined by $p(x) = F(C(x))$. Use the formulas for F and C to determine an expression for $p(x)$ and simplify this expression as much as possible. What do you observe?

d. Now, let $r(y) = C(F(y))$. Use the formulas for F and C to determine an expression for $r(y)$ and simplify this expression as much as possible. What do you observe?

e. What is the value of $C'(x)$? of $F'(y)$? How do these values appear to be related?

2.9.1 Basic facts about inverse functions

A function f is a rule that associates each element in its *domain* to one and only one element in its *range* . If the relationship g comprised of all points of the form $(f(x), x)$ is still a function, then we say that g is the *inverse* of f.

We often use the notation f^{-1} (read "f-inverse") to denote the inverse of f. The inverse function undoes the work of f. Indeed, if $y = f(x)$, then

$$f^{-1}(y) = f^{-1}(f(x)) = x.$$

Thus, the equations $y = f(x)$ and $x = f^{-1}(y)$ say the same thing. The only difference between the two equations is one of perspective — one is solved for x, while the other is solved for y.

Here we briefly remind ourselves of some key facts about inverse functions.

Note 2.9.1 For a function f,

- provided f^{-1} exists, the domain of f^{-1} is the range of f, and the range of f^{-1} is the domain of f;
- $f^{-1}(f(x)) = x$ for every x in the domain of f and $f(f^{-1}(y)) = y$ for every y in the range of f;
- $y = f(x)$ if and only if $x = f^{-1}(y)$.

The last fact reveals a special relationship between the graphs of f and f^{-1}. If a point (x, y) that lies on the graph of $y = f(x)$, then it is also true that $x = f^{-1}(y)$, which means that the point (y, x) lies on the graph of f^{-1}. This shows us that the graphs of f and f^{-1} are the reflections of each other across the line $y = x$, because this reflection is precisely the geometric action that swaps the coordinates in an ordered pair. In Figure 2.9.2, we see this illustrated by the function $y = f(x) = 2^x$ and its inverse, with the points $(-1, \frac{1}{2})$ and $(\frac{1}{2}, -1)$ highlighting the reflection of the curves across $y = x$.

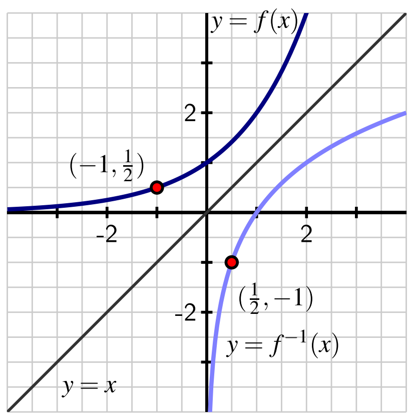


Figure 2.9.2 A graph of a function $y = f(x)$ along with its inverse, $y = f^{-1}(x)$.

To close our review of important facts about inverses, we recall that the natural exponential function $y = f(x) = e^x$ has an inverse function, namely the natural logarithm, $x = f^{-1}(y) = \ln(y)$. Thus, writing $y = e^x$ is interchangeable with $x = \ln(y)$, plus $\ln(e^x) = x$ for every real number x and $e^{\ln(y)} = y$ for every positive real number y.

2.9.2 The derivative of the natural logarithm function

In what follows, we find a formula for the derivative of $g(x) = \ln(x)$. To do so, we take advantage of the fact that we know the derivative of the natural exponential function, the inverse of g. In particular, we know that writing $g(x) = \ln(x)$ is equivalent to writing $e^{g(x)} = x$. Now we differentiate both sides of this equation and observe that

$$\frac{d}{dx}\left[e^{g(x)}\right] = \frac{d}{dx}[x].$$

The righthand side is simply 1; by applying the chain rule to the left side, we find that

$$e^{g(x)}g'(x) = 1.$$

Next we solve for $g'(x)$, to get

$$g'(x) = \frac{1}{e^{g(x)}}.$$

Finally, we recall that $g(x) = \ln(x)$, so $e^{g(x)} = e^{\ln(x)} = x$, and thus

$$g'(x) = \frac{1}{x}.$$

Natural Logarithm.

For all positive real numbers x, $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$.

This rule for the natural logarithm function now joins our list of basic derivative rules. Note that this rule applies only to positive values of x, as these are the only values for which $\ln(x)$ is defined.

Also notice that for the first time in our work, differentiating a basic function of a particular type has led to a function of a very different nature: the derivative of the natural logarithm is not another logarithm, nor even an exponential function, but rather a rational one.

Derivatives of logarithms may now be computed in concert with all of the rules known to date. For instance, if $f(t) = \ln(t^2 + 1)$, then by the chain rule, $f'(t) = \frac{1}{t^2+1} \cdot 2t$.

There are interesting connections between the graphs of $f(x) = e^x$ and $f^{-1}(x) = \ln(x)$.

In Figure 2.9.3, we are reminded that since the natural exponential function has the property that its derivative is itself, the slope of the tangent to $y = e^x$ is equal to the height of the curve at that point. For instance, at the point $A = (\ln(0.5), 0.5)$, the slope of the tangent line is $m_A = 0.5$, and at $B = (\ln(5), 5)$, the tangent line's slope is $m_B = 5$.

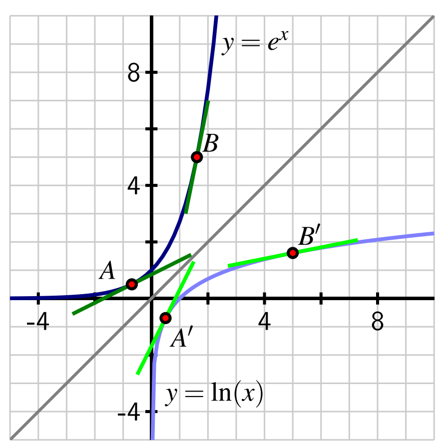


Figure 2.9.3 A graph of the function $y = e^x$ along with its inverse, $y = \ln(x)$, where both functions are viewed using the input variable x.

At the corresponding points A' and B' on the graph of the natural logarithm function (which come from reflecting A and B across the line $y = x$), we know that the slope of the tangent line is the reciprocal of the x-coordinate of the point (since $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$). Thus, at $A' = (0.5, \ln(0.5))$, we have $m_{A'} = \frac{1}{0.5} = 2$, and at $B' = (5, \ln(5))$, $m_{B'} = \frac{1}{5}$.

In particular, we observe that $m_{A'} = \frac{1}{m_A}$ and $m_{B'} = \frac{1}{m_B}$. This is not a coincidence, but in fact holds for any curve $y = f(x)$ and its inverse, provided the inverse exists. This is due to the reflection across $y = x$. It changes the roles of x and y, thus reversing the rise and run, so the slope of the inverse function at the reflected point is the reciprocal of the slope of the original function.

Activity 2.9.2 For each function given below, find its derivative.

a. $h(x) = x^2 \ln(x)$

b. $p(t) = \dfrac{\ln(t)}{e^t + 1}$

c. $s(y) = \ln(\cos(y) + 2)$

d. $z(x) = \tan(\ln(x))$

e. $m(z) = \ln(\ln(z))$

2.9.3 The link between the derivative of a function and the derivative of its inverse

In Figure 2.9.3, we saw an interesting relationship between the slopes of tangent lines to the natural exponential and natural logarithm functions at points reflected across the line $y = x$. In particular, we observed that at the point $(\ln(2), 2)$ on the graph of $f(x) = e^x$, the slope of the tangent line is $f'(\ln(2)) = 2$, while at the corresponding point $(2, \ln(2))$ on the graph of $f^{-1}(x) = \ln(x)$, the slope of the tangent line is $(f^{-1})'(2) = \frac{1}{2}$, which is the reciprocal of $f'(\ln(2))$.

That the two corresponding tangent lines have reciprocal slopes is not a coincidence. If f and g are differentiable inverse functions, then $y = f(x)$ if and only if $x = g(y)$, then$f(g(x)) = x$ for every x in the domain of f^{-1}. Differentiating both sides of this equation, we have

$$\frac{d}{dx}[f(g(x))] = \frac{d}{dx}[x],$$

and by the chain rule,

$$f'(g(x))g'(x) = 1.$$

Solving for $g'(x)$, we have $g'(x) = \frac{1}{f'(g(x))}$. Here we see that the slope of the tangent line to the inverse function g at the point $(x, g(x))$ is precisely the reciprocal of the slope of the tangent line to the original function f at the point $(g(x), f(g(x))) = (g(x), x)$.

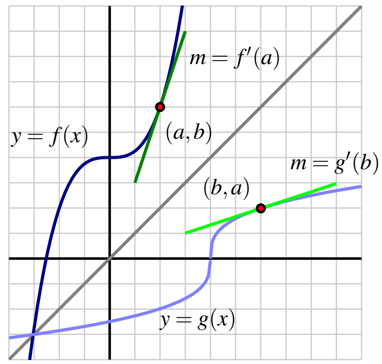


Figure 2.9.4 A graph of function $y = f(x)$ along with its inverse, $y = g(x) = f^{-1}(x)$. Observe that the slopes of the two tangent lines are reciprocals of one another.

To see this more clearly, consider the graph of the function $y = f(x)$ shown in Figure 2.9.4, along with its inverse $y = g(x)$. Given a point (a, b) that lies on the graph of f, we know that (b, a) lies on the graph of g; because $f(a) = b$ and $g(b) = a$. Now, applying the rule that $g'(x) = 1/f'(g(x))$ to the value $x = b$, we have

$$g'(b) = \frac{1}{f'(g(b))} = \frac{1}{f'(a)},$$

which is precisely what we see in the figure: the slope of the tangent line to g at (b, a) is the reciprocal of the slope of the tangent line to f at (a, b), since these two lines are reflections of one another across the line $y = x$.

Derivative of an inverse function.

Suppose that f is a differentiable function with inverse g and that (a, b) is a point that lies on the graph of f at which $f'(a) \neq 0$. Then

$$g'(b) = \frac{1}{f'(a)}.$$

More generally, for any x in the domain of g', we have $g'(x) = 1/f'(g(x))$.

The rule we derived for $\ln(x)$ is just a specific example of this general property of the derivative of an inverse function. Indeed, with $g(x) = \ln(x)$ and $f(x) = e^x$, it follows that

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{e^{\ln(x)}} = \frac{1}{x}.$$

2.9.4 Summary

- **Question 2.9.5** What is the derivative of the natural logarithm function? □

  For all positive real numbers x, $\frac{d}{dx}[\ln(x)] = \frac{1}{x}$.

- **Question 2.9.6** If g is the inverse of a differentiable function f, how is g' computed in terms of f, f', and g? □

If g is the inverse of a differentiable function f, then for any point x in the domain of g', $g'(x) = \frac{1}{f'(g(x))}$.

2.9.5 Exercises

1. Determine the general rule for derivatives of logarithmic functions. That is, let $b > 0$, $b \neq 1$, and $\ell(x) = \log_b(x)$. Determine $\ell'(x)$.

HINT: Let $f(x) = b^x$. Then $f^{-1}(x) = \ell(x) = \log_b(x)$.

2. Consider the graph of $y = f(x)$ provided in Figure 2.9.7 and use it to answer the following questions.

a. Use the provided graph to estimate the value of $f'(1)$.

b. Sketch an approximate graph of $y = f^{-1}(x)$. Label at least three distinct points on the graph that correspond to three points on the graph of f.

c. Based on your work in (a), what is the value of $(f^{-1})'(-1)$? Why?

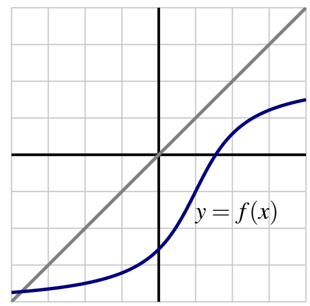


Figure 2.9.7 A function $y = f(x)$

3. Let $f(x) = \frac{1}{4}x^3 + 4$.

a. Sketch a graph of $y = f(x)$ and explain why f is an invertible function.

b. Let g be the inverse of f and determine a formula for g.

c. Compute $f'(x)$, $g'(x)$, $f'(2)$, and $g'(6)$. What is the special relationship between $f'(2)$ and $g'(6)$? Why?

4. Let $h(x) = x + \sin(x)$.

a. Use technology to sketch a graph of $y = h(x)$ and explain why h must be invertible.

b. Explain why it does not appear to be algebraically possible to determine a formula for h^{-1}.

c. Observe that the point $(\frac{\pi}{2}, \frac{\pi}{2} + 1)$ lies on the graph of $y = h(x)$. Determine the value of $(h^{-1})'\left(\frac{\pi}{2} + 1\right)$.

Chapter 3

Using the Derivative

3.1 Linear and Quadratic Approximation

Motivating Questions

- What is the formula for the general tangent line approximation to a differentiable function $y = f(x)$ at the point $(a, f(a))$?
- What is the principle of local linearity and what is the linear approximation (or local linearization) of a differentiable function f at a point $(a, f(a))$?
- How does knowing just the tangent line approximation tell us information about the behavior of the original function itself near the point of approximation?
- How does knowing the second derivative's value at this point provide us additional knowledge of the original function's behavior?

Among all functions, linear functions are simplest. One of the powerful consequences of a function $y = f(x)$ being differentiable at a point $(a, f(a))$ is that, up close, the function $y = f(x)$ is locally linear and looks like its tangent line at that point. In certain circumstances, this allows us to approximate the original function f with a simpler function L that is linear: this can be advantageous when we have limited information about f or when f is computationally or algebraically complicated. We will explore all of these situations in what follows.

It is essential to recall that when f is differentiable at $x = a$, the value of $f'(a)$ provides the slope of the tangent line to $y = f(x)$ at the point $(a, f(a))$. If we know both a point on the line and the slope of the line we can find the equation of the tangent line and write the equation in point-slope form[1].

Warm-Up 3.1.1 Consider the function $y = g(x) = -x^2 + 3x + 2$.

a. Use derivative rules to compute a formula for $y = g'(x)$.

b. Determine the slope of the tangent line to $y = g(x)$ at the value $x = 2$.

c. Compute $g(2)$.

d. Find an equation for the tangent line to $y = g(x)$ at the point $(2, g(2))$. Write your result in point-slope form.

[1] Recall that a line with slope m that passes through (x_0, y_0) has equation $y - y_0 = m(x - x_0)$, and this is the *point-slope form* of the equation.

e. On the axes provided in Figure 3.1.1, sketch an accurate, labeled graph of $y = g(x)$ along with its tangent line at the point $(2, g(2))$.

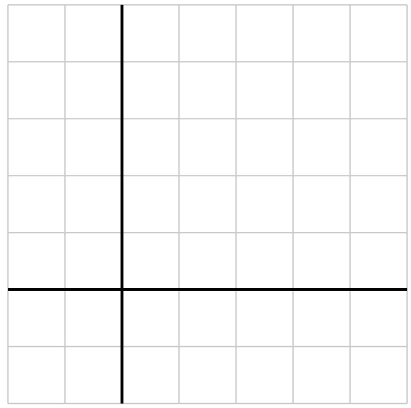


Figure 3.1.1 Axes for plotting $y = g(x)$ and its tangent line to the point $(2, g(2))$.

3.1.1 The tangent line

Given a function f that is differentiable at $x = a$, we know that we can determine the slope of the tangent line to $y = f(x)$ at $(a, f(a))$ by computing $f'(a)$. The equation of the resulting tangent line is given in point-slope form by

$$y - f(a) = f'(a)(x - a) \quad \text{or} \quad y = f'(a)(x - a) + f(a).$$

Note well: there is a major difference between $f(a)$ and $f(x)$ in this context. The former is a constant that results from using the given fixed value of a (called the **center**), while the latter is the general expression for the rule that defines the function. The same is true for $f'(a)$ and $f'(x)$: we must carefully distinguish between these expressions. Each time we find the tangent line, we need to evaluate the function and its derivative at the *center*: a fixed a-value.

In Figure 3.1.2, we see the graph of a function f and its tangent line at the point $(a, f(a))$. Notice how when we zoom in we see the local linearity of f more clearly highlighted. The function and its tangent line are nearly indistinguishable up close.

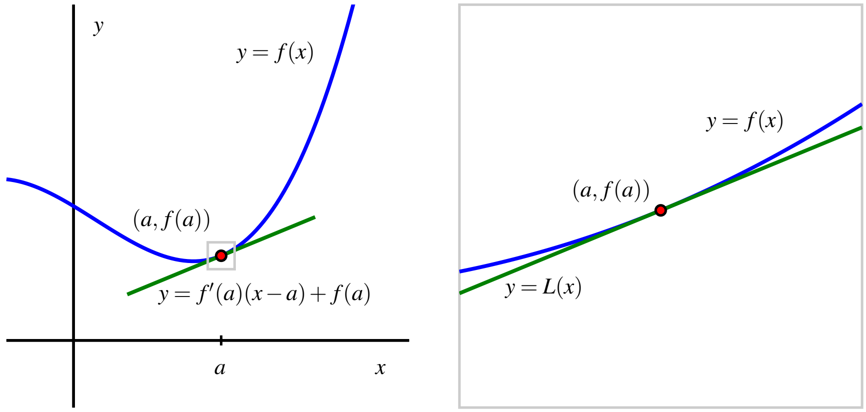


Figure 3.1.2 A function $y = f(x)$ and its tangent line at the point $(a, f(a))$: at left, from a distance, and at right, up close. At right, we label the tangent line function by $y = L(x)$ and observe that for x near a, $f(x) \approx L(x)$.

Use the interactive provided to see local linearity dynamically. Zoom in towards the center a and describe the relationship between the tangent line at a and the function values near a.

www.desmos.com/calculator/oie9mpmacg

3.1.2 Linear Approximation

A slight change in perspective and notation will enable us to be more precise in discussing how the tangent line approximates f near $x = a$. By solving for y, we can write the equation for the tangent line as

$$y = f'(a)(x - a) + f(a)$$

This line is itself a function of x. Replacing the variable y with the expression $L(x)$, we call

$$L(x) = f'(a)(x - a) + f(a)$$

the **linear approximation of f centered at the point** $(a, f(a))$. We may also use the terms **tangent line approximation of f** or **local linearization of f** to reference the same thing.

In this notation, $L(x)$ is nothing more than a new name for the tangent line. As we saw above, for x close to the center a, $f(x) \approx L(x)$.

Example 3.1.3 Suppose that a function $y = f(x)$ has its tangent line approximation given by $L(x) = 3 - 2(x - 1)$ centered at the point $(1, 3)$, but we do not know anything else about the function f. To estimate a value of $f(x)$ for x near 1, such as $f(1.2)$, we can use the fact that $f(1.2) \approx L(1.2)$ and hence

$$f(1.2) \approx L(1.2) = 3 - 2(1.2 - 1) = 3 - 2(0.2) = 2.6.$$

□

We emphasize that $y = L(x)$ is simply a new name for the tangent line function. Using this new notation and our observation that $L(x) \approx f(x)$ for x near a, it follows that we can write

$$f(x) \approx f(a) + f'(a)(x - a) \text{ for } x \text{ near } a.$$

Activity 3.1.2 Suppose it is known that for a given differentiable function $y = g(x)$, its local linearization at the point where $a = -1$ is given by $L(x) = -2 + 3(x + 1)$.

a. Compute the values of $L(-1)$ and $L'(-1)$.

b. What must be the values of $g(-1)$ and $g'(-1)$? Why?

c. Do you expect the value of $g(-1.03)$ to be greater than or less than the value of $g(-1)$? Why?

d. Use the local linearization to estimate the value of $g(-1.03)$.

e. Suppose that you also know that $g''(-1) = 2$. What does this tell you about the graph of $y = g(x)$ at $a = -1$?

f. For x near -1, sketch the graph of the local linearization $y = L(x)$ as well as a possible graph of $y = g(x)$ on the axes provided in Figure 3.1.4.

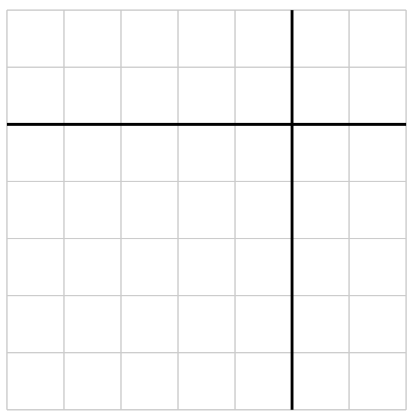


Figure 3.1.4 Axes for plotting $y = L(x)$ and $y = g(x)$.

From Activity 3.1.2, we see that the local linearization $y = L(x)$ is a linear function that shares two important values with the function $y = f(x)$ that it is derived from. In particular,

- because $L(x) = f(a) + f'(a)(x - a)$, it follows that $L(a) = f(a)$; and
- because L is a linear function, its derivative is its slope.

Hence, $L'(x) = f'(a)$ for every value of x, and specifically $L'(a) = f'(a)$. Therefore, we see that L is a linear function that has both the same value and the same slope as the function f at the point $(a, f(a))$.

Thus, if we know the linear approximation $y = L(x)$ for a function, we know the original function's value and its slope at the point of tangency. What remains unknown, however, is the shape of the function f at the point of tangency. There are essentially four possibilities, as shown in Figure 3.1.5.

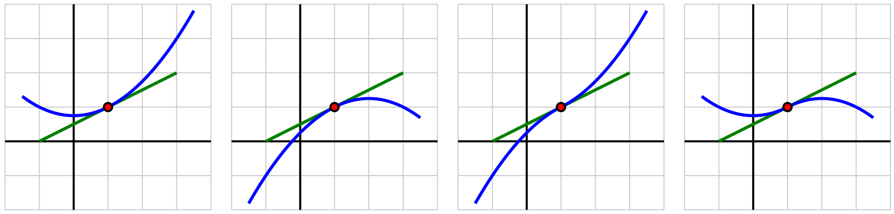


Figure 3.1.5 Four possible graphs for a nonlinear differentiable function and how it can be situated relative to its tangent line at a point.

These possible shapes result from the fact that there are three options for the value of the second derivative: either $f''(a) < 0$, $f''(a) = 0$, or $f''(a) > 0$.

- If $f''(a) > 0$, then we know the graph of f is concave up, and we see the first possibility on the left, where the tangent line lies entirely below the curve.
- If $f''(a) < 0$, then f is concave down and the tangent line lies above the curve, as shown in the second figure.
- If $f''(a) = 0$ and f'' changes sign at $x = a$, the concavity of the graph will change, and we will see either the third or fourth figure.[2].
- A fifth option (which is not very interesting) can occur if the function f itself is linear, so that $f(x) = L(x)$ for all values of x.

The plots in Figure 3.1.5 highlight yet another important thing that we can learn from the concavity of the graph near the point of tangency: whether the tangent line lies above or below the curve itself. This is key because it tells us whether or not the tangent line approximation's values will be too large (an overestimate) or too small (an underestimate) in comparison to the true value of f. For instance, in the first situation in the leftmost plot in Figure 3.1.5 where $f''(a) > 0$, because the tangent line falls below the curve, we know that $L(x) \le f(x)$ for all values of x near a.

Activity 3.1.3 This activity concerns a function $f(x)$ about which the following information is known:

- f is a differentiable function defined at every real number x
- $f(2) = -1$
- $y = f'(x)$ has its graph given in Figure 3.1.6

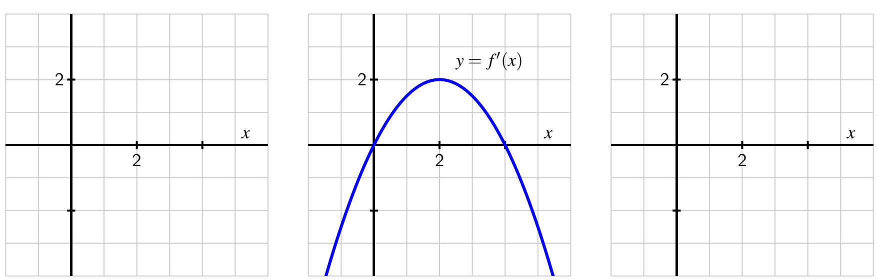


Figure 3.1.6 At center, a graph of $y = f'(x)$; at left, axes for plotting $y = f(x)$; at right, axes for plotting $y = f''(x)$.

[2]It is possible that $f''(a) = 0$ but f'' does not change sign at $x = a$, in which case the graph will look like one of the first two options.

Your task is to determine as much information as possible about f (especially near the value $a = 2$) by responding to the questions below.

a. Find a formula for the tangent line approximation, $L(x)$, to f at the point $(2, -1)$.

b. Use the tangent line approximation to estimate the value of $f(2.07)$. Show your work carefully and clearly.

c. Sketch a graph of $y = f''(x)$ on the righthand grid in Figure 3.1.6; label it appropriately.

d. Is the slope of the tangent line to $y = f(x)$ increasing, decreasing, or neither when $x = 2$? Explain.

e. Sketch a possible graph of $y = f(x)$ near $x = 2$ on the lefthand grid in Figure 3.1.6. Include a sketch of $y = L(x)$ (found in part (a)). Explain how you know the graph of $y = f(x)$ looks like you have drawn it.

f. Does your estimate in (b) over- or under-estimate the true value of $f(2.07)$? Why?

The idea that a differentiable function looks linear and can be well-approximated by a linear function is an important one that finds wide application in calculus. Many applications that we will see in the remainder of this chapter rely on the concept of local linearity. One example is that it can help us to make further sense of certain challenging limits. For instance, we will see in Section 3.6 that the limit

$$\lim_{x \to 0} \frac{\sin(x)}{x}$$

is an *indeterminate form*, because both its numerator and denominator tend to 0 This can make it challenging to know the value of the limit algebraically. While there is no algebra that we can do to simplify $\frac{\sin(x)}{x}$, it is straightforward to show that the linearization of $f(x) = \sin(x)$ at the point $(0, 0)$ is given by $L(x) = x$. Hence, for values of x near 0, $\sin(x) \approx x$, and therefore

$$\frac{\sin(x)}{x} \approx \frac{x}{x} = 1,$$

which makes plausible the fact that

$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1.$$

Indeed, we will develop further tools in Section 3.6 to see this is in fact the value of the limit.

3.1.3 Quadratic Approximation

We have already seen that the second derivative can be useful information in determining the shape (concavity) of a function near a point when using the tangent line approximation to estimate a function's value near that point. We have also seen that the tangent line approximation $L(x)$ is a "good" approximation of a function $f(x)$ near its center because it has the same y value at the center ($f(a) = L(a)$) and it has the same slope at the center ($f'(a) = L'(a)$).

Instead of restricting ourselves to approximating with a line, we can similarly ask for the best *quadratic* function to use to approximate a function near a point. A benefit of using the graph of a parabola in approximation is that we can now ask not only that the approximation match the function's y value and first derivative, but also match the function's second derivative (concavity). The following activity will help us determine a general formula for the best quadratic function to use when approximating a function near a point.

Activity 3.1.4 Quadratic Approximation. Let $f(x)$ be a twice differentiable function and 1 the center we would like to use for an approximation. Suppose $Q(x) = d + c(x-1) + b(x-1)^2$ is the best quadratic approximation, meaning

- $Q(1) = f(1)$,
- $Q'(1) = f'(1)$, and
- $Q''(1) = f''(1)$.

1. Determine the value $Q(1)$. What does this tell you about the value of the parameter d?
2. Compute $Q'(x)$ and determine the value $Q'(1)$. What does this tell you about the value of the parameter c?
3. Compute $Q''(x)$ and determine the value $Q''(1)$. What does this tell you about the value of the parameter b?
4. Write down the general formula for $Q(x)$ in terms of the function f and its derivatives. How does it compare to the linear approximation $L(x)$ of f centered at 1?

We summarize our findings from Activity 3.1.4 below:

Quadratic Approximation.

The **quadratic approximation** of a twice differentiable function f centered at a is

$$Q(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2.$$

Example 3.1.7 Linear and Quadratic Approximation. Let $f(x) = \cos(x)$. We will determine the linear and quadratic approximations of $f(x)$ with center $x = 0$.

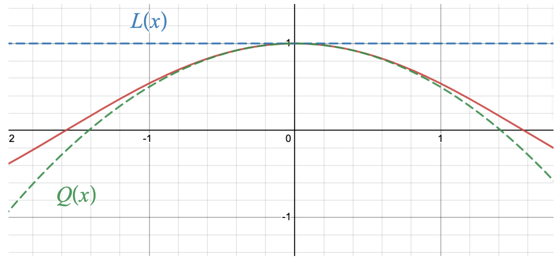
The linear approximation is given by

$$\begin{aligned} L(x) &= f(0) + f'(0)(x-0) \\ &= 1 - \sin(0)x \\ &= 1. \end{aligned}$$

The quadratic approximation is given by

$$\begin{aligned} Q(x) &= f(0) + f'(0)(x-0) + \frac{f''(0)}{2}(x-0)^2 \\ &= 1 - \sin(0)x - \frac{\cos(0)}{2}x^2 \\ &= 1 - \frac{1}{2}x^2. \end{aligned}$$

While both $L(x)$ and $Q(x)$ can be used to approximate $cos(x)$ for x values close to 0, the graph below shows that in general, $Q(x)$ will result in better approximations because it is able to “bend” with the graph of $\cos(x)$.



$\square$

3.1.4 Summary

- **Question 3.1.8** What is the formula for the general tangent line approximation to a differentiable function $y = f(x)$ at the point $(a, f(a))$? $\square$

  The tangent line to a differentiable function $y = f(x)$ at the point $(a, f(a))$ is given in point-slope form by the equation

  $$y - f(a) = f'(a)(x - a).$$

- **Question 3.1.9** What is the principle of local linearity and what is the linear approximation (or local linearization) of a differentiable function f at a point $(a, f(a))$? $\square$

  The principle of local linearity tells us that if we zoom in on a point where a function $y = f(x)$ is differentiable, the function will be indistinguishable from its tangent line. That is, a differentiable function looks linear when viewed up close. We rename the tangent line to be the function $y = L(x)$, where $L(x) = f(a) + f'(a)(x - a)$. Thus, $f(x) \approx L(x)$ for all x near $x = a$.

- **Question 3.1.10** How does knowing just the tangent line approximation tell us information about the behavior of the original function itself near the point of approximation? $\square$

  If we know the tangent line approximation $L(x) = f(a) + f'(a)(x - a)$ to a function $y = f(x)$, then because $L(a) = f(a)$ and $L'(a) = f'(a)$, we also know the values of both the function and its derivative at the point where $x = a$. In other words, the linear approximation tells us the height and slope of the original function.

- **Question 3.1.11** How does knowing the second derivative's value at this point provide us additional knowledge of the original function's behavior? $\square$

  If, in addition, we know the value of $f''(a)$, we then know whether the tangent line lies above or below the graph of $y = f(x)$, depending on the concavity of f. We also can compute the quadratic approximation of f centered at a,

  $$Q(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2,$$

  which may provide us with even better approximations near the center a.

3.1.5 Exercises

1. A certain function $y = p(x)$ has its local linearization at $a = 3$ given by $L(x) = -2x + 5$.

a. What are the values of $p(3)$ and $p'(3)$? Why?

b. Estimate the value of $p(2.79)$.

c. Suppose that $p''(3) = 0$ and you know that $p''(x) < 0$ for $x < 3$. Is your estimate in (b) an overestimate or underestimate of the actual value?

d. Suppose that $p''(x) > 0$ for $x > 3$. Use this fact and the additional information above to sketch an accurate graph of $y = p(x)$ near $x = 3$. Include a sketch of $y = L(x)$ in your work.

e. Does the quadratic approximation of p centered at 3 provide better, worse, or equal approximations as compared to $L(x)$ in this case? How do you know?

2. An object moving along a straight line path has a differentiable position function $y = s(t)$; $s(t)$ measures the object's position relative to the origin at time t. It is known that at time $t = 9$ seconds, the object's position is $s(9) = 4$ feet (i.e., 4 feet to the right of the origin). Furthermore, the object's instantaneous velocity at $t = 9$ is -1.2 feet per second, and its acceleration at the same instant is 0.08 feet per second per second.

a. Use local linearity to estimate the position of the object at $t = 9.34$.

b. Is your estimate from (a) likely an overestimate or an underestimate of the actual value? Why?

c. Use quadratic approximation centered at 9 to estimate the position of the object at $t = 9.34$.

d. In everyday language, describe the behavior of the moving object at $t = 9$. Is it moving toward the origin or away from it? Is its velocity increasing or decreasing?

3. For a certain function f, its derivative is known to be $f'(x) = (x - 1)e^{-x^2}$. Note that you do not know a formula for $y = f(x)$.

a. At what x-value(s) is $f'(x) = 0$? Justify your answer algebraically, but include a graph of f' to support your conclusion.

b. Reasoning graphically, for what intervals of x-values is $f''(x) > 0$? What does this tell you about the behavior of the original function f? Explain.

c. Assuming that $f(2) = -3$, estimate the value of $f(1.88)$ by finding and using the tangent line approximation to f at $x = 2$. Is your estimate larger or smaller than the true value of $f(1.88)$? Justify your answer.

d. Determine the quadratic approximation of f centered at 2.

3.2 The Stability Theorem

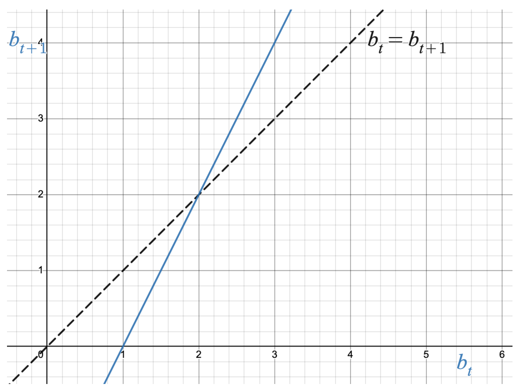
Motivating Questions

- How can we use the first derivative to help classify equilibrium values as stable or unstable?
- Can the first derivative always tell us whether an equilibrium value is stable or unstable?

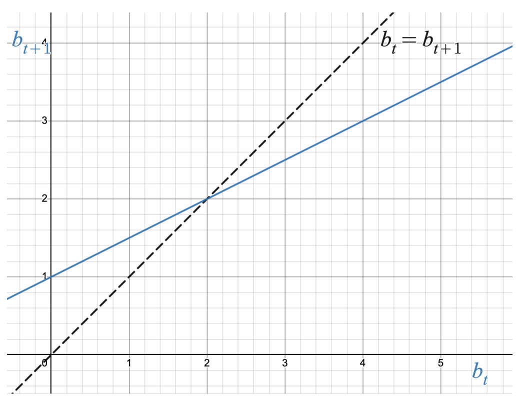
In this section we will revisit an old question from Section 1.8: given a DTDS with equilibrium value x^*, is x^* *stable* or *unstable*? Previously we used Cob-webbing to answer this question, which required graphing the updating function rule with the diagonal $y = x$. However, in this section we will see that the *derivative* is a useful tool that we can use to determine stability with or without the graph of the updating function rule.

Warm-Up 3.2.1 The graph of the updating function rule f is given below for four different DTDS's. Each DTDS has an equilibrium value at $b^* = 2$ (why?). Use cob-webbing to classify the equilibrium value as stable or unstable. What can you say about the derivative of the updating function at the equilibrium value, $f'(2)$?

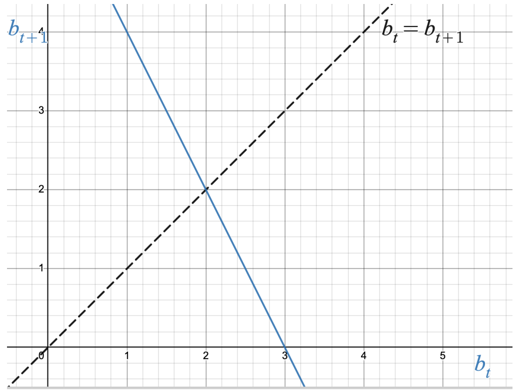
1.



