MA172: Calculus II

Dylan C. Beck

Acknowledgements

Primarily, the contents of this document were created in the Fall 2020 semester at the University of Kansas. Later, during the Spring 2023 semester at Baker University, the notes were overhauled with additional examples. I would like to express my gratitude toward my students in MA172 — especially those who assisted in the enhancement of these notes with comments and suggestions. I extend my thanks also to the user Ignasi on TeX Stack Exchange for some graphics in Chapter 2.

Contents

Chapter 1

Differentiation

Broadly speaking, differential calculus is the study of instantaneous change. Early on in a first calculus course, students learn that the derivative of a function at a point measures the slope of the line tangent at that point; the slope of the tangent line at a point is simply limit of the slopes of the secant lines passing through the specified point, and these slopes measure the average rate of change of the function. Consequently, the derivative measures the instantaneous change of a function. Bearing this in mind, calculus is immediately applicable in a wide range of fields — from physics and engineering to biology, chemistry, and medicine. Conversely, it is the aim of integral calculus to quantify change over time given the instantaneous rate of change. Combined, differential and integral calculus constitute an indispensable tool in many applied sciences today.

1.1 Limits and Continuity

Calculus is the study of change in functions. Essentially, a **function** is simply a rule that assigns to each input x one and only one output $y = f(x)$. Often, in this course, we will simply consider **real functions**, i.e., functions that are defined such that their inputs and outputs are **real numbers**. We are unwittingly very familiar with real numbers: the real numbers $\mathbb{R}$ include zero, all positive and negative whole numbers, all positive and negative rational numbers (or fractions), all positive and negative square roots of positive rational numbers, and transcendental numbers like π and e.

We will use the notation $f : \mathbb{R} \to \mathbb{R}$ to express that f is a function whose **domain** is the real numbers $\mathbb{R}$ and whose **codomain** is the real numbers $\mathbb{R}$. Explicitly, the domain of a function is the set of all possible inputs of a function, and the codomain of a function is the set of all possible outputs of the function. Even more, the collection of all possible outputs of a function is the **range** of the function. We will adopt the **set-builder** notation for the domain and range of a function f.

$D_f = \{x \in \mathbb{R} \mid f(x) \text{ is a real number}\}$ consists of real numbers x such that $f(x)$ is a real number.

$R_f = \{f(x) \in \mathbb{R} \mid x \in D_f\}$ consists of real numbers $f(x)$ such that x lies in the domain of f.

Example 1.1.1. Consider the real function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$. By definition, this function outputs the real number x that is input. We refer to this as the **identity function** on the real numbers. Consequently, the domain of f is $D_f = \mathbb{R}$ because the output of any real number is a real number, and the range of f is $R_f = \mathbb{R}$ because every real number is the output of itself.

Caution: the domain of a real function might not be all real numbers; the range of a real function might not be all real numbers, either, as our next pair of examples illustrate.

Example 1.1.2. Consider the real function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. By definition, this function outputs the square x^2 of the real number x that is input. Certainly, the square of any real number is a real number, hence the domain of f is $D_f = \mathbb{R}$; on the other hand, the only real numbers that are the square of another real number are the non-negative real numbers. Explicitly, for any real number x, the real number $f(x) = x^2$ is a non-negative real number, i.e., we have that $x^2 \geq 0$. Consequently, the codomain of f is $\mathbb{R}$, but the range of f is $R_f = \mathbb{R}_{\geq 0} = \{y \in \mathbb{R} \mid y \geq 0\}$.

Example 1.1.3. Consider the real function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sqrt{x}$. By definition, this function outputs the square root $\sqrt{x}$ of the real number x that is input. We cannot take the square root of a negative real number, hence the domain of f consists of all non-negative real numbers, i.e., we have that $D_f = \mathbb{R}_{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$; on the other hand, every non-negative real number can be realized as the square root of a non-negative real number. Explicitly, for any non-negative real number y, the real number y^2 satisfies that $y = \sqrt{y^2} = f(y^2)$. Consequently, the codomain of f is $\mathbb{R}$, but once again, the range of f is $R_f = \mathbb{R}_{\geq 0} = \{y \in \mathbb{R} \mid y \geq 0\}$.

Generally, the restrictions on the domain of a real function consist of the following situations.

(a.) We cannot divide by zero.

(b.) We cannot take the even root of a negative real number.

(c.) We cannot take the logarithm of a non-positive real number.

Occasionally, it is necessary to split the domain or the range of a function into distinct chunks of the real number line. By the above rule, the domain of the real function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^{-1}$ consists of all nonzero real numbers. Consequently, we can certainly realize the domain of f as $D_f = \{x \in \mathbb{R} \mid x \neq 0\}$, but it is sometimes more convenient to describe this set using the **union** symbol $\cup$. Put simply, the union symbol $\cup$ functions as the logical connective "or." Clearly, a nonzero real number is either positive or negative, hence we can partition the domain of f into those real numbers that are positive and those real numbers that are negative. We achieve this with the union symbol as $D_f = \{x \in \mathbb{R} \mid x > 0\} \cup \{x \in \mathbb{R} \mid x < 0\}$. Even more, we learn in college algebra (or earlier) that the set of real numbers x satisfying the **inequalities** $x > 0$ and $x < 0$ can be described respectively using the **open intervals** $(0, \infty)$ and $(-\infty, 0)$. Consequently, in **interval notation**, the domain of the real function $f(x) = x^{-1}$ is given by $D_f = (-\infty, 0) \cup (0, \infty)$.

Exercise 1.1.4. Compute the domain and range of the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$.

Exercise 1.1.5. Compute the domain and range of the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^{-3}$.

Exercise 1.1.6. Compute the domain and range of the function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x) = \frac{1}{\ln(x)}$.

Consider a function $f : \mathbb{R} \to \mathbb{R}$ whose domain is D_f. Given any real number a in D_f, we say that the **limit** of $f(x)$ as x approaches a is the quantity L (if it exists) such that for every real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that $|x - a| < \delta$ implies that $|f(x) - L| < \varepsilon$. Put another way, the quantity L can be made arbitrarily close to the value of $f(x)$ by taking x to be sufficiently close in value to a. Conveniently, if the quantity L exists, then we write $L = \lim_{x \to a} f(x)$.

Example 1.1.7. Let us compute the limit of $f(x) = x^2$ as x approaches $a = 1$ using the definition. Computing the limit is essentially like playing a game of limbo: we are handed a real number $\varepsilon > 0$ (the limbo bar), and our challenge is to find a real number $\delta > 0$ such that $|x^2 - 1| < \varepsilon$ whenever we assume that $|x - 1| < \delta$. Of course, we are at liberty to take δ as small as necessary to ensure that $|x^2 - 1| < \varepsilon$. We may therefore assume that $0 < \delta \leq 1$. Considering that $x^2 - 1 = (x - 1)(x + 1)$, if we assume that $|x - 1| < \delta \leq 1$, then we must have that $0 < x < 2$, from which it follows that $|x + 1| \leq |x| + 1 = x + 1 < 3$ by the Triangle Inequality. Consequently, we have that

$$|x^2 - 1| = |(x - 1)(x + 1)| = |x - 1||x + 1| < 3\delta.$$

Last, if we wish to have that $|x^2 - 1| < \varepsilon$, then we should choose δ to be the minimum of 1 and $\frac{\varepsilon}{3}$.

One-sided limits can be defined analogously to the limit above: the **left-hand limit** of $f(x)$ as x approaches a is the quantity L^- (if it exists) such that for every real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that $-\delta < x - a < 0$ implies that $|f(x) - L^-| < \varepsilon$. Likewise, the **right-hand limit** of $f(x)$ as x approaches a is the quantity L^+ (if it exists) such that for every real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that $0 < x - a < \delta$ implies that $|f(x) - L^+| < \varepsilon$.

$L^- = \lim_{x \to a^-} f(x)$ is the symbolic way to express the left-hand limit of $f(x)$ as x approaches a.

$L^+ = \lim_{x \to a^+} f(x)$ is the symbolic way to express the right-hand limit of $f(x)$ as x approaches a.

Ultimately, the two-sided limit exists if and only if the left- and right-hand limits exist and are equal; thus, the two-sided limit is equal to the common value of the left- and right-hand limits.

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

Graphically, it is possible to compute the two-sided limit L of some functions $f(x)$ as x approaches a by tracing one's finger along the graph of $f(x)$ from the left- and right-hand sides.

Example 1.1.8. Let us graphically compute the limit of $f(x) = x^2$ as x approaches $a = 1$. Using the graph of $f(x) = x^2$, we find that the limit is 1. Particularly, if we trace the graph with our left pointer finger, moving from left to right toward the point $x = 1$, our finger stops at $y = f(1) = 1$. Likewise, if we trace the graph with our right pointer finger moving from right to left toward $x = 1$, our finger stops at $y = f(1) = 1$. Put in the language of calculus, we have that $L^- = 1 = L^+$.

We say that a function $f : \mathbb{R} \to \mathbb{R}$ is **continuous** at a real number a if and only if

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

Explicitly, we require three things to be true of the function $f(x)$ in this case.

1.) We must have that f is defined at the real number a, i.e., $f(a)$ must be in the range of f.

2.) We must have that $\lim_{x \to a^-} f(x) = f(a)$, i.e., the left-hand limit of f at a must be $f(a)$.

3.) We must have that $\lim_{x \to a^+} f(x) = f(a)$, i.e., the right-hand limit of f at a must be $f(a)$.

Consequently, if any of these criteria is violated, then the function f cannot be continuous at a.

Example 1.1.9. One of the easiest ways to detect that a function is not continuous at a real number a is to observe that the function is not defined at a. Explicitly, the function $f(x) = \frac{1}{x}$ is not continuous at $a = 0$ because the domain of f excludes $a = 0$ (since we cannot divide by zero).

Example 1.1.10. Consider the function $f : \mathbb{R} \to \mathbb{R}$ that is defined **piecewise** as follows.

$$f(x) = \begin{cases} 1 & \text{if } x \geq 0 \text{ and} \\ -1 & \text{if } x < 0 \end{cases}$$

Graphically, if we trace our fingers along f from the left-hand side, when we arrive at $a = 0$ from the left-hand side, we find that the limiting value here is -1; however, if we trace our fingers along f from the right-hand side, when we arrive at $a = 0$ from the right-hand side, we find that the limiting value here is 1. Consequently, the function $f(x)$ is not continuous at $a = 0$.

Example 1.1.11. Let us prove by definition that the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$ is continuous for all real numbers a. Observe that f is defined piecewise as follows.

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \text{ and} \\ -x & \text{if } x < 0 \end{cases}$$

Consequently, it suffices to show that $g(x) = x$ and $h(x) = -x$ are everywhere continuous. Given real numbers $\varepsilon_1, \varepsilon_2 > 0$, we must find real numbers $\delta_1, \delta_2 > 0$ such that $|x - a| < \varepsilon_1$ whenever $|x - a| < \delta_1$ and $|-x - (-a)| < \varepsilon_2$ whenever $|x - a| < \delta_2$. Considering that the absolute value is multiplicative, we have that $|-x - (-a)| = |-x + a| = |-(x - a)| = |x - a|$, we may simply take the real numbers $\delta_1 = \varepsilon_1$ and $\delta_2 = \varepsilon_2$. We conclude that $g(x) = x$ and $h(x) = -x$ are continuous for all real numbers a so that $f(x) = |x|$ is continuous for all nonzero real numbers by the piecewise definition of $f(x)$ prescribed above. We are done as soon as we show that

$$\lim_{x \to 0^-} |x| = \lim_{x \to 0^-} f(x) = 0 = \lim_{x \to 0^+} f(x) = \lim_{x \to 0^+} |x|.$$

By continuity of the functions $g(x)$ and $h(x)$ and by definition of $|x|$, the left-hand limit is given by $\lim_{x \to 0^-} |x| = \lim_{x \to 0^-} h(x) = h(0) = 0$, and the right-hand limit is $\lim_{x \to 0^+} |x| = \lim_{x \to 0^+} g(x) = g(0) = 0$.

Generally, continuity can be defined as a property of a function on any **subset** of its domain, i.e., on any collection of real numbers that lie in the domain. Often, we will consider functions that are continuous on their entire domain, but it is possible that a function is not continuous at some point in its domain. We say that a function f is **discontinuous** at a real number a if f is not continuous at the real number a. By the above three criteria, we can classify these **discontinuities**.

- We say that f has a **removable discontinuity** at a real number a if a is not in the domain of f but the left- and right-hand limits of f at a exist and are equal, i.e., $\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x)$.
- We say that f has a **jump discontinuity** at a real number a if both of the left- and right-hand limits of f at a exist but are not equal, i.e., $\lim_{x \to a^-} f(x) = L^- \neq L^+ = \lim_{x \to a^+} f(x)$.
- We say that f has an **essential discontinuity** at a real number a if either the left- or the right-hand limit of f at a does not exist, i.e., either $\lim_{x \to a^-} f(x)$ or $\lim_{x \to a^+} f(x)$ does not exist.

Often, if a function f is continuous for every real number in its domain D_f, we will say that the function is **continuous** on its domain. Explicitly, if the domain of a function f is all real numbers and f is continuous on its domain, then we will say that f is **everywhere continuous**. Graphically, we may detect that a function is continuous if we can draw it without lifting our pencil.

Example 1.1.12. We can graph $|x|$ without lifting our pencil, hence it is everywhere continuous.

Example 1.1.13. We cannot graph x^{-2} without lifting our pencil at $x = 0$, hence x^{-2} is not continuous at $a = 0$. On the other hand, for all real numbers a other than $a = 0$, we can graph this function without lifting our pencil, hence x^{-2} is continuous on its domain $(-\infty, 0) \cup (0, \infty)$.

Continuous functions abound: **polynomial** functions such as $x^3 - 2x^2 + x - 7$ and **exponential** functions such as e^x are defined for all real numbers and are everywhere continuous. Likewise, the **trigonometric** functions $\sin(x)$ and $\cos(x)$ are defined for all real numbers and are everywhere continuous. **Logarithmic** functions such as $\ln(x)$ and $\log(x)$ and **algebraic** functions such as $\sqrt{x}$ and $x^{3/2}$ are defined for all positive real numbers and are continuous on their domains. Further, addition, subtraction, multiplication, division, composition, and any finite combination of these operations on continuous functions result in functions that are typically continuous on their domains.

1.2 Differentiation and L'Hôpital's Rule

Given any real numbers a and $h > 0$ and any real function $f(x)$ such that $f(a)$ and $f(a+h)$ are defined, consider the closed interval $[a, a+h]$ consisting of all real numbers x with $a \leq x \leq a+h$. We define the **secant line** of $f(x)$ over this interval as the line passing through the points $(a, f(a))$ and $(a+h, f(a+h))$. Observe that the slope of the secant line is given by the **difference quotient**

$$Q_a(h) = \frac{f(a+h) - f(a)}{(a+h) - a} = \frac{f(a+h) - f(a)}{h}.$$

By taking the limit of $Q_a(h)$ as h approaches 0, we obtain the **derivative** of $f(x)$ at a

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

Of course, this limit might not exist; however, if it does, we interpret it geometrically as the slope of the line tangent to $f(x)$ at the point $(a, f(a))$. Given that the quantity $f'(a)$ exists, we say that $f(x)$ is **differentiable** at a. One fundamental interpretation of the derivative in the context of a function that measures something physical (e.g., velocity) is as the instantaneous rate of change.

Exercise 1.2.1. Use the limit definition of the derivative to compute $f'(x)$ for $f(x) = x^3$.

Exercise 1.2.2. Use the limit definition of the derivative to compute $g'(x)$ for $g(x) = \frac{1}{x}$.

Exercise 1.2.3. Use the limit definition of the derivative to compute $h'(x)$ for $h(x) = \sqrt{x}$.

One of the most important properties of differentiable real functions is the following.

Proposition 1.2.4. *If a real function f is differentiable at a real number a, then f is continuous at a. Explicitly, a function that is differentiable at a point in its domain is necessarily continuous there. Conversely, there exists a function that is continuous but not differentiable on its domain.*

Proof. We will assume that f is differentiable at a real number a. Consequently, the limit

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

exists. Using the substitution $x = a + h$, we have that $h = x - a$. Crucially, under this substitution, the limit of any function $g(h)$ as h approaches 0 is equal to the limit of the function $g(x - a)$ as x approaches a. (Verify this by definition of the limit.) Consequently, the following identity holds.

$$f'(a) = \lim_{x \to a} Q_a(x - a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

Considering that $x - a$ is a polynomial function, it is continuous at a, and we conclude that

$$\lim_{x \to a}(x - a) = a - a = 0.$$

Using the fact that the limit of a product is the product of limits (when both limits exist),

$$0 = f'(a) \cdot \lim_{x \to a}(x - a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \cdot \lim_{x \to a}(x - a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \cdot x - a = \lim_{x \to a}[f(x) - f(a)]$$

yields the result that $\lim_{x \to a} f(x) = \lim_{x \to a}[f(a) + f(x) - f(a)] = \lim_{x \to a} f(a) + \lim_{x \to a}[f(x) - f(a)] = f(a)$.

Conversely, the function $|x|$ is continuous on its domain, but it is not differentiable at $a = 0$: indeed, by Example 1.1.10, the piecewise function $f(x)$ satisfying that $f(x) = 1$ for $x \geq 0$ and $f(x) = -1$ for $x < 0$ is not continuous because the left- and right-hand limits do not agree at 0. One can readily verify that this function is exactly the derivative of $|x|$, hence the claim holds. □

Computing limits by definition is even more tedious than it looks, but luckily, there are plenty of tools that allow us to compute derivatives of functions without ever touching a limit. Particularly,

- the **Power Rule** says that if $f(x) = x^r$ for some real number r, then $f'(x) = rx^{r-1}$;
- the **Product Rule** says that if $f(x)$ and $g(x)$ are both differentiable, then

$$\frac{d}{dx}[f(x) \cdot g(x)] = f'(x)g(x) + f(x)g'(x);$$

- the **Quotient Rule** says that if $f(x)$ and $g(x)$ are both differentiable, then

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}; \text{ and}$$

- the **Chain Rule** says that if $f(x)$ and $g(x)$ are both differentiable, then

$$\frac{d}{dx}[f \circ g(x)] = \frac{d}{dx}[f(g(x))] = f'(g(x)) \cdot g'(x) = [f' \circ g(x)] \cdot g'(x).$$

Computing the limit of a function that is continuous is quite easy: we may simply "plug and chug;" however, there exist functions that are not continuous. Even worse, when evaluating limits, we can encounter situations that result in an **indeterminate form** when the limit is the form

$$\frac{0}{0} \text{ or } \frac{\infty}{\infty}.$$

Theorem 1.2.5 (L'Hôpital's Rule)**.** *Given any real functions $f(x)$ and $g(x)$ that are differentiable for all real numbers x such that $a < x < b$ (with the possible exception of one point $x = c$ for some real number $a \leq c \leq b$), consider the following conditions.*

(1.) *We have that* $\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0$ *or* $\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = \pm\infty$.

(2.) *We have that* $g'(x) \neq 0$ *for any real number x such that $a < x < b$ and $x \neq c$.*

(3.) *We have that* $\lim_{x \to c} \frac{f'(x)}{g'(x)}$ *exists.*

Granted that each of the above conditions holds, it follows that $\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$.

Exercise 1.2.6. Compute the limit of $f(x) = \dfrac{\ln(x)}{x^3 - 1}$ as x approaches $a = 1$.

Exercise 1.2.7. Compute the limit of $g(x) = (2x - \pi)\sec(x)$ as x approaches $a = \frac{\pi}{2}$ from the left.

Exercise 1.2.8. Compute the limit of $h(x) = \dfrac{\sin(x)}{\sin(x) + \tan(x)}$ as x approaches $a = 0$.

Exercise 1.2.9. If $\frac{d}{dx}\sin(x) = \cos(x)$, compute the limit of $f(x) = \dfrac{\sin(x)}{x}$ as x approaches $a = 0$.

Caution: Unfortunately, the above example is not a valid proof of this limit identity: in fact, this limit identity is needed to prove that $\frac{d}{dx}\sin(x) = \cos(x)$, so in order to prove this identity in a rigorous and non-circular manner, we must use tools from trigonometry and the Squeeze Theorem.

1.3 Implicit Differentiation

Curves in the Cartesian plane can be represented by an equation involving a function of two variables. Explicitly, we are familiar with such curves as $xy = 1$ and $y - x^2 = 0$; they are respectively the functions $y = f(x) = x^{-1}$ and $y = g(x) = x^2$. We refer to the functions $f(x)$ and $g(x)$ as the **explicit** forms of the curves. Unfortunately, it is not possible to write every curve in the Cartesian plane as a function of one variable: curves such as the unit circle $x^2 + y^2 = 1$ or the hyperbola $y^2 - x = 0$ cannot be represented as functions because they fail the **Vertical Line Test**; however, we will see throughout this semester that these curves provide important models in calculus. Curves that do not admit closed-form expressions of the form $y = f(x)$ can be written **implicitly**.

Under certain conditions, it is possible to find a "small enough" region in the Cartesian plane in which an implicit curve can be represented by a function; thus, in this "window," the slope and tangent line of such curves are well-defined. Consequently, we may define the **implicit derivative** by assuming that y is a function of x (on some "small window" in the plane) with derivative $y' = \frac{dy}{dx}$.

Example 1.3.1. Compute $\frac{dy}{dx}$ for the unit circle $x^2 + y^2 = 1$.

Solution. Considering the variable y as some function $y = f(x)$ of x and using the convention that $y' = \frac{dy}{dx}$, we may invoke the Chain Rule in order to determine that

$$0 = \frac{d}{dx}1 = \frac{d}{dx}(x^2 + y^2) = 2x + 2yy'.$$

Crucially, each time the derivative operator $\frac{d}{dx}$ encounters the variable y, we differentiate y as we would the function $y = f(x)$ that represents y locally. Consequently, if y is nonzero, then

$$\frac{dy}{dx} = y' = -\frac{2x}{2y} = -\frac{x}{y}.$$

Otherwise, the tangent line does not exist if $y = 0$ because $2x + 2yy' = 0$ has no solution if $y = 0$. $\diamond$

Example 1.3.2. Compute $\frac{dy}{dx}$ for the parabola $y^2 - x = 0$.

Solution. By the Chain Rule applied to $y = f(x)$, we have that

$$0 = \frac{d}{dx} 0 = \frac{d}{dx}(y^2 - x) = 2yy' - 1$$

so that $\frac{dy}{dx} = y' = (2y)^{-1}$ for all points (x, y) on the hyperbola such that y is nonzero. $\diamond$

1.4 Exponential and Logarithmic Functions

Given any positive real number a, the **exponential** function with **base** a is given by $\exp_a(x) = a^x$. Crucially, the most important exponential function is simply $\exp(x) = e^x$: here, the base is **Euler's number** $e \approx 2.72$. Later, we will concern ourselves with the definition of Euler's number; for now, we need only recall the following properties of exponential functions for any real numbers x and y.

1.) $a^{x+y} = a^x a^y$

2.) $a^{x-y} = a^x a^{-y}$

3.) $a^{xy} = (a^x)^y$

4.) $(ab)^x = a^x b^x$ for any real number $b > 0$

We do not yet have the machinery available to use to prove the following, but it is true that

$$\frac{d}{dx} e^x = e^x.$$

Considering that $e^x > 0$ for all real numbers x, it follows that e^x is a strictly increasing function, hence it passes the **Horizontal Line Test** and must therefore admit an **inverse** function; we refer to this function as the **natural logarithmic** function $\ln(x)$. Put another way, we have that

$$e^{\ln(x)} = x \text{ for all real numbers } x > 0 \text{ and } \ln(e^x) = x \text{ for all real numbers } x.$$

Observe that the range of e^x is $(0, \infty)$, hence the domain of $\ln(x)$ is $(0, \infty)$. Conversely, the domain of e^x is $(-\infty, \infty)$, hence the range of $\ln(x)$ is $(-\infty, \infty)$. We will also simply assert that

$$\frac{d}{dx} \ln|x| = \frac{1}{x}.$$

We may also deduce the following properties of logarithmic functions for any real numbers $x, y > 0$.

1.) $\log_a(xy) = \log_a(x) + \log_a(y)$

2.) $\log_a(x^r) = r \log_a(x)$ for all real numbers r

3.) $\log_a(xy^{-1}) = \log_a(x) - \log_a(y)$

4.) $\log_a(x) = \frac{\ln(x)}{\ln(a)}$

Even more, for any real number $a > 0$, the exponential function $\exp_a(x) = a^x$ is differentiable for all real numbers x. Further, observe that $y = a^x$ is strictly positive for all real numbers x, hence the function $\ln(y) = x\ln(a)$ is well-defined. Using the Chain Rule, we find that

$$\frac{1}{y} \cdot y' = \frac{d}{dx}\ln(y) = \frac{d}{dx}[x\ln(a)] = \ln(a) \cdot \frac{d}{dx}x = \ln(a) \text{ and } \frac{d}{dx}a^x = \frac{d}{dx}y = \frac{dy}{dx} = y' = y\ln(a) = a^x\ln(a).$$

By a similar rationale as before, one can define the **logarithmic** function $\log_a(x)$ **base** a for any positive real number a as the function inverse of a^x; its domain is $(0, \infty)$, and its range is $(-\infty, \infty)$.

Exercise 1.4.1. Compute the derivative of $y = \log_a(x)$ by using the fact that $a^y = x$.

1.5 Inverse Trigonometric Functions

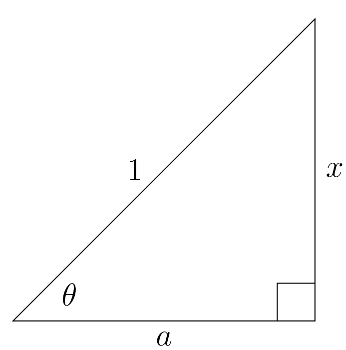
Even though the trigonometric functions like $\sin(x)$, $\cos(x)$, and $\tan(x)$ are **periodic**, we can find a region on the x-axis in which these functions pass the Horizontal Line Test and admit function inverses. Explicitly, the inverse trigonometric functions are denoted as follows.

$$\begin{array}{lll} \arcsin(x) = \sin^{-1}(x) & \text{domain: } [-1, 1] & \text{range: } \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ \arccos(x) = \cos^{-1}(x) & \text{domain: } [-1, 1] & \text{range: } [0, \pi] \\ \arctan(x) = \tan^{-1}(x) & \text{domain: } (-\infty, \infty) & \text{range: } \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \end{array}$$

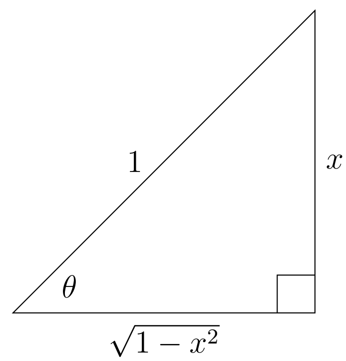
Considering that the input of the sine function is an angle, the output of the arcsine function is an angle. Consequently, if $x = \sin(\theta)$, then it follows by definition that $\theta = \arcsin(x)$ so that

$$\frac{d}{dx}\arcsin(x) = \frac{d\theta}{dx}.$$

Observe that $\sin(\theta)$ is the ratio of the opposite side and the hypotenuse of a right triangle, so we may construct a right triangle whose opposite side has length x and whose hypotenuse has length 1 in order to obtain $\sin(\theta) = x$. Our right triangle therefore has the following form.



By the **Pythagorean Theorem**, we must have that $x^2 + a^2 = 1$ so that $a = \sqrt{1 - x^2}$.

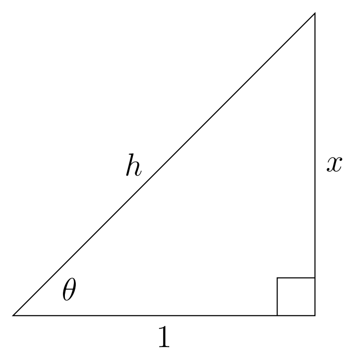


Using the Chain Rule, we can compute $\frac{d\theta}{dx}$. Explicitly, we have that

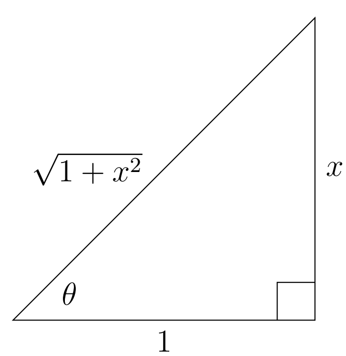
$$\cos(\theta) \cdot \frac{d\theta}{dx} = \frac{d}{dx}\sin(\theta) = \frac{d}{dx}x = 1 \text{ so that } \frac{d}{dx}\arcsin(x) = \frac{d\theta}{dx} = \frac{1}{\cos(\theta)} = \frac{1}{\sqrt{1-x^2}}.$$

Exercise 1.5.1. Use a right triangle involving 1, x, and $\sqrt{1-x^2}$ to compute $\frac{d}{dx}\arccos(x)$.

Using a similar idea as the one we employed to compute the derivative of $\arcsin(x)$ and $\arccos(x)$, we will set up a triangle with $\tan(\theta) = x$. Observe that $\tan(\theta)$ is the ratio of the opposite side and the adjacent side of a right triangle, so we may construct a right triangle whose opposite side has length x and whose adjacent side has length 1 in order to obtain $\tan(\theta) = x$.



Once again, by the Pythagorean Theorem, we find that $h^2 = x^2 + 1^2$ so that $h = \sqrt{1+x^2}$.

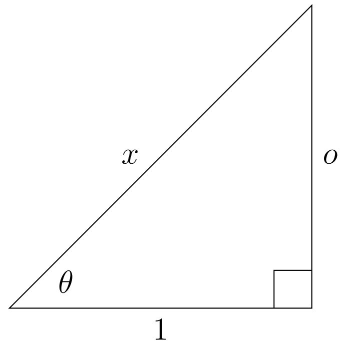


Using the Chain Rule, we can compute $\frac{d\theta}{dx}$. Explicitly, we have that

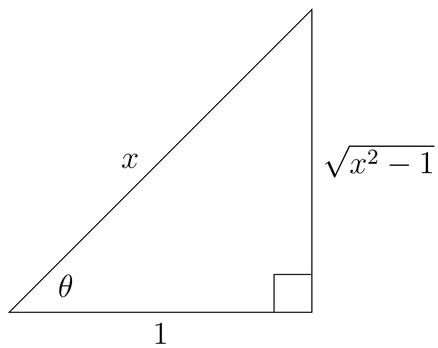
$$\sec^2(\theta) \cdot \frac{d\theta}{dx} = \frac{d}{dx}\tan(\theta) = \frac{d}{dx}x = 1 \text{ so that } \frac{d}{dx}\arctan(x) = \frac{d\theta}{dx} = \cos^2(\theta) = \frac{1}{1+x^2}.$$

Exercise 1.5.2. Use a right triangle involving 1, x, and $\sqrt{1+x^2}$ to compute $\frac{d}{dx}\operatorname{arccot}(x)$.

Last but not least, we will set up a triangle with $\sec(\theta) = x$. Observe that $\sec(\theta)$ is the ratio of the hypotenuse to the adjacent side of a right triangle, so we obtain the following diagram.



Once again, by the Pythagorean Theorem, we find that $x^2 = o^2 + 1^2$ so that $o = \sqrt{x^2-1}$.



Using the Chain Rule, we can compute $\dfrac{d\theta}{dx}$. Explicitly, we have that

$$\sec(\theta)\tan(\theta)\cdot\frac{d\theta}{dx} = \frac{d}{dx}\sec(\theta) = \frac{d}{dx}x = 1 \text{ so that } \frac{d}{dx}\operatorname{arcsec}(x) = \frac{d\theta}{dx} = \cos(\theta)\cot(\theta) = \frac{1}{x\sqrt{x^2-1}}.$$

Exercise 1.5.3. Use a right triangle involving 1, x, and $\sqrt{x^2-1}$ to compute $\frac{d}{dx}\operatorname{arccsc}(x)$.

Chapter 2

Integration

2.1 Antidifferentiation

Considering that a derivative is a rate of change, it is natural in the applied sciences to begin with a rate of change and use it to estimate the net change of a process over time. Explicitly, if we observe that the velocity of a body is given by a function $f(x)$ over some interval of time, then we may seek a function $F(x)$ such that $F'(x) = f(x)$ over this interval of time. Given that such a function $F(x)$ exists and satisfies that $F'(x) = f(x)$, we refer to $F(x)$ as an **antiderivative** of $f(x)$.

Exercise 2.1.1. Prove that the function $F(x) = \frac{1}{3}x^3$ is an antiderivative of $f(x) = x^2$.

Exercise 2.1.2. Prove that the function $G(x) = x\ln(x) - x$ is an antiderivative of $g(x) = \ln(x)$.

Exercise 2.1.3. Prove that the function $H(x) = xe^x - e^x$ is an antiderivative of $h(x) = xe^x$.

Observe that for any antiderivative $F(x)$ of a function $f(x)$, there exists a family of antiderivatives indexed by the real numbers. Particularly, the function $G(x) = F(x) + C$ is an antiderivative of $f(x)$ for every real number C. Even more, by the **Mean Value Theorem**, every antiderivative of $f(x)$ is of the form $F(x) + C$ for some antiderivative $F(x)$ of $f(x)$ and some real number C. Consequently, we may define the **general antiderivative** or **indefinite integral** of $f(x)$ to be

$$\int f(x)\,dx = F(x) + C$$

for any real number C. By the familiar derivative rules, we obtain

- the **Power Rule**, i.e., $\int x^r\,dx = \frac{1}{r+1}x^{r+1} + C$ for all real numbers $r \neq -1$ and
- the **Chain Rule**, i.e., $\int f'(g(x))g'(x)\,dx = f(g(x)) + C$.

Further, indefinite integration is **linear**: for all real functions $f(x)$ and $g(x)$, we have

- the **Multiples Rule** $\int kf(x)\,dx = k\big(\int f(x)\,dx\big)$ for all real numbers k and
- the **Sum Rule** $\int [f(x) + g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx$

Exercise 2.1.4. Compute the indefinite integral of $f(x) = x^{-1}$.

Exercise 2.1.5. Compute the indefinite integral of $g(x) = 2xe^{x^2}$.

Exercise 2.1.6. Compute the indefinite integral of $h(x) = 2\sin(x)\cos(x)$.

Circling back to the opening remarks of this section, we will assume that the velocity of a body over an interval of time is a continuous function $v(t)$. Even more, suppose that we note the position $s(t)$ of the particle at time $t = 0$, i.e., the quantity $s(0)$ is known. Considering that $s'(t) = v(t)$, it follows that $s(t)$ must differ from $\int v(t)\,dt$ by a constant C that depends on the quantity $s(0)$. We refer to this scenario as an **initial value problem** of the **differential equation** $s'(t) = v(t)$.

Example 2.1.7. Consider the velocity function $v(t) = 3t^2 - 4t + 2$ of a body whose position $s(t)$ at time $t = 0$ is given by $s(0) = 7$. Give an explicit formula for $s(t)$.

Solution. Observe that $s(t) = \int v(t)\,dt = \int 3t^2\,dt - \int 4t\,dt = \int 2\,dt = t^3 - 2t^2 + 2t + C$. By plugging in our initial value of $s(0) = 7$, we find that $7 = s(0) = C$ so that $s(t) = t^3 - 2t^2 + 2t + 7$. ◇

Exercise 2.1.8. Consider tossing a ball upward with an initial velocity of 48 feet per second and constant acceleration of -32 feet per second from the edge of a cliff of height 432 feet. Compute the maximum height of the ball; then, find the time it takes for the ball to reach the ground.

2.2 Computing Area Bounded by a Curve of One Variable

Continuing in the theme of extrapolating data from intermittent observations, suppose that we observe the velocity $v(t)$ of a particle over a period of time $0 \leq t \leq 25$, taking care to mark down the velocity of the particle every five seconds. Consider along these lines the following table.

t	0	5	10	15	20	25
$v(t)$	25	31	35	43	47	46

We can roughly approximate the total distance traveled by the body for $0 \leq t \leq 25$ by assuming (incorrectly) that the body maintains a constant velocity each time we see it. Computing the total distance travelled by the particle during our observation amounts to finding the **displacement** of the body over each time interval and adding these quantities together. Explicitly, we have that

$$\text{total distance traveled} = 25 \cdot 5 + 31 \cdot 5 + 35 \cdot 5 + 43 \cdot 5 + 47 \cdot 5 + 46 \cdot 5 = 1135.$$

Certainly, we can improve this estimation by taking more measurements: even recording one more observation will give us a better understanding of the behavior of the particle over the specified interval of time. Better yet, the more observations we record, the more accurate our understanding of the total distance traveled; however, this also requires adding more numbers together. Consequently, it will be convenient to develop notation to take sums of arbitrarily large quantities of data.

Let us assume for the moment that we have a collection of n real numbers $a_1, a_2, \dots, a_n$ for some positive integer n. Certainly, the sum of these real numbers can be realized as

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n.$$

We refer to this as **sigma notation**: indeed, the Greek letter sigma Σ is used as a mnemonic device for "sum"; the subscript $i = 1$ denotes the **index of summation** and informs us of the first term

a_1 in our collection of data; and the superscript n tells us that the sum terminates with the last term a_n in our collection of data. We refer to the real number a_i as the ith **summand** for each integer $1 \leq i \leq n$; the entire sum $\sum_{i=1}^n a_i = a_1 + a_2 + \cdots + a_n$ is called a **finite sum**.

Often, we will consider finite sums whose ith summand can be conveniently expressed in **closed-form**. Explicitly, this means that there exists a function $f(x)$ such that $a_i = f(i)$.

Example 2.2.1. Consider the finite sum $1 + 2 + 3 + \cdots + 10$ of the first ten positive integers. Observe that the ith summand is simply the positive integer i, hence we have that $a_i = i$ and

$$1 + 2 + 3 + \cdots + 10 = \sum_{i=1}^{10} i.$$

Crucially, we point out another way to **index** the given sum — namely, we have that

$$\sum_{i=1}^{10} i = 1 + 2 + 3 + \cdots + 10 = 0 + 1 + 2 + 3 + \cdots + 10 = \sum_{i=0}^{10} i.$$

Often, if a sum involves a summand of zero, we will simply omit it (unless it is more convenient to include it). We could have also written this sum in a third way as follows.

$$\sum_{i=1}^{10} i = 1 + 2 + 3 + \cdots + 10 = (1 + 2 + 3 + \cdots + 20) - (11 + 12 + 13 + \cdots + 20) = \sum_{i=1}^{20} i - \sum_{i=11}^{20} i.$$

Example 2.2.2. Consider the finite sum $1 + 4 + 9 + \cdots + 100$ of squares of the first ten positive integers in which the ith summand is simply the positive integer i^2. We have that $a_i = i^2$ and

$$1 + 4 + 9 + \cdots + 100 = \sum_{i=1}^{10} i^2.$$

Example 2.2.3. Express the finite sum $1^3 + 2^3 + 3^3 + \cdots + 1000^3$ of cubes of the first 1000 positive integers in summation notation, identifying the closed-form expression for the ith summand a_i.

Quite importantly, finite sums admit a convenient arithmetic of their own.

Proposition 2.2.4 (Properties of Finite Sums)**.** *Given any positive integer n and any real numbers $a_1, \dots, a_n, b_1, \dots, b_n$, and C, the following identities hold.*

(i.) (Empty Sum Law) *We have that $\sum_{i=n}^m a_i = 0$ for all integers $m < n$.*

(ii.) (Constant Sum Formula) *We have that $\sum_{i=m}^n C = C(n - m + 1)$ for all integers $m \leq n$.*

(iii.) (Linearity of a Finite Sum I) *We have that $\sum_{i=1}^n Ca_i = C(\sum_{i=1}^n a_i)$.*

(iv.) (Linearity of a Finite Sum II) *We have that $\sum_{i=1}^n (a_i + b_i) = \sum_{i=1}^n a_i + \sum_{i=1}^n b_i$.*

One can easily prove the above formulas by expanding and comparing the expressions on both sides of the equation. We will not endeavor to prove the following identities because these details are beyond the scope of this course; however, they will be indispensable in what follows.

Proposition 2.2.5. *Consider any positive integer n.*

(i.) *We have that* $\displaystyle\sum_{i=1}^n i = \frac{n(n+1)}{2}$.

(ii.) *We have that* $\displaystyle\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$.

(iii.) *We have that* $\displaystyle\sum_{i=1}^n i^3 = \left[\frac{n(n+1)}{2}\right]^2$.

Going back to our example of tracking a particle over a period of time, if we know the velocity $v(t)$ of the particle at any time $0 \leq t \leq 25$, then we can approximate the total distance traveled by the particle by recording the velocity a positive integer n times and computing the total displacement of the particle over each interval of time. Explicitly, if we observe the particle for some real numbers $0 = t_0 < t_1 < \cdots < t_n = 25$ and we assume that the particle has constant velocity $v(t_i)$ for each integer $0 \leq i \leq n$, then the total distance traveled by the particle between time t_{i-1} and time t_i is given by the real number $\Delta t_i = t_i - t_{i-1}$ and the total displacement of the particle on this closed interval $[t_{i-1}, t_i]$ is $v(t_i)\,\Delta t_i$ (rate $\times$ time). Consequently, in sigma notation, we have that

$$\text{total distance traveled} = v(t_1)\,\Delta t_1 + v(t_2)\,\Delta t_2 + \cdots + v(t_n)\,\Delta t_n = \sum_{i=1}^n v(t_i)\,\Delta t_i.$$

By viewing the points $(t_i, v(t_i))$ as lying on the graph of the velocity curve $v(t)$, we may recognize $\sum_{i=1}^n v(t_i)\,\Delta t_i$ as an approximation of the area between the curve $v(t)$ and the t-axis, i.e., the net area bounded by the curve $v(t)$ of one variable. We will now generalize this idea.

Consider any real function $f(x)$ that is continuous on a closed and bounded interval $[a, b]$. Choose any positive integer n; then, choose n real numbers $a = x_0 < x_1 < \cdots < x_n = b$. Consider the closed and bounded intervals $[x_{i-1}, x_i]$ for each integer $1 \leq i \leq n$. We refer to the collection $\mathcal{P}$ of such closed and bounded intervals as a **partition** of $[a, b]$, and we denote by $\Delta x_i = x_i - x_{i-1}$ the length of the interval $[x_{i-1}, x_i]$. Choosing **sample points** x_i^* such that $x_{i-1} \leq x_i^* \leq x_i$ yields a so-called **tagged partition** $(\mathcal{P}, x_i^*)$ consisting of closed and bounded intervals and sample points within them. We associate to each tagged partition a **Riemann sum** (or **Riemann approximation**)

$$\sum_{i=1}^n f(x_i^*)\,\Delta x_i = f(x_1^*)\,\Delta x_1 + f(x_2^*)\,\Delta x_2 + \cdots + f(x_n^*)\,\Delta x_n.$$

Geometrically, we may realize $f(x_i^*)$ as the height of a rectangle with base Δx_i, hence the above Riemann sum provides an approximation of the **net area** bounded by the curve $f(x)$ over the closed interval $[a, b]$. Common tagged partitions are formed by taking x_i^* to be the left- or right-**endpoint** or the **midpoint** of $[x_{i-1}, x_i]$. Each of these tagged partitions uses $n + 1$ **equally-spaced** points $a = x_0 < x_1 < \cdots < x_n = b$; the common length of each interval $[x_{i-1}, x_i]$ is Δx. Considering that

$$b - a = \Delta x_1 + \Delta x_2 + \cdots + \Delta x_n = \sum_{i=1}^n \Delta x_i = \sum_{i=1}^n \Delta x = n\,\Delta x$$

by the second part of Proposition 2.2.4, we conclude that $\Delta x = \frac{b-a}{n}$.

- We denote by $\mathcal{L}_n$ the **left-endpoint Riemann approximation** of the function $f(x)$ on the closed interval $[a, b]$ with $\Delta x_i = \Delta x = \frac{b-a}{n}$ and sample points $x_i^* = \ell_i = a + (i-1)\,\Delta x$.

- We denote by $\mathcal{R}_n$ the **right-endpoint Riemann approximation** of the function $f(x)$ on the closed interval $[a, b]$ with $\Delta x_i = \Delta x = \frac{b-a}{n}$ and sample points $x_i^* = r_i = a + i\,\Delta x$.

- We denote by $\mathcal{M}_n$ the **midpoint Riemann approximation** of the function $f(x)$ on the closed interval $[a, b]$ with $\Delta x_i = \Delta x = \frac{b-a}{n}$ and sample points $x_i^* = m_i = a + \frac{2i-1}{2}\,\Delta x$.

Example 2.2.6. Compute the left- and right-endpoint and midpoint Riemann approximations of the curve $f(x) = x$ on the closed and bounded interval $[0, 4]$ using four equally-spaced points.

Solution. By recognizing that $a = 0$ and $b = 4$, the length of each interval of the partition is

$$\Delta x_i = \Delta x = \frac{4-0}{4} = \frac{4}{4} = 1.$$

Consequently, the left-endpoint approximation satisfies that $\ell_i = 0 + (i-1)1 = i - 1$; the right-endpoint approximation satisfies that $r_i = 0 + i = i$; and the midpoint approximation satisfies that $m_i = 0 + \frac{2i-1}{2}(1) = \frac{2i-1}{2}$ for each integer $1 \leq i \leq 4$. We conclude therefore that the following hold.

$$\mathcal{L}_4 = \sum_{i=1}^{4} f(\ell_i)\,\Delta x = \sum_{i=1}^{4} \ell_i = \sum_{i=1}^{4}(i-1) = \sum_{i=1}^{4} i - \sum_{i=1}^{4} 1 = \frac{4(4+1)}{2} - 4 = 6$$

$$\mathcal{R}_4 = \sum_{i=1}^{4} f(r_i)\,\Delta x = \sum_{i=1}^{4} r_i = \sum_{i=1}^{4} i = \frac{4(4+1)}{2} = 10$$

$$\mathcal{M}_4 = \sum_{i=1}^{4} f(m_i)\,\Delta x = \sum_{i=1}^{4} \frac{2i-1}{2} = \sum_{i=1}^{4} i - \sum_{i=1}^{4} \frac{1}{2} = 10 - \frac{1}{2}(4) = 8$$ ◇

Example 2.2.7. Compute the left- and right-endpoint and midpoint Riemann approximations of the curve $g(x) = x^2$ on the closed and bounded interval $[0, 1]$ using five equally-spaced points.

Solution. Like before, we find that $a = 0$ and $b = 1$ so that the length of each interval is

$$\Delta x = \frac{1-0}{5} = \frac{1}{5}.$$

By the above, the left-endpoint approximation uses the sample points $\ell_i = 0 + (i-1)\,\Delta x = \frac{i-1}{5}$; the right-endpoint approximation uses the sample points $r_i = 0 + i\Delta x = \frac{i}{5}$; and the midpoint approximation uses the sample points $m_i = 0 + \frac{2i-1}{2} \Delta x = \frac{2i-1}{10}$. We conclude that

$$\mathcal{L}_5 = \sum_{i=1}^{5} g(\ell_i)\,\Delta x = \sum_{i=1}^{5} \frac{\ell_i^2}{5} = \frac{1}{5} \sum_{i=1}^{5} \left(\frac{i-1}{5}\right)^2 = \frac{1}{5}\left(0 + \frac{1}{25} + \frac{4}{25} + \frac{9}{25} + \frac{16}{25}\right) = \frac{30}{75} = \frac{6}{25}$$

$$\mathcal{R}_5 = \sum_{i=1}^{5} g(\ell_i)\,\Delta x = \sum_{i=1}^{5} \frac{r_i^2}{5} = \frac{1}{5} \sum_{i=1}^{5} \left(\frac{i}{5}\right)^2 = \frac{1}{5}\left(\frac{1}{25} + \frac{4}{25} + \frac{9}{25} + \frac{16}{25} + \frac{25}{25}\right) = \frac{55}{75} = \frac{11}{25}$$

$$\mathcal{M}_5 = \sum_{i=1}^{5} g(m_i)\,\Delta x = \sum_{i=1}^{5} \frac{m_i^2}{5} = \frac{1}{5} \sum_{i=1}^{5} \left(\frac{2i-1}{10}\right)^2 = \frac{1}{5}\left(\frac{1}{100} + \frac{9}{100} + \frac{25}{100} + \frac{49}{100} + \frac{81}{100}\right) = \frac{33}{100} \quad \diamond$$

Exercise 2.2.8. Compute the left- and right-endpoint and midpoint Riemann approximations of the curve $h(x) = x^3$ on the closed and bounded interval $[0, 2]$ using eight equally-spaced points.

By allowing the number of sample points to grow arbitrarily large, the error of approximating the area bounded by a curve of one variable by a Riemann sum shrinks to zero, hence we define

$$\text{area bounded by the curve } f(x) \text{ on the closed and bounded interval } [a, b] = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i,$$

where x_i^* are sample points of a partition $\mathcal{P}$ of $[a, b]$ and $\Delta x_i = x_i - x_{i-1}$ for each integer $1 \leq i \leq n$.

Example 2.2.9. Compute the area bounded by $f(x) = x^2$ on the closed interval $[0, 1]$.

