# MAT134 Lecture Notes

Tyler Holden, ©2014-2017

## Contents

# 1 Introduction

We begin with a very rapid review of some pre-calculus concepts. It is essential that the student master these ideas before we begin studying calculus proper.

## 1.1 The Real Numbers

The real numbers are our focus in this course, or rather functions which eat real numbers. Roughly speaking, the real numbers consist of all possible decimal expansions of numbers, and are denoted $\mathbb{R}$. For example,

$$\pi, \qquad \sqrt{2}, \qquad -17.37\overline{125}$$

are real numbers.

As functions often act on large collections of numbers, we need a system for concisely describing intervals of such numbers. If $a, b$ are real numbers with $a < b$, we use the following notation:

$$\begin{aligned} &\text{All } x \text{ satisfying } a < x < b && \text{by} \quad (a,b), \\ &\text{All } x \text{ satisfying } a \leq x < b && \text{by} \quad [a,b), \\ &\text{All } x \text{ satisfying } a < x \leq b && \text{by} \quad (a,b], \\ &\text{All } x \text{ satisfying } a \leq x \leq b && \text{by} \quad [a,b]. \end{aligned}$$

In particular, a parenthesis means that the endpoint is *not* included in the interval, while a square bracket indicates that the endpoint is contained in the interval. We say that the interval $(a, b)$ is an *open* interval and $[a, b]$ is a *closed* interval. The intervals $(a, b]$ and $[a, b)$ may be referred to as either half open or half closed. When we wish to indicate that $x$ is simply less than or larger than a number, we include a $\pm\infty$ sign in the appropriate spot. For example,

$$\begin{aligned} &\text{All } x \text{ satisfying } x < a && \text{by} \quad (-\infty,a), \\ &\text{All } x \text{ satisfying } x \leq a && \text{by} \quad (-\infty,a], \\ &\text{All } x \text{ satisfying } x > b && \text{by} \quad (b,\infty), \\ &\text{All } x \text{ satisfying } x \geq g && \text{by} \quad [b,\infty). \end{aligned}$$

Note that the infinity sign is always used in conjunction with an open bracket. If you are familiar with the notion of unions and intersections, you can use these to combine intervals in a convenient way.

## 1.2 Functions

We think of a function as a machine which eats a number and produces another number. It is important that a function only produce a single output for each input. For example, the function $f(x) = x^2$ takes in an input $x$ and produces the output $x^2$.

| Input | Output |
|---|---|
| $-2$ | $4$ |
| $0$ | $0$ |
| $\sqrt{2}$ | $2$ |
| $\pi$ | $\pi^2$ |

A function has a *domain* and a *range*. The domain is the set of all things which can be put into the function, while the range is the set of all things which come out of the function.

**Example 1.1**

Determine the domain and range of each of the following functions:

1. $f(x) = \dfrac{1}{x}$,
2. $g(x) = \sqrt{x-2}$,
3. $h(x) = (x-1)^2 - 3$,
4. $r(x) = \dfrac{1}{\sqrt{(x-1)(x+1)}}$.

*Solution.*

1. Let's begin by looking at the function $f(x)$. We may divide by every number except 0, hence the domain of this function is $(-\infty, 0) \cup (0, \infty)$. For the range, we notice that $1/x$ can never be zero, since if so then $1/x = 0$ implies that $1 = 0$, and this cannot be true. Hence the range is also $(-\infty, 0) \cup (0, \infty)$.

2. We turn our attention to $g(x)$. Since we may not take the square root of a negative number, we require that $x - 2 \geq 0$ or rather, $x \geq 2$. Hence $f$ has domain $[2, \infty)$. On the other hand, the square root function is always non-negative, with minimum occurring at $x = 2$, showing that the range of $g$ is $[0, \infty)$.

3. We have no restrictions on what numbers can be input into $h$, so the domain of $h$ is $\mathbb{R}$. The range requires a bit more thought. Notice that the value of $(x-1)^2$ is always non-negative, regardless of the input, so $(x-1)^2 - 3 \geq -3$. This is in fact the range $[-3, \infty)$.

4. Since we cannot divide by zero, the points $x = \pm 1$ cannot be in the domain of $r(x)$. Similarly, we cannot take the square root of a negative number. We can determine where $(x+1)(x-1) > 0$ with the following table:

   | | $x < -1$ | $-1 < x < 1$ | $x > 1$ |
   |---|---|---|---|
   | $x-1$ | $-$ | $-$ | $+$ |
   | $x+1$ | $-$ | $+$ | $+$ |
   | $(x-1)(x+1)$ | $+$ | $-$ | $+$ |

   so that $(x-1)(x+1)$ is positive when $x < -1$ and $x > 1$; that is, on the interval $(-\infty, 1) \cup (1, \infty)$. Hence this is the domain of $r(x)$. The range is much tougher! Try it on your own. ■

A useful way to visualize functions is in terms of a graph. This is the collection of points in the $xy$-plane with coordinate $(x, f(x))$; that is, the collection of points in the $xy$-plane whose horizontal

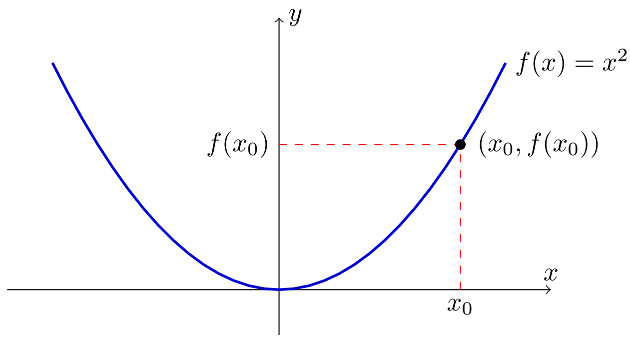



Figure 1: The graph of the function $f(x) = x^2$. The dark blue line is graph($f$) itself, embedded in the plane. In the plane, its coordinates are just $(x, f(x))$.

distance is the $x$-value and whose vertical distance is $f(x)$, seen in Figure 1. The graph of a function gives us the ability to discern qualitative properties of a function by visualizing its behaviour.

Given a curve in the $xy$-plane, there is a simple way of determining whether that curve is the graph of a function, called the *vertical line test*. Since a function must send each input $x$ to a *unique* output $f(x)$, this tells us that each $x$-value must have one and only one point above it in the graph. Given a fixed $x_0$, we may determine the corresponding $f(x_0)$ by drawing a vertical line through $x_0$ and seeing where this line intersects the given curve. If there is any point on a graph where a vertical line intersects the curve twice, the curve cannot correspond to a function.

**Example 1.2**

Consider the curves given in Figure 2. Determine which are given by functions, and which fail to be functions.

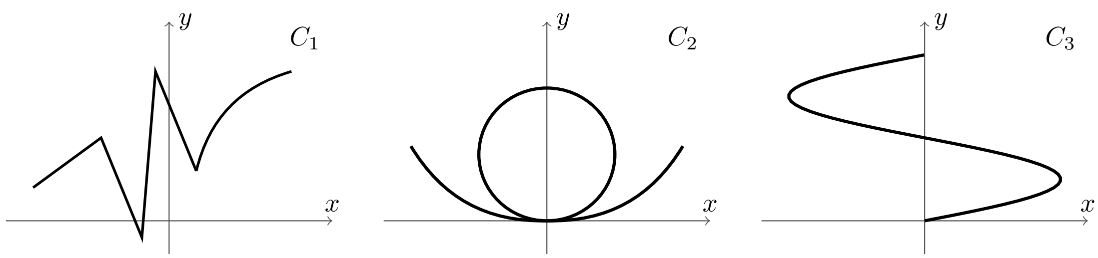



Figure 2: A collection of curves. Which of these are given by functions and which cannot possibly be given by functions?

*Solution.* The curve corresponding to $C_1$ is not particularly appealing, but is nonetheless given by a function. Regardless of where we choose to draw a vertical line, it will intersect the graph at only one point.

The curves $C_2$ and $C_3$ cannot possibly be the graphs of functions, as they fail the vertical line test. While there are many points at which the test fails, perhaps the most obvious place is the $y$-axis. This axis is indeed a vertical line and intersects each of $C_2$ and $C_3$ in 2 and 3 points respectively. ■

### 1.2.1 Operations on Functions

Functions may be added and multiplied in a *pointwise* manner. For example, if $f(x)$ and $g(x)$ are functions, then

$$(f+g)(x) = f(x) + g(x), \qquad (fg)(x) = f(x)g(x).$$

When the range of one function includes into the domain of another, we may combine the two to form a new function $g \circ f$, known as the *composition*:

$$(g \circ f)(s) = g(f(s)).$$

One may think of this as chaining together two functions. Note that not all functions can be composed, and the restriction that the range of $f$ be part of the domain of $g$ is essential. For example, the function $f(x) = -x^2 - 1$ has range $(-\infty, -1]$ and the function $g(x) = \sqrt{x}$ has domain $[0, \infty)$. The composition $(g \circ f)(x) = \sqrt{-x^2-1}$ does not make sense, since any input of $x$ would require that we take the square root of a negative number, which we cannot do.

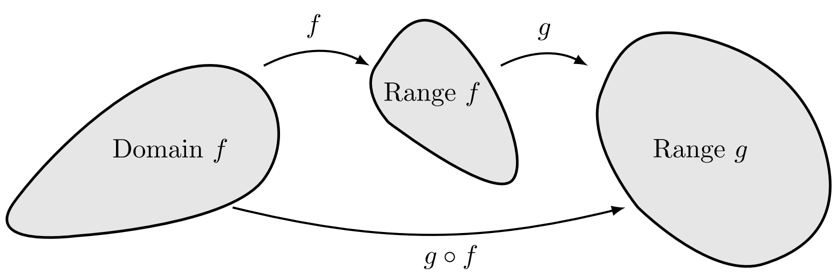



Figure 3: Given two functions $f(x)$ and $g(x)$, their composition $g \circ f$.