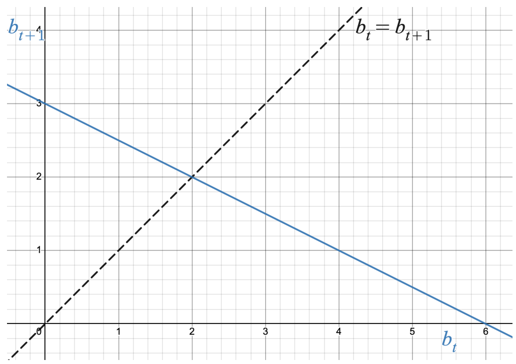
2.



3.



4.



3.2.1 Testing stability of equilibria using the derivative

Section 3.1 introduced the concept of *local linearity*: a function whose derivative exists at an input value a resembles the tangent line at a for x values close to a. The stability of an equilibrium value is also a *local* property, since it is dependent on the cob-webbing behavior when using initial values that are close to the equilibrium value. Therefore, even though most updating function rules are not linear functions, when asking about the stability of an equilibrium value we can approximate the updating function rule using its tangent line at the associated equilibrium point. We expect the cob-webbing behavior using the tangent line to be qualitatively the same as if we used the updating function when using initial values that are close to the equilibrium value.

Warm-Up 3.2.1 illustrates some examples of how we expect cob-webbing to behave with linear updating function rules, which is dependent on the slope of the linear updating function rule. We see that the stability of an equilibrium value depends on the *steepness* of the slope, and not on its *direction*. Indeed, playing with a few more examples may convince you that the slope of the diagonal line ($m = 1$) is the right value to compare steepness to in order to determine the stability of an equilibrium value: if the slope of a linear updating function rule is less steep than 1, the equilibrium value will be stable; if the slope of a linear updating function rule is steeper than 1, the equilibrium value will be unstable. This observation, along with local linearity, gives us the following stability criteria based on the first derivative:

Let f be the updating function rule for a DTDS with equilibrium value x^*.

- If $|f'(x^*)| < 1$, then x^* is a stable equilibrium value.
- If $|f'(x^*)| > 1$, then x^* is an unstable equilibrium value.

Remark 3.2.1

1. We do not compare the actual value of the derivative at the equilibrium value to 1, but the *absolute value* of the derivative at the equilibrium value. This reflects our observation that it is only the steepness of the line that impacts the stability, and not the direction of the line.

2. The stability criteria given so far does not address the case when $|f'(x^*)| = 1$. We will address this case in Subsection 3.2.2.

While the direction of the tangent line (the sign of the slope) does not impact stability, it does tell us about other qualitative behavior of the solution function. You may notice in the examples of Warm-Up 3.2.1 that the cob-webbing associated with the negative-sloped updating functions looks different than that of the positive-sloped updating functions. We explore this difference in more detail in the following example.

Example 3.2.2 Derivative Signs and Solution Behavior. In Warm-Up 3.2.1 number 1, the updating function is $b_{t+1} = 2b_t - 2$. Using an initial value of $b_0 = 2.5$ and iterating the updating function produces the following table of values:

Table 3.2.3

t	b_t	b_{t+1}
0	2.5	3
1	3	4
2	4	6
3	6	10

The table shows that the solution function values are trending away from $b^* = 2$, which is unstable behavior. This is because the updating function rule is $f(x) = 2x - 2$, which means $f'(x) = 2$. Using the stability criteria, we see that $|f'(2)| = |2| = 2$, which is greater than 1, and so the equilibrium value $b^* = 2$ is unstable.

Further, we see the solution function is moving in one direction (increasing in this case) with each iteration. This is because $f'(2) > 0$.

In Warm-Up 3.2.1 number 4, the updating function is $b_{t+1} = -0.5b_t + 3$. Using an initial value of $b_0 = 2.5$ and iterating the updating function produces the following table of values:

Table 3.2.4

t	b_t	b_{t+1}
0	2.5	1.75
1	1.75	2.125
2	2.125	1.9375
3	1.9375	2.03125

The table shows that the solution function values are trending towards $b^* = 2$, which is stable behavior. This is because the updating function rule is $f(x) = -0.5x + 3$, which means $f'(x) = -0.5$. Using the stability criteria, we see that $|f'(2)| = |-0.5| = 0.5$, which is less than 1, and so the equilibrium value $b^* = 2$ is stable.

Further, we see the solution function is going back and forth between above and below the equilibrium value of 2 with each iteration. This is because $f'(2) < 0$. □

The qualitative behavior of a solution going back and forth between above and below the equilibrium value is called **oscillation**, and we've seen that the first derivative of an updating function rule can also help us determine when solution functions will oscillate around an equilibrium value.

Oscillation Criteria.

Let f be the updating function rule for a DTDS with equilibrium value x^*. If $f'(x^*) < 0$, then the solution function will oscillate around x^* when the initial value is close to x^*.

It's important to note that the behaviors of "stability" and "oscillation" are not mutually exclusive. That is, it is possible for an equilibrium value to be stable with oscillatory behavior, unstable with oscillatory behavior, stable with non-oscillatory behavior, and unstable with non-oscillatory behavior.

Activity 3.2.2

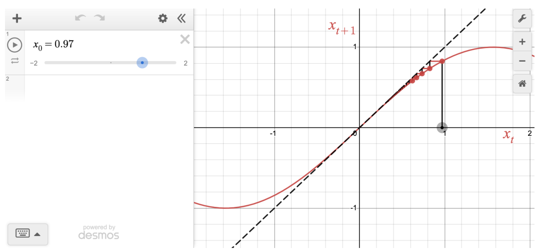
1. Consider a DTDS with updating function $b_{t+1} = b_t(b_t - 4)$.
   (a) Verify that $b^* = 0$ and $b^* = 5$ are equilibrium values of this system.
   (b) Use the stability criteria to classify each equilibrium value as stable or unstable.
   (c) For which equilibrium value(s), if any, will a solution function display oscillatory behavior for initial values that are close to that equilibrium value?
   (d) Use technology to graph the updating function rule. Apply the stability criteria using your graph to classify each equilibrium value as stable or unstable. Do your answers match what you found in part (b)?
   (e) Use your graph and cob-webbing to confirm all of your previous answers.
2. Consider a DTDS with updating function $p_{t+1} = \dfrac{2p_t}{p_t + 1}$.
   (a) Verify that $p^* = 0$ and $p^* = 1$ are equilibrium values of this system.
   (b) Use the stability criteria to classify each equilibrium value as stable or unstable.
   (c) For which equilibrium value(s), if any, will a solution function display oscillatory behavior for initial values that are close to that equilibrium value?
   (d) Use technology to graph the updating function rule. Apply the stability criteria using your graph to classify each equilibrium value as stable or unstable. Do your answers match what you found in part (b)?
   (e) Use your graph and cob-webbing to confirm all of your previous answers.

3.2.2 An inconclusive case using the derivative

The stability criteria provided so far does not tell us what happens when the derivative of an updating function rule at an equilibrium value is equal to 1 in absolute value. For the sake of examples, we will focus on the case when the derivative is equal to 1 (a similar discussion can be had when the derivative is equal to -1).

This is an interesting case because if a DTDS has an equilibrium value x^*, then the graph of the updating function rule will intersect the diagonal line $x_{t+1} = x_t$ at the point (x^*, x^*). If the derivative of the updating function rule at x^* is equal to 1, its tangent line has the same slope as the diagonal line. This means the tangent line intersects the diagonal line and has the same slope, which means the diagonal line is the tangent line to the updating function rule at the equilibrium point. We have seen in Section 3.1 that the way a function behaves around its tangent line at a point can vary depending on the second derivative at that point. We will use this as motivation in generating the two examples below.

First consider the updating function $x_{t+1} = \sin(x_t)$. It can be verified algebraically that $x^* = 0$ is an equilibrium value. Note also that the updating function rule is $f(x) = \sin(x)$, so that $f'(x) = \cos(x)$. Therefore, $f'(0) = \cos(0) = 1$, which means this is a case in which the derivative of the updating function rule at the equilibrium value is equal to 1. Use the interactive below to answer the questions before viewing the answers.



www.desmos.com/calculator/l7lzh9l5g8

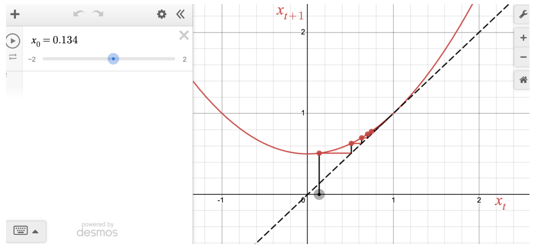
- For initial values close to and greater than the equilibrium value, how does the solution function behave? For initial values greater than 0, cob-webbing shows that the solution function decreases towards 0.

- For initial values close to and less than the equilibrium value, how does the solution function behave? For initial values less than 0, cob-webbing shows that the solution function increases towards 0.

- Would you classify the equilibrium value as stable or unstable? Since the solution function values get closer and closer to the equilibrium value for initial values on either side of the equilibrium value, we would classify the equilibrium value as stable.

Because of this example, we may be tempted to conclude that when the derivative of an updating function rule at an equilibrium value is equal to 1, the equilibrium value is stable. A single example, however, cannot prove to us that this will always occur. The next example will show us that we can, in fact, get different behavior if we change the second derivative value of the updating function at the equilibrium value.

Now we consider the updating function $x_{t+1} = 0.5(x_t)^2 = 0.5$. It can be verified algebraically that $x^* = 1$ is an equilibrium value. Note also that the updating function rule is $f(x) = 0.5x^2 + 0.5$, so that $f'(x) = x$. Therefore, $f'(1) = 1$, which means this is a case in which the derivative of the updating function rule at the equilibrium value is equal to 1. Use the interactive below to answer the questions before viewing the answers.



www.desmos.com/calculator/igtnpjb0du

- For initial values close to and greater than the equilibrium value, how does the solution function behave? For initial values greater than 1, cob-webbing shows that the solution function increases away from 1.
- For initial values close to and less than the equilibrium value, how does the solution function behave? For initial values less than 1, cob-webbing shows that the solution function increases towards 1.
- Would you classify the equilibrium value as stable or unstable? This is a case we have not seen yet. The behavior is different depending on which side of the equilibrium value the initial value is on. Since the solution function values do not approach the equilibrium value for initial values on both sides of the equilibrium value, we would classify this equilibrium value as having a special case of unstable behavior.

While we technically classify the last example as unstable behavior, we can be more descriptive by calling it **half-stable**, since it displays stable behavior on one side of the equilibrium value and unstable behavior on the other side.

We have found two examples that both have updating functions rules whose derivative is equal to 1 at the equilibrium value, but they display very different behaviors in terms of stability. This means that in this particular case, the first derivative of the updating function does not hold enough information to tell us whether the equilibrium value is stable or unstable. In this case, we must use another method (like cob-webbing, or generating an appropriate table of values) to determine stability. We end this section by stating the stability criteria given earlier along with the inconclusive case in what we will call the **Stability Theorem**:

Stability Theorem.

Let f be the updating function rule for a DTDS with equilibrium value x^*.

- If $|f'(x^*)| < 1$, then x^* is a stable equilibrium value.
- If $|f'(x^*)| > 1$, then x^* is an unstable equilibrium value.
- If $|f'(x^*)| = 1$, then the stability theorem is inconclusive. Another method must be used to determine the stability of x^*.

3.2.3 Summary

- **Question 3.2.5** How can we use the first derivative to help classify equilibrium values as stable or unstable? ☐

  Due to local linearity of differentiable functions, the cob-webbing behavior for initial values near an equilibrium value can be analyzed by understanding the cob-webbing behavior of linear updating functions. This results in the following stability criteria for an updating function rule $f(x)$ and an equilibrium value x^*:

- If $|f'(x^*)| < 1$, then x^* is a stable equilibrium value.
- If $|f'(x^*)| > 1$, then x^* is an unstable equilibrium value.

Though the sign of the derivative does not impact stability, it can tell us whether a solution function will oscillate around its equilibrium value. Oscillation around x^* occurs when $f'(x^*) < 0$.

- **Question 3.2.6** Can the first derivative always tell us whether an equilibrium value is stable or unstable? ☐

  No. If $|f'(x^*)| = 1$, then the stability theorem is inconclusive. Another method must be used to determine the stability of x^*.

3.2.4 Exercises

1. Draw a graph of an updating function with an equilibrium value at $x^* = 1$ such that

(a) $x^* = 1$ is stable and non-oscillatory

(b) $x^* = 1$ is unstable and non-oscillatory

(c) $x^* = 1$ is stable and oscillatory

(d) $x^* = 1$ is unstable and oscillatory

2. Let $f(x)$ be an updating function rule for a DTDS that has an equilibrium value at $x^* = 2$ such that $f'(2) = 1$.

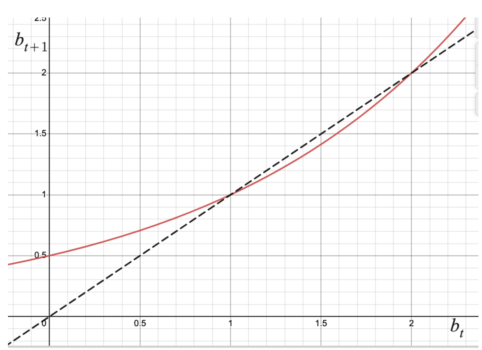
(a) Draw a graph of an updating function that satisfies the above criteria such that $x^* = 2$ is an unstable (not half-stable) equilibrium value. Illustrate why your example works.

(b) Draw a graph of an updating function that satisfies the above criteria such that $x^* = 2$ is unstable for initial values less than 2 but stable for initial values greater than 2. Illustrate why your example works.

3. For each updating function below, identify all equilibrium values. Then classify each equilibrium value as stable or unstable. You should justify your answers using the Stability Theorem.

(a) $y_{t+1} = \dfrac{2y_t}{1 + 0.001y_t}$

(b)



3.3 The Logistic Discrete-Time Dynamical System

Motivating Questions

- What is the logistic dynamical system and what can it be used for?
- How can we use calculus tools we've developed so far to understand how the logistic dynamical system behaves?

In Section 1.9, we analyzed two discrete-time dynamical systems, the lung model and the competing species model, using tools from Chapter 1. In this section, we will again analyze an important discrete-time dynamical system, the logistic DTDS, using our newly developed tools involving the derivative from Section 3.2.

In Subsection 1.9.2, we modeled population growth with the updating function $p_{t+1} = rp_t$, where r represents the per capita reproduction rate. We know from Section 1.7 that a solution function associated with this updating function is exponential. While exponential growth can be a good model for population growth at first, over time it becomes unrealistic since exponential growth does not reflect limited resources or a carrying capacity that we observe with real populations over time. One use case for the logistic dynamical system is as a population model that takes into account the carrying capacity of a population.

Warm-Up 3.3.1 Consider the updating function $x_{t+1} = rx_t(1 - x_t)$.

1. Write the updating function rule. Identify which letters are parameters and which letters are variables.
2. Compute the derivative of the updating function rule.

   Note: you can use the product rule, or do some algebraic simplification first and use the power and sum rules.
3. $x^* = 0$ is an equilibrium value (why?). Compute $f'(0)$..
4. According to the Stability Theorem, for what values of r is $x^* = 0$ stable?

3.3.1 The logistic map

Let p_t be the population size in year t, and M the maximum possible population size. We would like a model that reflects our observation that population growth slows as it gets nearer to M. We can achieve this by letting the per capita reproduction rate change as the population changes, as opposed to remaining constant as in the exponential model. To that end, if $r > 0$ is the maximum possible per capita reproduction rate, we set the per capita reproduction rate at year t to be

$$r_t = r\left(1 - \frac{p_t}{M}\right).$$

Note that this has the desired effect of decreasing the per capita reproduction rate when the population is large (close to M). The updating function for the population then becomes

$$\begin{aligned} p_{t+1} &= \text{per capita reproduction at time } t \cdot \text{population at time } t \\ &= r_t \cdot p_t \end{aligned}$$

$$= r\left(1 - \frac{p_t}{M}\right) \cdot p_t.$$

Thus, $p_{t+1} = r\left(1 - \frac{p_t}{M}\right) \cdot p_t$ is an updating function which models population size at year t, taking into account a maximum population size. Since we typically will not care about the actual value of M, but just the size of the population relative to M, we define the variable $x_t = \frac{p_t}{M}$, which gives the decimal percentage of the maximum population at year t. We can then re-write our updating function as

$$\begin{aligned} p_{t+1} &= r\left(1 - \frac{p_t}{M}\right) \cdot p_t \\ \frac{1}{M} \cdot p_{t+1} &= r\left(1 - \frac{p_t}{M}\right) \cdot p_t \cdot \frac{1}{M} \\ x_{t+1} &= r(1 - x_t) \cdot x_t \end{aligned}$$

The logistic map.

The **logistic map** is the recursive function $x_{t+1} = rx_t(1 - x_t)$.

In the context of population growth, r is a positive constant representing the maximum possible per capita reproduction rate, and x_t is the percentage of the maximum population in year t.

3.3.2 Equilibrium values and stability of the logistic map

We now have a new model for population growth in the logistic map, but we don't know many details of how this system behaves outside of the general decrease in per capita reproduction with increase in population size that we built into the model. What are the equilibrium values of this system? How does the maximum per capita reproduction rate r impact their stability? What do the equilibrium values and their stability represent in the context of the model?

We know from Section 1.8 that we can find equilibrium values graphically or algebraically. Since our model contains a parameter and thus has no single graph to analyze, we will compute equilibrium values algebraically. Letting $x_{t+1} = x_t = x^*$ in the updating function $x_{t+1} = rx_t(1 - x_t)$ gives $x^* = rx^*(1 - x^*)$, or equivalently $x^* = rx^* - r(x^*)^2$. We now solve for x^*:

$$\begin{aligned} x^* &= rx^* - r(x^*)^2 \\ r(x^*)^2 + x^* - rx^* &= 0 \\ r(x^*)^2 + (1 - r)x^* &= 0 \\ x^*[rx^* + (1 - r)] &= 0 \end{aligned}$$

Thus we have two factors, x^* and $rx^* + (1 - r)$ that multiply to 0, meaning the zeros of both factors can give us an equilibrium value. The first equilibrium value is $x^* = 0$, and the second equilibrium value is the solution to the equation $rx^* + (1 - r) = 0$, which we solve as

$$\begin{aligned} rx^* + (1 - r) &= 0 \\ rx^* &= r - 1 \\ x^* &= \frac{r - 1}{r}. \end{aligned}$$

Division by r is valid since $r > 0$, and the second equilibrium value is only valid for $r > 1$ since x^* represents a percentage and cannot be negative.

We are now interested in classifying the stability of the equilibrium values $x^* = 0$ and $x^* = \dfrac{r-1}{r}$. Cob-webbing is not an ideal method, since our model contains a parameter and thus these equilibrium values represent many different systems (and graphs). Therefore, we will use the first derivative of the updating function rule to analyze stability using the Stability Theorem.

The updating function rule of $x_{t+1} = rx_t(1 - x_t)$ is $f(x) = rx(1 - x)$, or equivalently $f(x) = rx - rx^2$. The first derivative is then $f'(x) = r - 2rx$. For the equilibrium value $x^* = 0$, we compute

$$|f'(0)| = |r - 2r(0)| = |r|.$$

By the Stability Theorem, $x^* = 0$ is stable when $|r| < 1$, which is equivalent to $-1 < r < 1$. Since $r > 0$ in our model, $x^* = 0$ is stable when $0 < r < 1$.

For the equilibrium value $x^* = \dfrac{r-1}{r}$, we compute

$$\left| f'\left(\frac{r-1}{r}\right)\right| = \left| r - 2r\left(\frac{r-1}{r}\right)\right| = |r - 2(r-1)| = |2 - r|.$$

By the Stability Theorem, $x^* = \dfrac{r-1}{r}$ is stable when $|2 - r| < 1$. To determine the exact bounds on r, we compute

$$\begin{aligned} |2 - r| &< 1, \text{ or} \\ -1 < 2 - r &< 1, \text{ or} \\ -3 < -r &< -1, \text{ or} \\ 3 > r &> 1. \end{aligned}$$

Thus, $x^* = \dfrac{r-1}{r}$ is stable when $1 < r < 3$.

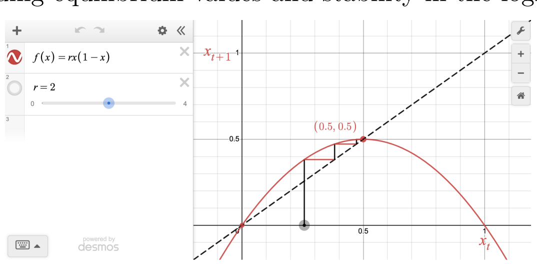
We can summarize our analysis of the equilibrium values in the following table:

Table 3.3.1 Stability of Equilibrium Values for the logistic map

	$x^* = 0$	$x^* = \dfrac{r-1}{r}$
Stable	$0 < r < 1$	$1 < r < 3$
Unstable	$r > 1$	$0 < r < 1$ and $r > 3$

Note that our analysis does not say what happens when $r = 1$ or $r = 3$, since these are cases in which the Stability Theorem is inconclusive. Cob-webbing or numerical methods may be used to determine the stability of each equilibrium value for these values of r.

Take a moment with the interactive below to visually verify our results regarding equilibrium values and stability in the logistic dynamical system.



www.desmos.com/calculator/an1jr9y0yt

Activity 3.3.2 The following questions are in reference to the equilibrium values of the logistic map discussed above.

1. Will the solution function ever oscillate around either equilibrium value? If so, for which value(s) of r does this occur?

2. In the context of modeling population growth, write a carefully worded sentence describing what it means for $x^* = 0$ to be a stable equilibrium value.

3. In the context of modeling population growth, write a carefully worded sentence describing what it means for $x^* = 0.5$ to be a stable equilibrium value.

There are many similar models to the logistic dynamical system with different equilibrium values and stability behaviors. In the final activity, you will perform a similar analysis as we have in this section on a DTDS whose updating function is a modified logistic map.

Activity 3.3.3 Consider a modified logistic dynamical system whose updating function is $y_{t+1} = ry_t(1 - (y_t)^2)$, where $r > 0$.

1. Determine all of the positive equilibrium values of this system.

2. For each positive equilibrium value, determine the values of r which make that value stable.

3. Modify the equation in the logistic map interactive to verify and visualize your results.

3.3.3 Summary

- **Question 3.3.2** What is the logistic dynamical system and what can it be used for? ☐

  The logistic dynamical system is the DTDS with updating function $x_{t+1} = rx_t(1 - x_t)$. It can be used as a model for population growth which takes into account a maximum possible population, among other things.

  The logistic map has many other applications outside of population modeling, and its behavior contains some interesting and deep mathematics, particularly in the types of unstable behavior that can occur around equilibrium values. Though it is outside the scope of this course, you can see a brief introduction to the other aspects of the logistic map in this video from Veritasium[1].

- **Question 3.3.3** How can we use calculus tools we've developed so far to understand how the logistic dynamical system behaves? ☐

  Since the logistic DTDS contains a parameter r, we cannot look at a single graph to analyze its behavior. Instead, we can compute equilibrium values algebraically and compute the derivative of the updating function, which gives us answers in terms of the parameter r. We can then analyze these expressions to determine how the parameter r effects the system's behavior. For example, we can use the Stability Theorem to determine which values of r make a given equilibrium value stable.

[1] veritasium.com

3.3.4 Exercises

1. Consider the logistic map with $r = 2$: $x_{t+1} = 2x_t(1 - x_t)$. We will explore what the solution function looks like when $x_0 = 0.01$.

(a) Based on our analysis in this section, what are the equilibrium values of this system? Classify each as stable or unstable.

(b) Iterate the updating function to determine the first 10 points of the solution function (up to x_9). Use Desmos[2] to plot the points on a graph.

(c) Looking at your graph, how is the solution function similar to an exponential solution function? How is it different?

(d) How are the stable equilibrium value(s) of this system represented in the graph of the solution function?

3.4 Identifying Extreme Values of Functions

Motivating Questions

- What are the critical numbers of a function f and how are they connected to identifying the most extreme values the function achieves?
- How does the first derivative of a function reveal important information about the behavior of the function, including the function's extreme values?
- How can the second derivative of a function be used to help identify extreme values of the function?

In many different settings, we are interested in knowing where a function achieves its least and greatest values. These can be important in applications — say to identify a point at which maximum profit or minimum cost occurs — or in theory to characterize the behavior of a function or a family of related functions.

Consider the simple and familiar example of a parabolic function such as $s(t) = -16t^2 + 32t + 48$ (shown at left in Figure 3.4.2) that represents the height of an object tossed vertically: its maximum value occurs at the vertex of the parabola and represents the greatest height the object reaches. This maximum value is an especially important point on the graph, the point at which the curve changes from increasing to decreasing.

Definition 3.4.1 Given a function f, we say that $f(c)$ is a **global** or **absolute maximum** of f provided that $f(c) \geq f(x)$ for all x in the domain of f, and similarly we call $f(c)$ a **global** or **absolute minimum** of f whenever $f(c) \leq f(x)$ for all x in the domain of f. ◊

For instance, in Figure 3.4.2, g has a global maximum of $g(c)$, but g does not appear to have a global minimum, as the graph of g seems to decrease without bound. Note that the point $(c, g(c))$ marks a fundamental change in the behavior of g, where g changes from increasing to decreasing; similar things happen at both $(a, g(a))$ and $(b, g(b))$, although these points are not global minima or maxima.

[2] `Desmos.com`

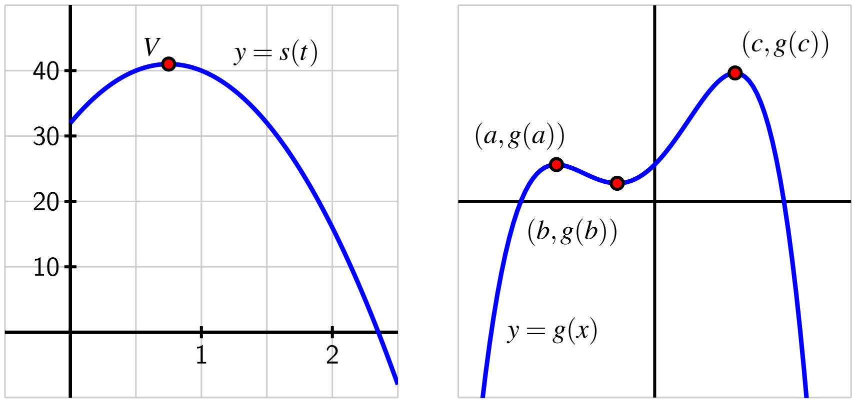


Figure 3.4.2 At left, $s(t) = -16t^2 + 24t + 32$ whose vertex is $(\frac{3}{4}, 41)$; at right, a function g that demonstrates several high and low points.

Definition 3.4.3 We say that $f(c)$ is a **local maximum** or **relative maximum** of f provided that $f(c) \geq f(x)$ for all x near c, and $f(c)$ is called a **local** or **relative minimum** of f whenever $f(c) \leq f(x)$ for all x near c. ◊

For example, in Figure 3.4.2, g has a relative minimum of $g(b)$ at the point $(b, g(b))$ and a relative maximum of $g(a)$ at $(a, g(a))$. We have already identified the global maximum of g as $g(c)$; it can also be considered a relative maximum. Any maximum or minimum may also be called an *extreme value* of f.

We would like to use calculus ideas to identify and classify key function behavior, including the location of relative extremes. Of course, if we are given a graph of a function, it is often straightforward to locate these important behaviors visually.

Warm-Up 3.4.1 Consider the function h given by the graph in Figure 3.4.4. Use the graph to answer each of the following questions.

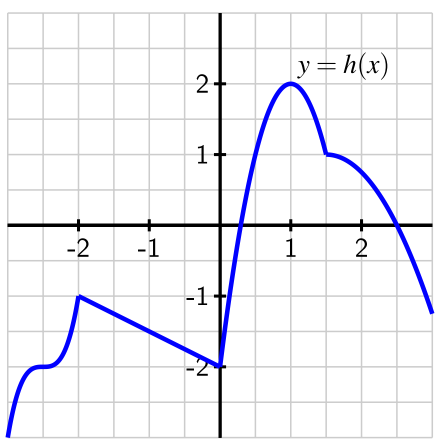


Figure 3.4.4 The graph of a function h on the interval $[-3, 3]$.

a. Identify all of the values of c such that $-3 < c < 3$ for which $h(c)$ is a local maximum of h.

b. Identify all of the values of c such that $-3 < c < 3$ for which $h(c)$ is a local minimum of h.

c. Does h have a global maximum on the interval $[-3, 3]$? If so, what is the value of this global maximum?

d. Does h have a global minimum on the interval $[-3, 3]$? If so, what is its value?

e. Identify all values of c for which $h'(c) = 0$.

f. Identify all values of c for which $h'(c)$ does not exist.

g. True or false: every relative maximum and minimum of h occurs at a point where $h'(c)$ is either zero or does not exist.

h. True or false: at every point where $h'(c)$ is zero or does not exist, h has a relative maximum or minimum.

3.4.1 Critical numbers and the first derivative test

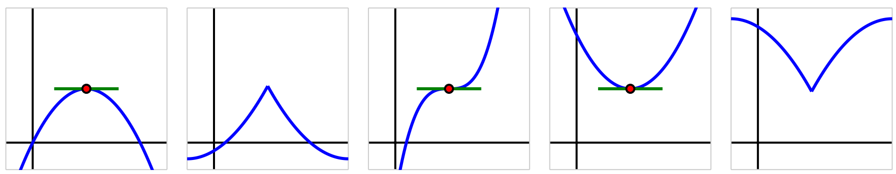


Figure 3.4.5 From left to right, a function with a relative maximum where its derivative is zero; a function with a relative maximum where its derivative is undefined; a function with neither a maximum nor a minimum at a point where its derivative is zero; a function with a relative minimum where its derivative is zero; and a function with a relative minimum where its derivative is undefined.

If a continuous function has a relative maximum at c, then it is both necessary and sufficient that the function change from being increasing just before c to decreasing just after c. A continuous function has a relative minimum at c if and only if the function changes from decreasing to increasing at c. (See Figure 3.4.5.) There are only two possible ways for these changes in behavior to occur: either $f'(c) = 0$ or $f'(c)$ is undefined. Because these values of c are so important, we call them *critical numbers*.

Definition 3.4.6 We say that a function f has a **critical number** at $x = c$ provided that c is in the domain of f, and $f'(c) = 0$ or $f'(c)$ is undefined. ◊

Critical numbers are the only possible locations where the function f may have relative extremes. Note that not every critical number produces a maximum or minimum; in the middle graph of Figure 3.4.5, the function pictured there has a horizontal tangent line at the noted point, but the function is increasing before and increasing after, so the critical number does not yield a maximum or minimum.

When c is a critical number, we say that $(c, f(c))$ is a **critical point** of the function, or that $f(c)$ is a **critical value** . The *first derivative test* summarizes how sign changes in the first derivative (which can only occur at critical numbers) indicate the presence of a local maximum or minimum for a given function.

First Derivative Test.

If p is a critical number of a continuous function f that is differentiable near p (except possibly at $x = p$), then f has a relative maximum at p if and only[1] if f' changes sign from positive to negative at p, and f has a relative minimum at p if and only if f' changes sign from negative to positive at p.

Example 3.4.7 Let f be a function whose derivative is given by the formula $f'(x) = e^{-2x}(3-x)(x+1)^2$. Determine all critical numbers of f and decide whether a relative maximum, relative minimum, or neither occurs at each.

Solution. Since we already have $f'(x)$ written in factored form, it is straightforward to find the critical numbers of f. Because $f'(x)$ is defined for all values of x, we need only determine where $f'(x) = 0$. From the equation

$$e^{-2x}(3-x)(x+1)^2 = 0$$

and the zero product property, it follows that $x = 3$ and $x = -1$ are critical numbers of f. (There is no value of x that makes $e^{-2x} = 0$.)

Next, to apply the first derivative test, we'd like to know the sign of $f'(x)$ at inputs near the critical numbers. Because the critical numbers are the only locations at which f' can change sign, it follows that the sign of the derivative is the same on each of the intervals created by the critical numbers: for instance, the sign of f' must be the same for every $x < -1$. We create a first derivative sign chart to summarize the sign of f' on the relevant intervals, along with the corresponding behavior of f.

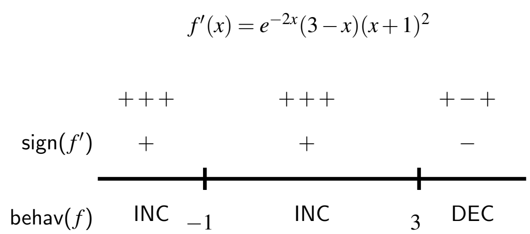


Figure 3.4.8 The first derivative sign chart for a function f whose derivative is given by the formula $f'(x) = e^{-2x}(3-x)(x+1)^2$.

To produce the first derivative sign chart in Figure 3.4.8 we identify the sign of each factor of $f'(x)$ at one selected point in each interval. For instance, for $x < -1$, we could determine the sign of e^{-2x}, $(3-x)$, and $(x+1)^2$ at the value $x = -2$. We note that both e^{-2x} and $(x+1)^2$ are positive regardless of the value of x, while $(3-x)$ is also positive at $x = -2$. Hence, each of the three terms in f' is positive, which we indicate by writing "$+ + +$." Taking the product of three positive terms results in a positive value for f', which we denote by the "+" in the interval to the left of $x = -1$. And, since f' is positive on that interval, we know that f is increasing, so we write "INC" to represent the behavior of f. In a similar way, we find that f' is positive and f is increasing on $-1 < x < 3$, and f' is negative and f is decreasing for $x > 3$.

Now we look for critical numbers at which f' changes sign. In this example, f' changes sign only at $x = 3$, from positive to negative, so f has a relative maximum at $x = 3$. Although f has a critical number at $x = -1$, since f is increasing both before and after $x = -1$, f has neither a minimum nor a maximum at $x = -1$. □

[1] Technically, we also have to assume that f is not piecewise constant on any intervals. This is because every point on a horizontal line is a relative maximum (and relative minimum) despite the fact that the derivative doesn't change sign at any point along the horizontal line.

Activity 3.4.2 Suppose that $h(x) = x^2 - 2x + 7$.

Suppose that $g(x)$ is a continuous function whose first derivative is $g'(x) = (x + 4)(x - 1)^2$.

a. Determine all critical numbers of h and all the critical numbers of g.

b. By developing a carefully labeled first derivative sign chart for each function, decide whether h and g have a local maximum, local minimum, or neither at each critical number.

3.4.2 The second derivative test

Recall that the second derivative of a function tells us several important things about the behavior of the function itself. For instance, if f'' is positive on an interval, then we know that f' is increasing on that interval and, consequently, that f is concave up, so throughout that interval the tangent line to $y = f(x)$ lies below the curve at every point. At a point where $f'(p) = 0$, the sign of the second derivative determines whether f has a local minimum or local maximum at the critical number p.

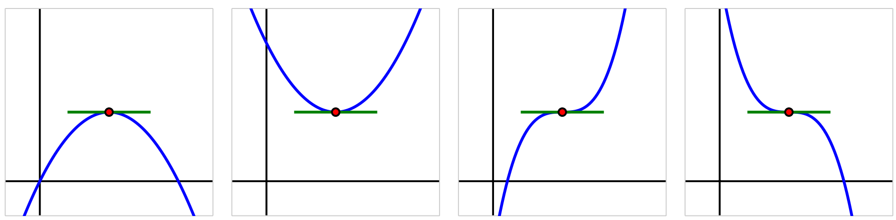


Figure 3.4.9 Four possible graphs of a function f with a horizontal tangent line at a critical point.

In Figure 3.4.9, we see the four possibilities for a function f that has a critical number p at which $f'(p) = 0$, provided $f''(p)$ is not zero on an interval including p (except possibly at p). On either side of the critical number, f'' can be either positive or negative, and hence f can be either concave up or concave down. In the first two graphs, f does not change concavity at p, and in those situations, f has either a local minimum or local maximum. In particular, if $f'(p) = 0$ and $f''(p) < 0$, then f is concave down at p with a horizontal tangent line, so f has a local maximum there. This fact, along with the corresponding statement for when $f''(p)$ is positive, is the substance of the *second derivative test*.

Second Derivative Test.