Solution. Crucially, the above definition of the area does not depend on the choice sample points x_i^* or the partition $\mathcal{P}$ of $[0, 1]$, so we may carefully construct these to make things as convenient as possible. Given any choice of equally-spaced points $a = x_0 < x_1 < \cdots < x_n = b$, we have that $\Delta x = \frac{1-0}{n} = \frac{1}{n}$. We may choose the right-endpoint approximation so that $x_i^* = \frac{i}{n}$ and

$$\mathcal{R}_n = \sum_{i=1}^{n} f(x_i^*)\,\Delta x = \sum_{i=1}^{n} \left(\frac{i}{n}\right)^2 \left(\frac{1}{n}\right) = \sum_{i=1}^{n} \frac{i^2}{n^3} = \frac{1}{n^3} \sum_{i=1}^{n} i^2 = \frac{1}{n^3} \cdot \frac{n(n+1)(2n+1)}{6}$$

by the second part of Proposition 2.2.5. By taking the limit as $n \to \infty$, we conclude that

$$\text{area bounded by } x^2 \text{ on } [0, 1] = \lim_{n \to \infty} \mathcal{R}_n = \lim_{n \to \infty} \frac{n(n+1)(2n+1)}{6n^3} = \frac{2}{6} = \frac{1}{3}. \quad \diamond$$

2.3 Definite Integration

Given any real function $f(x)$ and any real numbers a and b, consider any collection of points $(x_n, f(x_n))$ on the graph of $f(x)$ with $a = x_0 < x_1 < \cdots < x_n = b$ and $\Delta x_i = x_i - x_{i-1}$ for each integer $1 \leq i \leq n$. Each of the closed and bounded intervals $[x_{i-1}, x_i]$ gives rise to a partition $\mathcal{P}$ of the closed interval $[a, b]$, and we may choose sample points x_i^* for each integer $1 \leq i \leq n$ such that $x_{i-1} \leq x_i^* \leq x_i$ and $x_1^* < x_2^* < \cdots < x_n^*$. Crucially, we are not assuming here that the points $x_0, x_1, \ldots, x_n$ are equally-spaced, hence we may denote $\|\mathcal{P}\| = \max\{\Delta x_i \mid 1 \leq i \leq n\}$. We define

$$\int_a^b f(x)\,dx = \lim_{\|\mathcal{P}\| \to 0} \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i$$

as the **definite integral** of $f(x)$ on the closed and bounded interval $[a, b]$. Provided that the above limit exists, we say that $f(x)$ is **integrable** on $[a, b]$. We refer to the function $f(x)$ in this case as the **integrand**; the real numbers a and b are the **limits of integration**. By our work in the previous section, we may interpret the definite integral $\int_a^b f(x)\,dx$ as the net area bounded by $f(x)$: indeed, $\sum_{i=1}^n f(x_i^*)\,\Delta x_i$ is a Riemann sum representing rectangles of height $f(x_i^*)$ and width Δx_i.

Example 2.3.1. Express the following as the definite integral of a function on the interval $[1, 8]$.

$$\lim_{\|\mathcal{P}\|\to 0} \sum_{i=1}^n \sqrt{2x_i^* + (x_i^*)^2}\,\Delta x_i$$

Solution. Considering that we do not know the partition $\mathcal{P}$ or the sample points x_i^*, there is not much we can do other than recognize the function $f(x)$. Comparing the limit with the definition above, we recognize that $f(x) = \sqrt{2x + x^2}$ so that the limit in question is $\int_1^8 \sqrt{2x + x^2}\,dx$. ◇

Exercise 2.3.2. Express the following as the definite integral of a function on the interval $[0, \pi]$.

$$\lim_{\|\mathcal{P}\|\to 0} \sum_{i=1}^n x_i^* \sin(x_i^*)\,\Delta x_i$$

Often, it is most simple to work with a **regular partition** $\mathcal{P}$, i.e., a partition of $[a, b]$ with $n+1$ equally-spaced points $a = x_0 < x_1 < \cdots < x_n = b$ such that $\Delta x_1 = \Delta x_2 = \cdots = \Delta x_n = \Delta x = \frac{b-a}{n}$. Under this identification, we have that $\Delta x_1 = x_1 - x_0$ so that $x_1 = x_0 + \Delta x_1 = a + \Delta x$, from which it follows that $x_2 = x_1 + \Delta x_2 = (a + \Delta x) + \Delta x = a + 2\Delta x$ and $x_i = a + i\Delta x$ for each integer $1 \leq i \leq n$. Choosing our sample points such that $x_i^* = x_i = a + i\Delta x$ and using the fact that

$$\|\mathcal{P}\| = \max\{\Delta x_i \mid 1 \leq i \leq n\} = \Delta x = \frac{b-a}{n}$$

approaches zero if and only if n approaches ∞, we conclude that

$$\int_a^b f(x)\,dx = \lim_{\|\mathcal{P}\|\to 0} \sum_{i=1}^n f(x_i^*)\,\Delta x_i = \lim_{n\to\infty} \sum_{i=1}^n f(a + i\Delta x)\left(\frac{b-a}{n}\right).$$

Example 2.3.3. Express the following as the definite integral of a function on a closed interval.

$$\lim_{n\to\infty} \sum_{i=1}^n \cos\left(-\pi + i\frac{2\pi}{n}\right)\left(\frac{2\pi}{n}\right)$$

Solution. Considering that $\Delta x = \frac{2\pi}{n} = \frac{b-a}{n}$ and $a = -\pi$, we must have $b = a + n\Delta x = -\pi + 2\pi = \pi$. Even more, the integrand is $\cos(x)$, hence the limit describes the quantity $\int_{-\pi}^{\pi} \cos(x)\,dx$. ◇

Before we endeavor to compute any definite integrals by the limit definition provided above, it is conceptually important to note that the definite integral can be computed by hand in some cases without appealing to any limits. Explicitly, for any real numbers c and d, we have that $\int_a^b (cx+d)\,dx$ represents the net area bounded by the line $cx + d$ and the coordinate axes. Consequently, this area can be computed geometrically as a linear combination of areas of triangles and rectangles.

Exercise 2.3.4. Compute the definite integral $\int_{-2}^3 (3x - 2)\,dx$ using geometry.

Exercise 2.3.5. Compute the definite integral $\int_{-3}^{2}(5-2x)\,dx$ using geometry.

Likewise, for any function of the form $y = f(x) = \sqrt{r^2 - x^2}$, it follows that $x^2 + y^2 = r^2$ yields a circle of radius r, hence we can determine an integral of the form $\int_{-r}^{r} \sqrt{r^2 - x^2}\,dx$.

Exercise 2.3.6. Compute the definite integral $\int_{-1}^{1} \sqrt{1-x^2}\,dx$ using geometry.

Often, we will deal with definite integrals that cannot be computed by geometry; for now, if we encounter this situation, we can sometimes use the limit definition of the definite integral.

Example 2.3.7. Compute the definite integral $\int_0^1 x^2\,dx$ as the limit of a Riemann approximation as the number n of subintervals tends to infinity.

Solution. Considering that $a = 0$ and $b = 1$, we have that

$$\Delta x = \frac{b-a}{n} = \frac{1-0}{n} = \frac{1}{n}$$

so that $a + i\Delta x = 0 + \frac{i}{n} = \frac{i}{n}$. Consequently, it follows that

$$\int_0^1 x^2\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{i}{n}\right)^2 \left(\frac{1}{n}\right) = \lim_{n\to\infty} \frac{1}{n^3} \sum_{i=1}^{n} i^2 = \lim_{n\to\infty} \frac{1}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{1}{3}. \qquad \diamond$$

Example 2.3.8. Compute the definite integral $\int_0^3 (x^3 - 6x)\,dx$ as the limit of a Riemann approximation as the number n of subintervals tends to infinity.

Solution. Considering that $a = 0$ and $b = 3$, we have that

$$\Delta x = \frac{b-a}{n} = \frac{3-0}{n} = \frac{3}{n}$$

so that $a + i\Delta x = 0 + \frac{3i}{n} = \frac{3i}{n}$. Consequently, it follows that

$$\int_0^3 (x^3 - 6x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} \left[\left(\frac{3i}{n}\right)^3 - 6\left(\frac{3i}{n}\right)\right]\left(\frac{3}{n}\right) = \lim_{n\to\infty} \frac{3}{n^2} \sum_{i=1}^{n} \left(\frac{27i^3}{n^2} - 18i\right).$$

Granted that the limit of each of these Riemann sums exists, the limit of their difference is given by the difference of their limits, hence it suffices to compute these limits separately.

$$\lim_{n\to\infty} \frac{3}{n^2} \sum_{i=1}^{n} \frac{81i^3}{n^2} = \lim_{n\to\infty} \frac{81}{n^4} \sum_{i=1}^{n} i^3 = \lim_{n\to\infty} \frac{81}{n^4} \cdot \left[\frac{n(n+1)}{2}\right]^2 = \frac{81}{4}$$

$$\lim_{n\to\infty} \frac{3}{n^2} \sum_{i=1}^{n} 18i = \lim_{n\to\infty} \frac{54}{n^2} \sum_{i=1}^{n} i = \lim_{n\to\infty} \frac{54}{n^2} \cdot \frac{n(n+1)}{2} = \frac{54}{2} = \frac{108}{4}$$

Consequently, we have that $\int_0^3 (x^3 - 6x)\,dx = \frac{81}{4} - \frac{108}{4} = -\frac{27}{4}$. $\diamond$

Based on the definition of the definite integrals and the summation properties outlined in the previous section, we can extrapolate the following properties of definite integrals.

Proposition 2.3.9 (Properties of Definite Integrals)**.** *Given any real function $f(x)$ that is integrable on a closed and bounded interval $[a, b]$, the following properties hold for $\int_a^b f(x)\,dx$.*

(i.) (Empty Integral Law) $\int_a^a f(x)\,dx = 0$

(ii.) (Reversing the Limits of Integration) $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$

(iii.) (Additivity of Adjacent Intervals) $\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$ *for all real numbers c*

(iv.) (Constant Integral Formula) $\int_a^b k\,dx = k(b-a)$ *for all real numbers k*

(v.) (Linearity of a Definite Integral I) $\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$ *for all real numbers k*

(vi.) (Linearity of a Definite Integral II) $\int_a^b [f(x)+g(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$

Example 2.3.10. Compute the definite integral $\int_0^1 (3x^2+4)\,dx$.

Solution. By appealing to Example 2.3.7 and Proposition 2.3.9, we have that

$$\int_0^1 (3x^2+4)\,dx = \int_0^1 3x^2\,dx + \int_0^1 4\,dx = 3\int_0^1 x^2\,dx + 4(1-0) = 3\left(\frac{1}{3}\right) + 4 = 5. \qquad \diamond$$

Example 2.3.11. Given any pair of real functions $f(x)$ and $g(x)$ such that $\int_{-1}^1 f(x)\,dx = 2$ and $\int_{-1}^1 g(x)\,dx = -1$, compute the definite integral $\int_{-1}^1 [3f(x) - g(x)]\,dx$.

Solution. By appealing to Proposition 2.3.9, we have that

$$\begin{aligned}\int_{-1}^1 [3f(x)-g(x)]\,dx &= \int_{-1}^1 (3f(x) + [-g(x)])\,dx \\ &= \int_{-1}^1 3f(x)\,dx + \int_{-1}^1 [-g(x)]\,dx \\ &= 3\int_{-1}^1 f(x)\,dx - \int_{-1}^1 g(x)\,dx = 3(2) - (-1) = 7. \qquad \diamond\end{aligned}$$

Example 2.3.12. Given any real function $f(x)$ such that $\int_0^4 f(x)\,dx = 1$, $\int_{-2}^3 f(x)\,dx = 3$, and $\int_{-2}^0 f(x)\,dx = 5$, compute the definite integral $\int_3^4 f(x)\,dx$.

Solution. By appealing to Proposition 2.3.9, we have that

$$\begin{aligned}\int_3^4 f(x)\,dx &= \int_3^{-2} f(x)\,dx + \int_{-2}^4 f(x)\,dx \\ &= -\int_{-2}^3 f(x)\,dx + \int_{-2}^4 f(x)\,dx \\ &= -\int_{-2}^3 f(x)\,dx + \int_{-2}^0 f(x)\,dx + \int_0^4 f(x)\,dx = -3+5+1 = 3. \qquad \diamond\end{aligned}$$

2.4 The Fundamental Theorem of Calculus

Calculus can be divided into two topics — differentiation and integration — that are connected by the Fundamental Theorem of Calculus. Essentially, the Fundamental Theorem of Calculus says that differentiation and integration are inverse operations: if $f(x)$ is continuous on an open interval, then $f(x)$ admits an antiderivative by the definite integral, and conversely, the definite integral of $f(x)$ over a closed interval measures the **net change** of any antiderivative over that interval.

Theorem 2.4.1 (Fundamental Theorem of Calculus, Part I)**.** *Given any real function $f(x)$ that is integrable with a continuous antiderivative $F(x)$ on a closed interval $[a, b]$, we have that*

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Even more, this quantity measures the net area bounded by the curve $f(x)$ from $x = a$ to $x = b$.

Proof. Observe that the quantity $F(b)-F(a)$ measures the net change of $F(x)$ on the closed interval $[a, b]$. Given any collection of n real numbers $a = x_0 < x_1 < \cdots < x_n = b$, we have that

$$F(b) - F(a) = F(b) - F(x_{n-1}) + F(x_{n-1}) - F(x_{n-2}) + \cdots + F(x_1) - F(a)$$

by adding and subtracting $F(x_i)$ for each integer $1 \leq i \leq n-1$. Grouping each consecutive pair of differences and using the fact that $a = x_0$ and $b = x_n$, it follows that

$$F(b) - F(a) = \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})].$$

By the Mean Value Theorem applied to $F(x)$, for each integer $1 \leq i \leq n$, there exists a real number x_i^* such that $x_{i-1} \leq x_i^* \leq x_i$ and $F(x_i) - F(x_{i-1}) = F'(x_i^*)(x_i - x_{i-1})$. By assumption that $F(x)$ is an antiderivative of $f(x)$ on the closed interval $[a, b]$, we have that $F'(x) = f(x)$, hence we can rewrite each of these equations as $F(x_i) - F(x_{i-1}) = f(x_i^*)\,\Delta x_i$ for the quantity $\Delta x_i = x_i - x_{i-1}$. Going back to our above displayed equation with this new identity, we have that

$$F(b) - F(a) = \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i.$$

By taking the limit as n approaches ∞ on both sides, we conclude the desired result that

$$F(b) - F(a) = \lim_{n\to\infty} [F(b) - F(a)] = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i = \int_a^b f(x)\,dx. \qquad \square$$

Consequently, if $v(t)$ measures the velocity of a particle over time, then the (definite) integral of $v(t)$ over $[a, b]$ measures the total distance travelled by the particle from time $t = a$ to time $t = b$.

Exercise 2.4.2. Compute the net area bounded by the curve $f(x) = x^3$ from $x = -1$ to $x = 1$.

Exercise 2.4.3. Compute the net area bounded by the curve $g(x) = \sin(x)$ from $x = -\frac{\pi}{2}$ to $x = \frac{\pi}{2}$.

Exercise 2.4.4. Compute the net area bounded by the curve $h(x) = \frac{1}{x}$ from $x = 1$ to $x = e$.

Remark 2.4.5. Like we previously mentioned, if $F(x)$ is an antiderivative of a real function $f(x)$ on a closed interval $[a, b]$, the Mean Value Theorem implies that every antiderivative of $f(x)$ over $[a, b]$ is of the form $F(x) + C$ for some real number C. Consequently, the choice of antiderivative of $f(x)$ does not matter when it comes to computing the definite integral of $f(x)$ on $[a, b]$:

$$\int_a^b f(x)\,dx = [F(b) + C] - [F(a) + C] = F(b) - F(a)$$

holds for all real numbers C by the Fundamental Theorem of Calculus, Part I.

One other way to interpret the first part of the Fundamental Theorem of Calculus is as follows.

Corollary 2.4.6 (Net Change Theorem)**.** *Given any differentiable function $f(x)$ on an open interval (a, b) such that $f(a)$ and $f(b)$ are defined, we have that*

$$f(b) - f(a) = \int_a^b f'(x)\,dx.$$

Put another way, the net change of $f(x)$ over the closed interval $[a, b]$ is $\int_a^b f'(x)\,dx$.

Exercise 2.4.7. Consider a leaky water heater that loses $2 + 5t$ gallons of water per hour for each hour after 7 AM. Compute the total amount of water leaked between the time of 9 AM and 12 PM.

Exercise 2.4.8. Consider any medication that disperses into a patient's bloodstream at a rate of $50 - 2\sqrt{t}$ milligrams per hour from the time it is administered. Compute the amount of medication dispersed into a patient's bloodstream one hour after it is administered. Given that one full dose is 50 milligrams, what percentage of the dose reaches the patient's bloodstream in an hour?

Exercise 2.4.9. Consider any particle that moves with velocity $t^3 - 10t^2 + 24t$ meters per second after initial observation at time $t = 0$. Compute the total displacement of and the total distance travelled by the particle from time $t = 0$ to time $t = 6$; then, compare the values.

Conversely, the second part of the Fundamental Theorem of Calculus states that every continuous function on a closed interval $[a, b]$ admits an antiderivative in the form of a definite integral.

Theorem 2.4.10 (Fundamental Theorem of Calculus, Part II)**.** *Given any real function $f(x)$ that is continuous on a closed interval $[a, b]$, for all real numbers $a < x < b$, we have that*

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x).$$

Proof. Considering that $f(x)$ is continuous on $[a, b]$, it is integrable on $[a, b]$, hence we may define

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

for all real numbers $a \leq x \leq b$. We must demonstrate that for all real numbers $a < x < b$, the limit

$$\frac{d}{dx}F(x) = \lim_{h \to 0} \frac{F(x + h) - F(x)}{h}$$

exists. By the second and third parts of Proposition 2.3.9, it follows that

$$F(x+h)-F(x)=\int_a^{x+h} f(t)\,dt-\int_a^x f(t)\,dt=\int_a^{x+h} f(t)\,dt+\int_x^a f(t)\,dt=\int_x^{x+h} f(t)\,dt.$$

By the **Mean Value Theorem for Definite Integrals**, there exists a real number c (depending upon h) such that $x<c<x+h$ and $\int_x^{x+h} f(t)\,dt=f(c)[(x+h)-x]=f(c)h$ so that

$$f(c)=\frac{F(x+h)-F(x)}{h}.$$

Considering that $f(x)$ is continuous on the closed interval $[a,b]$, it follows that

$$f\Big(\lim_{h\to 0} c\Big)=\lim_{h\to 0} f(c)=\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}=F'(x),$$

hence it suffices to compute the limit of c as h approaches 0. By the Squeeze Theorem, we have

$$x=\lim_{h\to 0} x\leq \lim_{h\to 0} c\leq \lim_{h\to 0}(x+h)=x$$

so that $\lim\limits_{h\to 0} c=x$ and $F'(x)=f(x)$ for all real numbers $a<x<b$, as desired. □

Exercise 2.4.11. Compute the derivative of $\int_0^x \sin(t)\,dt$ for any real number $x>0$.

Exercise 2.4.12. Compute the derivative of $\int_{-1}^x e^t\,dx$ for any real number $x>-1$.

Exercise 2.4.13. Compute the derivative of $\int_1^x \ln(t)\,dt$ for any real number $x>1$.

Exercise 2.4.14. Given any differentiable real functions $f(x)$, $g(x)$, and $h(x)$, use the Fundamental Theorem of Calculus, Part II and the Chain Rule for derivatives to prove that

$$\frac{d}{dx}\int_{g(x)}^{h(x)} f(t)\,dt=f(h(x))h'(x)-f(g(x))g'(x).$$

Exercise 2.4.15. Compute the derivative of $\int_0^{x^2}\sin(\cos(t))\,dt$ for any real number $x>0$.

Exercise 2.4.16. Compute the derivative of $\int_{\ln(x)}^{10}\sqrt{t^2+1}\,dt$ for any real number $0<x<e^{10}$.

Exercise 2.4.17. Compute the derivative of $\int_{x^3}^{x^2}\sqrt{t}\,dt$ for any real number $0<x<1$.

2.5 u-Substitution

Until now, we have managed to find the antiderivatives of many functions by viewing antidifferentiation as the inverse to differentiation (in the sense of the Fundamental Theorem of Calculus, Part I) and subsequently using the appropriate analog of the familiar rules for differentiation such as the Power Rule and the Chain Rule. Explicitly, given any real number $r\neq -1$, we have that

$$\int x^r\,dx=\frac{1}{r+1}x^{r+1}+C$$

by the Power Rule. Further, for any differentiable functions $f(x)$ and $g(x)$, we have that

$$\int f'(g(x))g'(x)\,dx=f(g(x))+C$$

by the Chain Rule. Essentially, if we make the assignment $u = g(x)$, then it follows that $\frac{du}{dx} = g'(x)$ and $f'(g(x))g'(x) = f'(u)\frac{du}{dx}$. Conventionally, this relationship is written as $du = g'(x)\,dx$ so that $f'(g(x))g'(x)\,dx = f'(u)\,du$. Considering that $f(u)$ is an antiderivative of $f'(u)$, it follows that

$$\int f'(g(x))g'(x)\,dx = \int f'(u)\,du = f(u) + C = f(g(x)) + C.$$

Colloquially, we refer to this technique (and its broader applications) as u-**substitution**.

Exercise 2.5.1. Compute the indefinite integral of $f(x) = (x+1)^{100}$.

Exercise 2.5.2. Compute the indefinite integral of $g(x) = x\cos(2x^2)$.

Exercise 2.5.3. Compute the indefinite integral of $h(x) = x^2 e^{x^3}$.

Exercise 2.5.4. Compute the indefinite integral of $k(x) = x\sqrt{2x-1}$.

Even more, the technique of u-substitution can be used to evaluate definite integrals. Explicitly, suppose that $f'(x)$ is integrable on the closed interval $[g(a), g(b)]$ and $f'(g(x))g'(x)$ is integrable on the closed interval $[a, b]$. By performing the substitution $u = g(x)$, we have that $du = g'(x)\,dx$ and $f'(g(x))g'(x)\,dx = f'(u)\,du$. Even more, if $x = a$, then $u = g(a)$, and if $x = b$, then $u = g(b)$ so that

$$\int_a^b f'(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f'(u)\,du.$$

Exercise 2.5.5. Compute the definite integral $\int_0^1 x^4(x^5-1)^{10}\,dx$.

Exercise 2.5.6. Compute the definite integral $\int_{-\pi/4}^{\pi/4} 2x\sec^2(x^2)\,dx$

Exercise 2.5.7. Compute the definite integral $\int_1^e \frac{\ln(x)}{x}\,dx$.

We say that a real function $f(x)$ is **even** if it holds that $f(-x) = f(x)$ for all real numbers x in the domain of f such that $-x$ is in the domain of f. Consequently, the polynomial $3x^4 - x^2 + 2$ and the trigonometric function $\cos(x)$ are even functions. Conversely, we say that $f(x)$ is **odd** if it holds that $f(-x) = -f(x)$ for all real numbers x in the domain of f such that $-x$ is in the domain of f. We note that the polynomial $4x^5 + x + 1$ and the trigonometric function $\sin(x)$ are odd functions. We refer to the property that a function is even or odd as the **parity** of the function. We note that a function need not have parity, as illustrated by the fact that $f(x) = x^2 + x$ does not satisfy either $f(-x) = f(x)$ or $f(-x) = -f(x)$; however, the parity of a function is always well-defined.

Exercise 2.5.8. Explain whether $f(x) = \tan(x)$ is even, odd, or neither.

Exercise 2.5.9. Explain whether $g(x) = x^2 e^x$ is even, odd, or neither.

Exercise 2.5.10. Explain whether $h(x) = \sin^2(x)$ is even, odd, or neither.

Proposition 2.5.11 (Properties of Function Parity)**.** *Consider any real functions $f(x)$ and $g(x)$.*

(i.) (Preservation of Parity Under Nonzero Scalar Multiple) *If $f(x)$ has parity, then for all nonzero real numbers α, the scalar multiple $\alpha f(x)$ of $f(x)$ by α has the same parity as $f(x)$.*

(ii.) (Preservation of Parity Under Sum) *If $f(x)$ and $g(x)$ have the same parity, then their sum $f(x) + g(x)$ has the same parity as both $f(x)$ and $g(x)$.*

(iii.) (Preservation of Parity Under Product) *If $f(x)$ and $g(x)$ have the same parity, then their product $f(x)g(x)$ has the same parity as both $f(x)$ and $g(x)$.*

(iv.) (Products of Functions of Opposite Parity) *If $f(x)$ and $g(x)$ have opposite parity, then their product $f(x)g(x)$ is an odd function.*

(v.) (Preservation of Parity Under Quotient) *If $f(x)$ and $g(x)$ have the same parity, then their quotient $f(x)/g(x)$ has the same parity as both $f(x)$ and $g(x)$.*

(vi.) (Quotients of Functions of Opposite Parity) *If $f(x)$ and $g(x)$ have opposite parity, then their quotient $f(x)/g(x)$ is an odd function.*

(vii.) (Preservation of Parity Under Composition) *If $f(x)$ and $g(x)$ have the same parity, then their composite $f(g(x))$ has the same parity as both $f(x)$ and $g(x)$.*

(viii.) (Composition of Functions of Opposite Parity) *If $f(x)$ and $g(x)$ have opposite parity, then their composite $f(g(x))$ is an even function.*

(ix.) (Parity of the Derivative of a Function) *If $f(x)$ is differentiable and $f(x)$ has parity, then the derivative $f'(x)$ has the opposite parity of $f(x)$.*

Proposition 2.5.12 (Definite Integral of an Even Function on a Symmetric Interval)**.** *Consider any even real function $f(x)$ that is integrable on a closed interval $[-a, a]$. We have that*

$$\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx.$$

Proof. By the third property of Proposition 2.3.9, it follows that

$$\int_{-a}^{a} f(x)\,dx = \int_{-a}^{0} f(x)\,dx + \int_{0}^{a} f(x)\,dx.$$

Consider the substitution $u = -x$ with $du = -dx$. By assumption that $f(-x) = f(x)$, we have that

$$\int_{-a}^{0} f(x)\,dx = \int_{-a}^{0} f(-x)\,dx = \int_{a}^{0} f(u)(-du) = -\int_{a}^{0} f(u)\,du = \int_{0}^{a} f(u)\,du = \int_{0}^{a} f(x)\,dx.$$

Consequently, by the above two displayed equations, the desired identity holds. □

Proposition 2.5.13 (Definite Integral of an Odd Function on a Symmetric Interval)**.** *Consider any odd real function $f(x)$ that is integrable on a closed interval $[-a, a]$. We have that*

$$\int_{-a}^{a} f(x)\,dx = 0.$$

Proof. By the third property of Proposition 2.3.9, it follows that

$$\int_{-a}^{a} f(x)\,dx = \int_{-a}^{0} f(x)\,dx + \int_{0}^{a} f(x)\,dx.$$

By assumption that $f(-x) = -f(x)$, the substitution $u = -x$ with $du = -dx$ yields that

$$\int_{-a}^{0} f(x)\,dx = \int_{-a}^{0} -f(-x)\,dx = \int_{a}^{0} -f(u)(-du) = \int_{a}^{0} f(u)\,du = -\int_{0}^{a} f(u)\,du = -\int_{0}^{a} f(x)\,dx.$$

Consequently, by the above two displayed equations, the desired identity holds. □

2.6 Integration by Parts

We turn our attention next to an analog of the Product Rule for antidifferentiation. We adopt the shorthand notation $u = f(x)$ and $v = g(x)$ for some differentiable functions $f(x)$ and $g(x)$ so that $\frac{du}{dx} = f'(x)$ and $\frac{dv}{dx} = g'(x)$ or $du = f'(x)\,dx$ and $dv = g'(x)\,dx$. By the Product Rule, we have that

$$\frac{d}{dx}[uv] = \frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + g(x)f'(x).$$

Considering that uv is clearly an antiderivative of $\frac{d}{dx}[uv] = f(x)g'(x) + g(x)f'(x)$, it follows that

$$uv = \int [f(x)g'(x) + g(x)f'(x)]\,dx = \int f(x)g'(x)\,dx + \int g(x)f'(x)\,dx = \int u\,dv + \int v\,du.$$

By rearranging, we obtain an analog to the Product Rule for antidifferentiation.

Theorem 2.6.1 (Integration by Parts Formula)**.** *Given any differentiable functions $u = f(x)$ and $v = g(x)$, under the convention that $du = f'(x)\,dx$ and $dv = g'(x)\,dx$, we have that*

$$\int u\,dv = uv - \int v\,du.$$

Colloquially, we refer to this technique as the method of **integration by parts** because the rule allows us to identify two parts of the integrand — namely, u and dv — in such a way that

(i.) the antiderivative of u is difficult to determine and its derivative du is simpler;

(ii.) the antiderivative v of dv is readily obtained; and

(iii.) the antiderivative of $v\,du$ is known or can be found by the method of integration by parts.

Exercise 2.6.2. Use integration by parts to compute the antiderivative of $x\cos(x)$.

Exercise 2.6.3. Use integration by parts to compute the antiderivative of $\ln(x)$.

Exercise 2.6.4. Use integration by parts to compute the antiderivative of xe^x.

Once again, the advantage of the method of integration by parts is that it allows us to trade an expression $u\,dv$ that is difficult to antidifferentiate for an expression $v\,du$ whose antiderivative is known or can be found by integration by parts. Consequently, we may identify families of functions whose antiderivatives are unknown to us at this time — e.g., logarithmic and inverse trigonometric functions — and use these as candidates for u. On the other hand, we may identify functions whose antiderivatives are easily found — e.g., algebraic, trigonometric, and exponential functions — and use these as candidates for dv. Ultimately, this gives rise to the following acronym.

Logarithmic Inverse Trigonometric Algebraic Trigonometric Exponential

Essentially, this acronym is intended to help us remember how to prioritize the assignments of u and dv to our integrand: if the function is further left on the list, then it should be made u; if the function is further right on the list, it should be made dv. Consequently, we have the following.

Algorithm 2.6.5 (Using LIATE)**.** Given any pair of functions $f(x)$ and $g(x)$ such that

(a.) $f(x)$ is a logarithmic, inverse trigonometric, or algebraic function and

(b.) $g(x)$ is an algebraic, trigonometric, or exponential function,

in order to compute $\int f(x)g(x)\,dx$, we may assign $u = f(x)$ and $dv = g(x)\,dx$.

Exercise 2.6.6. Use integration by parts once to compute the antiderivative of $x^3 \ln(x)$.

Exercise 2.6.7. Use integration by parts twice to compute the antiderivative of $x^2 \sin(x)$.

Exercise 2.6.8. Use integration by parts three times to compute the antiderivative of $x^3 e^x$.

Exercise 2.6.9. Explain the difficulty in using integration by parts with $u = x^3$ and $dv = e^{x^2}\,dx$ to compute the antiderivative of $x^3 e^{x^2}$. Group the terms differently, and try again successfully.

Observe that in two of the above examples, we were required to use integration by parts multiple times in order to find the antiderivatives of the given functions. Generally, if we wish to evaluate the antiderivative of the product of a function $f(x)$ and a polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$, we may use a shorthand version of integration by parts known as the **tabular method**.

Theorem 2.6.10 (Tabular Method for Integration)**.** *Given any function $f(x)$ whose antiderivatives are known and any polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$ with nonzero a_n, we have that*

$$\int p(x)f(x)\,dx = \sum_{k=0}^{n} (-1)^k p^{(k)}(x) I^{k+1} f(x),$$

where $p^{(k)}(x)$ denotes the kth derivative of $p(x)$ and $I^k f(x)$ denotes the k-fold antiderivative of $f(x)$.

Proof. Observe that the nth derivative of $p(x)$ is given by $p^{(n)}(x) = a_n n!$ so that $p^{(n+1)}(x) = 0$. By the method of Integration by Parts Formula with $u = p(x)$ and $dv = f(x)\,dx$, we have that

$$\int p(x)f(x)\,dx = p(x)F(x) - \int p'(x)F(x)\,dx$$

for some real function $F(x)$ such that $\frac{d}{dx}F(x) = f(x)$. By hypothesis, the antiderivative of $F(x)$ is known, hence we may use integration by parts with $u = p'(x)$ and $dv = F(x)\,dx$ to find that

$$\int p'(x)F(x)\,dx = p'(x)I^2 f(x) - \int p''(x)I^2 f(x)\,dx,$$

where $I^2 f(x)$ denotes the antiderivative of $F(x)$, i.e., $\frac{d}{dx} I^2 f(x) = F(x)$ so that $\frac{d^2}{dx^2} I^2 f(x) = f(x)$. Combined with the above displayed equation, we have that

$$\begin{aligned}\int p(x)f(x)\,dx &= p(x)F(x) - \left(p'(x)I^2 f(x) - \int p''(x)I^2 f(x)\,dx\right) \\ &= p(x)F(x) - p'(x)I^2 f(x) + \int p''(x)I^2 f(x)\,dx.\end{aligned}$$

Using integration by parts once again with $u = p''(x)$ and $dv = I^2 f(x)\,dx$, we have that

$$\int p''(x) I^2 f(x)\,dx = p''(x) I^3 f(x) - \int p'''(x) I^3 f(x)\,dx.$$

Combined with the above displayed equation, we find that

$$\int p(x) f(x)\,dx = p(x)F(x) - p'(x)I^2 f(x) + p''(x)I^3 f(x) - \int p'''(x) I^3 f(x)\,dx.$$

Continue in this manner until $u = p^{(n)}(x)$. By our opening remarks, we have that $du = p^{(n+1)}(x) = 0$ so that $\int v\,du = 0$. Observing the pattern and using $F(x) = \int f(x)\,dx = I^1 f(x)$, we are done. $\square$

Graphically, we can quite simply implement the tabular method by writing out a table with four columns: the first column consists of the index k; the second column consists of the sign $(-1)^k$; the third column consists of the consecutive derivatives of $p(x)$ up to and including 0; and the fourth column consists of the consecutive antiderivatives $I^{k+1} f(x)$ of $f(x)$. Once we have these, the tabular method guarantees that $\int p(x) f(x)\,dx$ can be found by adding the consecutive products of the kth row of the second and third columns by the $(k+1)$th row of the fourth column.

Example 2.6.11. We will illustrate the tabular method to compute the antiderivative of $x^2 \sin(x)$ as in Example 2.6.7. Construct the following table with $p(x) = x^2$ and $f(x) = \sin(x)$.

k	$(-1)^k$	$p^{(k)}(x)$	$I^{k+1} f(x)$
0	$+$	x^2	$\sin(x)$
1	$-$	$2x$	$-\cos(x)$
2	$+$	2	$-\sin(x)$
3	$-$	0	$\cos(x)$

Consequently, we find that $\int x^2 \sin(x)\,dx = x^2(-\cos(x)) - 2x(-\sin(x)) + 2\cos(x)$, as desired.

Exercise 2.6.12. Use the tabular method to verify your solution to Example 2.6.8.

Exercise 2.6.13. Use the tabular method to compute the antiderivative of $x^{10}(2x+1)^4$.

2.7 Trigonometric Integrals

Given positive integers (or whole numbers) m and n, we refer to an integral of the form

$$\int \sin^m(x) \cos^n(x)\,dx$$

as a **trigonometric integral**: indeed, the integrand is a product of powers of basic trigonometric functions. Quickly, one can glean that u-substitution fails, and integration by parts is hopelessly complicated. Using basic trigonometry, however, we are able to evaluate these integrals by converting them to a form in which we can use the tried-and-true methods of yore. Given a right triangle with hypotenuse of length $h > 0$, base of length a, and height of length o, the Pythagorean Theorem states that $o^2 + a^2 = h^2$. By dividing each term in this equation by h, we have that $\frac{o^2}{h^2} + \frac{a^2}{h^2} = 1$.

Using x to represent the angle whose opposite side has length o and whose adjacent side has length a, the Pythagorean Theorem yields the so-called **Pythagorean Identity**

$$\sin^2(x) + \cos^2(x) = 1.$$

Consequently, we may convert any even power of $\cos(x)$ into a power of $1-\sin^2(x)$ (and vice-versa). Considering that $\frac{d}{dx}\sin(x) = \cos(x)$ and $\frac{d}{dx}\cos(x) = -\sin(x)$, we have the following stratagem.

Strategy 2.7.1 (Trigonometric Integration, Case I)**.** Consider the case that either m or n is odd.

(a.) Given that m is odd, we may write $m = 2k+1$ for some positive integer k so that

$$\int \sin^m(x)\cos^n(x)\,dx = \int \sin^{2k+1}(x)\cos^n(x)\,dx = \int [\sin^2(x)]^k \cos^n(x)(\sin(x)\,dx).$$

Considering that $\sin^2(x) = 1-\cos^2(x)$ and $\frac{d}{dx}\cos(x) = -\sin(x)$, letting $u = \cos(x)$ yields that

$$\int \sin^m(x)\cos^n(x)\,dx = -\int (1-u^2)^k u^n\,du.$$

Expanding the polynomial $(1-u^2)^k$ and using the Power Rule, we can find the antiderivative.

(b.) Given that n is odd, we may write $n = 2\ell+1$ for some positive integer ℓ so that

$$\int \sin^m(x)\cos^n(x)\,dx = \int \sin^m(x)\cos^{2\ell+1}(x)\,dx = \int \sin^m(x)(\cos^2(x))^\ell(\cos(x)\,dx).$$

Considering that $\cos^2(x) = 1-\sin^2(x)$ and $\frac{d}{dx}\sin(x) = \cos(x)$, letting $v = \sin(x)$ yields that

$$\int \sin^m(x)\cos^n(x)\,dx = \int v^m(1-v^2)^\ell\,dv.$$

Expanding the polynomial $(1-v^2)^\ell$ and using the Power Rule, we can find the antiderivative.

Example 2.7.2. Compute the indefinite integral of $\sin^3(x)\cos^2(x)$.

Solution. Observe that $\sin^3(x)\cos^2(x)\,dx = \sin^2(x)\cos^2(x)(\sin(x)\,dx)$. By the Pythagorean Identity, we have that $\sin^2(x) = 1-\cos^2(x)$, from which it follows that

$$\sin^3(x)\cos^2(x)\,dx = (1-\cos^2(x))\cos^2(x)(\sin(x)\,dx).$$

Using the substitution $u = \cos(x)$, we have that $du = -\sin(x)\,dx$ so that

$$\sin^3(x)\cos^2(x)\,dx = (1-u^2)u^2(-du) = (u^2-1)u^2\,du = (u^4-u^2)\,du.$$

Consequently, we find that

$$\int \sin^3(x)\cos^2(x)\,dx = \int (u^4-u^2)\,du = \frac{1}{5}u^5 - \frac{1}{3}u^3 + C = \frac{1}{5}\cos^5(x) - \frac{1}{3}\cos^3(x) + C. \qquad \diamond$$

Exercise 2.7.3. Compute the indefinite integral of $\sin^5(x)$.

Unfortunately, this method fails in the case that both m and n are even. Consider the trigonometric integral of $\sin^2(x)\cos^2(x)$. By setting $u = \sin(x)$, we find that $du = \cos(x)\,dx$ so that

$$\sin^2(x)\cos^2(x)\,dx = u^2\cos(x)\,du.$$

But the lingering factor of $\cos(x)$ obstructs our efforts to take the indefinite integral. Likewise, a similar obstruction appears if we attempt to let $u = \cos(x)$. Luckily, we have more trigonometric tools at our disposal. Recall the following **angle addition formulas**.

$$\begin{aligned}\sin(x+y) &= \sin(x)\cos(y) + \sin(y)\cos(x)\\ \cos(x+y) &= \cos(x)\cos(y) - \sin(x)\sin(y).\end{aligned}$$

Using these, we can derive the **double-angle formulas** by plugging in $x = y$.

$$\begin{aligned}\sin(2x) &= 2\sin(x)\cos(x)\\ \cos(2x) &= \cos^2(x) - \sin^2(x)\end{aligned}$$

Considering that $\sin^2(x) + \cos^2(x) = 1$, we may simplify these identities as follows.

$$\begin{aligned}\cos(2x) &= \cos^2(x) - \sin^2(x) = [1 - \sin^2(x)] - \sin^2(x) = 1 - 2\sin^2(x)\\ \cos(2x) &= \cos^2(x) - \sin^2(x) = \cos^2(x) - (1 - \cos^2(x)) = 2\cos^2(x) - 1\end{aligned}$$

By solving for $\sin^2(x)$ and $\cos^2(x)$ above, we obtain the **power-reduction formulas**.

$$\sin^2(x) = \frac{1-\cos(2x)}{2} \qquad\qquad \cos^2(x) = \frac{1+\cos(2x)}{2}$$

One way to memorize the distinction is to "remember your sign" when using sine. Or as my former student Ronald Heminway so eloquently put it, we may use the mnemonic device "sinus minus."

Strategy 2.7.4 (Trigonometric Integration, Case II)**.** Consider the case that neither of the integers m and n is odd. Put another way, consider the case that both of the integers m and n are even.

(a.) Given that $m = n = 2k$ for some positive integer k, we have that

$$\int \sin^m(x)\cos^n(x)\,dx = \int \sin^{2k}(x)\cos^{2k}(x)\,dx = \int [\sin(x)\cos(x)]^{2k}\,dx.$$

Using the double-angle formula $\sin(2x) = 2\sin(x)\cos(x)$, we have that

$$[\sin(x)\cos(x)]^{2k} = \left[\frac{\sin(2x)}{2}\right]^{2k} = \frac{[\sin^2(2x)]^k}{4^k}.$$

Using the power-reduction formula $\sin^2(2x) = \frac{1}{2}[1 - \cos(4x)]$, we can then obtain a polynomial in $\cos(4x)$. Continue using the power-reduction formula for cosine to obtain a linear combination of $\cos(4x)$, $\cos(8x)$, $\cos(16x)$, etc. Each of these has an elementary antiderivative.

(b.) Given that $m = 2i$ and $n = 2j$ for some distinct positive integers i and j, use the power-reduction formulas repeatedly to express $\sin^{2i}(x)\cos^{2j}(x) = [\sin^2(x)]^i[\cos^2(x)]^j$ as a linear combination of $\cos(2x)$, $\cos(4x)$, $\cos(8x)$, etc. Each of these has an elementary antiderivative.

Example 2.7.5. Compute the indefinite integral of $\sin^2(x)\cos^2(x)$.

Solution. By the double-angle formula, we have that

$$\sin^2(x)\cos^2(x)\,dx = [\sin(x)\cos(x)]^2\,dx = \left(\frac{1}{2}\sin(2x)\right)^2 dx = \frac{1}{4}\sin^2(2x)\,dx.$$

Using the power-reduction formula, we find that

$$\sin^2(x)\cos^2(x)\,dx = \frac{1}{4}\sin^2(2x)\,dx = \frac{1}{4}\cdot\frac{1}{2}[1-\cos(4x)]\,dx$$

has an elementary antiderivative. Consequently, we conclude that

$$\int \sin^2(x)\cos^2(x)\,dx = \frac{1}{8}\int[1-\cos(4x)]\,dx = \frac{1}{8}\left[x - \frac{1}{4}\sin(4x)\right] + C. \qquad \diamond$$

Exercise 2.7.6. Compute the indefinite integral of $\cos^4(x)$.

Using the Pythagorean Identity $\sin^2(x) + \cos^2(x) = 1$, we can obtain another identity

$$\tan^2(x) + 1 = \sec^2(x)$$

by dividing each term by $\cos^2(x)$ and recalling that $\tan(x) = \frac{\sin(x)}{\cos(x)}$ and $\sec(x) = \frac{1}{\cos(x)}$. Consequently, we can adapt our stratagem from Trigonometric Integration, Case I to evaluate integrals of the form

$$\int \tan^m(x)\sec^n(x)\,dx.$$

Crucially, toward achieving this end, we must observe the following facts.

1.) By the Quotient Rule, we have that

$$\frac{d}{dx}\tan(x) = \frac{d}{dx}\left[\frac{\sin(x)}{\cos(x)}\right] = \frac{\cos^2(x) - (-\sin(x))(\sin(x))}{\cos^2(x)} = \frac{\sin^2(x)+\cos^2(x)}{\cos^2(x)} = \sec^2(x).$$

2.) By the Chain Rule, we have that

$$\frac{d}{dx}\sec(x) = \frac{d}{dx}[\cos(x)]^{-1} = -[\cos(x)]^{-2}[-\sin(x)] = \frac{\sin(x)}{\cos^2(x)} = \sec(x)\tan(x).$$

3.) Using the substitution $u = \cos(x)$ with $du = -\sin(x)\,dx$, we have that

$$\int \tan(x)\,dx = \int \frac{\sin(x)}{\cos(x)}\,dx = \int \frac{-du}{u} = -\ln|u| + C = -\ln|\cos(x)| + C = \ln|\sec(x)| + C.$$

4.) Using the substitution $u = \sec(x) + \tan(x)$ with $du = [\sec(x)\tan(x) + \sec^2(x)]\,dx$, we have

$$\int \frac{\sec(x)[\sec(x)+\tan(x)]}{\sec(x)+\tan(x)}\,dx = \int \frac{\sec^2(x)+\sec(x)\tan(x)}{\tan(x)+\sec(x)}\,dx = \ln|\sec(x)+\tan(x)| + C.$$

Strategy 2.7.7 (Trigonometric Integration, Case III)**.** Consider the case that $n \geq 2$ is an even integer. Explicitly, assume that $n = 2k$ for some positive integer k, from which it follows that

$$\int \tan^m(x)\sec^n(x)\,dx = \int \tan^m(x)\sec^{2k}(x)\,dx = \int \tan^m(x)[\sec^2(x)]^{k-1}(\sec^2(x)\,dx).$$

Considering that $\sec^2(x) = 1 + \tan^2(x)$ and $\frac{d}{dx}\tan(x) = \sec^2(x)$, letting $u = \tan(x)$ yields that

$$\int \tan^m(x)\sec^n(x)\,dx = \int \tan^m(x)(1+\tan^2(x))^{k-1}(\sec^2(x)\,dx) = \int u^m(1+u^2)^{k-1}\,du.$$

Expanding the polynomial $(1+u^2)^{k-1}$ and using the Power Rule, we can compute the integral.

Exercise 2.7.8. Compute the indefinite integral of $\tan^2(x)\sec^2(x)$.

Exercise 2.7.9. Compute the indefinite integral of $\tan^5(x)\sec^4(x)$.

Strategy 2.7.10 (Trigonometric Integration, Case IV)**.** Consider the case that $m \geq 1$ is odd and $n \geq 1$. Explicitly, assume that $m = 2\ell + 1$ for some positive integer ℓ, from which it follows that

$$\int \tan^m(x)\sec^n(x)\,dx = \int \tan^{2\ell+1}(x)\sec^n(x)\,dx = \int [\tan^2(x)]^\ell \sec^{n-1}(x)(\sec(x)\tan(x)\,dx).$$

Considering that $\tan^2(x) = \sec^2(x) - 1$ and $\frac{d}{dx}\sec(x) = \sec(x)\tan(x)$, letting $v = \sec(x)$ yields that

$$\int \tan^m(x)\sec^n(x)\,dx = \int [\sec^2(x)-1]^\ell \sec^{n-1}(x)(\sec(x)\tan(x)\,dx) = \int (v^2-1)^\ell v^{n-1}\,dv.$$

Expanding the polynomial $(v^2-1)^{\ell-1}$ and using the Power Rule, we can compute the integral.

Exercise 2.7.11. Compute the indefinite integral of $\tan(x)\sec^2(x)$.

Exercise 2.7.12. Compute the indefinite integral of $\tan^3(x)\sec^3(x)$.

Unfortunately, it is difficult to compute the indefinite integral of the function $\tan^m(x)\sec^n(x)$ when $m \geq 2$ is an even integer and $n \geq 1$ is an odd integer; however, in this case, it is possible to use integration by parts and the Pythagorean Identity to transform the integrand into one that falls into either Trigonometric Integration, Case III or Trigonometric Integration, Case IV as follows.

Example 2.7.13. Consider the trigonometric function $\tan^2(x)\sec(x)$. Observe that if $u = \tan(x)$ and $dv = \sec(x)\tan(x)\,dx$, then by the method of Integration by Parts Formula, we have that

$$\int \tan^2(x)\sec(x)\,dx = \sec(x)\tan(x) - \int \sec^3(x)\,dx.$$

We are now in a position to compute the indefinite integral by evaluating the indefinite integral of $\sec^3(x)$. By the Pythagorean Identity $1 + \tan^2(x) = \sec^2(x)$, we have that

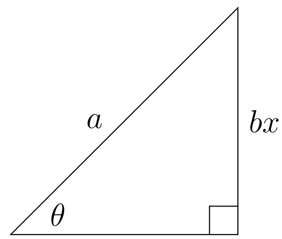
$$\int \sec^3(x)\,dx = \int \sec(x)[\tan^2(x)+1]\,dx = \int \tan^2(x)\sec(x)\,dx + \int \sec(x)\,dx.$$

By plugging this back into our above displayed equation and rearranging, it follows that

$$\int \tan^2(x)\sec(x) = \frac{1}{2}\left[\sec(x)\tan(x) - \int \sec(x)\,dx\right] = \frac{1}{2}\sec(x)\tan(x) - \frac{1}{2}\ln|\sec(x)+\tan(x)| + C.$$

2.8 Trigonometric Substitution

Beyond their extensive applications in geometry and physics, the trigonometric functions yield a very powerful substitution method for integration. Consider the following right triangle.