**Example 1.3**

Consider the following three functions which map $\mathbb{R} \to \mathbb{R}$: $f(x) = 4x, g(x) = 1/(x^2+1)$ and $h(x) = x + 7$. Compute $f \circ g, f \circ h, g \circ h$ and compare these to $g \circ f, h \circ f$ and $h \circ g$.

*Solution.* Computing compositions can be as easy as substituting one function into the argument of another. Hence

$$(f \circ g)(x) = f(g(x)) = f\left(\frac{1}{x^2+1}\right) = \frac{4}{x^2+1}.$$

Continuing in this fashion for all other examples, we find that

$$\begin{aligned} (f \circ g)(x) &= \frac{4}{x^2+1} & (g \circ f) &= \frac{1}{16x^2+1} \\ (f \circ h)(x) &= 4x + 28 & (h \circ f) &= 4x + 7 \\ (g \circ h)(x) &= \frac{1}{x^2+14x+50} & (h \circ g) &= \frac{7x^2+8}{x^2+1}. \end{aligned}$$

Note that in general, the order of the compositions matters. ∎

### 1.2.2 Symmetries

Function which exhibit symmetric behaviour are desirable, since their analysis is often much simpler. Here we will talk about what it means for a function to be even or odd.

**Definition 1.4**

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *even* if $f(-x) = f(x)$. The function is said to be *odd* if $f(-x) = -f(x)$.

Let us take a moment to determine what is happening. Assume first that $f(x)$ is an even function, so that $f(-x) = f(x)$. This means that for a fixed value $x_0$, the height of the graph of $f(x)$ is the same at both $x_0$ and $-x_0$: Even functions are symmetric about reflections in the $y$-axis. On the other hand, if $f(x)$ is an odd function so that $f(-x) = -f(x)$, then the height of the graph at $-x_0$ is the same as at $x_0$ but now negative: Odd functions are symmetric about rotations of $180°$. Figure 4 gives examples of even and odd functions.

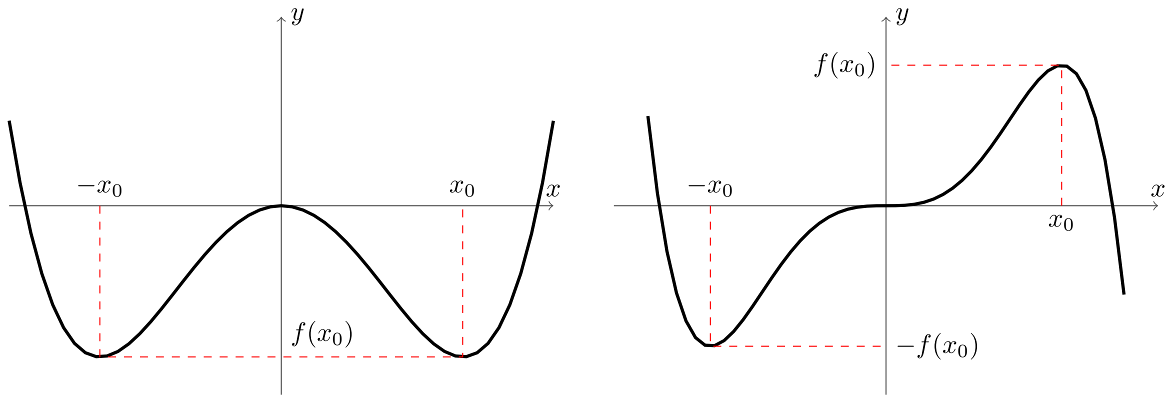



Figure 4: Examples of even (left) and odd (right) functions. Notice that the left figure exhibits reflectional symmetry about the $y$-axis, while the right figure exhibits a rotational symmetry of $180°$ (or equivalently, in both the $x$- and $y$-axes).

**Example 1.5**

Determine whether the following functions are even, odd, or neither:

$$f(x) = x^2, \quad g(x) = x^3, \quad h(x) = f(x) + g(x).$$

*Solution.* To determine whether a function has any of these symmetries, substitute $-x$ into its argument and see if you can relate it to the original function. For $f(x)$ we have

$$f(-x) = (-x)^2 = x^2 = f(x)$$

implying that $f(x)$ is even. For $g(x)$ we have

$$g(-x) = (-x)^2 = -x^3 = -g(x)$$

implying that $g(x)$ is odd. Finally, for $h(x)$ we have

$$h(-x) = f(-x) + g(-x) = f(x) - g(x).$$

However, there is no natural way to relate $f-g$ to $f+g$ by using only a single minus sign. Hence $h(x)$ is neither even nor odd. ■

### 1.2.3 Roots

Mathematically, the number 0 is one of the most interesting (and troublesome) numbers. The student is likely familiar with the fact that $0 \times a = 0$ and $0 + a = a$ for any value of $a$, and that division by 0 is strictly prohibited. Hence it is unsurprising that we are often interested in the places where functions $f : \mathbb{R} \to \mathbb{R}$ attain the value 0.

**Definition 1.6**

Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that $\alpha \in \mathbb{R}$ is a *root* of $f(x)$ if $f(\alpha) = 0$.

Geometrically, roots correspond to the places at which the graph of a function passes through the $x$-axis.

**Example 1.7**

Find the roots of the functions

$$f_1(x) = x - 5, \qquad f_2(x) = 0, \qquad f_3(x) = \frac{1}{x}.$$

*Solution.* We begin by looking at $f_1(x)$. We want to find values $\alpha$ for which $f_1(\alpha) = \alpha - 5 = 0$. We may simply solve this equation for $\alpha$ to find that $\alpha = 5$. This is the only possible root of $f_1(x)$. It is easy to see that given any function of the form $g(x) = x - r$, the root of $g(x)$ will be $r$.

For $f_2(x)$, we want the collection of $\alpha$ satisfying $f_2(\alpha) = 0$. Since $f_2$ is just the function which sends everything to zero, it turns out that every real number is a root of $f_2(x)$. This turns out to be clear when we realize that the graph of $f_2(x)$ is just the $x$-axis itself.

Finally, for $f_3(x)$ we want $\alpha$ such that $f_3(\alpha) = 1/\alpha = 0$. In order to solve this equation for $\alpha$, we would need to take a reciprocal of both sides, but this would require us to divide by zero! Hence $1/\alpha = 0$ has no solutions, implying that $f_3(x)$ has not roots. Again, try plotting $f_3(x)$ and this will become obvious. ■

### 1.2.4 Piecewise Functions

Piecewise functions are described by gluing together two functions to form a new one. For example, if $g(x)$ and $h(x)$ are two functions and $a \in \mathbb{R}$, we may define

$$f(x) = \begin{cases} g(x) & x \leq a \\ h(x) & x > a \end{cases}.$$

This means that if $x \leq a$ then $f(x) = g(x)$ and if $x > a$ then $f(x) = h(x)$.

**Example 1.8**

Graph the function $f(x) = \begin{cases} x^2 & x \leq 0 \\ 3 & x > 0 \end{cases}$.

*Solution.* The graphs of $y = x^2$, $y = 3$, and $f(x)$ are given in Figure 5. It is our hope that this illustrates the idea of a piecewise function in terms of cutting and pasting; namely, we cut the graphs of $x^2$ and 3 at the line $x = 0$ and then re-attach them in the way described by $f(x)$. ■

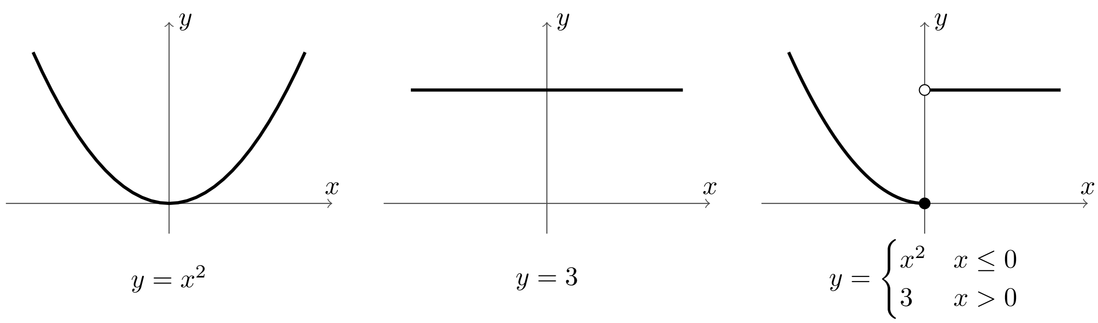



Figure 5: A piecewise function is a way of gluing two functions together to form a new function. This figure illustrates how we have taken the functions $y = x^2$ and $y = 3$, cut each along the line $x = 0$, and then glued them together to get the function $f(x)$. Notice the exaggerated hole in $f(x)$, used to indicate that the value 3 is not actually attained at $x = 0$.

While our examples above utilized two functions, there is no limit on the number of functions which we may splice together, so long as that number is finite. For example, the following piecewise function has four components:

$$f(x) = \begin{cases} 2 - x^2 & x < 1 \\ 0 & x = 1 \\ 4 - x & 1 < x < 2 \\ 1/x & x > 2. \end{cases}$$

Try graphing this piecewise function.