If p is a critical number of a continuous function f such that $f'(p) = 0$ and $f''(p) \neq 0$, then f has a relative maximum at p if and only if $f''(p) < 0$, and f has a relative minimum at p if and only if $f''(p) > 0$.

In the event that $f''(p) = 0$, the second derivative test is inconclusive. That is, the test doesn't provide us any information. This is because if $f''(p) = 0$, it is possible that f has a local minimum, local maximum, or neither.[2]

Just as a first derivative sign chart reveals all of the increasing and decreasing behavior of a function, we can construct a second derivative sign chart that demonstrates all of the important information involving concavity.

[2]Consider the functions $f(x) = x^4$, $g(x) = -x^4$, and $h(x) = x^3$ at the critical point $p = 0$.

Example 3.4.10 Let $f(x)$ be a function whose first derivative is $f'(x) = 3x^4 - 9x^2$. Construct both first and second derivative sign charts for f, fully discuss where f is increasing and decreasing and concave up and concave down, identify all relative extreme values, and sketch a possible graph of f.

Solution. Since we know $f'(x) = 3x^4 - 9x^2$, we can find the critical numbers of f by solving $3x^4 - 9x^2 = 0$. Factoring, we observe that

$$0 = 3x^2(x^2 - 3) = 3x^2(x + \sqrt{3})(x - \sqrt{3}),$$

so that $x = 0, \pm\sqrt{3}$ are the three critical numbers of f. The first derivative sign chart for f is given in Figure 3.4.11.

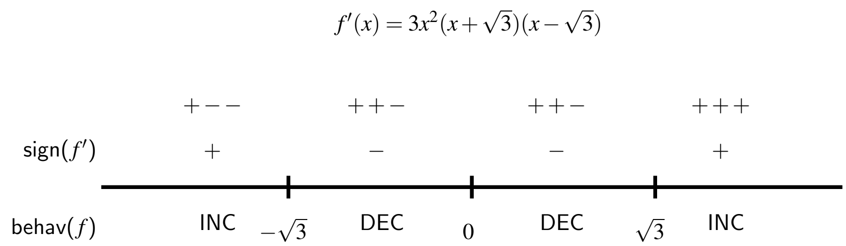


Figure 3.4.11 The first derivative sign chart for f when $f'(x) = 3x^4 - 9x^2 = 3x^2(x^2 - 3)$.

We see that f is increasing on the intervals $(-\infty, -\sqrt{3})$ and $(\sqrt{3}, \infty)$, and f is decreasing on $(-\sqrt{3}, 0)$ and $(0, \sqrt{3})$. By the first derivative test, this information tells us that f has a local maximum at $x = -\sqrt{3}$ and a local minimum at $x = \sqrt{3}$. Although f also has a critical number at $x = 0$, neither a maximum nor minimum occurs there since f' does not change sign at $x = 0$.

Next, we move on to investigate concavity. Differentiating $f'(x) = 3x^4 - 9x^2$, we see that $f''(x) = 12x^3 - 18x$. Since we are interested in knowing the intervals on which f'' is positive and negative, we first find where $f''(x) = 0$. Observe that

$$0 = 12x^3 - 18x = 12x\left(x^2 - \frac{3}{2}\right) = 12x\left(x + \sqrt{\frac{3}{2}}\right)\left(x - \sqrt{\frac{3}{2}}\right).$$

This equation has solutions $x = 0, \pm\sqrt{\frac{3}{2}}$. Building a sign chart for f'' in the exact same way we do for f', we see the result shown in Figure 3.4.12.

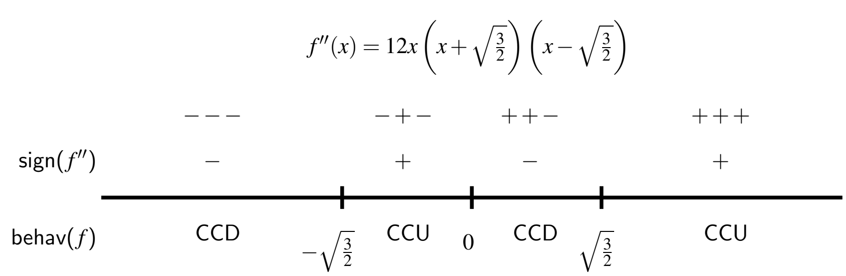


Figure 3.4.12 The second derivative sign chart for f when $f''(x) = 12x^3 - 18x = 12x^2\left(x^2 - \frac{3}{2}\right)$.

Therefore, f is concave down on the intervals $(-\infty, -\sqrt{\frac{3}{2}})$ and $(0, \sqrt{\frac{3}{2}})$, and concave up on $(-\sqrt{\frac{3}{2}}, 0)$ and $(\sqrt{\frac{3}{2}}, \infty)$.

Putting all of this information together, we now see a complete and accurate possible graph of f in Figure 3.4.13.

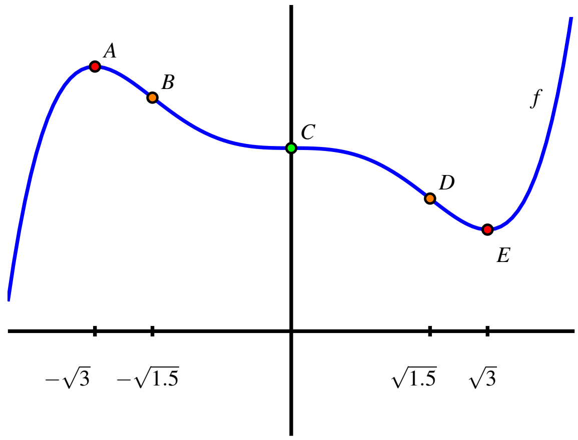


Figure 3.4.13 A possible graph of the function f in Example 3.4.10.

The point $A = (-\sqrt{3}, f(-\sqrt{3}))$ is a local maximum, because f is increasing prior to A and decreasing after; similarly, the point $E = (\sqrt{3}, f(\sqrt{3}))$ is a local minimum. Note, too, that f is concave down at A and concave up at E, which is consistent both with our second derivative sign chart and the second derivative test. At points B and D, concavity changes, as we saw in the results of the second derivative sign chart in Figure 3.4.12. Finally, at point C, f has a critical point with a horizontal tangent line, but neither a maximum nor a minimum occurs there, since f is decreasing both before and after C. It is also the case that concavity changes at C.

While we completely understand where f is increasing and decreasing, where f is concave up and concave down, and where f has relative extremes, we do not know any specific information about the y-coordinates of points on the curve. For instance, while we know that f has a local maximum at $x = -\sqrt{3}$, we don't know the value of that maximum because we do not know $f(-\sqrt{3})$. Any vertical translation of our sketch of f in Figure 3.4.13 would satisfy the given criteria for f. □

Points B, C, and D in Figure 3.4.13 are locations at which the concavity of f changes. We give a special name to any such point.

Definition 3.4.14 If p is a value in the domain of a continuous function f at which f changes concavity, then we say that $(p, f(p))$ is an **inflection point** (or **point of inflection**) of f. ♢

Just as we look for locations where f changes from increasing to decreasing at points where $f'(p) = 0$ or $f'(p)$ is undefined, so too we find where $f''(p) = 0$ or $f''(p)$ is undefined to see if there are points of inflection at these locations.

At this point in our study, it is important to remind ourselves of the big picture that derivatives help to paint: the sign of the first derivative f' tells us *whether* the function f is increasing or decreasing, while the sign of the second derivative f'' tells us *how* the function f is increasing or decreasing.

Activity 3.4.3 Suppose that g is a function whose second derivative, g'', is given by the graph in Figure 3.4.15.

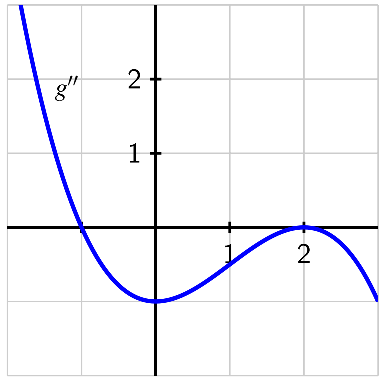


Figure 3.4.15 The graph of $y = g''(x)$.

a. Find the x-coordinates of all points of inflection of g.

b. Fully describe the concavity of g by making an appropriate sign chart.

c. Suppose you are given that $g'(-1.67857351) = 0$. Is there is a local maximum, local minimum, or neither (for the function g) at this critical number of g, or is it impossible to say? Why?

3.4.3 Summary

- **Question 3.4.16** What are the critical numbers of a function f and how are they connected to identifying the most extreme values the function achieves? □

  The critical numbers of a continuous function f are the values of p for which $f'(p) = 0$ or $f'(p)$ does not exist. These values are important because they identify horizontal tangent lines or corner points on the graph, which are the only possible locations at which a local maximum or local minimum can occur.

- **Question 3.4.17** How does the first derivative of a function reveal important information about the behavior of the function, including the function's extreme values? □

  Given a differentiable function f, whenever f' is positive, f is increasing; whenever f' is negative, f is decreasing. The first derivative test tells us that at any point where f changes from increasing to decreasing, f has a local maximum, while conversely at any point where f changes from decreasing to increasing f has a local minimum.

- **Question 3.4.18** How can the second derivative of a function be used to help identify extreme values of the function? □

  Given a twice differentiable function f, if we have a horizontal tangent line at $x = p$ and $f''(p)$ is nonzero, the sign of f'' tells us the concavity of f and hence whether f has a maximum or minimum at $x = p$. In particular, if $f'(p) = 0$ and $f''(p) < 0$, then f is concave down at p and f has a local maximum there, while if $f'(p) = 0$ and $f''(p) > 0$, then f has a local minimum at p. If $f'(p) = 0$ and $f''(p) = 0$, then the second derivative does not tell us whether f has a local extreme at p or not.

3.4.4 Exercises

1. This problem concerns a function about which the following information is known:

- f is a differentiable function defined at every real number x
- $f(0) = -1/2$
- $y = f'(x)$ has its graph given at center in Figure 3.4.19

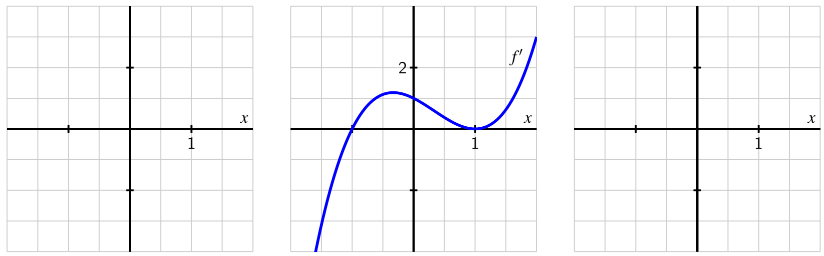


Figure 3.4.19 At center, a graph of $y = f'(x)$; at left, axes for plotting $y = f(x)$; at right, axes for plotting $y = f''(x)$.

a. Construct a first derivative sign chart for f. Clearly identify all critical numbers of f, where f is increasing and decreasing, and where f has local extrema.

b. On the right-hand axes, sketch an approximate graph of $y = f''(x)$.

c. Construct a second derivative sign chart for f. Clearly identify where f is concave up and concave down, as well as all inflection points.

d. On the left-hand axes, sketch a possible graph of $y = f(x)$.

2. Suppose that g is a differentiable function and $g'(2) = 0$. In addition, suppose that on $1 < x < 2$ and $2 < x < 3$ it is known that $g'(x)$ is positive.

a. Does g have a local maximum, local minimum, or neither at $x = 2$? Why?

b. Suppose that $g''(x)$ exists for every x such that $1 < x < 3$. Reasoning graphically, describe the behavior of $g''(x)$ for x-values near 2.

c. Besides being a critical number of g, what is special about the value $x = 2$ in terms of the behavior of the graph of g?

3. Let p be a function whose second derivative is $p''(x) = (x + 1)(x - 2)e^{-x}$.

a. Construct a second derivative sign chart for p and determine all inflection points of p.

b. Suppose you also know that $x = \frac{\sqrt{5}-1}{2}$ is a critical number of p. Does p have a local minimum, local maximum, or neither at $x = \frac{\sqrt{5}-1}{2}$? Why?

c. If the point $(2, \frac{12}{e^2})$ lies on the graph of $y = p(x)$ and $p'(2) = -\frac{5}{e^2}$, find the equation of the tangent line to $y = p(x)$ at the point where $x = 2$. Does the tangent line lie above the curve, below the curve, or neither at this value? Why?

3.5 Global Optimization and Applications

Motivating Questions

- What are the differences between finding relative extreme values and global extreme values of a function?
- How is the process of finding the global maximum or minimum of a function over the function's entire domain different from determining the global maximum or minimum on a restricted domain?
- For a function that is guaranteed to have both a global maximum and global minimum on a closed, bounded interval, what are the possible points at which these extreme values occur?

We have seen that we can use the first derivative of a function to determine where the function is increasing or decreasing, and the second derivative to know where the function is concave up or concave down. This information helps us determine the overall shape and behavior of the graph, as well as whether the function has relative extrema.

Remember the difference between a relative maximum and a global maximum: there is a *relative* maximum of f at $x = p$ if $f(p) \geq f(x)$ for all x *near* p, while there is a *global* maximum at p if $f(p) \geq f(x)$ for *all* x in the domain of f.

For instance, in Figure 3.5.1, we see a function f that has a global maximum at $x = c$ and a relative maximum at $x = a$, since $f(c)$ is greater than $f(x)$ for every value of x, while $f(a)$ is only greater than the value of $f(x)$ for x near a. Since the function appears to decrease without bound, f has no global minimum, though clearly f has a relative minimum at $x = b$.

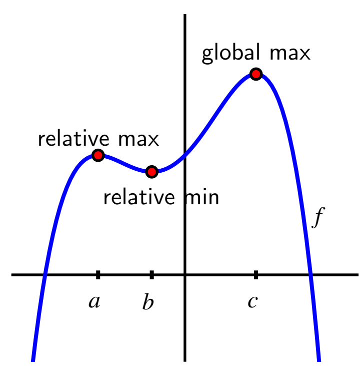


Figure 3.5.1 A function f with a global maximum, but no global minimum.

Our emphasis in this section is on finding the global extreme values of a function (if they exist), either over its entire domain or on some restricted portion.

Warm-Up 3.5.1 Let $f(x) = 2 + \frac{3}{1+(x+1)^2}$.

a. Determine all of the critical numbers of f.

b. Construct a first derivative sign chart for f and thus determine all intervals on which f is increasing or decreasing. Classify all critical numbers from part (a) as relative maxima, relative minima, or neither.

c. Do you think f has a global maximum? If so, why, and what is its value and where is the maximum attained? If not, explain why.

3.5.1 Global Optimization

In Figure 3.5.1 and Warm-Up 3.5.1, we were interested in finding the global maximum for f on its entire domain. This example highlights a fact that will be useful for identifying global extrema on open intervals (for example, $(-\infty, \infty)$):

Fact 3.5.2 Single Critical Point. *If f is a continuous function on an interval I which contains a single critical number $x = c$ in I, then*

- *if f has a local maximum at c, f also has a* global *maximum at c, and*
- *if f has a local minimum at c, f also has a* global *minimum at c.*

In other contexts, however, we will focus on some restriction of the domain. When the restricted domain is closed (i.e., includes its endpoints), there is a mechanical process for determining the extreme values.

For example, rather than considering $f(x) = 2 + \frac{3}{1+(x+1)^2}$ for every value of x, perhaps instead we are only interested in those x for which $0 \leq x \leq 4$, and we would like to know which values of x in the interval $[0, 4]$ produce the largest possible and smallest possible values of f. We are accustomed to critical numbers playing a key role in determining the location of extreme values of a function; now, by restricting the domain to an interval, it makes sense that the endpoints of the interval will also be important to consider, as we see in the following activity. When limiting ourselves to a particular interval, we will often refer to the *absolute* maximum or minimum value, rather than the *global* maximum or minimum.

Activity 3.5.2 Let $g(x) = \frac{1}{3}x^3 - 2x + 2$.

a. Find all critical numbers of g that lie in the interval $-2 \leq x \leq 3$.

b. Use a graphing utility to construct the graph of g on the interval $-2 \leq x \leq 3$.

c. From the graph, determine the x-values at which the absolute minimum and absolute maximum of g occur on the interval $[-2, 3]$.

d. How do your answers change if we instead consider the interval $-2 \leq x \leq 2$?

e. What if we instead consider the interval $-2 \leq x \leq 1$?

In Activity 3.5.2, we saw how the absolute maximum and absolute minimum of a function on a closed, bounded interval $[a, b]$, depend not only on the critical numbers of the function, but also on the values of a and b. These observations demonstrate several important facts that hold more generally. First, we state an important result called the Extreme Value Theorem.

The Extreme Value Theorem.

If f is a continuous function on a closed interval $[a, b]$, then f attains both an absolute minimum and absolute maximum on $[a, b]$. That is, for some value x_m such that $a \leq x_m \leq b$, it follows that $f(x_m) \leq f(x)$ for all x in $[a, b]$. Similarly, there is a value x_M in $[a, b]$ such that $f(x_M) \geq f(x)$ for all x in $[a, b]$. Letting $m = f(x_m)$ and $M = f(x_M)$, it follows that $m \leq f(x) \leq M$ for all x in $[a, b]$.

The Extreme Value Theorem tells us that on any closed interval $[a, b]$, a continuous function has to achieve both an absolute minimum and an absolute maximum. The theorem does not tell us where these extreme values occur, but rather only that they must exist. As we saw in Activity 3.5.2, the only possible locations for absolute extremes are at the endpoints of the interval or at a critical number.

Note 3.5.3 Thus, we have the following approach to finding the absolute maximum and minimum of a continuous function f on the interval $[a, b]$:

- find all critical numbers of f that lie in the interval;
- evaluate the function f at each critical number in the interval and at each endpoint of the interval;
- from among those function values, the smallest is the absolute minimum of f on the interval, while the largest is the absolute maximum.

Activity 3.5.3 Find the *exact* absolute maximum and minimum of each function on the stated interval.

a. $h(x) = xe^{-x}$, $[0, 3]$

b. $p(t) = \sin(t) + \cos(t)$, $[-\frac{\pi}{2}, \frac{\pi}{2}]$

c. $q(x) = \frac{x^2}{x-2}$, $[3, 7]$

d. $f(x) = 4 - e^{-(x-2)^2}$, $(-\infty, \infty)$ (*Find the global minimum only.*)

e. $h(x) = xe^{-ax}$, $[0, \frac{2}{a}]$ $(a > 0)$

The interval we choose has nearly the same influence on extreme values as the function under consideration. Consider, for instance, the function pictured in Figure 3.5.4.

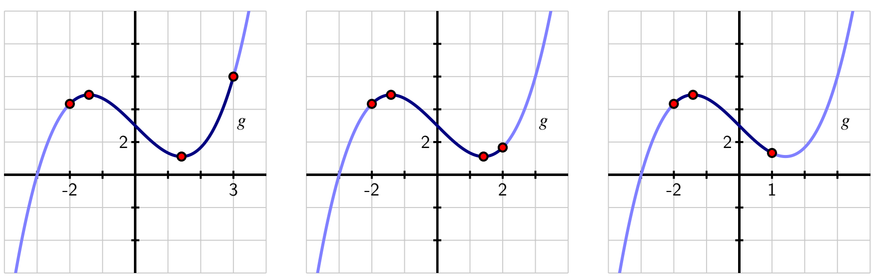


Figure 3.5.4 A function g considered on three different intervals.

In sequence, from left to right, the interval under consideration is changed from $[-2, 3]$ to $[-2, 2]$ to $[-2, 1]$.

- On the interval $[-2, 3]$, there are two critical numbers, with the absolute minimum at one critical number and the absolute maximum at the right endpoint.
- On the interval $[-2, 2]$, both critical numbers are in the interval, with the absolute minimum and maximum at the two critical numbers.
- On the interval $[-2, 1]$, just one critical number lies in the interval, with the absolute maximum at one critical number and the absolute minimum at one endpoint.

Remember to consider only the critical numbers that lie within the interval.

3.5.2 Applications

We conclude this section with some applications of optimization, highlighting the methods discussed so far in determining global extrema.

Example 3.5.5 A 20 ft piece of wood is cut into two pieces. One piece will be used to form a square garden box and the other to form a triangular garden box which is an equilateral triangle. How should the wood be cut to maximize the total area enclosed by the square and triangle boxes? to minimize the area?

Solution. We begin by sketching a picture that illustrates the situation. The variable in the problem is where we decide to cut the wood. We thus label the cut point at a distance x from one end of the wood, and note that the remaining portion of the wood then has length $20 - x$

As shown in Figure 3.5.6, we see that the x ft of wood that is used to form the equilateral triangle with three sides of length $\frac{x}{3}$. For the remaining $20 - x$ ft of wood, the square that results will have each side of length $\frac{20-x}{4}$.

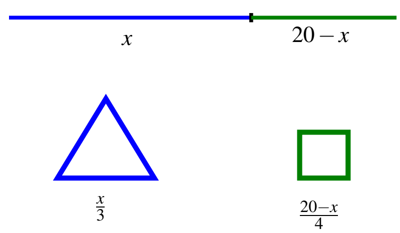


Figure 3.5.6 A 20 ft piece of wood cut into two pieces, one of which forms an equilateral triangle, the other which yields a square.

At this point, we note that there are obvious restrictions on x: in particular, $0 \leq x \leq 20$. In the extreme cases, all of the wire is being used to make just one figure. For instance, if $x = 0$, then all 20 ft of wire are used to make a square that is 5×5.

Now, our overall goal is to find the minimum and maximum areas that can be enclosed. Because the height of an equilateral triangle is $\sqrt{3}$ times half the length of the base, the area of the triangle is

$$A_{\Delta} = \frac{1}{2}bh = \frac{1}{2} \cdot \frac{x}{3} \cdot \frac{x\sqrt{3}}{6}.$$

The area of the square is $A_{\square} = s^2 = \left(\frac{20-x}{4}\right)^2$. Therefore, the total area function is

$$A(x) = \frac{\sqrt{3}x^2}{36} + \left(\frac{20-x}{4}\right)^2.$$

Remember that we are considering this function only on the restricted domain $[0, 20]$.

Differentiating $A(x)$, we have

$$A'(x) = \frac{\sqrt{3}x}{18} + 2\left(\frac{20-x}{4}\right)\left(-\frac{1}{4}\right) = \frac{\sqrt{3}}{18}x + \frac{1}{8}x - \frac{5}{2}.$$

When we set $A'(x) = 0$, we find that $x = \frac{180}{4\sqrt{3}+9} \approx 11.3007$ is the only critical number of A in the interval $[0, 20]$.

Evaluating A at the critical number and endpoints, we see that

- $A\left(\frac{180}{4\sqrt{3}+9}\right) = \frac{\sqrt{3}(\frac{180}{4\sqrt{3}+9})^2}{4} + \left(\frac{20 - \frac{180}{4\sqrt{3}+9}}{4}\right)^2 \approx 10.8741$

- $A(0) = 25$
- $A(20) = \frac{\sqrt{3}}{36}(400) = \frac{100}{9}\sqrt{3} \approx 19.2450$

Thus, the absolute minimum occurs when $x \approx 11.3007$ and results in the minimum area of approximately 10.8741 square feet. The absolute maximum occurs when we invest all of the wood in the square (and none in the triangle), resulting in 25 square feet of area. These results are confirmed by a plot of $y = A(x)$ on the interval $[0, 20]$, as shown in Figure 3.5.7.

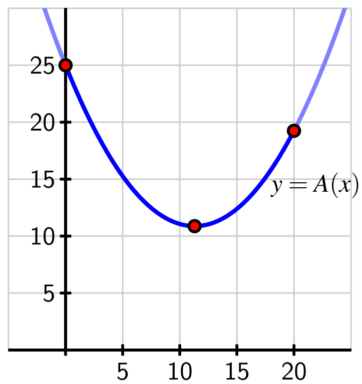


Figure 3.5.7 A plot of the area function from Example 3.5.5.

□

Activity 3.5.4 A piece of cardboard that is 10×15 (each measured in inches) is being made into a box without a top. To do so, squares are cut from each corner of the box and the remaining sides are folded up. If the box needs to be at least 1 inch deep and no more than 3 inches deep, what is the maximum possible volume of the box? what is the minimum volume? Justify your answers using calculus.

a. Draw a labeled diagram that shows the given information. What variable should we introduce to represent the choice we make in creating the box? Label the diagram appropriately with the variable, and write a sentence to state what the variable represents.

b. Determine a formula for the function V (that depends on the variable in (a)) that tells us the volume of the box.

c. What is the domain of the function V? That is, what values of x make sense for input? Are there additional restrictions provided in the problem?

d. Determine all critical numbers of the function V.

e. Evaluate V at each of the endpoints of the domain and at any critical numbers that lie in the domain.

f. What is the maximum possible volume of the box? the minimum?

Example 3.5.5 and Activity 3.5.4 illustrate standard steps that we undertake in almost every applied optimization problem: we draw a picture to demonstrate the situation, introduce one or more variables to represent quantities that are changing, find a function that models the quantity to be optimized, and then decide on an appropriate domain for that function. Once that is done, we are in the familiar situation of finding the absolute minimum and maximum of a function over a particular domain, so we apply the calculus ideas that we have been studying in this section.

We complete this section with a final optimization application relevant to foraging individuals in a system where food is obtained in discrete spaces which require time to travel between. One example is a bee traveling between flowers for nectar. If the bee's goal is to maximize its average food collection rate per flower, it must decide whether to stay on the current flower, where its rate of food collection is decreasing as the nectar is depleted, or travel to another flower with more nectar, but lose the time it takes to travel there.

To find the optimal amount of time the bee should stay on each flower, we must determine a model for the quantity we'd like to maximize, the bee's average food collection rate per flower after t seconds, which we will call $r(t)$. This includes the non-zero travel time it takes to get to another flower, which we will call τ. Assume $f(t)$ is the amount of food collected after t seconds on a flower, and that $f(t)$ is continuous and satisfies $f''(t) < 0$. This last assumption means f is concave down, or that the bee's food collection rate on the flower decreases as time goes on and nectar is depleted. We can then write $r(t)$ as

$$r(t) = \frac{\text{food collected after } t \text{ seconds on flower}}{\text{total time before next flower}} = \frac{f(t)}{t+\tau}.$$

The domain of $r(t)$ is the interval $[0, \infty)$. To find the maximum value of r on this domain we first search for critical numbers of r in $[0, \infty)$. We compute the first derivative of r as

$$r'(t) = \frac{f'(t)(t+\tau) - f(t) \cdot 1}{(t+\tau)^2}$$

Note that the denominator is always positive, so $r'(t)$ is not undefined in the domain $[0, \infty)$. Thus, t_c is a critical number if $r'(t_c) = 0$, which is true when $f'(t_c)(t_c + \tau) - f(t_c) = 0$. This is equivalent to $f'(t_c)(t_c + \tau) = f(t_c)$, which is equivalent to

$$f'(t_c) = \frac{f(t_c)}{t_c+\tau} = r(t_c).$$

In words, this says that if a time t_c is a critical number, then the instantaneous rate of change of the bee's food collection on the flower at time t_c is the same as the bee's average collection rate over the $(t_c + \tau)$ seconds it takes to be on the next flower. To verify that a maximum value occurs at a critical number t_c, the assumption that $f''(t) < 0$ is important. It is a good exercise in computing derivatives to verify that if we compute $r''(t)$, evaluate $r''(t_c)$ and then simplify, we get

$$r''(t_c) = \frac{f''(t_c)}{t_c+\tau}.$$

Since $t_c + \tau > 0$ and under the assumption that $f''(t_c) < 0$, this means that $r''(t_c) < 0$. Hence, by the second derivative test, a local maximum of $r(t)$ occurs at the critical number t_c. Further, every critical number t_c in the domain $[0, \infty)$ satisfies this property, and so every critical number in $[0, \infty)$ is a local maximum. Since $r(t)$ is a continuous function on the domain $[0, \infty)$, there can only be one such critical point. By Fact 3.5.2, a global maximum of $r(t)$ also occurs at the critical number t_c.

This general result is known as the *Marginal Value Theorem*, and is summarized below:

Fact 3.5.8 Marginal Value Theorem. *We have a system of foraging individuals in which there is a constant travel time $\tau > 0$ to get from one source to the next. Let $f(t)$ represent the amount of intake of a resource after time t*

spent on a source. Assume $f(t)$ is continuous and $f''(t) < 0$. Let $r(t) = \dfrac{f(t)}{t+\tau}$ be the average collection rate after spending time t on a single source. Then the time t_c an individual should spend on a single source in order to maximize its average collection rate $r(t)$ satisfies

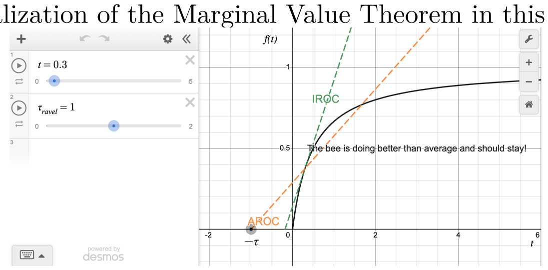
$$f'(t_c) = r(t_c).$$

Intuitively, the Marginal Value Theorem says that a bee should leave a flower the instant before its instantaneous collection rate drops below the average collection rate; that is, the bee should leave when it can do better on a different flower.

Example 3.5.9 Marginal Value Theorem. Suppose the food collection function for a bee is $f(t) = \dfrac{t}{0.5+t}$, and the travel time between flowers is $\tau = 1$ second. Then the average collection rate after t seconds is given by

$$r(t) = \frac{t}{(0.5+t)(t+1)}.$$

It is a good exercise to verify that the only positive critical number of $r(t)$ is $t_c = \sqrt{0.5}$, and that $f'(\sqrt{0.5}) = r(\sqrt{0.5})$. The interactive below shows a visualization of the Marginal Value Theorem in this example:



www.desmos.com/calculator/xen906mmtt

At times t when the bee is instantaneously doing better than average ($0 \le t < \sqrt{0.5}$), the bee should stay on the flower. This is when the slope of the tangent line at t is steeper than the slope of the secant line over $t+1$ seconds. At times t when the bee is instantaneously doing worse than average ($t > \sqrt{0.5}$), the bee would have done better to leaver earlier. This is when the slope of the tangent line at t is less steep than the slope of the secant line over $t+1$ seconds. Type in "$t = \sqrt{0.5}$" into the interactive. This is when the bee's instantaneous collection rate and average collection rate are equal, and is when the bee should leave.

How does the travel time τ impact when the bee should leave a flower? Answer the questions below based on your intuition, then use the interactive to confirm or deny your answer.

- If the travel time is larger than 1, should the bee leave sooner or later than it does when the travel time is 1 second? When the travel time is larger, there is more incentive to stay on the current flower, so the bee should leave later than when the travel time is 1 second.

- If the travel time is smaller than 1, should the bee leave sooner or later than it does when the travel time is 1 second? When the travel time is smaller, there is more incentive to leave the current flower, so the bee should leave sooner than when the travel time is 1 second.

□

3.5.3 Summary

- **Question 3.5.10** What are the differences between finding relative extreme values and global extreme values of a function? ☐

  To find relative extreme values of a function, we use a first derivative sign chart and classify all of the function's critical numbers. If instead we are interested in absolute extreme values, we first decide whether we are considering the entire domain of the function or a particular interval.

- **Question 3.5.11** How is the process of finding the global maximum or minimum of a function over the function's entire domain different from determining the global maximum or minimum on a restricted domain? ☐

  In the case of finding global extremes over the function's entire domain, we will typically see models that contain a single critical number in the domain. In this case we use Fact 3.5.2 to help determine the global extrema we're interested in. If we are working to find absolute extremes on a restricted interval, then we first identify all critical numbers of the function that lie in the interval.

- **Question 3.5.12** For a function that is guaranteed to have both a global maximum and global minimum on a closed, bounded interval, what are the possible points at which these extreme values occur? ☐

  For a continuous function on a closed, bounded interval, the only possible points at which absolute extreme values occur are the critical numbers and the endpoints. Thus, we simply evaluate the function at each endpoint and each critical number in the interval, and compare the results to decide which is largest (the absolute maximum) and which is smallest (the absolute minimum).

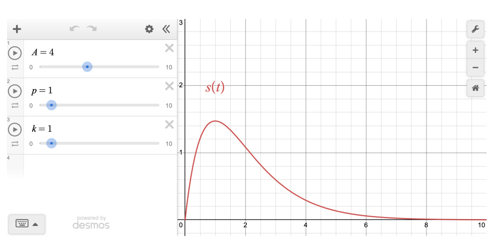
3.5.4 Exercises

1. Let $s(t)$ be the family of functions given by $s(t) = At^p e^{-kt}$, where A, p, and k are positive parameters. These functions are called **surge functions**, and can be used to model quantities that experience a "surge" before decaying slowly, such as the concentration of a medication after injection. An interactive is provided for you to verify the answers you compute below.

(a) Determine when the global maximum of $s(t)$ occurs on the domain $(0, \infty)$. Be sure to justify how you know it is a global maximum that occurs.

(b) Does the parameter A effect *where* the global maximum of $s(t)$ occurs? Explain how you know.

(c) Does the parameter A effect *the value* of the global maximum of $s(t)$? Explain how you know.



www.desmos.com/calculator/azlzzf2xc2

2. Let $f(t) = 6t + \dfrac{24}{t}$. Find the exact absolute maximum and minimum of f on the provided intervals by testing the endpoints and finding and evaluating all relevant critical numbers of f.

a. $[1, 3]$

b. $[1, 6]$

c. $[3, 6]$

3. Sketch the graph of a function that satisfies each set of criteria below, if one exists. If such a graph does not exist, explain how you know.

(a) Sketch a graph of $f(x)$ whose domain is $[-2, 3]$ such that f achieves an absolute maximum of $y = 3$ at $x = 0$ and $x = 2$, and an absolute minimum of $y = -5$ at $x = 3$.

(b) Sketch a graph of $h(x)$ whose domain is $[0, 10]$ such that h achieves an absolute maximum of $y = 0$ at $x = 5$ and does not have an absolute minimum value.

(c) Sketch a graph of $k(x)$ whose domain is $[0, 10]$ such that k has a single critical number at $x = 5$, and achieves its absolute maximum and minimum values at the endpoints.

(d) Sketch a graph of $j(x)$ whose domain is $(-\infty, \infty)$ such that j has a global maximum value and two critical points but does not have a global minimum value.

3.6 Limits: L'Hôpital's Rule

Motivating Questions

- How can derivatives be used to help us evaluate indeterminate limits of the form $\frac{0}{0}$?
- What does it mean to say that $\lim_{x\to\infty} f(x) = L$ and $\lim_{x\to a} f(x) = \infty$?
- How can derivatives assist us in evaluating indeterminate limits of the form $\frac{\infty}{\infty}$?

Because differential calculus is based on the definition of the derivative, and the definition of the derivative involves a limit, there is a sense in which all of calculus rests on limits. In addition, the limit involved in the definition of the derivative always generates the *indeterminate form* $\frac{0}{0}$. If f is a differentiable function, then in the definition

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h},$$

not only does $h \to 0$ in the denominator, but also $(f(x+h) - f(x)) \to 0$ in the numerator, since f is continuous. Remember, saying that a limit has an indeterminate form only means that we don't yet know its value and have more work to do: indeed, limits of the form $\frac{0}{0}$ can take on any value, as is evidenced by evaluating $f'(x)$ for varying values of x for a function such as $f'(x) = x^2$.

In Chapter 2, we learned many techniques for evaluating the limits that result from the derivative definition, including a large number of shortcut rules. In this section, we turn the situation upside-down: instead of using limits to evaluate derivatives, we explore how to use derivatives to evaluate certain limits.