By the Pythagorean Theorem, the side adjacent to the interior angle θ has length $\sqrt{a^2 - b^2x^2}$ so that $a\cos(\theta) = \sqrt{a^2 - b^2x^2}$. Observe that $bx = a\sin(\theta)$ so that $b\,dx = a\cos(\theta)\,d\theta$, and we have that

$$\begin{aligned}
\int \sqrt{a^2 - b^2x^2}\,dx &= \int a\cos(\theta)\Big(\frac{a}{b}\cos(\theta)\,d\theta\Big) \\
&= \frac{a^2}{b}\int \cos^2(\theta)\,d\theta \\
&= \frac{a^2}{2b}\int [1 + \cos(2\theta)]\,d\theta && \text{(power-reduction formula)} \\
&= \frac{a^2}{2b}\left[\theta + \frac{1}{2}\sin(2\theta)\right] + C \\
&= \frac{a^2}{2b}[\theta + \sin(\theta)\cos(\theta)] + C && \text{(double-angle formula)} \\
&= \frac{a^2}{2b}\left[\arcsin\left(\frac{bx}{a}\right) + \frac{bx}{a^2}\sqrt{a^2 - b^2x^2}\right] + C,
\end{aligned}$$

where the last equality comes from the substitution $bx = \sin(\theta)$ and the above triangle.

Strategy 2.8.1 (Trigonometric Substitution, Case I)**.** Given a function $f(x)$ that can be written as $g(x)\sqrt{a^2 - b^2x^2}$ for some nonzero real numbers a and b and some function $g(x)$, we may attempt to compute $\int f(x)\,dx$ by making the substitution $bx = a\sin(\theta)$ so that $b\,dx = a\cos(\theta)\,d\theta$.

Example 2.8.2. Use a trigonometric substitution to compute the indefinite integral of $x^2\sqrt{1 - x^2}$.

Solution. Considering that this function has a factor of $\sqrt{1 - x^2}$, we may make the trigonometric substitution $x = \sin(\theta)$ so that $dx = \cos(\theta)\,d\theta$. Observe that $x^2 = \sin^2(\theta)$ so that by the Pythagorean Identity, we have that $\sqrt{1 - x^2} = \sqrt{1 - \sin^2(\theta)} = \sqrt{\cos^2(\theta)} = \cos(\theta)$. Consequently, we find that

$$\int x^2\sqrt{1 - x^2}\,dx = \int \sin^2(\theta)\cos(\theta)(\cos(\theta)\,d\theta) = \int \sin^2(\theta)\cos^2(\theta)\,d\theta.$$

By Example 2.7.5 above and the double-angle formulas, we have that

$$\begin{aligned}\int \sin^2(\theta)\cos^2(\theta)\,d\theta &= \frac{1}{8}\left[\theta - \frac{1}{4}\sin(4\theta)\right] + C \\ &= \frac{1}{8}\left[\theta - \frac{1}{2}\sin(2\theta)\cos(2\theta)\right] + C \\ &= \frac{1}{8}(\theta - \sin(\theta)\cos(\theta)[\cos^2(\theta) - \sin^2(\theta)]) + C \\ &= \frac{1}{8}[\theta - \sin(\theta)\cos^3(\theta) + \sin^3(\theta)\cos(\theta)] + C.\end{aligned}$$

Using the substitution $x = \sin(\theta)$ and the fact that $\sqrt{1-x^2} = \cos(\theta)$, we conclude that

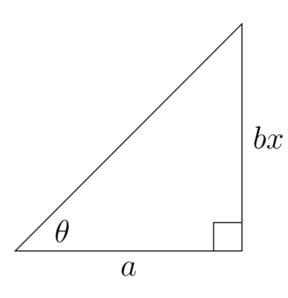
$$\int x^2\sqrt{1-x^2}\,dx = \frac{1}{8}\left[\arcsin(x) - x(1-x^2)^{3/2} + x^3\sqrt{1-x^2}\right] + C. \qquad \diamond$$

Exercise 2.8.3. Use a trigonometric substitution to compute the indefinite integral of $\frac{x}{\sqrt{1-x^2}}$.

Exercise 2.8.4. Use a trigonometric substitution to compute the indefinite integral of $x^5\sqrt{1-9x^2}$.

Exercise 2.8.5. Use a trigonometric substitution to compute the indefinite integral of $\frac{x^2}{\sqrt{9-x^2}}$.

Certainly, it is possible to consider other possibilities for our initial right triangle. Explicitly, suppose that the altitude and base of a right triangle are given as follows.



By the Pythagorean Theorem, the hypotenuse of the above right triangle has length $\sqrt{a^2+b^2x^2}$.

Observe that $bx = a\tan(\theta)$ so that $b\,dx = a\sec^2(\theta)\,d\theta$. Consequently, we have that

$$\begin{aligned}
\int \frac{dx}{\sqrt{a^2+b^2x^2}} &= \int \frac{\frac{a}{b}\sec^2(\theta)\,d\theta}{\sqrt{a^2+a^2\tan^2(\theta)}} \\
&= \frac{a}{b}\int \frac{\sec^2(\theta)\,d\theta}{\sqrt{a^2(1+\tan^2(\theta))}} \\
&= \frac{a}{b}\int \frac{\sec^2(\theta)\,d\theta}{\sqrt{a^2\sec^2(\theta)}} && \text{(Pythagorean Identity)} \\
&= \frac{1}{b}\int \sec(\theta)\,d\theta && (a>0 \text{ and } \sec(\theta)>0) \\
&= \frac{1}{b}\ln|\sec(\theta)+\tan(\theta)| + C \\
&= \frac{1}{b}\ln\left|\frac{\sqrt{a^2+b^2x^2}+bx}{a}\right| + C,
\end{aligned}$$

where the last equality comes from the substitution $bx = a\tan(\theta)$ and the above triangle.

Strategy 2.8.6 (Trigonometric Substitution, Case II)**.** Given a function $f(x)$ that can be written as $g(x)\sqrt{a^2+b^2x^2}$ for some nonzero real numbers a and b and some function $g(x)$, we may attempt to compute $\int f(x)\,dx$ by making the substitution $bx = a\tan(\theta)$ so that $b\,dx = a\sec^2(\theta)\,d\theta$.

Example 2.8.7. Use a trigonometric substitution to compute the indefinite integral of $x^3\sqrt{1+x^2}$.

Solution. Considering that this function has a factor of $\sqrt{1+x^2}$, we may make the trigonometric substitution $x = \tan(\theta)$ with $dx = \sec^2(\theta)\,d\theta$. Observe that $x^2 = \tan^2(\theta)$ so that by the Pythagorean Identity, we have that $\sqrt{1+x^2} = \sqrt{1+\tan^2(\theta)} = \sqrt{\sec^2(\theta)} = \sec(\theta)$. Consequently, we find that

$$\int x^3\sqrt{1+x^2}\,dx = \int \tan^3(\theta)\sec(\theta)(\sec^2(\theta)\,d\theta) = \int \tan^3(\theta)\sec^3(\theta)\,d\theta.$$

We are now in a position to evaluate a trigonometric integral. By the technique outlined in Trigonometric Integration, Case IV, we may borrow a factor of $\tan(\theta)$ and a factor of $\sec(\theta)$ and use the Pythagorean Identity $\tan^2(\theta) = \sec^2(\theta) - 1$ to simplify the integrand $\tan^3(\theta)\sec^3(\theta)\,d\theta$ as follows.

$$\int \tan^3(\theta)\sec^3(\theta)\,d\theta = \int (\sec^2(\theta)-1)\sec^2(\theta)(\sec(\theta)\tan(\theta)\,d\theta)$$

We now employ the substitution $u = \sec(\theta)$ with $du = \sec(\theta)\tan(\theta)\,d\theta$ to obtain the following.

$$\int (\sec^2(\theta)-1)\sec^2(\theta)(\sec(\theta)\tan(\theta)\,d\theta) = \int (u^2-1)u^2\,du = \int (u^4-u^2)\,du = \frac{1}{5}u^5 - \frac{1}{3}u^3 + C.$$

Considering that $u = \sec(\theta) = \sqrt{1+x^2}$, it follows that

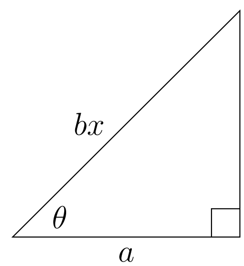
$$\int x^3\sqrt{1+x^2}\,dx = \frac{1}{5}(1+x^2)^{5/2} - \frac{1}{3}(1+x^2)^{3/2} + C. \qquad \diamond$$

Exercise 2.8.8. Use a trigonometric substitution to compute the indefinite integral of $(x^2+1)^{-3/2}$.

Exercise 2.8.9. Use a trigonometric substitution to compute the indefinite integral of $x^2(x^2+9)^{3/2}$.

Exercise 2.8.10. Use a trigonometric substitution to compute the indefinite integral of $x^5\sqrt{4+x^2}$.

Last, consider the following right triangle in which the base and hypotenuse are given.



By the Pythagorean Theorem, the side opposite the interior angle θ has length $\sqrt{b^2x^2-a^2}$. Observe that $bx = a\sec(\theta)$ so that $b\,dx = a\sec(\theta)\tan(\theta)\,d\theta$. Consequently, we have that

$$\begin{aligned}
\int \frac{dx}{b^2x^2-a^2} &= \int \frac{\frac{a}{b}\sec(\theta)\tan(\theta)\,d\theta}{a^2\tan^2(\theta)} \\
&= \frac{1}{ab}\int \frac{\sec(\theta)\,d\theta}{\tan(\theta)} \\
&= \frac{1}{ab}\int \csc(\theta)\,d\theta \\
&= -\frac{1}{ab}\ln|\csc(\theta)+\cot(\theta)| + C \\
&= -\frac{1}{ab}\ln\left|\frac{bx+a}{\sqrt{b^2x^2-a^2}}\right| + C,
\end{aligned}$$

where the last equality comes from the substitution $bx = a\sec(\theta)$ and the above triangle.

Strategy 2.8.11 (Trigonometric Substitution, Case III)**.** Given a function $f(x)$ that can be written as $g(x)\sqrt{b^2x^2-a^2}$ for some nonzero real numbers a and b and some function $g(x)$, we may attempt to compute $\int f(x)\,dx$ via the substitution $bx = a\sec(\theta)$ so that $b\,dx = a\sec(\theta)\tan(\theta)\,d\theta$.

Example 2.8.12. Use a trigonometric substitution to compute the indefinite integral of $x^3\sqrt{x^2-1}$.

Solution. Considering that this function has a factor of $\sqrt{x^2-1}$, we may make the substitution $x = \sec(\theta)$ so that $dx = \sec(\theta)\tan(\theta)\,d\theta$. Observe that $x^2 = \sec^2(\theta)$ so that by the Pythagorean Identity, we have that $\sqrt{x^2-1} = \sqrt{\sec^2(\theta)-1} = \sqrt{\tan^2(\theta)} = \tan(\theta)$. Consequently, we find that

$$\int x^3\sqrt{x^2-1}\,dx = \int \sec^3(\theta)\tan(\theta)(\sec(\theta)\tan(\theta)\,d\theta) = \int \tan^2(\theta)\sec^4(\theta)\,d\theta.$$

Observe that we may use the substitution $u = \tan(\theta)$ with $du = \sec^2(\theta)\,d\theta$ to obtain

$$\begin{aligned}
\int \tan^2(\theta)\sec^4(\theta)\,d\theta &= \int \tan^2(\theta)\sec^2(\theta)(\sec^2(\theta)\,d\theta) \\
&= \int \tan^2(\theta)(1+\tan^2(\theta))(\sec^2(\theta)\,d\theta) && \text{(Pythagorean Identity)} \\
&= \int u^2(1+u^2)\,du \\
&= \int (u^2+u^4)\,du \\
&= \frac{1}{3}u^3 + \frac{1}{5}u^5 + C.
\end{aligned}$$

Considering that $u = \tan(\theta) = \sqrt{x^2-1}$, it follows that

$$\int x^3\sqrt{x^2-1}\,dx = \frac{1}{3}(x^2-1)^{3/2} + \frac{1}{5}(x^2-1)^{5/2} + C. \qquad \diamond$$

Exercise 2.8.13. Use a trigonometric substitution to compute $\int (x^2-4)^{-3/2}\,dx$.

Exercise 2.8.14. Use a trigonometric substitution to compute $\int \sqrt{4x^2-9}\,dx$.

Exercise 2.8.15. Use a trigonometric substitution to compute $\int x^5\sqrt{x^2-16}\,dx$.

2.9 Partial Fraction Decomposition

We have thus far discussed several satisfactory techniques for integrating power functions, algebraic functions, exponential functions, logarithmic functions, trigonometric functions, and their products; however, we have not yet uniformly dealt with the problem of integrating **rational functions**. By definition, a rational function is a quotient of two polynomial expressions, e.g., the rational functions

$$\frac{1}{x^2+2x} \text{ and } \frac{x-2}{x-5} \text{ and } \frac{x^3-1}{x^2+1}.$$

We say that a rational function is **proper** if and only if the degree of the polynomial in the numerator is less than the degree of the polynomial in the denominator. Of the displayed rational functions above, only the first is a proper rational function. By performing **polynomial long division**, we may convert any **improper** rational function into a linear combination of proper rational functions. Explicitly, we have that $x-2=(x-5)+3$ so that dividing each side by $x-5$ yields that

$$\frac{x-2}{x-5}=1+\frac{3}{x-2}.$$

We may subsequently compute the antiderivative of this rational function by elementary methods.

$$\int \frac{x-2}{x-5}\,dx=\int\left(1+\frac{3}{x-2}\right)dx=\int 1\,dx+\int\frac{3}{x-2}\,dx=x+3\ln|x-2|+C$$

Likewise, by polynomial long division, we find that $x^3-1=x(x^2+1)-(x+1)$ so that the improper rational function can be written as the following linear combination of proper rational functions.

$$\frac{x^3-1}{x^2+1}=x-\frac{x+1}{x^2+1}=x-\frac{x}{x^2+1}-\frac{1}{x^2+1}$$

Once again, the antiderivative of this rational function can be found with relative ease.

$$\int\frac{x^3-1}{x^2+1}\,dx=\int x\,dx-\int\frac{x}{x^2+1}\,dx-\frac{1}{x^2+1}\,dx=\frac{1}{2}x^2-\frac{1}{2}\ln(x^2+1)-\arctan(x)+C$$

Unfortunately, the antiderivative of the proper rational function $(x^2+2x)^{-1}$ cannot be obtained by any technique we have discussed so far; however, it is possible to integrate this function by noticing (quite cleverly) that it can be written as a difference of proper rational functions as follows.

$$\int\frac{1}{x^2+2x}\,dx=\int\left(\frac{1}{2x}-\frac{1}{2(x+2)}\right)dx=\frac{1}{2}\int\frac{1}{x}\,dx-\frac{1}{2}\int\frac{1}{x+2}\,dx=\frac{1}{2}(\ln|x|-\ln|x+2|)+C$$

Essentially, the content of this observation is the method of **partial fraction decomposition**.

Before we delve into the method of partial fraction decomposition, we must continue to recall some important notions from college algebra. We say that a polynomial is **irreducible** if it cannot be written as a product of two polynomials of strictly lesser degree. Consequently, a linear polynomial $ax+b$ is irreducible; it can be shown that a quadratic polynomial is irreducible if and only if it has no roots. By the Quadratic Equation, the roots of a real quadratic polynomial ax^2+bx+c are

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

so that ax^2+bx+c is irreducible if and only if $b^2-4ac<0$. We refer to the real number b^2-4ac as the **discriminant** of the quadratic: if this quantity is negative, the quadratic has only imaginary roots. One of the most useful (and nontrivial) facts about real polynomials is that the only irreducible polynomials with real coefficients are linear or quadratic. Put another way, it turns out that every real polynomial factors as a product of linear and irreducible quadratic polynomials.

Theorem 2.9.1 (Partial Fraction Decomposition Theorem)**.**

(a.) (Distinct Linear Factors) *Given any real numbers a, b, c, and d such that a and c are nonzero and $ax+b$ and $cx+d$ are distinct, there exist nonzero real numbers A and B such that*

$$\frac{1}{(ax+b)(cx+d)}=\frac{A}{ax+b}+\frac{B}{cx+d}.$$

(b.) (Powers of Distinct Linear Factors) *Given any real numbers a, b, c, and d such that a and c are nonzero and $ax+b$ and $cx+d$ are distinct and any pair of positive integers m and n, there exist real numbers $A_1, A_2, \dots, A_m$ and $B_1, B_2, \dots, B_n$ not all of which are zero such that*

$$\frac{1}{(ax+b)^m(cx+d)^n} = \sum_{i=1}^{m} \frac{A_i}{(ax+b)^i} + \sum_{j=1}^{n} \frac{B_j}{(cx+d)^j}.$$

(c.) (Linear and Irreducible Quadratic Factors) *Given any real numbers a, b, c, d, and e such that a and c are nonzero and $d^2 - 4ce < 0$, there exist real numbers A, B, C not all zero such that*

$$\frac{1}{(ax+b)(cx^2+dx+e)} = \frac{A}{ax+b} + \frac{Bx+C}{cx^2+dx+e}.$$

(d.) (Distinct Irreducible Quadratic Factors) *Given any real numbers a, b, c, d, e, and f such that a and d are nonzero, ax^2+bx+c and dx^2+ex+f are distinct, $b^2-4ac<0$, and $e^2-4df<0$, there exist real numbers A, B, C, and D not all of which are zero such that*

$$\frac{1}{(ax^2+bx+c)(dx^2+ex+f)} = \frac{Ax+B}{ax^2+bx+c} + \frac{Cx+D}{dx^2+ex+f}.$$

(e.) (Powers of Distinct Irreducible Quadratic Factors) *Given any pair of positive integers m and n and any real numbers a, b, c, d, e, and f such that a and d are nonzero, $b^2-4ac<0$, $e^2-4df<0$, and ax^2+bx+c and dx^2+ex+f are distinct, there exist real numbers $A_1, \dots, A_m, B_1, \dots, B_m, C_1, \dots, C_n$, and $D_1, D_2, \dots, D_n$ not all of which are zero such that*

$$\frac{1}{(ax^2+bx+c)^m(dx^2+ex+f)^n} = \sum_{i=1}^{m} \frac{A_ix+B_i}{(ax^2+bx+c)^i} + \sum_{j=1}^{n} \frac{C_jx+D_j}{(dx^2+ex+f)^j}.$$

Even more, these are all of the possible cases of proper rational functions with numerator 1.

Example 2.9.2. Use the Partial Fraction Decomposition Theorem to compute $\displaystyle\int \frac{1}{x^2-5x-6}\,dx$.

Solution. Observe that $x^2-5x-6 = (x-6)(x+1)$ is a factorization of x^2-5x-6 into distinct linear factors, hence the method of partial fraction decomposition yields that

$$\frac{1}{x^2-5x-6} = \frac{A}{x-6} + \frac{B}{x+1}.$$

Clearing denominators and using the fact that $(x-6)(x+1) = x^2-5x-6$, we find that

$$1 = A(x+1) + B(x-6).$$

By setting $x=6$, we find that $1 = 7A$ so that $A = \frac{1}{7}$. By setting $x=-1$, we find that $1 = -7B$ so that $B = -\frac{1}{7}$. Consequently, the method of partial fraction decomposition reveals that

$$\frac{1}{x^2-5x-6} = \frac{\frac{1}{7}}{x-6} + \frac{-\frac{1}{7}}{x+1}.$$

We may therefore return to compute our indefinite integral with elementary techniques.

$$\int \frac{1}{x^2-5x-6}\,dx = \frac{1}{7}\int \frac{1}{x-6}\,dx - \frac{1}{7}\int \frac{1}{x+1}\,dx = \frac{1}{7}\ln|x-6| - \frac{1}{7}\ln|x+1| + C. \qquad \diamond$$

Example 2.9.3. Use the method of partial fraction decomposition to compute $\int (x^4-1)^{-1}\,dx$.

Solution. Observe that $x^4-1=(x^2-1)(x^2+1)=(x-1)(x+1)(x^2+1)$ is a factorization of x^4-1 into distinct linear and quadratic factors. Considering that $0-4(1)(1)=-4<0$, it follows that x^2+1 is irreducible. Using the method of partial fraction decomposition, it follows that

$$\frac{1}{x^4-1}=\frac{A}{x-1}+\frac{B}{x+1}+\frac{Cx+D}{x^2+1}.$$

Clearing denominators and using the fact that $(x-1)(x+1)=x^2-1$, we find that

$$1=A(x+1)(x^2+1)+B(x-1)(x^2+1)+(Cx+D)(x^2-1).$$

Considering that this identity holds for all x, it follows that $4A=1$ by plugging in $x=1$, $-4B=1$ by plugging in $x=-1$, and $A-B-D=1$ by plugging in $x=0$. We find immediately that

$$A=\frac{1}{4},\ B=-\frac{1}{4},\ \text{and } D=A-B-1=\frac{1}{2}-1=-\frac{1}{2}.$$

Expanding the polynomial on the right in the second-to-last displayed equation, we find that

$$0x^3+1=1=(A+B+C)x^3+\text{some polynomial of degree at most two.}$$

Comparing coefficients gives that $A+B+C=0$ so that $C=0$. We conclude that

$$\frac{1}{x^4-1}=\frac{\frac{1}{4}}{x-1}-\frac{\frac{1}{4}}{x+1}-\frac{\frac{1}{2}}{x^2+1},$$

from which it follows that

$$\int\frac{1}{x^4+1}\,dx=\frac{1}{4}\int\frac{1}{x-1}\,dx-\frac{1}{4}\int\frac{1}{x+1}-\frac{1}{2}\int\frac{1}{x^2+1}\,dx$$

$$=\frac{1}{4}\ln|x-1|-\frac{1}{4}\ln|x+1|-\frac{1}{2}\arctan(x)+C. \qquad \diamond$$

Caution: it is not necessarily always possible to eliminate variables by plugging in carefully chosen values $x=a$ when implementing the method of partial fraction decomposition. Ultimately, it is in fact best to use the **method of undetermined coefficients**, as outlined in our next example.

Example 2.9.4. Use the Partial Fraction Decomposition Theorem to compute $\displaystyle\int\frac{2x+1}{x^4+2x^2+1}\,dx$.

Solution. Observe that $x^4+4x^2+3=(x^2+1)(x^2+3)$ is a factorization of x^4+4x^2+3 into distinct irreducible factors. Using the method of partial fraction decomposition, we have that

$$\frac{2x+1}{(x^2+1)(x^2+3)}=\frac{Ax+B}{x^2+1}+\frac{Cx+D}{x^2+3}.$$

Clearing denominators, we find that

$$2x+1=(Ax+B)(x^2+3)+(Cx+D)(x^2+1).$$

Considering that x^2+1 and x^2+3 are irreducible, we cannot eliminate either of these quadratic factors by substituting $x=a$ for any real number a. Consequently, we must compare coefficients. Expanding the right-hand side in the second-to-last displayed equation, we find that

$$2x+1=(A+C)x^3+(B+D)x^2+(3A+C)x+3B+D,$$

from which we obtain the following linear system of equations.

$$\begin{aligned} A+C&=0 \qquad\qquad & 3A+C&=2\\ B+D&=0 & 3B+D&=1\end{aligned}$$

We have therefore that $A=-C$ and $B=-D$ so that $2=-3C+C=-2C$ and $1=-3D+D=-2D$. We conclude that $A=1$, $B=\frac{1}{2}$, $C=-1$, and $D=-\frac{1}{2}$, from which it follows that

$$\begin{aligned}\int \frac{2x+1}{x^4+4x^2+3}\,dx &= \int \left(\frac{x+\frac{1}{2}}{x^2+1}-\frac{x+\frac{1}{2}}{x^2+3}\right)dx\\
&=\frac{1}{2}\int \frac{2x+1}{x^2+1}\,dx-\frac{1}{2}\int\frac{2x+1}{x^2+3}\,dx\\
&=\frac{1}{2}\int\frac{2x}{x^2+1}\,dx+\frac{1}{2}\int\frac{1}{x^2+1}\,dx-\frac{1}{2}\int\frac{2x}{x^2+3}\,dx-\frac{1}{2}\int\frac{1}{x^2+3}\,dx\\
&=\frac{1}{2}\ln|x^2+1|+\frac{1}{2}\arctan(x)-\frac{1}{2}\ln|x^2+3|-\frac{1}{2\sqrt{3}}\arctan\left(\frac{x}{\sqrt{3}}\right)+C,\end{aligned}$$

where the last integral is determined by $x^2+3=3\left[\left(\frac{x}{\sqrt{3}}\right)^2+1\right]$ and the substitution $u=\frac{x}{\sqrt{3}}$. ◇

Example 2.9.5. Use the Partial Fraction Decomposition Theorem to compute $\displaystyle\int \frac{1}{x^2-1}\,dx$.

Exercise 2.9.6. Use the method of partial fraction decomposition to compute $\displaystyle\int\frac{2x+3}{x^3-2x^2+4x-8}\,dx$.

Observe that the method of partial fraction decomposition applies to proper rational functions; however, by polynomial long division, every rational function induces a proper rational function.

Example 2.9.7. Use polynomial long division to express the following rational function as the sum of a polynomial and a proper rational function; then, compute its indefinite integrals.

$$f(x)=\frac{x^3+1}{x^2+x+1}$$

Solution. We proceed by polynomial long division. Our task is to sequentially eliminate the largest power of x in each polynomial that appears as the **dividend** in the long division.

1.) Our dividend is x^3+1, and our **divisor** is x^2+x+1. Observe that

$$(x^3+1)-x(x^2+x+1)=(x^3+1)-(x^3+x^2+x)=-x^2-x+1.$$

2.) Our dividend is now $-x^2 - x + 1$, and our divisor is $x^2 + x + 1$. Observe that

$$(-x^2 - x + 1) - (-1)(x^2 + x + 1) = (-x^2 - x + 1) + (x^2 + x + 1) = 2.$$

3.) Our dividend of 2 has lesser degree than $x^2 + x + 1$, so the division terminates.

$$
\begin{array}{rrrrr}
 & & & x & -\,1\\
\cline{2-5}
x^2+x+1 \big) & x^3 & & & +\,1\\
 & -\,x^3 & -\,x^2 & -\,x & \\
\cline{2-4}
 & & -\,x^2 & -\,x & +\,1\\
 & & x^2 & +\,x & +\,1\\
\cline{3-5}
 & & & & 2
\end{array}
$$

Ultimately, we find that $x^3 + 1 = (x - 1)(x^2 + x + 1) + 2$ so that

$$\frac{x^3 + 1}{x^2 + x + 1} = x - 1 + \frac{2}{x^2 + x + 1}.$$

Considering that $1^2 - 4(1)(1) = -3 < 0$, it follows that $x^2 + x + 1$ is an irreducible quadratic polynomial, hence the method of partial fraction decomposition fails to improve the situation here; rather, we may revert to the method of **completing the square** to find that

$$x^2 + x + 1 = x^2 + x + \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 + 1 = \left(x + \frac{1}{2}\right)^2 + \frac{3}{4}.$$

By setting $u = x + \frac{1}{2}$, we find that $du = dx$ so that

$$\int \frac{2}{x^2 + x + 1}\,dx = 2\int \frac{1}{\left(x + \frac{1}{2}\right)^2 + \frac{3}{4}}\,dx = 2\int \frac{1}{u^2 + \frac{3}{4}}\,du = \frac{8}{3}\int \frac{1}{\left(\frac{2}{\sqrt{3}}u\right)^2 + 1}\,du.$$

One can perform a substitution $t = \frac{2}{\sqrt{3}}u$ with $dt = \frac{2}{\sqrt{3}}\,du$ or simply recognize this integral as

$$\frac{8}{3}\int \frac{1}{\left(\frac{2}{\sqrt{3}}u\right)^2 + 1}\,du = \frac{4}{\sqrt{3}}\arctan\left(\frac{2}{\sqrt{3}}u\right) + C = \frac{4}{\sqrt{3}}\arctan\left(\frac{2x+1}{\sqrt{3}}\right) + C.$$

Ultimately, we conclude that the function has the following general antiderivative.

$$\int \frac{x^3 + 1}{x^2 + x + 1}\,dx = \int \left(x - 1 + \frac{2}{x^2 + x + 1}\right)dx = \frac{1}{2}x^2 - x + \frac{4}{\sqrt{3}}\arctan\left(\frac{2x+1}{\sqrt{3}}\right) + C \qquad \diamond$$

Example 2.9.8. Use polynomial long division to express the following rational functions as the sum of a polynomial and a proper rational function; then, compute their indefinite integrals.

(a.) $\dfrac{x^3 + 1}{x^2 + x + 1}$ (b.) $\dfrac{x^4 - x^2 + 1}{x^2 - 1}$ (c.) $\dfrac{x^5 - 4x^4 + 9x^2 - 6}{x^3 + x^2 - x - 1}$

2.10 Improper Integration

Our interest in integrals so far has been to find the net area bounded by the curve $f(x)$. Because of this, we have restricted ourselves to closed and bounded intervals of the form $[a, b]$. Often, we are interested in how a mathematical model behaves in the long-run, i.e., as x grows arbitrarily large (or approaches $\pm\infty$). Under this framework, we develop the concept of the improper integral.

Given a function $f(x)$ that is integrable over the closed region $[a, b]$ for every real number $b > a$, the **improper integral** of $f(x)$ over the interval $[a, \infty)$ is defined (if it exists) as

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx.$$

By the Fundamental Theorem of Calculus, Part I, for any antiderivative $F(x)$ of $f(x)$, we have that

$$\lim_{b\to\infty} \int_a^b f(x)\,dx = \lim_{b\to\infty} [F(b) - F(a)].$$

One can analogously define the improper integral of $f(x)$ over the interval $(-\infty, b]$ as

$$\int_{-\infty}^b f(x)\,dx = \lim_{a\to-\infty} \int_a^b f(x)\,dx$$

whenever $f(x)$ is integrable over the closed and bounded interval $[a, b]$ for every real numbers $a < b$. Even more, the doubly improper integral of $f(x)$ over $(-\infty, \infty)$ is defined as

$$\int_{-\infty}^\infty f(x)\,dx = \lim_{b\to\infty} \left(\lim_{a\to-\infty} \int_a^b f(x)\,dx \right) = \lim_{a\to-\infty} \left(\lim_{b\to\infty} \int_a^b f(x)\,dx \right)$$

whenever $f(x)$ is integral over the closed and bounded interval $[a, b]$ for all real numbers a and b.

Exercise 2.10.1. Compute the improper integral $\int_1^\infty x^{-2}\,dx$.

Exercise 2.10.2. Compute the improper integral $\int_{-\infty}^1 e^x\,dx$.

Exercise 2.10.3. Compute the improper integral $\int_0^\infty xe^{-x}\,dx$.

Exercise 2.10.4. Compute the improper integral $\int_{-\infty}^\infty (1 + x^2)^{-1}\,dx$.

Exercise 2.10.5. Compute the improper integral $\int_{-\infty}^\infty xe^{-x^2}\,dx$.

Each of the above functions admits horizontal asymptotes, hence the improper integrals we computed were all finite, and the ends of our computations justified the means.

One can also consider the improper integral of a function with a vertical asymptote. Given that $f(x)$ is continuous on the half-open interval $[a, b)$ and $\lim_{x\to b^-} f(x) = \pm\infty$, we have that

$$\int_a^b f(x)\,dx = \lim_{t\to b^-} \int_a^t f(x)\,dx = \lim_{t\to b^-} [F(t) - F(a)]$$

for any antiderivative $F(x)$ of $f(x)$ (if this limit exists). One can analogously define the improper integral of $f(x)$ over the half-open interval $(a, b]$ whenever $\lim_{x\to a^+} f(x) = \pm\infty$ (provided it exists) as

$$\int_a^b f(x)\,dx = \lim_{u\to a^+} \int_u^b f(x)\,dx = \lim_{u\to a^+} [F(b) - F(u)].$$

Even if the integrand $f(x)$ is unbounded as $x > a$ approaches a and as $x < b$ approaches b, it is still possible to define the doubly improper integral of $f(x)$ over (a, b) as

$$\int_a^b f(x)\,dx = \lim_{t \to b^-} \left(\lim_{u \to a^+} \int_u^t f(x)\,dx \right) = \lim_{u \to a^+} \left(\lim_{t \to b^-} \int_u^t f(x)\,dx \right)$$

provided that $f(x)$ is integrable over the closed interval $[u, t]$ for all real numbers $a < u < t < b$.

Exercise 2.10.6. Compute the improper integral $\int_0^1 (x-1)^{-1}\,dx$.

Exercise 2.10.7. Compute the improper integral $\int_0^1 x^{-1/2}\,dx$.

Exercise 2.10.8. Compute the improper integral $\int_{-1}^1 x^{-2/3}\,dx$.

Conventionally, we say that an improper integral **converges** whenever the limit of definition exists, and we say that it **diverges** if the limit does not exist. Even if we cannot explicitly compute an improper integral, the **Comparison Theorem** allows us to say whether it converges or diverges.

Theorem 2.10.9 (Comparison Theorem for Improper Integrals)**.** *Consider any pair of continuous functions $f(x)$ and $g(x)$ such that $f(x) \geq g(x) \geq 0$ for all real numbers $x \geq a$.*

(a.) *If $\int_a^\infty f(x)\,dx$ converges, then $\int_a^\infty g(x)\,dx$ converges.*

(b.) *If $\int_a^\infty g(x)\,dx$ diverges, then $\int_a^\infty f(x)\,dx$ diverges.*

One can make analogous statements for the improper integrals $\int_{-\infty}^b f(x)\,dx$ and $\int_{-\infty}^b g(x)\,dx$, doubly improper integrals, and improper integrals of a function with a vertical asymptote.

Exercise 2.10.10. Determine if the improper integral $\int_0^\infty xe^x\,dx$ converges.

Exercise 2.10.11. Determine if the improper integral $\int_0^\infty x^{-2}\sin^2(x)\,dx$ converges.

Chapter 3

Physical Applications of Integration

3.1 Regions and Areas Bounded by Curves

Our introduction to the notion of integration already gave us an interpretation of the definite integral $\int_a^b f(x)\,dx$ as the net area bounded by the graph of the curve $f(x)$ and the x-axis. Consequently, there are myriad benefits of using a definite integral to capture information about real-life observations: the Net Change Theorem states that if $f'(x)$ is the derivative of an integrable function $f(x)$, then the definite integral $\int_a^b f'(x)\,dx = f(b) - f(a)$ measures the net change of $f(x)$ over the closed interval $[a, b]$. For instance, if $f'(t)$ is the velocity of a particle observed from time $t = a$ to time $t = b$, then the definite integral $f(b) - f(a) \int_a^b f'(t)\,dt$ is the net displacement of the particle (i.e., the net distance traveled by the particle) during the time frame in which we observed it.

Crucially, we can view the x-axis of the Cartesian plane as the curve $y = g(x) = 0$, hence if $f(x)$ satisfies that $f(x) \geq g(x) = 0$ (i.e., $f(x)$ is non-negative) for all real numbers x such that $a \leq x \leq b$, then the definite integral $\int_a^b f(x)\,dx = \int_a^b [f(x) - g(x)]\,dx$ measures the area between the curves $f(x)$ and $g(x)$. Generalizing this notion gives us a way to measure the area between any two curves $f(x)$ and $g(x)$ satisfying $f(x) \geq g(x)$ for all real numbers x such that $a \leq x \leq b$.

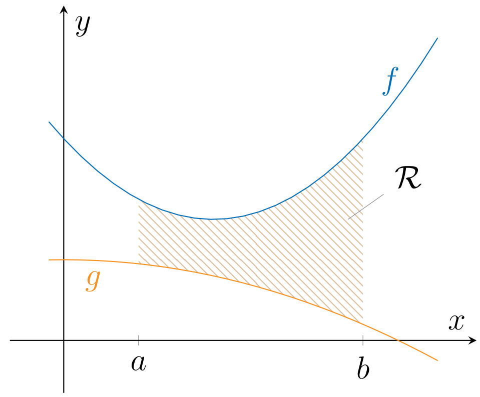
Formula 3.1.1 (Area Formula for a Region Bounded by Four Curves)**.** *Consider any pair of functions $f(x)$ and $g(x)$ satisfying that $f(x) \geq g(x)$ for all real numbers x such that $a \leq x \leq b$. Provided that $f(x)$ and $g(x)$ are both integrable on $[a, b]$, the curves $f(x)$, $g(x)$, $x = a$, and $x = b$ bound a region $\mathcal{R}$ in the Cartesian plane of finite area. Explicitly, the area of this region is given by*

$$\text{area}(\mathcal{R}) = \int_a^b [f(x) - g(x)]\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx.$$

Basically, the proof of this formula boils down to the fact if $f(x)$ and $g(x)$ are both integrable functions, then for any choice of partition $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ of the interval $[a, b]$ and any choice of sample points x_i^*, the limit that defines the integral of $f(x) - g(x)$ over $[a, b]$ exists.

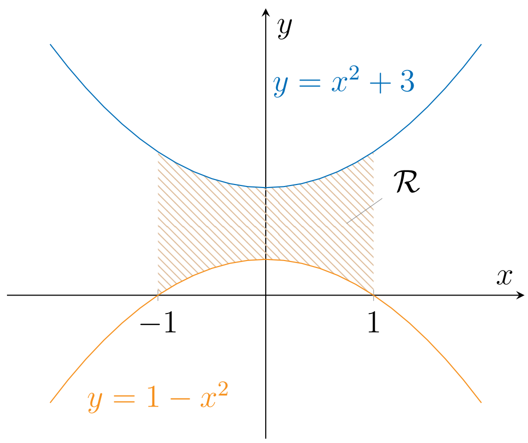
$$\int_a^b [f(x) - g(x)]\,dx = \lim_{n\to\infty} \sum_{i=1}^n [f(x_i^*) - g(x_i^*)]\,\Delta x_i = \lim_{n\to\infty} \sum_{i=1}^n [f(x_i^*)]\,\Delta x_i - \lim_{n\to\infty} \sum_{i=1}^n g(x_i^*)\,\Delta x_i$$

We will repeatedly see this kind of rationale applied throughout this chapter of the lecture notes: if we want to compute area, volume, moment of inertia, work, etc., we can approximate by a Riemann sum and take a limit to reduce the error of our approximation to zero.



Example 3.1.2. Compute the area of the region $\mathcal{R}$ bounded by $y = x^2 - 3$, $y = 1 - x^2$, $x = -1$, and $x = 1$.

Solution. We begin by drawing each of the four curves $y = x^2 + 3$, $y = 1 - x^2$, $x = -1$, and $x = 1$.

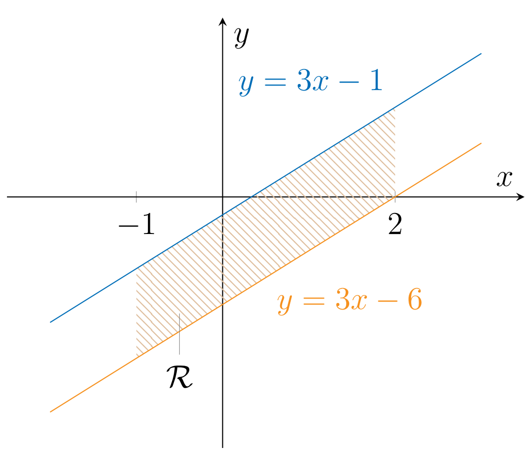


Consequently, we may label the curves as $f(x) = x^2 - 3$ and $g(x) = 1 - x^2$; then, according to the Area Formula for a Region Bounded by Four Curves, we have that

$$\operatorname{area}(\mathcal{R}) = \int_{-1}^{1} [(x^2 - 3) - (1 - x^2)]\, dx = \int_{-1}^{1} (2x^2 - 4)\, dx = \left[\frac{2}{3}x^3 - 4x\right]_{-1}^{1} = \frac{4}{3} + 4. \qquad \diamond$$

Example 3.1.3. Compute the area of the region $\mathcal{R}$ bounded by the curves $y = 3x - 1$, $y = 3x - 6$, $x = -1$, and $x = 2$. Explain what would happen if the curves $x = -1$ and $x = 2$ were not given.

Solution. We begin by drawing each of the four curves $y = 3x - 1$, $y = 3x - 6$, $x = -1$, and $x = 1$.



Consequently, we may label the curves as $f(x) = 3x - 1$ and $g(x) = 3x - 6$; then, according to the Area Formula for a Region Bounded by Four Curves, we have that

$$area(\mathcal{R}) = \int_{-1}^{1} [(3x-1) - (3x-6)]\, dx = \int_{-1}^{1} 5\, dx = 5x\Big|_{-1}^{1} = 5(1) - 5(-1) = 10.$$

Last, observe that if the curves $x = -1$ and $x = 1$ were not given in the problem statement, then the area bounded by the parallel lines $y = 3x - 1$ and $y = 3x - 6$ would be infinite! ⋄

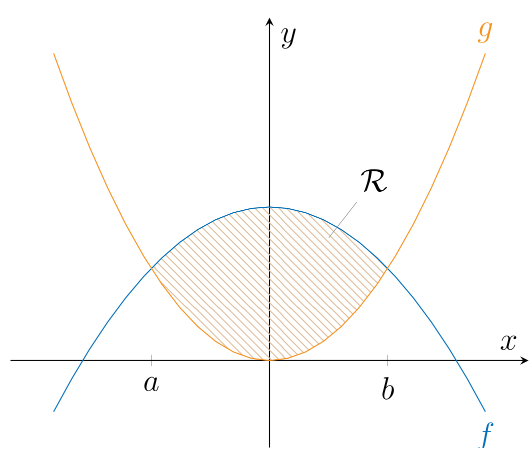
Example 3.1.4. Consider the functions $f(x) = x^2 + 1$ and $g(x) = 2x$. Observe that $f(x) = g(x)$ if and only if $x^2 + 1 = 2x$ if and only if $x^2 - 2x + 1 = 0$ if and only if $(x-1)^2 = 0$ if and only if $x = 1$. By the same rationale, we have that $f(x) \geq g(x)$ for all real numbers $x \geq 1$ because $f(x) - g(x) = (x-1)^2 \geq 0$ for all real numbers $x \geq 1$. We can therefore consider the region $\mathcal{R}$ bounded by the curves $f(x) = x^2 + 1$, $g(x) = 2x$, and $x = 4$; the area of this region is given by

$$\text{area}(\mathcal{R}) = \int_1^4 [f(x) - g(x)]\, dx = \int_1^4 (x-1)^2\, dx = \left[\frac{(x-1)^3}{3}\right]_1^4 = 9.$$

Example 3.1.5. Consider the functions $f(x) = e^{2x}$ and $g(x) = e^x$. Observe that $f(x) = g(x)$ if and only if $e^{2x} = e^x$ if and only if $e^{2x} - e^x = 0$ if and only if $e^x(e^x - 1) = 0$ if and only if $e^x - 1 = 0$ if and only if $e^x = 1$ if and only if $x = 0$. Even more, we have that $e^{2x} \geq e^x$ for all real numbers $x \geq 0$ because $e^{2x} - e^x = e^x(e^x - 1)$ and $e^x - 1 \geq 0$ for all real numbers $x \geq 0$. Consequently, the region $\mathcal{R}$ bounded by the curves $f(x) = e^{2x}$, $g(x) = e^x$, and $x = \ln(10)$ has area given by

$$\text{area}(\mathcal{R}) = \int_0^{\ln(10} [f(x) - g(x)]\, dx = \int_0^{10} (e^{2x} - e^x)\, dx = \left[\frac{e^{2x}}{2} - e^x\right]_0^{\ln(10)} = \frac{81}{2}.$$

Often, we will be interested in the region bounded by two curves $f(x)$ and $g(x)$ satisfying that $f(a) = g(a)$, $f(b) = g(b)$, and $f(x) \geq g(x)$ for all real numbers x such that $a \leq x \leq b$.



Below, we outline the basic strategy to find the area bounded by curves $f(x)$ and $g(x)$ such that $f(a) = g(a)$, $f(b) = g(b)$, and $f(x) \geq g(x)$ for all real numbers x such that $a \leq x \leq b$.

Algorithm 3.1.6 (Determining the Region Bounded by four Curves)**.** Complete the following steps to determine the area bounded by the graphs of some curves $f_1(x)$, $f_2(x)$, $x = a$, and $x = b$.

(i.) Provided that the equation $f_1(x) = f_2(x)$ is readily solved by algebraic methods, one can find the intersection point(s) of $f_1(x)$ and $f_2(x)$ by solving this equation in terms of x.

(ii.) By plugging in points, we can determine whether if $f_1(x) \geq f_2(x)$ or $f_1(x) \leq f_2(x)$ on the interval. Label the larger function as $f(x)$, and label the smaller function as $g(x)$.

(iii.) By the Area Formula for a Region Bounded by Four Curves, the area of the region $\mathcal{R}$ bounded by the curves can be found using $x = a$, $x = b$, and the intersection points of $f_1(x)$ and $f_2(x)$.

(iv.) Conversely, if the equation $f_1(x) = f_2(x)$ is difficult to solve, then choose several x-values so that the value of the function $f_1(x)$ is known (or can easily approximate and accurately plotted on a graph). Use the rule of thumb that if f is a polynomial of degree n, it is best to choose $n + 1$ different x-values to plot $f(x)$; use at least four points for other functions.

(v.) Plot the corresponding points $(x, f_1(x))$, and use these to sketch the graph of $f_1(x)$.

(vi.) Repeat the second and third steps of the algorithm for the function $f_2(x)$.

(vii.) Label the top function as $f(x)$ and the bottom function as $g(x)$ based on the graph.

(viii.) Use the Area Formula for a Region Bounded by Four Curves to compute the area of $\mathcal{R}$.

Example 3.1.7. Compute the area of the region $\mathcal{R}$ bounded by $y = -x^2 + 4$ and $y = x^2 - 4$.

Solution. We must first determine the intersection points of the curves $y = -x^2+4$ and $y = x^2-4$. Consequently, we solve the equation $-x^2 + 4 = x^2 - 4$. We find that $2x^2 = 8$ so that $x^2 = 4$ and $x = -2$ or $x = 2$. Even more, the inequality $-x^2+4 \geq x^2-4$ holds for all real numbers x such that $-2 \leq x \leq 2$ because $-0^2 + 4 = 4 > -4 = 0^2 - 4$. We conclude therefore that

$$\operatorname{area}(\mathcal{R}) = \int_{-2}^{2} [(-x^2 + 4) - (x^2 - 4)]\,dx = \int_{-2}^{2} (-2x^2 + 8)\,dx = \left[-\frac{2}{3}x^3 + 8x\right]_{-2}^{2} = \frac{64}{3}. \qquad \diamond$$

Exercise 3.1.8. Compute the area of the region $\mathcal{R}$ bounded by the curves $y = \sqrt{x}$ and $y = x^2$.

Generally, we say that a region $\mathcal{R}$ in the Cartesian plane is **vertically simple** if there exist curves $y = f_1(x)$, $y = f_2(x)$, $x = a$, and $x = b$ such that every point (x, y) in the region $\mathcal{R}$ satisfies that $a \leq x \leq b$ and $f_1(x) \leq y \leq f_2(x)$. Graphically, the curve $y = f_1(x)$ can be viewed as the "bottom" of the region $\mathcal{R}$, and the curve $y = f_2(x)$ can be viewed as the "top" of $\mathcal{R}$. Commonly, we will refer to $f_1(x)$ as y_{bottom} and $f_2(x)$ as y_{top}. Below is a reformulation of the above area formula.

Formula 3.1.9 (Area Formula for a Vertically Simple Region)**.** *Given a vertically simple region* $\mathcal{R}$ *bounded by the curves* $y_{\text{top}} = f_2(x)$, $y_{\text{bottom}} = f_1(x)$, $x = a$, *and* $x = b$, *we have that*

$$\operatorname{area}(\mathcal{R}) = \int_a^b (y_{\text{top}} - y_{\text{bottom}})\,dx = \int_a^b [f_2(x) - f_1(x)]\,dx.$$

Our regions have been thus far vertically simple, hence we have been able to compute their areas using the above formula. Unfortunately, there exist regions that are not vertically simple.

Exercise 3.1.10. Prove that the region $\mathcal{R}$ bounded by the curves $y = x$, $y = -x$, and $y = -2$ is not vertically simple; then, express $\mathcal{R}$ as the union of two vertically simple regions $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$, and find the area of $\mathcal{R}$ by using the fact that $\text{area}(\mathcal{R}_1 \cup \mathcal{R}_2) = \text{area}(\mathcal{R}_1) + \text{area}(\mathcal{R}_2)$.