### 1.2.5 Inverse Functions

The word "inverse" has many different meanings depending on the context in which it is used. For example, what if we were to ask the student to find the inverse of the number 2? What does this mean? To what are we taking the inverse? To properly understand this, we need to understand the following: Given a binary operator (an operator which takes in two things and produces a single thing in return, such as addition and multiplication), we say that a number id is the *identity* of that operator if operating against it does nothing to the input. For example, in the case of addition, the operator will satisfy $x + \mathrm{id}_+ = x$ for all possible $x$; for example,

$$2 + \mathrm{id}_+ = 2, \qquad -5 + \mathrm{id}_+ = -5.$$

Our experience tells us that $\mathrm{id}_+ = 0$. Similarly, for multiplication the identity $\mathrm{id}_\times$ will satisfy $x \times \mathrm{id}_\times = x$ for all $x$; for example,

$$3 \times \mathrm{id}_\times = 3, \qquad \pi \times \mathrm{id}_\times = \pi.$$

Again our experience tells us that $\mathrm{id}_\times = 1$. We thus say that 0 is the additive identity and 1 is the multiplicative identity. We say that the *inverse of* $x$ is an element which, when paired against $x$, gives the identity. Hence the additive inverse of 2 is the number $y$ such that $2 + y = \mathrm{id}_+ = 0$, or rather $-2$. In general, the additive inverse of $n$ is $-n$, and this always exists! For multiplication, it is not too hard to convince ourselves that the multiplicative inverse of $x$ is $\frac{1}{x}$; for example, $2 \times \frac{1}{2} = 1 = \mathrm{id}_\times$. Notice that there is no multiplicative inverse for the number 0, so in this case the inverse does not always exist.

Function composition $f \circ g$ is another example of a binary operator. What is the identity for this operation? Well, we would like a function $\mathrm{id}_\circ$ such that

$$\begin{aligned} f(\mathrm{id}_\circ(x)) &= f(x) \\ &= \mathrm{id}_\circ(f(x)). \end{aligned}$$

If we this about this for a moment, the identity function is the function $\mathrm{id}_\circ(x) = x$, the function which does nothing to the argument! Now what is the inverse of a function? The inverse of a function $f(x)$ is a <u>function</u> $f^{-1}(x)$ such that $f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}_\circ$.

To compute the inverse of $y = f(x)$, notice that by applying $f^{-1}$ to both sides we get

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

Hence by switching $x$ and $y$ and solving for $y$, we get $y = f^{-1}(x)$.

**Example 1.9**

Determine the inverse of the function $y = f(x) = (x-1)/(x+1)$.

*Solution.* As recommended above, we interchange $y$ and $x$ and solve for $y$, so we get

$$\begin{aligned} x = \frac{y-1}{y+1} \quad &\Leftrightarrow \quad (y+1)x = y - 1 \\ &\Leftrightarrow \quad yx - y = -(x+1) \\ &\Leftrightarrow \quad y(x-1) = -(x+1) \\ &\Leftrightarrow \quad y = \frac{x+1}{1-x} \end{aligned}$$

So $f^{-1}(x) = (x+1)/(1-x)$. Indeed we can check this by composing $f \circ f^{-1}$ and $f^{-1} \circ f$ to find that

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

and the other direction is left as an exercise. ■

Note that not all functions are invertible. For example, the function $f(x) = x^2$ is not invertible in general. It is tempting to say that $g(x) = \sqrt{x}$ is the inverse to $f$, but this is not the case. Indeed, while we do have that

$$(f \circ g) = \left(\sqrt{x}\right)^2 = x,$$

the opposite composition gives

$$(g \circ f)(x) = \sqrt{x^2} = |x|$$

which is not the identity function. To test whether a function can be inverted, it must satisfy the *horizontal line test*; that is, every horizontal line must intersect the graph of $f$ in at most one place.

**Exercise**

Determine the inverses of each of the following functions. What special property do $f, g$, and $h$ all share?

$$f(x) = \frac{1}{x}, \qquad g(x) = 1 - x, \qquad h(x) = \frac{x}{x-1}.$$

## 1.3 Polynomials and Rational Functions

Polynomials are the collection of all objects of the form

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

for some natural number $n > 0$ and real numbers $a_0, a_1, \ldots, a_n$. We say that the *degree* of a polynomial is the highest power whose coefficient is non-zero. For example, the following functions are polynomials

$$p(x) = 3x^4 + 8x^3 - 2x, \qquad q(x) = 39x^{66} - 5x^2 + 1$$

and the degree of $p(x)$ is 4 while the degree of $q(x)$ is 66. Some degrees occur so frequently that they even have special names:

| Degree | Name |
|---|---|
| 1 | Linear |
| 2 | Quadratic |
| 3 | Cubic |
| 4 | Quartic |
| 5 | Quintic |

Factoring polynomials is the process by which we reverse the act of multiplying a polynomial: We would like to write a single polynomial as a product of polynomials with strictly smaller degree.

There are some very easy factorization which involve simply removing a power of $x$. If there is no constant term (the coefficient of $x^0$ is 0), then we may remove at least one power of $x$ from the polynomial. For example,

$$x^4 + x^2 = x^2(x^2 + 1), \qquad x^5 + x^4 + x = x(x^4 + x^3 + 1).$$

In general, factoring polynomials with constant terms can be incredibly difficult. For most purposes however, we may limit ourselves to factoring quadratic polynomials. Given a quadratic polynomial of the form $x^2+ax+b$, the trick is to try and find two numbers $p, q$ such that $a = p+q$ and $b = pq$. This is because

$$(x+p)(x+q) = x^2 + (p+q)x + pq.$$

**Example 1.10**

Factor the following polynomials:

$$x^2 + 2x + 1, \qquad 3x^2 + 15x + 18, \qquad x^2 - 1, \qquad x^3 - x^2 - 2x.$$

*Solution.* We begin with $x^2 + 2x + 1$. To factor this, we try to think of two numbers $p, q$ such that $p + q = 2$ and $pq = 1$. Hopefully, the choice $p = 1, q = 1$ springs to our minds and we guess $(x + 1)(x + 1) = x^2 + 2x + 1$. A quick check verifies that this is the case.

For $3x^2 + 15x + 18$ we are not quite in the situation described above as the coefficient in front of $x^2$ is not 1. However, we may first factor out a 3 to get $3x^2 + 15x + 18 = 3(x^2 + 5x + 6)$. Now we would like to find $p, q$ such that $p + q = 5$ and $pq = 6$. The choice $p = 2$ and $q = 3$ jumps to mind, and a quick calculation verifies that $(x + 2)(x + 3) = x^2 + 5x + 6$. Thus

$$3x^2 + 15x + 18 = 3(x + 2)(x + 3).$$

The polynomial $x^2 - 1$ look tricky: what do we do if we have no $x$ term? Instead of panicking, let's try our usual technique; that is, find $p, q$ such that $p + q = 0$ and $pq = -1$. We could actually solve this equation, or just guess that $p = 1$ and $q = -1$ work. Indeed, it turns out that $x^2 - 1 = (x + 1)(x - 1)$.

Finally, $x^3 - x^2 - 2x$ is not a quadratic polynomial. However, the lack of a constant term means we can first factor out an $x$ term to get $x^3 - x^2 - 2x = x(x^2 - x - 2)$. We hence content ourselves to find $p, q$ such that $p + q = -1$ and $pq = -2$. This one is a bit tricky, but some though reveals that $p = -2$ and $q = 1$ will do the trick, and indeed $(x - 2)(x + 1) = x^2 - x - 2$ so that

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

Rational functions are quotients of polynomials; that is, they are functions which can be written as $f(x) = p(x)/q(x)$ where $p(x)$ and $q(x)$ are both polynomials. The following are examples of rational functions:

$$f(x) = \frac{x^2 + 2x + 1}{x - 1}, \qquad g(x) = \frac{1}{x^2 + 1}, \qquad h(x) = \frac{x^3 + 2x - 1}{4x^4 - x^2 + 13}.$$

## 1.4 Absolute Values

One of the most important concepts in mathematics is that of length, historically motivated by applications and the Greek obsession with compasses and rulers. One finds that there are a hierarchy of structures, each more powerful than the next, that endow a space with a measure of length: metrics, norms, and inner products.

We shall not study such structures in this course, but the student who ventures into the study of linear algebra will quickly find him/herself acquainted with norms and inner products.

### 1.4.1 The Absolute Value

Given a number $x \in \mathbb{R}$ we would like to discuss its "distance" from the number 0. Naively, we would like to say something along the lines of "4 is the same distance from 0 as $-4$" or perhaps "2 is the same distance from 4 as $-1$ is from $-3$." Figure 6 illustrates this idea.

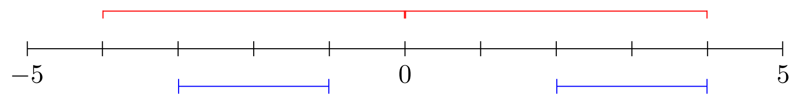



Figure 6: The real line from $-5$ to positive 5. We would like to define a system of measurement such that the red bars have the same length and the blue bars have the same length.

The formal way to talk about the concept of length is with absolute values:

**Definition 1.11**

For $x \in \mathbb{R}$ we define the *absolute value* of $x$ as

$$|x| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}.$$

Notice that the absolute value is always positive: If $x$ is already positive, the absolute value does not do anything, while if $x$ is negative we negate it again to make it positive. Geometrically, we may interpret $|x|$ as the distance from $x$ to 0. The distance from 4 to 0 is $|4| = 4$ while the distance from $-4$ to 0 is $|-4| = -(-4) = 4$. As we discussed above, this is precisely what we expected.