Warm-Up 3.6.1 Let h be the function given by $h(x) = \frac{x^5+x-2}{x^2-1}$.

a. What is the domain of h?

b. Explain why $\lim_{x\to 1} \frac{x^5+x-2}{x^2-1}$ results in an indeterminate form.

c. Next we will investigate the behavior of both the numerator and denominator of h near the point where $x = 1$. Let $f(x) = x^5 + x - 2$ and $g(x) = x^2 - 1$. Find the local linearizations (Section 3.1) of f and g at $a = 1$, and call these functions $L_f(x)$ and $L_g(x)$, respectively.

d. Explain why $h(x) \approx \frac{L_f(x)}{L_g(x)}$ for x near $a = 1$.

e. Using your work from (c) and (d), evaluate

$$\lim_{x\to 1} \frac{L_f(x)}{L_g(x)}.$$

What do you think your result tells us about $\lim_{x\to 1} h(x)$?

f. Investigate the function $h(x)$ graphically and numerically near $x = 1$. What do you think is the value of $\lim_{x\to 1} h(x)$?

3.6.1 Using derivatives to evaluate indeterminate limits of the form $\frac{0}{0}$.

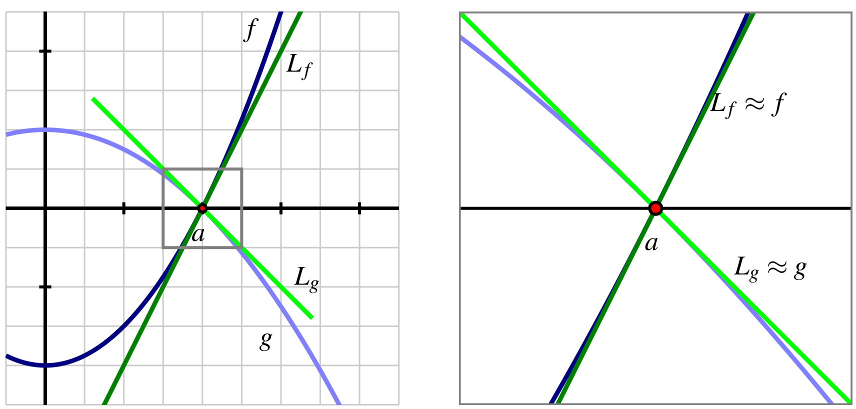


Figure 3.6.1 At left, the graphs of f and g near the value a, along with their tangent line approximations L_f and L_g at $x = a$. At right, zooming in on the point a and the four graphs.

The idea demonstrated in Warm-Up 3.6.1 — that we can evaluate an indeterminate limit of the form $\frac{0}{0}$ by replacing each of the numerator and denominator with their local linearizations at the point of interest — can be generalized in a way that enables us to evaluate a wide range of limits. We have a function $h(x)$ that can be written as a quotient $h(x) = \frac{f(x)}{g(x)}$, where f and g are both differentiable at $x = a$ and for which $f(a) = g(a) = 0$. We would like to evaluate the indeterminate limit given by $\lim_{x\to a} h(x)$. Figure 3.6.1 illustrates the situation. We see that both f and g have an x-intercept at $x = a$. Their respective tangent line approximations L_f and L_g at $x = a$ are also shown in the figure. We can take advantage of the fact that a function and its tangent line approximation become indistinguishable as $x \to a$.

First, let's recall that $L_f(x) = f'(a)(x-a) + f(a)$ and $L_g(x) = g'(a)(x-a) + g(a)$. Because x is getting arbitrarily close to a when we take the limit, we can replace f with L_f and replace g with L_g, and thus we observe that

$$\begin{aligned}\lim_{x\to a} \frac{f(x)}{g(x)} &= \lim_{x\to a} \frac{L_f(x)}{L_g(x)} \\ &= \lim_{x\to a} \frac{f'(a)(x-a)+f(a)}{g'(a)(x-a)+g(a)}.\end{aligned}$$

Next, we remember that both $f(a) = 0$ and $g(a) = 0$, which is precisely what makes the original limit indeterminate. Substituting these values for $f(a)$ and $g(a)$ in the limit above, we now have

$$\begin{aligned}\lim_{x\to a} \frac{f(x)}{g(x)} &= \lim_{x\to a} \frac{f'(a)(x-a)}{g'(a)(x-a)} \\ &= \lim_{x\to a} \frac{f'(a)}{g'(a)},\end{aligned}$$

where the latter equality holds because $\frac{x-a}{x-a} = 1$ when x is approaching (but not equal to) a. Finally, we note that $\frac{f'(a)}{g'(a)}$ is constant with respect to x, and thus

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

This result holds as long as $g'(a)$ is not equal to zero. The formal name of the result is L'Hôpital's Rule.

L'Hôpital's Rule.

Let f and g be differentiable at $x = a$, and suppose that $f(a) = g(a) = 0$ and that $g'(a) \neq 0$. Then $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}$.

In practice, we typically work with a slightly more general version of L'Hôpital's Rule, which states that if f and g are differentiable near a (except possibly at a), $g'(x) \neq 0$ near a (except possibly at a), and $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)},$$

provided the righthand limit exists. This form reflects the basic idea of L'Hôpital's Rule: if $\frac{f(x)}{g(x)}$ produces an indeterminate limit of form $\frac{0}{0}$ as $x \to a$, that limit is equivalent to the limit of the quotient of the two functions' derivatives, $\frac{f'(x)}{g'(x)}$.

For example, if we consider the limit from Warm-Up 3.6.1,

$$\lim_{x\to 1} \frac{x^5+x-2}{x^2-1},$$

by L'Hôpital's Rule we have that

$$\lim_{x\to 1} \frac{x^5+x-2}{x^2-1} = \lim_{x\to 1} \frac{5x^4+1}{2x} = \frac{6}{2} = 3.$$

By replacing the numerator and denominator with their respective derivatives, we often replace an indeterminate limit with one whose value we can easily determine.

Activity 3.6.2 Evaluate each of the following limits. If you use L'Hôpital's Rule, indicate where it was used, and be certain its hypotheses are met before you apply it.

a. $\displaystyle\lim_{x\to 0} \frac{\ln(1+x)}{x}$

b. $\displaystyle\lim_{x\to \pi} \frac{\cos(x)}{x}$

c. $\displaystyle\lim_{x\to 1} \frac{2\ln(x)}{1-e^{x-1}}$

d. $\displaystyle\lim_{x\to 0} \frac{\sin(x)-x}{\cos(2x)-1}$

While L'Hôpital's Rule can be applied in an entirely algebraic way, it is important to remember that the justification of the rule is graphical: the main idea is that the slopes of the tangent lines to f and g at $x = a$ determine the value of the limit of $\frac{f(x)}{g(x)}$ as $x \to a$.

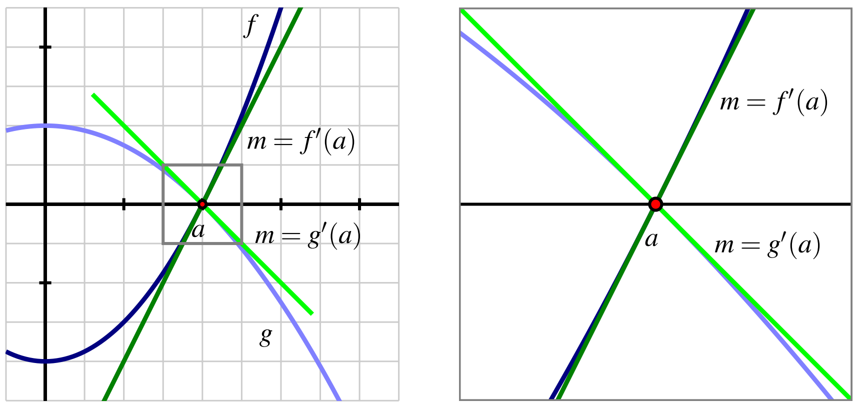


Figure 3.6.2 Two functions f and g that satisfy L'Hôpital's Rule.

We see this in Figure 3.6.2, where we can see from the grid that $f'(a) = 2$ and $g'(a) = -1$, hence by L'Hôpital's Rule,

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)} = \frac{2}{-1} = -2.$$

It's not the fact that f and g both approach zero that matters most, but rather the *rate* at which each approaches zero that determines the value of the limit. This is a good way to remember what L'Hôpital's Rule says: if $f(a) = g(a) = 0$, the the limit of $\frac{f(x)}{g(x)}$ as $x \to a$ is given by the ratio of the slopes of f and g at $x = a$.

Activity 3.6.3 In this activity, we reason graphically from the following figure to evaluate limits of ratios of functions about which some information is known.

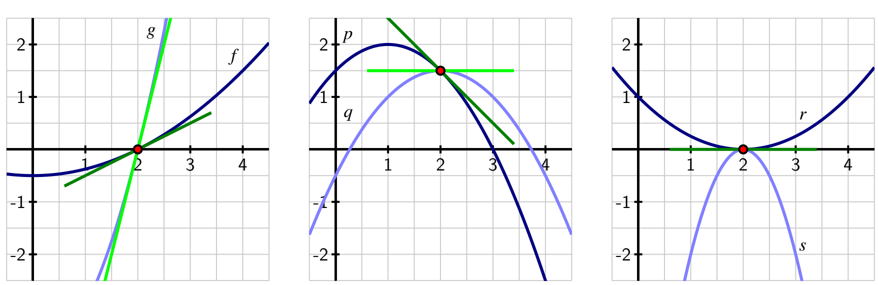


Figure 3.6.3 Three graphs referenced in the questions of Activity 3.6.3.

a. Use the left-hand graph to determine the values of $f(2)$, $f'(2)$, $g(2)$, and

$g'(2)$. Then, evaluate $\lim_{x \to 2} \frac{f(x)}{g(x)}$.

b. Use the middle graph to find $p(2)$, $p'(2)$, $q(2)$, and $q'(2)$. Then, determine the value of $\lim_{x \to 2} \frac{p(x)}{q(x)}$.

c. Assume that r and s are functions whose for which $r''(2) \neq 0$ and $s''(2) \neq 0$ Use the right-hand graph to compute $r(2)$, $r'(2)$, $s(2)$, $s'(2)$. Explain why you cannot determine the exact value of $\lim_{x \to 2} \frac{r(x)}{s(x)}$ without further information being provided, but that you can determine the sign of $\lim_{x \to 2} \frac{r(x)}{s(x)}$. In addition, state what the sign of the limit will be, with justification.

3.6.2 Limits involving ∞

The concept of infinity, denoted ∞, arises naturally in calculus, as it does in much of mathematics. It is important to note from the outset that ∞ is a concept, but not a number itself. Indeed, the notion of ∞ naturally invokes the idea of limits. Consider, for example, the function $f(x) = \frac{1}{x}$, whose graph is pictured in Figure 3.6.4.

We note that $x = 0$ is not in the domain of f, so we may naturally wonder what happens as $x \to 0$. As $x \to 0^+$, we observe that $f(x)$ *increases without bound*. That is, we can make the value of $f(x)$ as large as we like by taking x closer and closer (but not equal) to 0, while keeping $x > 0$. This is a good way to think about what infinity represents: a quantity is tending to infinity if there is no single number that the quantity is always less than.

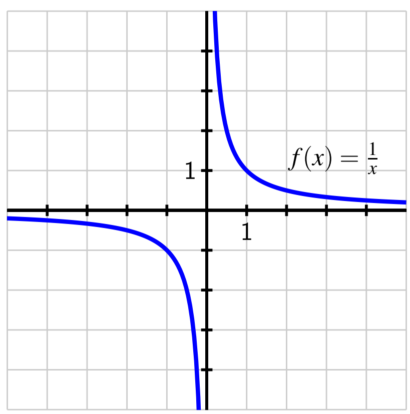


Figure 3.6.4 The graph of $f(x) = \frac{1}{x}$.

Recall that the statement $\lim_{x \to a} f(x) = L$, means that we can make $f(x)$ as close to L as we'd like by taking x sufficiently close (but not equal) to a. We now expand this notation and language to include the possibility that either L or a can be ∞. For instance, for $f(x) = \frac{1}{x}$, we now write

$$\lim_{x \to 0^+} \frac{1}{x} = \infty,$$

by which we mean that we can make $\frac{1}{x}$ as large as we like by taking x sufficiently close (but not equal) to 0. In a similar way, we write

$$\lim_{x \to \infty} \frac{1}{x} = 0,$$

since we can make $\frac{1}{x}$ as close to 0 as we'd like by taking x sufficiently large (i.e., by letting x increase without bound).

In general, the notation $\lim_{x\to a} f(x) = \infty$ means that we can make $f(x)$ as large as we like by taking x sufficiently close (but not equal) to a, and the notation $\lim_{x\to\infty} f(x) = L$ means that we can make $f(x)$ as close to L as we like by taking x sufficiently large. This notation also applies to left- and right-hand limits, and to limits involving $-\infty$. For example, returning to Figure 3.6.4 and $f(x) = \frac{1}{x}$, we can say that

$$\lim_{x\to 0^-} \frac{1}{x} = -\infty \quad \text{and} \quad \lim_{x\to -\infty} \frac{1}{x} = 0.$$

Finally, we write

$$\lim_{x\to\infty} f(x) = \infty$$

if we can make the value of $f(x)$ as large as we'd like by taking x sufficiently large. For example,

$$\lim_{x\to\infty} x^2 = \infty.$$

Limits involving infinity identify *vertical* and *horizontal asymptotes* of a function. If $\lim_{x\to a} f(x) = \infty$, then $x = a$ is a vertical asymptote of f, while if $\lim_{x\to\infty} f(x) = L$, then $y = L$ is a horizontal asymptote of f. Similar statements can be made using $-\infty$, and with left- and right-hand limits as $x \to a^-$ or $x \to a^+$.

In precalculus classes, it is common to study the *end behavior* of certain families of functions, by which we mean the behavior of a function as $x \to \infty$ and as $x \to -\infty$. Here we briefly examine some familiar functions and note the values of several limits involving ∞.

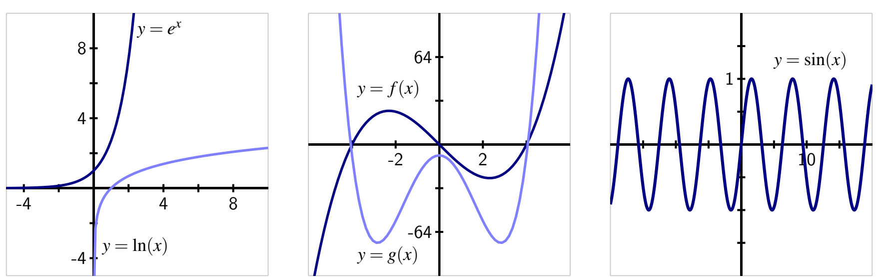


Figure 3.6.5 Graphs of some familiar functions whose end behavior as $x \to \pm\infty$ is known. In the middle graph, $f(x) = x^3 - 16x$ and $g(x) = x^4 - 16x^2 - 8$.

For the natural exponential function e^x, we note that $\lim_{x\to\infty} e^x = \infty$ and $\lim_{x\to-\infty} e^x = 0$. For the exponential decay function e^{-x}, these limits are reversed, with $\lim_{x\to\infty} e^{-x} = 0$ and $\lim_{x\to-\infty} e^{-x} = \infty$. Turning to the natural logarithm function, we have $\lim_{x\to 0^+} \ln(x) = -\infty$ and $\lim_{x\to\infty} \ln(x) = \infty$. While both e^x and $\ln(x)$ grow without bound as $x \to \infty$, the exponential function does so much more quickly than the logarithm function does. We'll soon use limits to quantify what we mean by "quickly."

For polynomial functions of the form

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0,$$

the end behavior depends on the sign of a_n and whether the highest power n is even or odd. If n is even and a_n is positive, then $\lim_{x\to\infty} p(x) = \infty$ and $\lim_{x\to-\infty} p(x) = \infty$, as in the plot of g in Figure 3.6.5. If instead a_n is negative, then $\lim_{x\to\infty} p(x) = -\infty$ and $\lim_{x\to-\infty} p(x) = -\infty$. In the situation where n

is odd, then either $\lim_{x\to\infty} p(x) = \infty$ and $\lim_{x\to-\infty} p(x) = -\infty$ (which occurs when a_n is positive, as in the graph of f in Figure 3.6.5), or $\lim_{x\to\infty} p(x) = -\infty$ and $\lim_{x\to-\infty} p(x) = \infty$ (when a_n is negative).

A function can fail to have a limit as $x \to \infty$. For example, consider the plot of the sine function at right in Figure 3.6.5. Because the function continues oscillating between -1 and 1 as $x \to \infty$, we say that $\lim_{x\to\infty} \sin(x)$ does not exist.

Finally, it is straightforward to analyze the behavior of any rational function as $x \to \infty$.

Example 3.6.6 Determine the limit of the function

$$q(x) = \frac{3x^2 - 4x + 5}{7x^2 + 9x - 10}$$

as $x \to \infty$.

Note that both $(3x^2 - 4x + 5) \to \infty$ as $x \to \infty$ and $(7x^2 + 9x - 10) \to \infty$ as $x \to \infty$. Here we say that $\lim_{x\to\infty} q(x)$ has indeterminate form $\frac{\infty}{\infty}$. We can determine the value of this limit through a standard algebraic approach. Multiplying the numerator and denominator each by $\frac{1}{x^2}$, we find that

$$\begin{aligned}\lim_{x\to\infty} q(x) &= \lim_{x\to\infty} \frac{3x^2 - 4x + 5}{7x^2 + 9x - 10} \cdot \frac{\frac{1}{x^2}}{\frac{1}{x^2}} \\ &= \lim_{x\to\infty} \frac{3 - 4\frac{1}{x} + 5\frac{1}{x^2}}{7 + 9\frac{1}{x} - 10\frac{1}{x^2}} = \frac{3}{7}\end{aligned}$$

since $\frac{1}{x^2} \to 0$ and $\frac{1}{x} \to 0$ as $x \to \infty$. This shows that the rational function q has a horizontal asymptote at $y = \frac{3}{7}$. A similar approach can be used to determine the limit of any rational function as $x \to \infty$. □

But how should we handle a limit such as

$$\lim_{x\to\infty} \frac{x^2}{e^x}?$$

Here, both $x^2 \to \infty$ and $e^x \to \infty$, but there is not an obvious algebraic approach that enables us to find the limit's value. Fortunately, it turns out that L'Hôpital's Rule extends to cases involving infinity.

L'Hôpital's Rule (∞).

If f and g are differentiable near a and both approach zero or both approach $\pm\infty$ as $x \to a$ (where a is allowed to be $\pm\infty$) , then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)},$$

provided the righthand limit exists.

(To be technically correct, we need to add the additional hypothesis that $g'(x) \neq 0$ near a; this is almost always met in practice.)

To evaluate $\lim_{x\to\infty} \frac{x^2}{e^x}$, we can apply L'Hôpital's Rule, since both $x^2 \to \infty$ and $e^x \to \infty$. Doing so, it follows that

$$\lim_{x\to\infty} \frac{x^2}{e^x} = \lim_{x\to\infty} \frac{2x}{e^x}.$$

This updated limit is still indeterminate and of the form $\frac{\infty}{\infty}$, but it is simpler since $2x$ has replaced x^2. Hence, we can apply L'Hôpital's Rule again, and find that

$$\lim_{x\to\infty} \frac{x^2}{e^x} = \lim_{x\to\infty} \frac{2x}{e^x} = \lim_{x\to\infty} \frac{2}{e^x}.$$

Now, since 2 is constant and $e^x \to \infty$ as $x \to \infty$, it follows that $\frac{2}{e^x} \to 0$ as $x \to \infty$, which shows that

$$\lim_{x\to\infty} \frac{x^2}{e^x} = 0.$$

Activity 3.6.4 Evaluate each of the following limits. If you use L'Hôpital's Rule, indicate where it was used, and be certain its hypotheses are met before you apply it.

a. $\lim_{x\to\infty} \frac{x}{\ln(x)}$

b. $\lim_{x\to\infty} \frac{e^x + x}{2e^x + x^2}$

c. $\lim_{x\to 0^+} \frac{\ln(x)}{\frac{1}{x}}$

d. $\lim_{x\to \frac{\pi}{2}^-} \frac{\tan(x)}{x - \frac{\pi}{2}}$

e. $\lim_{x\to\infty} xe^{-x}$

To evaluate the limit of a quotient of two functions $\frac{f(x)}{g(x)}$ that results in an indeterminate form of $\frac{\infty}{\infty}$, in essence we are asking which function is growing faster without bound. We say that the function g **dominates** the function f as $x \to \infty$ provided that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0,$$

whereas f dominates g provided that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \infty$. Finally, if the value of $\lim_{x\to\infty} \frac{f(x)}{g(x)}$ is finite and nonzero, we say that f and g *grow at the same rate*. For example, we saw that $\lim_{x\to\infty} \frac{x^2}{e^x} = 0$, so e^x dominates x^2, while $\lim_{x\to\infty} \frac{3x^2-4x+5}{7x^2+9x-10} = \frac{3}{7}$, so $f(x) = 3x^2 - 4x + 5$ and $g(x) = 7x^2 + 9x - 10$ grow at the same rate.

3.6.3 Summary

- **Question 3.6.7** How can derivatives be used to help us evaluate indeterminate limits of the form $\frac{0}{0}$? ☐

  Derivatives can be used to help us evaluate indeterminate limits of the form $\frac{0}{0}$ through L'Hôpital's Rule, by replacing the functions in the numerator and denominator with their tangent line approximations. In particular, if $f(a) = g(a) = 0$ and f and g are differentiable at a, L'Hôpital's Rule tells us that

  $$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

- **Question 3.6.8** What does it mean to say that $\lim_{x\to\infty} f(x) = L$ and $\lim_{x\to a} f(x) = \infty$? ☐

  When we write $x \to \infty$, this means that x is increasing without bound. Thus, $\lim_{x\to\infty} f(x) = L$ means that we can make $f(x)$ as close to L as we like by choosing x to be sufficiently large. Similarly, $\lim_{x\to a} f(x) = \infty$, means that we can make $f(x)$ as large as we like by choosing x sufficiently close to a.

- **Question 3.6.9** How can derivatives assist us in evaluating indeterminate limits of the form $\frac{\infty}{\infty}$? ☐

A version of L'Hôpital's Rule also helps us evaluate indeterminate limits of the form $\frac{\infty}{\infty}$. If f and g are differentiable and both approach zero or both approach $\pm\infty$ as $x \to a$ (where a is allowed to be ∞), then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

3.6.4 Exercises

1. Let f and g be differentiable functions about which the following information is known: $f(3) = g(3) = 0$, $f'(3) = g'(3) = 0$, $f''(3) = -2$, and $g''(3) = 1$. Let a new function h be given by the rule $h(x) = \frac{f(x)}{g(x)}$. On the same set of axes, sketch possible graphs of f and g near $x = 3$, and use the provided information to determine the value of

$$\lim_{x\to 3} h(x).$$

Provide explanation to support your conclusion.

2. Consider the function $g(x) = x^{2x}$, which is defined for all $x > 0$. Observe that $\lim_{x\to 0^+} g(x)$ is indeterminate due to its form of 0^0.

a. Let $h(x) = \ln(g(x))$. Explain why $h(x) = 2x\ln(x)$.

b. Next, explain why it is equivalent to write $h(x) = \frac{2\ln(x)}{\frac{1}{x}}$.

c. Use L'Hôpital's Rule and your work in (b) to compute $\lim_{x\to 0^+} h(x)$.

d. Explain why $g(x) = e^{h(x)}$. Then use the value of $\lim_{x\to 0^+} h(x)$ to determine $\lim_{x\to 0^+} g(x)$. Check your answer using technology by graphing $g(x)$

3. Recall we say that function g **dominates** function f provided that $\lim_{x\to\infty} f(x) = \infty$, $\lim_{x\to\infty} g(x) = \infty$, and $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$. Use limits to justify each of your answers to the questions below.

a. Which function dominates the other: x^3 or x^2?

b. Which function dominates the other: 2^x or 5^x?

c. Which function dominates the other: $\ln(x)$ or $\sqrt{x}$?

d. Which function dominates the other: $\ln(x)$ or $\sqrt[n]{x}$? (n can be any positive integer)

e. Explain why e^x will dominate any polynomial function.

f. Explain why x^n will dominate $\ln(x)$ for any positive integer n.

g. Give any example of two nonlinear functions such that neither dominates the other.

3.7 Limits: Leading Behaviors

Motivating Questions

- How can we evaluate complicated limits as $x \to \infty$?
- How can we evaluate complicated limits as $x \to -\infty$?

In this section we continue with the general goal of Section 3.6: evaluating limits. In this section we will focus on **limits at infinity**, which have the form

$$\lim_{x\to\infty} f(x) \text{ and } \lim_{x\to-\infty} f(x).$$

These are important for determining the *end behavior* of a function, which can tell us how a system behaves in the long run. They can help us answer questions like "is the population headed towards extinction?" and "what is the baseline concentration of a substance in the bloodstream?". We were introduced to basic limits of this form in Subsection 3.6.2. In this section, we will use the *method of leading behaviors* to help us evaluate more complicated limits at infinity.

Warm-Up 3.7.1

1. Graph $y = x^4 - 2x + \sqrt{x} - 10$ and $y = x^4$ in Desmos. Compare and contrast the graphs that you see. What is the same? What is different?
2. Graph $y = -3x^5 + 10x + \ln(x) - 50$ and $y = -3x^5$ in Desmos. Compare and contrast the graphs that you see. What is the same? What is different?
3. Graph $g(x) = x^{10} + 2x + \sqrt{x} - e^x$ in Desmos. What is $\lim_{x\to\infty} g(x)$? Why do you think this is the limit's value, and how could you determine this value just using the equation of g?

3.7.1 Leading Behavior at ∞

In Warm-Up 3.7.1 we observed a complicated function built as the sum of simpler functions can exhibit *local behavior* that is complicated to predict without computation, like its zeros and its local extrema. However, if we are interested in describing the *end behavior* of such a function, as we are in this section, it can be predicted based on the behavior of a single simpler function. Indeed, if $f(x) = f_1(x) + f_2(x) + f_3(x)$, and $|f_1(x)|$ dominates $|f_2(x)|$ and $|f_3(x)|$ as $x \to \infty$, then the graph of f will start to behave like the graph of f_1 as $x \to \infty$. For such a function, we give the dominant term a special name.

Definition 3.7.1 Let $f(x)$ be a sum of functions. If one function in the sum dominates all of the other functions in the sum in absolute value as $x \to \infty$, we call this function **the leading behavior of f at ∞**. We denote the leading behavior of f at ∞ as $f_\infty(x)$. ♢

Note that leading behavior is determined by which term dominates *in absolute value*. A term being positive or negative can impact the value of the limit, but it does not effect whether or not a term is the leading behavior. As we described above, the leading behavior of a function at ∞ can be useful in evaluating limits at infinity:

Leading Behavior at ∞.

Let $f(x)$ be a function with leading behavior at ∞ $f_\infty(x)$. Then

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} f_\infty(x).$$

The benefit of leading behaviors is that it allows us to replace a complicated function with a simpler one that has the same end behavior.

Example 3.7.2 Leading Behavior at ∞. Let $g(x) = -x^3 - \sqrt{x} + \ln(x)$ and $h(x) = e^{-x} + 2^x + x^{100}$. In Section 3.6 we determined that as $x \to \infty$, exponential growth dominates power growth dominates logarithmic growth. Therefore, we know that $g_\infty(x) = -x^3$, since $|-x^3|$ dominates the other terms as $x \to \infty$. Also, $h_\infty(x) = 2^x$ since exponential growth dominates power function growth. We must be careful to notice here that the leading behavior at infinity is not e^{-x}, since this function decays to 0 as $x \to \infty$.

We could also use leading behaviors to evaluate the limit of the ratio of g and h:

$$\lim_{x\to\infty} \frac{-x^3 - \sqrt{x} + \ln(x)}{e^{-x} + 2^x + x^{100}} = \lim_{x\to\infty} \frac{g_\infty(x)}{h_\infty(x)} = \lim_{x\to\infty} \frac{-x^3}{2^x} = 0,$$

since 2^x dominates $|-x^3|$ as $x \to \infty$. □

Activity 3.7.2 For each function pair f, g below, first determine $f_\infty(x)$ and $g_\infty(x)$. Then evaluate $\lim_{x\to\infty} \dfrac{f(x)}{g(x)}$.

1. $f(x) = e^{-2x} + x + e^{0.1x}$, $g(x) = x^2 - x + \ln(x)$
2. $f(x) = x^{-10} + \sqrt{x} + 0.5^x$, $g(x) = -\ln(x) + x^{-1}$
3. $f(x) = -3x^2 + \sqrt[3]{x^5} + 0.5^x$, $g(x) = x^2 + x^{-1}$

3.7.2 Leading Behavior at $-\infty$

Leading behavior at $-\infty$ is defined similarly to leading behavior at ∞:

Definition 3.7.3 Let $f(x)$ be a sum of functions. If one function in the sum dominates all of the other functions in the sum in absolute value as $x \to -\infty$, we call this function **the leading behavior of f at $-\infty$**. We denote the leading behavior of f at $-\infty$ as $f_{-\infty}(x)$. ◊

We can also use $f_{-\infty}(x)$ similarly to help us evaluate limit as $x \to -\infty$:

Leading Behavior at $-\infty$.

Let $f(x)$ be a function with leading behavior at $-\infty$ $f_{-\infty}(x)$. Then

$$\lim_{x\to-\infty} f(x) = \lim_{x\to-\infty} f_{-\infty}(x).$$

The main thing we must be careful of with leading behaviors at $-\infty$ is to pay attention to the behavior of the function as $x \to -\infty$; that is, as we move to the left on the graph of the function. This behavior is in general different than the behavior as $x \to \infty$ (as we move to the right on the graph of the function). For example, e^x is a common leading behavior at ∞ because it is exponential growth, and gets large quickly as $x \to \infty$. However, it is not a common leading behavior at $-\infty$ since it decays to 0 as $x \to -\infty$: $\lim_{x\to-\infty} e^x = 0$.

Example 3.7.4 Leading Behavior at $-\infty$. Let $g(x) = -x^3 - \sqrt[3]{x} + \ln(-x)$ and $h(x) = e^{-x} + 2^x + x^{100}$. Then $g_{-\infty}(x) = -x^3$, since $|-x^3|$ dominates the other terms as $x \to -\infty$. For $h(x)$, note that 2^x gets very small as $x \to -\infty$. The function x^{100} gets very large as $x \to -\infty$, but is power function growth. The function e^{-x} grows exponentially as $x \to -\infty$. Thus, $h_{-\infty}(x) = e^{-x}$.

We can use leading behaviors to evaluate the limit of the ratio of g and h:

$$\lim_{x\to-\infty} \frac{-x^3 - \sqrt[3]{x} + \ln(-x)}{e^{-x} + 2^x + x^{100}} = \lim_{x\to-\infty} \frac{g_{-\infty}(x)}{h_{-\infty}(x)} = \lim_{x\to-\infty} \frac{-x^3}{e^{-x}} = 0,$$

since e^{-x} dominates $|-x^3|$ as $x \to -\infty$. □

Activity 3.7.3 For each function pair f, g below, first determine $f_{-\infty}(x)$ and $g_{-\infty}(x)$. Then evaluate $\lim_{x\to-\infty} \dfrac{f(x)}{g(x)}$.

1. $f(x) = e^{-2x} + x + e^{0.1x}$, $g(x) = x^2 - x + \ln(-x)$
2. $f(x) = x^{-10} + \sqrt{-x} + 0.5^x$, $g(x) = -\ln(-x) + x^{-1}$
3. $f(x) = -3x^2 + \sqrt[3]{x^5} + 0.5^x$, $g(x) = x^2 + x^{-1}$

3.7.3 Summary

- **Question 3.7.5** How can we evaluate complicated limits as $x \to \infty$? □

  To evaluate limits of the form $\lim_{x\to\infty} f(x)$, we can determine if $f(x)$ has a leading behavior at ∞, $f_\infty(x)$. If so, we know

  $$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} f_\infty(x).$$

- **Question 3.7.6** How can we evaluate complicated limits as $x \to -\infty$? □

  To evaluate limits of the form $\lim_{x\to-\infty} f(x)$, we can determine if $f(x)$ has a leading behavior at $-\infty$, $f_{-\infty}(x)$. If so, we know

  $$\lim_{x\to-\infty} f(x) = \lim_{x\to-\infty} f_{-\infty}(x).$$

3.7.4 Exercises

1. Let $k(x) = e^x + e^{-x}$.

(a) Determine $k_\infty(x)$.

(b) Evaluate $\displaystyle\lim_{x\to\infty} \frac{e^x}{e^x + e^{-x}}$.

(c) Determine $k_{-\infty}(x)$.

(d) Evaluate $\displaystyle\lim_{x\to-\infty} \frac{e^x}{e^x + e^{-x}}$.

2. Determine the *equation* of a function that satisfies each set of criteria below. Justify why your equation satisfies the criteria using leading behaviors.

(a) A function h such that $\lim_{x\to\infty} h(x) = 3$ and $\lim_{x\to-\infty} h(x) = 0$.

(b) A function j such that $\lim_{x\to\infty} j(x) = 3$ and $\lim_{x\to-\infty} j(x) = 2$.

Chapter 4

Continuous-Time Dynamical Systems

4.1 Introduction to Differential Equations and Antiderivatives

Motivating Questions

- What is a differential equation and how do these models relate to discrete-time dynamical systems?
- What is a solution to a differential equation and how does this relate to the concept of an antiderivative?

In Section 1.7 we were introduced to *discrete-time dynamical systems*, which describe systems of data taken at equally spaced intervals. These systems have an *updating function* (which can be written as a *difference equation*) that describes how the system's average rate of change between discrete points behaves over time. If we have an *initial value* for the system, we can describe an associated *solution function*, which is an explicit description of how the system behaves over time.

For a discrete-time dynamical system, describing the *average* rate of change of the system makes sense since we have discrete points. However, if we are modeling *continuous-time dynamical systems*, we have enough data to describe the system's *instantaneous* rate of change. Such models are called *differential equations*. Similar to discrete-time dynamical systems, given an initial value an important question is to describe the associated solution function.

Warm-Up 4.1.1 Let $s(t)$ be a function such that $s'(t) = 4t + 1$.

1. What are the critical number(s) of $s(t)$? How do you know?
2. When is s increasing? decreasing? How do you know?
3. Sketch at least two possible graphs of $s(t)$ using your answers from above. How are they related?

4.1.1 Differential Equations

A **differential equation** is an equation that describes derivative values of a function that is unknown to us. For example, the equation $s'(t) = 4t + 1$

from Warm-Up 4.1.1 is a differential equation which describes a function $s(t)$. Note that $\frac{ds}{dt} = 4t + 1$ is the same differential equation which uses a different notation for the first derivative. Differential equations are a very broad class of equations, and there are courses devoted to just this topic. As such, this text will only explore the basics of analyzing specific types of differential equations.

Example 4.1.1 Differential Equations. The following equations are all examples of a differential equation:

1. $s''(t) = -9.8$
2. $\frac{dA}{dt} = \cos\left(\frac{\pi}{12}t\right)$
3. $\frac{dP}{dt} = 0.03P$
4. $\frac{dy}{dx} = y - x$

Equation number 1 is a **second-order differential equation** because the highest derivative involved is the second derivative. All other equations are an example of a **first-order differential equation** since the highest derivative involved is the first derivative.

Equations number 1 and number 2 are **pure-time differential equations** because the derivative value does not depend on the dependent variable (s in number 1 and A in number 2). It is common that differential equations are used to model relationships whose input quantity is "time", which is why we will often use t as the independent variable (and why we use the phrase "pure-time"). However, it is not required that time is the input quantity nor that we use t as the independent variable, as shown in number 4.

Equation number 3 is an **autonomous differential equation** because the derivative value depends on the dependent variable (P), but not the independent variable (t). number 4 is neither pure-time nor autonomous because the derivative value depends on both the independent (x) and dependent (y) variables. □

This text will largely focus on *first-order, pure-time* differential equations. We will, however, describe some approximation techniques that will work for many types of differential equations in Section 4.6.

We study differential equations because they arise naturally when describing phenomena that we observe in the real world. For instance:

- Example 4.1.1 number 1 can be used to model an object's position s after t seconds of free-fall.
- Example 4.1.1 number 2 can be used to model the outside temperature A after t hours.
- Example 4.1.1 number 3 can be used to model a population size P after t years.

To see many more examples of differential equations in use to describe real-world systems, take a moment to look back at the resources described in Exercise 1.1.4.2.