Exercise 3.1.10 exhibits a region $\mathcal{R}$ that is not vertically simple; however, if we tilt our head to the side, then we would see a vertically simple region. Explicitly, we say that the region $\mathcal{R}$ is **horizontally simple** if there exist curves $x = g_1(y)$, $x = g_2(y)$, $y = c$, and $y = d$ such that every point (x, y) in the region $\mathcal{R}$ satisfies that $g_1(y) \leq x \leq g_2(y)$ and $c \leq y \leq d$. Graphically, the curve $x = g_1(y)$ is the "left" function, and the curve $x = g_2(y)$ is the "right" function. Like before, we will refer to $g_1(y)$ as x_{left} and $g_2(y)$ as x_{right}, so we have the following area formula.

Formula 3.1.11 (Area Formula for a Horizontally Simple Region)**.** *Given a horizontally simple region $\mathcal{R}$ bounded by the curves $x_{\text{right}} = g_2(y)$ and $x_{\text{left}} = g_1(y)$, $y = c$, and $y = d$, we have that*

$$\text{area}(\mathcal{R}) = \int_c^d (x_{\text{right}} - x_{\text{left}})\, dy = \int_c^d [g_2(y) - g_1(y)]\, dy.$$

Example 3.1.12. Prove that the region $\mathcal{R}$ of Example 3.1.10 is horizontally simple by exhibiting well-defined curves $x_{\text{left}} = g_1(y)$, $x_{\text{right}} = g_2(y)$, $y = c$, and $y = d$; then, compute the area of $\mathcal{R}$.

Solution. Observe that the curves $x = g_1(y) = y$ and $x = g_2(y) = -y$ intersect at $y = 0$. Even more, for all real numbers y such that $-2 \leq y \leq 0$, we have that $0 \leq -y \leq 2$ so that $g_1(y) \leq x \leq g_2(y)$ for all real numbers $-2 \leq y \leq 0$. We conclude that the region $\mathcal{R}$ is horizontally simple with $x_{\text{left}} = y$ and $x_{\text{right}} = -y$. By the Area Formula for a Horizontally Simple Region, we conclude that

$$\text{area}(\mathcal{R}) = \int_{-2}^0 (x_{\text{right}} - x_{\text{left}})\, dy = \int_{-2}^0 (-y - y)\, dy = \int_{-2}^0 -2y\, dy = \left[-y^2\right]_{-2}^0 = 4. \qquad \diamond$$

Example 3.1.13. Compute the area of the region $\mathcal{R}$ bounded by $x = \sqrt{1 - y^2}$ and $x = 0$.

Solution. Observe that $\sqrt{1 - y^2} = 0$ if and only if $1 - y^2 = 0$ if and only if $y^2 = 1$ if and only if $y = -1$ or $y = 1$. Even more, for all real numbers y such that $-1 \leq y \leq 1$, we have that $\sqrt{1 - y^2} \geq 0$, hence the region $\mathcal{R}$ is horizontally simple with $x_{\text{left}} = 0$ and $x_{\text{right}} = \sqrt{1 - y^2}$. By the Area Formula for a Horizontally Simple Region, we conclude that

$$\text{area}(\mathcal{R}) = \int_{-1}^1 (x_{\text{right}} - x_{\text{left}})\, dy = \int_{-1}^1 \sqrt{1 - y^2}\, dy = \frac{\pi}{2}.$$

Explicitly, the integral can be evaluated by elementary geometry since $\mathcal{R}$ is half of the unit circle. $\diamond$

Unfortunately, it is also possible for a region to be neither vertically nor horizontally simple.

Example 3.1.14. Prove that the region $\mathcal{R}$ bounded by the curves $y = x - 2$, $y = 2 - x$, $y = -x + 2$, and $y = -x - 2$ is neither vertically nor horizontally simple; then, compute the area of $\mathcal{R}$.

Solution. Graphing the region, we find that $\mathcal{R}$ is not vertically simple: indeed, for all real numbers x such that $-2 \leq x \leq 0$, we have that $-x - 2 \leq y \leq x + 2$; however, for all real numbers x such that $0 \leq x \leq 2$, we have that $x - 2 \leq y \leq -x + 2$. Consequently, the top and bottom curves of $\mathcal{R}$ are not well-defined. By the symmetry of $\mathcal{R}$, it is not horizontally simple, either.

On the bright side, as our above analysis illustrates, we can represent $\mathcal{R}$ as the union of two vertically simple regions: indeed, we have that $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ for the vertically simple regions

$$\mathcal{R}_1 = \{(x, y) \mid -2 \leq x \leq 0 \text{ and } -x-2 \leq y \leq x+2\} \text{ and}$$
$$\mathcal{R}_2 = \{(x, y) \mid 0 \leq x \leq 2 \text{ and } x-2 \leq y \leq -x+2\}.$$

Consequently, in view of the fact that $\text{area}(\mathcal{R}) = \text{area}(\mathcal{R}_1) + \text{area}(\mathcal{R}_2)$, we find that

$$\text{area}(\mathcal{R}) = \int_{-2}^{0} [(x+2) - (-x-2)]\,dx + \int_{0}^{2} [(-x+2) - (x-2)]\,dx$$

$$= \int_{-2}^{0} (2x+4)\,dx + \int_{0}^{2} (-2x+4)\,dx = \left[x^2+4x\right]_{-2}^{0} + \left[-x^2+4x\right]_{0}^{2} = 8. \qquad \diamond$$

Our above exposition completely determines how to compute the area of a region as soon as we can identify it as vertically or horizontally simple; however, there remains some nuance to these types of problems. Our previous example establishes the existence of regions that are neither vertically nor horizontally simple, so the question remains as to how we deal with these. One strategy is to break up such a region into subregions that are either vertically or horizontally simple. (Later, in Calculus III, we will learn the change of variables method that will make this issue more manageable.)

On the other hand, it is also completely possible that we are handed a region that is both vertically and horizontally simple, and the description of the region as vertically simple renders the integral infeasible to compute. Our best bet in this case is to check the description of the region as horizontally simple and hope that the integrand works out to be nicer in this lens.

Example 3.1.15. Compute $-\int_0^1 \ln(x)\,dx$ by viewing it as the area of some region $\mathcal{R}$.

Solution. Considering that $\ln(x) \leq 0$ for all real numbers x such that $0 < x \leq 1$, it follows that $-\int_0^1 \ln x\,dx$ is the area of the region $\mathcal{R}$ bounded by the curves $y = 0$, $y = \ln(x)$, $x = 0$, and $x = 1$. Consequently, we may view $\mathcal{R}$ as the horizontally simple region bounded by the curves $x = 0$, $x = e^y$, and $y = 0$. By the Area Formula for a Horizontally Simple Region, we conclude that

$$\int_0^1 \ln(x)\,dx = \text{area}(\mathcal{R}) = \int_{-\infty}^{0} e^y\,dy = \lim_{a \to -\infty} \int_a^0 e^y\,dy = \lim_{a \to -\infty} \left[e^y\right]_a^0 = \lim_{a \to -\infty} (1 - e^a) = 1. \qquad \diamond$$

3.2 Volume, Density, and Average Value

Given any region $\mathcal{R}$ in the Cartesian plane that is bounded by several curves, by the intuition of the previous section, we can endeavor to find the area of $\mathcal{R}$ by viewing the region as the union of some (vertically or horizontally) simple subregions and summing the respective areas of each subregion. Consequently, we might suspect that a similar approach could be used to compute the volume of a three-dimensional solid. Explicitly, by taking n slices of equal width Δx perpendicular to the solid and using the fact that the volume of a solid of constant area is equal to the product of area times width, we may approximate the volume of the solid as the sum of the product of the **cross-sectional area** $\alpha(x_i)$ of the ith **cross section** of the solid and the width Δx of each slice

$$\text{volume of a solid of variable cross-sectional area } \alpha(x) \approx \sum_{i=1}^{n} \alpha(x_i)\Delta x.$$

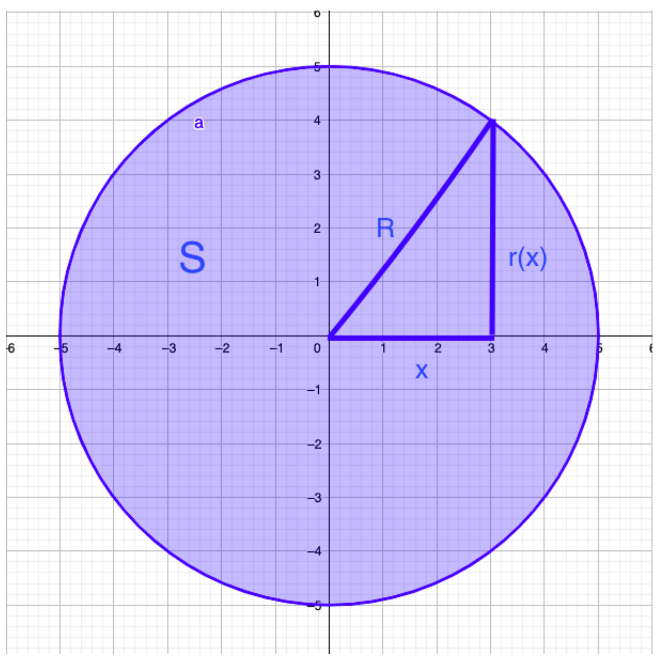
By viewing this quantity as a Riemann sum and taking the limit as n approaches ∞, we have that

$$\text{volume of a solid of variable cross-sectional area } \alpha(x) = \int_a^b \alpha(x)\,dx.$$

Formula 3.2.1 (Volume of a Three-Dimensional Solid with Variable Cross-Sectional Area)**.** *Consider any three-dimensional solid $\mathcal{S}$ whose cross-sectional area is determined by a continuous real function $\alpha(x)$ of a real variable x for all real numbers x such that $a \leq x \leq b$. We have that*

$$\text{volume}(\mathcal{S}) = \int_a^b \alpha(x)\,dx.$$

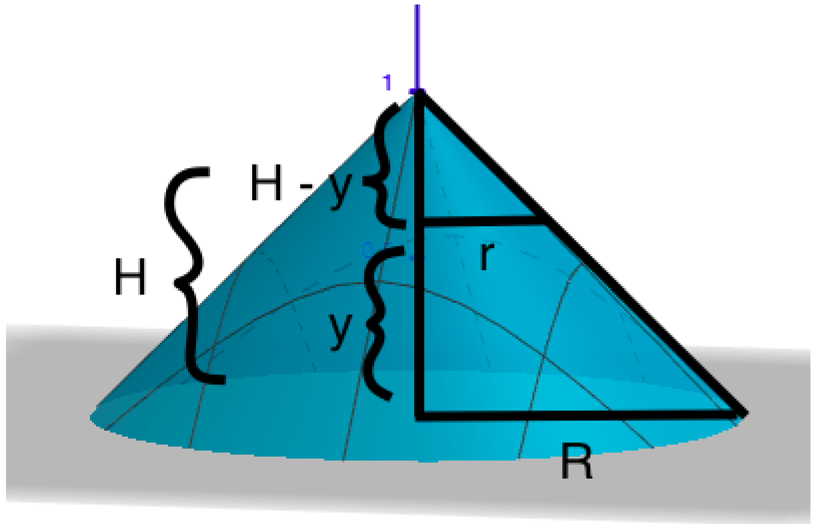
Example 3.2.2. Let us demonstrate that the volume of a sphere of radius $R > 0$ is $\frac{4}{3}\pi R^3$. Consider the sphere of radius $R > 0$ centered at the origin. Each of the cross sections of the sphere is a circle of radius $r(x)$ of each real number $-R \leq x \leq R$, hence the cross-sectional area of the cross section of the sphere at x is simply the area of a circle of radius $r(x)$, i.e., we have that $a(x) = \pi[r(x)]^2$. Consider the diagram below of the cross section of the sphere of radius R at the real number x.



By the Pythagorean Theorem, we have that $R^2 = x^2 + [r(x)]^2$ so that $[r(x)]^2 = R^2 - x^2$ and $\alpha(x) = \pi(R^2 - x^2)$. By the formula for the Volume of a Three-Dimensional Solid with Variable Cross-Sectional Area, we conclude that the volume of the sphere of radius $R > 0$ is given by

$$\int_a^b \alpha(x)\,dx = \int_{-R}^R \pi(R^2 - x^2)\,dx = \pi\left[R^2x - \frac{x^3}{3}\right]_{-R}^R = \frac{4}{3}\pi R^3.$$

Example 3.2.3. Likewise, we may demonstrate that the volume of a right-circular cone of radius $R > 0$ and height $H > 0$ is $\frac{1}{3}\pi R^2 H$. Consider the right-circular cone of radius $R > 0$ and height $R > 0$ as the three-dimensional solid whose base is a circle of radius $R > 0$ centered at the origin and whose horizontal cross sections from $y = 0$ to $y = H$ are circles of radius $r(y)$. Explicitly, the diagram below depicts the right-circular cone of radius $R > 0$ and $H > 0$ as we have described it.



Observe that in the diagram above, there are similar right triangles determined by the acute angle θ formed by the y-axis and the vertex of the cone. Consequently, we have that

$$\frac{r(y)}{H-y} = \tan(\theta) = \frac{R}{H} \text{ so that } r(y) = \frac{R}{H}(H-y) = R - \frac{R}{H}y.$$

Ultimately, it follows that the cross-sectional area of each horizontal slice of the cone at height y is $\pi[r(y)]^2$. By the formula for the Volume of a Three-Dimensional Solid with Variable Cross-Sectional Area, we conclude that the volume of a right-circular cone of radius $R > 0$ and height $H > 0$ is

$$\int_0^H \pi\left(R - \frac{R}{H}y\right)^2 dy = \frac{\pi H}{R}\int_0^R u^2\,du = \frac{\pi H}{R}\left[\frac{u^3}{3}\right]_0^R = \frac{1}{3}\pi R^2 H.$$

General physical principles dictate that the mass m of an object of length ℓ and constant linear density ρ is given by $m = \rho\ell$. Consider any object of variable linear density $\rho(x)$ for some real variable x such that $a \leq x \leq b$ for some real numbers a and b. Like before, if we split the object into n slices of equal length Δx, then we can approximate its mass by the Riemann sum

$$\text{mass of an object of variable linear density } \rho(x) \approx \sum_{i=1}^{n} \rho(x_i)\Delta x.$$

By taking the limit as the number of slices n tends to ∞, we reduce our error to zero, and we obtain

$$\text{mass of an object of variable linear density } \rho(x) = \int_a^b \rho(x)\,dx.$$

Formula 3.2.4 (Mass of an Object with Variable Linear Density)**.** *Consider any three-dimensional object $\mathcal{O}$ whose linear density is determined by a continuous real function $\rho(x)$ of a real variable x for all real numbers x such that $a \leq x \leq b$. We have that*

$$\text{mass}(\mathcal{O}) = \int_a^b \rho(x)\,dx.$$

Example 3.2.5. Let us compute the mass of a rod of unit length and linear density $\rho(x) = xe^{x^2}$. By the formula for the Mass of an Object with Variable Linear Density, the mass of the rod is

$$\int_0^1 xe^{x^2}\,dx = \int_0^1 \frac{1}{2}e^u\,du = \left[\frac{1}{2}e^u\right]_0^1 = \frac{1}{2}(e-1).$$

Given any list of n real number $a_1, \dots, a_n$, recall that the **average** of these real numbers is

$$\frac{a_1 + \cdots + a_n}{n} = \frac{1}{n}\sum_{i=1}^n a_i.$$

Consequently, we may use this approach if we wish to approximate the average value of a function $f(x)$ that is integrable on a closed interval $[a, b]$. Explicitly, we may choose n values $f(x_1), \dots, f(x_n)$ for some equally-spaced real numbers $a = x_1 \leq \cdots \leq x_n = b$. Using the fact that $\Delta x = \frac{b-a}{n}$ is the distance between any two consecutive x-values, our above displayed equation gives that

$$\text{average value of } f(x) \text{ on } [a,b] \approx \frac{1}{n}\sum_{i=1}^n f(x_i) = \frac{1}{b-a}\cdot\frac{b-a}{n}\sum_{i=1}^n f(x_i) = \frac{1}{b-a}\sum_{i=1}^n f(x_i)\Delta x.$$

By recognizing this as a Riemann sum as taking the limit as n approaches ∞, we find that

$$\text{average value of } f(x) \text{ on } [a,b] = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

Formula 3.2.6 (Average Value of an Integrable Real Function on a Closed and Bounded Interval)**.** *Given any real function $f(x)$ that is integrable on a closed and bounded interval $[a, b]$ for some real numbers $a < b$, the average value of $f(x)$ on $[a, b]$ is given by*

$$\textit{average value of } f(x) \textit{ on } [a,b] = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

Example 3.2.7. We will compute in this example the average value of the function $f(x) = x^{-1}$ on the interval $\left[\frac{1}{e}, 1\right]$. Observe that $f(x)$ is continuous for all real numbers $x \neq 0$, hence it is integrable on the closed and bounded interval in question. By the formula for the Average Value of an Integrable Real Function on a Closed and Bounded Interval, we conclude that

$$\text{average value of } x^{-1} \text{ on } \left[\frac{1}{e}, 1\right] = \int_{1/e}^1 \frac{1}{x}\,dx = [\ln(x)]_{1/e}^1 = -\ln\left(\frac{1}{e}\right) = \ln(e) = 1.$$

One of the most important applications of the average value of a function is the following.

Theorem 3.2.8 (Mean Value Theorem for Integrals)**.** *Given any real function $f(x)$ that is continuous on a closed interval $[a, b]$, there exists a real number c satisfying that $a \leq c \leq b$ and*

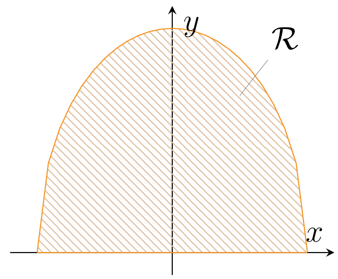
$$f(c) = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

Exercise 3.2.9. Prove that if a vehicle travels through a 325 unit-long tunnel in four minutes and the speed limit in the tunnel is 80 units per minute, then the vehicle broke the speed limit.

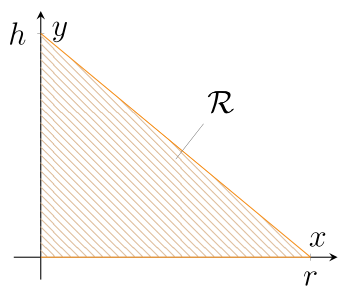
3.3 Disk and Washer Method

Last section, we developed strategies to find the area of a two-dimensional region bounded by several curves in the Cartesian plane. We prefer regions that are either vertically or horizontally simple. Our next task expounds upon this theme to find the volume of a three-dimensional object $\mathcal{S}$ called a **solid of revolution** that is obtained by rotating a region of the Cartesian plane about an axis.

Example 3.3.1. We can obtain any ball (i.e., a filled sphere) of radius r by rotating the semicircular region $\mathcal{R}$ bounded by the curves $y = \sqrt{r^2 - x^2}$ and $y = 0$ about the x-axis.



Example 3.3.2. We can obtain any right-circular cone of radius r and height h by rotating the triangular region $\mathcal{R}$ bounded by the curves $y = -\frac{h}{r}x + h$, $x = 0$, and $y = 0$ about the y-axis.



Consider any function $f(x)$ that is continuous on a closed interval $[a, b]$ and satisfies that $f(x) \geq 0$ for all real numbers x such that $a \leq x \leq b$. Observe that if we rotate $f(x)$ about the x-axis from $x = a$ to $x = b$, then we obtain a solid of revolution $\mathcal{S}$. Each vertical cross section of $\mathcal{S}$ at x_i^* is a disk of radius $f(x_i^*)$, hence the cross-sectional area of each slice of $\mathcal{S}$ is $\pi[f(x_i^*)]^2$. Consequently, approximating the volume of $\mathcal{S}$ via the cross-sectional area of n slices of thickness Δx_i yields that

$$\text{volume}(\mathcal{S}) \approx \sum_{i=1}^{n} \pi[f(x_i^*)]^2 \, \Delta x_i = \pi \sum_{i=1}^{n} [f(x_i^*)]^2 \, \Delta x_i.$$

By taking the limit as n tends to infinity, we reduce our error to zero and obtain the following.

Formula 3.3.3 (Disk Method)**.** *Given any function $f(x)$ that is continuous on a closed interval $[a, b]$ and satisfies that $f(x) \geq 0$ for all real numbers x such that $a \leq x \leq b$, the solid of revolution $\mathcal{S}$ obtained by rotating $f(x)$ about the x-axis from $x = a$ to $x = b$ has the following volume.*

$$\text{volume}(\mathcal{S}) = \pi \int_a^b [f(x)]^2 \, dx$$

Likewise, if $g(y)$ is any function that is continuous on a closed interval $[c, d]$ along the y-axis and satisfies that $g(y) \geq 0$ for all real numbers y such that $c \leq y \leq d$, the solid of revolution $\mathcal{S}$ obtained by rotating $g(y)$ about the y-axis from $y = c$ to $y = d$ has the following volume.

$$\operatorname{volume}(\mathcal{S}) = \pi \int_c^d [g(y)]^2 \, dy$$

Example 3.3.4. We can obtain any ball of radius r by rotating the semicircular region $\mathcal{R}$ bounded by the curves $y = \sqrt{r^2 - x^2}$ and $y = 0$ about the x-axis. Certainly, the function $f(x) = \sqrt{r^2 - x^2}$ is non-negative, hence the Disk Method yields that the volume of a ball of radius r is given by

$$\pi \int_{-r}^{r} (r^2 - x^2) \, dx = \pi \left[r^2 x - \frac{x^3}{3} \right]_{-r}^{r} = \frac{4\pi r^3}{3}.$$

Example 3.3.5. We can obtain any right-circular cone of radius r and height h by rotating the triangular region $\mathcal{R}$ bounded by the curves $y = -\frac{h}{r}x + h$, $x = 0$, and $y = 0$ about the y-axis. Observe that the function $g(y) = -\frac{r}{h}y + r$ is positive for all real numbers y such that $0 \leq y \leq h$. By the Disk Method, the volume of a right-circular cone of radius r and height h is given by

$$\pi \int_0^h \left(-\frac{r}{h}y + r \right)^2 dy = \frac{\pi h}{r} \int_0^r u^2 \, du = \frac{\pi h}{r} \left[\frac{u^3}{3} \right]_0^r = \frac{\pi}{3} r^2 h.$$

Example 3.3.6. We can obtain a right-circular cylinder of radius r and height h by rotating the rectangular region $\mathcal{R}$ bounded by the curves $y = 0$, $y = h$, $x = 0$, and $x = r$ about the y-axis. By the Disk Method, we can find the volume of this solid of revolution by computing the integral

$$\pi \int_0^h r^2 \, dy = \pi \left[r^2 y \right]_0^h = \pi r^2 h.$$

Example 3.3.7. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the curve $f(x) = \sqrt{x}$ about the x-axis from $x = 0$ to $x = 4$.

Solution. Considering that $f(x) \geq 0$ for all real numbers x such that $0 \leq x \leq 4$, we have that

$$\operatorname{volume}(\mathcal{S}) = \pi \int_0^4 x \, dx = \pi \left[\frac{x^2}{2} \right]_0^4 = 8\pi. \qquad \diamond$$

Given any pair of functions $f(x)$ and $g(x)$ that are continuous on a closed interval $[a, b]$ and satisfy that $f(x) \geq g(x) \geq 0$ for all real numbers x such that $a \leq x \leq b$, we may consider the region $\mathcal{R}$ bounded by the curves $y = f(x)$, $y = g(x)$, $x = a$, and $x = b$. By rotating $\mathcal{R}$ about the x-axis, we obtain a solid of revolution $\mathcal{S}$. Each vertical cross section of $\mathcal{S}$ at x_i^* is the difference between a disk of radius $f(x_i^*)$ and a disk of radius $g(x_i^*)$ — called a **washer** of inner radius $g(x_i^*)$ and outer radius $f(x_i^*)$ — hence the cross-sectional area of each slice is $\pi[f(x_i^*)]^2 - \pi[g(x_i^*)]^2$. Consequently, approximating the volume of $\mathcal{S}$ via the cross-sectional area of n slices of thickness Δx_i yields that

$$\operatorname{volume}(\mathcal{S}) \approx \sum_{i=1}^{n} (\pi[f(x_i^*)]^2 - \pi[g(x_i^*)]^2) \, \Delta x_i = \pi \sum_{i=1}^{n} ([f(x_i^*)]^2 - [g(x_i^*)]^2) \, \Delta x_i.$$

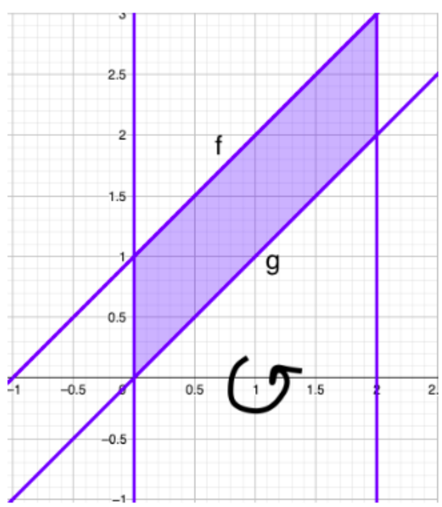
By taking the limit as n tends to infinity, we reduce our error to zero and obtain the following.

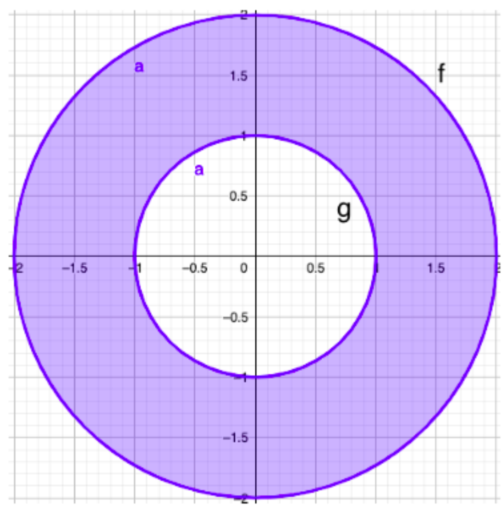
Formula 3.3.8 (Washer Method)**.** *Given any functions $f(x)$ and $g(x)$ that are continuous on a closed interval $[a, b]$ and satisfy that $f(x) \geq g(x) \geq 0$ for all real numbers x such that $a \leq x \leq b$, the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by $y = f(x)$, $y = g(x)$, $x = a$, and $x = b$ about the x-axis has cross-sections that are washers. Even more, it holds that*

$$\text{volume}(\mathcal{S}) = \pi \int_a^b ([f(x)]^2 - [g(x)]^2)\, dx.$$

Likewise, if $h(y)$ and $k(y)$ are any functions that are continuous on a closed interval $[c, d]$ along the y-axis and satisfy that $h(y) \geq k(y) \geq 0$ for all real numbers y such that $c \leq y \leq d$, the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by $x = h(y)$, $x = k(y)$, $y = c$, and $y = d$ about the y-axis has cross-sections that are washers. Even more, it holds that

$$\text{volume}(\mathcal{S}) = \pi \int_c^d ([h(y)]^2 - [k(y)]^2)\, dy.$$





One can see by the diagram above that the Washer Method is so named because the cross sections of the solid of revolution are shaped like washers. Below are some examples for illustration.

Example 3.3.9. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by the curves $y = 2x$, $y = 2x - \frac{1}{2}$, $x = 1$, and $x = 3$ about the x-axis.

Solution. We must first determine the outer radius and the inner radius of each washer that constitutes the cross-section of $\mathcal{S}$. Considering that $2x > 2x - \frac{1}{2}$ for all real numbers x, it follows that $f(x) = 2x$ and $g(x) = 2x - \frac{1}{2}$. By the Washer Method, we conclude that

$$\text{volume}(\mathcal{S}) = \pi \int_1^3 \left[(2x)^2 - \left(2x - \frac{1}{2}\right)^2\right] dx = \pi \int_1^3 \left(2x - \frac{1}{4}\right) dx = \pi \left[x^2 - \frac{1}{4}x\right]_1^3 = \frac{15\pi}{2}. \qquad \diamond$$

Example 3.3.10. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by the curves $y = 4 - x^2$ and $y = 3$ about the x-axis.

Solution. We are not given the bounds on x, hence we must determine the x-values of the intersection points of the two curves $y = 4 - x^2$ and $y = 3$. Explicitly, we must solve the equation $4 - x^2 = 3$. We find that $x^2 - 1 = 0$ so that $x = \pm 1$. Observe that for every real number x such that $-1 \leq x \leq 1$, we have that $4 - x^2 \geq 3$. By the Washer Method, the volume of $\mathcal{S}$ is given by

$$\text{volume}(\mathcal{S}) = \pi \int_{-1}^{1} [(4 - x^2)^2 - 3^2] \, dx = \pi \int_{-1}^{1} (x^4 - 8x^2 + 7) \, dx = \pi \left[\frac{x^5}{5} - \frac{8x^3}{3} + 7x \right]_{-1}^{1} = \frac{136\pi}{15}. \diamond$$

We have thus far seen that we can obtain a surface of revolution by rotating a region in the Cartesian plane about some axis; however, we have limited our attention to the coordinate axes (either the x-axis or the y-axis). Essentially, the method for determining the volume of any solid of revolution obtained by rotating a region about an axis is analogous to what we have done previously. We will assume to this end that C is any real number such that $f(x)$ and $g(x)$ are continuous on a closed interval $[a, b]$ and satisfy that $f(x) \geq g(x) \geq C$ for all real numbers x such that $a \leq x \leq b$. Consider the region $\mathcal{R}$ bounded by the curves $y = f(x)$, $y = g(x)$, $x = a$, and $x = b$. By rotating $\mathcal{R}$ about the line $y = C$, we obtain a solid of revolution $\mathcal{S}$. Each vertical cross section of $\mathcal{S}$ at x_i^* is the difference between a disk of radius $f(x_i^*) - C$ and a disk of radius $g(x_i^*) - C$, hence the cross-sectional area of each slice is $\pi[f(x_i^*) - C]^2 - \pi[g(x_i^*) - C]^2$. Consequently, approximating the volume of $\mathcal{S}$ via the cross-sectional area of n slices of thickness Δx_i yields that

$$\text{volume}(\mathcal{S}) \approx \sum_{i=1}^{n} (\pi[f(x_i^*) - C]^2 - \pi[g(x_i^*) - C]^2) \, \Delta x_i = \pi \sum_{i=1}^{n} ([f(x_i^*) - C]^2 - [g(x_i^*) - C]^2) \, \Delta x_i.$$

By taking the limit as n tends to infinity, we reduce our error to zero and obtain the following.

Formula 3.3.11 (Washer Method for a Non-Coordinate Axis)**.** *Given any real number C and any functions $f(x)$ and $g(x)$ that are continuous on a closed interval $[a, b]$ and satisfy that $f(x) \geq g(x) \geq C$ for all real numbers x such that $a \leq x \leq b$, the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by $y = f(x)$, $y = g(x)$, $x = a$, and $x = b$ about the line $y = C$ satisfies that*

$$\text{volume}(\mathcal{S}) = \pi \int_{a}^{b} ([f(x) - C]^2 - [g(x) - C]^2) \, dx.$$

Likewise, if C is any real number and $h(y)$ and $k(y)$ are any functions that are continuous on a closed interval $[c, d]$ along the y-axis and satisfy that $h(y) \geq k(y) \geq C$ for all real numbers y such that $c \leq y \leq d$, the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by $x = h(y)$, $x = k(y)$, $y = c$, and $y = d$ about the line $x = C$ satisfies that

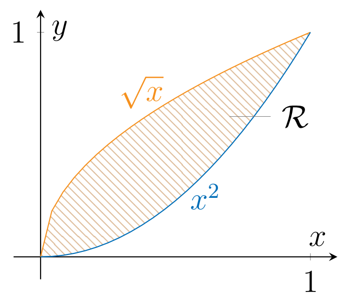
$$\text{volume}(\mathcal{S}) = \pi \int_{c}^{d} ([h(y) - C]^2 - [k(y) - C]^2) \, dy.$$

Example 3.3.12. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region $\mathcal{R}$ bounded by $f(x) = \sqrt{x}$ and $g(x) = x^2$ about the line $y = -1$.

Solution. We will use the Washer Method for a Non-Coordinate Axis. We are not provided with the limits of integration for the definite integral here, so we must compute them. We achieve this by noticing that $f(x) = g(x)$ if and only if $\sqrt{x} = x^2$ if and only if $x = x^4$ if and only if $x = 0$ or

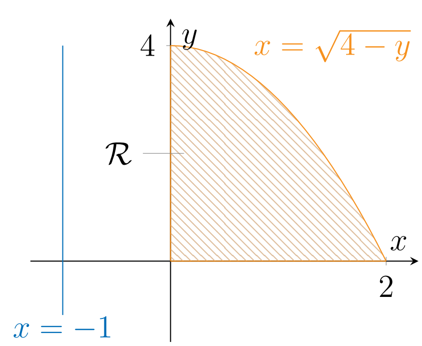
$x^3 = 1$ if and only if $x = 0$ or $x = 1$. Put another way, the curves $f(x) = \sqrt{x}$ and $g(x) = x^2$ intersect when $x = 0$ and $x = 1$. Considering that $\sqrt{x} \geq x^2$ for all real numbers x such that $0 \leq x \leq 1$,

$$\begin{aligned}
\text{volume}(\mathcal{S}) &= \pi \int_0^1 ([\sqrt{x}+1]^2 - [x^2+1]^2)\,dx \\
&= \pi \int_0^1 (x + 2\sqrt{x} + 1 - x^4 - 2x^2 - 1)\,dx \\
&= \pi \int_0^1 (-x^4 - 2x^2 + x + 2\sqrt{x})\,dx \\
&= \pi \left[-\frac{x^5}{5} - \frac{2x^3}{3} + \frac{x^2}{2} + \frac{4x^{3/2}}{3} \right]_0^1 = \frac{29}{30}\pi.
\end{aligned}$$ ◇



Example 3.3.13. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region $\mathcal{R}$ bounded by $f(x) = 4 - x^2$, $g(x) = 0$, and $x = 0$ about the line $x = -1$.

Solution. Care must be taken: the functions we are given are defined with respect to x, but our axis of revolution $x = -1$ is vertical. Consequently, we must solve for $y = f(x) = 4 - x^2$ in terms of y. Observe that $y = 4 - x^2$ if and only if $x^2 = 4 - y$ if and only if $x = h(y) = \sqrt{4-y}$. Crucially, we use the positive square root because the region is bounded on the left by $x = 0$, hence x must be positive. Observe that $x = 0$ is the left-hand curve and $x = \sqrt{4-y}$ is the right-hand curve; the bounds of integration are $y = g(x) = 0$ and $y = 4$ because when $x = 0$, we have that $y = f(0) = 4$.



By the Washer Method for a Non-Coordinate Axis, we conclude that

$$\begin{aligned}
\text{volume}(\mathcal{S}) &= \pi \int_0^4 \left(\left[\sqrt{4-y}+1 \right]^2 - [0+1]^2 \right) dy \\
&= \pi \int_0^4 \left(4 - y + 2\sqrt{4-y} + 1 - 1 \right) dy \\
&= \pi \int_0^4 (-y + 2(4-y)^{1/2} + 4)\, dy \\
&= \pi \left[-\frac{y^2}{2} - \frac{4(4-y)^{3/2}}{3} + 4y \right]_0^4 = \frac{56}{3}\pi.
\end{aligned}$$

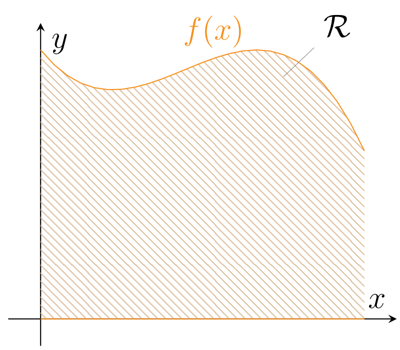
◇

3.4 Shell Method

Unfortunately, the Disk Method and Washer Method work best (and sometimes only) when the axis of revolution under consideration is perpendicular to the region we are rotating. Quite technically, if we wish to revolve a vertically simple region about a line of the form $y = C$ for some real number C (or likewise if we wish to revolve a horizontally simple region about a line of the form $x = C$), then the Disk Method or the Washer Method can be used; however, if we wish to rotate a vertically simple region about a line of the form $x = C$, then we would need to determine the function inverse of the curves in question. Explicitly, Examples 3.3.5 and 3.3.12 bear witness to this process.

Often, the inverse of a function is quite difficult (or impossible) to compute, hence the Disk Method and the Washer Method fail to produce the volume of the resulting solid of revolution.

Example 3.4.1. Observe that the continuous function $f(x) = -x^3 + 2x^2 - x + 1$ has no function inverse on the closed interval $\left[0, \frac{3}{2}\right]$ because $f(x)$ fails the Horizontal Line Test.



Consequently, the Disk Method cannot be used to find the volume of the solid of revolution obtained by rotating $\mathcal{R}$ about the y-axis without splitting the interval $\left[0, \frac{3}{2}\right]$ so that $\mathcal{R}$ is the union of two horizontally-simple regions. Even with this partitioning of $\mathcal{R}$ accomplished, it would subsequently be rather cumbersome to find the inverse function $x = f^{-1}(y)$ on each of these intervals.

Using Example 3.4.1 as motivation, we develop another method to compute the volume of a solid of revolution $\mathcal{S}$. Consider any real numbers $b > a \geq 0$ and any function $f(x)$ that is continuous on the closed interval $[a, b]$ and satisfies that $f(x) \geq 0$ for all real numbers x such that $a \leq x \leq b$. Given any real number $a \leq x_i^* \leq b$, if we revolve any point $(x_i^*, f(x_i^*))$ about the y-axis, we obtain a cylindrical shell of radius x_i^*, height $f(x_i^*)$, and thickness Δx_i. Observe that the surface area of such a cylindrical shell is product of its circumference and its height. Explicitly, the surface area of a cylindrical shell is $2\pi x_i^* f(x_i^*)$, hence each cylindrical shell has volume $2\pi x_i^* f(x_i^*) \Delta x_i$. Consequently, approximating the volume of $\mathcal{S}$ as the sum of the volumes of n cylindrical shells yields that

$$\text{volume}(\mathcal{S}) \approx \sum_{i=1}^{n} 2\pi x_i^* f(x_i^*)\, \Delta x_i = 2\pi \sum_{i=1}^{n} x_i^* f(x_i^*)\, \Delta x_i.$$

By taking the limit as n tends to infinity, we reduce our error to zero and obtain the following.

Formula 3.4.2 (Shell Method I)**.** *Given any function $f(x)$ that is continuous on a closed interval $[a, b]$ and satisfies that $f(x) \geq 0$ for all real numbers x such that $a \leq x \leq b$, the solid of revolution $\mathcal{S}$ obtained by rotating $f(x)$ about the y-axis from $x = a$ to $x = b$ has the following volume.*

$$\text{volume}(\mathcal{S}) = 2\pi \int_a^b x f(x)\, dx$$

Likewise, if $g(y)$ is any function that is continuous on a closed interval $[c, d]$ along the y-axis and satisfies that $g(y) \geq 0$ for all real numbers y such that $c \leq y \leq d$, the solid of revolution $\mathcal{S}$ obtained by rotating $g(y)$ about the x-axis from $y = c$ to $y = d$ has the following volume.

$$\text{volume}(\mathcal{S}) = 2\pi \int_c^d y g(y)\, dy$$

Example 3.4.3. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by the curves $y = -x^3 + 2x^2 - x + 1$, $y = 0$, $x = 0$, and $x = \frac{3}{2}$ about the y-axis.

Solution. Considering that we are revolving a vertically simple region about a vertical axis, let us employ the Shell Method I. By a simple matter of plug-and-chug, we have that

$$\begin{aligned} \text{volume}(\mathcal{S}) &= 2\pi \int_0^{3/2} x(-x^3 + 2x^2 - x + 1)\, dx \\ &= 2\pi \int_0^{3/2} (-x^4 + 2x^3 - x^2 + x)\, dx \\ &= 2\pi \left[-\frac{x^5}{5} + \frac{x^4}{2} - \frac{x^3}{3} + \frac{x^2}{2} \right]_0^{3/2} = \frac{81}{40}\pi. \end{aligned}$$ ◇

Like with the Washer Method, there is an obvious analog to the aforementioned Shell Method for regions bounded by more general curves. Explicitly, for the region bounded by some curves $y = f(x)$, $y = g(x)$, $x = a$, and $x = b$ such that $f(x) \geq g(x)$ for all real numbers x such that $a \leq x \leq b$, the height of a cylindrical shell is the difference of the top and bottom curves.

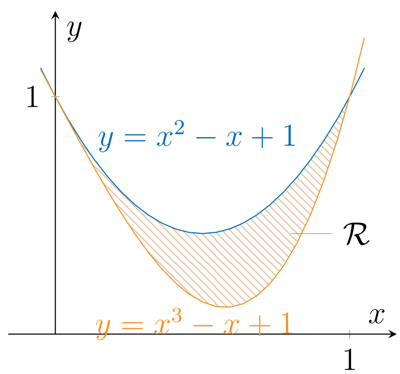
Formula 3.4.4 (Shell Method II)**.** *Given any real number C and any functions $f(x)$ and $g(x)$ that are continuous on a closed interval $[a, b]$ and satisfy that $f(x) \geq g(x)$ for all real numbers x such that $a \leq x \leq b$ and either $a \geq C$ or $b \leq C$, the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by $y = f(x)$, $y = g(x)$, $x = a$, and $x = b$ about the line $y = C$ has the following volume.*

$$\operatorname{volume}(\mathcal{S}) = 2\pi \int_a^b |x - C|[f(x) - g(x)]\,dx$$

Likewise, if C is any real number and $h(y)$ and $k(y)$ are any functions that are continuous on a closed interval $[c, d]$ along the y-axis and satisfies that $h(y) \geq k(y) \geq C$ for all real numbers y such that $c \leq y \leq d$ and either $c \geq C$ or $d \leq C$, the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by $x = h(y)$, $x = k(y)$, $y = c$, and $y = d$ about the line $y = C$ has the following volume.

$$\operatorname{volume}(\mathcal{S}) = 2\pi \int_c^d |y - C|[h(y) - k(y)]\,dy$$

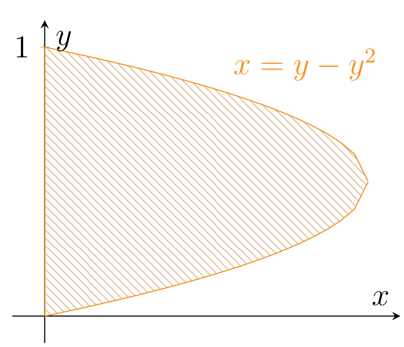
Example 3.4.5. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by the curves $y = x^3 - x + 1$ and $y = x^2 - x + 1$ about the y-axis.



Solution. By the graph provided above, it follows that $f(x) = x^2 - x + 1$ and $g(x) = x^3 - x + 1$ for $x = 0$ and $x = 1$. We can verify these bounds of integration algebraically: indeed, we have that $f(x) = g(x)$ if and only if $x^2 - x + 1 = x^3 - x + 1$ if and only if $x^2 = x^3$ if and only if $x = 0$ or $x = 1$. Consequently, by the Shell Method II, we conclude that the volume of the solid of revolution is

$$\operatorname{volume}(\mathcal{S}) = 2\pi \int_0^1 x(x^2 - x^3)\,dx = 2\pi \int_0^1 (x^3 - x^4)\,dx = 2\pi \left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 = \frac{\pi}{10}. \qquad \diamond$$

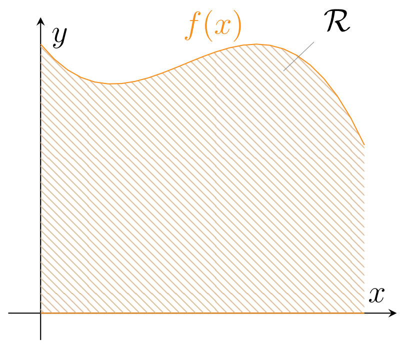
Example 3.4.6. Explain the difficulty in using the Washer Method to compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by the curves $x = y - y^2$ and $x = 0$ about the x-axis; then, use the Shell Method I to compute this volume.



Solution. Our curve $x = y - y^2$ of x is a function of y, and our axis of rotation is horizontal. Consequently, in order to use the Washer Method, we would first need to determine two functions $f(x)$ and $g(x)$ and a real number b such that $f(x) \geq g(x)$ for all real numbers x such that $0 \leq x \leq b$. We could achieve this because the region above is vertically simple, but because it is also horizontally simple, it is indeed easier to use the Shell Method. Explicitly, with the Shell Method, the height of a cylindrical shell is $h(y) = y - y^2$; the radius of a cylindrical shell is y; hence, we find that

$$\text{volume}(\mathcal{S}) = 2\pi \int_0^1 y(y - y^2)\,dy = 2\pi \int_0^1 (y^2 - y^3)\,dy = 2\pi \left[\frac{y^3}{3} - \frac{y^4}{4}\right]_0^1 = \frac{\pi}{6}. \qquad \diamond$$

Example 3.4.7. Compute the volume of the solid of revolution $\mathcal{S}$ obtained by rotating the region bounded by the curves $y = -x^3 + 2x^2 - x + 1$, $y = 0$, $x = 0$, and $x = \frac{3}{2}$ about the line $x = 2$.



Solution. Observe that the height of a cylindrical shell is $-x^3 + 2x^2 - x + 1$, and the radius of a cylindrical shell is $2 - x$ because for each real number x such that $0 \leq x \leq \frac{3}{2}$, we have that $|x - 2| = -(x - 2) = 2 - x$. Consequently, by the Shell Method II, we conclude that

$$\begin{aligned}
\text{volume}(\mathcal{S}) &= 2\pi \int_0^{3/2} (2 - x)(-x^3 + 2x^2 - x + 1)\,dx \\
&= 2\pi \int_0^{3/2} (x^4 - 4x^3 + 5x^2 - 3x + 2)\,dx \\
&= 2\pi \left[\frac{x^5}{5} - x^4 - \frac{5x^3}{3} - \frac{3x^2}{2} + 2x\right]_0^{3/2} = \frac{273}{80}\pi. \qquad \diamond
\end{aligned}$$

3.5 Work

One of the foremost applications of integration is in the applied sciences of physics and chemistry. Explicitly, the definite integral can be applied to compute a physical quantity called **work** that is defined as the amount of energy expended to displace an object a distance d units by a **force** of F units. By **Newton's Second Law of Motion**, if an object of mass m moves along a straight path according to the position function $s(t)$ measured with respect to time t, then we have that

$$F = m\frac{d^2 s}{dt^2} \text{ (force = mass × acceleration).}$$

Customarily, the mass of an object is measured in kilograms (kg) and the position of an object is measured in meters (m), hence the force of an object is measured in **Newtons** ($\mathrm{N} = \mathrm{kg} \cdot \mathrm{m/s^2}$); however, it is also possible to measure force in terms of pounds (lbs). Either way, if we denote by W the work required to displace an object d meters by a constant force of F Newtons, then

$$W = Fd \text{ (work = force} \times \text{distance).}$$

We note that work W is measured in Newton-meters or **joules** (J). Often, we will perform work against the force of gravity; in this case, it will typically be useful to know that the **gravitational acceleration constant** (or more simply, the acceleration of an object due to gravity) is given by

$$g = 9.8 \frac{\mathrm{m}}{\mathrm{s}^2}.$$

Even more, in the context of work against the force of gravity, it is possible to measure the force in terms of pounds (lbs) and the distance in terms of feet (ft); under these conventions, the resulting work W is measured in **foot-pounds** (ft-lbs). Context ought to make it clear which units to use.

Exercise 3.5.1. Compute the work done to move an object 10 meters with a force of 3.5 Newtons.

Solution. By definition, we have that $W = Fd = (3.5)(10) = 35$ J. ◇

Exercise 3.5.2. Compute the work done to lift a 50-kg object a height of 1.5 meters.

Solution. Considering that we are not provided with the force, we must find it. We are performing work against the force of gravity, hence we have that $F = mg = (50)(9.8) = 490$ N. Once we have the force required to lift the object, we may use it to compute $W = Fd = (490)(1.5) = 735$ J. ◇

Exercise 3.5.3. Compute the work done to lift a 20-lb basket of towels a height of 3 feet.

Solution. We are given the force required to lift the object in the form of a weight (20-lb), and we are given a distance in terms of feet. Consequently, our units of work will be measured in ft-lbs; using the formula for work as force times distance, we find that $W = (20)(3) = 60$ ft-lbs. ◇

Exercise 3.5.4. Compute the weight of an upright piano if it requires 9000 ft-lbs to lift it 20 feet.