**Proposition 1.12: Properties of the Absolute Value**

If $a, b \in \mathbb{R}$ then

1. $|ab| = |a||b|$ (Multiplicative)
2. $|a + b| \leq |a| + |b|$ (Triangle Inequality)
3. $|a| = 0$ if and only if $a = 0$ (Non-degenerate).

### 1.4.2 Relation to Intervals

Instead of looking at the distance from $a$ to 0, we can look at the distance between $a$ and $b$, given by $|a - b|$. Therefore, we may use absolute values combined with inequalities to describe intervals. For example, consider the statement $|x - c| < a$. Using the definition of the absolute value, we can write this as

$$|x - c| = \begin{cases} x - c & x \geq c \\ c - x & x < c \end{cases}.$$

Now $|x - c| < a$ implies that both $x - c < a$ and $c - x < a$ for all values of $x$. If we multiply $c - x < a$ by $-1$ we get $x - c > -a$, which we may combine with $x - c < a$ to conclude that

$$|x - c| < a \quad \Longleftrightarrow \quad -a < x - c < a.$$

We may read $|x - c| < a$ geometrically as

"The distance from $x$ to $c$ is less than $a$."

Intuitively, the set of all $x$ which satisfy this will lie in the interval $(c-a, c+a)$. We can show this more concretely by realizing that

$$\begin{aligned} |x-c| < a \quad &\iff \quad -a < x-c < a \\ &\iff \quad c-a < x < c+a \end{aligned} \tag{1.1}$$

**Example 1.13**

Find the intervals corresponding to all $x$ which satisfy the following inequalities:

$$|x| \leq 1, \qquad |2x-5| < 3, \qquad |x+7| > 5.$$

*Solution.* If $|x| \leq 1$ then $-1 \leq x \leq 1$ and this corresponds to the interval $[-1, 1]$. The next example is $|x-2| < 3$ and proceeding by the same argument in (1.1) we find that

$$\begin{aligned} |2x-5| < 3 \quad &\iff \quad -3 < 2x-5 < 3 \\ &\iff \quad 2 < 2x < 8 \\ &\iff \quad 1 < x < 4 \end{aligned}$$

so that the corresponding interval is $(1, 4)$.

Expressions of the form $|x+7| > 5$ will occur far less frequently than the examples considered above, but should still be solvable if we go back to the definition of absolute value. Intuitively we see that the $x$ which satisfy this will be a distance of at least 5 from $-7$; that is, $(-\infty, -12) \cup (-2, \infty)$. Let us check that this is the case.

The condition that $|x+7| > 5$ implies that both $x+7 > 5$ and $-x-7 > 5$. Solving the former for $x$ we find that $x > -2$ while the latter reveals that $x < -12$, precisely as we expected. ■

### 1.4.3 Algebra with Inequalities

Students often have troubles with absolute values, especially when encountering them while solving for a variable. Whenever absolute values are encountered, the best strategy in each case is to remove the absolute values by considering cases in which the absolute values can be removed.

**Example 1.14**

Find all $x$ for which $|x+7| < 4x+10$.

*Solution.* The equation $|x+7| < 4x+10$ is untenable in this form, so we break it into the case $x < -7$ where $|x+7| = -x-7$ and $x \geq -7$ where $|x+7| = x+7$.

**Case $x < -7$:** If we restrict ourselves to $x < -7$ then $|x+7| < 4x+10$ becomes

$$-x-7 < 4x+10.$$

Some quick algebraic manipulation shows us that $x > -\frac{17}{5}$, which combined with $x < -7$ tells us that no $x$ satisfy this equation.

**Case $x \geq -7$:** In this case $|x+7| < 4x+10$ becomes $x+7 < 4x+10$. Some algebraic work shows us that $x > -1$. Hence both $x > -1$ and $x \geq -7$ implies that $x > -1$ is the solution.

Combining the results from both cases, we see that $|x+7| < 4x+10$ if $x > -1$; that is, $x \in (-1, \infty)$. ■

**Example 1.15**

Find all $x$ for which
$$|x-3| \geq |x+1| - 2. \tag{1.2}$$

*Solution.* The expression $|x-3|$ will change signs at $x = 3$ while $|x+1|$ will switch signs at $x = -1$. This implies that we should consider three cases: $x < -1$, $-1 < x < 3$, and $x > 3$.

**Case $x < -1$:** Equation (1.2) becomes
$$-x+3 \geq -x-1-2.$$

The $x$'s will actually cancel giving the expression $3 \geq -3$, which is clearly true, so $x < -1$ always satisfies the equation.

**Case $-1 < x < 3$:** In this case equation (1.2) becomes
$$-x+3 \geq x+1-2$$

which is solved to find $x \leq 2$. Hence $x$ must satisfy both $-1 < x < 3$ and $x \leq 2$ implying that $-1 < x \leq 2$.

**Case $x > 3$:** Now equation (1.2) becomes
$$x-3 \geq x+1-2$$

which yields $-3 \geq -1$, a false expression. This means that no $x$ in this region satisfies the equation.

Finally, we check the switch points $x = -1, 3$ themselves. Substituting $x = -1$ into (1.2) we get
$$|(-1)+3| \geq |(-1)+1| + 2 \quad \Rightarrow \quad 2 \geq 2$$

which is true, so that $-1$ satisfies the equation. On the other hand, $x = 3$ yields
$$|3-3| \geq |3+1| + 2 \quad \Rightarrow \quad 0 \geq 6$$

which is not true, so $x = 2$ does not satisfy the equation. Combining all of our information, the total solution is
$$\{x < -1\} \cup \{-1 < x \leq 2\} \cup \{-1\} = \{x \leq 2\}$$

or more concisely, the interval $(-\infty, 2]$ ■

## 1.5 Trigonometric Functions

Trigonometry seems to be ubiquitously alluded to as the bane of every secondary school student, but it should not be feared. A solid foundation and understanding of some simple trigonometry will be essential for the study of calculus as well as crucial for many applications in industry.

In case the student is unfamiliar, we recall that trigonometric functions arise in the study of angles, and in particular we typically associate them with triangles.

### 1.5.1 Degrees versus radians

When one is first introduced to angles, one typically learns that we measure them in *degrees*; namely, we say that a full rotation of a circle is equal to 360° and then divide the circle into 360 equal portions. The angle made by each of these portions is a single degree.

The choice of the number 360 is an artifact of the ancient Babylonians who worked in base 60. A "metric" version of an angle wherein a right-angle is assigned a unit of 100 might seem more natural and gives rise to a unit called a *gradian*. Again though, even a metric system is based off of the (ostensibly) arbitrary fact that modern society uses base 10. Mathematically, we would like to find an angle measurement which is intrinsic to the circle, and this is where the idea of a *radian* is born.

The definition of the number $\pi$ is that it is the ratio of a circle's circumference to its diameter; that is, a circle of diameter $d$ has circumference $c = \pi d$. If $r$ is the radius of the circle then $d = 2r$ and the circumference may be written as $c = 2\pi r$. Given a circle of radius $r$, we define a radian to be the angle subtended by an arc of length $r$. It should be clear that this is intrinsic to the circle, does not depend on $r$, and is free of arbitrary choices such as those made in defining degrees and gradians.

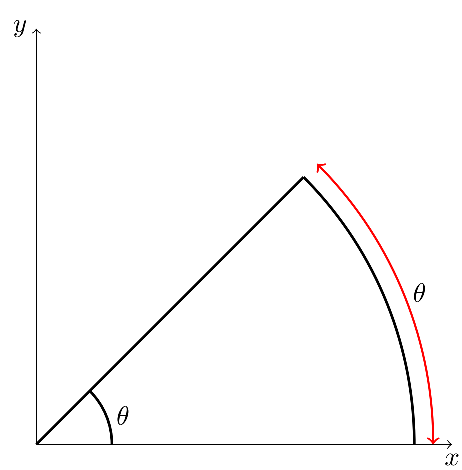



Figure 7: The definition of a radian is that the angle is the same as the length of the arc of the unit circle it subtends.

So how many radians are contained in a full rotation of the circle? The solution is the number of times an arc of length $r$ divides the circumference $c$; namely, $\frac{c}{r} = 2\pi$. In fact, we already have enough tools to discuss the length of a circular sector.

**Proposition 1.16**

Consider a circle of radius $r$ and a circular sector subtended by an angle of $\theta$. The outer perimeter $s$ of the sector is given by the formula $s = r\theta$.

*Proof.* The idea is as follows: we know that the sector given by an angle to $2\pi$ is the whole circle, giving a sector length of $c = 2\pi r$. If $0 < \theta < 2\pi$ then the sector subtended by the angle $\theta$ will have a sector length equal to the ratio of $\theta$ to the whole circle, $2\pi$. Hence

$$s = \frac{2\pi r}{\frac{2\pi}{\theta}} = 2\pi r \frac{\theta}{2\pi} = r\theta. \qquad \square$$

It is essential at this point to remark that the usual formulas derived and used in the study of calculus **assume the use of radians.** The wayward student who attempts to use the theory of calculus to make computations, but substitutes degrees instead of radians, may find themselves sending lunar rockets to the sun instead of the moon. As such, it is necessary to discuss the conversion between radians and degrees.

There are $\pi$ radians for every 180 degrees, so we may convert between the two via the following formulas:

$$\text{degrees} = \frac{180}{\pi} \times \text{radians}, \qquad \text{radians} = \frac{\pi}{180} \times \text{degrees}.$$

These formulas are easy to remember if the student thinks about "canceling" the units, demonstrated in the following example.