4.1.2 Solutions to Differential Equations and Antiderivatives

When working with a discrete-time dynamical system, an updating function was a convenient way to record observations. However, to analyze and understand the system being observed, it was often necessary to describe a solution function, which showed the explicit relationship between the input and output quantities of the system.

Similarly, when working with a continuous-time dynamical system, a differential equation is a convenient way to record how the system behaves. However, to analyze and understand the system, we would like to describe a **solution**. By a *solution* to a differential equation, we mean simply a function that satisfies this description.

For instance, the first differential equation we looked at in Warm-Up 4.1.1 is

$$\frac{ds}{dt} = 4t + 1,$$

which describes an unknown function $s(t)$. We may check that $s(t) = 2t^2 + t$ is a solution because it satisfies this description:

$$\frac{ds}{dt} = \frac{d}{dt}\left(2t^2 + t\right) = 4t + 1$$

Notice that $s(t) = 2t^2 + t + 4$ is also a solution.

When working with pure-time differential equations, checking whether a function candidate is a solution resembles the process above: take the derivative of the function candidate, and check whether it matches the differential equation. We give solutions to pure-time differential equations a special name:

Definition 4.1.2 Let $f(x)$ and $F(x)$ be functions such that $F'(x) = f(x)$. We say that $F(x)$ is an **antiderivative** of $f(x)$. ♢

Remark 4.1.3

1. For a pure-time differential equation $\frac{dy}{dt} = f(t)$, any antiderivative of f, $y = F(t)$, is a solution.

2. Since the derivative of a constant is zero, if a function has at least one antiderivative it has infinitely many. This means that a differential equation may have infinitely many possible solutions.

We will focus on computing antiderivatives of various functions in Section 4.2. For the remainder of this section, we will emphasize verifying and visualizing solutions to pure-time differential equations.

Activity 4.1.2 Consider the differential equation $\frac{dy}{dx} = e^{2x} - \sin(x) + \sqrt{x}$.

1. Determine which function below is a solution to this differential equation:

$$y = e^{2x} - \cos(x) + x^{\frac{3}{2}}$$

OR

$$y = 0.5e^{2x} + \cos(x) + \frac{2}{3}x^{\frac{3}{2}}$$

2. Give an example of one more solution to the differential equation that is different from the solution in part 1.

3. Suppose we also know $y(0) = 2$. How many solutions will satisfy this initial value condition? How do you know?

As we've seen, a differential equation in and of itself can describe many solution functions. Given a differential equation $\frac{dy}{dt}$, we call the family of all possible solutions $y(t)$ the **general solution** of the differential equation. If we are also provided an initial value $y(a) = b$ (note the initial value does not need to be the function's value at 0), then the differential equation/initial value pair will have one solution. A differential equation/initial value pair is called an **initial value problem**, which will have exactly one solution called the **specific solution** of the initial value problem.[1]

Example 4.1.4 Initial Value Problem. Consider the differential equation $\frac{ds}{dt} = 4t + 1$ from Warm-Up 4.1.1. We can verify that the general solution to this differential equation is $s(t) = 2t^2 + t + C$, where C is any constant, by computing

$$\frac{ds}{dt} = \frac{d}{dt}\left(2t^2 + t + C\right) = 4t + 1.$$

If we are instead given the initial value problem $\frac{ds}{dt} = 4t + 1$, $s(1) = 10$, then we can find the specific solution by solving for C:

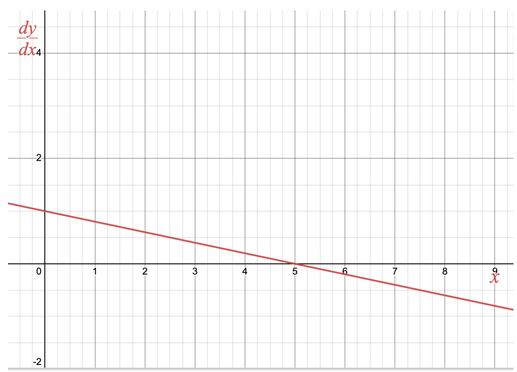
$$10 = 2(1)^2 + 1 + C,$$

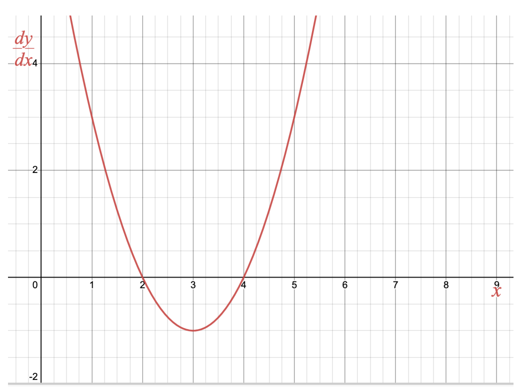
so

$$C = 10 - 2 - 1 = 7.$$

□

Activity 4.1.3 For each graph of a differential equation $\frac{dy}{dx}$ below, sketch two solution functions: one with initial value $y(0) = 0$ and one with initial value $y(0) = 2$.

1. 

2. 

[1] Technically, the differential equation must be "well-behaved" for this to be true, though we will not worry about this technical condition as it will always be true in the differential equations we study.

4.1.3 Summary

- **Question 4.1.5** What is a differential equation and how do these models relate to discrete-time dynamical systems? □

  A differential equation is an equation that describes derivative values of a function that is unknown to us. A first-order differential equation describes the instantaneous rate of change of a system, and so is the continuous counterpart of an updating function (or difference equation), which describes the average rate of change of a discrete-time dynamical system.

- **Question 4.1.6** What is a solution to a differential equation and how does this relate to the concept of an antiderivative? □

  A solution to a differential equation is any function whose derivative satisfies the differential equation. When we are working with a pure-time differential equation $\frac{dy}{dt} = f(t)$, a solution is an antiderivative of the function $f(t)$; that is, a solution is a function $y(t) = F(t)$ such that $F'(t) = f(t)$.

4.1.4 Exercises

1. Let $\frac{dq}{dt} = \frac{1}{t} + t^3 + -2^t$, where $t > 0$. Which family of functions below is the general solution to this differential equation? Explain how you know.

- $q(t) = \ln(t) + 0.25t^4 - \dfrac{2^t}{\ln(2)}$
- $q(t) = \ln(t) + 0.25t^4 - \dfrac{2^t}{\ln(2)} + C$
- $q(t) = -t^{-2} + 3t^2 - \ln(2) \cdot 2^t + C$

2. Let $\frac{dp}{dt} = \cos(t) + t^{\frac{3}{2}}$, where $t \geq 0$ and $p(0) = 1$. Which family of functions below is the specific solution to this initial value problem? Explain how you know.

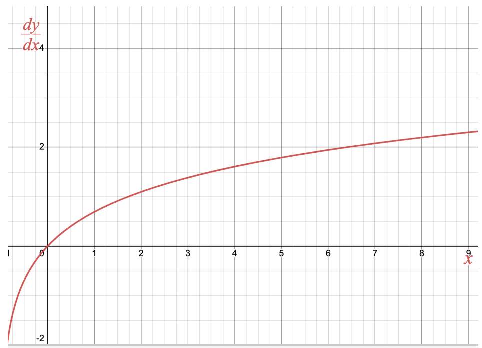
- $p(t) = \sin(t) + \dfrac{2}{5}t^{\frac{5}{2}}$
- $p(t) = \sin(t) + \dfrac{2}{5}t^{\frac{5}{2}} + 1$
- $p(t) = -\sin(t) + \dfrac{3}{2}t^{\frac{1}{2}} + 1$

3. Suppose $\frac{dr}{dt} = \ell(t)$, where $\ell(t)$ is a linear function. Write one thoughtful sentence explaining how you know that any solution $r(t)$ must be a quadratic function.

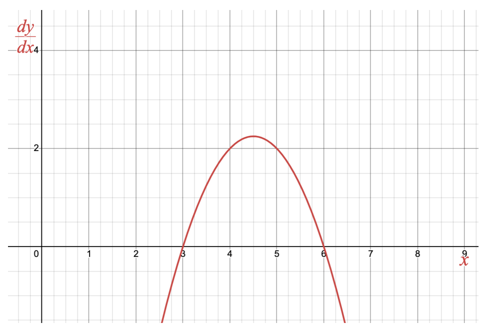
4. For each graph of a differential equation $\frac{dy}{dx}$ below, sketch one possible solution function on the same coordinate system. Identify an initial value of the solution function that you drew.

(a)

(b)



(c)



4.2 Solving Pure-Time Differential Equations

Motivating Questions

- How can we solve pure-time differential equations that are sums of basic functions?
- How can we solve more complicated pure-time differential equations?

We have seen that finding the family of solutions to a pure-time differential equation $\frac{dy}{dt} = f(t)$ amounts to answering the following question: which functions have a derivative equal to $f(t)$? In other words, we must find the *family of antiderivatives* of $f(t)$. In Section 4.1 we focused on verifying whether functions were solutions and sketching graphs of solutions to differential equations that were represented graphically. In this section, we will gain methods for computing a family of antiderivatives. We first give some common vocabulary and notation for representing a family of antiderivatives.

Indefinite Integral.

Given a function $f(x)$, the **indefinite integral** of f is the family of antiderivatives of f, written as

$$\int f(x)dx.$$

In Example 4.1.4 we saw that the family of antiderivatives of $4t + 1$ was $2t^2 + t + C$. Equivalently, we could write this as

$$\int (4t + 1)dt = 2t^2 + t + C.$$

When we find an antiderivative, we will often say that we *evaluate an indefinite integral*. Just as the notation $\frac{d}{dx}[\square]$ means "take the derivative of $\square$ with respect to the variable x", the notation $\int \square dx$ means "find a function of x whose derivative is $\square$." It is important to pay attention to the notation to know what letter is considered as the independent variable (dx, dt, etc.).

Warm-Up 4.2.1 Use your knowledge of derivatives of basic functions to complete Table 4.2.1 of antiderivatives. For each entry, your task is to find a family of functions whose derivatives are the given function f.

Table 4.2.1 Familiar basic functions and their antiderivatives.

$f(x)$	$\int f(x)dx$
k, (k is constant)	
x^2	
$x^{0.5}$	
x^n, $n \neq -1$	
$\frac{1}{x}$, $x \neq 0$	
$\sin(x)$	
$\cos(x)$	
e^x	
2^x	
b^x $(b > 0, b \neq 1)$	

4.2.1 Basic Antiderivatives

Evaluating the indefinite integral of a single basic function involves reversing a single derivative rule. We summarize the basic antiderivative rules from Warm-Up 4.2.1 below.

Basic Antiderivative Rules.

1. (Reverse Power Rule) $\displaystyle\int x^n dx = \frac{1}{n+1}x^{n+1} + C$, for $n \neq -1$
2. $\int x^{-1}dx = \ln|x| + C$, for $x \neq 0$
3. (Reverse Exponential Rule) $\displaystyle\int b^x dx = \frac{1}{\ln(b)}b^x + C$, for $b > 0, b \neq 1$
4. (Reverse Trig Rules) $\displaystyle\int \sin(x)dx = -\cos(x) + C$ and $\displaystyle\int \cos(x)dx =$

$\sin(x) + C$

Similar to the constant multiple and sum rules for derivatives in Subsection 2.5.3, we have constant multiple and sum rules for antiderivatives that allow us to compute indefinite integrals for many functions using just these basic rules.

Constant Multiple Rule.

For any real number k,

$$\int kf(x)dx = k\int f(x)dx.$$

Sum Rule.

For two functions $f(x)$ and $g(x)$,

$$\int (f(x) + g(x))dx = \int f(x)dx + \int g(x)dx.$$

Example 4.2.2 Solving Basic Pure-Time Differential Equations. Let $\frac{ds}{dt} = e^t - 2\sin(t) + 4t^3$, $s(0) = 2$. To solve this initial value problem, we will first find the general solution by evaluating

$$\begin{aligned}\int (e^t - 2\sin(t) + 4t^3)dt &= \int e^t dt + \int -2\sin(t)dt + \int 4t^3 dt\\ &= \int e^t dt - 2\int \sin(t)dt + 4\int t^3 dt\\ &= e^t - 2(-\cos(t)) + 4\cdot\frac{1}{4}t^4 + C\\ &= e^t + 2\cos(t) + t^4 + C.\end{aligned}$$

[1] Now we use the initial value $s(0) = 2$ to find the specific solution:

$$2 = e^0 + 2\cos(0) + 0^4 + C = 1 + 2 + C,$$

which implies $C = -1$. Thus, the specific solution to this initial value problem is $s(t) = e^t + 2\cos(t) + t^4 - 1$. □

Activity 4.2.2 Solve each initial value problem below. Practice using correct notation for evaluating indefinite integrals.

1. $\frac{dr}{dt} = t^{-1} - 3\sqrt{t}$, $r(1) = 5$
2. $\frac{dp}{dt} = \cos(t) + 2\cdot 1.5^t$, $p(0) = \ln(1.5)$
3. $\frac{dy}{dt} = e^{0.5t}$, $y(0) = 0$

[1] Though technically each individual integral has its own constant, in the end we can add them all together to a single constant. In practice, we typically just add one constant at the very end as we've done in this example.

4.2.2 u-Substitution

The initial value problem in Activity 4.2.2 part 3 is slightly different in nature than the others, and required one additional small step for computing the general solution. A good first guess would be that $\int e^{0.5t} dt = e^{0.5t} + C$. However, we can see that this is incorrect by computing that $\frac{d}{dt}[e^{0.5t} + C] = 0.5e^{0.5t}$ (using the Chain Rule), which is not the same function that we started with under the integral. However, since it is only off by a constant multiple, we can make an easy correction to our original guess by dividing by this constant multiple to get

$$\int e^{0.5t} dt = \frac{1}{0.5} \cdot e^{0.5t} + C = 2e^{0.5t} + C.$$

The general reason this problem was more difficult was because we were integrating a composition: if $f(t) = e^t$ and $g(t) = 0.5t$, then $e^{0.5t} = f(g(t))$. The general reason our correction worked is that the "inside function" $g(t) = 0.5t$ is *linear*, and so we just needed to divide by its derivative (slope). This is a correction that will work every time in this context:

> If $\int f(x)dx = F(x) + C$, then
>
> $$\int f(ax + b)dx = \frac{1}{a} \cdot F(ax + b) + C.$$

Activity 4.2.3 Compute each indefinite integral below.

1. $\displaystyle\int e^{-t} dt$
2. $\displaystyle\int (5t - 11)^{10} dt$
3. $\displaystyle\int \frac{1}{2x - 1} dx$
4. $\displaystyle\int \cos(-0.1\theta + 4) d\theta$

How can we compute more complicated integrals like this when the inside function is not linear? For example, how might we compute $\int xe^{x^2} dx$? Similar to the case when the inside function is linear, evaluating integrals like this is related to the chain rule for taking derivatives.

Recall that the Chain Rule states that

$$\frac{d}{dx}[f(g(x))] = f'(g(x) \cdot g'(x).$$

Restating this in terms of antiderivatives this says that

$$\int f'(g(x) \cdot g'(x)dx = f(g(x)) + C. \tag{4.2.1}$$

Equation (4.2.1) tells us that if we can view a given function as $f'(g(x))g'(x)$ for some appropriate choices of f and g, then we can antidifferentiate the function by reversing the Chain Rule. Note that both $g(x)$ and $g'(x)$ appear in the form of $f'(g(x))g'(x)$; we will sometimes say that we seek to *identify a function-derivative pair* ($g(x)$ and $g'(x)$) when trying to apply the rule in Equation (4.2.1).

If we can identify a function-derivative pair, we will introduce a new variable u to represent the function $g(x)$. With $u = g(x)$, it follows that $\frac{du}{dx} = g'(x)$, so that in terms of differentials[2], $du = g'(x)\,dx$. Now converting the indefinite integral to a new one in terms of u, we have

$$\int f'(g(x))g'(x)\,dx = \int f'(u)\,du.$$

Provided that f' is an elementary function whose antiderivative is known, we can easily evaluate the indefinite integral in u, and then go on to determine the desired overall antiderivative of $f'(g(x))g'(x)$. We call this process *u-substitution*, and summarize the rule as follows:

With the substitution $u = g(x)$,

$$\int f'(g(x))g'(x)\,dx = \int f'(u)\,du = f(u) + C = f(g(x)) + C.$$

To see u-substitution at work, we consider the following example.

Example 4.2.3 Evaluate the indefinite integral

$$\int x^3 \cdot \sin(7x^4 + 3)\,dx$$

and check the result by differentiating.

Solution. We can make two algebraic observations regarding the integrand, $x^3 \cdot \sin(7x^4 + 3)$. First, $\sin(7x^4 + 3)$ is a composite function; as such, we know we'll need a more sophisticated approach to antidifferentiating. Second, x^3 is almost the derivative of $(7x^4 + 3)$; the only issue is a missing constant. Thus, x^3 and $(7x^4 + 3)$ are nearly a function-derivative pair. Furthermore, we know the antiderivative of $f(u) = \sin(u)$. The combination of these observations suggests that we can evaluate the given indefinite integral by reversing the chain rule through u-substitution.

Letting u represent the inner function of the composite function $\sin(7x^4+3)$, we have $u = 7x^4 + 3$, and thus $\frac{du}{dx} = 28x^3$. In differential notation, it follows that $du = 28x^3\,dx$, and thus $x^3\,dx = \frac{1}{28}\,du$. The original indefinite integral may be slightly rewritten as

$$\int \sin(7x^4 + 3) \cdot x^3\,dx,$$

and so by substituting u for $7x^4 + 3$ and $\frac{1}{28}\,du$ for $x^3\,dx$, it follows that

$$\int \sin(7x^4 + 3) \cdot x^3\,dx = \int \sin(u) \cdot \frac{1}{28}\,du.$$

Now we may evaluate the easier integral in u, and then replace u by the expression $7x^4 + 3$. Doing so, we find

$$\begin{aligned}\int \sin(7x^4 + 3) \cdot x^3\,dx &= \int \sin(u) \cdot \frac{1}{28}\,du \\ &= \frac{1}{28}\int \sin(u)\,du\end{aligned}$$

[2] If we recall from the definition of the derivative that $\frac{du}{dx} \approx \frac{\Delta u}{\Delta x}$ and use the fact that $\frac{du}{dx} = g'(x)$, then we see that $g'(x) \approx \frac{\Delta u}{\Delta x}$. Solving for Δu, $\Delta u \approx g'(x)\Delta x$. It is this last relationship that, when expressed in "differential" notation enables us to write $du = g'(x)\,dx$ in the change of variable formula.

$$= \frac{1}{28}(-\cos(u)) + C$$
$$= -\frac{1}{28}\cos(7x^4 + 3) + C.$$

To check our work, we observe by the Chain Rule that

$$\frac{d}{dx}\left[-\frac{1}{28}\cos(7x^4+3)\right] = -\frac{1}{28}\cdot(-1)\sin(7x^4+3)\cdot 28x^3 = \sin(7x^4+3)\cdot x^3,$$

which is indeed the original integrand. □

The u-substitution worked because the function multiplying $\sin(7x^4 + 3)$ was x^3. If instead that function was x^2 or x^4, the substitution process would not have worked. This is one of the primary challenges of antidifferentiation: slight changes in the integrand make tremendous differences. For instance, we can use u-substitution with $u = x^2$ and $du = 2xdx$ to find that

$$\int xe^{x^2}\,dx = \int e^u \cdot \frac{1}{2}\,du$$
$$= \frac{1}{2}\int e^u\,du$$
$$= \frac{1}{2}e^u + C$$
$$= \frac{1}{2}e^{x^2} + C.$$

However, for the similar indefinite integral

$$\int e^{x^2}\,dx,$$

the u-substitution $u = x^2$ is no longer possible because the factor of x is missing. Hence, part of the lesson of u-substitution is just how specialized the process is: it only applies to situations where, up to a missing constant, the integrand is the result of applying the Chain Rule to a different, related function.

Activity 4.2.4 Evaluate each of the following indefinite integrals by using these steps:

- Find two functions within the integrand that form (up to a possible missing constant) a function-derivative pair;
- Make a substitution and convert the integral to one involving u and du;
- Evaluate the new integral in u;
- Convert the resulting function of u back to a function of x by using your earlier substitution;
- Check your work by differentiating the function of x. You should come up with the integrand originally given.

a. $\displaystyle\int \frac{x^2}{5x^3+1}\,dx$

b. $\displaystyle\int e^x \sin(e^x)\,dx$

c. $\displaystyle\int \frac{\cos(\sqrt{x})}{\sqrt{x}}\,dx$

4.2.3 Summary

- **Question 4.2.4** How can we solve pure-time differential equations that are sums of basic functions? ☐

  To find the general solution of a pure-time differential equation $\frac{dy}{dt} = f(t)$, we need to find the family of antiderivatives $f(t)$, $\int f(t)dt$. When $f(t)$ is a single basic function, we can reverse the appropriate rule using the basic antiderivative rules. When $f(t)$ is made up of a sum and/or constant multiple of these basic functions, we can break it up using the constant multiple rule and the sum rule.

- **Question 4.2.5** How can we solve more complicated pure-time differential equations? ☐

  For pure-time differential equations that are defined with an appropriate "function-derivative pair", we can use the method of u-substitution to find the general solution.

4.2.4 Exercises

1. Given that $s''(t) = \sin(t) + t$, $s'(0) = -2$, and $s(0) = 5$, determine the equation for $s(t)$.

2. Given that $\int g(x)dx = x^2 + 4x + C$, evaluate each indefinite integral below.

(a) $\displaystyle\int xg(x^2)dx$

(b) $\displaystyle\int \frac{g(\ln(x))}{x}dx$

3. Evaluate $\int \tan(x)dx$ using u-substitution. (Hint: start by writing $\tan(x)$ as $\frac{\sin(x)}{\cos(x)}$)

4. For the town of Calculus, CO, residential carbon emissions have shown certain trends over recent years. Based on data reflecting average emission rates, the town's environmental agency has modeled emission rates as

$$\frac{dA}{dt} = 4 + \sin(0.263t + 4.7) + \cos(0.526t + 9.4).$$

Here, t measures time in hours after midnight on a typical weekday, and $\frac{dA}{dt}$ is the rate of carbon emissions in pounds per hour.

a. Use technology to sketch a graph of $\frac{dA}{dt}$ on the interval [0,24] and explain its meaning. Why is this a reasonable model of carbon emissions?

b. Determine the general solution of this differential equation. What are the units of $A(t)$?

4.3 Riemann Sums

Motivating Questions

- How does the area between a velocity curve and the horizontal time axis over a particular interval give us information about position?
- How can we use a Riemann sum to estimate the area between a given curve and the horizontal axis over a particular interval?
- What are the differences among left, right, and middle Riemann sums?
- How can we write Riemann sums in an abbreviated form?

In Chapter 4, a general goal has been to describe how a continuous-time dynamical system is changing using a differential equation that models that system. In this section, we will see there is a connection between describing change in position and computing area under a velocity curve. We will then begin discussing how we go about computing areas under curves.

Warm-Up 4.3.1 Let $s(t)$ be the position of an object (in feet) relative to its starting position after t seconds. Suppose we know $s(t)$ is changing according to the differential equation $\frac{ds}{dt} = 5$.

1. What is a common name for the quantity $\frac{ds}{dt}$? What are the units of $\frac{ds}{dt}$?
2. Compute the distance the object has traveled after 10 seconds.
3. Sketch a graph of the differential equation: $\frac{ds}{dt}$ on the vertical axis and t on the horizontal axis.
4. Compute the area under the $\frac{ds}{dt}$ curve and above the t-axis on the interval $[0, 10]$. How does this area relate to your answer in part 2? Write one careful sentence explaining why you think your answers relate this way.
5. Suppose instead the differential equation describing s is $\frac{ds}{dt} = -5$. How does this change your answers to the previous questions? What is the physical difference between these two differential equations?

4.3.1 Areas, Distance, and Displacement

In Warm-Up 4.3.1, we learned that if an object moves with positive velocity $v = \frac{ds}{dt}$, the area between $y = v(t)$ and the t-axis over a given time interval tells us the distance traveled by the object over that time period. If $v(t)$ is sometimes negative and we view the area of any region below the t-axis as having an associated negative sign, then the sum of these signed areas tells us the moving object's *change in position*, or **displacement** over a given time interval.

For instance, for the velocity function given in Figure 4.3.1, if the areas of shaded regions are A_1, A_2, and A_3 as labeled, then the total distance D traveled by the moving object on $[a, b]$ is

$$D = A_1 + A_2 + A_3,$$

while the total change in the object's position (or displacement) on $[a, b]$ is

$$s(b) - s(a) = A_1 - A_2 + A_3.$$

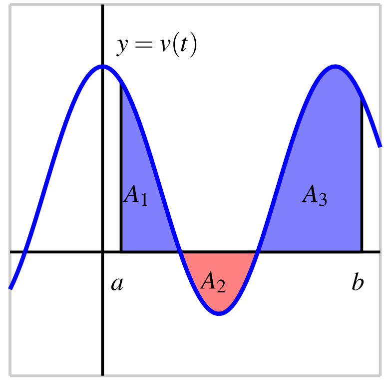


Figure 4.3.1 A velocity function that is sometimes negative.

Because the motion is in the negative direction on the interval where $v(t) < 0$, we subtract A_2 to determine the object's total change in position.

Of course, finding D and $s(b) - s(a)$ for the graph in Figure 4.3.1 presumes that we can actually find the areas A_1, A_2, and A_3. So far, we have worked with velocity functions that were constant, so that the area bounded by the velocity function and the horizontal axis is the area of a rectangle, and we can find the area exactly. If the velocity function were any linear function, we could similarly compute areas of triangles to find the area exactly. But when the curve bounds a region that is not a familiar geometric shape, we cannot find its area exactly. Indeed, this is one of our biggest goals in Chapter 4: to learn how to find the exact area bounded between a curve and the horizontal axis for as many different types of functions as possible.

In Activity 4.3.2, we will see how we might approximate the area under a nonlinear velocity function using rectangles.

Activity 4.3.2 A person walking along a straight path has her velocity in miles per hour at time t given by the function $v(t) = 0.25t^3 - 1.5t^2 + 3t + 0.25$, for times in the interval $0 \leq t \leq 2$. The graph of this function is also given in each of the three diagrams in Figure 4.3.2.

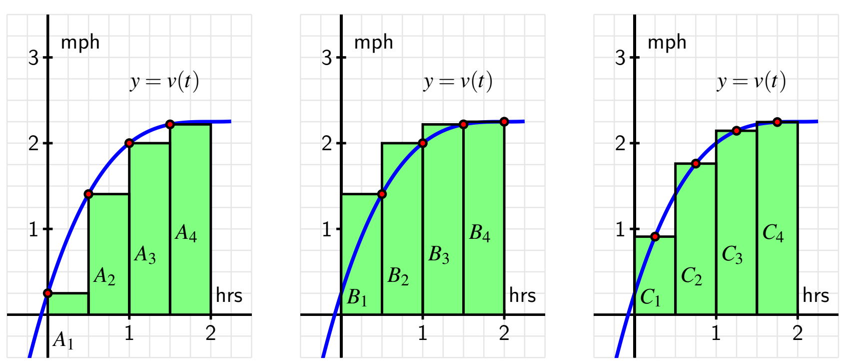


Figure 4.3.2 Three approaches to estimating the area under $y = v(t)$ on the interval $[0, 2]$.

Note that in each diagram, we use four rectangles to estimate the area under $y = v(t)$ on the interval $[0, 2]$, but the method by which the four rectangles' respective heights are decided varies among the three individual graphs.

a. How are the heights of rectangles in the left-most diagram being chosen? Explain, and hence determine the value of

$$S = A_1 + A_2 + A_3 + A_4$$

by evaluating the function $y = v(t)$ at appropriately chosen values and observing the width of each rectangle. Note, for example, that

$$A_3 = v(1) \cdot \frac{1}{2} = 2 \cdot \frac{1}{2} = 1.$$

b. Explain how the heights of rectangles are being chosen in the middle diagram and find the value of

$$T = B_1 + B_2 + B_3 + B_4.$$

c. Likewise, determine the pattern of how heights of rectangles are chosen in the right-most diagram and determine

$$U = C_1 + C_2 + C_3 + C_4.$$

d. Let D be the total distance the person traveled on $[0, 2]$. Of the estimates S, T, and U,

- which do you think is clearly an over-estimate of D?
- which do you think is clearly an under-estimate of D?
- which do you think is the best approximation of D?

4.3.2 Sigma Notation

We have used sums of areas of rectangles to approximate the area under a curve. Intuitively, we expect that using a larger number of thinner rectangles will provide a better estimate for the area. Consequently, we anticipate dealing with sums of a large number of terms. To do so, we introduce *sigma notation*, named for the Greek letter Σ, which is the capital letter S in the Greek alphabet.

For example, say we are interested in the sum

$$1 + 2 + 3 + \cdots + 100,$$

the sum of the first 100 natural numbers. In sigma notation we write

$$\sum_{k=1}^{100} k = 1 + 2 + 3 + \cdots + 100.$$

We read the symbol $\sum_{k=1}^{100} k$ as "the sum from k equals 1 to 100 of k." The variable k is called the index of summation, and any letter can be used for this variable. The pattern in the terms of the sum is denoted by a function of the index; for example,

$$\sum_{k=1}^{10} (k^2 + 2k) = (1^2 + 2 \cdot 1) + (2^2 + 2 \cdot 2) + (3^2 + 2 \cdot 3) + \cdots + (10^2 + 2 \cdot 10),$$

and more generally,

$$\sum_{k=1}^{n} f(k) = f(1) + f(2) + \cdots + f(n).$$

Sigma notation allows us to vary easily the function being used to describe the terms in the sum, and to adjust the number of terms in the sum simply by changing the value of n. We test our understanding of this new notation in the following activity.

Activity 4.3.3 For each sum written in sigma notation, write the sum longhand and evaluate the sum to find its value. For each sum written in expanded form, write the sum in sigma notation.

a. $\sum_{k=1}^{5}(k^2+2)$

b. $\sum_{i=3}^{6}(2i-1)$

c. $3+7+11+15+\cdots+27$

d. $4+8+16+32+\cdots+256$

e. $\sum_{i=1}^{6}\frac{1}{2^i}$

4.3.3 Riemann Sums

When a moving body has a positive velocity function $y = v(t)$ on a given interval $[a, b]$, the area under the curve over the interval gives the total distance the body travels on $[a, b]$. We are also interested in finding the exact area bounded by $y = f(x)$ on an interval $[a, b]$, regardless of the meaning or context of the function f. For now, we continue to focus on finding an accurate estimate of this area by using a sum of the areas of rectangles. Unless otherwise indicated, we assume that f is continuous and non-negative on $[a, b]$.

The first choice we make in such an approximation is the number of rectangles.

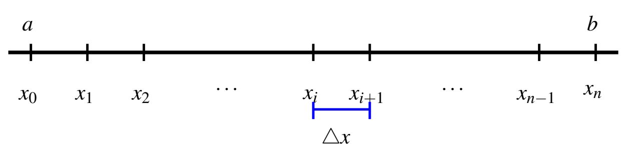


Figure 4.3.3 Subdividing the interval $[a, b]$ into n subintervals of equal length Δx.

If we desire n rectangles of equal width to subdivide the interval $[a, b]$, then each rectangle must have width $\Delta x = \frac{b-a}{n}$. We let $x_0 = a$, $x_n = b$, and define $x_i = a + i\Delta x$, so that $x_1 = x_0 + \Delta x$, $x_2 = x_0 + 2\Delta x$, and so on, as pictured in Figure 4.3.3.

We use each subinterval $[x_i, x_{i+1}]$ as the base of a rectangle, and next choose the height of the rectangle on that subinterval. There are three standard choices: we can use the left endpoint of each subinterval, the right endpoint of each subinterval, or the midpoint of each. These are precisely the options encountered in Activity 4.3.2 and seen in Figure 4.3.2. We next explore how these choices can be described in sigma notation.

Consider an arbitrary positive function f on $[a, b]$ with the interval subdivided as shown in Figure 4.3.3, and choose to use left endpoints. Then on each interval $[x_i, x_{i+1}]$, the area of the rectangle formed is given by

$$A_{i+1} = f(x_i) \cdot \Delta x,$$

as seen in Figure 4.3.4.

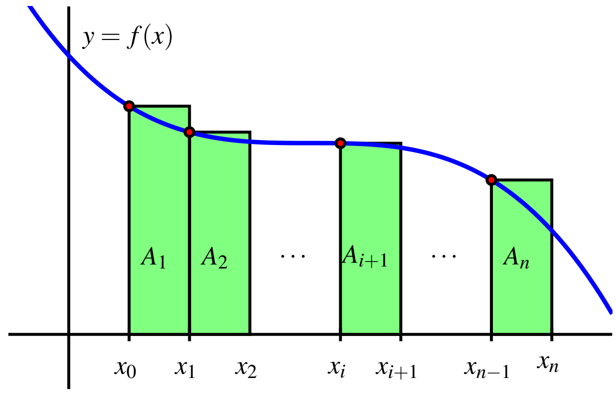


Figure 4.3.4 Subdividing the interval $[a, b]$ into n subintervals of equal length Δx and approximating the area under $y = f(x)$ over $[a, b]$ using left rectangles.

If we let L_n denote the sum of the areas of these rectangles, we see that

$$\begin{aligned} L_n &= A_1 + A_2 + \cdots + A_{i+1} + \cdots + A_n \\ &= f(x_0) \cdot \Delta x + f(x_1) \cdot \Delta x + \cdots + f(x_i) \cdot \Delta x + \cdots + f(x_{n-1}) \cdot \Delta x. \end{aligned}$$

In the more compact sigma notation, we have

$$L_n = \sum_{i=0}^{n-1} f(x_i)\Delta x.$$

Note that since the index of summation begins at 0 and ends at $n-1$, there are indeed n terms in this sum. We call L_n the **left Riemann sum** for the function f on the interval $[a, b]$ using n rectangles.

To see how the Riemann sums for right endpoints and midpoints are constructed, we consider Figure 4.3.5.

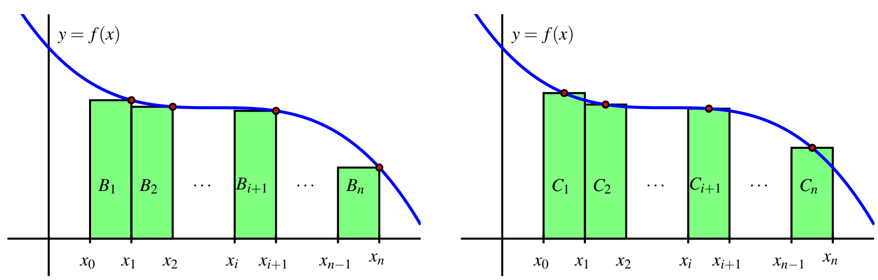


Figure 4.3.5 Riemann sums using right endpoints and midpoints.

For the sum with right endpoints, we see that the area of the rectangle on an arbitrary interval $[x_i, x_{i+1}]$ is given by $B_{i+1} = f(x_{i+1}) \cdot \Delta x$, and that the sum of all such areas of rectangles is given by

$$\begin{aligned} R_n &= B_1 + B_2 + \cdots + B_{i+1} + \cdots + B_n \\ &= f(x_1) \cdot \Delta x + f(x_2) \cdot \Delta x + \cdots + f(x_{i+1}) \cdot \Delta x + \cdots + f(x_n) \cdot \Delta x \\ &= \sum_{i=1}^{n} f(x_i)\Delta x. \end{aligned}$$

We call R_n the **right Riemann sum** for the function f on the interval $[a, b]$ using n rectangles.