Solution. We are provided with the work $W = 9000$ ft-lbs and the distance of 20 feet. By solving for F in the formula $W = Fd$, we find that $F = W/d = 9000/20 = 450$ pounds (lbs). ◇

Often, in real world scenarios, the force required to displace an object is not constant but rather depends upon some physical constraint x such as distance along a vertical or horizontal axis. Explicitly, we may view this variable force $F(x)$ as a function of x. Consequently, if we wish to determine the work W required to move an object from $x = a$ to $x = b$, then we may first approximate W by splitting the interval $[a, b]$ into n subintervals $[x_{i-1}, x_i]$ each of length Δx_i for some real numbers $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. Observe that if the number n of subintervals is large, then the length Δx_i of each subinterval is small, so we may assume the force $F(x_i^*)$ is constant on the interval $[x_{i-1}, x_i]$; the work required to move the object a distance of Δx_i is $W_i = F(x_i^*)\,\Delta x_i$. By summing each of these approximations W_i, we obtain a rough estimate of the total work

$$W \approx \sum_{i=1}^{n} W_i = \sum_{i=1}^{n} F(x_i^*)\,\Delta x_i$$

Considering that the force exerted on an object must be a continuous function of x, it follows that we can approximate the work in such a way that our error is reduced to zero as n tends to infinity.

$$W = \lim_{n \to \infty} \sum_{i=1}^{n} F(x_i^*)\, \Delta x_i = \int_a^b F(x)\, dx.$$

Formula 3.5.5 (Work Done by a Variable Force)**.** *Given any continuous function $F(x)$ that measures the force required to displace an object from a point $x = a$ to a point $x = b$ along the x-axis, the work done in moving the object from $x = a$ to $x = b$ is given by the definite integral*

$$W = \int_a^b F(x)\, dx.$$

Exercise 3.5.6. Compute the work required to move an object from the point $x = -1$ to the point $x = 1$ if a force of $F(x) = x^3 + 3x^2$ Newtons acts on the object at a point x meters from the origin.

Solution. By the formula for Work Done by a Variable Force, we have that

$$W = \int_{-1}^{1} (x^3 + 3x^2)\, dx = \left[\frac{x^4}{4} + x^3\right]_{-1}^{1} = \left(\frac{1}{4} + 1\right) - \left(\frac{1}{4} - 1\right) = 2 \text{ J.} \qquad \diamond$$

Exercise 3.5.7. Compute the work required to move an object from the point $x = 0$ to the point $x = \frac{\pi}{3}$ if a force of $F(x) = \tan^2(x)$ pounds acts on the object at a point x feet from the origin.

Solution. By the formula for work done by a variable force, we have that

$$W = \int_0^{\pi/3} \tan^2(x)\, dx = \int_0^{\pi/3} [\sec^2(x) - 1]\, dx = [\tan(x) - x]_0^{\pi/3} = \sqrt{3} - \frac{\pi}{3} \text{ ft-lbs.} \qquad \diamond$$

Exercise 3.5.8. Compute the work required to move an object from the point $x = -1$ to the point $x = 3$ if a force of $F(x) = x^2(x-1)^{99}$ Newtons acts on the object at x meters from the origin.

Solution. By the formula for work done by a variable force, we have that

$$W = \int_{-1}^{3} x^2(x-1)^{99}\, dx.$$

We could approach this integral using integration by parts with $u = x^2$ and $dv = (x-1)^{99}\, dx$, but it is much simpler to notice that if $u = x - 1$, then $du = dx$ and

$$\int_{-1}^{3} x^2(x-1)^{99}\, dx = \int_{-2}^{2} (u+1)^2 u^{99}\, du = \int_{-2}^{2} (u^2 + 2u + 1)u^{99}\, du = \int_{-2}^{2} (u^{101} + 2u^{100} + u^{99})\, du.$$

Crucially, observe that the real polynomial $u^{101} + u^{99}$ is an odd function because each monomial in this polynomial has odd degree, hence the part of the integral involving each of these functions is zero because it is an odd function over a symmetric interval. Consequently, we conclude that

$$W = \int_{-1}^{3} x^2(x-1)^9 9\, dx = \int_{-2}^{2} 2u^{100}\, du = \left[\frac{2u^{101}}{101}\right]_{-2}^{2} = \frac{2^{103}}{101} \text{ J.} \qquad \diamond$$

Chemistry students often think about work in the context of fluids, pressure, and the force required to compress some volume of fluid. By simplifying our view of the system in question, this process can be modelled by the force required to stretch or compress a string. Crucially, **Hooke's Law** states that the force $F(x)$ required to maintain a spring at a constant distance of x units beyond **equilibrium** is equal and opposite to the force exerted by the string against the compression (or stretching); the force the spring exerts to remain in equilibrium is called the **spring constant**, and it is denoted by k. Explicitly, under these conventions, Hooke's Law states that

$$F(x) = kx$$

so that k is measured in Newtons per meter. Consequently, the formula for Work Done by a Variable Force and Hooke's Law together give rise to the following formula for work done on a spring.

Formula 3.5.9 (Work Done on a Spring)**.** *Given a spring of spring constant k Newtons per meter, the work required to move a spring from $x = a$ to $x = b$ meters beyond equilibrium is given by*

$$W = \int_a^b kx\,dx.$$

Bear in mind that if equilibrium is taken as $x = 0$ meters, then if we wish to compress a spring a distance of 1 cm beyond equilibrium, we need to consider the points $x = 0$ and $x = -0.01$ meters. Conversely, to stretch a spring 2 cm beyond equilibrium, we consider the points $x = 0$ and $x = 0.02$.

Exercise 3.5.10. Compute the work required to compress a spring a distance of 1.7 meters beyond equilibrium if the spring constant is known to be 1400 Newtons per meter.

Solution. By the formula for Work Done on a Spring, we have that

$$W = \int_0^{1.7} 1400x\,dx = \left[700x^2\right]_0^{1.7} = 2023 \text{ J.}$$ ◇

Exercise 3.5.11. Compute the work required to stretch a spring from -8 cm beyond equilibrium to 2 cm beyond equilibrium if the spring constant is known to be 100 Newtons per meter.

Solution. By the formula for work done on a spring, we have that

$$W = \int_{-0.08}^{0.02} 100x\,dx = \left[50x^2\right]_{-0.08}^{0.02} = -0.3 \text{ J.}$$

We note that this makes sense intuitively: indeed, it requires negative work to return to the spring to equilibrium, and we are not stretching the spring as far as it was compressed in the first place. ◇

Exercise 3.5.12. Consider a spring of length 7 inches. Compute the spring constant of the spring if it requires a force of 72 ft-lbs to stretch the spring to a length of 16 inches; then, use this to compute the work required to compress the spring 3 inches beyond equilibrium.

Solution. Observe that to stretch the spring to a length of 16 inches, we must stretch the spring 9 inches (or 0.75 feet) beyond equilibrium. By the formula for Work Done on a Spring, we have that

$$72 = \int_0^{0.75} kx\,dx = \left[\frac{kx^2}{2}\right]_0^{0.75} = \frac{9k}{64}.$$

By solving for k, we find that $k = 256$ pounds per foot. Consequently, by the aforementioned formula, we conclude that the work required to compress the spring 3 inches beyond equilibrium is

$$W = \int_0^{-0.25} 256x \, dx = \left[128x^2\right]_0^{-0.25} = 8 \text{ ft-lbs.} \qquad \diamond$$

Exercise 3.5.13. Consider a 10-cm spring. Compute the work required to stretch the spring 5 cm beyond equilibrium if it requires a force of 6000 Newtons to maintain the spring at length 5 cm.

Solution. By Hooke's Law, the force required to maintain the spring at a length of x meters beyond equilibrium is given by kx. We are given that if we compress the spring to a length of 5 cm, it requires a force of 6000 Newtons to maintain the spring at this length, hence we have that

$$0.05k = 6000 \text{ so that } k = 120000 \text{ Newtons per meter.}$$

We next stretch the spring 5cm beyond equilibrium. By the formula for work done on a spring,

$$W = \int_0^{0.05} 120000x \, dx = \left[60000x^2\right]_0^{0.05} = 150 \text{ J.} \qquad \diamond$$

Even more, it is typical in real world applications (such as chemical engineering and civil engineering) to consider work done against the force of gravity. General physical principles dictate that the **mass** of an object is equal to the product of the density and volume of the object. Explicitly, if the m is the mass, ρ is the density, and v is the volume of an object, then we have that

$$m = \rho v \text{ (mass = density} \times \text{volume).}$$

We will typically use kilogram (kg) units to measure mass; volume will be measured in cubic meters (m^3); and density is therefore given by kilograms per cubic meter (kg/m^3). Often, the volume of an object is not known but can be measured as the product of its cross-sectional area and its thickness. Explicitly, if the cross-sectional area of an object is some function $a(x)$ of the distance x on some axis and the thickness of a cross section of the object is Δx, then the volume of the object is

$$v(x) = a(x) \, \Delta x \text{ (volume = cross-sectional area} \times \text{thickness).}$$

Combined, these observations are typically used to perform the required **force analysis** to determine the force function $F(x)$ that needs to be integrated. Let us illustrate with some examples.

Example 3.5.14. Consider a 50-foot hanging chain of **weight-density** 10 lb/ft. We will compute the work done against the force of gravity to wind up the chain. Be careful to notice that pounds measures the force, so our final work should be in foot-pounds. Given that the weight-density of the chain is 10 lb/ft, the force exerted to lift a point x_i^* on the chain a distance of Δx_i feet is given by $F(x_i^*) = 10 \, \Delta x_i$ pounds whenever Δx_i is small. Overall, we must lift the chain a distance of 50 feet, so the total distance travelled by a point x_i^* on the chain is $50 - x_i^*$ feet. Consequently, the work required to lift a point x_i^* on the chain a distance of $50 - x_i^*$ feet is given by $10(50 - x_i^*) \, \Delta x_i$. We conclude that the work required to wind the entire 50 feet of chain is given by the definite integral

$$W = \lim_{n \to \infty} \sum_{i=1}^n 10(50 - x_i^*) \, \Delta x_i = \int_0^{50} 10(50 - x) \, dx = 10 \left[50x - \frac{x^2}{2}\right]_0^{50} = 12500 \text{ ft-lbs.}$$

Example 3.5.15. Consider a 15-meter hanging rope of uniform **linear density** that weighs 105 kg and has a 20-kg weight attached at the bottom. We will compute the work done against the force of gravity to lift the weight from the ground to a height of 10 meters. We will adopt an approach in this example that is substantially different from our technique in Example 3.5.14; the reader can determine their preferred method. Considering that the rope has length 15 meters, uniform linear density, and weighs 105 kg, we conclude that the linear density of the rope is $7 = 105/15$ kilograms per meter. Consider a point x along the length of the rope such that $x = 0$ is the bottom of the rope and $x = 15$ is the top of the rope. We must pull on the rope to raise the weight off the ground; the length of rope remaining after pulling the rope a distance of x meters is $15 - x$ meters, hence the mass of rope remaining at that point is $7(15 - x)$ kilograms; and the total mass of the rope and the weight combined is $125 - 7x = 7(15 - x) + 20$ kilograms. By taking into account the acceleration due to gravity, the force exerted on a point on the rope a distance of x meters from the ground is $F(x) = 1225 - 68.6x$ Newtons. We conclude by the Work Done by a Variable Force formula that

$$W = \int_0^{10} (1225 - 68.6x)\, dx = 8820 \text{ J}.$$

Example 3.5.16. Consider an upside-down right-circular conical tank that is 60-meters tall with a radius of 10 meters. We wish to pump heavy crude oil out of the tank through a spout located 3 meters above the tank. We will assume that the density of heavy crude oil is 900 kg/m^3. We begin with a force analysis to compute the work to empty the tank. Consider a point x along the height of the tank such that $x = 0$ is the bottom of the tank. Each layer of heavy crude oil at a height of x_i^* meters must travel a total of $60 - x_i^*$ meters to get to the top of the tank; then, the layer must travel a distance of 3 meters through the spout. Consequently, a layer of crude oil at height x_i^* must travel a distance of $63 - x_i^*$ meters. Certainly, the work we are performing is against the force of gravity, so in order to lift a layer of crude oil at a height of x_i^* meters, we must exert a force of

$$F(x_i^*) = 9.8m(x_i^*)$$

for the mass $m(x_i^*)$ of heavy crude oil. Considering that mass is density times volume, we have that

$$m(x_i^*) = 900v(x_i^*)$$

for the volume $v(x_i^*)$ of heavy crude oil. Crucially, we have used here that the density of the oil is 900 kg/m^3. Considering the shape of the tank, the cross-sectional area $a(x_i^*)$ of a layer of oil at a height of x_i^* meters depends upon x_i^*: indeed, each cross-section is a circle of radius $r(x_i^*)$ because the tank is an upside-down right-circular cone, hence the cross-sectional area of a layer of oil is

$$a(x_i^*) = \pi[r(x_i^*)]^2.$$

We use similar triangles to deduce that the radius $r(x_i^*)$ of the cross section at height x_i^* is given by

$$r(x_i^*) = \frac{x_i^*}{6}.$$

Consequently, if the thickness of each layer of oil is Δx_i, then our force analysis yields that

$$W_i = (9.8)(900)\left(\frac{\pi}{36}(x_i^*)^2\right) \Delta x_i = 245\pi(x_i^*)^2\, \Delta x_i.$$

We conclude by the Work Done by a Variable Force formula that the work to empty the tank is

$$W = \int_0^{10} 245\pi x^2(63 - x)\,dx = \pi \int_0^{10} (21735x^2 - 245x)\,dx = \pi \left[7245x^3 - \frac{245x^2}{2}\right]_0^{10} = 4532500\pi \text{ J}.$$

Example 3.5.17. Consider a spherical tank with a radius of 3 meters. Given that the density of maple syrup is 1043 kg/m^3, let us compute the work done against the force of gravity to pump orange soda out of the tank through a spout located 5 meters above the tank. Consider a point x along the height of the tank such that $x = 0$ is perpendicular with the center of the tank so that $x = -3$ is located at the bottom of the tank and $x = 3$ is located at the top of the tank. Each layer of orange soda at a height of x_i^* meters must travel a total distance of $3 - x_i^*$ meters to reach the top of the tank; then, the layer must travel a distance of 5 meters through the spout. Consequently, a layer of orange soda at height x_i^* must travel a distance of $8 - x_i^*$ meters. We are performing work against the force of gravity, so our force function is given by

$$F(x_i^*) = 9.8m(x_i^*)$$

for the mass $m(x_i^*)$ of orange soda at height x_i^*; the mass of orange soda at height x_i^* is given by

$$m(x_i^*) = 1043v(x_i^*)$$

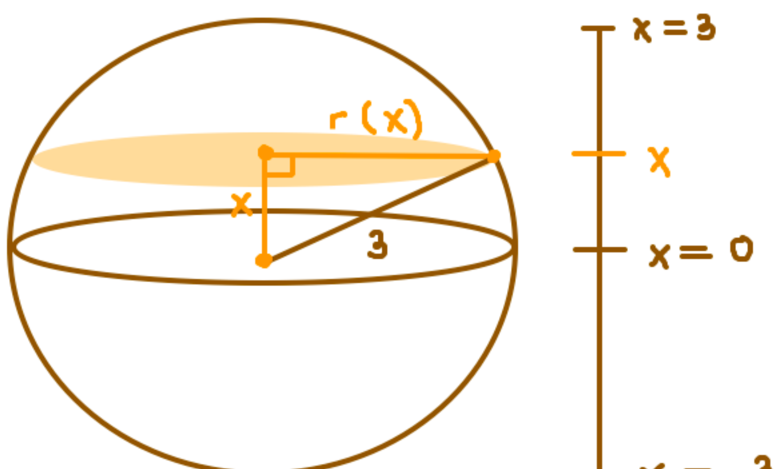
for the volume $v(x_i^*)$ of orange soda at height x_i^*; and this volume of orange soda is given by

$$v(x_i^*) = a(x_i^*)\,\Delta x_i$$

for the cross-sectional area of a layer of orange soda at height x_i^* of thickness Δx_i. Each cross section of the spherical tank is a circle of radius $r(x_i^*)$, hence the cross-sectional area at height x_i^* is

$$a(x_i^*) = \pi[r(x_i^*)]^2.$$

We use the Pythagorean Theorem to determine the radius $r(x_i^*)$ of a layer of orange soda at height x_i^*. Extend a line segment from the center of the spherical tank to the boundary of the tank: explicitly, by drawing a line of length x from the center of the tank to our layer at height x, we form a right triangle with hypotenuse of length 3, height of length x, and base of length $r(x)$.



Consequently, by the Pythagorean Theorem, we have that $x^2 + [r(x)]^2 = 3^2$ so that $[r(x)]^2 = 9 - x^2$. Putting this all together, we find the work function that we must subsequently integrate.

$$W_i = (9.8)(1043)[\pi(9 - (x_i^*)^2)]\,\Delta x_i = 10221.4\pi(9 - (x_i^*)^2)\,\Delta x$$

We conclude by the Work Done by a Variable Force formula that the work to empty the tank is

$$W = \int_{-3}^{3} 10221.4\pi(9 - x^2)\,dx = 10221.4\pi\left[9x - \frac{x^3}{3}\right]_{-3}^{3} = 367970.4\pi.$$

Chapter 4

Parametrization and Polar Coordinates

4.1 Parametric Equations

Consider a dust particle floating through space. We can track the location of the particle in the xy-plane at time t by recording its position $x(t)$ in the east-west direction and $y(t)$ in the north-south direction. Combined, these two position functions uniquely determine the location of the particle in the xy-plane. We may refer to the variable t as the (time) **parameter** and to the collection of all points $\mathcal{C} = \{(x(t), y(t)) \mid t \geq 0\}$ as the **parametric curve** defined by $x = x(t)$ and $y = y(t)$. Graphically, the parametric curve $\mathcal{C}$ is the path in the xy-plane travelled by the dust particle as it moves through space. Each of the functions $x(t)$ and $y(t)$ is called a **parametric equation**, and we refer to the ordered pair $(x, y) = (x(t), y(t))$ as a **parametrization** of x and y in terms of t.

Example 4.1.1. One of the most formative experiences from my childhood was to watch a samara (better known as a "helicopter seed," "whirigig," or "spinning Jenny") fall from a tree. I always found it truly fascinating how they drift down in a manner such that the tip of the winged seed spins in a circle. Recall that the equation of a circle of radius $r > 0$ centered at (h, k) in the xy-plane is

$$(x - h)^2 + (y - k)^2 = r^2.$$

By setting $x(t) = r\cos(t) + h$ and $y(t) = r\sin(t) + k$, the Pythagorean Identity yields that

$$[x(t) - h]^2 + [y(t) - k]^2 = [r\cos(t)]^2 + [r\sin(t)]^2 = r^2\cos^2(t) + r^2\sin^2(t) = r^2[\cos^2(t) + \sin^2(t)] = r^2$$

is a parametrization of the circle of radius $r > 0$ centered at (h, k); the parametric equations

$$\begin{cases} x(t) = r\cos(t) - h \\ y(t) = r\sin(t) - k \end{cases}$$

in terms of our parameter t together completely determine this circle. Back to the example of the whirigig, if we consider the center of the seed as the origin $(0, 0)$ and the distance to the tip of the whirigig is $r > 0$ units, then the position of a point (x, y) on the tip of the "helicopter seed" at time t can be described by $(x, y) = (x(t), y(t)) = (r\cos(t), r\sin(t))$ for some real number $r > 0$.

Example 4.1.2. Given a real number $r > 0$, the parametric equations

$$\begin{cases} x(t) = r\sin(t) \\ y(t) = r\cos(t) \end{cases}$$

with $0 \leq t \leq 2\pi$ give rise to a circle of radius r centered at the origin $(0,0)$ in the xy-plane.

$$(x-0)^2 + (y-0)^2 = [x(t)]^2 + [y(t)]^2 = r^2 \sin^2(t) + r^2 \cos^2(t) = r^2[\sin^2(t) + \cos^2(t)] = r^2$$

Observe that as the parameter t increases from the point $t = 0$ to $t = \frac{\pi}{2}$, the x-coordinate of the circle increases from $x(0) = 0$ to $x\left(\frac{\pi}{2}\right) = r$ and the y-coordinate of the circle decreases from $y(0) = r$ to $y\left(\frac{\pi}{2}\right) = 0$. Consequently, as t increases from $t = 0$ to $t = 2\pi$, we move along the circumference of the circle in a clockwise fashion. Compare this with the parametrization of the very same circle

$$\begin{cases} x(t) = r\cos(t) \\ y(t) = r\sin(t) \end{cases}$$

in which the motion from the point $t = 0$ to $t = 2\pi$ is counterclockwise around the circle.

Example 4.1.3. We can parametrize the parabola $y = x^2$ on the domain $-1 \leq x \leq 1$ in two ways. By letting s denote our first parameter so that $-1 \leq s \leq 1$, we have the following parametrization.

$$\begin{cases} x(s) = s \\ y(s) = s^2 \end{cases}$$

By letting t denote our second parameter so that $0 \leq t \leq \pi$, we have the following parametrization.

$$\begin{cases} x(t) = \cos(t) \\ y(t) = \cos^2(t) \end{cases}$$

Observe that the parametrization in s traces the parabola $y = x^2$ from the point $(-1,1)$ to the point $(1,1)$. Compare this with the parametrization in t that traces the curve from $(1,1)$ to $(-1,1)$.

We have discussed thus far the process of determining the graph of a **plane curve** $f(x,y) = 0$ from a parametrization $x = x(t)$ and $y = y(t)$ of the variables x and y as functions of a parameter t. Conversely, if we are given a pair of parametric equations $x = x(t)$ and $y = y(t)$, then we can sometimes determine the underlying plane curve by a process called **elimination**. Explicitly, the idea of elimination is to solve (if possible) one of the equations $x = x(t)$ or $y = y(t)$ for the parameter t and subsequently plug that equation in to the parametric equation for the other variable.

Example 4.1.4. Consider the parametric equations $x = t^2 - 2t$ and $y = t + 1$. By solving for the parameter t in the equation involving y, we find that $t = y-1$ so that $x = t^2 - 2t = (y-1)^2 - 2(y-1)$. Consequently, the plane curve parametrized by these equations is a parabola with horizontal axis of symmetry (because x is a quadratic function of y). We can be even more specific about the curve if we find the vertex form of this parabola: observe that $x = y^2 - 2y + 1 - 2y + 2 = y^2 - 4y + 3 = (y-2)^2 - 1$, hence this is a parabola in y that opens toward the right and has vertex $(2,-1)$.

Example 4.1.5. Consider the parametric equations $x = \cos(t) - 2$ and $y = \sin(t) + 3$. Observe that $x + 2 = \cos(t)$ and $y - 3 = \sin(t)$, hence by the Pythagorean Identity, we have that

$$(x+2)^2 + (y-3)^2 = \cos^2(t) + \sin^2(t) = 1.$$

Consequently, the plane curve described by these equations is a circle of radius 1 centered at $(-2,3)$.

Example 4.1.6. Consider the parametric equations $x = 2\sin(3t)$ and $y = 3\cos(3t)$. Observe that $\frac{x}{2} = \cos(3t)$ and $\frac{y}{3} = \sin(3t)$, hence by the Pythagorean Identity, we have that

$$\left(\frac{x}{2}\right)^2 + \left(\frac{y}{3}\right)^2 = \sin^2(t) + \cos^2(t) = 1.$$

Consequently, the plane curve parametrized by these equations is an ellipse centered at $(0, 0)$.

Example 4.1.7. Consider the parametric equations $x = 3t - 5$ and $y = 2t + 1$. By solving the equation in x for the parameter t, we find that $t = \frac{x}{3} + \frac{5}{3}$ so that $y = 2t + 1 = \frac{2}{3}x + \frac{13}{3}$. Consequently, the plane curve parametrized by these equations is a line of slope $\frac{2}{3}$ with y-intercept $\left(0, \frac{13}{3}\right)$.

Example 4.1.8. Consider the parametric equations $x = t^3$ and $y = t^2$. Observe that if we square x and cube y, then we have that $x^2 = t^6 = y^3$, from which it follows that the plane curve parametrized by these equations is the cuspidal cubic $x^2 - y^3 = 0$. We remark that this is an important curve in algebraic geometry and commutative algebra because of the vertical tangents at $t = 0$.

Conversely, we might wish to parametrize a plane curve in a variable t over some domain.

Example 4.1.9. Consider the plane curve defined by the equation $(x - 1)^2 + (y + 2)^2 = 9$. We recognize this as a circle of radius 3 centered at the point $(1, -2)$ in the xy-plane. Observe that if

$$\begin{cases} x = 3\sin(t) + 1 \text{ and} \\ y = 3\cos(t) - 2, \end{cases}$$

then the Pythagorean Identity $\sin^2(t) + \cos^2(t) = 1$ yields that

$$9 = 9\sin^2(t) + 9\cos^2(t) = [3\sin^2(t) + 1 - 1]^2 + [3\cos^2(t) - 2 + 2]^2 = (x - 1)^2 + (y + 2)^2,$$

hence the assignments of $x = 3\sin(t) + 1$ and $y = 3\cos(t) - 2$ provide a valid parametrization.

Example 4.1.10. Consider the plane curve defined by the equation $\dfrac{x^2}{25} + \dfrac{y^2}{4} = 1$. Observe that if

$$\begin{cases} x = 5\cos(t) \text{ and} \\ y = 2\sin(t), \end{cases}$$

then the Pythagorean Identity $\sin^2(t) + \cos^2(t) = 1$ yields that

$$1 = \cos^2(t) + \sin^2(t) = \left(\frac{x}{5}\right)^2 + \left(\frac{y}{2}\right)^2 = \frac{x^2}{25} + \frac{y^2}{4},$$

hence the assignments of $x = 5\cos(t)$ and $y = 2\sin(t)$ provide a valid parametrization.

Example 4.1.11. Consider the plane curve defined by the equation $x^2 - y^2 = -1$. We may rewrite this as a sum of squares $x^2 + 1 = y^2$. By the Pythagorean Identity $\tan^2(t) + 1 = \sec^2(t)$, the assignments of $x = \tan(t)$ and $y = \sec(t)$ provide a valid parametrization.

Example 4.1.12. Given any plane curve that can be defined by some function $y = f(x)$, we can obtain a very simple parametrization by setting $x = t$ and noticing that $y = f(x) = f(t)$.

Even if we are not able to untangle the parametric equations $x = x(t)$ and $y = y(t)$ to find a plane curve $f(x, y) = 0$, under some mild assumptions about the relationship between x and y, we are still able to determine important characteristics of the parametric curve by elementary calculus.

Formula 4.1.13 (Slope of the Tangent Line of a Parametric Curve)**.** *Consider any parametric curve defined by the parametric equations $x = x(t)$ and $y = y(t)$. Given that $y = f(x)$ is a differentiable function of x and $x = x(t)$ is a differentiable function of the parameter t, then we have that*

$$y'(t) = \frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = f'(x)x'(t).$$

Consequently, if $x'(t)$ is nonzero, then the slope of the tangent line to the parametric curve is

$$f'(x) = \frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{y'(t)}{x'(t)}.$$

We note that the proof of the above fact follows directly from the Chain Rule applied to $y = f(x(t))$.

Example 4.1.14. Consider the parametric curve defined by the parametric equations $x(t) = 3t - 5$ and $y(t) = 2t + 3$. By the formula for the Slope of the Tangent Line of a Parametric Curve,

$$\frac{dy}{dx} = \frac{y'(t)}{x'(t)} = \frac{2}{3}$$

implies that the slope of the tangent line to this parametric curve is $\frac{2}{3}$ for all real numbers t.

Example 4.1.15. Consider the parametric curve defined by the parametric equations $x(t) = t^2 + 2t - 1$ and $y(t) = t - 4$. By the formula for the slope of the tangent line of a parametric curve,

$$\frac{dy}{dx} = \frac{y'(t)}{x'(t)} = \frac{1}{2t+2}$$

implies that the slope of the tangent line to this parametric curve is $\frac{1}{2t+2}$ for all real numbers t such that $2t + 2$ is nonzero. Observe that at $t = -1$, we have that $2t + 2 = 0$, so the curve possesses a vertical tangent at $t = -1$ because the denominator of $\frac{dy}{dx}$ is zero but the numerator is nonzero.

Example 4.1.16. Consider the parametric curve defined by the parametric equations $x(t) = \sin(t)$ and $y(t) = \cos(t)$. By the formula for the slope of the tangent line of a parametric curve,

$$\frac{dy}{dx} = \frac{y'(t)}{x'(t)} = \frac{-\sin(t)}{\cos(t)} = -\tan(t)$$

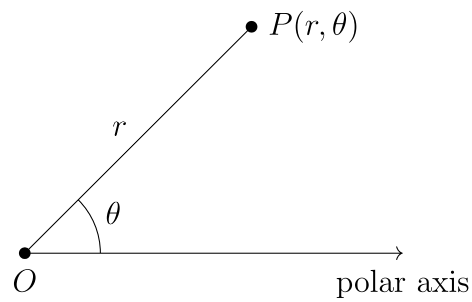
implies that the slope of the tangent line to this parametric curve is $-\tan(t)$ for all real numbers t such that $\cos(t)$ is nonzero. Observe that if $\sin(t) = 0$, then the parametric curve has horizontal tangent line because the numerator of $\frac{dy}{dx}$ is zero and the denominator is nonzero. Consequently, the parametric curve has horizontal tangent lines for all real numbers $t = k\pi$ for some integer k. Conversely, if $\cos(t) = 0$, then the denominator of $\frac{dy}{dx}$ is zero and the numerator is nonzero, hence the parametric curve has vertical tangent lines for all real number $t = k\pi + \frac{\pi}{2}$ for some integer k.

4.2 Polar Coordinates

Cartesian coordinates are constructed from lines and rectangles; alternatively, Cartesian coordinates are called “rectangular coordinates.” We turn our attention next to a coordinate system that is constructed from arcs and circles. Explicitly, the **polar coordinate system** consists of (a.) a **pole** O centered at the origin $(0, 0)$ and

(b.) a **polar axis**, i.e., a ray extending indefinitely from the pole O in the positive x-direction.

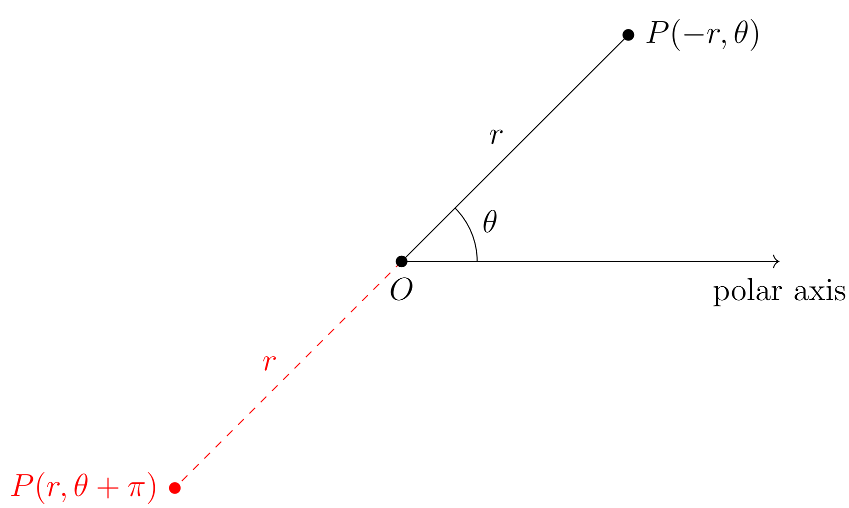
Given any point P in the Cartesian plane, we denote by r the distance from the pole O to the point P, and we denote by θ the angle subtended by the arc from the polar axis to the line segment OP. Consequently, we may represent the point $P = P(r, \theta)$ as an ordered pair in **polar coordinates** by specifying its **radial coordinate** r and its **angular coordinate** θ. Below is a diagram.



Conventionally, the following identifications are made for points in polar coordinates.

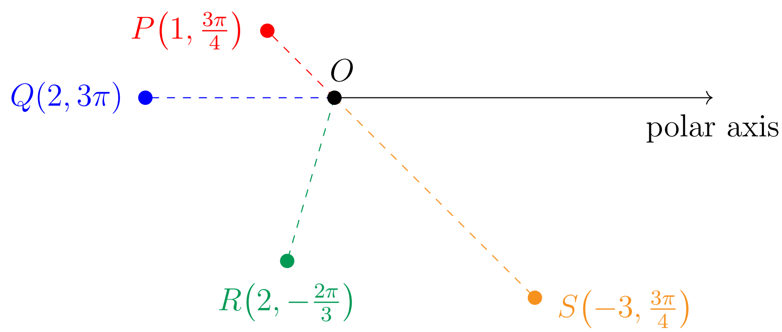
(a.) We have that $P(-r, \theta) = P(r, \theta + \pi)$. Put another way, the point $P(-r, \theta)$ in polar coordinates a distance of $-r$ units from the origin at an angle of θ radians from the polar axis is the same as the point $P(r, \theta + \pi)$ in polar coordinates a distance of r units from the origin at an angle of $\theta + \pi$ radians from the polar axis. Consequently, we may assume that r is non-negative.

(b.) We have that $P(r, \theta) = P(r, \theta + 2k\pi)$ for any integer k. Put another way, the point in polar coordinates a distance of r units from the origin at an angle of θ radians from the polar axis is the same as the point in polar coordinates a distance of r units from the origin at an angle of $\theta + 2\pi k$ radians from the polar axis for any integer k. Consequently, we may assume our angular coordinate θ satisfies that $0 \leq \theta < 2\pi$ (i.e., it is positive and does not exceed 360°).



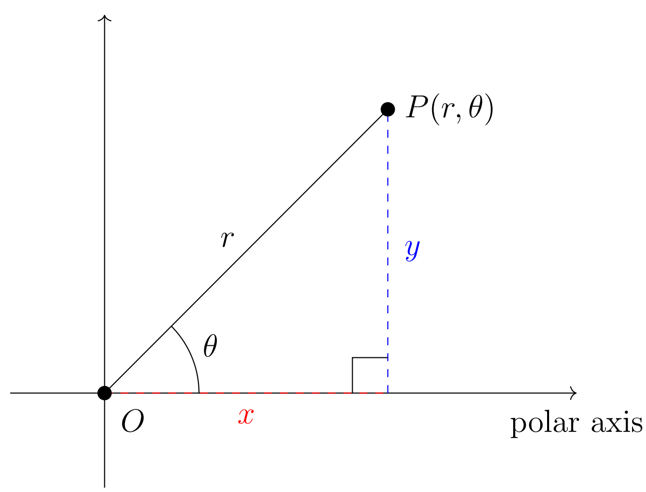
We illustrate in our next example the difference between plotting points in Cartesian coordinates and plotting points in polar coordinates. Explicitly, in order to plot the point $P(r, \theta)$ using polar coordinates, one must first determine the distance r from the origin; then, one can plot the point $P(r, \theta)$ by rotating this distance an angle of θ radians from the polar axis. Care should be taken if $r < 0$ because in this case, by the aforementioned convention, we use the angle $\theta + \pi$, instead.

Example 4.2.1. Consider the following points in polar coordinates.



Explicitly, to plot the point Q, we use the convention that $(2, 3\pi) = (2, 3\pi - 2\pi) = (2, \pi)$; to plot the point R, we use the convention that $\left(2, -\frac{2\pi}{3}\right) = \left(2, -\frac{2\pi}{3} + 2\pi\right) = \left(2, \frac{4\pi}{3}\right)$; and to plot the point S, we use the convention that $\left(-3, \frac{3\pi}{4}\right) = \left(3, \frac{3\pi}{4} + \pi\right) = \left(3, \frac{7\pi}{4}\right)$.

Crucially, the following illuminates the relationship between Cartesian and polar coordinates.



$$x = r\cos(\theta) \qquad\qquad r = \sqrt{x^2 + y^2}$$
$$y = r\sin(\theta) \qquad\qquad \tan(\theta) = \frac{y}{x}$$

Explicitly, if we view the radial coordinate r of the point $P(r, \theta)$ in polar coordinates as the hypotenuse of a right triangle with side of length x adjacent to θ and side of length y opposite to θ, then it follows that $\cos(\theta) = \frac{x}{r}$ so that $x = r\cos(\theta)$ and $\sin(\theta) = \frac{y}{r}$ so that $y = r\sin(\theta)$. Even more, by the Pythagorean Theorem, we have that $x^2 + y^2 = r^2$ so that $r = \sqrt{x^2 + y^2}$ (because we may assume that the radial coordinate r is non-negative) and $\tan(\theta) = \frac{y}{x}$ so long as x is nonzero.

Formula 4.2.2 (Polar to Cartesian Coordinates Conversion)**.** *Given any point (r, θ) in polar coordinates, the corresponding point (x, y) in Cartesian coordinates is given by*

$$x = r\cos(\theta) \text{ and } y = r\sin(\theta).$$

Formula 4.2.3 (Cartesian to Polar Coordinates Conversion)**.** *Given any point (x, y) in Cartesian coordinates such that x is nonzero, the corresponding point (r, θ) in polar coordinates is given by*

$$r = \sqrt{x^2 + y^2} \text{ and } \tan(\theta) = \frac{y}{x}.$$

Exercise 4.2.4. Convert the following points in polar coordinates to Cartesian coordinates.

(a.) $P\big(2, \frac{\pi}{3}\big)$ (b.) $Q\big(3, \frac{\pi}{2}\big)$ (c.) $R\big(2\sqrt{2}, \frac{3\pi}{4}\big)$ (d.) $S(4, 3\pi)$

Solution. (a.) We have that $r = 2$ and $\theta = \frac{\pi}{3}$, hence the formula for Polar to Cartesian Coordinates Conversion yields that $x = r\cos(\theta) = (2)\big(\frac{1}{2}\big) = 1$ and $y = r\sin(\theta) = (2)\Big(\frac{\sqrt{3}}{2}\Big) = \sqrt{3}$.

(b.) We have that $r = 3$ and $\theta = \frac{\pi}{2}$, hence the formula for converting from polar to Cartesian coordinates yields that $x = r\cos(\theta) = (3)(0) = 0$ and $y = r\sin(\theta) = (3)(1) = 3$.

(c.) We have that $r = 2\sqrt{2}$ and $\theta = \frac{3\pi}{4}$, hence the formula for converting from polar to Cartesian coordinates yields that $x = r\cos(\theta) = \big(2\sqrt{2}\big)\Big(-\frac{\sqrt{2}}{2}\Big) = -2$ and $y = r\sin(\theta) = \big(2\sqrt{2}\big)\Big(\frac{\sqrt{2}}{2}\Big) = 2$.

(d.) We have that $r = 4$ and $\theta = 3\pi$, hence the formula for converting from polar to Cartesian coordinates yields that $x = r\cos(\theta) = (4)(-1) = -4$ and $y = r\sin(\theta) = (4)(0) = 0$. $\diamond$

Exercise 4.2.5. Convert the following points in Cartesian coordinates to polar coordinates.

(a.) $P(1, 1)$ (b.) $Q\big(2\sqrt{3}, -2\big)$ (c.) $R(-2, 3)$ (d.) $S(0, -1)$ (e.) $T(-2, 0)$

Solution. (a.) We have that $x = 1$ and $y = 1$, hence the formula for Cartesian to Polar Coordinates Conversion yields that $r = \sqrt{x^2 + y^2} = \sqrt{2}$ and $\tan(\theta) = \frac{y}{x} = 1$. Considering that (x, y) lies in Quadrant I and $\tan(\theta) = 1$, we conclude that $\theta = \frac{\pi}{4}$.

(b.) We have that $x = 2\sqrt{3}$ and $y = -2$, hence the formula for converting from Cartesian to polar coordinates yields that $r = \sqrt{x^2 + y^2} = \sqrt{12 + 4} = 4$ and $\tan(\theta) = \frac{y}{x} = -\frac{\sqrt{3}}{3}$. Considering that (x, y) lies in Quadrant IV and $\tan(\theta) = -\frac{\sqrt{3}}{3}$, we conclude that $\theta = \frac{11\pi}{6}$.

(c.) We have that $x = -2$ and $y = 3$, hence the formula for converting from Cartesian to polar coordinates yields that $r = \sqrt{x^2 + y^2} = \sqrt{13}$ and $\tan(\theta) = \frac{y}{x} = -\frac{3}{2}$. Considering that (x, y) lies in Quadrant II and $\tan(\theta) = -\frac{3}{2}$, we conclude that $\theta = \arctan\big(-\frac{3}{2}\big) + \pi$.

(d.) We have that $x = 0$ and $y = -1$, hence we cannot use the formula for converting from Cartesian to polar coordinates because x is zero; however, observe that $(0, -1)$ is the point on the unit circle that lies at an angle $\theta = \frac{3\pi}{2}$, hence in polar coordinates, we have that $S(0, -1) = \big(1, \frac{3\pi}{2}\big)$.

(e.) We have that $x = -2$ and $y = 0$, hence the formula for converting from Cartesian to polar coordinates yields that $r = \sqrt{x^2 + y^2} = \sqrt{2}$ and $\tan(\theta) = \frac{y}{x} = 0$ so that $\theta = \pi$ because $x < 0$. $\diamond$

Often, in polar coordinates, we consider functions $r = f(\theta)$ of the radial coordinate r in terms of the angular coordinate θ. By using the formula for converting from Cartesian to polar coordinates, we can occasionally derive an equation in Cartesian coordinates for a given **polar curve** $r = f(\theta)$.

Example 4.2.6. Consider the polar curve $r = 2$. Observe that $x^2 + y^2 = r^2 = 4$, hence the polar curve $r = 2$ is a circle of radius 2 centered at the origin in polar coordinates.

Example 4.2.7. Consider the polar curve $\theta = \frac{\pi}{4}$. Observe that $y = \tan(\theta)x = \tan\big(\frac{\pi}{4}\big)x = x$, hence $\theta = \frac{\pi}{4}$ is the line through the origin in polar coordinates that forms a $45°$ angle with the polar axis.

Consider a function $r = f(\theta)$ of the radial coordinate r in terms of the angular coordinate θ in polar coordinates. By the formula for Polar to Cartesian Coordinates Conversion, we have that

$$\begin{cases} x(\theta) = r\cos(\theta) = f(\theta)\cos(\theta) \text{ and} \\ y(\theta) = r\sin(\theta) = f(\theta)\sin(\theta) \end{cases}$$

is a parametrization of x and y in terms of the parameter θ. By the formula for the Slope of the Tangent Line of a Parametric Curve, the slope of the tangent line to the polar curve $r = f(\theta)$ is

$$\frac{dy}{dx} = \frac{\frac{dy}{d\theta}}{\frac{dx}{d\theta}} = \frac{y'(\theta)}{x'(\theta)} = \frac{f'(\theta)\cos(\theta) - f(\theta)\sin(\theta)}{f'(\theta)\sin(\theta) + f(\theta)\cos(\theta)}$$

provided that $x'(\theta)$ is nonzero. Like before, the polar curve $r = f(\theta)$ has horizontal tangents when $y'(\theta) = 0$ and $x'(\theta)$ is nonzero and vertical tangents when $x'(\theta) = 0$ and $y'(\theta)$ is nonzero.

Example 4.2.8. Consider the polar curve $r = 1 + \sin(\theta)$. Observe that

$$\begin{aligned} x(\theta) &= r\cos(\theta) = \cos(\theta) + \sin(\theta)\cos(\theta) \text{ and} \\ y(\theta) &= r\sin(\theta) = \sin(\theta) + \sin^2(\theta) \end{aligned}$$

together yield that the slope of the polar curve in question is

$$\frac{dy}{dx} = \frac{y'(\theta)}{x'(\theta)} = \frac{\cos(\theta) + 2\sin(\theta)\cos(\theta)}{-\sin(\theta) - \sin^2(\theta) + \cos^2(\theta)} = \frac{\cos(\theta) + 2\sin(\theta)\cos(\theta)}{1 - \sin(\theta) - 2\sin^2(\theta)}$$

by the Pythagorean Identity $\cos^2(\theta) = 1 - \sin^2(\theta)$. Let us find all values of θ in $[0, 2\pi]$ such that the tangent line to the given polar curve is vertical or horizontal. We achieve this by determining the values of θ for which one (but not both) of the functions in the denominator or numerator of $\frac{dy}{dx}$ is zero. By setting $u = \sin(\theta)$, the denominator of $\frac{dy}{dx}$ is zero if and only if $1 - u - 2u^2 = 0$ if and only if $2u^2 + u - 1 = 0$ if and only if $(2u - 1)(u + 1) = 0$ if and only if $u = \frac{1}{2}$ or $u = -1$. Consequently, the values of θ in $[0, 2\pi]$ for which $x'(\theta) = 0$ are $\theta = \frac{\pi}{6}, \frac{5\pi}{6}$, and $\frac{3\pi}{2}$. Likewise, we have that $y'(\theta) = 0$ if and only if $\cos(\theta) + 2\sin(\theta)\cos(\theta) = 0$ if and only if $\cos(\theta)(1 + \sin(\theta)) = 0$ if and only if $\cos(\theta) = 0$ or $1 + 2\sin(\theta) = 0$ if and only if $\cos(\theta) = 0$ or $\sin(\theta) = -\frac{1}{2}$ if and only if $\theta = \frac{\pi}{2}, \frac{7\pi}{6}, \frac{3\pi}{2}$, or $\frac{11\pi}{6}$. Excluding the case that $y'(\theta) = x'(\theta) = 0$, we conclude that $r = 1 + \sin(\theta)$ has vertical tangents when $\theta = \frac{\pi}{2}, \frac{7\pi}{6}$, and $\frac{11\pi}{6}$ and horizontal tangents when $\theta = \frac{\pi}{6}$ and $\theta = \frac{5\pi}{6}$.

Exercise 4.2.9. Compute the slope of the tangent line to any point on the polar curve $r = 2\sin(\theta)$ with $0 \leq \theta \leq 2\pi$; then, use this to determine all points with vertical or horizontal tangents.

Example 4.2.10. Compute the slope of the tangent line to any point on the polar curve $r = \cos(2\theta)$ with $0 \leq \theta \leq 2\pi$; then, use this to determine all points with vertical or horizontal tangents.

Solution. We must first realize the polar curve as a parametric curve in the parameter θ. We achieve this via the Polar Coordinates to Cartesian Coordinates Conversion

$$\begin{aligned} x(\theta) &= r\cos(\theta) = \cos(\theta)\cos(2\theta) \text{ and} \\ y(\theta) &= r\sin(\theta) = \sin(\theta)\cos(2\theta). \end{aligned}$$

By the formula for the Slope of the Tangent Line of a Parametric Curve, we have that

$$\frac{dy}{dx} = \frac{\frac{dy}{d\theta}}{\frac{dx}{d\theta}} = \frac{y'(\theta)}{x'(\theta)} = \frac{-\sin(\theta)\cos(2\theta) - 2\cos(\theta)\sin(2\theta)}{\cos(\theta)\cos(2\theta) - 2\cos(\theta)\sin(2\theta)}.$$

We obtain the points on the polar curve with vertical tangent line by finding all θ-values for which the numerator is zero and the denominator is nonzero (and vice-versa for the points on the polar curve with horizontal tangent line). Consequently, we must solve the following trigonometric equations.

$$\begin{aligned}
y'(\theta) &= 0 \\
-\sin(\theta)\cos(2\theta) - 2\cos(\theta)\sin(2\theta) &= 0 \\
-\sin(\theta)\cos(2\theta) - 2\sin(\theta)\cos^2(\theta) &= 0 && (\sin(2\theta) = 2\sin(\theta)\cos(\theta)) \\
-\sin(\theta)[\cos(2\theta) + 2\cos^2(\theta)] &= 0 \\
-\sin(\theta)\left[\frac{1}{2} + \frac{5}{2}\cos^2(\theta)\right] &= 0 && (\cos(2\theta) = \tfrac{1}{2} + \tfrac{1}{2}\cos^2(\theta)) \\
\sin(\theta) &= 0 \\
x'(\theta) &= 0 \\
\cos(\theta)\cos(2\theta) - 2\cos(\theta)\sin(2\theta) &= 0 \\
\cos(\theta)[\cos(2\theta) - 2\sin(2\theta)] &= 0
\end{aligned}$$

Consequently, we find that $y'(\theta) = 0$ if and only if $\sin(\theta) = 0$ if and only if $\theta = 0$, $\theta = \pi$, or $\theta = 2\pi$. Likewise, we find that $x'(\theta) = 0$ if and only if $\cos(\theta) = 0$ or $\cos(2\theta) = 2\sin(2\theta)$ if and only if $\theta = \frac{\pi}{2}$ or $\theta = \frac{3\pi}{2}$ or $\tan(2\theta) = \frac{1}{2}$. One solution of $\tan(2\theta) = \frac{1}{2}$ is $\theta = \frac{1}{2}\arctan\left(\frac{1}{2}\right)$, hence the general solution to this equation is $\theta = \frac{1}{2}\arctan\left(\frac{1}{2}\right) + k\pi$ for some integer k. We conclude the following.

(1.) The polar curve $r = \cos(2\theta)$ has horizontal tangents if and only if $\theta = 0, \pi$, or 2π.