**Example 1.17**

Convert $\frac{\pi}{3}$ radians to degrees, and 45° to radians.

*Solution.* We use our formula to compute the conversions. We have

$$\begin{aligned} \text{degrees} &= \frac{180 \text{ degrees}}{\pi \cancel{\text{ radians}}} \times \frac{\pi}{3} \cancel{\text{ radians}} \\ &= \frac{180}{3} \text{ degrees} \\ &= 60 \text{ degrees}. \end{aligned}$$

Similarly converting from degrees to radians we have

$$\begin{aligned} \text{radians} &= \frac{\pi \text{ radians}}{180 \cancel{\text{ degrees}}} \times 45 \cancel{\text{ degrees}} \\ &= \frac{45\pi}{180} \text{ radians} \\ &= \frac{\pi}{4} \text{ radians}. \end{aligned} \qquad \blacksquare$$

### 1.5.2 Trigonometric Functions

The trigonometric functions describe the ratio of lengths of a right-triangle and there are three with which we will primarily concern ourselves. Consider the right-triangle given in Figure 8:

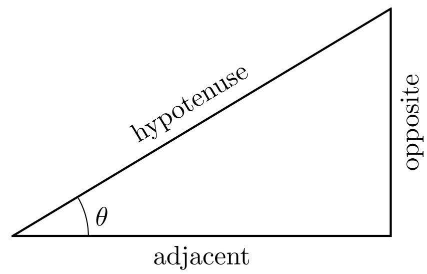



Figure 8: Given a vertex of a right triangle which is not the right-angle, we define the opposite and adjacent edges of that vertex.

Given an angle $\theta$ which is not the right-angle[1], we define the *hypotenuse* as the edge which has the longest length, the *adjacent* edge as the edge which forms the angle $\theta$ but is not the hypotenuse, and the *opposite* edge to be the remaining edge. Our trigonometric functions are the ratios of these edges as follows:

**Definition 1.18**

Given a right angle triangle and an angle $\theta$ as in Figure 8, we define

$$\sin(\theta) = \frac{\text{opposite}}{\text{hypotenuse}}, \quad \cos(\theta) = \frac{\text{adjacent}}{\text{hypotenuse}}, \quad \tan(\theta) = \frac{\text{opposite}}{\text{adjacent}}.$$

We say that $\sin(\theta)$ is the *sine* function, $\cos(\theta)$ is the *cosine* function, and $\tan(\theta)$ is the *tangent* function.

Given an arbitrary value of $\theta$, an analytic expression for $\sin(\theta)$ or $\cos(\theta)$ is often not attainable. However there are a few special triangles which are of particular importance as the trigonometric functions evaluate to nice expressions. These are shown in Figure 9.

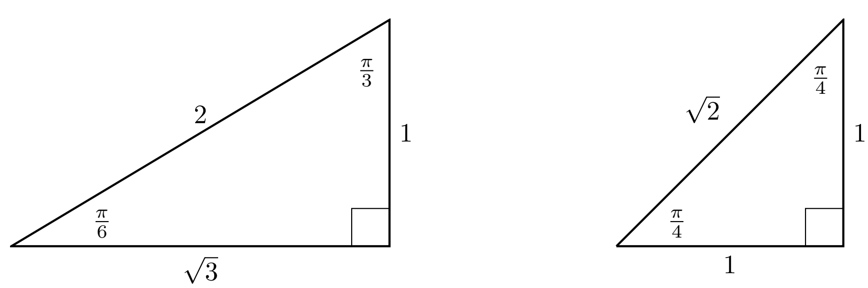



Figure 9: Two special triangles which allow for simple explicit expressions of the trigonometric functions.

[1] A right angle is an angle of $\pi/2$ radians or $90^\circ$.

**Example 1.19**

Find the values of $\sin\left(\frac{\pi}{3}\right)$ and $\cos\left(\frac{\pi}{4}\right)$.

*Solution.* Using Definition 1.18 and the special triangles shown in Figure 9, we can easily read off

$$\sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}, \qquad \cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}. \qquad \blacksquare$$

### 1.5.3 Between circles and triangles

It was mentioned in the beginning of this section that we associate trigonometric functions with triangles, while our discussion concerning radians concerns circles. What is the relationship between them?

The circle, despite being drawn in the two-dimensional plane, is really only a one dimensional object. Indeed, if you lived on a circle you could only move forwards or backwards, hence one dimension. This means that mathematically we should be able to write points on the circle in terms of a single parameter.

Our goal is to write a point $(x, y)$ on the unit circle purely in terms of the angle $\theta$ made by the positive $x$-axis and the line connecting $(x, y)$ to the origin. If the radius of the circle is $r$, Definition 1.18 tells us that $x = r\cos(\theta)$ and $y = r\sin(\theta)$. The Pythagorean theorem says that $x^2 + y^2 = r^2$, so in turn we have the identity[2]

$$r^2 \sin^2(\theta) + r^2\cos^2(\theta) = r^2.$$

As $r \neq 0$ we may divide through to get the *Pythagorean trigonometric identity*

$$\sin^2(\theta) + \cos^2(\theta) = 1. \tag{1.3}$$

This equation can be seen directly by setting the radius of the circle to be $r = 1$ and is shown in Figure 10.

Measuring angles on the circle gives us a useful technique for determining the sign of a trigonometric function. For example, consider the values

$$\sin\left(\frac{\pi}{6}\right), \quad \sin\left(\frac{5\pi}{6}\right), \quad \sin\left(\frac{7\pi}{6}\right), \quad \sin\left(\frac{11\pi}{6}\right).$$

Figure 11 shows the angles as inscribed in a circle. The hypotenuse (corresponding to the radius of the circle) is the same for each triangle drawn, and the length of the opposite angle is also the same. However, notice that in the quadrants I and II the opposite length (corresponding to the

[2] The notation $\sin^2(x)$ can be confusing, but this is how the expression $(\sin(x))^2$ is always denoted; that is, $\sin^2(x) = \sin(x)\sin(x)$. To exacerbate matters, many authors use $f^2(x)$ to denote the two-fold composition $f(f(x))$, but this is not what is meant with trigonometric functions.

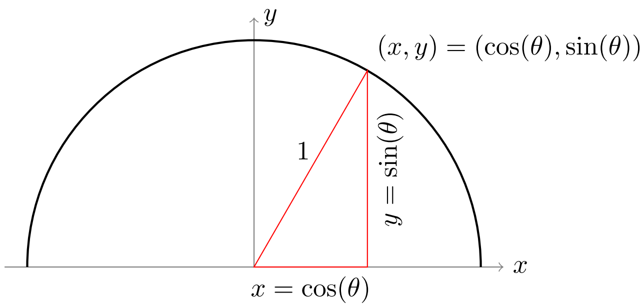



Figure 10: The angle $\theta$ made between the positive $x$-axis and a arc-of-radius of the circle gives us a way of writing each point on the circle in terms of a single parameter $\theta$.

$y$-value) is positive, while in quadrants III and IV the opposite angle is negative. This tells us that the sine of $\frac{5\pi}{6}$ is the same as $\frac{\pi}{6}$, while the sine of $\frac{7\pi}{6}$ and $\frac{11\pi}{6}$ are the negative of $\frac{\pi}{6}$:

$$\sin\left(\frac{\pi}{6}\right) = \sin\left(\frac{5\pi}{6}\right) = \frac{1}{2}, \qquad \sin\left(\frac{7\pi}{6}\right) = \sin\left(\frac{11\pi}{6}\right) = -\frac{1}{2}.$$

More generally, this argument shows that $\sin(\theta)$ is always positive for $0 < \theta < \pi$ and negative for $\pi < \theta < 2\pi$.

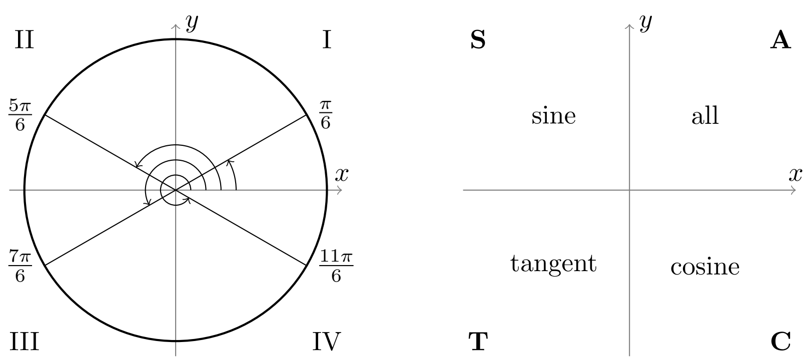



Figure 11: *Left:* Angles inscribed in a circle. Notice that $\frac{5\pi}{6}$ is really just an angle of $\frac{\pi}{6}$ but with respect to the negative $x$-axis, while $\frac{7\pi}{6}$ and $\frac{11\pi}{6}$ are angles of $-\frac{\pi}{6}$ with respect to the negative and positive $x$-axes respectively. *Right:* A mnemonic for remembering which trigonometric functions are positive on which quadrant.

Similar arguments can be made for the cosine and tangent functions and are left to the student as exercises. However, the right image in Figure 11 provides a useful mnemonic for remembering where each trigonometric function is positive. The 'A' in quadrant I stands for **All**, meaning that all functions are positive in that quadrant. The 'S','C', and 'T' stand for **Sine**, **Cosine**, and **Tangent** respectively and indicate the quadrants where those functions are positive. We shall refer to this as the "CAST" diagram.

**Example 1.20**

Determine the values of the tangent function evaluated on $\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$.