For the sum that uses midpoints, we introduce the notation

$$\overline{x}_{i+1} = \frac{x_i + x_{i+1}}{2}$$

so that $\overline{x}_{i+1}$ is the midpoint of the interval $[x_i, x_{i+1}]$. For instance, for the rectangle with area C_1 in Figure 4.3.5, we now have

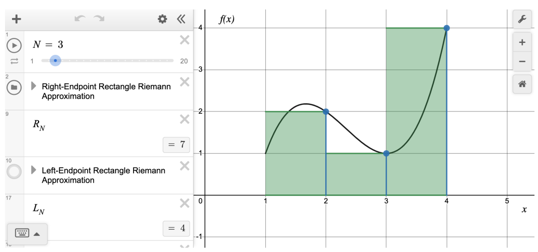
$$C_1 = f(\overline{x}_1) \cdot \Delta x.$$

Hence, the sum of all the areas of rectangles that use midpoints is

$$\begin{aligned} M_n &= C_1 + C_2 + \cdots + C_{i+1} + \cdots + C_n \\ &= f(\overline{x_1}) \cdot \Delta x + f(\overline{x_2}) \cdot \Delta x + \cdots + f(\overline{x}_{i+1}) \cdot \Delta x + \cdots + f(\overline{x}_n) \cdot \Delta x \\ &= \sum_{i=1}^{n} f(\overline{x}_i) \Delta x, \end{aligned}$$

and we say that M_n is the **middle Riemann sum** for f on $[a, b]$ using n rectangles.

Thus, we have two variables to explore: the number of rectangles and the height of each rectangle. We can explore these choices dynamically using the interactive provided below.



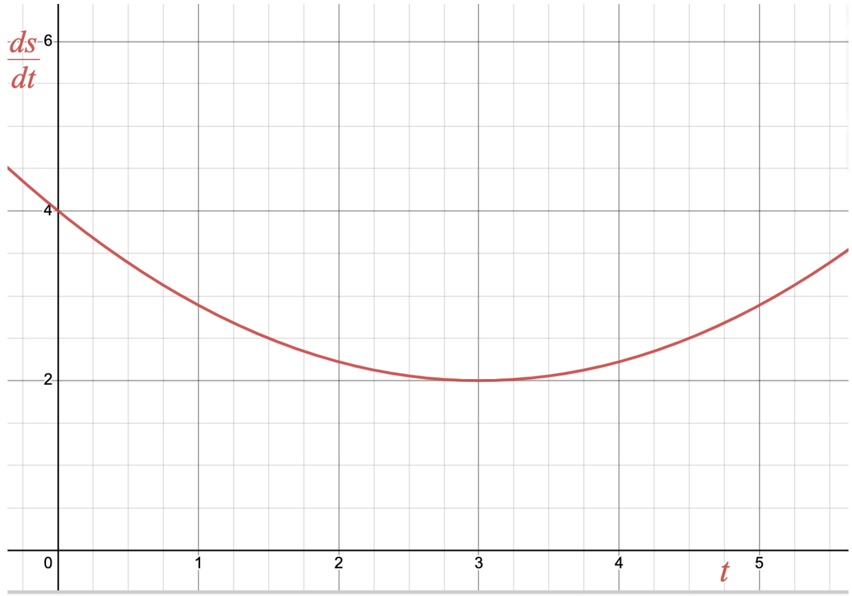
www.desmos.com/calculator/gb82c6uzmo

By moving the slider for N, we can see how the estimates change using more or less rectangles. By highlighting different types of estimates, we can see how the heights of the rectangles change as we consider left endpoints, midpoints, and right endpoints. As we change both of these variables, we can compare the estimate values to the actual value of the area under the curve.

When $f(x) \geq 0$ on $[a, b]$, each of the Riemann sums L_n, R_n, and M_n provides an estimate of the area under the curve $y = f(x)$ over the interval $[a, b]$. We also recall that in the context of a non-negative velocity function $y = v(t) = \frac{ds}{dt}$, the corresponding Riemann sums approximate the distance traveled on $[a, b]$ by a moving object with velocity function v.

Activity 4.3.4 Suppose that an object moving along a straight line path has its velocity in feet per second at time t in seconds given by $\frac{ds}{dt} = \frac{2}{9}(t - 3)^2 + 2$.

a. Use the graph below to illustrate the exact area that will tell you the value of the distance the object traveled on the time interval $2 \leq t \leq 5$.



b. Estimate the distance traveled on $[2, 5]$ by computing L_4, R_4, and M_4.

c. For this question, think about an arbitrary function f, rather than the particular function given above. If f is positive and increasing on $[a, b]$, will L_n over-estimate or under-estimate the exact area under f on $[a, b]$? Will R_n over- or under-estimate the exact area under f on $[a, b]$? Explain.

4.3.4 When the function is sometimes negative

For a Riemann sum such as

$$L_n = \sum_{i=0}^{n-1} f(x_i)\Delta x,$$

we can of course compute the sum even when f takes on negative values. We know that when f is positive on $[a, b]$, a Riemann sum estimates the area bounded between f and the horizontal axis over the interval.

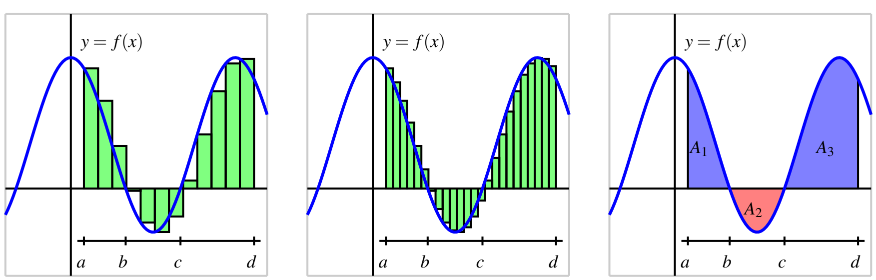


Figure 4.3.6 At left and center, two left Riemann sums for a function f that is sometimes negative; at right, the areas bounded by f on the interval $[a, d]$.

For the function pictured in the first graph of Figure 4.3.6, a left Riemann sum with 12 subintervals over $[a, d]$ is shown. The function is negative on the interval $b \leq x \leq c$, so at the four left endpoints that fall in $[b, c]$, the terms $f(x_i)\Delta x$ are negative. This means that those four terms in the Riemann sum produce an estimate of the *opposite* of the area bounded by $y = f(x)$ and the x-axis on $[b, c]$.

In the middle graph of Figure 4.3.6, we see that by increasing the number of rectangles the approximation of the area (or the opposite of the area) bounded by the curve appears to improve.

In general, any Riemann sum of a continuous function f on an interval $[a, b]$ approximates the difference between the area that lies above the horizontal axis on $[a, b]$ and under f and the area that lies below the horizontal axis on $[a, b]$ and above f. In the notation of Figure 4.3.6, we may say that

$$L_{24} \approx A_1 - A_2 + A_3,$$

where L_{24} is the left Riemann sum using 24 subintervals shown in the middle graph. A_1 and A_3 are the areas of the regions where f is positive, and A_2 is the area where f is negative. We will call the quantity $A_1 - A_2 + A_3$ the *net signed area* bounded by f over the interval $[a, d]$, where by the phrase "signed area" we indicate that we are attaching a minus sign to the areas of regions that fall below the horizontal axis.

Finally, we recall that if the function f represents the velocity of a moving object, the sum of the areas bounded by the curve tells us the total distance traveled over the relevant time interval, while the net signed area bounded by the curve computes the object's change in position, or displacement, on the interval.

Activity 4.3.5 Suppose that an object moving along a straight line path has its velocity $\frac{ds}{dt}$ (in feet per second) at time t (in seconds) given by

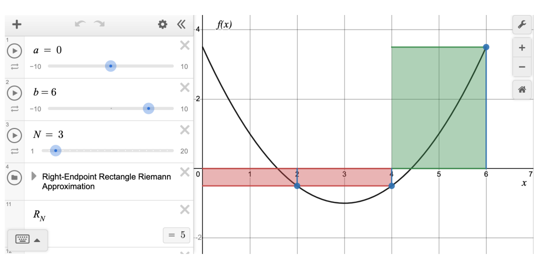
$$\frac{ds}{dt} = \frac{1}{2}t^2 - 3t + \frac{7}{2}.$$

a. Compute R_4, the rigth Riemann sum, with 4 rectangles for $\frac{ds}{dt}$ on the time interval $[1, 5]$. Be sure to clearly identify the value of Δt as well as the locations of t_0, t_1, $\cdots$, t_4. Use the interactive provided to check your answer, as well as to help you with the remaining questions.

b. Building on your work in (a), estimate the displacement of the object on the interval $[1, 5]$.

c. Building on your work in (a) and (b), estimate the total distance traveled by the object on $[1, 5]$.

d. Use the interactive to make a guess as to the exact displacement and the exact total distance traveled by the object on the interval $[1, 5]$.



www.desmos.com/calculator/nd78ac8d7x

4.3.5 Summary

- **Question 4.3.7** How does the area between a velocity curve and the horizontal time axis over a particular interval give us information about position? □

  The (signed) area between a velocity curve and the horizontal time axis on a particular interval tells us the total change in position (displacement) of the object over that time interval. By treating all area as positive, we can also compute the total distance traveled by the object over that time interval.

- **Question 4.3.8** How can we use a Riemann sum to estimate the area between a given curve and the horizontal axis over a particular interval? □

  A Riemann sum is simply a sum of products of the form $f(x_i^*)\Delta x$ that estimates the area between a positive function and the horizontal axis over a given interval. If the function is sometimes negative on the interval, the Riemann sum estimates the difference between the areas that lie above the horizontal axis and those that lie below the axis.

- **Question 4.3.9** What are the differences among left, right, and middle Riemann sums? □

  The three most common types of Riemann sums are left, right, and middle sums, but we can also work with a more general Riemann sum. The only difference among these sums is the location of the point at which the function is evaluated to determine the height of the rectangle whose area is being computed. For a left Riemann sum, we evaluate the function at the left endpoint of each subinterval, while for right and middle sums, we use right endpoints and midpoints, respectively.

- **Question 4.3.10** How can we write Riemann sums in an abbreviated form? □

  The left, right, and middle Riemann sums are denoted L_n, R_n, and M_n, with formulas

$$L_n = f(x_0)\Delta x + f(x_1)\Delta x + \cdots + f(x_{n-1})\Delta x = \sum_{i=0}^{n-1} f(x_i)\Delta x,$$

$$R_n = f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x = \sum_{i=1}^{n} f(x_i)\Delta x,$$

$$M_n = f(\overline{x}_1)\Delta x + f(\overline{x}_2)\Delta x + \cdots + f(\overline{x}_n)\Delta x = \sum_{i=1}^{n} f(\overline{x}_i)\Delta x,$$

  where $x_0 = a$, $x_i = a + i\Delta x$, and $x_n = b$, using $\Delta x = \frac{b-a}{n}$. For the midpoint sum, $\overline{x}_i = (x_{i-1} + x_i)/2$.

4.3.6 Exercises

1. Consider the function $f(x) = 3x + 4$.

a. Compute L_4 and R_4 for $y = f(x)$ on the interval $[2, 5]$. Be sure to clearly identify the value of Δx, as well as the locations of $x_0, x_1, \ldots, x_4$. Include a careful sketch of the function and the corresponding rectangles being used in the sum.

b. Use your sketch from above to determine whether L_4 is an over or under-estimate. Do the same for R_4. Then use a familiar geometric formula to determine the exact value of the area of the region bounded by $y = f(x)$ and the x-axis on $[2, 5]$ to verify your answers.

2. Let S be the sum given by

$$S = ((1.4)^2+1)\cdot 0.4+((1.8)^2+1)\cdot 0.4+((2.2)^2+1)\cdot 0.4+((2.6)^2+1)\cdot 0.4+((3.0)^2+1)\cdot 0.4.$$

a. Assume that S is a right Riemann sum. For what function f and what interval $[a, b]$ is S this function's Riemann sum? Why?

b. How does your answer to (a) change if S is a left Riemann sum? a middle Riemann sum?

c. Suppose that S really is a right Riemann sum. What geometric quantity does S approximate?

d. Use sigma notation to write a new sum R that is the right Riemann sum for the same function, but that uses twice as many subintervals as S.

3. A car traveling along a straight road is braking and its velocity is measured at several different points in time, as given in the following table.

Table 4.3.11 Data for the braking car.

seconds, t	0	0.3	0.6	0.9	1.2	1.5	1.8
Velocity in ft/sec, $v(t)$	100	88	74	59	40	19	0

a. Plot the given data on a set of axes with time on the horizontal axis and the velocity on the vertical axis.

b. Estimate the total distance traveled during the time the car brakes using a middle Riemann sum with 3 subintervals.

c. Estimate the total distance traveled on $[0, 1.8]$ by computing L_6, R_6, and $\frac{1}{2}(L_6 + R_6)$.

d. Assuming that $v(t)$ is always decreasing on $[0, 1.8]$, which estimate from above do we know is an overestimate of the total distance traveled? Why?

4. The rate at which pollution escapes a scrubbing process at a manufacturing plant increases over time as filters and other technologies become less effective. For this particular example, assume that the rate of pollution (in tons per week) is given by the function r that is pictured in Figure 4.3.12.

a. Use the graph to estimate the value of M_4 on the interval $[0, 4]$.

b. What is the meaning of M_4 in terms of the pollution discharged by the plant? (HINT: keep track of the units in your computation of M_4)

c. Suppose that $r(t) = 0.5e^{0.5t}$. Use this formula for r to compute L_5 on $[0, 4]$.

d. Determine an over-estimate of the total amount of pollution that can escape the plant during the pictured four week time period using 4 rectangles. Explain how you know it is an over-estimate.

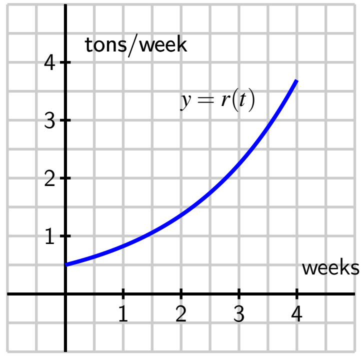


Figure 4.3.12 The rate, $r(t)$, of pollution in tons per week.

4.4 The Definite Integral

Motivating Questions

- How does increasing the number of subintervals affect the accuracy of the approximation generated by a Riemann sum?
- What is the definition of the definite integral of a function f over the interval $[a, b]$?
- What does the definite integral measure exactly, and what are some of the key properties of the definite integral?

In Figure 4.4.1, we see evidence that increasing the number of rectangles in a Riemann sum improves the accuracy of the approximation of the net signed area bounded by the given function.

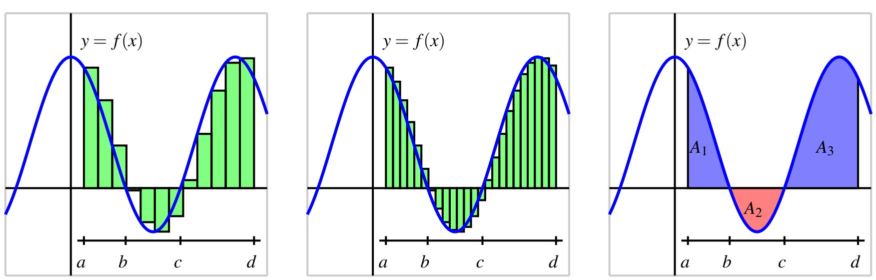
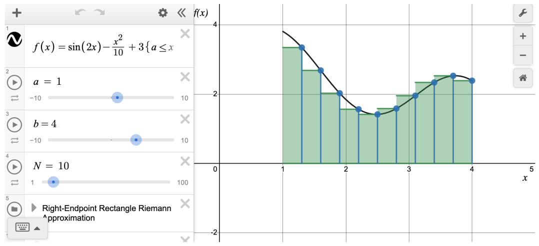


Figure 4.4.1 At left and center, two left Riemann sums for a function f that is sometimes negative; at right, the exact areas bounded by f on the interval $[a, d]$.

We therefore explore the natural idea of allowing the number of rectangles to increase without bound. In an effort to compute the exact net signed area we also consider the differences among left, right, and middle Riemann sums and the different results they generate as the value of n increases.

Warm-Up 4.4.1 We will use the following interactive:



www.desmos.com/calculator/xkfelsg0tj

The current function being graphed is $f(x) = \sin(2x) - \frac{x^2}{10} + 3$ on the domain $[1, 4]$, and we are estimating the signed area between the curve and x-axis using $R_{10} \approx 6.55$. Note that the value of the chosen Riemann sum is displayed right below each option, and that you can change the type of Riemann sum being computed by activating the appropriate folder.

Explore to see how you can change the domain that a sum is being computed over, the number of rectangles being used for the sum, as well as the function itself. You can also change the viewing window in Desmos as normal (zooming in/out with a mouse, or manually changing the settings with the wrench icon).

Work accordingly to adjust the interactive so that it uses a left Riemann sum with $n = 5$ subintervals for the function $f(x) = 2x + 1$ on the domain $[1, 4]$. You should see the updated figure shown in Figure 4.4.2. Then, answer the following questions.

a. Update the interactive (and view window, as needed) so that the function being considered is $f(x) = 2x + 1$ on $[1, 4]$, as directed above. For this function on this interval, compute L_n, M_n, R_n for $n = 5$, $n = 25$, and $n = 100$. What appears to be the exact area bounded by $f(x) = 2x + 1$ and the x-axis on $[1, 4]$?

b. Use basic geometry to determine the exact area bounded by $f(x) = 2x + 1$ and the x-axis on $[1, 4]$.

c. Based on your work in (a) and (b), what do you observe occurs when we increase the number of subintervals used in the Riemann sum?

d. Update the interactive to consider the function $f(x) = x^2 + 1$ on the interval $[1, 4]$ (note that you need to enter "`x ^ 2 + 1`" for the function formula). Use the interactive to compute L_n, M_n, R_n for $n = 5$, $n = 25$, and $n = 100$. What do you conjecture is the exact area bounded by $f(x) = x^2 + 1$ and the x-axis on $[1, 4]$?

e. Why can we not compute the exact value of the area bounded by $f(x) = x^2 + 1$ and the x-axis on $[1, 4]$ using a formula like we did in (b)?

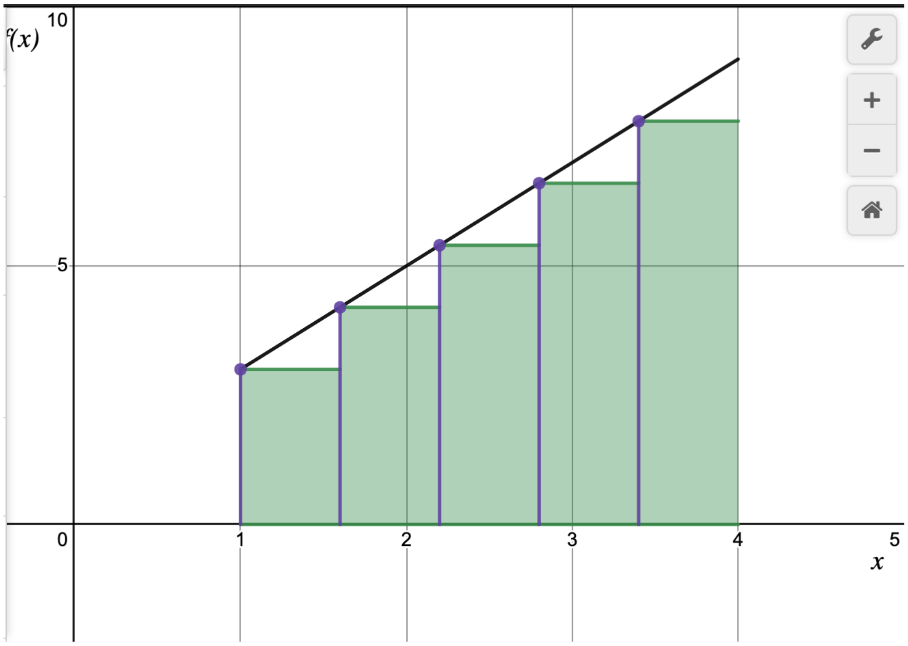


Figure 4.4.2 A left Riemann sum with 5 subintervals for the function $f(x) = 2x + 1$ on the interval $[1, 4]$ using a window with $-2 \leq y \leq 10$. The value of the sum is $L_5 = 16.2$.

4.4.1 The definition of the definite integral

In Warm-Up 4.4.1, we saw that as the number of rectangles got larger and larger, the values of L_n, M_n, and R_n all grew closer and closer to the same value. It turns out that this occurs for any continuous function on an interval $[a, b]$, and also for a Riemann sum using any point x_i^* in the interval $[x_i, x_{i+1}]$. Thus, as we let $n \to \infty$, it doesn't really matter where we choose to evaluate the function within a given subinterval, because

$$\lim_{n\to\infty} L_n = \lim_{n\to\infty} R_n = \lim_{n\to\infty} M_n = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x.$$

That these limits always exist (and share the same value) when f is continuous[1] allows us to make the following definition.

Definition 4.4.3 The **definite integral** of a continuous function f on the interval $[a, b]$, denoted $\int_a^b f(x)\,dx$, is the real number given by

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x,$$

where $\Delta x = \frac{b-a}{n}$, $x_i = a+i\Delta x$ (for $i = 0, \ldots, n$), and x_i^* satisfies $x_{i-1} \leq x_i^* \leq x_i$ (for $i = 1, \ldots, n$). ♢

We call the values a and b the **limits of integration** and the function f the **integrand**. The process of determining the real number $\int_a^b f(x)\,dx$ is called

[1] It turns out that a function need not be continuous in order to have a definite integral. For our purposes, we assume that the functions we consider are continuous on the interval(s) of interest. It is straightforward to see that any function that is piecewise continuous on an interval of interest will also have a well-defined definite integral.

evaluating the definite integral. While there are several different interpretations of the definite integral, for now the most important is that $\int_a^b f(x)\,dx$ measures the net signed area bounded by $y = f(x)$ and the x-axis on the interval $[a, b]$.

You may recognize the symbol $\int$ (called the *integral sign*,) from the definition of indefinite integrals. This notation suggests there is a connection between these concepts. Indeed, this connection is the main goal of Section 4.5.

For example, if f is the function pictured in Figure 4.4.4, and A_1, A_2, and A_3 are the exact areas bounded by f and the x-axis on the respective intervals $[a, b]$, $[b, c]$, and $[c, d]$, then

$$\int_a^b f(x)\,dx = A_1, \int_b^c f(x)\,dx = -A_2,$$
$$\int_c^d f(x)\,dx = A_3,$$
$$\text{and} \int_a^d f(x)\,dx = A_1 - A_2 + A_3.$$

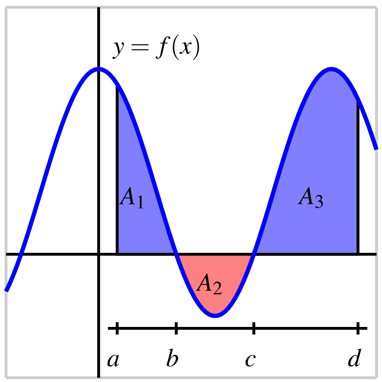


Figure 4.4.4 A continuous function f on the interval $[a, d]$.

We can also use definite integrals to express the change in position (displacement) and the distance traveled by a moving object. If v is a velocity function on an interval $[a, b]$, then the change in position of the object, $s(b) - s(a)$, is given by

$$s(b) - s(a) = \int_a^b v(t)\,dt.$$

If the velocity function is nonnegative on $[a, b]$, then $\int_a^b v(t)\,dt$ tells us the distance the object traveled. If the velocity is sometimes negative on $[a, b]$, we can use definite integrals to find the areas bounded by the function on each interval where v does not change sign, and the sum of these areas will tell us the distance the object traveled.

To compute the value of a definite integral from the definition, we have to take the limit of a sum. While this is possible to do in select circumstances, it is also tedious and time-consuming, and does not offer much additional insight into the meaning or interpretation of the definite integral. Instead, in Section 4.5, we will learn the Fundamental Theorem of Calculus, which provides a shortcut for evaluating a large class of definite integrals. This will enable us to determine the exact net signed area bounded by a continuous function and the x-axis in many circumstances.

For now, our goal is to understand the meaning and properties of the definite integral, rather than to compute its value. To do this, we will rely on the net signed area interpretation of the definite integral. So we will use as examples curves that produce regions whose areas we can compute exactly through area formulas. We can thus compute the exact value of the corresponding integral.

For instance, if we wish to evaluate the definite integral $\int_1^4 (2x+1)\,dx$, we observe that the region bounded by this function and the x-axis is the trapezoid shown in Figure 4.4.5. By the formula for the area of a trapezoid, $A = \frac{1}{2}(3+9)\cdot 3 = 18$, so

$$\int_1^4 (2x+1)\,dx = 18.$$

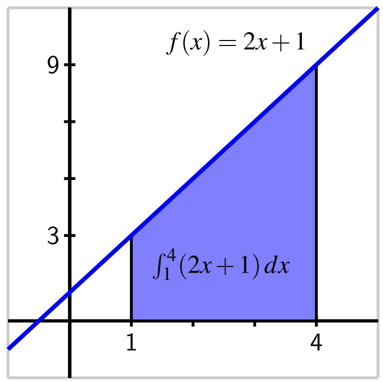


Figure 4.4.5 The area bounded by $f(x) = 2x+1$ and the x-axis on the interval $[1,4]$.

Activity 4.4.2 Use known geometric formulas and the net signed area interpretation of the definite integral to evaluate each of the definite integrals below.

a. $\displaystyle\int_0^1 3x\,dx$

b. $\displaystyle\int_{-1}^4 (2-2x)\,dx$

c. $\displaystyle\int_{-1}^1 \sqrt{1-x^2}\,dx$

d. $\int_{-3}^4 g(x)\,dx$, where g is the function pictured in Figure 4.4.6. Assume that each portion of g is either part of a line or part of a circle.

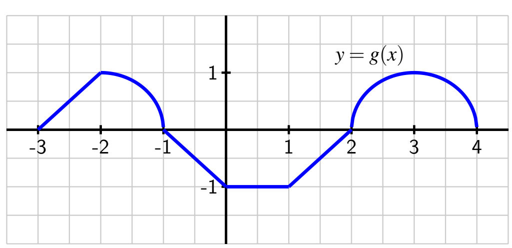


Figure 4.4.6 The function g for part (d). Note that g is piecewise defined, and each piece of the function is part of a circle or part of a line.

4.4.2 Some properties of the definite integral

Regarding the definite integral of a function f over an interval $[a, b]$ as the net signed area bounded by f and the x-axis, we discover several standard properties of the definite integral. It is helpful to remember that the definite integral is defined in terms of Riemann sums, which consist of the areas of rectangles.

For any real number a and the definite integral $\int_a^a f(x)\,dx$ it is evident that no area is enclosed, because the interval begins and ends with the same point. Hence,

> If f is a continuous function and a is a real number, then $\int_a^a f(x)\,dx = 0$.

Next, we consider the result of subdividing the interval of integration. In Figure 4.4.7, we see that

$$\int_a^b f(x)\,dx = A_1, \int_b^c f(x)\,dx = A_2,$$
$$\text{and } \int_a^c f(x)\,dx = A_1 + A_2,$$

which illustrates the following general rule.

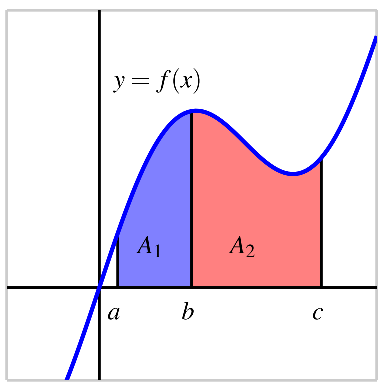


Figure 4.4.7 The area bounded by $y = f(x)$ on the interval $[a, c]$.

> If f is a continuous function and a, b, and c are real numbers, then
>
> $$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

While this rule is easy to see if $a < b < c$, it in fact holds in general for any values of a, b, and c. Another property of the definite integral states that if we reverse the order of the limits of integration, we change the sign of the integral's value.

If f is a continuous function and a and b are real numbers, then

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

This result makes sense because if we integrate from a to b, then in the defining Riemann sum we set $\Delta x = \frac{b-a}{n}$, while if we integrate from b to a, we have $\Delta x = \frac{a-b}{n} = -\frac{b-a}{n}$, and this is the only change in the sum used to define the integral.

There are two additional useful properties of the definite integral which mimic properties we have already seen in the context of derivatives and indefinite integrals. When we worked with derivative rules in Chapter 2, we formulated the Constant Multiple Rule and the Sum Rule. Recall that the Constant Multiple Rule says that if f is a differentiable function and k is a constant, then

$$\frac{d}{dx}[kf(x)] = kf'(x),$$

and the Sum Rule says that if f and g are differentiable functions, then

$$\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x).$$

These rules are useful because they allow us to deal individually with the simplest parts of certain functions by taking advantage of addition and multiplying by a constant. In other words, the process of taking the derivative respects addition and multiplying by constants in the simplest possible way.

It turns out that similar rules hold for the definite integral. First, let's consider the functions pictured in Figure 4.4.8.

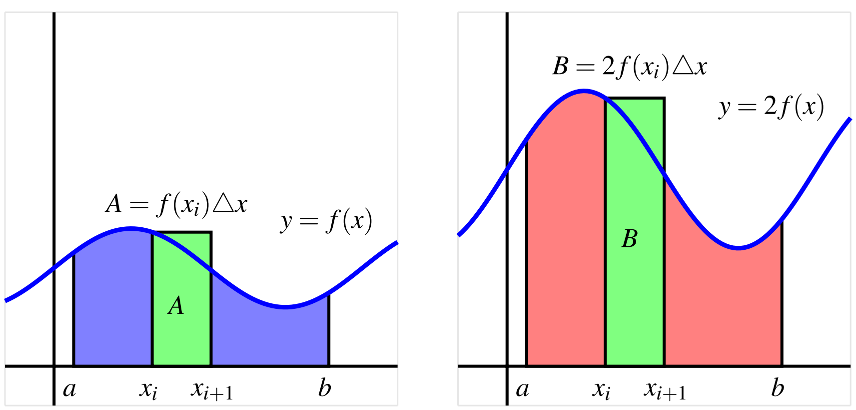


Figure 4.4.8 The areas bounded by $y = f(x)$ and $y = 2f(x)$ on $[a, b]$.

Because multiplying the function by 2 doubles its height at every x-value, we see that the height of each rectangle in a left Riemann sum is doubled, $f(x_i)$ for the original function, versus $2f(x_i)$ in the doubled function. For the areas A and B, it follows $B = 2A$. As this is true regardless of the value of n or the type of sum we use, we see that in the limit, the area of the red region bounded by $y = 2f(x)$ will be twice the area of the blue region bounded by $y = f(x)$. As there is nothing special about the value 2 compared to an arbitrary constant k, the following general principle holds.

Constant Multiple Rule.

If f is a continuous function and k is any real number, then

$$\int_a^b k \cdot f(x)\,dx = k \int_a^b f(x)\,dx.$$

We see a similar situation with the sum of two functions f and g.

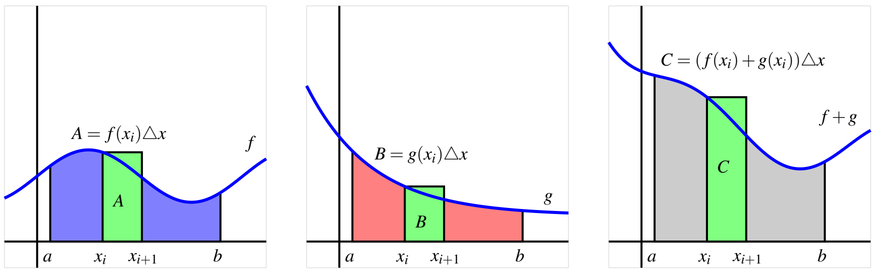


Figure 4.4.9 The areas bounded by $y = f(x)$ and $y = g(x)$ on $[a, b]$, as well as the area bounded by $y = f(x) + g(x)$.

If we take the sum of two functions f and g at every point in the interval, the height of the function $f + g$ is given by $(f + g)(x_i) = f(x_i) + g(x_i)$. Hence, for the pictured rectangles with areas A, B, and C, it follows that $C = A + B$. Because this will occur for every such rectangle, in the limit the area of the gray region will be the sum of the areas of the blue and red regions. In terms of definite integrals, we have the following general rule.

Sum Rule.

If f and g are continuous functions, then

$$\int_a^b [f(x) + g(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

The Constant Multiple and Sum Rules can be combined to say that for any continuous functions f and g and any constants c and k,

$$\int_a^b [cf(x) \pm kg(x)]\,dx = c \int_a^b f(x)\,dx \pm k \int_a^b g(x)\,dx.$$

Activity 4.4.3 Suppose that the following information is known about the functions f, g, x^2, and x^3:

- $\int_0^2 f(x)\,dx = -3$; $\int_2^5 f(x)\,dx = 2$
- $\int_0^2 g(x)\,dx = 4$; $\int_2^5 g(x)\,dx = -1$
- $\int_0^2 x^2\,dx = \frac{8}{3}$; $\int_2^5 x^2\,dx = \frac{117}{3}$
- $\int_0^2 x^3\,dx = 4$; $\int_2^5 x^3\,dx = \frac{609}{4}$

Use the provided information and the rules discussed in the preceding section to evaluate each of the following definite integrals.

a. $\int_5^2 f(x)\,dx$

b. $\int_0^5 g(x)\,dx$

c. $\int_0^5 (f(x)+g(x))\,dx$

d. $\int_2^5 (3x^2-4x^3)\,dx$

e. $\int_5^0 (2x^3-7g(x))\,dx$

4.4.3 How the definite integral is connected to a function's average value

One of the most valuable applications of the definite integral is that it provides a way to discuss the average value of a function, even for a function that takes on infinitely many values. Recall that if we wish to take the average of n numbers y_1, y_2, ..., y_n, we compute

$$AVG = \frac{y_1+y_2+\cdots+y_n}{n}.$$

Since integrals arise from Riemann sums in which we add n values of a function, it should not be surprising that evaluating an integral is similar to averaging the output values of a function. Consider, for instance, the right Riemann sum R_n of a function f, which is given by

$$R_n = f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x = (f(x_1)+f(x_2)+\cdots+f(x_n))\Delta x.$$

Since $\Delta x = \frac{b-a}{n}$, we can thus write

$$\begin{aligned} R_n &= (f(x_1)+f(x_2)+\cdots+f(x_n))\cdot\frac{b-a}{n} \\ &= (b-a)\frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{n}. \end{aligned} \tag{4.4.1}$$

We see that the right Riemann sum with n subintervals is just the length of the interval $(b-a)$ times the average of the n function values found at the right endpoints. And just as with our efforts to compute area, the larger the value of n we use, the more accurate our average will be. Indeed, we will define the average value of f on $[a,b]$ to be

$$f_{\text{AVG}[a,b]} = \lim_{n\to\infty}\frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{n}.$$

But we also know that for any continuous function f on $[a,b]$, taking the limit of a Riemann sum leads precisely to the definite integral. That is, $\lim_{n\to\infty} R_n = \int_a^b f(x)\,dx$, and thus taking the limit as $n\to\infty$ in Equation (4.4.1), we have that

$$\int_a^b f(x)\,dx = (b-a)\cdot f_{\text{AVG}[a,b]}. \tag{4.4.2}$$

Solving Equation (4.4.2) for $f_{\text{AVG}[a,b]}$, we have the following general principle.

Average value of a function.

If f is a continuous function on $[a,b]$, then its average value on $[a,b]$ is given by the formula

$$f_{\text{AVG}[a,b]} = \frac{1}{b-a} \cdot \int_a^b f(x)\,dx.$$

Equation (4.4.2) tells us another way to interpret the definite integral: the definite integral of a function f from a to b is the length of the interval $(b-a)$ times the average value of the function on the interval. In addition, when the function f is nonnegative on $[a, b]$, Equation (4.4.2) has a natural visual interpretation.

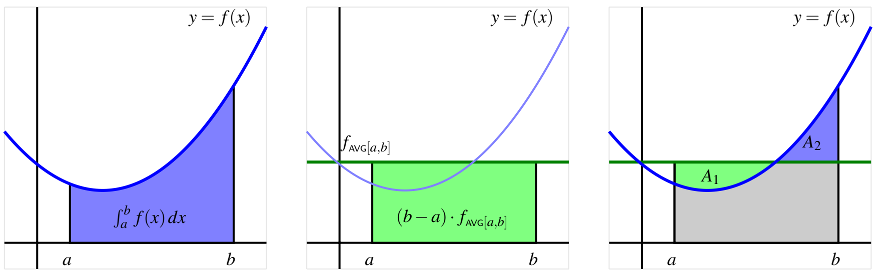


Figure 4.4.10 A function $y = f(x)$, the area it bounds, and its average value on $[a, b]$.