(2.) The polar curve $r = \cos(2\theta)$ has vertical tangents if and only if $\theta = \frac{\pi}{2}, \frac{3\pi}{2}$, or $\frac{1}{2}\arctan\left(\frac{1}{2}\right) + k\pi$ for some integer k. By taking k small enough, we can ensure θ satisfies that $0 \leq \theta \leq 2\pi$. ◇

One of the most important reasons to care about polar coordinates is the notion of integration in polar coordinates. Essentially, the idea is that some curves in Cartesian coordinates can be more easily described in polar coordinates. Certainly, this is reasonable because Cartesian coordinates are designed to handle rectilinear curves, whereas polar coordinates are built from arcs and circles. Consequently, if we can parametrize a plane curve via some polar function $r = f(\theta)$, then it would be most natural to to find the area bounded by the parametric curve in polar coordinates as opposed to Cartesian coordinates. We note toward this end that for any polar curve $r = f(\theta)$ for which the radial coordinate r is a continuous function of the angular coordinate θ in polar coordinates for some real numbers a and b such that $a \leq \theta \leq b$, we can partition $[a, b]$ by choosing real numbers $a = \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_n = b$ with $\Delta\theta_i = \theta_i - \theta_{i-1}$ for each integer $1 \leq i \leq n$. Choosing sample points θ_i^* such that $\theta_{i-1} \leq \theta_i^* \leq \theta_i$ for each integer $1 \leq i \leq n$ and $\theta_1^* < \theta_2^* < \cdots < \theta_n^*$, we can approximate the area bounded by $r = f(\theta)$ and the polar axis using n sectors. Explicitly, each sector is subtended by an angle $\Delta\theta_i$ and possesses a radius of $r = f(\theta_i^*)$, hence the area of each sector is $a(\theta_i^*) = \frac{1}{2}[f(\theta_i^*)]\,\Delta\theta_i$. Consequently, if we sum the area of these sectors, we obtain that

$$\text{area bounded by the polar curve } f(\theta) \text{ and the polar axis from } a \text{ to } b \approx \sum_{i=1}^{n} \frac{1}{2}[f(\theta_i^*)]^2\,\Delta\theta_i.$$

By recognizing this as a Riemann sum and taking the limit as n approaches ∞, we conclude that

$$\text{area bounded by the polar curve } f(\theta) \text{ and the polar axis from } a \text{ to } b = \int_a^b \frac{1}{2}[f(\theta)]^2\,d\theta.$$

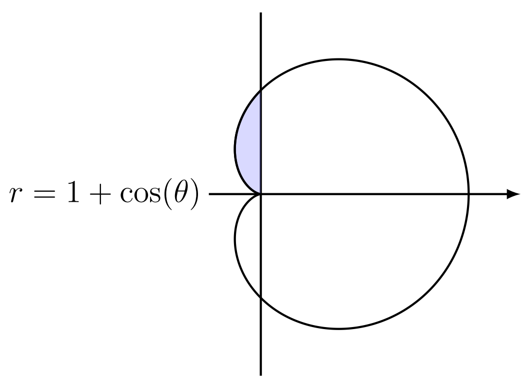
Formula 4.2.11 (Area Bounded by Polar Curves)**.** *Given any region $\mathcal{R}$ in polar coordinates that is bounded by any continuous polar curve $r = f(\theta)$ and any pair of rays $\theta = a$ and $\theta = b$ for any real numbers a and b such that $0 \leq b - a \leq 2\pi$, then area of $\mathcal{R}$ is given by the definite integral*

$$\text{area}(\mathcal{R}) = \frac{1}{2}\int_a^b [f(\theta)]^2\,d\theta.$$

Particularly, if $r = g(\theta)$ is any continuous polar curve such that $f(\theta) \geq g(\theta)$ for all real numbers θ such that $a \leq \theta \leq b$, then the area of the region $\mathcal{R}$ bounded by $f(\theta)$, $g(\theta)$, $\theta = a$, and $\theta = b$ is

$$\text{area}(\mathcal{R}) = \frac{1}{2}\int_a^b ([f(\theta)]^2 - [g(\theta)]^2)\,d\theta.$$

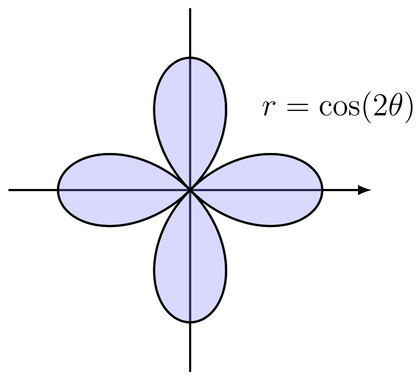
Example 4.2.12. We will compute in this example the area bounded by the polar curve below.



We begin by determining the rays $\theta = a$ and $\theta = b$ that bound the region in question. Observe that as θ ranges from $\theta = 0$ to $\theta = \pi$, the curve $r = f(\theta) = 1+\cos(\theta)$ is traced out counterclockwise from the point $(2, 0)$ to the point $(0, \pi)$: indeed, we have that $f(0) = 1 + \cos(0) = 2$, $f\left(\frac{\pi}{2}\right) = 1 + \cos\left(\frac{\pi}{2}\right) = 1$, and $f(\pi) = 1 + \cos(\pi) = 0$ so that the points $(2, 0)$, $\left(1, \frac{\pi}{2}\right)$, and $(0, \pi)$ lie on the curve. By definition of polar coordinates, the vertical axis pictured above corresponds to the ray $\theta = \frac{\pi}{2}$, and the left-hand side of the polar axis corresponds to the ray $\theta = \pi$. Consequently, the region $\mathcal{R}$ in question is bounded by $f(\theta) = 1 + \cos(\theta)$ for all real numbers θ such that $\frac{\pi}{2} \leq \theta \leq \pi$. We conclude that

$$\begin{aligned}\operatorname{area}(\mathcal{R}) = \frac{1}{2}\int_{\pi/2}^{\pi} [1+\cos(\theta)]^2\, d\theta &= \frac{1}{2}\int_{\pi/2}^{\pi} [1 + 2\cos(\theta) + \cos^2(\theta)]\, d\theta \\ &= \frac{1}{2}\int_{\pi/2}^{\pi} \left[\frac{3}{2} + 2\cos(\theta) + \frac{1}{2}\cos(2\theta)\right] d\theta \\ &= \frac{1}{2}\left[\frac{3}{2}\theta + 2\sin(\theta) + \frac{1}{4}\sin(2\theta)\right]_{\pi/2}^{\pi} = \frac{3\pi}{8} - 1.\end{aligned}$$

Example 4.2.13. We will compute in this example the area bounded by the polar curve below.

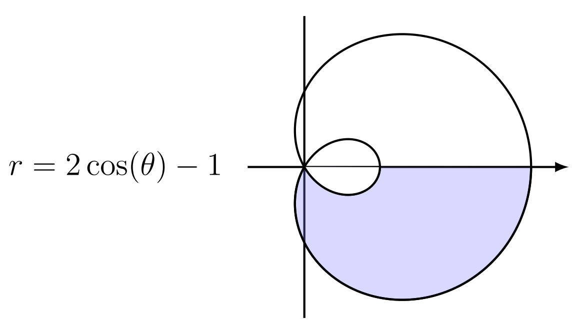


By symmetry, it suffices to compute the area bounded by one leaf of the above four-leaf clover. We begin by determining the rays $\theta = a$ and $\theta = b$ that bound the right-hand leaf. Observe that as θ ranges from $\theta = -\frac{\pi}{4}$ to $\theta = \frac{\pi}{4}$, the curve $r = f(\theta) = \cos(2\theta)$ is traced out counterclockwise from the point $\left(0, -\frac{\pi}{4}\right)$ to the point $(1, 0)$ and finally to the point $\left(0, \frac{\pi}{4}\right)$: indeed, we have that $f\left(-\frac{\pi}{4}\right) = \cos\left(-\frac{\pi}{2}\right) = 0$, $f(0) = \cos(0) = 1$, and $f\left(\frac{\pi}{4}\right) = \cos\left(\frac{\pi}{2}\right) = 0$ so that the points $\left(0, -\frac{\pi}{4}\right)$, $(1, 0)$, and $\left(0, \frac{\pi}{4}\right)$ lie on the curve. Even more, the right-hand leaf is bounded by $f(\theta) = \cos(2\theta)$ for all real numbers θ such that $-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}$. We conclude that the area of one leaf is given by

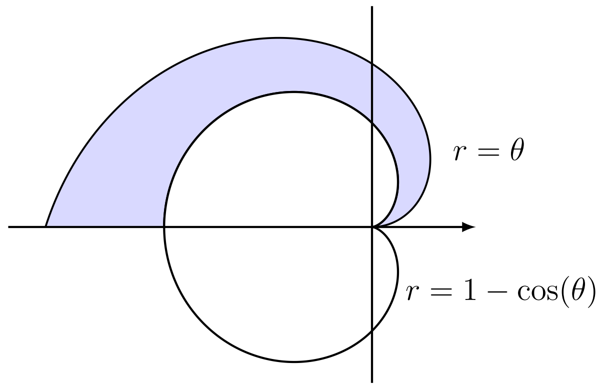
$$\frac{1}{2}\int_{-\pi/4}^{\pi/4} \cos^2(2\theta)\, d\theta = \frac{1}{2}\int_{-\pi/4}^{\pi/4} \left[\frac{1}{2} + \frac{1}{2}\cos(4\theta)\right] d\theta = \frac{1}{4}\left[\theta + \frac{1}{4}\sin(4\theta)\right]_{-\pi/4}^{\pi/4} = \frac{\pi}{8}.$$

Each of the four leaves of the clover has area $\frac{\pi}{8}$, hence the total area bounded by the curve is $\frac{\pi}{2}$.

Exercise 4.2.14. Compute the area bounded by the polar curve $r = f(\theta)$ pictured below.



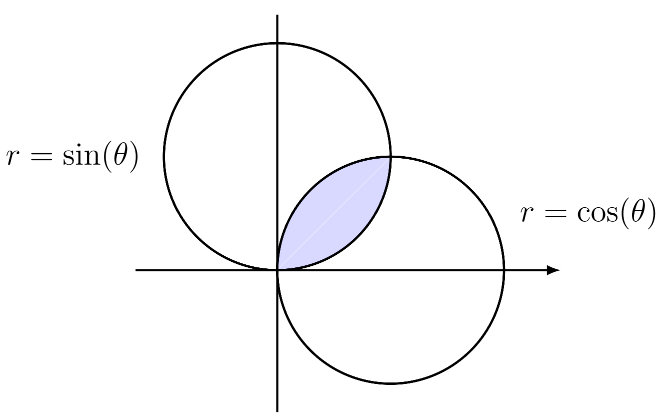
Example 4.2.15. We will compute in this example the area bounded by the polar curves below.



Considering that the shaded region $\mathcal{R}$ is bounded by the polar curves $f(\theta) = \theta$, $g(\theta) = 1 - \cos(\theta)$, $\theta = a$, and $\theta = b$ for some real numbers $0 \leq a \leq b \leq 2\pi$ and $f(\theta) \geq g(\theta)$ for all real numbers $\theta \geq 0$, by the formula for Area Bounded by Polar Curves, it suffices to determine the real numbers a and b. Certainly, as θ ranges from $\theta = 0$ to $\theta = \pi$, the curve $r = f(\theta) = \theta$ is traced out counterclockwise from the point $(0, 0)$ to the point (θ, θ). Likewise, the curve $r = g(\theta) = 1 - \cos(\theta)$ is traced out counterclockwise from the point $(0, 0)$ to the point $(\pi, 2)$ since we have that $g(0) = 1 - \cos(0) = 0$ and $g(\pi) = 1 - \cos(\pi) = 2$. We conclude that the area of the shaded region $\mathcal{R}$ is given by

$$\begin{aligned}
\text{area}(\mathcal{R}) &= \frac{1}{2}\int_0^\pi (\theta^2 - [1 - \cos(\theta)]^2)\, d\theta \\
&= \frac{1}{2}\int_0^\pi [\theta^2 - 1 + 2\cos(\theta) - \cos^2(\theta)]\, d\theta \\
&= \frac{1}{2}\int_0^\pi \left[\theta^2 - \frac{3}{2} + 2\cos(\theta) - \frac{1}{2}\cos(2\theta)\right] d\theta \\
&= \frac{1}{2}\left[\frac{\theta^3}{3} - \frac{3}{2}\theta + 2\sin(\theta) - \frac{1}{4}\sin(2\theta)\right]_0^\pi = \frac{\pi^3}{6} - \frac{3\pi}{4}
\end{aligned}$$

Exercise 4.2.16. Compute the area bounded by the polar curves $r = f(\theta)$ pictured below.



Chapter 5

Sequences and Series

5.1 Sequences

One of the main focuses of any Calculus II course is to understand sequences and series. Unwittingly, we have all encountered sequences in our lives at some point: if you have ever counted while holding your breath, then you have recited a sequence; if you have ever attempted to memorize some of the digits in the decimal expansion of π, then you have attempted to memorize a sequence; or if you have ever entered a telephone number to place a call, then you have a sequence entered into your phone. Basically, a **sequence** is an ordered list of objects. Put more precisely, a sequence is an ordered list $\{a_n\}_{n=1}^k$ of k objects $a_1, a_2, \dots, a_k$, where k is a positive **integer** (or whole number). We use the subscript n as an **index** so that the symbol a_n is the nth object that appears in the sequence. We will typically consider sequences of real numbers that start with $n = 0$, but it is possible to think about sequences that begin with any non-negative (or even negative!) whole number index.

Unfortunately, the digits of a telephone number are often quite random, and there is no formula for the nth digit in the decimal expansion of π, so it is impossible to come up with formulas for these sequences; however, there are plenty of sequences for which the nth term is formulaic.

Example 5.1.1. Consider the **natural numbers** $\mathbb{N}$ that are obtained by counting up from 1, adding 1 each time. We can list these sequentially as $1, 2, 3, \dots, n, \dots$, hence the infinite sequence

$$\{n\}_{n=1}^\infty = \lim_{k\to\infty} \{n\}_{n=1}^k = \lim_{k\to\infty} \{1, 2, 3, \dots, k\} = \{1, 2, 3, \dots, n, \dots\}$$

consists of all natural numbers. We could also write this sequence as $a_n = n$ for each integer $n \geq 1$.

Our interest lies in sequences (finite or infinite) for which there exists a real function $f(x)$ such that $a_n = f(n)$ for all elements n of some **index set** N consisting of non-negative whole numbers. We refer to the function $f(x)$ such that $a_n = f(n)$ for all indices n as the **closed form** of the sequence a_n, or we may alternatively say that $f(n)$ is a **closed-form expression** for a_n. Certainly, the benefit of obtaining a closed-form expression for a sequence a_n of real numbers is that if we wish to determine a_0, a_1, or a_2, then it suffices to compute $f(0)$, $f(1)$, or $f(2)$, respectively.

Example 5.1.2. Consider the infinite sequence that begins $1, -1, 1, -1, \dots$. Observe that the nth term a_n is obtained from the previous term a_{n-1} by multiplying by -1. Consequently, we have that

$$a_n = (-1)a_{n-1} = (-1)^2 a_{n-2} = \cdots = (-1)^n a_0.$$

Considering that $a_0 = 1$, we conclude that $a_n = (-1)^n$ for all integers $n \geq 0$.

Example 5.1.3. Consider the infinite sequence that begins $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$. Observe that in order to move from one term of the sequence to the next, we divide by 2. Consequently, the relationship between consecutive terms of the sequence is given by $a_n = \frac{a_{n-1}}{2}$. By repeating this, we find that

$$a_n = \frac{1}{2}a_{n-1} = \left(\frac{1}{2}\right)^2 a_{n-2} = \cdots = \left(\frac{1}{2}\right)^n a_0.$$

Considering that $a_0 = 1$, we conclude that $a_n = \left(\frac{1}{2}\right)^n$ for all integers $n \geq 0$.

One can define a sequence **recursively** by providing a formula $a_n = f(a_0, a_1, \dots, a_{n-1})$ for the nth term of the sequence in terms of some of the preceding terms of the sequence.

Example 5.1.4. Observe that $a_n = n$ is a recursive sequence defined by $a_n = a_{n-1} + 1$ for all n.

Example 5.1.5. Consider the recursive sequence defined by $a_n = na_{n-1}$ for all integers $n \geq 0$. By repeatedly applying the recursive definition of a_n, we find that

$$a_n = na_{n-1} = n(n-1)a_{n-2} = n(n-1)(n-2)a_{n-3} = \cdots = n(n-1)(n-2)\cdots(n-k+1)a_{n-k}.$$

Consequently, if we assume that $n \geq 0$ and $a_0 = 1$, then we obtain that

$$a_n = n(n-1)(n-2)\cdots 2 \cdot 1 \cdot a_0 = n(n-1)\cdots 2 \cdot 1.$$

We refer to this recursive sequence as n-**factorial**, and we write $n! = n(n-1)(n-2)\cdots 2 \cdot 1$.

Example 5.1.6. Consider the recursive sequence $a_n = 2a_{n-1}$ defined for each integer $n \geq 1$ with $a_0 = 1$. By repeatedly substituting the recursive formula for a_n, we obtain that

$$a_n = 2a_{n-1} = 2^2 a_{n-2} = \cdots = 2^n a_0.$$

Considering that $a_0 = 1$, we conclude that $a_n = 2^n$ for all integers $n \geq 0$.

Example 5.1.7. Curiously enough, even the most simple-looking recursive sequences can admit surprisingly complicated closed forms. Consider the famed **Fibonacci sequence** $a_n = a_{n-1} + a_{n-2}$ for each integer $n \geq 2$ with $a_0 = 0$ and $a_1 = 1$. One can prove that the closed form for this is

$$a_n = f(n) = \frac{(-1)^{n-1}\phi^{-n} + \phi^n}{\sqrt{5}} \text{ for the Golden Ratio } \phi = \frac{1+\sqrt{5}}{2}.$$

Computing closed forms for recursive sequences is crucial in the field of computer science, and a rigorous treatment of the subject is often supplied in any course on numerical analysis, but we will try to consider recursive sequences with an understanding of our limitations in this course.

Given a sequence a_n for some index set N that consists of positive whole numbers, we can always "reindex" the sequence so that it is defined for each integer $n \geq 1$, so we will assume henceforth that our sequences are all of the form $\{a_n\}_{n=1}^{\infty}$. We say that the sequence $\{a_n\}_{n=1}^{\infty}$ **converges** if there exists some real number L such that for every real number $\varepsilon > 0$, there exists a positive integer m with the property that $|a_n - L| < \varepsilon$ whenever we have that $n \geq m$. Put another way, the quantity L can be made arbitrarily close to the value of a_n by taking n to be sufficiently large. Given that no such real number L exists, we say that $\{a_n\}_{n=1}^{\infty}$ **diverges**. Further, if the terms of a_n increase (or decrease) without bound, then a_n **diverges to infinity** (or negative infinity).

Example 5.1.8. Consider the sequence $a_n = \frac{1}{n}$ for all integers $n \geq 1$. Observe that as n grows arbitrarily large, its reciprocal $\frac{1}{n}$ becomes arbitrarily close to 0. Consequently, we suspect that

$$\lim_{n \to \infty} \frac{1}{n} = 0.$$

Let us prove this by definition. Given any real number $\varepsilon > 0$, we want to find a positive integer m such that whenever $n \geq m$, we have that $\left|\frac{1}{n}\right| < \varepsilon$. Considering that $n \geq 1$, we have that $\frac{1}{n} > 0$ so that $\left|\frac{1}{n}\right| = \frac{1}{n}$. We can ensure that $\frac{1}{n} < \varepsilon$ by taking $n > \frac{1}{\varepsilon}$, hence our choice for M is quite intuitive: we should simply take $M = \frac{1}{\varepsilon}$. Unravelling this thought process gives a formal proof.

Proof. We claim that $\lim_{n \to \infty} \frac{1}{n} = 0$. Given any real number $\varepsilon > 0$, if we have that $n > M = \frac{1}{\varepsilon}$, then

$$\left|\frac{1}{n}\right| = \frac{1}{n} < \frac{1}{M} = \frac{1}{\frac{1}{\varepsilon}} = \varepsilon. \qquad \square$$

By definition, the limit of an infinite sequence a_n depends only on the values that a_n takes for sufficiently large indices n. Given some arbitrarily large (but fixed) positive integer m, we refer to the values of a_n for all indices $n \geq m$ as the **m-tail** of the sequence a_n. Consequently, the limit of an infinite sequence depends only on the m-tail of a_n, and as such, it will not be altered if we change (or omit) finitely many terms — namely, all of those terms a_n for which $n \leq m$. Further, if there exists a real number C such that $a_n = C$ for all indices $n \geq m$, then $\lim_{n \to \infty} a_n = C$.

Other than the Fibonacci sequence, we have studied (and will primarily study) only sequences $\{a_n\}_{n=1}^\infty$ with a closed form, i.e., infinite sequences for which there exists a function $f(n)$ such that $a_n = f(n)$ for each integer $n \geq 1$. Consequently, we can think about sequences as functions whose domains have been restricted to the positive whole numbers. Using the tools that we have from Calculus I — limits, derivatives, L'Hôpital's Rule, etc. — we can better understand sequences with closed forms in terms of the functions that define them. Particularly, the following holds.

Proposition 5.1.9 (Limit of a Sequence with a Closed Form)**.** *Given any sequence of real numbers a_n such that there exists a real function $f(x)$ with $a_n = f(n)$ for all sufficiently large integers n,*

$$\lim_{n \to \infty} a_n = \lim_{x \to \infty} f(x).$$

Proof. Compare the definitions to see that this is true. Explicitly, if there exists a real number L such that $\lim_{x \to \infty} f(x) = L$, then by definition, given a real number $\varepsilon > 0$, there exists a positive integer M such that $|f(x) - L| < \varepsilon$ for all real numbers $x > M$. But if this is true for all real numbers $x > M$, then it is certainly true for all positive whole numbers $n > M$ so that $\lim_{n \to \infty} a_n = L$. Use the analogous argument in the case that $\lim_{x \to \infty} f(x) = \pm\infty$ to show that $\lim_{n \to \infty} a_n = \pm\infty$. $\square$

Exercise 5.1.10. Compute the limit of the sequence $a_n = \dfrac{1}{2^n}$, or prove that it does not exist.

Solution. Considering that $a_n = f(n) = \dfrac{1}{2^n}$, it follows by Proposition 5.1.9 that

$$\lim_{n \to \infty} a_n = \lim_{n \to \infty} \frac{1}{2^n} = \lim_{x \to \infty} \frac{1}{2^x} = 0. \qquad \diamond$$

Exercise 5.1.11. Compute the limit of the sequence $a_n = (-1)^n$, or prove that it does not exist.

Proof. Observe that $a_n = (-1)^n$ diverges: if n is even, then $a_n = 1$ because -1 to an even power is 1. On the other hand, if n is odd, then $a_n = -1$ because -1 to an odd power is -1. Consequently, there is no real number L arbitrarily close to a_n for all sufficiently large integers n. $\square$

Exercise 5.1.12. Compute the limit of the sequence $a_n = \dfrac{\ln(n)}{n}$, or prove that it does not exist.

Exercise 5.1.13. Compute the limit of the sequence $a_n = \dfrac{n^5 + 3n^2 + 1}{3n^4 + n + 1}$, or prove it does not exist.

Given any real number c and any nonzero real number r, we refer to any sequence of the form $a_n = cr^n$ as a **geometric sequence**. Observe that if $r = 1$, then $r^n = 1$ so that $a_n = cr^n = c$ is a constant sequence, hence every constant sequence is geometric. Crucially, a nonzero sequence a_n of real numbers is geometric if and only if the ratio of consecutive terms of the sequence satisfy that

$$\frac{a_n}{a_{n-1}} = r$$

for some nonzero real number r. Observe that if the above identity holds, then we have that

$$a_n = ra_{n-1} = r^2 a_{n-2} = \cdots = r^n a_0.$$

Consequently, the 0th term a_0 of the geometric sequence is its **coefficient** $a_0 = c$, and we refer to the nonzero real number r as the **common ratio** of the geometric sequence $a_n = cr^n$. Our next proposition classifies the convergence of geometric sequences based on their common ratio.

Proposition 5.1.14 (Convergence of Geometric Sequences)**.** *Given any real number c and any nonzero real number r, the geometric sequence cr^n obeys the following rule for convergence.*

$$\lim_{n \to \infty} cr^n = \begin{cases} 0 & \text{if } -1 < r < 1 \\ c & \text{if } r = 1 \\ \text{diverges} & \text{if } r > 1 \text{ or } r \leq -1 \end{cases}$$

Proof. Consider the real function $f(x) = cr^x$. Observe that $f(n) = cr^n$ for each integer $n \geq 0$, hence

$$\lim_{n \to \infty} cr^n = \lim_{x \to \infty} f(x) = \lim_{x \to \infty} cr^x = c\Big(\lim_{x \to \infty} r^x\Big).$$

We will assume first that $r > 0$ so that $r^x > 0$ for all real numbers x. Consequently, the real number $\ln(r^x)$ satisfies that $r^x = e^{\ln(r^x)} = e^{x \ln(r)}$. Observe that if $r < 1$, then $\ln(r) < 0$ so that $t = x\ln(r) < 0$ for all real numbers $x > 0$. Even more, as x approaches ∞, we have that t approaches $-\infty$, hence

$$\lim_{n \to \infty} cr^n = c\Big(\lim_{x \to \infty} r^x\Big) = c\Big(\lim_{x \to \infty} e^{x \ln(r)}\Big) = c\Big(\lim_{t \to -\infty} e^t\Big) = 0.$$

Conversely, if $r > 1$, then $\ln(r) > 0$ so that $x \ln(r) > 0$ for all real numbers $x > 0$. We conclude that

$$\lim_{n \to \infty} cr^n = c\Big(\lim_{x \to \infty} r^x\Big) = c\Big(\lim_{x \to \infty} e^{x \ln(r)}\Big) = c\Big(\lim_{t \to \infty} e^t\Big) = \infty.$$

Observe that if $-1 < r < 0$, then $0 < |r| < 1$. By Proposition 5.1.21, the geometric sequence cr^n converges to 0. Conversely, if $r \leq -1$, then $-r \geq 1$, hence the sequence $cr^n = c(-1)^n(-r)^n$ diverges because it takes both positive and negative values for infinitely many integers $n \geq 0$. $\square$

Exercise 5.1.15. Explain whether each of the following sequences is geometric; if so, compute the coefficient c and the common ratio r; and determine with justification if a_n converges or diverges.

(a.) $a_n = 3^{2-3n}$

(b.) $6, -3, \frac{3}{2}, -\frac{3}{4}, \dots$

(c.) $a_n = (-1)^n$

(d.) $-1, 2, -3, 4, \dots$

(e.) $a_n = \ln(e^{2\pi n})$

(f.) $\frac{2}{3}, -\frac{2}{9}, -\frac{2}{27}, \frac{2}{81}, \dots$

(g.) $a_n = \sin(\pi n)$

(h.) $3, 6, 9, 12, \dots$

Solution. (a.) By simplifying the expression 3^{3-2n}, we find that

$$a_n = 3^{2-3n} = (3^2)(3^{-3n}) = \frac{9}{3^{3n}} = \frac{9}{27^n} = 9\left(\frac{1}{27}\right)^n.$$

Consequently, the sequence is geometric with coefficient $c = 9$ and common ratio $r = \frac{1}{27}$. Considering that $-1 < r < 1$, we conclude by Proposition 5.1.14 that $a_n = 3^{2-3n}$ converges.

(b.) Observe that the ratios of successive terms is constant and satisfies that

$$\frac{-3}{6} = \frac{\frac{3}{2}}{-3} = \frac{-\frac{3}{4}}{\frac{3}{2}} = -\frac{1}{2}.$$

Consequently, the sequence is geometric with common ratio $r = -\frac{1}{2}$ and coefficient $c = 6$. Considering that $-1 < r < 1$, we conclude by the above proposition that the sequence converges. ◇

Considering that the definition of the limit of a sequence is closely related to the definition of the limit of a function at $\pm\infty$, it is not surprising that the familiar limit laws holds for sequences.

Proposition 5.1.16 (Limit Laws for Sequences)**.** *Given any pair of convergent infinite sequences of real numbers a_n and b_n, the following limit properties hold.*

(i.) (**Additive Property of Limits**) *Limits distribute across sums and differences.*

$$\lim_{n \to \infty} (a_n \pm b_n) = \left(\lim_{n \to \infty} a_n\right) \pm \left(\lim_{n \to \infty} b_n\right)$$

(ii.) (**Multiplicative Property of Limits**) *Limits distribute across products.*

$$\lim_{n \to \infty} (a_n b_n) = \left(\lim_{n \to \infty} a_n\right)\left(\lim_{n \to \infty} b_n\right)$$

(iii.) (**Quotient Property of Limits**) *Limits distribute across quotients if* $\lim_{n \to \infty} b_n$ *is nonzero.*

$$\lim_{n \to \infty} \frac{a_n}{b_n} = \frac{\lim_{n \to \infty} a_n}{\lim_{n \to \infty} b_n}$$

Likewise, we point out the following analog of the Squeeze Theorem from Calculus I.

Proposition 5.1.17 (Squeeze Theorem for Sequences)**.** *Given any infinite sequences of real numbers a_n, b_n, and c_n satisfying that*

(1.) *b_n and c_n are convergent with $\lim_{n \to \infty} b_n = \lim_{n \to \infty} c_n$ and*

(2.) *$b_n \leq a_n \leq c_n$ for all sufficiently large integers n,*

the sequence a_n is convergent with $\lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n = \lim_{n \to \infty} c_n$.

Exercise 5.1.18. Compute the limit of the sequence $a_n = \dfrac{(-1)^n}{n}$.

Solution. Observe that $-1 \leq (-1)^n \leq 1$ for all integers $n \geq 1$, hence we have that

$$-\frac{1}{n} \leq \frac{(-1)^n}{n} \leq \frac{1}{n}$$

for all integers $n \geq 1$. By the Squeeze Theorem for Sequences, we conclude that $\lim_{n \to \infty} a_n = 0$. ⋄

Exercise 5.1.19. Compute the limit of the sequence $a_n = \dfrac{\sin(n)}{n}$.

Solution. Observe that $-1 \leq \sin(n) \leq 1$ for all integers $n \geq 0$, hence we have that

$$-\frac{1}{n} \leq \frac{\sin(n)}{n} \leq \frac{1}{n}$$

for all integers $n \geq 0$. By the Squeeze Theorem for Sequences, we conclude that $\lim_{n \to \infty} a_n = 0$. ⋄

Exercise 5.1.20. Compute the limit of the sequence $a_n = \dfrac{(-2)^n + 2^n}{3^n}$.

Solution. Observe that $0 \leq (-2)^n + 2^n \leq 2^{n+1}$ for all integers $n \geq 0$. Explicitly, if n is even, then $(-2)^n + 2^n = 2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$. Likewise, if n is odd, then $(-2)^n + 2^n = -(2^n) + 2^n = 0$. Consequently, by dividing each expression in this inequality by 3^n, we find that

$$0 \leq \frac{(-2)^n + 2^n}{3^n} \leq \frac{2^{n+1}}{3^n}$$

for all integers $n \geq 0$. Considering that the upper bound of this inequality

$$\frac{2^{n+1}}{3^n} = \frac{2(2^n)}{3^n} = 2\left(\frac{2^n}{3^n}\right) = 2\left(\frac{2}{3}\right)^n$$

is a convergent geometric sequence, we conclude by the Squeeze Theorem that $\lim_{n \to \infty} a_n = 0$. ⋄

Proposition 5.1.21 (Convergence to Zero in Absolute Value Implies Convergence to Zero)**.** *Given any infinite sequence of real numbers a_n, if $\lim_{n\to\infty} |a_n| = 0$, then $\lim_{n\to\infty} a_n = 0$.*

Proof. By definition of the absolute value function, we have that $|a_n| = -a_n$ if $a_n < 0$ and $|a_n| = a_n$ if $a_n \geq 0$. Consequently, we find that $-|a_n| \leq a_n \leq |a_n|$ for all integers $n \geq 0$. Considering that $\lim_{n\to\infty} |a_n| = 0$ by hypothesis, the Squeeze Theorem for Sequences ensures that $\lim_{n\to\infty} a_n = 0$. $\square$

Proposition 5.1.22 (Convergence of Exponential by Factorial)**.** *Given any real number r, we have*

$$\lim_{n\to\infty} \frac{r^n}{n!} = 0.$$

Proof. By Proposition 5.1.21, it suffices to prove that

$$\lim_{n\to\infty} \frac{|r|^n}{n!} = \lim_{n\to\infty} \frac{|r^n|}{n!} = \lim_{n\to\infty} \left|\frac{r^n}{n!}\right| = 0.$$

Certainly, this holds if $-1 < r < 1$ because the numerator is a convergent geometric sequence and the limit of the denominator is infinity; otherwise, if $|r| \geq 1$, then we may find an integer $a \geq 1$ such that $a \leq |r| \leq a + 1$. Consider the nth term of the sequence for any integer $n \geq a + 2$.

$$\frac{|r|^n}{n!} = \frac{|r| \cdot |r| \cdots |r| \cdot |r| \cdot |r| \cdots |r| \cdot r}{1 \cdot 2 \cdots a(a+1)(a+2)\cdots(n-1)n} = \underbrace{\frac{|r|}{1} \cdot \frac{|r|}{2} \cdots \frac{|r|}{a}}_{\text{Call this } C.} \cdot \underbrace{\frac{|r|}{a+1} \cdot \frac{|r|}{a+2} \cdots \frac{|r|}{n-1}}_{\text{Each factor here is } \leq 1.} \cdot \frac{|r|}{n}$$

Consequently, we can bound the sequence at hand below by 0 (because all terms are non-negative), and we can bounded the sequence above by the product of the red and black terms as follows.

$$0 \leq \frac{|r|^n}{n!} \leq \frac{C|r|}{n}$$

Considering that $\lim_{n\to\infty} \frac{C|r|}{n} = 0$, the Squeeze Theorem for Sequences ensures that $\lim_{n\to\infty} \frac{|r|^n}{n!} = 0$. $\square$

Continuous functions are characterized by the property that for any real number L, we have

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

Luckily, the same holds for limits of continuous functions of sequences.

Proposition 5.1.23 (Commutative Property of Limits and Continuous Functions)**.** *Given any continuous real function $f(x)$ and any convergent infinite sequence of real numbers a_n such that $\lim_{n\to\infty} a_n = L$, the infinite sequence of real numbers $f(a_n)$ is convergent, and its limit is given by*

$$\lim_{n\to\infty} f(a_n) = f\Big(\lim_{n\to\infty} a_n\Big) = f(L).$$

Exercise 5.1.24. Compute the limit of the sequence $a_n = \sin(e^{-n})$.

Solution. Observe that $\lim_{n\to\infty} e^{-n} = \lim_{n\to\infty} \frac{1}{e^n} = 0$. Considering that $\sin(x)$ is a continuous function, the Commutative Property of Limits and Continuous Functions implies that

$$\lim_{n\to\infty} \sin(e^{-n}) = \sin\Big(\lim_{n\to\infty} e^{-n}\Big) = \sin(0) = 0. \qquad \diamond$$

Exercise 5.1.25. Compute the limit of the sequence $a_n = \sqrt{4 + \frac{1}{n}}$.

Solution. By the Commutative Property of Limits and Continuous Functions, we find that

$$\lim_{n \to \infty} \sqrt{4 + \frac{1}{n}} = \sqrt{\lim_{n \to \infty} \left(4 + \frac{1}{n}\right)} = \sqrt{4 + \lim_{n \to \infty} \frac{1}{n}} = \sqrt{4} = 2. \qquad \diamond$$

Exercise 5.1.26. Compute the limit of the sequence $a_n = \arctan\left(\frac{\cos(n)}{n}\right)$.

By Proposition 5.1.14, the convergence of geometric sequences is fully classified. Consequently, one naturally wonders if it is in fact possible to classify the convergence of all infinite sequences of real numbers. Our first step toward that lofty goal is to study sequences that are bounded. We say that a sequence a_n is **bounded above** if there exists a real number M^+ such that $a_n \leq M^+$ for all sufficiently large integers n. Likewise, we say that a sequence is **bounded below** if there exists a real number M^- such that $a_n \geq M^-$ for all sufficiently large integers n. Combining these two notions, we say that a sequence is **bounded** if it is bounded above and bounded below. Conversely, any sequence that is either not bounded above or not bounded below is said to be **unbounded**.

Proposition 5.1.27 (Convergent Sequences Are Bounded)**.** *Every convergent infinite sequence of real numbers is bounded. Explicitly, if the sequence a_n converges, then a_n must be bounded.*

Proof. By definition, if a_n converges, then there exists a real number L such that $\lim_{n \to \infty} a_n = L$. By definition of the limit, there exists a positive real number N such that $|a_n - L| < 1$ for all $n > N$. Unraveling this gives that $L - 1 < a_n < L + 1$ for all $n > N$. We can choose M^+ to be larger than $a_1, a_2, \ldots, a_N$ and $L + 1$ and choose M^- to be smaller than $a_1, a_2, \ldots, a_N$ and $L - 1$. $\square$

By the contrapositive, Proposition 5.1.27 guarantees that if a_n is unbounded, then a_n diverges. On the other hand, Exercise 5.1.11 provides an example of a bounded but divergent sequences. We say that a sequence is **oscillating** if it is takes (at least) two distinct values for infinitely many indices. Quite generally, every oscillating sequence is divergent — even those that are bounded.

Given that an infinite sequence of real numbers a_n is (eventually) monotone, then its boundedness is in fact sufficient to conclude its convergence. We say that a_n is **monotone** if it is either **increasing** or **decreasing**. Particularly, we say that a_n is increasing if $a_{n+1} \geq a_n$ for all sufficiently large integers n, and we say that a_n is decreasing if $a_{n+1} \leq a_n$ for all sufficiently large integers n. By Calculus I, we may recall that a differentiable real function $f(x)$ is increasing if $f'(x) > 0$ for all sufficiently large real numbers x and decreasing if $f'(x) < 0$ for all sufficiently large real numbers x. Consequently, it is sometimes possible to determine the increasing or decreasing nature of a sequence a_n if it admits a closed form $a_n = f(n)$ such that $f(x)$ is a differentiable real function.

Proposition 5.1.28 (Criterion for Monotonicity)**.** *Given any infinite sequence of real numbers a_n such that $a_n = f(n)$ for some differentiable real function $f(x)$,*

(a.) *if $f'(x) > 0$ for all sufficiently large real numbers x, then a_n is increasing and*

(b.) *if $f'(x) < 0$ for all sufficiently large real numbers x, then a_n is decreasing.*

Exercise 5.1.29. Explain whether the sequence $a_n = \sin\left(\frac{1}{n}\right)$ is increasing, decreasing, or neither.

Solution. By the Criterion for Monotonicity, it suffices to check the derivative of the function that defines a_n. Observe that $a_n = f(n)$ for the differentiable real function $f(x) = \sin\left(\frac{1}{x}\right)$ with

$$f'(x) = \frac{d}{dx}\sin\left(\frac{1}{x}\right) = \cos\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right) = -\frac{\cos\left(\frac{1}{x}\right)}{x^2}.$$

Considering that $x^2 > 0$ and $\cos\left(\frac{1}{x}\right) > 0$ for all real numbers $x > 0$, we conclude that $f'(x) < 0$ for all real numbers $x > 0$ so that the sequence a_n is decreasing. ◇

Exercise 5.1.30. Explain whether the sequence $a_n = -ne^{-n^2}$ is increasing, decreasing, or neither.

Solution. Like in Exercise 5.1.29, we may determine if a_n is increasing or decreasing by checking the derivative of the differentiable real function $f(x) = -xe^{-x^2}$ such that $a_n = f(n)$.

$$f'(x) = \frac{d}{dx}\left(-xe^{-x^2}\right) = -xe^{-x^2}(-2x) - e^{-x^2} = e^{-x^2}(2x^2 - 1)$$

Considering that $e^{-x^2} > 0$ and $2x^2 - 1 > 0$ for all real numbers $x \geq 1$, we conclude that $f'(x) > 0$ for all real numbers $x \geq 1$, hence the sequence a_n is increasing. ◇

Exercise 5.1.31. Explain whether the sequence $a_n = \cos(\pi n)$ is increasing, decreasing, or neither.

Solution. Observe that $\cos(\pi n)$ oscillates between 1 and -1; it is neither increasing nor decreasing.

$$\cos(\pi n) = \begin{cases} 1 & \text{if } n \text{ is even} \\ -1 & \text{if } n \text{ is odd} \end{cases}$$ ◇

One of the most important theorems in the study of infinite sequences is the following.

Theorem 5.1.32 (Monotone Convergence Theorem)**.** *An infinite sequence of real numbers a_n that is increasing or decreasing for all sufficiently large integers n converges if and only it is bounded.*

(a.) *If a_n is increasing and bounded above by M^+, then a_n converges and $\lim_{n\to\infty} a_n \leq M^+$.*

(b.) *If a_n is decreasing and bounded below by M^-, then a_n converges and $\lim_{n\to\infty} a_n \geq M^-$.*

Exercise 5.1.33. Compute the limit of the recursive sequence $a_n = \sqrt{2a_{n-1}}$ with $a_0 = \sqrt{2}$.

Solution. By plugging in $n = 1$, we find that $a_1 = \sqrt{2\sqrt{2}} \geq \sqrt{2} = a_0$, hence we suspect that a_n is increasing. We note also that $a_1 < \sqrt{2}$, hence we will prove that a_n is bounded above by 2. Observe that if $a_n < 2$, then $a_{n+1} = \sqrt{2a_n} < \sqrt{2 \cdot 2} = 2$. By the Principle of Mathematical Induction, we conclude that $a_n < 2$. We turn our attention back to the increasing property of a_n. We will assume that $a_{n+1} \geq a_n$. By definition of the sequence and the fact that $a_{n+1} < 2$, we have that

$$a_{n+2} = \sqrt{2a_{n+1}} > \sqrt{a_{n+1}^2} = a_{n+1}.$$

Consequently, the sequence is increasing and bounded above by 2, hence by the Monotone Convergence Theorem, it follows that there exists a real number L such that

$$L = \lim_{n \to \infty} a_{n+1} = \lim_{n \to \infty} \sqrt{2a_n} = \sqrt{\lim_{n \to \infty} (2a_n)} = \sqrt{2\Big(\lim_{n \to \infty} a_n\Big)} = \sqrt{2L}.$$

By squaring both sides, we find that $L^2 = 2L$ so that $L^2 - 2L = 0$ and $L(L-2) = 0$. Considering that a_n is increasing and bounded below by $\sqrt{2}$, we conclude that $L = 2$. ⋄

Exercise 5.1.34. Compute the limit of the recursive sequence $a_n = \sqrt{2 + a_{n-1}}$ with $a_0 = \sqrt{2}$.

5.2 Basics of Infinite Series

One of the most powerful and important tools in all of mathematics is the infinite series. Countless applications for series abound in approximation theory, real analysis, complex analysis, combinatorics, probability, and statistics. Concretely, an infinite series can be used to approximate π (and many other irrational numbers) to any desired degree of accuracy: we will eventually learn that

$$\pi = 4 - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} - \frac{4}{11} + \cdots.$$

Of course, we have already familiarized ourselves with finite series: the Riemann sum

$$\sum_{k=0}^{n} f(x_k)\,\Delta x_k = f(x_0)\,\Delta x_0 + f(x_1)\,\Delta x_1 + \cdots + f(x_n)\,\Delta x_n$$

of the real function $f(x)$ for sequence of points $x_0, x_1, \dots, x_n$ with $\Delta x_k = x_k - x_{k-1}$ for each integer $1 \leq k \leq n$ is a finite series from integral calculus. We refer to this presentation of the sum as sigma notation (named for the Greek letter sigma Σ). Certainly, a finite series can be evaluated by simply adding up all of its terms, hence we are interested in evaluating infinite series (when possible).

We define an **infinite series** as the limit of the finite sums of its **general term** a_n, i.e.,

$$a_0 + a_1 + a_2 + \cdots = \sum_{k=0}^{\infty} a_k = \lim_{n \to \infty} \sum_{k=0}^{n} a_k = \lim_{n \to \infty} s_n.$$

We call the sequence $s_n = \sum_{k=0}^{n} a_k = a_0 + a_1 + a_2 + \cdots + a_n$ the n**th partial sum** of the infinite series. Consequently, an infinite series is the limit of the sequence $\{s_n\}_{n=1}^{\infty}$ of its partial sums.

Example 5.2.1. Consider the following infinite series.

$$\sum_{k=1}^{\infty} \frac{1}{k}$$

By definition, the nth partial sum of this series is the finite sum

$$s_n = \sum_{k=1}^{n} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Example 5.2.2. Consider the following infinite series.

$$\sum_{k=1}^{\infty} \frac{1}{k^2+k}$$

By definition, the nth partial sum of this series is the finite sum

$$s_n = \sum_{k=1}^{n} \frac{1}{k^2+k} = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \cdots + \frac{1}{n^2+n}.$$

By evaluating the first four partial sums explicitly, we can deduce a closed form expression for s_n.

$$s_1 = \frac{1}{2}$$

$$s_2 = \frac{1}{2} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

$$s_3 = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} = \frac{9}{12} = \frac{3}{4}$$

$$s_4 = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \frac{1}{20} = \frac{4}{5}$$

$$s_n = \frac{n}{n+1}$$

Consequently, we can evaluate the infinite series at hand by taking the limit of this sequence.

$$\sum_{k=1}^{\infty} \frac{1}{k^2+k} = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k^2+k} = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{n}{n+1} = 1$$

Considering an infinite series as the limit of its sequence of partial sums enables us to apply all of the techniques of sequences we studied in Section 5.1 to our study of infinite series. Even still, certain types of infinite series are easier to compute than others. One of these is the **telescoping series** whose nth partial sum can be written as $s_n = C + f(n)$ for some real number C and some real function $f(x)$. Consequently, the telescoping series converges if and only if $s_n = C + f(n)$ converges if and only if $f(n)$ converges, and the value of the telescoping series is equal to

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} [C + f(n)] = C + \lim_{n\to\infty} f(n).$$

Example 5.2.3. Consider the following infinite series.

$$\sum_{k=2}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1}\right)$$

By plugging in several values of the index k, we can deduce the value of the series.

$$\sum_{k=2}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1}\right) = \underbrace{\left(\frac{1}{2} - \frac{1}{3}\right)}_{\text{Plug in } k=2.} + \underbrace{\left(\frac{1}{3} - \frac{1}{4}\right)}_{\text{Plug in } k=3.} + \underbrace{\left(\frac{1}{4} - \frac{1}{5}\right)}_{\text{Plug in } k=4.} + \underbrace{\left(\frac{1}{5} - \frac{1}{6}\right)}_{\text{Plug in } k=5.} + \cdots = \frac{1}{2}.$$

Explicitly, each of the terms with colored font will be added to itself with the opposite sign, hence everything will cancel except the first term of the infinite series. Bearing this in mind, it is clearer now why such a series is referred to as a telescoping series because a telescope retracts to a point.

Using the technique of partial fraction decomposition from the Partial Fraction Decomposition Theorem, certain infinite series are revealed to be telescoping. We illustrate as follows.

Example 5.2.4. Consider the following infinite series.

$$\sum_{k=0}^{\infty} \frac{1}{k^2+3k+2}$$

Observe that $k^2+3k+2=(k+1)(k+2)$, hence we seek real numbers A and B such that

$$\frac{1}{k^2+3k+2}=\frac{A}{k+1}+\frac{B}{k+2}.$$

Clearing the denominators yields an equation $A(k+2)+B(k+1)=1$. By plugging in $k=-1$, we find that $A=1$. By plugging in $k=-2$, we find that $B=-1$. Consequently, we have that

$$\sum_{k=0}^{\infty} \frac{1}{k^2+3k+2}=\sum_{k=0}^{\infty}\left(\frac{1}{k+1}-\frac{1}{k+2}\right)=\underbrace{\left(\frac{1}{1}-\frac{1}{2}\right)}_{\text{Plug in } k=0.}+\underbrace{\left(\frac{1}{2}-\frac{1}{3}\right)}_{\text{Plug in } k=1.}+\underbrace{\left(\frac{1}{3}-\frac{1}{4}\right)}_{\text{Plug in } k=2.}+\cdots=1.$$

Exercise 5.2.5. Compute the value of the infinite series $\displaystyle\sum_{k=1}^{\infty}\frac{k}{k^2+5k+3}$, or state that it diverges.

Given any real numbers c and r, the geometric sequence cr^n gives rise to the **geometric series**

$$\sum_{k=0}^{\infty} cr^k.$$

Like with Proposition 5.1.14 on geometric sequences, we can completely classify the convergence geometric series based on the common ratio r: in fact, as the following illustrates, the sequence of partial sums of a geometric sequence is the sum of a constant sequence and a geometric sequence.