*Solution.* The angle $\frac{\pi}{4}$ refers to one of the special triangles, so we immediately gauge that $\tan\left(\frac{\pi}{4}\right) = 1$. To determine the other three angles we refer to Figure 11. The figure tells us that the tangent function is positive in quadrants I and III and negative everywhere else indicated that

$$\tan\left(\frac{\pi}{4}\right) = \tan\left(\frac{5\pi}{4}\right) = 1, \qquad \tan\left(\frac{3\pi}{4}\right) = \tan\left(\frac{7\pi}{4}\right) = -1. \qquad \blacksquare$$

### 1.5.4 The Reciprocal Trigonometric Functions

There are in fact many trigonometric functions relating circles and triangles as introduced by the ancient Greeks. The majority of these functions no longer appear in the literature, with the exception of the following three:

$$\csc(\theta) = \frac{1}{\sin(\theta)}, \qquad \sec(\theta) = \frac{1}{\cos(\theta)}, \qquad \cot(\theta) = \frac{1}{\tan(\theta)} \tag{1.4}$$

referred to as the *cosecant*, *secant* and *cotangent* functions. Classically, these functions were defined via their geometric interpretations, shown in Figure 12, but are now often just referred to as the reciprocal trigonometric functions.

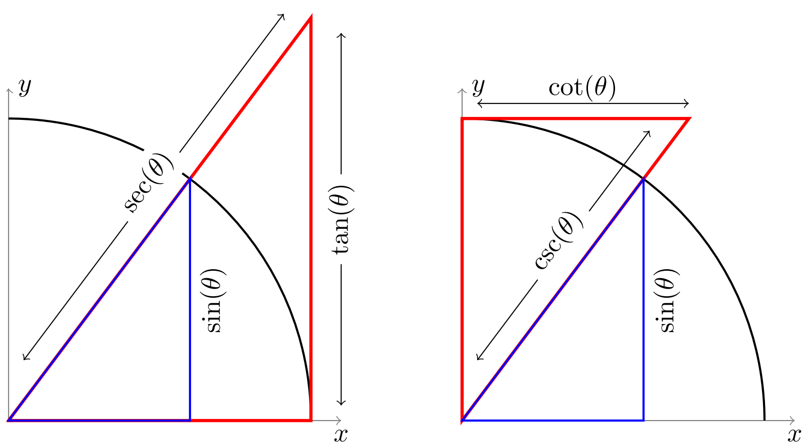



Figure 12: The geometric definition of the (reciprocal) trigonometric functions.

**Example 1.21**

If $\sin(\theta) = 1/2$ and $\cos(\theta) = -1/4$ compute $\csc(\theta), \sec(\theta)$ and $\cot(\theta)$.

*Solution.* The expressions for $\csc(\theta)$ and $\sec(\theta)$ are as simple as directly exploiting equation (1.4). Indeed,

$$\csc(\theta) = \frac{1}{\sin(\theta)} = \frac{1}{1/2} = 2, \qquad \sec(\theta) = \frac{1}{\cos(\theta)} = \frac{1}{-1/4} = -4.$$

For cotangent, we may exploit any of the following equivalent expressions:

$$\cot(\theta) = \frac{1}{\tan(\theta)} = \frac{\cos(\theta)}{\sin(\theta)} = \frac{\csc(\theta)}{\sec(\theta)}.$$

Doing this we find $\cot(\theta) = -\frac{1}{2}$. $\blacksquare$

Expanding trigonometric function across addition is not as well behaved as that of polynomials, though with some geometric reasoning we may still write the sum and differences analytically as follows:

**Proposition 1.22**

If $\theta, \psi$ are two angles, then sine and cosine satisfy the following *angle sum* identities:

$$\sin(\theta+\psi) = \sin(\theta)\cos(\psi) + \sin(\psi)\cos(\theta), \tag{1.5a}$$
$$\sin(\theta-\psi) = \sin(\theta)\cos(\psi) - \sin(\psi)\cos(\theta), \tag{1.5b}$$
$$\cos(\theta+\psi) = \cos(\theta)\cos(\psi) - \sin(\theta)\sin(\psi), \tag{1.5c}$$
$$\cos(\theta-\psi) = \cos(\theta)\cos(\psi) + \sin(\theta)\sin(\psi). \tag{1.5d}$$

The proof of this fact is rather technical and not particularly enlightening.

**Example 1.23**

Find an algebraic expression for sine and cosine evaluated upon $\frac{\pi}{12}$.

*Solution.* Unlike the angles corresponding to $\frac{2\pi}{12} = \frac{\pi}{6}$, $\frac{3\pi}{12} = \frac{\pi}{4}$, and $\frac{3\pi}{12} = \frac{\pi}{3}$, we cannot simply read off the corresponding sine and cosine values from special triangles. However, we cleverly realize that we may write $\frac{\pi}{12}$ in terms of the angles which appear in the special triangles as follows:

$$\frac{\pi}{12} = \frac{\pi}{3} - \frac{\pi}{4}.$$

From here the angle sum identities can be exploited to compute $\sin\left(\frac{\pi}{12}\right)$ and $\cos\left(\frac{\pi}{12}\right)$. We begin by using Equation (1.5b) to see that

$$\begin{aligned}\sin\left(\frac{\pi}{12}\right) &= \sin\left(\frac{\pi}{3} - \frac{\pi}{4}\right) \\ &= \sin\left(\frac{\pi}{3}\right)\cos\left(\frac{\pi}{4}\right) - \sin\left(\frac{\pi}{4}\right)\cos\left(\frac{\pi}{3}\right) \\ &= \left(\frac{\sqrt{3}}{2}\right)\left(\frac{1}{\sqrt{2}}\right) - \left(\frac{1}{\sqrt{2}}\right)\left(\frac{1}{2}\right) \\ &= \frac{\sqrt{3}-1}{2\sqrt{2}},\end{aligned}$$

while Equation (1.5d) reveals

$$\begin{aligned}\cos\left(\frac{\pi}{12}\right) &= \cos\left(\frac{\pi}{3} - \frac{\pi}{4}\right) \\ &= \cos\left(\frac{\pi}{3}\right)\cos\left(\frac{\pi}{4}\right) + \sin\left(\frac{\pi}{4}\right)\sin\left(\frac{\pi}{3}\right) \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{2}}\right)\left(\frac{\sqrt{3}}{2}\right) \\ &= \frac{\sqrt{3}+1}{2\sqrt{2}}.\end{aligned}$$

These expression are exact, but are rather unwieldy to work with and might have been hard to deduce without the angle sum identities. ■

The trigonometric functions have evolved beyond their simple geometric interpretation and are now ubiquitous throughout mathematics. They satisfy some very appealing properties such as[3]:

- [Translation:] Cosine and sine are related to one another through simple shifts:
$$\cos\left(\frac{\pi}{2} - \theta\right) = \sin(\theta), \qquad \sin\left(\frac{\pi}{2} - \theta\right) = \cos(\theta).$$
- [Periodicity:] For every $\theta \in \mathbb{R}$ we have
$$\sin(\theta + 2\pi) = \sin(\theta), \qquad \cos(\theta + 2\pi) = \cos(\theta).$$
- [Symmetry] Cosine is an even function and sine is an odd function:
$$\cos(-\theta) = \cos(\theta), \qquad \sin(-\theta) = -\sin(\theta).$$

Combining all of this information allows us to easily discern the graphs of $\sin(\theta)$ and $\cos(\theta)$ shown in Figure 13. The graph of $\tan(\theta)$ is somewhat more difficult to discern. Our knowledge of $\tan(\theta)$ tells us that the plot will also have singularities wherever $\cos(\theta) = 0$, but the graph also reveals that the tangent function is actually $\pi$-periodic.

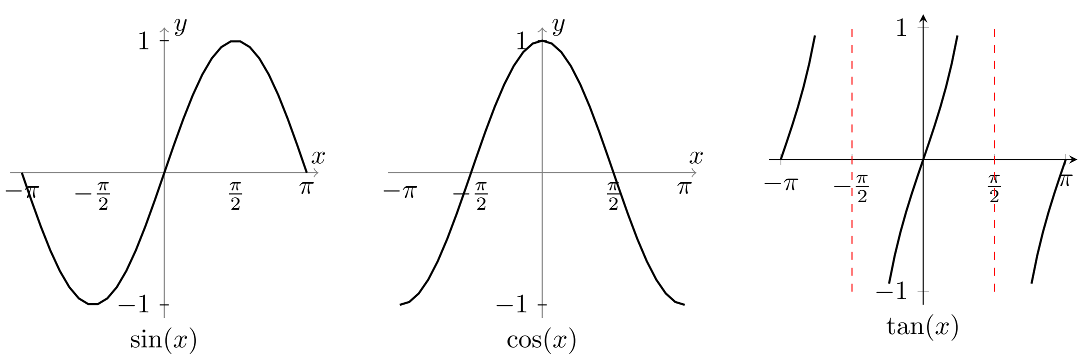



Figure 13: The graphs of the trigonometric functions sine, cosine, and tangent.

The functions $\sec(\theta)$ and $\csc(\theta)$ are the reciprocals of $\cos(\theta)$ and $\sin(\theta)$ and so also share the above properties. They are plotted together with $\cot(\theta)$ in Figure 14. We are not surprised to see that $\cot(\theta)$ is again $\pi$-periodic, just like $\tan(\theta)$. Notice that the plots of tangent and cotangent look very similar: indeed they are the same up to a translation by $\frac{\pi}{2}$ and a reflection about the $x$-axis.