Consider Figure 4.4.10, where we see at left the shaded region whose area is $\int_a^b f(x)\,dx$, at center the shaded rectangle whose dimensions are $(b-a)$ by $f_{\text{AVG}[a,b]}$, and at right these two figures superimposed. Note that in dark green we show the horizontal line $y = f_{\text{AVG}[a,b]}$. Thus, the area of the green rectangle is given by $(b-a) \cdot f_{\text{AVG}[a,b]}$, which is precisely the value of $\int_a^b f(x)\,dx$. The area of the blue region in the left figure is the same as the area of the green rectangle in the center figure. We can also observe that the areas A_1 and A_2 in the rightmost figure appear to be equal. Thus, knowing the average value of a function enables us to construct a rectangle whose area is the same as the value of the definite integral of the function on the interval. This applet[2] provides an opportunity to explore how the average value of the function changes as the interval changes, through an image similar to that found in Figure 4.4.10.

Activity 4.4.4 Suppose that $v(t) = \sqrt{4-(t-2)^2}$ tells us the instantaneous velocity of a moving object on the interval $0 \le t \le 4$, where t is measured in minutes and v is measured in meters per minute.

a. Use technology to sketch an accurate graph of $y = v(t)$. What kind of curve is $y = \sqrt{4-(t-2)^2}$?

b. Evaluate $\int_0^4 v(t)\,dt$ exactly.

c. In terms of the physical problem of the moving object with velocity $v(t)$, what is the meaning of $\int_0^4 v(t)\,dt$? Include units on your answer.

d. Determine the exact average value of $v(t)$ on $[0, 4]$. Include units on your answer.

e. Sketch a rectangle whose base is the line segment from $t = 0$ to $t = 4$ on the t-axis such that the rectangle's area is equal to the value of $\int_0^4 v(t)\,dt$. What is the rectangle's exact height?

[2] gvsu.edu/s/az

f. How can you use the average value you found in (d) to compute the total distance traveled by the moving object over $[0, 4]$?

4.4.4 Summary

- **Question 4.4.11** How does increasing the number of subintervals affect the accuracy of the approximation generated by a Riemann sum? □

  Any Riemann sum of a continuous function f on an interval $[a, b]$ provides an estimate of the net signed area bounded by the function and the horizontal axis on the interval. Increasing the number of subintervals in the Riemann sum improves the accuracy of this estimate, and letting the number of subintervals increase without bound results in the values of the corresponding Riemann sums approaching the exact value of the enclosed net signed area.

- **Question 4.4.12** What is the definition of the definite integral of a function f over the interval $[a, b]$? □

  When we take the limit of Riemann sums, we arrive at what we call the definite integral of f over the interval $[a, b]$. In particular, the symbol $\int_a^b f(x)\,dx$ denotes the definite integral of f over $[a, b]$, and this quantity is defined by the equation

  $$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x,$$

  where $\Delta x = \frac{b-a}{n}$, $x_i = a + i\Delta x$ (for $i = 0, \ldots, n$), and x_i^* satisfies $x_{i-1} \le x_i^* \le x_i$ (for $i = 1, \ldots, n$).

- **Question 4.4.13** What does the definite integral measure exactly, and what are some of the key properties of the definite integral? □

  The definite integral $\int_a^b f(x)\,dx$ measures the exact net signed area bounded by f and the horizontal axis on $[a, b]$; in addition, the value of the definite integral is related to what we call the average value of the function on $[a, b]$: $f_{\mathrm{AVG}[a,b]} = \frac{1}{b-a} \cdot \int_a^b f(x)\,dx$. In the setting where we consider the integral of a velocity function v, $\int_a^b v(t)\,dt$ measures the exact change in position of the moving object on $[a, b]$; when v is nonnegative, $\int_a^b v(t)\,dt$ is the object's distance traveled on $[a, b]$.

  The definite integral is a sophisticated sum, and thus has some of the same natural properties that finite sums have. Perhaps most important of these is how the definite integral respects sums and constant multiples of functions, which can be summarized by the rule

  $$\int_a^b [cf(x) \pm kg(x)]\,dx = c\int_a^b f(x)\,dx \pm k\int_a^b g(x)\,dx$$

  where f and g are continuous functions on $[a, b]$ and c and k are arbitrary constants.

4.4.5 Exercises

1. The velocity of an object moving along an axis is given by the piecewise linear function v that is pictured in Figure 4.4.14. Assume that the object is moving to the right when its velocity is positive, and moving to the left when its velocity is negative. Assume that the given velocity function is valid for $t = 0$ to $t = 4$.

a. Write an expression involving definite integrals whose value is the total change in position (displacement) of the object on the interval $[0, 4]$.

b. Use the provided graph of v to determine the value of the total change in position (displacement) on $[0, 4]$.

c. Write an expression involving definite integrals whose value is the total distance traveled by the object on $[0, 4]$. What is the exact value of the total distance traveled on $[0, 4]$?

d. What is the object's exact average velocity on $[0, 4]$?

e. Find an algebraic formula for the object's position function on $[0, 1.5]$ that satisfies $s(0) = 0$.

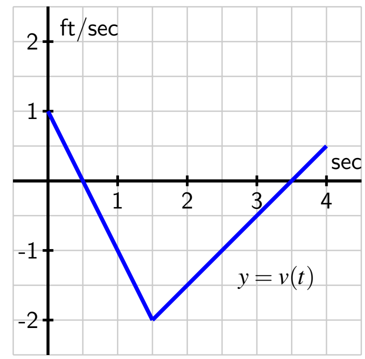


Figure 4.4.14 The velocity function of a moving object.

2. Suppose that the velocity of a moving object is given by $v(t) = t(t-1)(t-3)$, measured in feet per second, and that this function is valid for $0 \le t \le 4$.

a. Write an expression involving definite integrals whose value is the total change in position of the object on the interval $[0, 4]$.

b. Use appropriate technology (such as the interactive in Warm-Up 4.4.1) to compute Riemann sums to estimate the object's total change in position on $[0, 4]$. Work to get an estimate that you think is accurate to two decimal places, and explain why you think this to be the case.

c. Write an expression involving definite integrals whose value is the total distance traveled by the object on $[0, 4]$.

d. Use appropriate technology to compute Riemann sums to estimate the object's total distance travelled on $[0, 4]$. Work to get an estimate that you think is accurate to two decimal places, and explain why you think this to be the case.

e. Use your work above to determine an estimate of the object's average velocity on $[0, 4]$.

3. Consider the graphs of two functions f and g that are provided in Figure 4.4.15. Each piece of f and g is either part of a straight line or part of a circle.

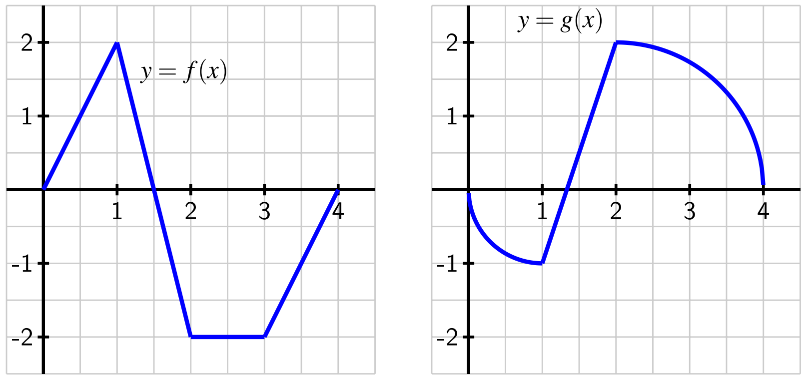


Figure 4.4.15 Two functions f and g.

a. Determine the exact value of $\int_0^1 [f(x) + g(x)]\, dx$.

b. Determine the exact value of $\int_1^4 [2f(x) - 3g(x)]\, dx$.

c. Find the exact average value of $h(x) = g(x) - f(x)$ on $[0, 4]$.

d. For what constant c does the following equation hold?

$$\int_0^4 c\, dx = \int_0^4 [f(x) + g(x)]\, dx$$

4. Let $f(x) = 3 - x^2$ and $g(x) = 2x^2$.

a. On the interval $[-1, 1]$, sketch a labeled graph of $y = f(x)$ and write a definite integral whose value is the exact area bounded by $y = f(x)$ on $[-1, 1]$.

b. On the interval $[-1, 1]$, sketch a labeled graph of $y = g(x)$ and write a definite integral whose value is the exact area bounded by $y = g(x)$ on $[-1, 1]$.

c. Write an expression involving a difference of definite integrals whose value is the exact area that lies between $y = f(x)$ and $y = g(x)$ on $[-1, 1]$.

d. Explain why your expression in (c) has the same value as the single integral $\int_{-1}^1 [f(x) - g(x)]\, dx$.

4.5 The Fundamental Theorem of Calculus

Motivating Questions

- How can we find the exact value of a definite integral without taking the limit of a Riemann sum?
- What is the statement of the Fundamental Theorem of Calculus, and how do antiderivatives of functions play a key role in applying the theorem?
- What is the meaning of the definite integral of a rate of change in contexts other than when the rate of change represents velocity?

Much of our work in Chapter 4 has been motivated by the velocity-distance problem: if we know the instantaneous velocity function, $v(t)$, for a moving object on a given time interval $[a, b]$, can we determine the distance it traveled on $[a, b]$? If the velocity function is nonnegative on $[a, b]$, the area bounded by $y = v(t)$ and the t-axis on $[a, b]$ is equal to the distance traveled. This area is also the value of the definite integral $\int_a^b v(t)\,dt$. If the velocity is sometimes negative, the total area bounded by the velocity function still tells us distance traveled, while the net signed area tells us the object's change in position (displacement).

For instance, for the velocity function in Figure 4.5.1, the total distance D traveled by the moving object on $[a, b]$ is

$$D = A_1 + A_2 + A_3,$$

and the total change in the object's position is

$$s(b) - s(a) = A_1 - A_2 + A_3.$$

The areas A_1, A_2, and A_3 are each given by definite integrals, which may be computed by limits of Riemann sums (and in special circumstances by geometric formulas).

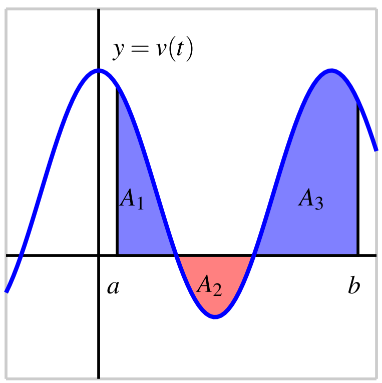


Figure 4.5.1 A velocity function that is sometimes negative.

We turn our attention to an alternate approach.

Warm-Up 4.5.1 A student with a third floor dormitory window 32 feet off the ground tosses a water balloon straight up in the air with an initial velocity of 16 feet per second. It turns out that the instantaneous velocity of the water balloon is given by $v(t) = -32t + 16$, where v is measured in feet per second and t is measured in seconds.

a. Let $s(t)$ represent the height of the water balloon above ground at time t, and note that s is an antiderivative of v. That is, v is the derivative of s: $s'(t) = v(t)$. Find a formula for $s(t)$ that satisfies the initial condition that the balloon is tossed from 32 feet above ground. In other words, make your formula for s satisfy $s(0) = 32$.

b. When does the water balloon reach its maximum height? When does it land?

c. Compute $s(\frac{1}{2}) - s(0)$, $s(2) - s(\frac{1}{2})$, and $s(2) - s(0)$. What do these represent?

d. What is the total vertical distance traveled by the water balloon from the time it is tossed until the time it lands?

e. Sketch a graph of the velocity function $y = v(t)$ on the time interval $[0, 2]$. What is the total net signed area bounded by $y = v(t)$ and the t-axis on $[0, 2]$? Answer this question in two ways: first by using your work above, and then by using a familiar geometric formula to compute areas of certain relevant regions.

4.5.1 The Fundamental Theorem of Calculus

Suppose we know the position function $s(t)$ and the velocity function $v(t)$ of an object moving in a straight line, and for the moment let us assume that $v(t)$ is positive on $[a, b]$. Then, as shown in Figure 4.5.2, we know two different ways to compute the distance, D, the object travels: one is that $D = s(b) - s(a)$, the object's change in position. The other is the area under the velocity curve, which is given by the definite integral, so $D = \int_a^b v(t)\,dt$.

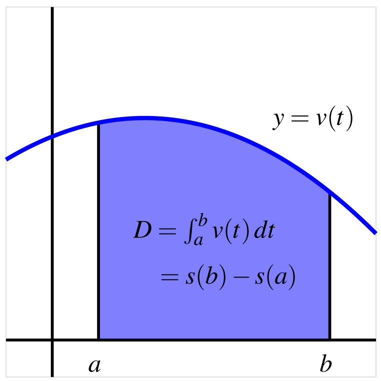


Figure 4.5.2 Finding distance traveled when we know a velocity function v.

Since both of these expressions tell us the distance traveled, it follows that they are equal, so

$$s(b) - s(a) = \int_a^b v(t)\,dt. \tag{4.5.1}$$

Equation (4.5.1) holds even when velocity is sometimes negative, because $s(b) - s(a)$,the object's change in position, is also measured by the net signed area on $[a, b]$ which is given by $\int_a^b v(t)\,dt$.

Perhaps the most powerful fact Equation (4.5.1) reveals is that we can compute the integral's value if we can find a formula for s. Remember, s and v are related by the fact that v is the derivative of s, or equivalently that s is an antiderivative of v.

Example 4.5.3 Determine the exact distance traveled on $[1, 5]$ by an object with velocity function $v(t) = 3t^2 + 40$ feet per second. The distance traveled on the interval $[1, 5]$ is given by

$$D = \int_1^5 v(t)\,dt = \int_1^5 (3t^2 + 40)\,dt = s(5) - s(1),$$

where s is an antiderivative of v. Now, the derivative of t^3 is $3t^2$ and the derivative of $40t$ is 40, so it follows that $s(t) = t^3 + 40t$ is an antiderivative of v. Therefore,

$$D = \int_1^5 3t^2 + 40\,dt = s(5) - s(1)$$

$$= (5^3 + 40 \cdot 5) - (1^3 + 40 \cdot 1) = 284 \text{ feet.}$$

□

Note the key lesson of Example 4.5.3: to find the distance traveled, we need to compute the area under a curve, which is given by the definite integral. But to evaluate the integral, we can find an antiderivative, s, of the velocity function, and then compute the total change in s on the interval. In particular, we can evaluate the integral without computing the limit of a Riemann sum.

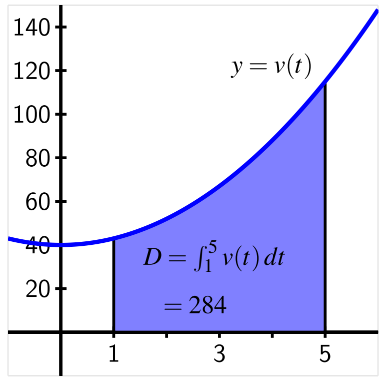


Figure 4.5.4 The exact area of the region enclosed by $v(t) = 3t^2 + 40$ on $[1, 5]$.

It will be convenient to have a shorthand symbol for a function's antiderivative. For a continuous function f, we will often denote an antiderivative of f by F, so that $F'(x) = f(x)$ for all relevant x. Using the notation V in place of s (so that V is an antiderivative of v) in Equation (4.5.1), we can write

$$V(b) - V(a) = \int_a^b v(t)\,dt. \tag{4.5.2}$$

Now, to evaluate the definite integral $\int_a^b f(x)\,dx$ for an arbitrary continuous function f, we could certainly think of f as representing the velocity of some moving object, and x as the variable that represents time. But Equations (4.5.1) and (4.5.2) hold for any continuous velocity function, even when v is sometimes negative. So Equation (4.5.2) offers a shortcut route to evaluating any definite integral, provided that we can find an antiderivative of the integrand. The Fundamental Theorem of Calculus (FTC) summarizes these observations.

Fundamental Theorem of Calculus.

If f is a continuous function on $[a, b]$, and F is any antiderivative of f, then $\int_a^b f(x)\,dx = F(b) - F(a)$.

A common alternate notation for $F(b) - F(a)$ is

$$F(b) - F(a) = F(x)\big|_a^b,$$

where we read the righthand side as "the function F evaluated from a to b." In this notation, the FTC says that

$$\int_a^b f(x)\,dx = F(x)\big|_a^b.$$

The FTC opens the door to evaluating a wide range of integrals if we can find an antiderivative F for the integrand f. For instance since $\frac{d}{dx}[\frac{1}{3}x^3] = x^2$, the FTC tells us that

$$\begin{aligned}\int_0^1 x^2\,dx &= \left.\frac{1}{3}x^3\right|_0^1 \\ &= \frac{1}{3}(1)^3 - \frac{1}{3}(0)^3 \\ &= \frac{1}{3}.\end{aligned}$$

But finding an antiderivative can be far from simple; it is often difficult or even impossible. While we can differentiate just about any function, even some relatively simple functions don't have an elementary antiderivative. A significant portion of integral calculus (which is often the main focus of second semester college calculus) is devoted to the problem of finding antiderivatives.

Activity 4.5.2 Use the Fundamental Theorem of Calculus to evaluate each of the following integrals exactly. For each, sketch a graph of the integrand on the relevant interval and write one sentence that explains the meaning of the value of the integral in terms of the (net signed) area bounded by the curve.

a. $\displaystyle\int_{-1}^{4} (2-2x)\,dx$

b. $\displaystyle\int_{0}^{\frac{\pi}{2}} \sin(x)\,dx$

c. $\displaystyle\int_{0}^{1} e^x\,dx$

d. $\displaystyle\int_{-1}^{1} x^5\,dx$

e. $\displaystyle\int_{0}^{2} (3x^3-2x^2-e^x)\,dx$

4.5.2 Evaluating Definite Integrals via u-substitution

In Subsection 4.2.2 we introduced u-substitution as a means to evaluate indefinite integrals of functions that can be written, up to a constant multiple, in the form $f(g(x))g'(x)$. This same technique can be used to evaluate definite integrals involving such functions, though we need to be careful with the corresponding limits of integration. Consider, for instance, the definite integral

$$\int_2^5 xe^{x^2}\,dx.$$

Whenever we write a definite integral, it is implicit that the limits of integration correspond to the variable of integration. To be more explicit, observe that

$$\int_2^5 xe^{x^2}\,dx = \int_{x=2}^{x=5} xe^{x^2}\,dx.$$

When we execute a u-substitution, we change the *variable* of integration; it is essential to note that this also changes the *limits* of integration. For instance, with the substitution $u = x^2$ and $du = 2x\,dx$, it also follows that when $x = 2$, $u = 2^2 = 4$, and when $x = 5$, $u = 5^2 = 25$. Thus, under the change of variables of u-substitution, we now have

$$\begin{aligned}\int_{x=2}^{x=5} xe^{x^2}\,dx &= \int_{u=4}^{u=25} e^u\cdot\frac{1}{2}\,du \\ &= \left.\frac{1}{2}e^u\right|_{u=4}^{u=25}\end{aligned}$$

$$= \frac{1}{2}e^{25} - \frac{1}{2}e^{4}.$$

Alternatively, we could consider the related indefinite integral $\int xe^{x^2}\,dx$, find the antiderivative $\frac{1}{2}e^{x^2}$ through u-substitution, and then evaluate the original definite integral. With that method, we'd have

$$\begin{aligned}\int_2^5 xe^{x^2}\,dx &= \frac{1}{2}e^{x^2}\Big|_2^5 \\ &= \frac{1}{2}e^{25} - \frac{1}{2}e^{4},\end{aligned}$$

which is, of course, the same result.

Activity 4.5.3 Evaluate each of the following definite integrals exactly through an appropriate u-substitution.

a. $\displaystyle\int_1^2 \frac{x}{1+4x^2}\,dx$

b. $\displaystyle\int_0^1 e^{-x}(2e^{-x}+3)^9\,dx$

c. $\displaystyle\int_{2/\pi}^{4/\pi} \frac{\cos\left(\frac{1}{x}\right)}{x^2}\,dx$

4.5.3 The total change theorem

Let us review three interpretations of the definite integral.

- For a moving object with instantaneous velocity $v(t)$, the object's change in position on the time interval $[a,b]$ is given by $\int_a^b v(t)\,dt$, and whenever $v(t) \ge 0$ on $[a,b]$, $\int_a^b v(t)\,dt$ tells us the total distance traveled by the object on $[a,b]$.

- For any continuous function f, its definite integral $\int_a^b f(x)\,dx$ represents the net signed area bounded by $y = f(x)$ and the x-axis on $[a,b]$, where regions that lie below the x-axis have a minus sign associated with their area.

- The value of a definite integral is linked to the average value of a function: for a continuous function f on $[a,b]$, its average value $f_{\text{AVG}[a,b]}$ is given by

$$f_{\text{AVG}[a,b]} = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

The Fundamental Theorem of Calculus now enables us to evaluate exactly (without taking a limit of Riemann sums) any definite integral for which we are able to find an antiderivative of the integrand.

A slight change in perspective allows us to gain even more insight into the meaning of the definite integral. Recall Equation (4.5.2), where we wrote the Fundamental Theorem of Calculus for a velocity function v with antiderivative V as

$$V(b) - V(a) = \int_a^b v(t)\,dt.$$

If we instead replace V with s (which represents position) and replace v with s' (since velocity is the derivative of position), Equation (4.5.2) then reads as

$$s(b) - s(a) = \int_a^b s'(t)\,dt. \tag{4.5.3}$$

In words, this version of the FTC tells us that the total change in an object's position function on a particular interval is given by the definite integral of the position function's derivative over that interval.

Of course, this result is not limited to only the setting of position and velocity. Writing the result in terms of a more general function f, we have the Total Change Theorem, which is a useful interpretation for gaining insight about a general pure-time differential equation.

Total Change Theorem.

If f is a continuously differentiable function on $[a, b]$ with derivative f', then $f(b) - f(a) = \int_a^b f'(x)\,dx$. That is, the definite integral of the rate of change of a function on $[a, b]$ is the total change of the function itself on $[a, b]$.

The Total Change Theorem tells us more about the relationship between the graph of a function and that of its derivative. Recall that heights on the graph of the derivative function are equal to slopes on the graph of the function itself. If instead we know f' and are seeking information about f, as is the case when modeling using differential equations, we can say the following:

> *differences in heights on f correspond to net signed areas bounded by f'.*

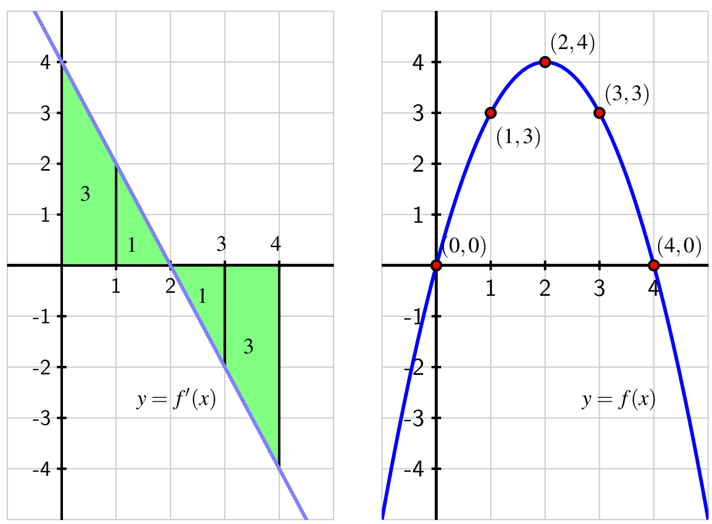


Figure 4.5.5 The graphs of $f'(x) = 4 - 2x$ (at left) and an antiderivative $f(x) = 4x - x^2$ at right. Differences in heights on f correspond to net signed areas bounded by f'.

To see why this is so, consider the difference $f(1) - f(0)$. This value is 3, because $f(1) = 3$ and $f(0) = 0$, but also because the net signed area bounded by $y = f'(x)$ on $[0, 1]$ is 3. That is,

$$f(1) - f(0) = \int_0^1 f'(x)\,dx.$$

In addition to this observation about area, the Total Change Theorem enables us to answer questions about a function whose rate of change we know.

Example 4.5.6 Suppose that pollutants are leaking out of an underground storage tank at a rate of $r(t)$ gallons/day, where t is measured in days. It is conjectured that $r(t)$ is given by the formula $r(t) = 0.0069t^3 - 0.125t^2 + 11.079$ over a certain 12-day period. The graph of $y = r(t)$ is given in Figure 4.5.7. What is the meaning of $\int_4^{10} r(t)\,dt$ and what is its value? What is the average rate at which pollutants are leaving the tank on the time interval $4 \le t \le 10$?

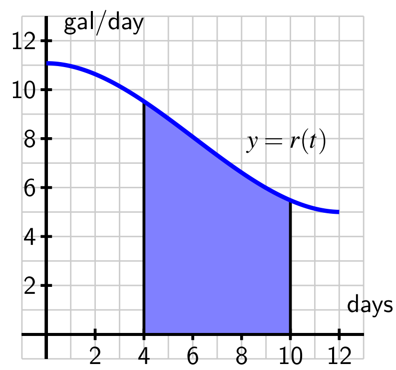


Figure 4.5.7 The rate $r(t)$ of pollution leaking from a tank, measured in gallons per day.

Solution. Since $r(t) \ge 0$, the value of $\int_4^{10} r(t)\,dt$ is the area under the curve on the interval $[4, 10]$. A Riemann sum for this area will have rectangles with heights measured in gallons per day and widths measured in days, so the area of each rectangle will have units of

$$\frac{\text{gallons}}{\text{day}} \cdot \text{days} = \text{gallons}.$$

Thus, the definite integral tells us the total number of gallons of pollutant that leak from the tank from day 4 to day 10. The Total Change Theorem tells us the same thing: if we let $R(t)$ denote the total number of gallons of pollutant that have leaked from the tank up to day t, then $R'(t) = r(t)$, and

$$\int_4^{10} r(t)\,dt = R(10) - R(4),$$

the number of gallons that have leaked from day 4 to day 10.

To compute the exact value of the integral, we use the Fundamental Theorem of Calculus. Antidifferentiating $r(t) = 0.0069t^3 - 0.125t^2 + 11.079$, we find that

$$\int_4^{10} 0.0069t^3 - 0.125t^2 + 11.079\,dt = \left. 0.0069 \cdot \frac{1}{4}t^4 - 0.125 \cdot \frac{1}{3}t^3 + 11.079t \right|_4^{10}$$
$$\approx 44.282.$$

Thus, approximately 44.282 gallons of pollutant leaked over the six day time period.

To find the average rate at which pollutant leaked from the tank over $4 \le t \le 10$, we compute the average value of r on $[4, 10]$. Thus,

$$r_{\text{AVG}[4,10]} = \frac{1}{10-4}\int_4^{10} r(t)\,dt \approx \frac{44.282}{6} = 7.380$$

gallons per day. □

Activity 4.5.4 During a 40-minute workout, a person riding an exercise machine burns calories at a rate of c calories per minute, where the function $y = c(t)$ is given in Figure 4.5.8. On the interval $0 \le t \le 10$, the formula for c is $c(t) = -0.05t^2 + t + 10$, while on $30 \le t \le 40$, its formula is $c(t) = -0.05t^2 + 3t - 30$.

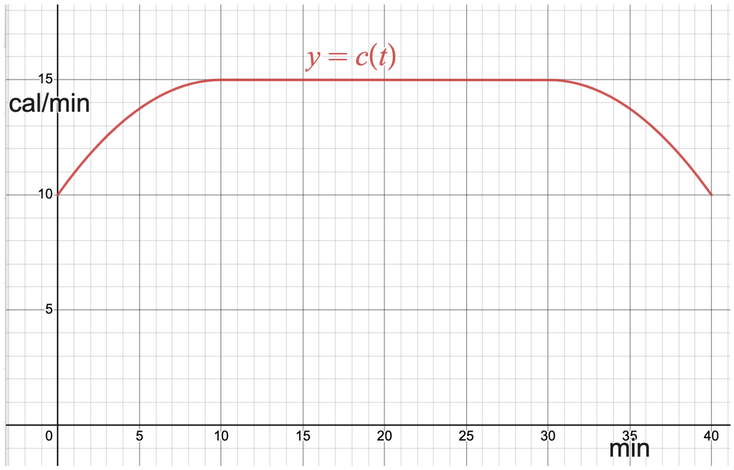


Figure 4.5.8 The rate $c(t)$ at which a person exercising burns calories, measured in calories per minute.

a. What is the exact total number of calories the person burns during the first 10 minutes of her workout?

b. Let $C(t)$ be an antiderivative of $c(t)$. What is the meaning of $C(40) - C(0)$ in the context of the person exercising? Include units on your answer.

c. Determine the exact average rate at which the person burned calories during the 40-minute workout.

d. At what time(s), if any, is the instantaneous rate at which the person is burning calories equal to the average rate at which she burns calories, on the time interval $0 \le t \le 40$?

4.5.4 Summary

- **Question 4.5.9** How can we find the exact value of a definite integral without taking the limit of a Riemann sum? □

  We can find the exact value of a definite integral without taking the limit of a Riemann sum or using a familiar area formula by finding the antiderivative of the integrand, and then applying the Fundamental Theorem of Calculus.

- **Question 4.5.10** What is the statement of the Fundamental Theorem of Calculus, and how do antiderivatives of functions play a key role in applying the theorem? □

  The Fundamental Theorem of Calculus says that if f is a continuous function on $[a, b]$ and F is an antiderivative of f, then

  $$\int_a^b f(x)\,dx = F(b) - F(a).$$

  Hence, if we can find an antiderivative for the integrand f, evaluating the definite integral comes from simply computing the change in F on $[a, b]$.

- **Question 4.5.11** What is the meaning of the definite integral of a rate of change in contexts other than when the rate of change represents velocity? □

A slightly different perspective on the FTC allows us to restate it as the Total Change Theorem, which says that

$$\int_a^b f'(x)\,dx = f(b) - f(a),$$

for any continuously differentiable function f. This means that the definite integral of the instantaneous rate of change of a function f on an interval $[a, b]$ is equal to the total change in the function f on $[a, b]$.

4.5.5 Exercises

1. The instantaneous velocity (in meters per minute) of a moving object is given by the function v as pictured in Figure 4.5.12. Assume that on the interval $0 \le t \le 4$, $v(t)$ is given by $v(t) = -\frac{1}{4}t^3 + \frac{3}{2}t^2 + 1$, and that on every other interval v is piecewise linear, as shown.

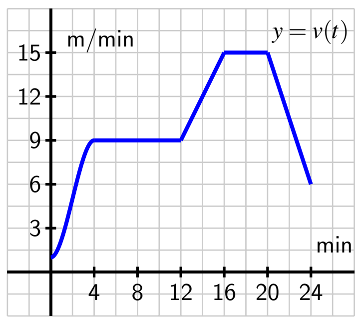


Figure 4.5.12 The velocity function of a moving body.

a. Determine the exact distance traveled by the object on the time interval $0 \le t \le 4$.

b. What is the object's average velocity on $[12, 24]$?

c. At what time is the object's acceleration greatest?

d. Suppose that the velocity of the object is increased by a constant value c for all values of t. What value of c will make the object's total distance traveled on $[12, 24]$ be 210 meters?

2. When an aircraft attempts to climb as rapidly as possible, its climb rate (in feet per minute) decreases as altitude increases, because the air is less dense at higher altitudes. Given below is a table showing performance data for a certain single engine aircraft, giving its climb rate at various altitudes, where $c(h)$ denotes the climb rate of the airplane at an altitude h.

h (feet)	0	1000	2000	3000	4000	5000	6000	7000	8000	9000	10,000
c (ft/min)	925	875	830	780	730	685	635	585	535	490	440

Let a new function called $m(h)$ measure the number of minutes required for a plane at altitude h to climb the next foot of altitude.

a. Determine a similar table of values for $m(h)$ and explain how it is related to the table above. Be sure to explain the units.

b. Give a careful interpretation of a function whose derivative is $m(h)$. Describe what the input is and what the output is. Also, explain intuitively what the function tells us.

c. Determine a definite integral whose value tells us exactly the number of minutes required for the airplane to ascend to 10,000 feet of altitude. Clearly explain why the value of this integral has the required meaning.

d. Use the Riemann sum M_5 to estimate the value of the integral you found in (c). Include units on your result.

3. In Chapter 1, we discussed that for an object moving along a straight line with position function $s(t)$, the object's "average rate of change on the interval $[a, b]$" is given by

$$AROC_{[a,b]} = \frac{s(b) - s(a)}{b - a}.$$

As we know, the rate of change of position describes velocity, so we may also say this is the object's "average velocity on the interval $[a, b]$".

More recently in Chapter 4, we found that for an object moving along a straight line with velocity function $v(t)$, the object's "average value of its velocity function on $[a, b]$" is

$$v_{\text{AVG}[a,b]} = \frac{1}{b - a} \int_a^b v(t)\, dt.$$

Are the "average velocity on the interval $[a, b]$" and the "average value of the velocity function on $[a, b]$" the same thing? Why or why not? Explain.

4.6 Approximations of Solutions

Motivating Questions

- What is Euler's method and how can we use it to approximate the solution to an initial value problem?
- How accurate is Euler's method?

In Chapter 4 we have been interested in describing solutions to differential equations. We have focused mostly on *pure-time* differential equations, which have the form $\frac{dy}{dt} = f(t)$. Given an initial value, we can find a specific solution if we can find the family of antiderivatives of $f(t)$. Even without an initial value, we can describe how the solution changes over time using the Fundamental Theorem of Calculus. Though we have seen several types of functions for which we can compute an antiderivative, this is generally a difficult task. How might we describe the solution to a pure-time differential equation when we cannot compute an antiderivative? Further, how might we describe the solution to a differential equation which is not pure-time?

In this section we will gain a tool for *approximating* solutions to differential equations. The main method we will use relies on linear approximation (Section 3.1).

Warm-Up 4.6.1 Consider the initial value problem

$$\frac{dy}{dt} = \frac{1}{2}(y+1), \; y(0) = 0.$$

Note that this differential equation is not pure-time; in fact, it is *autonomous* (Section 4.1), since the derivative value depends only on the dependent variable y. For example, the differential equation along with the initial value $(0,0)$ tells us that the derivative value of the solution function $y(t)$ at $t = 0$ is

$$\left.\frac{dy}{dt}\right|_{(0,0)} = \frac{1}{2}(y(0)+1) = \frac{1}{2}(0+1) = \frac{1}{2}.$$

a. Above we used the differential equation and initial value to find the slope of the tangent line to the solution $y(t)$ at $t = 0$. Now use the given initial value to find the linear approximation of $y(t)$ centered at $t = 0$.

b. Sketch the linear approximation of $y(t)$ centered at $t = 0$ on the axes provided in Figure 4.6.1 on the interval $0 \leq t \leq 2$ and use it to approximate $y(2)$, the value of the solution at $t = 2$.

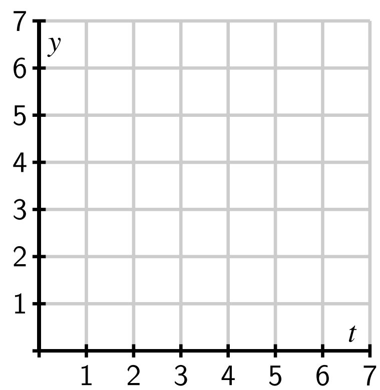


Figure 4.6.1 Grid for plotting the tangent line.

c. Assuming that your approximation for $y(2)$ is the actual value of $y(2)$, use the differential equation to find the slope of the linear approximation of $y(t)$ centered at $t = 2$. Then, write the equation of the linear approximation of $y(t)$ centered at $t = 2$.

d. Add a sketch of this linear approximation on the interval $2 \leq t \leq 4$ to your plot Figure 4.6.1; use this new linear approximation to approximate $y(4)$, the value of the solution at $t = 4$.

e. Repeat the same step to find an approximation for $y(6)$.