Proposition 5.2.6 (Convergence of Geometric Series)**.** *Given any real number c and any nonzero real number r, the geometric series below obeys the following rule for convergence.*

$$\sum_{k=0}^{\infty} cr^k=\begin{cases}\dfrac{c}{1-r} & \text{if } -1<r<1\\ \text{diverges} & \text{if } r\geq 1 \text{ or } r\leq -1\end{cases}$$

Proof. Certainly, if $r=1$, then by the second property of Proposition 2.2.4, we have that

$$\sum_{k=0}^{\infty} cr^k=\sum_{k=0}^{\infty} c=\lim_{n\to\infty}\sum_{k=0}^{n} c=\lim_{n\to\infty} nc=c\Big(\lim_{n\to\infty} n\Big)=\infty,$$

and the geometric series diverges to ∞. On the other hand, if $r\neq 1$, then the binomial $1-r^{n+1}$ is divisible by $1-r$ because $(1+r+\cdots+r^n)(1-r)=1-r^{n+1}$. Consequently, we find that

$$\sum_{k=0}^{n} cr^k=c+cr+cr^2+\cdots+cr^n=c(1+r+r^2+\cdots+r^n)=\frac{c(1-r^{n+1})}{1-r}.$$

By definition of an infinite series as the limit of a finite sum, we conclude that

$$\sum_{k=0}^{\infty} cr^k = \lim_{n\to\infty} \sum_{k=0}^{n} cr^k = \lim_{n\to\infty} \frac{c(1-r^{n+1})}{1-r} = \lim_{n\to\infty} \left(\frac{c}{1-r} - \frac{cr^{n+1}}{1-r} \right) = \frac{c}{1-r} - \frac{1}{1-r} \left(\lim_{n\to\infty} cr^{n+1} \right)$$

depends on the convergence of the geometric sequence cr^{n+1}. By Proposition 5.1.14, the sequence cr^{n+1} diverges if $r > 1$ or $r < -1$, hence the geometric series diverges in either of these cases. □

Proposition 5.2.7 (Formula for Convergent Geometric Series)**.** *Given any real numbers c and r such that $-1 < r < 1$, the following formula gives the value of the convergent geometric series.*

$$\sum_{n=k}^{\infty} cr^n = \frac{cr^k}{1-r}$$

Proof. By Proposition 5.2.6 and its proof, we have that

$$\sum_{n=k}^{\infty} cr^n = \sum_{n=0}^{\infty} cr^n - \sum_{n=0}^{k-1} cr^n = \frac{c}{1-r} - \frac{c(1-r^k)}{1-r} = \frac{cr^k}{1-r}.$$ □

Example 5.2.8. Consider the following infinite series.

$$\sum_{n=0}^{\infty} \frac{1}{2^n}$$

By rewriting the general termas $\frac{1}{2^n} = \left(\frac{1}{2}\right)^n$, we find ourselves dealing with a geometric series with $c = 1$ and $r = \frac{1}{2}$. By the Formula for Convergent Geometric Series, we conclude that

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = \frac{1}{1-\frac{1}{2}} = 2.$$

Exercise 5.2.9. Compute the value of the infinite series $\displaystyle\sum_{n=0}^{\infty} \left(-\frac{2}{3}\right)^n$, or state that it diverges.

Example 5.2.10. Consider the following infinite series.

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{3^{n+1}}$$

By rewriting the general term as $\frac{(-1)^n}{3^{n+1}} = \frac{(-1)^n}{3 \cdot 3^n} = \frac{1}{3} \cdot \frac{(-1)^n}{3^n} = \frac{1}{3}\left(-\frac{1}{3}\right)^n$, we obtain a geometric series with $c = \frac{1}{3}$ and $r = -\frac{1}{3}$. By the Formula for Convergent Geometric Series, we conclude that

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{3^{n+1}} = \sum_{n=1}^{\infty} \frac{1}{3}\left(-\frac{1}{3}\right)^n = \frac{\frac{1}{3}\left(-\frac{1}{3}\right)}{1+\frac{1}{3}} = -\frac{1}{12}.$$

Exercise 5.2.11. Compute the value of the infinite series $\displaystyle\sum_{n=2}^{\infty} \frac{35}{3^{2n}}$, or state that it diverges.

Exercise 5.2.12. Compute the value of the infinite series $\displaystyle\sum_{n=1}^{\infty}(\ln e^{\pi})^n$, or state that it diverges.

Later in the course, we will consider infinite series as the discrete analog of improper integrals. Like with convergent improper integrals, there are nice linearity properties for convergent series.

Proposition 5.2.13 (Linearity of Convergent Series)**.** *Given any convergent series $\sum a_n$ and $\sum b_n$ and any real number C, the following properties hold.*

(i.) (**Additive Property of Convergent Series**) $\sum(a_n \pm b_n) = \sum a_n \pm \sum b_n$

(ii.) (**Distributive Property of Convergent Series**) $\sum Ca_n = C(\sum a_n)$

Particularly, any linear combination of convergent series is a convergent series.

We have thus far determined when telescoping and geometric series are convergent. Conversely, we can determine when an infinite series is divergent by inspecting its general term a_n.

Theorem 5.2.14 (nth Term Divergence Test)**.** *Consider any infinite sum $\sum_{k=m}^{\infty} a_k$ of real numbers a_k beginning at any integer m. If $\lim\limits_{k\to\infty} a_k \neq 0$, then $\sum_{k=m}^{\infty} a_k$ diverges. Put another way, in order for an infinite series to converge, the sequence of its general terms must converge to 0.*

Proof. Observe that the sequence of partial sums of $\sum_{k=m}^{\infty} a_k$ is given by

$$s_n = \sum_{k=m}^{n} a_k = a_m + a_{m+1} + \cdots + a_{n-1} + a_n.$$

Consequently, the relationship between the consecutive terms in the sequence of partial sums is

$$s_n = a_m + a_{m+1} + \cdots + a_{n-1} + a_n = s_{n-1} + a_n$$

so that $a_n = s_n - s_{n-1}$. Observe that if $\sum_{k=m}^{\infty} a_k$ converges, then we must have that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty}(s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_n = 0$$

by the first part of the Limit Laws for Sequences because s_n must be a convergent sequence. □

On first glance, it might appear that the above proof did not actually establish what we intended: we proved that if $\sum_{k=m}^{\infty} a_k$ converges, then $\lim\limits_{n\to\infty} a_n = 0$. We refer to this as a **proof by contrapositive**.

Example 5.2.15. Consider the following infinite series.

$$\sum_{n=0}^{\infty} \frac{n}{\sqrt{n^2+1}}$$

Before we do anything with this infinite series, we must check if it diverges by computing the limit of its general term. By the nth Term Divergence Test, this infinite series diverges because

$$\lim_{n\to\infty} \frac{n}{\sqrt{n^2+1}} = \lim_{n\to\infty} \frac{n}{\sqrt{n^2\left(1+\frac{1}{n^2}\right)}} = \lim_{n\to\infty} \frac{n}{n\sqrt{1+\frac{1}{n^2}}} = \lim_{n\to\infty} \frac{1}{\sqrt{1+\frac{1}{n^2}}} = 1.$$

Example 5.2.16. Consider the following infinite series.

$$\sum_{n=0}^{\infty}(-1)^n$$

Even though it might appear naïvely that the infinite series could be telescoping, we can rule out this possibility by the nth Term Divergence Test: in fact, the series diverges because $(-1)^n$ diverges.

Exercise 5.2.17. Compute the value of the series $\displaystyle\sum_{n=1}^{\infty}\frac{n!}{10^n}$, or state that it diverges.

Exercise 5.2.18. Compute the value of the series $\displaystyle\sum_{n=7}^{\infty}\frac{n^3+n^2+n+1}{n^3-n^2+n-1}$, or state that it diverges.

Caution. Often, upon first learning the nth Term Divergence Test, students can easily become mixed up in the logic of what exactly the theorem guarantees. Plainly, the theorem says that

(a.) if the sequence a_n does not converge to 0, then the infinite series $\sum_{k=m}^{n} a_k$ diverges, and

(b.) if the infinite series $\sum_{k=m}^{\infty} a_k$ converges, then the sequence a_n converges to 0.

Consequently, using the nth Term Divergence Test, we are able to decipher when a series *diverges*; however, the drawback is that we cannot tell that a series *converges* by this test.

Proposition 5.2.19 (Converse of the Divergence Test Is False)**.** *There exists an infinite sequence of real numbers a_n such that $\lim_{n\to\infty} a_n = 0$ but the infinite series $\sum_{k=m}^{\infty} a_k$ diverges.*

Proof. Observe that the following infinite series diverges. Explicitly, we have that

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{n} &= \frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\cdots \\ &> \frac{1}{1}+\frac{1}{2}+\frac{1}{4}+\frac{1}{4}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\cdots \\ &= \frac{1}{1}+\frac{1}{2}+2\cdot\frac{1}{4}+4\cdot\frac{1}{8}+\cdots \\ &= \frac{1}{1}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\cdots = \frac{1}{1}+\sum_{n=1}^{\infty}\frac{1}{2}.\end{aligned}$$

Considering that the latter is a divergent geometric series, the series in question diverges; however, it is clear that the sequence of general terms $a_n = 1/n$ of the series converges to 0. □

Considering that this is our prototypical counterexample to the converse of the nth Term Divergence Test, we refer to the following divergent infinite series as the **harmonic series**.

$$\sum_{n=1}^{\infty}\frac{1}{n}$$

Crucially, the harmonic series will soon provide a useful point of comparison by which we can measure the divergence of other infinite series. We illustrate with a couple examples.

Example 5.2.20. Consider the following infinite series.

$$\sum_{n=1}^{\infty} \frac{n}{n^2+n}$$

Considering that $n \geq 1$, we may cancel a factor of n from the numerator and denominator to obtain

$$\sum_{n=1}^{\infty} \frac{n}{n^2+n} = \sum_{n=1}^{\infty} \frac{1}{n+1}.$$

Like with definite integrals, we may perform a substitution to change the **index of summation**. Explicitly, if $n = k-1$, then $k = n+1$ and $k = 2$ if $n = 1$. Changing the index of summation yields

$$\sum_{n=1}^{\infty} \frac{1}{n+1} = \sum_{k=2}^{\infty} \frac{1}{k} = -1 + \sum_{k=1}^{\infty} \frac{1}{k}.$$

Consequently, this infinite series differs from the divergent harmonic series by 1, hence it diverges.

Exercise 5.2.21. Consider the following infinite series.

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

Observe that $\sqrt{x}$ is an increasing function because its derivative $\frac{1}{2\sqrt{x}}$ is positive for all real numbers $x > 0$. Consequently, its reciprocal is a decreasing function, so we have that

$$s_n = \sum_{k=1}^{n} \frac{1}{\sqrt{k}} > \sum_{k=1}^{n} \frac{1}{\sqrt{n}} = \frac{n}{\sqrt{n}} = \sqrt{n}$$

for all integers $n \geq 1$. Considering that $\lim_{n \to \infty} \sqrt{n} = \infty$, it follows that $\lim_{n \to \infty} s_n = \infty$. By the nth Term Divergence Test, we conclude that the infinite series in question diverges to infinity.

5.3 The Integral Test and the p-Series Test

Given any infinite sequence of real numbers a_n such that there exists a non-negative integer m for which $a_n \geq 0$ for all integers $n \geq m$, one immediate interpretation of the value of the corresponding infinite series is the total area of rectangles with width 1 and height a_k for each integer $k \geq m$.

$$\sum_{k=m}^{\infty} a_k = a_m + a_{m+1} + a_{m+2} + \cdots = \text{total area of rectangles of with width 1 and height } a_k$$

Bearing this in mind, one can imagine that the sequence of the partial sums of the series is increasing.

$$s_{n+1} = \sum_{k=m}^{n+1} a_k = a_m + a_{m+1} + \cdots + a_n + a_{n+1} = a_{n+1} + \sum_{k=m}^{n} a_k = a_{n+1} + s_n \geq s_n$$

Consequently, if can deduce that the sequence of partial sums is bounded above, then we may apply the Monotone Convergence Theorem to determine the convergence of the infinite series since

$$\sum_{k=m}^{\infty} a_k = \lim_{n \to \infty} \sum_{k=m}^{n} a_k = \lim_{n \to \infty} s_n.$$

Proposition 5.3.1 (Convergence of Infinite Series with Non-Negative Terms)**.** *Consider any infinite sequence of real numbers a_n such that $a_n \geq 0$ for all integers $n \geq m$ for some integer m.*

(a.) *If the sequence $s_n = \sum_{k=m}^{n} a_k$ is bounded above, then the infinite series $\sum_{k=m}^{\infty} a_k$ converges.*

(b.) *If the sequence $s_n = \sum_{k=m}^{n} a_k$ is not bounded above, then the infinite series $\sum_{k=m}^{\infty} a_k$ diverges.*

Often, it is difficult in practice to find a closed form for the partial sums of a series, hence it is difficult to determine the an upper bound for Proposition 5.3.1. Luckily, we can say more.

Theorem 5.3.2 (Integral Test)**.** *Consider any infinite sequence of real numbers a_n such that for some integer m, we have that $a_n = f(n)$ for some function $f(x)$ with the following properties.*

(i.) *$f(x)$ is non-negative for all real numbers $x \geq m$.*

(ii.) *$f(x)$ is continuous for all real numbers $x \geq m$.*

(iii.) *$f(x)$ is decreasing for all real numbers $x \geq m$.*

Given that each of the above conditions holds, the convergence of $\sum_{k=m}^{\infty} a_m$ is determined as follows.

(a.) *If the improper integral $\int_m^\infty f(x)\,dx$ converges, then $\sum_{k=m}^{\infty} a_k$ converges, and*

(b.) *If the improper integral $\int_m^\infty f(x)\,dx$ diverges, then $\sum_{k=m}^{\infty} a_k$ diverges.*

Proof. By the exposition beginning this section, the series represents the total area of rectangles with width 1 and height $a_n = f(n)$ for all integers $n \geq m$. Considering that $f(x)$ is decreasing, it follows that the right-endpoint Riemann approximation of the area bounded by the curve $f(x)$ from $x = m$ to $x = n$ is an underestimate for all real numbers $n \geq m$. Put another way, we have that

$$\sum_{k=m}^{n} a_k = a_m + a_{m+1} + \cdots + a_n \leq \int_m^n f(x)\,dx$$

for all integers $n \geq m$. By taking the limit as n tends to infinity, we conclude that

$$\sum_{k=m}^{\infty} a_k = \lim_{n\to\infty} \sum_{k=m}^{n} a_k \leq \lim_{n\to\infty} \int_m^n f(x)\,dx = \int_m^\infty f(x)\,dx.$$

Consequently, if the improper integral on the right-hand side converges, then the infinite series converges, as well. Conversely, the left-endpoint Riemann approximation of the area bounded by the curve $f(x)$ from $x = m$ to $x = n$ is an overestimate for all real numbers $n \geq m$ so that

$$\sum_{k=m}^{\infty} a_k = \lim_{n\to\infty} \sum_{k=m}^{n} a_k \geq \lim_{n\to\infty} \int_m^n f(x)\,dx = \int_m^\infty f(x)\,dx. \qquad \square$$

Exercise 5.3.3. Use the Integral Test to prove that $\displaystyle\sum_{n=m}^{\infty} \frac{1}{n}$ diverges for any positive integer m.

Solution. Before we are able to use the Integral Test, we must verify the following properties.

(i.) Observe that $f(x) = \dfrac{1}{x}$ is positive for all real numbers $x \geq m \geq 1$.

(ii.) Observe that $f(x) = \dfrac{1}{x}$ is continuous for all real numbers $x \geq m \geq 1$.

(iii.) By the Criterion for Monotonicity, $f(x) = \dfrac{1}{x}$ is decreasing for all real numbers $x \geq m$ because

$$f'(x) = -\frac{1}{x^2} < 0$$

for all real numbers $x \geq 1$ since x^2 is positive for all real numbers $x \geq m \geq 1$.

Consequently, the Integral Test asserts that the harmonic series diverges because

$$\int_m^{\infty} \frac{1}{x}\,dx = \lim_{b \to \infty} \int_m^b \frac{1}{x}\,dx = \lim_{b \to \infty} [\ln(x)]_m^b = \lim_{b \to \infty} [\ln(b) - \ln(m)] = \infty. \qquad \diamond$$

Exercise 5.3.4. Use the Integral Test to determine the convergence of the series $\displaystyle\sum_{n=0}^{\infty} \frac{1}{n^2+1}$.

Solution. Before we are able to use the Integral Test, we must verify the following properties.

(i.) Observe that $f(x) = \dfrac{1}{x^2+1}$ is positive for all real numbers $x \geq 0$.

(ii.) Observe that $f(x) = \dfrac{1}{x^2+1}$ is continuous for all real numbers $x \geq 0$.

(iii.) By the Criterion for Monotonicity, $f(x) = \dfrac{1}{x^2}$ is decreasing for all real numbers $x \geq 1$ because

$$f'(x) = -\frac{2x}{(x^2+1)^2} < 0$$

for all real numbers $x \geq 1$ since $(x^2+1)^2 > 0$ and $-2x < 0$ for all real numbers $x \geq 1$.

Consequently, the Integral Test asserts that the infinite series in question converges because

$$\int_1^{\infty} \frac{1}{x^2+1}\,dx = \lim_{b \to \infty} \int_1^b \frac{1}{x^2+1}\,dx = \lim_{b \to \infty} [\arctan(x)]_1^b = \lim_{b \to \infty} \left[\arctan(b) - \frac{\pi}{4}\right] = \frac{\pi}{4}. \qquad \diamond$$

Exercise 5.3.5. Use the Integral Test to determine the convergence of the series $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n\ln(n)}$.

Solution. Before we are able to use the Integral Test, we must verify the following properties.

(i.) Observe that $f(x) = \dfrac{1}{x\ln(x)}$ is positive for all real numbers $x \geq 2$.

(ii.) Observe that $f(x) = \dfrac{1}{x\ln(x)}$ is continuous for all real numbers $x \geq 2$.

(iii.) By the Criterion for Monotonicity, $f(x) = \dfrac{1}{x\ln(x)}$ is decreasing for all real numbers $x \geq 2$:

$$f'(x) = \frac{1 - \ln(x)}{[x\ln(x)]^2} < 0$$

for all real numbers $x \geq 2$ since $[x\ln(x)]^2 > 0$ and $1 - \ln(x) < 0$ for all real numbers $x \geq 2$.

Consequently, the Integral Test asserts that the infinite series in question diverges because

$$\int_2^\infty \frac{1}{x\ln(x)}\,dx = \lim_{b\to\infty}\int_2^b \frac{1}{x\ln(x)}\,dx = \lim_{b\to\infty}\int_{\ln(2)}^\infty \frac{1}{u}\,du = \lim_{b\to\infty}\,[\ln(u)]_{\ln(2)}^b = \infty.$$

Explicitly, we have used the substitution $u = \ln(x)$ with $du = \frac{1}{x}\,dx$ to compute the integral. $\diamond$

Exercise 5.3.6. Use the Integral Test to determine the convergence of the series $\displaystyle\sum_{n=2}^\infty \frac{\ln(n)}{n}$.

Exercise 5.3.7. Use the Integral Test to determine the convergence of the series $\displaystyle\sum_{n=2}^\infty \frac{1}{n\sqrt{n^2-1}}$.

Solution. Before we are able to use the Integral Test, we must verify the following properties.

(i.) Observe that $f(x) = \dfrac{1}{x\sqrt{x^2-1}}$ is positive for all real numbers $x \geq 2$.

(ii.) Observe that $f(x) = \dfrac{1}{x\sqrt{x^2-1}}$ is continuous for all real numbers $x \geq 2$.

(iii.) By the Criterion for Monotonicity, $f(x) = \dfrac{1}{x\sqrt{x^2+1}}$ is decreasing for all real numbers $x \geq 2$:

$$f'(x) = -\frac{\sqrt{x^2-1} + \frac{x^2}{\sqrt{x^2-1}}}{x^2(x^2-1)} < 0$$

for all real numbers $x \geq 2$ since $x^2(x^2-1) > 0$ and $\sqrt{x^2-1} + \dfrac{x^2}{\sqrt{x^2-1}} > 0$ for all $x \geq 2$.

Consequently, the Integral Test asserts that the infinite sequence in question converges because

$$\int_2^\infty \frac{1}{x\sqrt{x^2-1}}\,dx = \lim_{b\to\infty}\int_2^b \frac{1}{x\sqrt{x^2-1}}\,dx = \lim_{b\to\infty}\,[\operatorname{arcsec}(x)]_2^b = \lim_{b\to\infty}\left[\operatorname{arcsec}(b) - \frac{\pi}{3}\right] = \frac{\pi}{6}.$$ $\diamond$

Exercise 5.3.8. Use the Integral Test to determine the convergence of the series $\displaystyle\sum_{n=1}^\infty \frac{n}{n^2+1}$.

Geometric and telescoping series are not the only common infinite series. Given any nonzero real number p and any integer $m \geq 1$, any infinite series of the following form is called a **p-series**.

$$\sum_{n=m}^{\infty} \frac{1}{n^p}$$

One of the foremost consequences of the Integral Test is the following test.

Theorem 5.3.9 (p-Series Test)**.** *Consider any nonzero real number p.*

(a.) *If $p > 1$, then the p-series $\displaystyle\sum_{n=m}^{\infty} \frac{1}{n^p}$ converges.*

(b.) *If $p \leq 1$, then the p-series $\displaystyle\sum_{n=m}^{\infty} \frac{1}{n^p}$ diverges.*

Proof. Certainly, if $p = 0$, the series diverges by the nth Term Divergence Test. Likewise, if $p < 0$, then $-p > 0$ implies that $\frac{1}{n^p} = n^{-p}$ has positive exponent. Once again, the nth Term Divergence Test can be applied to demonstrate that the infinite series diverges. Consequently, we may assume that $p > 0$. Observe that $f(x) = x^{-p}$ is positive, decreasing, and continuous for all real numbers $x \geq m$. Explicitly, we have that $f(x) = x^{-p}$ so that $f'(x) = -px^{-p-1} < 0$ for all real numbers $x \geq m$. By the Integral Test, it suffices to determine when the following improper integral converges.

$$\int_m^{\infty} \frac{1}{x^p}\,dx = \lim_{b \to \infty} \int_m^b x^{-p}\,dx = \begin{cases} \displaystyle\lim_{b \to \infty} \left[\frac{x^{1-p}}{1-p}\right]_m^b & \text{if } p \neq 1 \\ \displaystyle\lim_{b \to \infty} \left[\ln(x)\right]_m^b & \text{if } p = 1 \end{cases}$$

Consequently, if $p > 1$, then $1 - p < 0$ so that the limit converges. On the other hand, if $p < 1$, then $1 - p > 0$ so the limit diverges to infinity. Certainly, the limit of $\ln(x)$ diverges to infinity. □

Exercise 5.3.10. Use the p-Series Test to determine the convergence of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt[5]{n^7}}$.

Solution. Observe that $\sqrt[5]{n^7} = n^{7/5}$, hence this is a convergent p-series with $p = \frac{7}{5}$. ⋄

Exercise 5.3.11. Use the p-Series Test to determine the convergence of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^{2.7}}$.

Exercise 5.3.12. Use the p-Series Test to determine the convergence of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt[7]{n^5}}$.

Solution. Observe that $\sqrt[7]{n^5} = n^{5/7}$, hence this is a divergent p-series with $p = \frac{5}{7}$. ⋄

Exercise 5.3.13. Use the p-Series Test to determine the convergence of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^{0.31415}}$.

Of course, the p-Series Test only applies to infinite series of reciprocals of power functions, so it is not immediately applicable to determine the convergence of series the likes of

$$\sum_{n=1}^{\infty} \frac{n^2}{n^5+1}.$$

Even more, we would not endeavor to use the Integral Test on this series, either, because the antiderivative of $f(x) = \dfrac{x^2}{x^5+1}$ is absolutely horrendous. But rest assured, we are not out of luck!

5.4 Comparison Tests for Series

By the p-Series Test of the previous section, the convergence of infinite p-series is completely classified. Consequently, for any infinite series whose general term "looks like" a p-series, one might naturally suspect that it is intuitively possible to deduce the convergence of the series based on the convergence of attendant p-series. Explicitly, consider the infinite series

$$\sum_{n=1}^{\infty} \frac{n^2}{n^5+1}$$

from the previous section. Certainly, for all sufficiently large integers n, the sequence n^5+1 behaves similarly to the sequence n^5. Consequently, for all sufficiently large integers n, we have that

$$\frac{n^2}{n^5+1} \sim \frac{n^2}{n^5} \sim \frac{1}{n^3}.$$

We introduce the symbol $\sim$ to express that two infinite sequences of real numbers a_n and b_n are **asymptotically equivalent**, i.e., we have that $a_n \sim b_n$ if and only if

$$\lim_{n \to \infty} \frac{a_n}{b_n} = 1.$$

Ultimately, the upshot is that if the infinite sequences of real numbers a_n and b_n are asymptotically equivalent, then the infinite series $\sum a_n$ converges if and only if the infinite series $\sum b_n$ converges.

Even if the sequences a_n and b_n are not asymptotically equivalent, it is possible to use a similar intuition to deduce the convergence of the infinite series $\sum a_n$ from the convergence of infinite series $\sum b_n$ (and vice-versa). We illustrate and prove this fact in the following series test.

Theorem 5.4.1 (Direct Comparison Test)**.** *Consider any pair of infinite sequences of real numbers a_n and b_n such that there exists an integer m for which $0 \leq a_n \leq b_n$ for all integers $n \geq m$.*

(a.) *If $\displaystyle\sum_{n=m}^{\infty} b_n$ converges, then $\displaystyle\sum_{n=m}^{\infty} a_n$ converges.*

(b.) *If $\displaystyle\sum_{n=m}^{\infty} a_n$ diverges, then $\displaystyle\sum_{n=m}^{\infty} b_n$ diverges.*

Proof. By hypothesis that $0 \leq a_n \leq b_n$ for all integers $n \geq m$, we have that

$$\sum_{k=m}^{n} a_k = a_m + a_{m+1} + \cdots + a_n \leq b_m + b_{m+1} + \cdots + b_n = \sum_{k=m}^{n} b_k$$

for all integers $n \geq m$. Consequently, if the infinite series of b_n converges, then its sequence s_n of partial sums must be bounded above by the Monotone Convergence Theorem: indeed, the sequence

$$s_n = \sum_{k=m}^{n} b_k$$

is increasing because $b_n \geq 0$ for all integers $n \geq m$, and it converges. We conclude that the sequence

$$t_n = \sum_{k=m}^{n} a_k$$

of partial sums of the infinite series of a_n is bounded above and increasing because $a_n \geq 0$ for all integers $n \geq m$. Consequently, the Monotone Convergence Theorem guarantees that the infinite series of a_n converges. Conversely, the statement (b.) is the contrapositive of the statement (a.). □

Exercise 5.4.2. Use the Direct Comparison Test to determine the convergence of $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{n^5+1}$.

Solution. Like we mentioned at the beginning of this section, if we squint our eyes a bit, the general term of this infinite series is essentially a convergent p-series with $p = 3$. Consequently, we imagine it might be possible to directly compare with such a convergent infinite series. Observe that

$$n^5 + 1 \geq n^5 \text{ for all integers } n \geq 1 \text{ yields that}$$

$$\frac{1}{n^5+1} \leq \frac{1}{n^5} \text{ so that}$$

$$\frac{n^2}{n^5+1} \leq \frac{n^2}{n^5} \text{ and}$$

$$\frac{n^2}{n^5+1} \leq \frac{1}{n^3} \text{ for all integers } n \geq 1 \text{ and}$$

$$\sum_{n=1}^{\infty} \frac{n^2}{n^5+1} \leq \sum_{n=1}^{\infty} \frac{1}{n^3}.$$

Considering that $\displaystyle\sum_{n=1} \frac{1}{n^3}$ converges by the p-Series Test, we conclude that $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{n^5+1}$ converges by the Direct Comparison Test since it is bounded above by a convergent series. ◇

Exercise 5.4.3. Use the Direct Comparison Test to determine the convergence of $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{n^3+n^2+3}$.

Exercise 5.4.4. Use the Direct Comparison Test to determine the convergence of $\displaystyle\sum_{n=2}^{\infty} \frac{1}{\sqrt[7]{n^5-1}}$.

Solution. Once again, if we squint our eyes at the denominator, it resembles the divergent p-series with $p=\frac{5}{7}$. Consequently, we intuit that this infinite series diverges. By the Direct Comparison Test, we must find a positive integer m sufficiently large such that for all integers $n \geq m$, we have

$$\frac{1}{n^{5/7}} \leq \frac{1}{\sqrt[7]{n^5-1}}.$$

By working backwards from this inequality, we can determine the positive integer m as follows.

$$\frac{1}{n^{5/7}} \leq \frac{1}{\sqrt[7]{n^5-1}} \text{ yields that}$$

$$\frac{1}{n^5} \leq \frac{1}{n^5-1} \text{ by raising both sides to the seventh power and}$$

$$n^5 \geq n^5-1 \text{ for all integers } n \geq 2.$$

Consequently, the above analysis reveals that the desired inequality of infinite series holds.

$$\sum_{n=2}^{\infty} \frac{1}{n^{5/7}} \leq \sum_{n=2}^{\infty} \frac{1}{\sqrt[7]{n^5-1}}$$

By the Direct Comparison Test and the p-Series Test, we conclude that the series diverges. ◇

Exercise 5.4.5. Use the Direct Comparison Test to determine the convergence of $\displaystyle\sum_{n=0}^{\infty} \frac{1}{\sqrt[5]{n^7-1}}$.

Examples 5.4.2 and 5.4.4 reveal that the Direct Comparison Test can sometimes be employed to test for convergence of infinite series of any sequence that is the reciprocal of a power function composed with a polynomial; however, this application is not always the most straightforward.

Example 5.4.6. Consider the following infinite series.

$$\sum_{n=0}^{\infty} \frac{1}{\sqrt[7]{n^5+1}}.$$

Bearing in mind Example 5.4.4, our knee-jerk reaction is to directly compare this infinite series with the divergent p-series with $p=\frac{5}{7}$; however, this is not feasible because if we attempt to bound the series in question below by the aforementioned divergent p-series, we find that the inequalities here do not actually work out. Explicitly, we would begin by observing that

$$n^5 + 1 \geq n^5 \text{ for all integers } n \geq 0 \text{ implies that}$$

$$\frac{1}{n^5 + 1} \leq \frac{1}{n^5} \text{ for all integers } n \geq 1 \text{ so that}$$

$$\frac{1}{\sqrt[7]{n^5 + 1}} \leq \frac{1}{n^{5/7}} \text{ for all integers } n \geq 1 \text{ and}$$

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt[7]{n^5 + 1}} \leq \sum_{n=1}^{\infty} \frac{1}{n^{5/7}}.$$

But the divergence of the infinite series on the right-hand side of the infinite series has no bearing on the divergence of the infinite series on the left-hand side: indeed, we have that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n} \leq \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

because $n^2 \geq n \geq \sqrt{n}$ for all integers $n \geq 1$; however, the p-series with $p = 2$ converges and the harmonic series diverges. Consequently, the Direct Comparison Test fails in the case that a series is bounded above by a divergent series (or bounded below by a convergent series). Even still, because

$$\frac{1}{\sqrt[7]{n^5 + 1}} \sim \frac{1}{n^5} \text{ since we have that } \lim_{n \to \infty} \frac{\sqrt[7]{n^5 + 1}}{n^5} = 1,$$

we suspect that the infinite series in question diverges. Considering the end behavior of polynomials of odd degree, the clever reader might be able to piece together that $n^7 - n^5 - 1 \geq 0$ for all integers $n \geq m$ for some sufficiently large integer $m \geq 2$, from which it follows that

$$n^7 \geq n^5 + 1 \text{ so that } \frac{1}{n^7} \leq \frac{1}{n^5 + 1} \text{ yields that } \frac{1}{n} \leq \frac{1}{\sqrt[7]{n^5 + 1}} \text{ for all integers } n \geq m.$$

By the Direct Comparison Test, we conclude that the series in question diverges since

$$\sum_{n=m}^{\infty} \frac{1}{n} \leq \sum_{n=m}^{\infty} \frac{1}{\sqrt[7]{n^5 + 1}}.$$

Example 5.4.6 illustrates that even if there is an asymptotic equivalence between two sequences a_n and b_n, the Direct Comparison Test may fail if the inequality $0 \leq a_n \leq b_n$ does not hold. Even worse, it could be quite ad hoc and unintuitive to find another sequence for which the required inequality holds. But fortunately for us, the Direct Comparison Test does not use the full strength of the asymptotic equivalence of sequences. Our next series test makes good on this.

Theorem 5.4.7 (Limit Comparison Test)**.** *Consider any pair of infinite sequences of real numbers a_n and b_n such that there exists an integer m for which $a_n \geq 0$ and $b_n \geq 0$ for all integers $n \geq m$. Consider the following (possibly infinite) limit of the ratios of the sequences.*

$$L = \lim_{n \to \infty} \frac{a_n}{b_n}$$

(a.) *If* $L = 0$ *and* $\sum_{n=m}^{\infty} b_n$ *converges, then* $\sum_{n=m}^{\infty} a_n$ *converges.*

(b.) *If* $L > 0$ *and* L *is finite, then* $\sum_{n=m}^{\infty} a_n$ *converges if and only if* $\sum_{n=m}^{\infty} b_n$ *converges.*

(c.) *If* $L = \infty$ *and* $\sum_{n=m}^{\infty} b_n$ *diverges, then* $\sum_{n=m}^{\infty} a_n$ *diverges.*

Proof. We will assume first that $L \geq 0$ is finite. By definition of the limit, for all sufficiently large integers n, the ratio of a_n and b_n can be made as closed as desired to the value of L. Consequently, there exists a real number $\alpha > L$ such that for all sufficiently large integers n, we have that

$$0 \leq \frac{a_n}{b_n} \leq \alpha \text{ so that } 0 \leq a_n \leq \alpha b_n.$$

By the Direct Comparison Test, if the infinite series of b_n converges, then the infinite series of a_n converges; this proves statement (a.) and the "if" part of statement (b.). We will assume now that L is either positive ($L > 0$) or infinite ($L = \infty$). Either way, L is nonzero, hence the real number

$$K = \lim_{n \to \infty} \frac{b_n}{a_n} = \lim_{n \to \infty} \frac{1}{\frac{a_n}{b_n}} = \frac{1}{\lim_{n \to \infty} \frac{a_n}{b_n}} = \frac{1}{L}$$

is well-defined and satisfies that $K \geq 0$. By the previous paragraph, by simply reversing the roles of the infinite sequences a_n and b_n, we find that statement (c.) and the "only if" part of (b.) hold. $\square$

Unfortunately, though this proof is quite clever, it obscures the intuition by which we were initially lead to the Limit Comparison Test. We rephrase the three statements as follows.

(a.) If $L = 0$, then the terms of the infinite sequence b_n are eventually "significantly larger" than the terms of the infinite sequence a_n, hence by the Direct Comparison Test,

$$\text{if } \sum_{n=m}^{\infty} b_n \text{ converges, then } \sum_{n=m}^{\infty} a_n \text{ converges.}$$

(b.) If L is finite and $L > 0$, then the infinite sequences a_n and Lb_n are asymptotically equivalent:

$$\lim_{n \to \infty} \frac{a_n}{Lb_n} = \frac{1}{L}\left(\lim_{n \to \infty} \frac{a_n}{b_n}\right) = \frac{1}{L} \cdot L = 1.$$

Consequently, for all sufficiently large integers n, the infinite sequences of real numbers a_n and Lb_n can be made as close to each other in value as possible; thus, the infinite series

$$\sum_{n=m}^{\infty} a_n \text{ and } \sum_{n=m}^{\infty} Lb_n = L \sum_{n=m}^{\infty} b_n$$

can be made as close to each other in value as possible. We conclude that

$$\sum_{n=m}^{\infty} a_n \text{ converges if and only if } \sum_{n=m}^{\infty} b_n \text{ converges.}$$

(c.) If $L = \infty$, then the terms of the infinite sequence a_n are eventually "significantly larger" than the terms of the infinite sequence b_n, hence by the Direct Comparison Test,

$$\text{if } \sum_{n=m}^{\infty} a_n \text{ converges, then } \sum_{n=m}^{\infty} b_n \text{ converges.}$$

Exercise 5.4.8. Use the Limit Comparison Test to determine the convergence of $\displaystyle\sum_{n=0}^{\infty} \frac{1}{\sqrt[7]{n^5+1}}$.

Solution. By Example 5.4.4, we suspect it would be useful to compare with $n^{5/7}$. Observe that

$$\lim_{n\to\infty} \frac{n^{5/7}}{\sqrt[7]{n^5+1}} = \lim_{n\to\infty} \frac{n^{5/7}}{\sqrt[7]{n^5\left(1+\frac{1}{n^5}\right)}} = \lim_{n\to\infty} \frac{n^{5/7}}{n^{5/7}\sqrt[7]{1+\frac{1}{n^5}}} = \lim_{n\to\infty} \frac{1}{\sqrt[7]{1+\frac{1}{n^5}}} = 1.$$

By the Limit Comparison Test, the series in question diverges because the p-series $n^{-5/7}$ diverges. $\diamond$

Exercise 5.4.9. Use the Limit Comparison Test to determine the convergence of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt[2]{n^5+n}}$.

Exercise 5.4.10. Use the Limit Comparison Test to determine the convergence of $\displaystyle\sum_{n=0}^{\infty} \frac{n^3+n-1}{n^4-n^3+1}$.

Solution. By squinting our eyes, we suspect it would be useful to compare with n^{-1}. Observe that

$$\lim_{n\to\infty} \frac{\frac{n^3+n-1}{n^4-n^3+1}}{n^{-1}} = \lim_{n\to\infty} \frac{n^3+n-1}{n^{-1}(n^4-n^3+1)} = \lim_{n\to\infty} \frac{n^3+n-1}{n^3-n^2+\frac{1}{n}} = 1.$$

By the Limit Comparison Test, the series in question diverges because the p-series n^{-1} diverges. $\diamond$

Exercise 5.4.11. Use the Limit Comparison Test to determine the convergence of $\displaystyle\sum_{n=0}^{\infty} \frac{n^2+1}{n^4+16}$.

5.5 Alternating Series and Absolute Convergence

Our study of infinite series so far has culminated in the development of several tools with which to test convergence of many different types of series. Before we proceed, we summarize these results.

Summary 5.5.1 (Convergence Tests for Geometric Series and Series with Non-Negative Terms)**.**

(a.) Geometric series are of the form $\displaystyle\sum_{n=k}^{\infty} cr^n$ for some real numbers c and r. We have that

$$\sum_{n=k}^{\infty} cr^n = \begin{cases} \dfrac{cr^{k+1}}{1-r} & \text{if } -1 < r < 1 \text{ and} \\[2em] \text{diverges} & \text{if } r \geq 1 \text{ or } r \leq -1. \end{cases}$$

We refer to this loosely as the **Geometric Series Test**.

(b.) Given any infinite sequence of real numbers a_n such that there exists a positive integer m and a real function $f(x)$ with the properties that $a_n = f(n)$ for all integers $n \geq m$ and

- $f(x)$ is non-negative (i.e., $f(x) \geq 0$) for all real numbers $x \geq m$,
- $f(x)$ is continuous for all real numbers $x \geq m$, and
- $f(x)$ is decreasing (i.e., $f'(x) < 0$) for all real numbers $x \geq m$,

the Integral Test asserts that $\displaystyle\sum_{n=m}^{\infty} a_n$ converges if and only if $\displaystyle\int_m^{\infty} f(x)\,dx$ converges.

(c.) A p-series is of the form $\displaystyle\sum_{n=k}^{\infty} \frac{1}{n^p}$ for some positive integer k and real number p. We have that

$$\sum_{n=k}^{\infty} \frac{1}{n^p} \begin{cases} \text{converges} & \text{if } p > 1 \text{ and} \\ \text{diverges} & \text{if } p \leq 1 \end{cases}$$

by the p-Series Test. Generally, it is difficult to compute the value of a convergent p-series.

(d.) Given any infinite sequence of real numbers a_n such that there exists a positive integer m with $a_n \geq 0$ for all integers $n \geq m$, the Direct Comparison Test can be used to determine the convergence of an infinite series of a_n by directly comparing it to an infinite series of a sequence b_n such that either $0 \leq b_n \leq a_n$ or $0 \leq a_n \leq b_n$ for all integers $n \geq m$. Explicitly,

- if $0 \leq a_n \leq b_n$ for all integers $n \geq m$ and $\displaystyle\sum_{n=m}^{\infty} b_n$ converges, then $\displaystyle\sum_{n=m}^{\infty} a_n$ converges and
- if $0 \leq b_n \leq a_n$ for all integers $n \geq m$ and $\displaystyle\sum_{n=m}^{\infty} b_n$ diverges, then $\displaystyle\sum_{n=m}^{\infty} a_n$ diverges.

(e.) Given any infinite sequence of real numbers a_n such that there exists a positive integer m with $a_n \geq 0$ for all integers $n \geq m$, the Limit Comparison Test can be used to determine the convergence of an infinite series of a_n by considering the (possibly infinite) limit

$$L = \lim_{n \to \infty} \frac{a_n}{b_n}$$

for some infinite sequence of real numbers b_n such that $b_n \geq 0$ for all integers $n \geq m$. Explicitly,

- if $L = 0$ and $\displaystyle\sum_{n=m}^{\infty} b_n$ converges, then $\displaystyle\sum_{n=m}^{\infty} a_n$ converges;
- if $L > 0$ and L is finite, then $\displaystyle\sum_{n=m}^{\infty} a_n$ converges if and only if $\displaystyle\sum_{n=m}^{\infty} b_n$ converges; and
- if $L = \infty$ and $\displaystyle\sum_{n=m}^{\infty} b_n$ diverges, then $\displaystyle\sum_{n=m}^{\infty} a_n$ diverges.

Be sure to note that (other than for geometric series) the convergence tests that we have discussed so far pertain only to infinite series whose sequence of general terms possesses only finitely many negative terms. Curiously enough, the behavior of a general infinite series of a sequence a_n with infinitely many negative terms can stand in stark contrast to the infinite series of the sequence $|a_n|$ of the absolute value of the general term. We will soon see that the **alternating harmonic series**

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$$

converges in spite of the fact that the harmonic series (i.e., the p-series with $p = 1$) diverges!

Quite naturally, it is possible to obtain from any infinite series an infinite series with no negative terms: if a_n is the general term of the series, then we may simply consider infinite series of $|a_n|$. By doing this, for one, we allow ourselves the convenience of the Convergence Tests for Geometric Series and Series with Non-Negative Terms. Even more, we will say that the infinite series

$$\sum_{n=m}^{\infty} a_n$$

converges absolutely (or is **absolutely convergent**) if and only if the following series converges.

$$\sum_{n=m}^{\infty} |a_n|$$

Exercise 5.5.2. Prove that the infinite series $\displaystyle\sum_{n=1}^{\infty} (-1)^n n^{-\pi}$ converges absolutely.

Solution. Observe that in absolute value, we have that

$$\sum_{n=1}^{\infty} |(-1)^n n^{-\pi}| = \sum_{n=1}^{\infty} \frac{1}{n^{\pi}}$$

is a p-series with $p = \pi > 1$. Consequently, the series converges absolutely. ◇

Exercise 5.5.3. Prove that the infinite series $\displaystyle\sum_{n=0}^{\infty} (-1)^n (n^2+1)^{-1}$ converges absolutely.

Exercise 5.5.4. Prove that the infinite series $\displaystyle\sum_{n=1}^{\infty} (-1)^n n e^{-n^2}$ converges absolutely.

Solution. Observe that in absolute value, we have that

$$\sum_{n=1}^{\infty} \left|(-1)^n n e^{-n^2}\right| = \sum_{n=1}^{\infty} n e^{-n^2}.$$

Consider the continuous real function $f(x) = xe^{-x^2}$. By the Product Rule, we have that

$$f'(x) = e^{-x^2} + xe^{-x^2}(-2x) = e^{-x^2}(1 - 2x^2) < 0$$

because $1-2x^2 < 0$ and $e^{-x^2} > 0$ for all real numbers $x \geq 1$. Consequently, $f(x)$ is (i.) non-negative, (ii.) continuous, and (iii.) decreasing. By the Integral Test, it suffices to note that

$$\int_1^\infty xe^{-x^2}\,dx = \lim_{b\to\infty}\int_1^b xe^{-x^2}\,dx = \lim_{b\to\infty}\left[-\frac{1}{2}e^{-x^2}\right]_1^b = \lim_{b\to\infty}\left(\frac{1}{2}e^{-1} - \frac{1}{2}e^{-b^2}\right) = \frac{1}{2e}$$

converges, hence the series converges in absolute value, so it is absolutely convergent. ◇

Crucially, convergence in absolute value implies convergence of the original series.

Theorem 5.5.5 (Absolute Convergence Implies Convergence)**.** *Given any infinite sequence of real numbers a_n, if the infinite series $\sum_{n=m}^\infty |a_n|$ converges, then $\sum_{n=m}^\infty a_n$ converges.*

Proof. By definition of the absolute value function, we have that $-|a_n| \leq a_n \leq |a_n|$ for all integers n, hence we have that $0 \leq a_n + |a_n| \leq 2|a_n|$. By the Direct Comparison Test with

$$\sum_{n=m}^\infty (a_n + |a_n|) \leq \sum_{n=m}^\infty 2|a_n| = 2\sum_{n=m}^\infty |a_n|,$$

if the series on the right-hand side converges, then the series on the left-hand side converges. Even more, by the first property of Linearity of Convergent Series, we have that

$$\sum_{n=m}^\infty a_n = \sum_{n=m}^\infty (a_n + |a_n| - |a_n|) = \sum_{n=m}^\infty (a_n + |a_n|) - \sum_{n=m}^\infty |a_n|$$

converges because both of the infinite series in the difference converge. □

Exercise 5.5.6. Prove that the infinite series $\sum_{n=1}^\infty (-1)^n n^{-\pi}$ converges.

Proof. By Example 5.5.2, the series converges absolutely, so it converges. □

Exercise 5.5.7. Prove that the infinite series $\sum_{n=0}^\infty (-1)^n (n^2+1)^{-1}$ converges.

Exercise 5.5.8. Prove that the infinite series $\sum_{n=1}^\infty (-1)^n ne^{-n^2}$ converges.

Proof. By Example 5.5.4, the series converges absolutely, so it converges. □

Certainly, there exist infinite series that do not converge absolutely. Explicitly, for any divergent infinite series of a_n, the infinite series of $(-1)^n a_n$ cannot converge absolutely. Conversely, suppose that a_n is an infinite sequence of positive real numbers. We refer to the infinite series

$$\sum_{n=m}^\infty (-1)^n a_n$$

as an **alternating series** because its terms alternate in sign. Even more, we say that the infinite series of $(-1)^n a_n$ **converges conditionally** (or is **conditionally convergent**) if the alternating series of $(-1)^n a_n$ converges but the series of a_n diverges. Below, we provide a powerful theorem for testing convergence of alternating series. We invite the interested reader to note that the proof of the statements requires more care due to the alternating nature of its sequence of terms.

Theorem 5.5.9 (Alternating Series Test)**.** *Consider any sequence of real numbers a_n such that*

(i.) *a_n is non-negative for all integers $n \geq m$;*

(ii.) *a_n is decreasing for all integers $n \geq m$; and*

(iii.) *a_n converges to zero, i.e., $\lim_{n \to \infty} a_n = 0$.*

Given that each of the above conditions holds, the infinite series $\sum_{n=m}^{\infty} (-1)^n a_n$ converges.

Proof. Convergence of an infinite series depends explicitly upon the convergence of its sequence of partial sums, so we must prove that the following sequence of partial sums converges.

$$s_n = \sum_{k=m}^{n} (-1)^k a_k$$

Considering the alternating nature of the sequence $(-1)^k a_k$, we consider the cases that n is even and n is odd separately. We remind the reader that if n is even, then $n = 2\ell$ for some integer ℓ. Likewise, if n is odd, then $n = 2\ell + 1$ for some integer ℓ. Going forward, we will simply replace work with 2ℓ if n is even or $2\ell + 1$ if n is odd. Crucially, we have that $(-1)^{2\ell} = 1$ and $(-1)^{2\ell+1} = -1$. Even more, the ℓth term of the sequence $s_{2\ell}$ of even partial sums is followed by the $(\ell + 1)$th term $s_{2\ell+2}$ because $2(\ell + 1) = 2\ell + 2$. Likewise, the ℓth term of the sequence $s_{2\ell+1}$ of odd partial sums is followed by the $(\ell+1)$th term $s_{2\ell+3}$ because $2(\ell+1)+1 = 2\ell+3$. Bearing this in mind, we proceed.

(a.) Observe that for any integer ℓ such that $2\ell \geq m$, we have that

$$\begin{aligned} s_{2\ell+2} &= \sum_{k=m}^{2\ell+2} (-1)^k a_k \\ &= (-1)^m a_m + (-1)^{m+1} a_{m+1} + \cdots + (-1)^{2\ell} a_{2\ell} - a_{2\ell+1} + a_{2\ell+2} \\ &= a_{2\ell+2} - a_{2\ell+1} + \sum_{k=m}^{2\ell} (-1)^k a_k \\ &= a_{2\ell+2} - a_{2\ell+1} + s_{2\ell}. \end{aligned}$$

By assumption that a_n is decreasing, we have that $a_{2\ell+2} \leq a_{2\ell+1}$ so that $a_{2\ell+2} - a_{2\ell+1} \leq 0$ and

$$s_{2\ell+2} = a_{2\ell+2} - a_{2\ell+1} + s_{2\ell} \leq s_{2\ell}.$$

Consequently, the even partial sums of the infinite series form a decreasing sequence.