## 1.6 Exponential Functions

As multiplication was motivated as a tactic for abbreviating $n$-fold sums, exponentiation was originally shorthand for $n$-fold products. That is, if $a$ is a real number and $n$ is a natural number, then

[3]The student will verify these facts (and many more) in guided exercises at the end of the section.

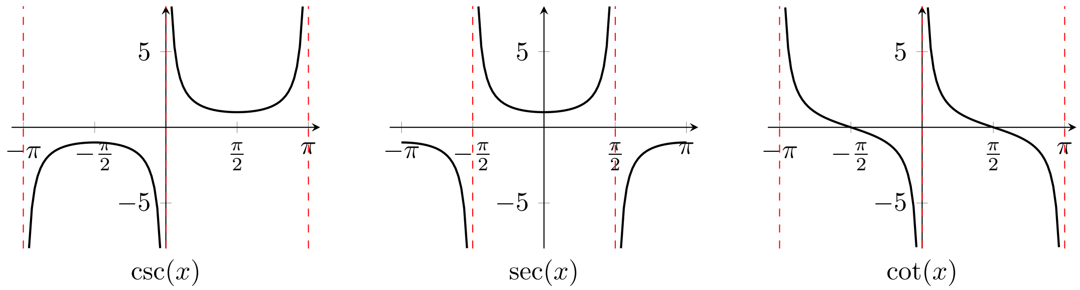



Figure 14: The graphs of the reciprocal trigonometric functions cosecant, secant, and cotangent.

we define $a^n$ as

$$\underbrace{a \times a \times \cdots \times a}_{n\text{-times}} = a^n. \tag{1.6}$$

As in the case of multiplication, we then define the exponent for negative numbers; $a^{-n} = 1/a^n$. Exponentiation then satisfies the following rules: If $a, b$ are real numbers and $n, m$ are integers, then

$$a^n a^m = a^{n+m}, \qquad (a^n)^m = a^{nm}, \qquad (ab)^n = a^n b^n.$$

### 1.6.1 Roots

In order to extend the idea of multiplication to rational numbers we exploited the notion of division, which is the "inverse" to multiplication. We will have to do something similar in order to exponentiation rational numbers.

**Definition 1.24**

If $a \geq 0$ is a real number and $n$ is a natural number, we define the $n^{th}$ *root* of $a$, denoted[a] $\sqrt[n]{a}$ to be the non-negative number $b$ such that $b^p = a$.

[a]When $n = 2$, we often omit the $n$ in $\sqrt[n]{a}$ and just write $\sqrt{a}$.

There are many subtleties in discussing roots as defined above, and in particular those subtleties can vary depending on the numbers one chooses to plug into the definition. First, roots do represent a partial inverse to exponentiation. By definition, if $b = \sqrt[n]{a}$ then $b$ satisfies $b^n = a$; that is,

$$b^n = (\sqrt[n]{a})^n = a.$$

For this reason, one often denotes $\sqrt[n]{a} = a^{1/n}$. The properties of power laws then implies that

$$(a^{1/n})^n = a^{n/n} = a^1 = a.$$

The next is that the definition clearly states that $a$ must be a non-negative number. Why is this the case? Consider the instance in which we are asked to determine $b = \sqrt{-1}$; so that $b$ must satisfy $b^2 = -1$. However, we know that any number multiplied by itself must be non-negative, so there can be no solution to this equation.

Furthermore, if we consider the case $b = \sqrt{4}$ we see that there are two numbers satisfying $b^2 = 4$: $b = 2$ and $b = -2$. Since we would like roots to define functions we can only choose one $b$ as our solution, so we establish the convention of always choosing the positive solution.

The problems discussed above were demonstrated in the case when $n = 2$, and it turns out that these pathological examples only occur when $n$ is even. When $n$ is odd, there is no issue with taking $n^{th}$ roots of negative numbers, nor with the existence of multiple solutions. As an example, consider $b = \sqrt[3]{-8}$. There is a unique number, $b = -2$, such that $b^3 = -8$. We summarize our discussion below.

1. If $n$ is even and $a \geq 0$, then $b^n = a$ will have multiple solutions. To avoid ambiguity in defining $\sqrt[n]{a}$, we demand that $b$ must be non-negative.
2. If $n$ is even, it is impossible to define the $n^{th}$ root of a negative number.
3. If $n$ is odd, then neither 1 nor 2 apply; that is, there is a unique solution to $b^n = a$ for any $a \in \mathbb{R}$.

**Example 1.25**

Determine the values of $\sqrt{9}$ and $\sqrt[3]{-64}$.

*Solution.* Starting with $\sqrt{9}$, our goal is to find a positive integer $b$ such that $b^2 = 9$. We know that there will be multiple solutions since $n = 2$ is even, and indeed $b = 3$ and $b = -3$ both work. As our definition stipulates that $b$ must be non-negative, we take $b = 3$ and conclude that $\sqrt{9} = 3$.

On the other hand, as $n = 3$ is odd we know that $b = \sqrt[3]{-64}$ is the unique solution to $b^3 = -64$. A bit of trial and error shows that $(-4)^3 = -64$ and so $\sqrt[3]{-64} = -4$. ■

By exploiting the identities given in Equation (1.6), we can immediately deduce the following for roots: If $a, b \in \mathbb{R}$ and $m, n \in \mathbb{Z}$ then

$$\sqrt[n]{a}\,\sqrt[m]{a} = \sqrt[\frac{m+n}{mn}]{a}, \qquad \sqrt[n]{\sqrt[m]{a}} = \sqrt[mn]{a}, \qquad \sqrt[n]{ab} = \sqrt[n]{a}\,\sqrt[n]{b}. \tag{1.7}$$

### 1.6.2 Logarithms

Given the equation $a^n = b$, we have discussed exponentiation and the process of taking roots. In essence, these ideas boil down to a two-out-of-three argument; that is, given two of variables solve for the third. For exponentiation, one is given $a$ and $n$ and told to determine $b$, while given $n$ and $b$ we may take $n^{th}$ roots to determine $a$. The remaining situation, given $a$ and $b$ determine $n$, is described by logarithms.

Why might we want to find such an $n$? There are many industrial reasons, the most often of which appear in pre-calculus courses as problems in finance. As an example, one is told that an asset appreciates at a fixed rate of 4% per annum and is tasked with determining the number of years until the asset's worth has doubled. This amounts to solving the equation $(1.04)^b = 2$, which we see is precisely the aforementioned problem which logarithms are designed to solve.

From a mathematical perspective, logarithms will arise as the inverse function to exponentiation. We saw that for a *fixed* natural number $n$ we could invert the process of exponentiation $x^n$ by taking an $n^{th}$ root. This is useful if we want to talk about inverses of polynomials: if $f(x) = x^3$ then $f^{-1}(x) = \sqrt[3]{x}$. If we now fix the base and let the exponent vary, taking roots is very much untenable; in fact, our goal is to find the exponent itself! Logarithms are the solution to the inversion problem.

$a^n = b$

| | | |
|---|---|---|
| Exponentiation | Given:<br>Determine: | $a$ and $n$<br>$c$ |
| Roots | Given:<br>Determine: | $n$ and $b$<br>$a$ |
| Logarithms | Given:<br>Determine: | $a$ and $b$<br>$n$ |

Table 1: A description of the possible "two-out-of-three" situations arising from the equation $a^n = b$.

**Definition 1.26**

If $a$ and $b$ are positive numbers, we define $\log_a b$ (read as *the base-a logarithm of b*) as the number $c$ satisfying $a^c = b$.

If the student is unfamiliar with logarithms the above definition can be a lot to take in. We would encourage the student to take a second and parse Definition 1.26 until it starts to make sense.

**Example 1.27**

Compute $\log_2 32$, $\log_3 27$.

*Solution.* Let $c = \log_2 32$ in which case Definition 1.26 implies that $c$ must satisfy $2^c = 32$. The student will hopefully recall (or easily compute) that $2^5 = 32$, so $c = 5$ and we conclude that $\log_2 32 = 5$.

Similarly, if $c = \log_3 27$ then $3^c = 27$. The student can then easily check that $3^3 = 27$ and so $\log_3 27 = 3$. ■

The manner in which we started this section should suggest that the logarithm is going to play the inverse role to exponentiation. Indeed, items 3 and 4 in the following proposition shed some light on the relationship between logarithms and exponentials.

**Proposition 1.28: I**

$a$ and $b$ are positive real numbers with $a \neq 1$, then

1. $\log_a(1) = 0$
2. $\log_a(a) = 1$
3. $\log_a(a^b) = b$
4. $a^{\log_a(b)} = b$

These results are very simple and, as always, the student should make an attempt to prove the results on their own before looking at the proof. This will not only build confidence in working with logarithms, but also expand the student's comprehension of the subject.

*Proof.* 1. Set $c = \log_a(1)$ so that $a^c = 1$. Since $a \neq 1$ by hypothesis, it must be the case that $c = 0$. Thus $\log_a(1) = 0$ as required.

2. Similar to part 1, we know that $c = \log_a(a)$ satisfies $a^c = a$. It is not to hard to see that $c = 1$ is the only possible solution and hence $\log_a(a) = 1$.

3. Let $c = \log_a(a^b)$ so that $c$ satisfies $a^c = a^b$. It should be clear[4] that $c = b$ is the solution, so that $\log_a(a^b) = b$.

4. Let $c = \log_a(b)$ so that $a^c = b$. However, simply substituting our first expression of $c$ into the latter expression, we get $a^{\log_a(b)} = b$ as required. ☐

### 1.6.3 The Exponential and Logarithmic Functions

The procedure for extending exponentiation from $a^n$ for natural numbers $n$, to $a^x$ for real numbers $x$, is quite difficult. It requires that we either have access to the mathematics of sequences (which we will not cover), or integration (which is not covered until the second half of the course). As a result, the student is going to have to take my word that such extensions exist.