4.6.1 Euler's Method

Warm-Up 4.6.1 demonstrates an algorithm known as Euler's[1] Method, which generates a numerical approximation to the solution of an initial value problem. In this algorithm, we will approximate the solution by taking horizontal steps of a fixed size that we denote by Δt.

[1] "Euler" is pronounced "Oy-ler." Among other things, Euler is the mathematician credited with the famous number e.

Before explaining the algorithm in detail, let's remember how we compute the slope of a line: the slope is the ratio of the vertical change to the horizontal change, as shown in Figure 4.6.2.
In other words, $m = \frac{\Delta y}{\Delta t}$. Solving for Δy, we see that the vertical change is the product of the slope and the horizontal change, or

$$\Delta y = m\Delta t.$$

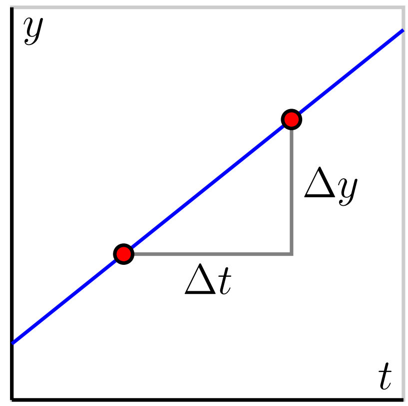


Figure 4.6.2 The role of slope in Euler's Method.

Now, suppose that we would like to solve the initial value problem

$$\frac{dy}{dt} = t - y,\ \ y(0) = 1.$$

There is an algorithm by which we can find an algebraic formula for the solution to this initial value problem, and we can check that this solution is $y(t) = t - 1 + 2e^{-t}$. But we are instead interested in generating an approximate solution by creating a sequence of points (t_i, y_i), where $y_i \approx y(t_i)$. For this first example, we choose $\Delta t = 0.2$.

Since we know that $y(0) = 1$, we will take the initial point to be $(t_0, y_0) = (0, 1)$ and move horizontally by $\Delta t = 0.2$ to the point (t_1, y_1). Thus, $t_1 = t_0 + \Delta t = 0.2$. Now, the differential equation tells us that the slope of the tangent line at this point is

$$m = \left.\frac{dy}{dt}\right|_{(0,1)} = 0 - 1 = -1,$$

so to move along the tangent line by taking a horizontal step of size $\Delta t = 0.2$, we must also move vertically by

$$\Delta y = m\Delta t = -1 \cdot 0.2 = -0.2.$$

We then have the approximation $y(0.2) \approx y_1 = y_0 + \Delta y = 1 - 0.2 = 0.8$. At this point, we have executed one step of Euler's method, as seen graphically in Figure 4.6.3.

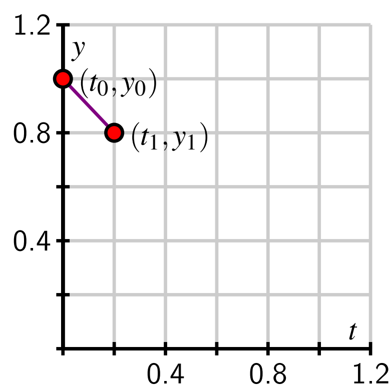


Figure 4.6.3 One step of Euler's method.

Now we repeat this process: at $(t_1, y_1) = (0.2, 0.8)$, the differential equation tells us that the slope is

$$m = \left.\frac{dy}{dt}\right|_{(0.2,0.8)} = 0.2 - 0.8 = -0.6.$$

If we move forward horizontally by Δt to $t_2 = t_1 + \Delta t = 0.4$, we must move vertically by

$$\Delta y = -0.6 \cdot 0.2 = -0.12.$$

We consequently arrive at $y_2 = y_1 + \Delta y = 0.8 - 0.12 = 0.68$, which gives $y(0.2) \approx 0.68$. Now we have completed the second step of Euler's method, as shown in Figure 4.6.4.

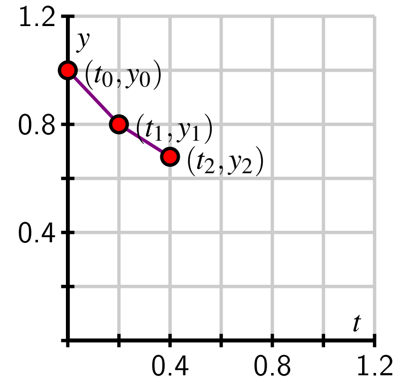


Figure 4.6.4 Two steps of Euler's method.

If we continue in this way, we may generate the points (t_i, y_i) shown in Figure 4.6.5. Because we can find a formula for the actual solution $y(t)$ to this differential equation, we can graph $y(t)$ and compare it to the points generated by Euler's method, as shown in Figure 4.6.6.

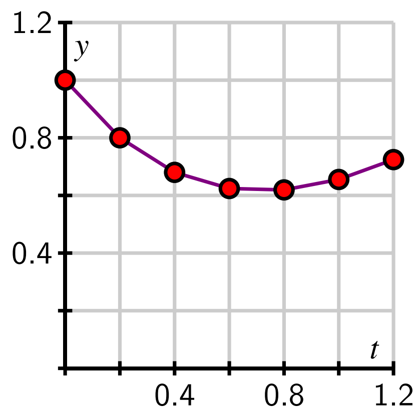


Figure 4.6.5 The points and piecewise linear approximate solution generated by Euler's method.

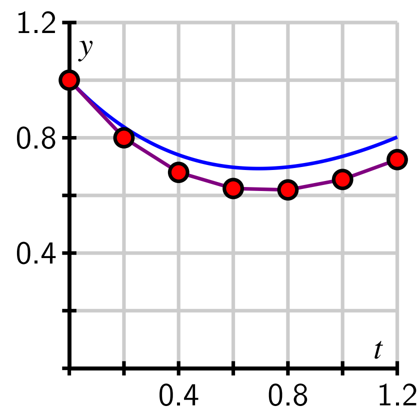


Figure 4.6.6 The approximate solution compared to the exact solution (shown in blue).

Because we need to generate a large number of points (t_i, y_i), it is convenient to organize the implementation of Euler's method in a table as shown. We begin with the given initial data.

t_i	y_i	dy/dt	Δy
0.0000	1.0000		

From here, we compute the slope of the tangent line $m = dy/dt$ using the formula for dy/dt from the differential equation, and then we find Δy, the change in y, using the rule $\Delta y = m\Delta t$.

t_i	y_i	dy/dt	Δy
0.0000	1.0000	-1.0000	-0.2000

Next, we increase t_i by Δt and y_i by Δy to get

t_i	y_i	dy/dt	Δy
0.0000	1.0000	−1.0000	−0.2000
0.2000	0.8000		

We continue the process for however many steps we decide, eventually generating a table like Table 4.6.7.

Table 4.6.7 Euler's method for 6 steps with $\Delta t = 0.2$.

t_i	y_i	dy/dt	Δy
0.0000	1.0000	−1.0000	−0.2000
0.2000	0.8000	−0.6000	−0.1200
0.4000	0.6800	−0.2800	−0.0560
0.6000	0.6240	−0.0240	−0.0048
0.8000	0.6192	0.1808	0.0362
1.0000	0.6554	0.3446	0.0689
1.2000	0.7243	0.4757	0.0951

Activity 4.6.2 Consider the initial value problem

$$\frac{dy}{dt} = 2t - 1, \ \ y(0) = 0$$

a. Use Euler's method with $\Delta t = 0.2$ to approximate the solution at $t_i = 0.2, 0.4, 0.6, 0.8,$ and 1.0. Record your work in the following table, and sketch the points (t_i, y_i) on the axes provided.

t_i	y_i	dy/dt	Δy
0.0000	0.0000		
0.2000			
0.4000			
0.6000			
0.8000			
1.0000			

Table 4.6.8 Table for recording results of Euler's method.

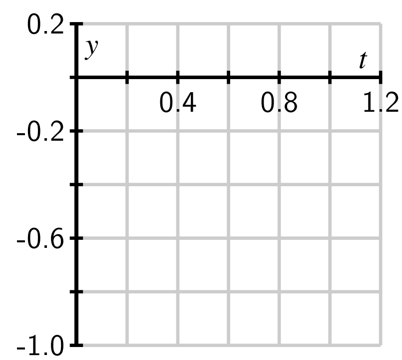


Figure 4.6.9 Grid for plotting points generated by Euler's method.

b. Find the exact solution to the original initial value problem and use this function to find the error in your approximation at each one of the points t_i.

c. Explain why the value y_5 generated by Euler's method for this initial value problem produces the same value as a left Riemann sum for the definite integral $\int_0^1 (2t - 1)\, dt$.

d. How would your computations differ if the initial value was $y(0) = 1$? What does this mean about different solutions to this differential equation?

Activity 4.6.3 Consider the differential equation $\frac{dy}{dt} = 6y - y^2$.

a. Using the initial value $y(0) = 1$, use Euler's method with $\Delta t = 0.2$ to approximate the solution at $t_i = 0.2, 0.4, 0.6, 0.8$, and 1.0. Record your results in Table 4.6.10 and sketch the corresponding points (t_i, y_i) on the axes provided in Figure 4.6.11.

t_i	y_i	dy/dt	Δy
0.0	1.0000		
0.2			
0.4			
0.6			
0.8			
1.0			

Table 4.6.10 Table for recording results of Euler's method with $\Delta t = 0.2$.

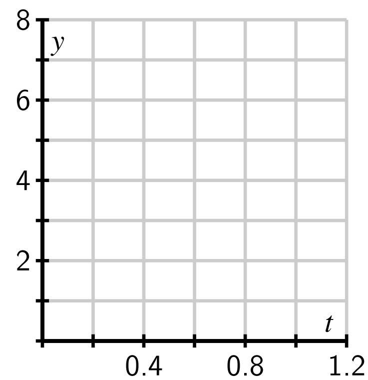


Figure 4.6.11 Axes for plotting the results of Euler's method.

b. What happens if we apply Euler's method to approximate the solution with $y(0) = 6$? What would you call the point $(0, 6)$ in this system, based on a similar concept in discrete-time dynamical systems?

4.6.2 The error in Euler's method

Since we are approximating the solutions to an initial value problem using local linearizations, we should expect that the error in the approximation will be smaller when the step size is smaller. Consider the initial value problem

$$\frac{dy}{dt} = y, \ y(0) = 1,$$

whose specific solution is possible find.

The question posed by this initial value problem is "what function do we know that is the same as its own derivative and has value 1 when $t = 0$?" It is not hard to see that the solution is $y(t) = e^t$. We now apply Euler's method to approximate $y(1) = e$ using several values of Δt. These approximations will be denoted by $E_{\Delta t}$, and we'll use them to see how accurate Euler's Method is.

To begin, we apply Euler's method with a step size of $\Delta t = 0.2$. In that case, we find that $y(1) \approx E_{0.2} = 2.4883$. The error is therefore

$$y(1) - E_{0.2} = e - 2.4883 \approx 0.2300.$$

Repeatedly halving Δt gives the following results, expressed in both tabular and graphical form.

Δt	$E_{\Delta t}$	Error
0.200	2.4883	0.2300
0.100	2.5937	0.1245
0.050	2.6533	0.0650
0.025	2.6851	0.0332

Table 4.6.12 Errors that correspond to different Δt values.

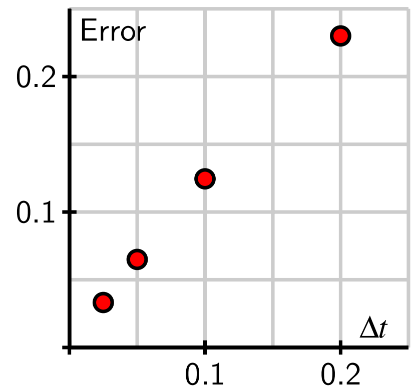


Figure 4.6.13 A plot of the error as a function of Δt.

Notice, both numerically and graphically, that the error is roughly halved when Δt is halved. This example illustrates the following general principle.

If Euler's method is used to approximate the solution to an initial value problem at a point $\bar{t}$, then the error is proportional to Δt. That is,

$$y(\bar{t}) - E_{\Delta t} \approx K\Delta t$$

for some constant of proportionality K.

4.6.3 Summary

- **Question 4.6.14** What is Euler's method and how can we use it to approximate the solution to an initial value problem? ☐

  Euler's method is an algorithm for approximating the solution to an initial value problem by following the tangent lines while we take horizontal steps across the t-axis.

- **Question 4.6.15** How accurate is Euler's method? ☐

  If we wish to approximate $y(\bar{t})$ for some fixed $\bar{t}$ by taking horizontal steps of size Δt, then the error in our approximation is proportional to Δt.

4.6.4 Exercises

1. Newton's Law of Cooling says that the rate at which an object, such as a cup of coffee, cools is proportional to the difference in the object's temperature and room temperature. If $T(t)$ is the object's temperature and T_r is room temperature, this law is expressed as

$$\frac{dT}{dt} = -k(T - T_r),$$

where k is a constant of proportionality. In this problem, temperature is measured in degrees Fahrenheit and time in minutes.

a. Two calculus students, Alice and Bob, enter a 70° classroom at the same time. Each has a cup of coffee that is 100°. The differential equation for Alice has a constant of proportionality $k = 0.5$, while the constant of proportionality for Bob is $k = 0.1$. What is the initial rate of change for Alice's coffee? What is the initial rate of change for Bob's coffee?

b. What feature of Alice's and Bob's cups of coffee could explain this difference?

c. As the heating unit turns on and off in the room, the temperature in the room is

$$T_r = 70 + 10\sin t.$$

Implement Euler's method with a step size of $\Delta t = 0.1$ to approximate the temperature of Alice's coffee over the time interval $0 \leq t \leq 50$. This will most easily be performed using a spreadsheet such as *Excel*. Graph the temperature of her coffee and room temperature over this interval.

d. In the same way, implement Euler's method to approximate the temperature of Bob's coffee over the same time interval. Graph the temperature of his coffee and room temperature over the interval.

e. Explain the similarities and differences that you see in the behavior of Alice's and Bob's cups of coffee.

2. Euler's Method is a way to use linear approximations in order to approximate solution values of an initial value problem. In this exercise, we'll see how this concept can be applied in a similar way using *quadratic* approximations.

(a) In Euler's Method using a step size of Δt, we use the slope m of the linear approximation at a given point to compute the corresponding increase in y as $\Delta y = m\Delta t$. If we used a quadratic approximation (where the second derivative value was m') instead of a linear approximation, explain why the corresponding increase in y would be $\Delta y = m\Delta t + \frac{m'}{2}(\Delta t)^2$. HINT: Review quadratic approximations in Section 3.1.

(b) Given the differential equation $\frac{dy}{dt} = t - y$, what is an expression for the second derivative of y with respect to t, $\frac{d^2y}{dt^2}$?

(c) Consider the initial value problem $\frac{dy}{dt} = t - y$, $y(0) = 1$, from the beginning of this section. Approximate $y(1)$ using quadratic approximations and $\Delta t = 0.2$. Compare your approximations to that of Euler's Method in Table 4.6.7. Given that the exact solution is $y(t) = t - 1 + 2e^{-t}$, which method produces a better approximation? Why do think this is?

Appendix A

Answers to Selected Exercises

Below you will find answers to selected exercises from the end of each section. Worked solutions are not provided, so you are encouraged to discuss your thought process and reasoning for exercises with others, even if your final answer matches what is listed here. Your process and reasoning are the most important things for deepening your understanding and ensuring you can use important concepts on your own in the future, so use this section as a guide, but not a replacement for engaging in the exercises with others.

1 · Functions As Models

1.1 · Biology and Calculus

1.1.4 · Exercises

1.1.4.1. 14.04 feet

1.2 · Functions

1.2.8 · Exercises

1.2.8.3.

(a) Parameters: k, m, c

(b) Parameters: k, t, c

1.2.8.5.

(a) -8

(b) 12

(c) -20

(d) -10

(e) 5

1.2.8.6.

(a) 3

(b) 0

(c) -2

1.3 · Units and Dimensions of Functions

1.3.5 · Exercises

1.3.5.2.

(a) °F

(b) °F/°C

1.4 · Linear Functions
1.4.5 · Exercises

1.4.5.2.

(a) $y - 10 = \frac{-13}{6}(x + 2)$

(b) $y = \frac{-13}{6}x + \frac{17}{3}$

1.4.5.3. $x = \frac{-4}{m-2}$, $m \neq 2$

1.4.5.4.

(a) dollars

(b) dollars/mile

1.5 · Exponential and Logarithmic Functions
1.5.5 · Exercises

1.5.5.1. Increasing for $k > 0$, decreasing for $k < 0$

1.5.5.2. $c = 4$

1.5.5.3. $h(x) = 1.5x^4$

1.5.5.4. slope: $\log(b)$, y-intercept: $(0, \log(a))$

1.6 · Trigonometric Functions
1.6.4 · Exercises

1.6.4.1. Yes

1.6.4.2. No

1.6.4.3.

(a) $c(t) = 2\cos\left(\frac{\pi}{12}t\right) + 3$, $u(t) = 0.5\cos\left(\frac{\pi}{2}t\right) + 1$

1.6.4.4. $x = \frac{\pi}{6} + 2\pi k$, $x = \frac{5\pi}{6} + 2\pi k$, where k is any integer
2 solutions in $[0, \pi]$
0 solutions in $[-\pi, 0]$

1.7 · Discrete-Time Dynamical Systems
1.7.4 · Exercises

1.7.4.2.

(a) $b_t = t + 5$

(b) $b_t = 5 \cdot (0.75)^t$

1.7.4.3.

(a) $b_{t+1} = 0.4b_t + 3$, $b_0 = 10$

(b)

(c) $b_t = 5 \cdot (0.4)^t + 5$

1.8 · Analyzing Discrete-Time Dynamical Systems
1.8.4 · Exercises

1.8.4.1. No

1.9 · Applications: The Lung Model and Competing Species

1.9.4 · Exercises

1.9.4.1.

(a)

(b) $c^* = \dfrac{p\beta}{p + \alpha(1-p)}$

(c) **Table A.0.1**

α	c^*
0	0.2
0.25	0.16
0.5	0.133
0.75	0.114
1	0.1

1.9.4.2.

(a) $c_t = -(0.4)^t + 2$

(b) About 5 breaths

1.9.4.4.

(a)

(b) $x_{t+1} = 1.6x_t + 0.3y_t$, $y_{t+1} = 0.4x_t + 0.7y_t$

(c) $p_{t+1} = \dfrac{1.3p_t + 0.3}{p_t + 1}$

(d) $p^* = 0.72$, stable

1.9.4.5.

(a) $k_{t+1} = 0.2k_t + 0.75h_t$

(b) $h_{t+1} = 0.8k_t + 0.25h_t$

(c) $p_{t+1} = 0.75 - 0.55p_t$

(d) $p^* = 0.484$, stable

2 · The Derivative

2.1 · Limits of Functions

2.1.4 · Exercises

2.1.4.1.

(a) $x \neq \pm 2$

(b)

(c) -8

(d) False

(e) False

2.1.4.2.

(a) $x \neq -3$

(b)

(c) Does not exist

(d) False

(e) False

2.1.4.4.

(a) $AROC_{[1,1+h]} = \dfrac{100\cos(0.75(1+h)) \cdot e^{-0.2(1+h)} - 100\cos(0.75) \cdot e^{-0.2}}{h}$

(b) -53.8

(c) feet/s

2.2 · The Derivative of a Function at a Point

2.2.3 · Exercises

2.2.3.1.

(a)

(b) $AROC_{[-3,-1]} \approx \frac{2.3}{2} = 1.15$

$AROC_{[0,2]} \approx \frac{-0.8}{2} = -0.4$

(c) $IROC_{x=-3} \approx \frac{5}{2} = 2.5$

$IROC_{x=0} \approx \frac{-1}{3}$

2.2.3.3.

(a) $P(7) - P(0) \approx .118$ billion people

$AROC_{[0,7]} \approx 0.017$ billion people per year

$AROC_{[0,7]} < IROC_{t=7}$

(b) $AROC_{[19,29]} \approx 0.0223$ billion people per year

(c) If today is July 1, 2022, then the limit would be $\lim\limits_{h\to 0} \dfrac{P(29.5+h) - P(29.5)}{h} \approx 0.024$ billion people per year

(d) If today is July 1, 2022, the equation of the tangent line would be $y = 1.733 + 0.024(t - 29.5)$

2.2.3.4.

(a) $f'(2) = 1$

(b) $f'(1) = -1$

(c) $f'(1) = 0.5$

(d) $f'(1)$ does not exist

(e) $f'\left(\dfrac{\pi}{2}\right) = 0$

2.3 · The Derivative Function

2.3.3 · Exercises

2.3.3.2.

(a) $g'(x) = 2x - 1$

(b)

(c) $p'(x) = 10x - 4$

2.3.3.3.

(a) g is linear with slope 1 on $(0, 2)$

(b) $(-\infty, -2)$, $(-2, 0)$,$(2, \infty)$

(c) $x = -2, 0, 2$

2.4 · The Second Derivative

2.4.5 · Exercises

2.4.5.1.

(a) f is increasing, concave down near $x = 2$

(b) It is likely that $f(2.1) > -3$

(c) It is likely that $f'(2.1) < 1.5$

2.4.5.2.

(a) $g'(2) \approx 1.4$

(b) g can have at most one real zero

(c) 9

(d) $g''(2) \approx 5$

2.5 · Elementary Derivative Rules

2.5.5 · Exercises

2.5.5.1.

(a) $h(2) = 27$, $h'(2) = \frac{-19}{2}$

(b) $y = 27 - \frac{19}{2}(x - 2)$

(c) Increasing

2.5.5.2.

(a) p' and q' both do not exist when $x = \pm 1$

(b) $r'(-2) = 4$, $r'(0) = 0.5$

(c) $y = 4$

2.6 · Derivatives of the Sine and Cosine Functions

2.6.3 · Exercises

2.6.3.1.

(a) $V'(2) = -0.638$ thousand dollars per year

(b) $V''(2) = -5.33$ thousand dollars per year per year

2.6.3.2.

(a) $f'\left(\frac{\pi}{4}\right) = \frac{-5\sqrt{2}}{2}$

(b) $y = 3 + 2(x - \pi)$

(c) Decreasing

(d) Above the curve

2.6.3.3.

(a)

(b) i. $s(\theta) = \cos(\theta)$

ii. $s'(\theta) = -\sin(\theta)$

iii. $s'(\theta) = \cos\left(\theta + \frac{\pi}{2}\right)$

2.7 · Derivatives of Products and Quotients

2.7.5 · Exercises

2.7.5.2.

(a) $h(2) = -15$, $h'(2) = \frac{23}{2}$

(b) $y = -15 + \frac{23}{2}(x-2)$

(c) Increasing

(d) $y = \frac{-3}{5} + \frac{17}{50}(x-2)$

2.7.5.3.

(a) $r'(-2) = 5$, $r'(0) = 1$

(b) $y = 2$

(c) $z'(0) = -4$, $z'(2) = -1$

(d) $\ell'(0) = 0$

2.7.5.4.

(a) $C(t) = A(t) \cdot Y(t)$

(b) $C(0) = 1,190,000$ bushels

(c) $C'(t) = A'(t)Y(t) + A(t)Y'(t)$

(d) $C'(0) = 158,000$ bushels per year

(e) $y = 1,190,000 + 158,000t$

2.7.5.5.

(a) $g(v) = \frac{1}{f(v)}$, $g(80) = 20$ km/L, $g'(v) = -0.16$ (km/L)/(km/h)

(b) $h(v) = f(v) \cdot v$, $h(80) = 4$ L/h, $h'(80) = 0.082$ (L/h)/(km/h)

2.8 · Derivatives of Compositions

2.8.5 · Exercises

2.8.5.1.

(a) $h'\left(\frac{\pi}{4}\right) = \frac{3\sqrt{2}}{4}$

(b) r is changing most rapidly

2.8.5.2.

(a) $p'(x) = e^{u(x)} \cdot u'(x)$

(b) $q'(x) = u'(e^x) \cdot e^x$

(c) $r'(x) = -\sin(u(x)) \cdot u'(x)$

(d) $s'(x) = u'(\cos(x)) \cdot (-\sin(x))$

(e) $a'(x) = u'(x^4) \cdot 4x^3$

(f) $b'(x) = 4(u(x))^3 \cdot u'(x)$

2.8.5.3.

(a) $C'(0) = 0$, $C'(3) = -0.5$

(b) $Y'(3) = 0$, $Z'(0) = 0$

2.8.5.4.

(a) 7π cubic feet per foot

(b) $h'(2) = \pi$ feet per hour

(c) $7\pi^2$ cubic feet per hour

2.9 · Derivatives of Inverse Functions

2.9.5 · Exercises

2.9.5.1. $\ell'(x) = \dfrac{1}{\ln(b)x}$

2.9.5.2.

(a) $f'(1) \approx 2$

(b)

(c) $(f^{-1})'(-1) \approx 0.5$

2.9.5.3.

(a)

(b) $g(x) = (4x - 16)^{\frac{1}{3}}$

(c) $f'(2) = 3$, $g'(6) = \frac{1}{3}$

2.9.5.4.

(a)

(b)

(c) $(h^{-1})'\left(\frac{\pi}{2} + 1\right) = 1$

3 · Using the Derivative

3.1 · Linear and Quadratic Approximation

3.1.5 · Exercises

3.1.5.1.

(a) $p(3) = -1$, $p'(3) = -2$

(b) $p'(2.79) \approx -0.58$

(c) Overestimate

(d)

(e) Equal approximations

3.1.5.2.

(a) $s(9.34) \approx 3.592$ feet

(b) Underestimate

(c) $Q(9.34) = 3.596624$ feet

(d) Moving towards the origin, slowing down

3.1.5.3.

(a) $x = 1$

(b) $(-0.35, 1.3)$

(c) $f(1.88) \approx -3.0022$, overestimate

(d) $Q(x) = -3 + e^{-4}(x-2) - \dfrac{3}{2}e^{-4}(x-2)^2$

3.2 · The Stability Theorem

3.2.4 · Exercises

3.2.4.3.

(a) $y^* = 0$ (unstable), $y^* = 1,000$ (stable)

(b) $b^* = 1$ (stable), $b^* = 2$ (unstable)

3.3 · The Logistic Discrete-Time Dynamical System

3.3.4 · Exercises

3.3.4.1.

(a) $x^* = 0$ (unstable), $x^* = 0.5$ (stable)

(b)

(c) The solution looks to increase exponentially at first (concave up), but then there is an inflection point where the solution continues to increase but is concave down

(d) $x^* = 0.5$ is a horizontal asymptote in the solution function graph

3.4 · Identifying Extreme Values of Functions

3.4.4 · Exercises

3.4.4.1.

(a) Critical numbers are $x = -1$ (local min) and $x = 1$ (neither)

(b)

(c) f is CCU on $(-\infty, -0.3) \cup (1, \infty)$ and CCD on $(-0.3, 1)$. f has inflection points at $x = -0.3, 1$.

3.4.4.2.

(a) Neither

(b) g'' will change from negative to positive at $x = 2$

(c) It is an inflection point

3.4.4.3.

(a) Inflection points at $x = -1, 2$

(b) Local maximum

(c) $y = \frac{12}{e^2} - \frac{5}{e^2}(x-2)$, neither above nor below the curve

3.5 · Global Optimization and Applications

3.5.4 · Exercises

3.5.4.1.

(a) $t = \dfrac{p}{k}$

(b) No

(c) Yes

3.5.4.2.

(a) Max of $y = 30$ at $x = 1$, Min of $y = 24$ at $x = 2$

(b) Max of $y = 40$ at $x = 6$, Min of $y = 24$ at $x = 2$

(c) Max of $y = 40$ at $x = 6$, Min of $y = 26$ at $x = 3$

3.5.4.3.

(a)

(b) Not possible

(c) Not possible

3.6 · Limits: L'Hôpital's Rule

3.6.4 · Exercises

3.6.4.1. $\lim_{x\to 3} h(x) = -2$

3.6.4.2.

(a)

(b)

(c) $\displaystyle\lim_{x\to 0^+} h(x) = 0$

(d) $\displaystyle\lim_{x\to 0^+} g(x) = 1$

3.6.4.3.

(a) x^3

(b) 5^x

(c) $\sqrt{x}$

(d) $\sqrt[n]{x}$

3.7 · Limits: Leading Behaviors

3.7.4 · Exercises

3.7.4.1.

(a) $k_\infty(x) = e^x$

(b) $\displaystyle\lim_{x\to\infty} \frac{e^x}{e^x + e^{-x}} = 1$

(c) $k_{-\infty}(x) = e^{-x}$

(d) $\lim_{x\to-\infty} \dfrac{e^x}{e^x+e^{-x}} = 0$

4 · Continuous-Time Dynamical Systems

4.1 · Introduction to Differential Equations and Antiderivatives

4.1.4 · Exercises

4.1.4.1. $q(t) = \ln(t) + 0.25t^4 - \frac{2^t}{\ln(2)} + C$

4.1.4.2. $p(t) = \sin(t) + \frac{2}{5}t^{\frac{5}{2}} + 1$

4.2 · Solving Pure-Time Differential Equations

4.2.4 · Exercises

4.2.4.1. $s(t) = -\sin(t) + \frac{1}{6}t^3 - t + 5$

4.2.4.2.

(a) $0.5(x^4 + 4x^2 + C)$

(b) $(\ln(x))^2 + 4\ln(x) + C$

4.2.4.3. $-\ln|\cos(x)| + C$

4.2.4.4.

(a)

(b) $A(t) = 4t - \frac{1}{0.263}\cos(0.263t + 4.7) + \frac{1}{0.526}\sin(0.526t + 9.4) + C$, measured in pounds

4.3 · Riemann Sums

4.3.6 · Exercises

4.3.6.1.

(a) $\Delta x = 0.75$, $L_4 = 40.125$, $R_4 = 46.875$

(b) L_4 is an underestimate and R_4 is an overestimate of the actual area of 43.5

4.3.6.2.

(a) $f(x) = x^2 + 1$ on $[1, 3]$

(b) Left: $f(x) = x^2 + 1$ on $[1.4, 3.4]$

Middle: $f(x) = x^2 + 1$ on $[1.2, 3.2]$

(c) The area under the graph of $x^2 + 1$ from $x = 1$ to $x = 3$

(d) $n = 10$, $\Delta x = 0.2$, $R_{10} = \sum_{i=1}^{10}[(1 + 0.2i)^2 + 1] \cdot 0.2$

4.3.6.3.

(a)

(b) $M_3 = 99.6$ feet

(c) $L_6 = 114$ feet, $R_6 = 84$ feet, $\frac{1}{2}(L_6 + R_6) = 99$ feet

(d) L_6

4.3.6.4.

(a) $M_4 \approx 6.45$

(b) Units of M_4 are tons

(c) $L_5 = 5.196$

(d) $R_4 = 8.119$

4.4 · The Definite Integral

4.4.5 · Exercises

4.4.5.1.

(a) $\displaystyle\int_0^4 v(t)dt$

(b) $\frac{-21}{8}$ feet

(c) $\int_0^{0.5} v(t)dt - \int_{0.5}^{3.5} v(t)dt + \int_{3.5}^4 v(t)dt = \frac{27}{8}$ feet

(d) $\frac{-21}{32}$ feet per second

(e) $s(t) = -t^2 + t$

4.4.5.2.

(a) $\displaystyle\int_0^4 t(t-1)(t-3)dt$

(b) 2.66 feet

(c) $\displaystyle\int_0^1 v(t)dt - \int_1^3 v(t)dt + \int_3^4 v(t)dt$

(d) 8.00 feet

(e) 0.665 feet per second

4.4.5.3.

(a) $1 - \dfrac{\pi}{4}$

(b) $\dfrac{-15}{2} - 3\pi$

(c) $\dfrac{5}{8} + \dfrac{3\pi}{16}$

(d) $c = \dfrac{-3}{8} + \dfrac{3\pi}{16}$

4.4.5.4.

(a) $\displaystyle\int_{-1}^1 3 - x^2 dx$

(b) $\displaystyle\int_{-1}^1 2x^2 dx$

(c) $\displaystyle\int_{-1}^1 3 - x^2 dx - \int_{-1}^1 2x^2 dx$

4.5 · The Fundamental Theorem of Calculus

4.5.5 · Exercises

4.5.5.1.

(a) 20 meters

(b) $\frac{25}{2}$ meters per minute

(c) $t = 2$ minutes

(d) $c = 5$

4.5.5.2.

(a) $m(h) = \frac{1}{c(h)}$, measured in minutes per foot

(b) Input: height in feet, Output: number of minutes

(c) $\displaystyle\int_0^{10,000} m(h)dh$

(d) $M_5 = 15.27$ minutes

4.5.5.3. Yes

4.6 · Approximations of Solutions

4.6.4 · Exercises

4.6.4.1.

(a) Alice: $\frac{dT}{dt} = -15$ degrees F per minute

Bob: $\frac{dT}{dt} = -3$ degrees F per minute

(b) Bob's cup is better insulated

4.6.4.2.

(a)

(b) $\dfrac{d^2y}{dt^2} = 1 - \dfrac{dy}{dt}$

(c) $y(1) \approx 0.7414796816$ using quadratic approximations, which is closer to the actual value of $y(1) = 0.735758882$ than using Euler's method

Appendix B

List of Symbols

Symbol	Description	Page
p_0	the initial value of a quantity p in a DTDS	3
p_t	the value of a quantity p after t steps in a DTDS	3
$f(x)$	the output value associated with an input value of x in the function relationship represented by f	9
$(f \circ g)(x)$	f composed with the function g, or $f(g(x))$	9
$f^{-1}(x)$	the inverse function of an invertible function $f(x)$	12
$AROC_{[a,b]}$	the average rate of change between x values a and b for a given function	14
t_d	the doubling time for a given exponential growth model	??
t_h	the half life for a given exponential decay model	??
b^*	an equilibrium value for a DTDS whose dependent variable is b	47
$\lim_{x\to a} f(x)$	The limit of $f(x)$ as x approaches a	58
$\lim_{x\to a^-} f(x)$	The limit of $f(x)$ as x approaches a from the left	59
$\lim_{x\to a^+} f(x)$	The limit of $f(x)$ as x approaches a from the right	59
$f'(a)$	The derivative of $f(x)$ at the x value a	67
$f'(x)$	The derivative function of $f(x)$ with respect to the variable x	77
$f''(x)$	The second derivative function of $f(x)$ with respect to the variable x	86
$\frac{dy}{dx}$	The derivative function of y with respect to the variable x	94
$\frac{d}{dx}[f]$	The derivative function of f with respect to the variable x	94
$\frac{d^2f}{dx^2}$	The second derivative function of f with respect to the variable x	94
$L(x)$	The linear approximation of a function centered at an x value	128

(Continued on next page)

Symbol	**Description**	**Page**
$Q(x)$	The quadratic approximation of a function centered at an x value	**??**
$\lim_{x\to\infty} f(x)$	The limit of $f(x)$ as x gets arbitrarily large	168
$f_\infty(x)$	The leading behavior of $f(x)$ at ∞	172
$f_{-\infty}(x)$	The leading behavior of $f(x)$ at $-\infty$	173
$\int f(x)dx$	The indefinite integral of $f(x)$ with respect to the variable x	**??**
$\sum_{k=1}^{n} f(k)$	The sum of expressions of the form $f(k)$ as k goes from 1 to n	189
L_n	A left Riemann sum for a function on an interval using n rectangles	191
R_n	A right Riemann sum for a function on an interval using n rectangles	191
M_n	A middle Riemann sum for a function on an interval using n rectangles	192
$\int_a^b f(x)dx$	The definite integral of $f(x)$ as x values range from a to b	199
$f_{\text{AVG}[a,b]}$	The average value of $f(x)$ on the interval $[a,b]$	205
$F(x)\big\|_a^b$	The function $F(x)$ evaluated from a to b: $F(b)-F(a)$	212
$\left.\frac{dy}{dx}\right\|_{(x,y)}$	The derivative of y with respect to x evaluated at the point (x,y)	220
$E_{\Delta t}$	An approximate value of the solution to an initial value problem using Euler's method with a step size of Δt	224

Index

Colophon

This book was authored in PreTeXt.