(b.) Conversely, for any integer ℓ such that $2\ell+1 \geq m$, we have that

$$\begin{aligned} s_{2\ell+3} &= \sum_{k=m}^{2\ell+3} (-1)^k a_k \\ &= (-1)^m a_m + (-1)^{m+1} a_{m+1} + \cdots + (-1)^{2\ell+1} a_{2\ell+1} + a_{2\ell+2} - a_{2\ell+3} \\ &= a_{2\ell+2} - a_{2\ell+3} + \sum_{k=m}^{2\ell+1} (-1)^k a_k \\ &= a_{2\ell+2} - a_{2\ell+3} + s_{2\ell+1}. \end{aligned}$$

By assumption that a_n is decreasing, we have that $a_{2\ell+2} \geq a_{2\ell+3}$ so that $a_{2\ell+2} - a_{2\ell+3} \geq 0$ and

$$s_{2\ell+3} = a_{2\ell+2} - a_{2\ell+3} + s_{2\ell+1} \geq s_{2\ell+1}.$$

Consequently, the odd partial sums of the infinite series form an increasing sequence.

By the Monotone Convergence Theorem, if $s_{2\ell}$ is bounded below and $s_{2\ell+1}$ is bounded above, then these sequences converge. Considering that a_n is non-negative for all integers $n \geq m$, we have that $s_{2\ell} - s_{2\ell+1} = -(-1)^{2\ell+1} a_{2\ell+1} = a_{2\ell+1} \geq 0$. Consequently, if m is even, then we have that

$$0 \leq a_m - a_{m+1} = s_{m+1} \leq s_{2\ell+1} \leq s_{2\ell} \leq s_m = a_m$$

for any integer ℓ such that $2\ell \geq m$. On the other hand, if m is odd, then we have that

$$0 \leq s_m \leq s_{2\ell+1} \leq s_{2\ell} \leq s_{m+1} = a_{m+1} - a_m$$

for any integer ℓ such that $2\ell + 1 \geq m$. Either way, this analysis illustrates that $s_{2\ell}$ is bounded below and $s_{2\ell+1}$ is bounded above, hence each of these sequences converges. Consequently, by the Limit Laws for Sequences and our assumption that $\lim_{n \to \infty} a_n = 0$, we conclude that

$$0 = \lim_{\ell \to \infty} a_{2\ell+1} = \lim_{n \to \infty} (s_{2\ell} - s_{2\ell+1}) = \lim_{\ell \to \infty} s_{2\ell} - \lim_{\ell \to \infty} s_{2\ell+1},$$

hence $s_{2\ell}$ and $s_{2\ell+1}$ converge to the same value L. Considering that $s_{2\ell+1} \leq s_n \leq s_{2\ell}$ for any integer ℓ such that $n = 2\ell$ or $n = 2\ell+1$, the Squeeze Theorem for Sequences yields that s_n converges. □

Corollary 5.5.10 (Estimation Formula for Alternating Series)**.** *Given any sequence of real numbers a_n that satisfies the hypotheses of the Alternating Series Test, the following inequality holds.*

$$\left| \sum_{k=m}^{\infty} (-1)^k a_k - \sum_{k=m}^{n} (-1)^k a_k \right| \leq a_{n+1}$$

Even more, we can approximate the value of $\sum_{n=m}^{\infty} (-1)^n a_n$ to any desired degree of accuracy.

Proof. Let L denote the value of the infinite series. Like before, we consider the following cases.

(a.) By the proof of the Alternating Series Test, if $n = 2\ell$ for some integer ℓ, then

$$\left|\sum_{k=m}^{\infty}(-1)^k a_k - \sum_{k=m}^{2\ell}(-1)^k a_k\right| = |L - s_{2\ell}| = s_{2\ell} - L \leq s_{2\ell} - s_{2\ell+1} = a_{2\ell+1}.$$

Explicitly, we have that $s_{2\ell+1} \leq L$ for all integers ℓ such that $2\ell + 1 \geq m$.

(b.) Likewise, by the proof of the Alternating Series Test, if $n = 2\ell + 1$ for some integer ℓ, then

$$\left|\sum_{k=m}^{\infty}(-1)^k a_k - \sum_{k=m}^{2\ell+1}(-1)^k a_k\right| = |L - s_{2\ell+1}| = L - s_{2\ell+1} \leq s_{2\ell+2} - s_{2\ell+1} = a_{2\ell+2}.$$

Explicitly, we have that $s_{2\ell+2} \geq L$ for all integers ℓ such that $2\ell + 2 \geq m$.

Considering that $\lim_{n\to\infty} a_n = 0$, the approximation grows more accurate as n grows larger. □

Example 5.5.11. Consider the alternating harmonic series

$$\sum_{n=1}^{\infty}\frac{(-1)^n}{n}.$$

By the Alternating Series Test, it suffices to note that the sequence $a_n = \dfrac{1}{n}$ satisfies that

(i.) a_n is non-negative for all integers $n \geq 1$;

(ii.) a_n is decreasing because $a_{n+1} = \dfrac{1}{n+1} \leq \dfrac{1}{n} = a_n$ for all integers $n \geq 1$; and

(iii.) a_n converges to zero, i.e., $\lim_{n\to\infty} a_n = 0$.

Consequently, the alternating harmonic series converges, as we suggested earlier in the section. Even more, by the Estimation Formula for Alternating Series, if we wish to approximate its value in a manner that is accurate to three decimal places, we seek to find an integer $n \geq 1$ such that

$$\left|\sum_{k=1}^{n}\frac{(-1)^k}{k} - \sum_{k=1}^{\infty}\frac{(-1)^k}{k}\right| \leq \frac{1}{n+1} \leq 0.0001.$$

Considering that $0.0001 = 10^{-4}$, it suffices to take $n \geq 10^4$. Using a computer algebra system, we find that the first three decimal places of the value of the alternating harmonic series are -0.693.

Exercise 5.5.12. Prove that the alternating series $\displaystyle\sum_{n=1}^{\infty}\frac{(-1)^n}{\sqrt{n}}$ converges; then, use the Estimation Formula for Alternating Series to approximate its value accurate to one decimal place.

Example 5.5.13. Consider the **alternating p-series**

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^p}.$$

We compute in this example all real numbers p such that the alternating p-series converges. Observe that for any real number $p > 0$, the differentiable power function $f(x) = x^{-p}$ is non-negative for all real numbers $x \geq 1$. Even more, we have that $-p < 0$ so that $f'(x) = -px^{p-1} < 0$ for all real numbers $x \geq 1$, hence $f(x)$ is decreasing for all real numbers $x \geq 1$. Last, we have that

$$\lim_{n \to \infty} \frac{1}{n^p} = \lim_{x \to \infty} \frac{1}{x^p} = 0.$$

By the Alternating Series Test, it follows that the alternating p-series converges for all real numbers $p > 0$. Conversely, if $p < 0$, then $-p > 0$ so that $f(x) = x^{-p}$ is increasing. Explicitly, we have that $\lim_{n \to \infty} n^{-p} \neq 0$, hence by the nth Term Divergence Test, the alternating p-series diverges.

Proposition 5.5.14 (Convergence of Alternating p-Series)**.** *Given any real number p, we have that*

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^p} = \begin{cases} \textit{converges} & \textit{if } p > 0 \textit{ and} \\ \textit{diverges} & \textit{if } p \leq 0. \end{cases}$$

Exercise 5.5.15. Prove that the alternating series $\displaystyle\sum_{n=2}^{\infty} \frac{(-1)^n \ln(n)}{n}$ converges.

Proof. By the Alternating Series Test, it suffices to prove that the non-negative function

$$f(x) = \frac{\ln(x)}{x}$$

is decreasing and converges to zero in the limit. By the Quotient Rule, we have that

$$f'(x) = \frac{1 - \ln(x)}{x^2} < 0$$

because $\ln(x) > 1$ and $x^2 > 0$ for all real numbers $x \geq 2$. By L'Hôpital's Rule, we have that

$$\lim_{x \to \infty} \frac{\ln(x)}{x} \stackrel{\text{L'H}}{=} \lim_{x \to \infty} \frac{1}{x} = 0. \qquad \square$$

Exercise 5.5.16. Prove that the alternating series $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n \sin(n)}{n}$ converges.

5.6 The Ratio Test

We have presented thus far many tests for determining the convergence of infinite series; however, we have made a distinction between series with positive and negative terms. Particularly, we cannot apply the Integral Test or either of the Comparison Tests to a series whose terms alternate in sign. On the other hand, we cannot apply the Alternating Series Test to a series with non-negative terms. Our last series test can be applied to any series regardless of the sign of the general term.

Theorem 5.6.1 (Ratio Test). *Given any infinite sequence of real numbers a_n, consider the limit*

$$L = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|.$$

(a.) *If $L < 1$, then the infinite series $\sum_{n=m}^{\infty} a_n$ converges absolutely.*

(b.) *If $L > 1$, then the infinite series $\sum_{n=m}^{\infty} a_n$ diverges.*

(c.) *If $L = 1$, then the Ratio Test is inconclusive: the infinite series $\sum_{n=m}^{\infty} a_n$ may diverge.*

Proof. We can easily dispense of the case (c.) that $L = 1$ by considering the infinite series

$$\sum_{n=1}^{\infty} \frac{1}{n} \text{ and } \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Both of these infinite series satisfy that $L = 1$; however, the former is the famously divergent harmonic series, and the latter converges by the p-Series Test since it is a p-series with $p = 2$.

Likewise, if $L = \infty$, then for all sufficiently large integers n, we have that $|a_{n+1}| > |a_n|$, hence the sequence a_n is eventually increasing without bound in absolute value. By the nth Term Divergence Test, we conclude that the infinite series diverges (because its sequence of terms diverges).

We may assume that L is finite. By definition of the limit L, given any real number $\varepsilon > 0$, there exists a positive integer m sufficiently large such that for all integers $n \geq m$, we have that

$$-\varepsilon < \left| \frac{a_{n+1}}{a_n} \right| - L < \varepsilon.$$

By simplifying this inequality, for all integers $n \geq m$, it follows that

$$L - \varepsilon < \left| \frac{a_{n+1}}{a_n} \right| < L + \varepsilon.$$

Consider the real numbers $r = L + \varepsilon$ and $s = L - \varepsilon$. We proceed by cases as follows.

(a.) If $L < 1$, we can ensure that $r < 1$ by taking ε to be sufficiently small. Observe that

$$\begin{aligned} |a_{m+1}| &< |a_M| r, \\ |a_{m+2}| &< |a_{m+1}| r < |a_M| r^2, \\ |a_{m+3}| &< |a_{m+2}| r < |a_{m+1}| r^2 < |a_M| r^3, \end{aligned}$$

and in general, it holds that $|a_{m+n}| < |a_M| r^n$. Consequently, we have that

$$\sum_{n=m}^{\infty} |a_n| = \sum_{k=0}^{\infty} |a_{m+k}| = \sum_{n=0}^{\infty} |a_{m+n}| < \sum_{n=0}^{\infty} |a_M| r^n = |a_M| \sum_{n=0}^{\infty} r^n.$$

By the Convergence of Geometric Series, the geometric series on the right-hand side converges by hypothesis that $0 \leq L < r < 1$. By the Direct Comparison Test, the series $\sum |a_n|$ converges, hence the series in question converges absolutely by definition of absolute convergence.

(b.) If $L > 1$, we can ensure that $s > 1$ by taking ε to be sufficiently small. By a similar argument as above, it follows that $|a_{m+n}| > |a_M| s^n$. Considering that $s > 1$, it follows that

$$\lim_{n\to\infty} |a_n| = \lim_{n\to\infty} |a_{m+n}| > \lim_{n\to\infty} |a_M| s^n = |a_M| \lim_{n\to\infty} s^n = \infty.$$

Consequently, the series in question diverges by the nth Term Divergence Test. □

Exercise 5.6.2. Use the Ratio Test to determine the convergence of the infinite series $\displaystyle\sum_{n=0}^{\infty} \frac{e^n}{n!}$.

Solution. By the Ratio Test, it suffices to compute the following limit and make an interpretation.

$$L = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n\to\infty} \frac{e^{n+1}}{(n+1)!} \cdot \frac{n!}{e^n} = \lim_{n\to\infty} \frac{e}{n+1} = 0.$$

Considering that $L < 1$, by the Ratio Test, we conclude that the series converges absolutely. ◇

Exercise 5.6.3. Use the Ratio Test to determine the convergence of the infinite series $\displaystyle\sum_{n=1}^{\infty} \frac{10^n}{3^{n^2}}$.

Exercise 5.6.4. Use the Ratio Test to determine the convergence of the infinite series $\displaystyle\sum_{n=1}^{\infty} \frac{2^{3^n}}{3^{2^n}}$.

Solution. By the Ratio Test, it suffices to compute the following limit and make an interpretation.

$$L = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n\to\infty} \frac{2^{3^{n+1}}}{3^{2^{n+1}}} \cdot \frac{3^{2^n}}{2^{3^n}} = \lim_{n\to\infty} \frac{2^{3^{n+1}-3^n}}{3^{2^{n+1}-2^n}} = \lim_{n\to\infty} \frac{4^{3^n}}{3^{2^n}} \geq \lim_{n\to\infty} \frac{4^{3^n}}{4^{2^n}} = \lim_{n\to\infty} 4^{3^n-2^n} = \infty.$$

Considering that $L > 1$, by the Ratio Test, we conclude that the series diverges. ◇

Exercise 5.6.5. Use the Ratio Test to determine the convergence of the infinite series $\displaystyle\sum_{n=1}^{\infty} \frac{n^n}{(n^2)!}$.

One of the great advantages of the Ratio Test is that there are no exclusionary rules or provisions to check in order to perform the test: indeed, we can completely carry out the analysis of the Ratio Test from start to finish with the infinite series that is handed to us. Compare this feature with either the Integral Test or Alternating Series Test in which we are required to verify some properties of the general term of the series before reaping the benefits. Even more, the Ratio Test requires no inspiration or divine intervention. Compare this with either of the Comparison Tests in which we are required to (perhaps miraculously) come up with another sequence to compare with the sequence at hand (as if the sequence we started with was not already enough to handle).

Unfortunately, as with all things that appear too good to be true, there is a substantial caveat to using the Ratio Test: namely, the Ratio Test cannot determine the convergence of any series whose general term converges "too slowly" to zero. For instance, if we consider any rational series

$$\sum_{n=m}^{\infty} \frac{p(n)}{q(n)}$$

for any real polynomial functions $p(x)$ and $q(x)$ such that $q(x)$ is nonzero for any integer $n \geq m$, the Ratio Test is inconclusive since the degree and leading coefficient of $p(n+1)$ are equal to the respective degree and leading coefficient of $p(n)$ (and likewise for q), hence we have that

$$L = \lim_{n \to \infty} \left| \frac{p(n+1)}{q(n+1)} \cdot \frac{q(n)}{p(n)} \right| = \lim_{n \to \infty} \left| \frac{p(n+1)}{p(n)} \cdot \frac{q(n)}{q(n+1)} \right| = \left(\lim_{n \to \infty} \frac{p(n+1)}{p(n)} \right) \left(\lim_{n \to \infty} \frac{q(n)}{q(n+1)} \right) = 1.$$

Likewise, the Ratio Test fails to detect the divergence of the infinite series

$$\sum_{n=2}^{\infty} \frac{1}{\ln(n)}$$

because the positive sequence $\ln(n)$ grows "too slowly." Explicitly, we have that

$$L = \lim_{n \to \infty} \left| \frac{\ln(n+1)}{\ln(n)} \right| = \lim_{n \to \infty} \frac{\ln(n+1)}{\ln(n)} \stackrel{\text{L'H}}{=} \lim_{n \to \infty} \frac{n}{n+1} = 1.$$

Even in spite of these shortcomings, we will soon see that the Ratio Test is an indispensable tool.

5.7 Power Series

Recall that a nonzero **polynomial** function of **degree** n is any function of the form

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

for some integer $n \geq 0$ and real numbers $a_0, a_1, \dots, a_n$ such that a_n is nonzero. We refer to a_i as the **coefficient** of the **monomial** x^i for each integer $0 \leq i \leq n$; the monomials $a_i x^i$ are the **terms** of the polynomial. Using the notion of infinite series, we obtain a generalization of polynomials that allows us to include terms of arbitrarily large degree. Explicitly, we define the **power series**

$$f(x) = \sum_{n=0}^{\infty} a_n (x-c)^n = a_0 + a_1(x-c) + a_2(x-c)^2 + \cdots + a_n(x-c)^n + \cdots .$$

We refer to the constant c as the **center** of the power series. Under this identification, a polynomial function is simply a power series for which the sequence of coefficients a_n is nonzero for only finitely many integers $n \geq 0$ (i.e., we have that $a_n = 0$ for all sufficiently large integers n).

Convergence of a power series depends not only on its coefficients but also its center c.

Example 5.7.1. Consider the power series centered at $x = 0$ with sequence of coefficients $a_n = n$.

$$f(x) = \sum_{n=0}^{\infty} a_n (x-c)^n = \sum_{n=0}^{\infty} n x^n$$

Observe that the value of $f(x)$ is determined by the convergence of the underlying power series.

Explicitly, in this case, by plugging in $x = -1$, $x = \frac{1}{2}$, and $x = 1$, we have that

$$f(-1) = \sum_{n=0}^{\infty} (-1)^n n,$$

$$f\left(\frac{1}{2}\right) = \sum_{n=0}^{\infty} n \left(\frac{1}{2}\right)^n = \sum_{n=0}^{\infty} \frac{n}{2^n}, \text{ and}$$

$$f(1) = \sum_{n=0}^{\infty} n.$$

By the nth Term Divergence Test, it follows that the infinite series corresponding to $f(-1)$ and $f(1)$ diverge. By the Ratio Test, we find that the infinite series corresponding to $f\left(\frac{1}{2}\right)$ converges.

$$L = \lim_{n \to \infty} \frac{n+1}{2^{n+1}} \cdot \frac{2^n}{n} = \lim_{n \to \infty} \frac{n+1}{2n} = \frac{1}{2}$$

Generally, the convergence of a power series can be determined by the Ratio Test as follows.

Theorem 5.7.2 (Convergence of Power Series)**.** *Consider the power series*

$$f(x) = \sum_{n=0}^{\infty} a_n (x-c)^n$$

centered at the real number c determined by the infinite sequence of real numbers a_n. There exists a (possibly infinite) real number $R \geq 0$ called the **radius of convergence** *of $f(x)$ such that*

(a.) *$f(x)$ converges absolutely for all real numbers x such that $c - R < x < c + R$ and*

(b.) *$f(x)$ diverges for all real numbers x such that $x > c + R$ or $x < c - R$.*

We refer to the interval $I = (c - R, c + R)$ where $f(x)$ converges as the **interval of convergence***. Even more, the radius of convergence and the interval of convergence satisfy the following properties.*

(a.) *If $R = \infty$, then $f(x)$ converges absolutely for all real numbers, i.e., $I = (-\infty, \infty)$.*

(b.) *If $R > 0$ is finite, then $f(x)$ may converge or diverge at $x = c - R$ and $x = c + R$.*

(c.) *If $R = 0$, then $f(x)$ diverges for all real numbers $x \neq c$ and $f(x)$ converges for $x = c$.*

Proof. By the Ratio Test, the convergence of the power series $f(x)$ is determined by the following.

$$L = \lim_{n \to \infty} \left| \frac{a_{n+1}(x-c)^{n+1}}{a_n(x-c)^n} \right| = \lim_{n \to \infty} \left| \frac{a_{n+1}(x-c)}{a_n} \right| = |x - c| \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

Consider the (possibly infinite) real number R such that the following equality holds.

$$\frac{1}{R} = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

Certainly, if the above limit tends to 0, then L tends to 0 and R tends to infinity. Even more, in this case, we have that $L = 0$, hence the Ratio Test ensures that $f(x)$ converges absolutely for all real numbers x. Conversely, if the above limit tends to infinity, then L tends to infinity and R tends to 0, hence the Ratio Test ensures that $f(x)$ diverges for all real numbers $x \neq c$ and $f(x)$ converges for $x = c$. Last, if the above limit is finite, then the real number $R \geq 0$ is finite. By the Ratio Test, the power series $f(x)$ converges absolutely if and only if $L < 1$ if and only if

$$\frac{|x-c|}{R} < 1$$

if and only if $|x-c| < R$ if and only if $-R < x - c < R$ if and only if $c - R < x < c + R$. Put another way, we have that $f(x)$ converges absolutely for all real numbers x such that $c - R < x < c + R$ and $f(x)$ diverges for all real numbers x such that $x > c + R$ or $x < c - R$. □

Caution. Be very careful to note that the Convergence of Power Series theorem does not guarantee anything about the convergence of $f(x)$ when $x = c - R$ or $x = c + R$ in the case that the radius of convergence R is finite and nonzero; rather, we must explicitly test for convergence at these points.

Exercise 5.7.3. Compute the radius and interval of convergence of the power series $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n!}$.

Solution. We proceed by the Ratio Test.

$$\begin{aligned}
L &= \lim_{n\to\infty} \left| \frac{x^{n+1}}{(n+1)!} \cdot \frac{n!}{x^n} \right| \\
&= \lim_{n\to\infty} \left| \frac{x^{n+1}}{x^n} \cdot \frac{n!}{(n+1)!} \right| && \text{(Group like terms.)} \\
&= |x| \lim_{n\to\infty} \frac{n!}{(n+1)!} && \text{(Cancel, and pull out constants.)} \\
&= |x| \lim_{n\to\infty} \frac{n!}{(n+1)n!} && \text{(Express } n! \text{ as a factor of } (n+1)!.\text{)} \\
&= |x| \lim_{n\to\infty} \frac{1}{n+1} && \text{(Cancel common factors.)} \\
&= 0.
\end{aligned}$$

We conclude that regardless of x, the power series in question converges absolutely. Consequently, the radius of convergence is $R = \infty$, and the interval of convergence is $I = (-\infty, \infty)$. ◇

Exercise 5.7.4. Compute the radius and interval of convergence of the power series $\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2^n n}$.

Solution. We proceed by the Ratio Test.

$$L = \lim_{n\to\infty} \left| \frac{x^{2n+3}}{2^{n+1}(n+1)} \cdot \frac{2^n n}{x^{2n+1}} \right| = \lim_{n\to\infty} \left| \frac{x^2(n+1)}{2n} \right| = \frac{x^2}{2} \lim_{n\to\infty} \frac{n+1}{n} = \frac{x^2}{2}$$

By the Ratio Test, the series converges if $L < 1$ if and only if $x^2 < 2$ if and only if $-\sqrt{2} < x < \sqrt{2}$. Even more, the series diverges if $L > 1$ if and only if $x^2 > 2$ if and only if $x > \sqrt{2}$ or $x < -\sqrt{2}$. Consequently, it suffices to determine convergence at $x = \pm\sqrt{2}$. Observe that if $x = \sqrt{2}$, then

$$\frac{x^{2n+1}}{2^n} = \frac{\sqrt{2}^{2n+1}}{\sqrt{2}^{2n}} = \sqrt{2} \text{ so that } \frac{(-1)^n x^{2n+1}}{2^n n} = \frac{\sqrt{2}}{n}.$$

Considering that this sequence is positive, decreasing, and converges to 0, by the Alternating Series Test, we conclude that the power series converges at $x = \sqrt{2}$. Likewise, if $x = -\sqrt{2}$, then

$$(-1)^n x^{2n+1} = (-1)^n(-\sqrt{2})^{2n+1} = (-1)^n(-1)^{2n+1}\sqrt{2}^{2n+1} = (-1)^{3n+1}\sqrt{2}^{2n+1} = (-1)^{n+1}\sqrt{2}^{2n+1}$$

is alternating. By the above rationale, the series converges at $x = -\sqrt{2}$. We conclude that the radius of convergence of the power series is $R = 2\sqrt{2}$, and the interval of convergence is $[-\sqrt{2}, \sqrt{2}]$. ⋄

Exercise 5.7.5. Compute the radius and interval of convergence of the power series $\displaystyle\sum_{n=0}^{\infty} n^n x^n$.

Solution. We proceed by the Ratio Test.

$$L = \lim_{n\to\infty} \left| \frac{(n+1)^{n+1} x^{n+1}}{n^n x^n} \right| = |x| \lim_{n\to\infty} \frac{(n+1)(n+1)^n}{n^n} = |x| \lim_{n\to\infty} (n+1)\left(\frac{n+1}{n}\right)^n$$

Consequently, it suffices to compute the limit of the terms involving n. We note first that if

$$y = \left(1 + \frac{1}{x}\right)^x, \text{ then}$$

$$\ln(y) = x \ln\left(1 + \frac{1}{x}\right) \text{ implies that}$$

$$\frac{1}{y}\frac{dy}{dx} = \frac{x}{1 + \frac{1}{x}}\left(-\frac{1}{x^2}\right) + \ln\left(1 + \frac{1}{x}\right) = -\frac{1}{x+1} + \ln\left(1 + \frac{1}{x}\right) \text{ and}$$

$$\frac{dy}{dx} = \left(1 + \frac{1}{x}\right)^x \left[-\frac{1}{x+1} + \ln\left(1 + \frac{1}{x}\right)\right] > 0 \text{ for all real numbers } x \geq 1.$$

Crucially, it follows that the sequence of n in the above limit is increasing and unbounded because it is a product of increasing sequences neither of which converges to 0. Consequently, we have that $L = \infty$. By the Ratio Test, we conclude that the power series diverges for all real numbers $x \neq 0$. ⋄

Exercise 5.7.6. Compute the radius and interval of convergence of the power series $\displaystyle\sum_{n=0}^{\infty} \frac{(3-x)^n}{3^n}$.

Solution. By grouping the terms of the power series with exponent n together under the same exponent, we can view this power series as a geometric series with common ratio

$$r = \frac{x-3}{-3}.$$

By the Convergence of Geometric Series, we conclude that the power series converges if and only if $-1 < r < 1$ if and only if $-3 < x - 3 < 3$ if and only if $0 < x < 6$. Consequently, the radius of convergence of the power series is $R = 3$, and the interval of convergence is $I = (0, 6)$. $\diamond$

Example 5.7.7. Given any nonzero real number c and any real function $g(x)$, consider the series

$$f(x) = \sum_{n=0}^{\infty} c[g(x)]^n.$$

By the Ratio Test, we have that $f(x)$ converges absolutely if and only if

$$|g(x)| = \lim_{n\to\infty} |g(x)| = \lim_{n\to\infty} \left|\frac{[g(x)]^{n+1}}{[g(x)]^n}\right| = \lim_{n\to\infty} \left|\frac{c[g(x)]^{n+1}}{c[g(x)]^n}\right| < 1.$$

Even more, by the Convergence of Geometric Series formula, if $|g(x)| < 1$, then

$$f(x) = \sum_{n=0}^{\infty} c[g(x)]^n = \frac{c}{1-g(x)}.$$

Consequently, we obtain a closed form of $f(x)$ in terms of $g(x)$ for all real numbers x with $|g(x)| < 1$.

Exercise 5.7.8. Use Example 5.7.7 to express each of the following functions as a power series; then, state the radius and interval of convergence for each power series.

(a.) $\dfrac{1}{1-x}$ (b.) $\dfrac{1}{1+x}$ (c.) $\dfrac{1}{1+x^2}$

Solution. (a.) Observe that for the real function $g(x) = x$, we obtain the power series representation

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

that is valid for all real numbers x such that $|x| < 1$. We conclude that the radius of convergence of the power series is $R = 1$, and the interval of convergence is $I = (-1, 1)$.

(c.) Observe that for the real function $g(x) = -x^2$, we obtain the power series representation

$$\frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{n=0}^{\infty} (-x^2)^n = \sum_{n=0}^{\infty} (-1)^n x^{2n}$$

that is valid for all real numbers x such that $0 \leq x^2 < 1$. Considering that $0 \leq x^2 < 1$ if and only if $-1 < x < 1$, the radius of convergence is $R = 1$, and the interval of convergence is $I = (-1, 1)$. $\diamond$

One of the most useful features of power series is that we may differentiate them term-by-term.

Theorem 5.7.9 (Power Series Are Differentiable)**.** *Consider the following power series centered at a real number c with (possibly infinite) radius of convergence $R > 0$.*

$$f(x) = \sum_{n=0}^{\infty} a_n (x-c)^n$$

(a.) *We have that $f(x)$ is differentiable on the interval $I = (c-R, c+R)$ with derivative*

$$f'(x) = \frac{d}{dx} f(x) = \frac{d}{dx} \sum_{n=0}^{\infty} a_n (x-c)^n = \sum_{n=0}^{\infty} a_n \frac{d}{dx} (x-c)^n = \sum_{n=1}^{\infty} n a_n (x-c)^{n-1}.$$

Consequently, $f'(x)$ is a power series centered at $x = c$ with radius of convergence R.

(b.) *We have that the antiderivative of $f(x)$ on the interval $I = (c-R, c+R)$ is given by*

$$F(x) + C = \int f(x)\, dx = \int \sum_{n=0}^{\infty} a_n (x-c)^n\, dx = \sum_{n=0}^{\infty} a_n \int (x-c)^n\, dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1} (x-c)^{n+1}.$$

Consequently, $\int f(x)\, dx$ is a power series centered at $x = c$ with radius of convergence R.

We note that in practice, the constant C can be found by used the fact that $F(c) + C = 0$.

Exercise 5.7.10. Use Example 5.7.8 to find a power series representation for the following functions; then, state the radius and interval of convergence for each power series.

(a.) $\dfrac{1}{(1-x)^2}$ (b.) $\dfrac{-2x}{(1+x^2)^2}$ (c.) $\ln(1+x)$ (d.) $\arctan(x)$

Solution. (a.) By the Power Rule and the Chain Rule for Derivatives, we have that

$$\frac{1}{(1-x)^2} = \frac{d}{dx} \frac{1}{1-x} = \frac{d}{dx} \sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} \frac{d}{dx} x^n = \sum_{n=0}^{\infty} n x^{n-1} = \sum_{n=1}^{\infty} n x^{n-1}$$

by Theorem 5.7.9; the radius of convergence is $R = 1$, and the interval of convergence is $(-1, 1)$.

(b.) By the Power Rule and Chain Rule for Derivatives, we have that

$$\frac{-2x}{(1+x^2)^2} = \frac{d}{dx} \frac{1}{1+x^2} = \frac{d}{dx} \sum_{n=0}^{\infty} (-1)^n x^{2n} = \sum_{n=0}^{\infty} (-1)^n \frac{d}{dx} x^{2n} = \sum_{n=1}^{\infty} (-1)^n 2n x^{2n-1}$$

with radius of convergence $R = 1$ and interval of convergence $(-1, 1)$.

(c.) Observe that if $-1 < x < 1$, then $0 < 1 + x < 2$ and

$$\ln(1+x) + C = \int \frac{1}{1+x}\, dx = \int \sum_{n=0}^{\infty} (-1)^n x^n\, dx = \sum_{n=0}^{\infty} (-1)^n \int x^n\, dx = \sum_{n=0}^{\infty} \frac{(-1)^n x^{n+1}}{n+1}.$$

By plugging in $x = 0$, we find that $C = \ln(1) + C = 0$, hence we conclude that

$$\ln(1+x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{n+1}}{n+1}$$

with radius of convergence $R = 1$ and interval of convergence $[-1, 1]$. Crucially, we achieve convergence at both endpoints $x = -1$ and $x = 1$ by the Alternating Series Test.

(d.) Last, we have that

$$\arctan(x) + C = \int \frac{1}{1+x^2}\,dx = \int \sum_{n=0}^{\infty} (-1)^n x^{2n}\,dx = \sum_{n=0}^{\infty} (-1)^n \int x^{2n}\,dx = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1}.$$

By plugging in $x = 0$, we find that $C = \arctan(0) + C = 0$, hence we conclude that

$$\arctan(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1}$$

with radius $R = 1$ and interval of convergence $[-1, 1]$ by the Alternating Series Test. $\diamond$

One immediate consequence of the previous example is that we are now able to approximate (via power series) the value of previously unknown quantities. By Example 5.7.10(c.), we have that

$$\frac{\pi}{4} = \arctan(1) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots.$$

Consequently, we can approximate π to any desired degree of accuracy, i.e., we can write the decimal expansion of π in a manner that is accurate to as many decimal places as desired! (We remark that the convergence of the arctangent series is very slow since it is asymptotically equivalent to the alternating harmonic series; indeed, there are better series approximations of π that are preferred in practice. Even still, this approximation of π is historically significant and quite remarkable.)

5.8 Taylor Series

Consider any real function $f(x)$ and any real number c such that the nth derivative $f^{(n)}(x)$ of $f(x)$ exists at $x = c$ for all integers $n \geq 0$. We refer to the power series

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n$$

as the **Taylor series** of $f(x)$ centered at $x = c$. Crucially, if a real function $f(x)$ is represented by a power series centered at $x = c$ on some open interval of the form $(c - R, c + R)$ for some real number $R > 0$, then that power series must be the Taylor series of $f(x)$ centered at $x = c$.

Theorem 5.8.1 (Uniqueness of Taylor Series)**.** *Given any real function $f(x)$ for which the Taylor series of $f(x)$ centered at $x = c$ exists, there exists a real number $R > 0$ such that the Taylor series of $f(x)$ centered at $x = c$ is the unique power series expansion of $f(x)$ on the interval $(c-R, c+R)$.*

Caution. Be very careful not to misinterpret the theorem. Explicitly, this does not guarantee that a real function $f(x)$ admits a power series expansion; rather, it says that if $f(x)$ can be represented as a power series center at $x = c$, then that power series must in fact be the Taylor series of $f(x)$.

We have in Examples 5.7.7, 5.7.8, and 5.7.10 provided power series expansions of the logarithmic function $\ln(1+x)$, the inverse trigonometric function $\arctan(x)$, and the rational functions

$$\frac{1}{1-x} \text{ and } \frac{1}{1+x^2}.$$

Consequently, by Theorem 5.8.1, these are in fact the Taylor series expansions of these functions centered at $x = 0$! Generally, if it exists, the Taylor series expansion of a real function $f(x)$ centered at $x = 0$ is referred to as the **Maclaurin series** of $f(x)$; the terminology is no doubt perplexing, but it is commonplace and remains in use due to historical considerations.

Exercise 5.8.2. Compute the Maclaurin series of $f(x) = e^x$.

Solution. We compute the nth derivative of e^x for each integer $n \geq 0$. Considering that $f(x) = e^x$ satisfies that $f'(x) = e^x$, we find that $f^{(n)}(x) = e^x$ for all integers $n \geq 0$. Consequently, we have that $f^{(n)}(0) = 1$ for all integers $n \geq 0$. By the formula for the Maclaurin series, we conclude that

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Even more, by the Ratio Test, the series converges absolutely for all real numbers since

$$L = \lim_{n \to \infty} \left| \frac{x^{n+1}}{(n+1)!} \cdot \frac{n!}{x^n} \right| = \lim_{n \to \infty} \left| \frac{xn!}{(n+1)n!} \right| = |x| \lim_{n \to \infty} \frac{1}{n+1} = 0. \qquad \diamond$$

Exercise 5.8.3. Compute the Maclaurin series of $f(x) = \cos(x)$.

Solution. We compute the nth derivative of $\cos(x)$ for each integer $n \geq 0$. Observe that

$$\begin{aligned} f(x) &= \cos(x), & f''(x) &= -\cos(x), \text{ and} \\ f'(x) &= -\sin(x), & f'''(x) &= \sin(x). \end{aligned}$$

Considering that $\cos(x)$ is the derivative of $\sin(x)$, the derivatives of $\cos(x)$ are periodic with

$$\begin{aligned} f^{(4k)}(x) &= \cos(x), & f^{(4k+2)}(x) &= -\cos(x), \text{ and} \\ f^{(4k+1)}(x) &= -\sin(x), & f^{(4k+3)}(x) &= \sin(x) \end{aligned}$$

for all integers $k \geq 0$. Consequently, for each integer $k \geq 1$, we have that

$$\begin{aligned} f^{(4k)}(0) &= 1, & f^{(4k+2)}(0) &= -1, \text{ and} \\ f^{(4k+1)}(0) &= 0, & f^{(4k+3)}(0) &= 0. \end{aligned}$$

Put another way, the even derivatives of $f(x) = \cos(x)$ are alternating in sign and the odd derivatives of $f(x) = \cos(x)$ are zero for $c = 0$. By the formula for the Maclaurin series, we conclude that

$$\cos(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.$$

Even more, by the Ratio Test, the series converges absolutely for all real numbers since

$$L = \lim_{n \to \infty} \left| \frac{x^{2n+2}}{(2n+2)!} \cdot \frac{(2n)!}{x^{2n}} \right| = \lim_{n \to \infty} \left| \frac{x^2(2n)!}{(2n+2)(2n+1)(2n)!} \right| = x^2 \lim_{n \to \infty} \frac{1}{4n^2+6n+2} = 0. \qquad \diamond$$

Exercise 5.8.4. Compute the Maclaurin series of $f(x) = \sin(x)$.

Solution. We compute the nth derivative of $\sin(x)$ for each integer $n \geq 0$. Observe that

$$f(x) = \sin(x), \qquad f''(x) = -\sin(x), \text{ and}$$
$$f'(x) = \cos(x), \qquad f'''(x) = -\cos(x).$$

Considering that $\sin(x)$ is the derivative of $-\cos(x)$, the derivatives of $\sin(x)$ are periodic with

$$f^{(4k)}(x) = \sin(x), \qquad f^{(4k+2)}(x) = --\sin(x), \text{ and}$$
$$f^{(4k+1)}(x) = \cos(x), \qquad f^{(4k+3)}(x) = -\cos(x)$$

for all integers $k \geq 0$. Consequently, for each integer $k \geq 1$, we have that

$$f^{(4k)}(0) = 0, \qquad f^{(4k+2)}(0) = 0, \text{ and}$$
$$f^{(4k+1)}(0) = 1, \qquad f^{(4k+3)}(0) = -1.$$

Put another way, the odd derivatives of $f(x) = \sin(x)$ are alternating in sign and the even derivatives of $f(x) = \sin(x)$ are zero for $c = 0$. By the formula for the Maclaurin series, we conclude that

$$\sin(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}.$$

Even more, by the Ratio Test, the series converges absolutely for all real numbers since

$$L = \lim_{n \to \infty} \left| \frac{x^{2n+3}}{(2n+3)!} \cdot \frac{(2n+1)!}{x^{2n+1}} \right| = \lim_{n \to \infty} \left| \frac{x^2 (2n+1)!}{(2n+3)(2n+2)(2n+1)!} \right| = x^2 \lim_{n \to \infty} \frac{1}{4n^2 + 10n + 6} = 0. \diamond$$

Theorem 5.8.5 (Convergence of Taylor Series)**.** *Consider any real numbers c and $R > 0$ and any real function $f(x)$ such that $f^{(n)}(x)$ is continuously differentiable for all integers $n \geq 0$ and all real numbers x such that $c - R < x < c + R$. Provided that there exists a real number K such that $|f^{(n)}(x)| \leq K$ for all integers $n \geq 0$ and all real numbers x such that $c - R < x < c + R$, the Taylor series of $f(x)$ centered at $x = c$ converges to $f(x)$, i.e., the following representation of $f(x)$ is valid.*

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!} (x - c)^n$$

Once we know the Taylor series expansion of some real function $f(x)$ centered at $x = c$, it is not overtly difficult to find the Taylor series expansion of the real functions $x^k f(x)$ or $f(x)/x^k$ for some any integer $k \geq 1$ or the Taylor series expansion of $(g \circ f)(x)$ for some real function $g(x)$; however, it is possible to change the center of a Taylor series when performing these operations.

Exercise 5.8.6. Compute the Taylor series expansion of each of the following.

(a.) $x^3 \cos(x)$ (b.) e^{1-x^2} (c.) e^{x-4} (d.) $\dfrac{x - \sin(x)}{x}$

Solution. (a.) By Example 5.8.3, the Maclaurin series for $\cos(x)$ is given by

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.$$

By multiplying this power series by x^3, we obtain the Maclaurin series expansion

$$x^3 \cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+3}}{(2n)!}.$$

By the Ratio Test, this series converges for all real numbers.

(b.) By Example 5.8.2, the Maclaurin series for e^x is given by

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Considering that $e^{1-x^2} = ee^{-x^2}$, plugging in $-x^2$ to the above yields the Maclaurin series

$$e^{1-x^2} = ee^{-x^2} = e \sum_{n=0}^{\infty} \frac{(-x^2)^n}{n!} = \sum_{n=0}^{\infty} \frac{e(-1)^n x^{2n}}{n!}.$$

By the Ratio Test, this series converges for all real numbers.

(c.) Likewise, by plugging in $x - 4$ to the Maclaurin series of e^x, we obtain the Taylor series

$$e^{x-4} = \sum_{n=0}^{\infty} \frac{(x-4)^n}{n!}$$

of e^x centered at $x = 4$. By the Ratio Test, this series converges for all real numbers.

(d.) By Example 5.8.4, the Maclaurin series for $\sin(x)$ is given as follows.

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

By multiplying this expansion by -1 and adding x, we obtain the following.

$$x - \sin(x) = x - \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right) = \frac{x^3}{3!} - \frac{x^5}{5!} + \frac{x^7}{7!} - \frac{x^9}{9!} + \cdots$$

Last, by dividing each side of this identity by x, we conclude that

$$\frac{x - \sin(x)}{x} = \frac{x^2}{3!} - \frac{x^4}{5!} + \frac{x^6}{7!} - \frac{x^8}{9!} + \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n+1} x^{2n}}{(2n+1)!}.$$

By the Ratio Test, this Maclaurin series converges for all real numbers. ◇

One of the most ingenious uses of power series is to compute limits and to find power series representations for the antiderivatives of certain functions that lack elementary antiderivatives.

Exercise 5.8.7. Verify that L'Hôpital's Rule can be used to compute the limit

$$\lim_{x \to 0} \frac{x - \sin x}{x^3 \cos x};$$

then, explain the difficulty in doing so. Ultimately, compute the limit using power series.

Exercise 5.8.8. Verify that L'Hôpital's Rule can be used to compute the limit

$$\lim_{x \to 0} \frac{\cos(\sqrt{x}) - 1}{2x};$$

then, explain the difficulty in doing so. Ultimately, compute the limit using power series.

5.9 Taylor Polynomials and Approximation

Given any real function $f(x)$ that is differentiable on an open interval (a, b) and any real number c such that $a < c < b$, recall that the **first-order** (or **linear**) **approximation** of $f(x)$ is given by

$$f(x) \approx f(c) + f'(c)(x - c)$$

for any real number x such that the distance $|x - c|$ between x and c is sufficiently small: indeed, by the limit definition of the derivative $f'(x)$ evaluated at the real number $x = c$, we have that

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c} \text{ so that } f'(c) \approx \frac{f(x) - f(c)}{x - c}$$

when x is sufficiently close to c. We refer to the linear polynomial $T_1(x) = f(c) + f'(c)(x - c)$ as the **linearization** of $f(x)$ at $x = c$. Considering that $T_1(c) = f(c)$ and $T_1'(c) = f'(c)$, it follows that $T_1(x)$ agrees with $f(x)$ and $T_1'(x)$ agrees with $f'(x)$ on a sufficiently small open interval containing $x = c$, so in practice, we can replace $f(x)$ with a linear polynomial for all real numbers x sufficiently close to $x = c$. Power series provide an even more powerful technique for approximation of differentiable real functions as polynomials. Explicitly, consider the following power series.

$$f(x) = \sum_{n=0}^{\infty} a_n (x - c)^n = a_0 + a_1(x - c) + a_2(x - c)^2 + a_3(x - c)^3 + \cdots$$

By Theorem 5.7.9, we may differentiate $f(x)$ term-by-term to obtain the following.

$$\begin{aligned} f(c) &= a_0 & \qquad f''(c) &= 2a_2 \\ f'(c) &= a_1 & \qquad f'''(c) &= 3 \cdot 2a_3 \end{aligned}$$

Continuing in this manner and identifying the pattern yields $f^{(n)}(c) = n!\, a_n$. Consequently, we obtain an n**th-order** approximation of $f(x)$ at $x = c$ by the following real polynomial of degree n.

$$T_n(x) = f(c) + f'(c)(x - c) + \frac{f''(c)}{2!}(x - c)^2 + \frac{f'''(c)}{3!}(x - c)^3 + \cdots + \frac{f^{(n)}(c)}{n!}(x - c)^n$$

We refer to the real polynomial $T_n(x)$ of degree n as the **nth Taylor polynomial** of $f(x)$ centered at $x = c$. Crucially, we observe that the nth Taylor polynomial of $f(x)$ centered at $x = c$ satisfies

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k = \frac{f^{(n)}(c)}{n!}(x-c)^n + \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k = \frac{f^{(n)}(c)}{n!}(x-c)^n + T_{n-1}(x).$$

Consequently, the difference between consecutive Taylor polynomials satisfies that

$$T_n(x) - T_{n-1}(x) = \frac{f^{(n)}(c)}{n!}(x-c)^n.$$

We will soon discover that this observation is absolutely indispensable to the development of the theory of Taylor series; however, before that, we point out the following fundamental fact.

Theorem 5.9.1 (Uniqueness of Taylor Polynomials)**.** *Given any integer $n \geq 0$ and any real function $f(x)$ such that $f^{(n)}(x)$ is continuous at $x = c$, the Taylor polynomial $T_n(x)$ of $f(x)$ centered at $x = c$ is the unique polynomial of degree (at most) n that approximates $f(x)$ to order n at $x = c$.*

We are already familiar with the Taylor series of several common functions; thus, these power series expansions will make short work of determining the Taylor polynomials of such functions.

Example 5.9.2. By Example 5.8.2, the Taylor series of $f(x) = e^x$ centered at $x = 0$ is given by

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Consequently, for any integer $n \geq 0$, the nth Taylor polynomial of e^x is given by

$$T_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}.$$

Explicitly, we have that $T_3(x) = 1 + x + \dfrac{x^2}{2} + \dfrac{x^3}{6}$.

Example 5.9.3. By Example 5.8.3, the Taylor series of $f(x) = \cos(x)$ centered at $x = 0$ is

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}.$$

Consequently, for any integer $n \geq 0$, the nth Taylor polynomial of $\cos(x)$ is

$$T_n(x) = \sum_{k=0}^{n} \frac{(-1)^k x^{2k}}{(2k)!}.$$

Explicitly, we have that $T_2(x) = 1 - \dfrac{x^2}{2} + \dfrac{x^4}{24}$.

Example 5.9.4. By Example 5.8.4, the Taylor series of $f(x) = \sin(x)$ centered at $x = 0$ is

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}.$$

Consequently, for any integer $n \geq 0$, the nth Taylor polynomial of $\sin(x)$ is

$$T_n(x) = \sum_{k=0}^{n} \frac{(-1)^k x^{2k+1}}{(2k+1)!}.$$

Explicitly, we have that $T_2(x) = x - \dfrac{x^3}{6} + \dfrac{x^5}{120}$.

Of course, we have said all along that our aim has been to use power series to *approximate*, and as with any approximation, there is some amount of error involved.

Theorem 5.9.5 (Error Bound Theorem)**.** *Consider any real function $f(x)$ such that $f^{(n+1)}(x)$ is continuous near $x = c$. Let $T_n(x)$ denote the nth Taylor polynomial of $f(x)$ centered at $x = c$. Provided that there exists a real number K such that $|f^{(n+1)}(a)| \leq K$ for all real numbers a between c and x, the error in approximating $f(x)$ by $T_n(x)$ for all real numbers near $x = c$ is bounded and*

$$|f(x) - T_n(x)| \leq \frac{K|x-c|^{n+1}}{(n+1)!}.$$

Exercise 5.9.6. Use the Error Bound Theorem to find the maximum error in approximating e^2 with $f(x) = e^x$ and its fourth Taylor polynomial $T_4(x)$ centered at $x = 0$.

Exercise 5.9.7. Use the Error Bound Theorem to approximate $\cos(1)$ to three decimal places.

Consider any real function $f(x)$ such that $f^{(n)}(x)$ is continuous for all integers $n \geq 0$ and all real numbers x in some open interval I. Let K be any real number such that $|f^{(n)}(x)| \leq K$ for all integers $n \geq 0$ and all real numbers x in I. By Proposition 5.1.22 and the Error Bound Theorem,

$$\lim_{n \to \infty} |f(x) - T_n(x)| \leq \lim_{n \to \infty} \frac{K|x-c|^{n+1}}{(n+1)!} = 0.$$

Consequently, as its degree n grows arbitrarily large, the nth Taylor polynomial centered at $x = c$ provides an increasingly better approximation of $f(x)$ near $x = c$ since the error in approximating $f(x)$ via $T_n(x)$ near $x = c$ converges to 0. Even more, the Taylor series of $f(x)$ satisfies that

$$\lim_{n \to \infty} T_n(x) = \lim_{n \to \infty} \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k = f(x).$$

Exercise 5.9.8. Explain the difficulty in trying to find the antiderivative of $\sin(x^2)$; then, compute the power series expansion of the antiderivative $\sin(x^2)$, and state its radius of convergence.

Exercise 5.9.9. Explain the difficulty in trying to find the antiderivative of e^{1-x^2}; then, compute the power series expansion of the antiderivative e^{1-x^2}, and state its radius of convergence.

References

[RA15] Jon Rogawski and Colin Adams. *Calculus: Early Transcendentals*. 3rd ed. W.H. Freeman, 2015.

[Ste07] James Stewart. *Essential Calculus*. Brooks/Cole, Cengage Learning, 2007.