We define an exponential function $f(x) = a^x$ whenever $a > 0$. This function has domain $\mathbb{R}$ and range $(0, \infty)$. There is a very special value of $a$ known as *Euler's number* and denoted by $e$. Unfortunately, the most intuitive definitions of this number require some notion of calculus, and so we only mention it here and define it later. For the interest of the student, $e$ is approximately

$$e \approx 2.7182818284...$$

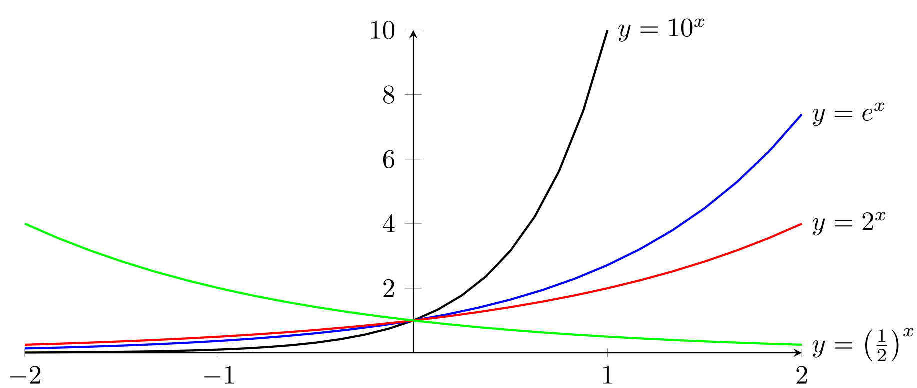



Figure 15: The graph of the exponential function for several choice of bases.

Figure 15 contains the graphs of several exponential functions. Notice that these satisfy the horizontal line test, and therefore should be invertible. The *logarithmic function with base* $a > 0$ is

[4] Here we technically require that $a^x$ is injective for positive $a \neq 1$. At this point, the proof of injectivity would be rather difficult, but will be a simple exercise once we have learned a little calculus.

the function $g(x) = \log_a(x)$, which is designed to act as the inverse function for $f(x) = a^x$. Indeed, using items 3 and 4 of Proposition 1.29 we see that

$$a^{\log_a(x)} = x, \qquad \log_a(a^x) = x$$

so that the functions are true inverses of one another. Recall that $a^x$ grows to infinity if $a > 0$ and shrinks to 0 if $0 < a < 1$. As such, we expect a similar dichotomy in the graphs of the logarithmic function. Since $a^x$ and $\log_a(x)$ are inverses, the graph of $\log_a(x)$ is just the reflection of $a^x$ about the line $y = x$ and is given in Figure 16. The domain of $\log_a(x)$ is $(0, \infty)$ while its range is all of $\mathbb{R}$.

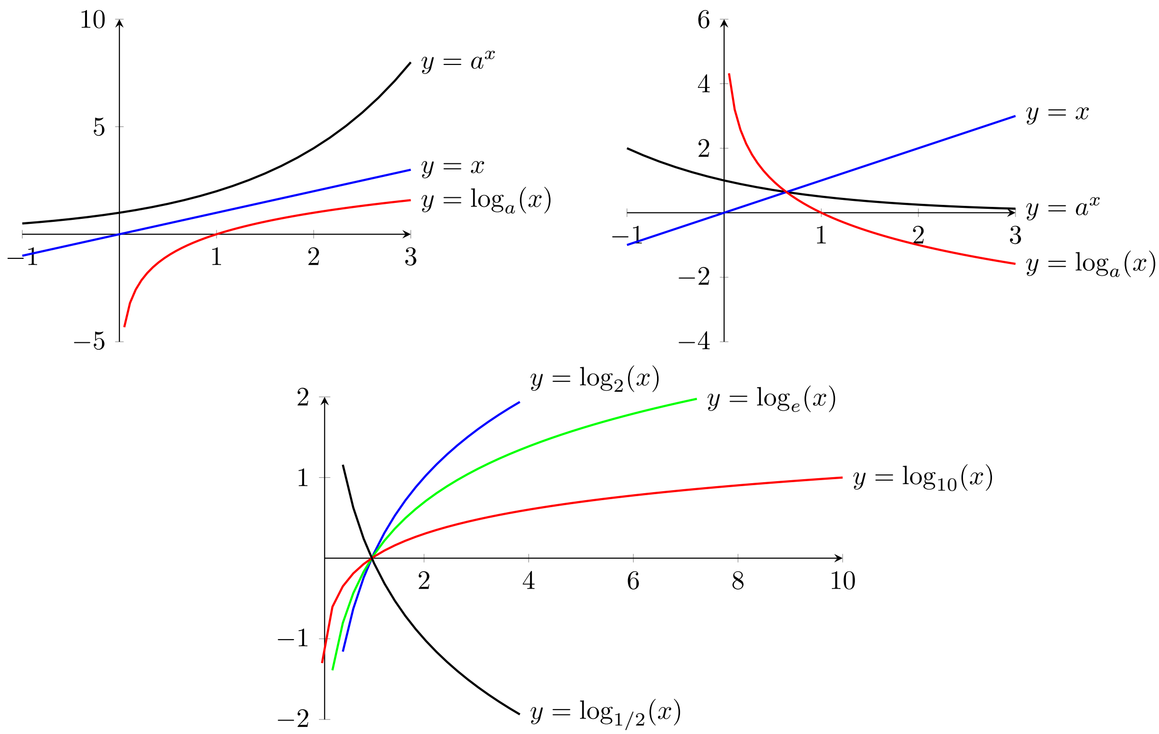



Figure 16: *Top:* The graph of the logarithm function $\log_a(x)$ for different values of $a$, viewed as the reflection of the exponential function $a^x$ about the line $y = x$. *Bottom:* The logarithmic function with several different bases.

Given the close relationship between logarithms and exponents, it should not be surprising that it is also a special base for the logarithmic function $\log_e(x)$. In fact, this function is so special that there are two competing mathematical conventions in writing it down. The first is to write $\ln(x)$, pronounced as "lawn of $x$," while the other is to simply omit the base and write $\log(x)$. The latter is typically used by mathematicians alone, while scientists and engineers prefer the $\ln(x)$ notation.

The next proposition gives a list of useful logarithmic identities, all of which may be proven by exploiting the relationship between the logarithm and the exponential. We shall provide the proof of the most difficult result, but the rest are left as exercises for the student.

**Proposition 1.29**

Let $d$ be any real number and $a$ be a positive number such that $a \neq 1$. For any $x, y > 0$ we then have

1. $\log_a(x^d) = d\log_a(x)$,
2. $\log_a(xy) = \log_a x + \log_a y$
3. $\log_a(x/y) = \log_a x - \log_a y$,
4. $\log_a b = \dfrac{\log_d b}{\log_d a}$.

*Proof.* The proofs of 1, 2, and 3 are exercises in applying the appropriate exponential identity and are left to the student. We prove here 4. Define

$$\begin{array}{rclcrclcrcl} c &=& \log_a b & \qquad & c_1 &=& \log_d b & \qquad & c_2 &=& \log_d a \\ a^c &=& b & & d^{c_1} &=& b & & d^{c_2} &=& a \end{array} \tag{1.8}$$

Now starting with $a^c = b$ we substitute the latter two expressions in (1.8) to get

$$\begin{aligned} a^c &= b \\ (d^{c_2})^c &= (d^{c_1}) && \text{since } a = d^{c_2} \text{ and } b = d^{c_1} \\ d^{c_2 \times c} &= d^{c_1} \end{aligned}$$

which implies that $c_2 \times c = c_1$. Solving for $c$ we get $c = c_1/c_2$ or rather

$$\log_a b = \frac{\log_d b}{\log_d a}$$

as required. ∎

# 2 Limits

## 2.1 Some Motivation

Limits are the method by which we, as manifestly finite beings, deal with concepts of infinities and infinitesimals. The goal towards which we are working is a description of instantaneous rate of change, so let's think on what this means.

The majority of us have been in a car at some point or another, and have afforded a casual glance at the speedometer. Let us say that at the instant we look down, the speedometer reads 90 km/hr. Have you ever thought about what it means, at that single instant in time, to be travelling at that speed? As suggested by its units, speed is an object which requires both distance and time to measure, but at a single moment, neither any time nor any distance has passed, so what does this mysterious quantity mean?

Despite my claims that the previous example should get you thinking about how the word "instantaneous" really affects a quantity, many of you will simply shrug aside my suggestions. In anticipation of this reaction, what if we change the associated quantities around and instead of the instantaneous speed of a car, we discuss shopping! At any given point of time, somebody on this planet is making a purchase. Assume that we were able to measure the rate at which people were spending money, and I told you that at this moment in time the human species was globally spending \$140 million dollars an hour? What does this mean?

Now on the other hand, what if you were asked to determine the instantaneous speed of a race car at the instant its front bumper passes a finish line? Being clever students, you decide to measure how far the car has travelled in the minute before it hits the finish line, and get a result of 1500 meters. Hence the car was travelling

$$\frac{1500 \text{ metres}}{1 \text{ minute}} \times \frac{1 \text{ kilometre}}{1000 \text{ metres}} \times \frac{60 \text{ minutes}}{1 \text{ hour}} = \frac{90 \text{ kilometres}}{1 \text{ hour}}.$$

But what if the cars speed was not constant during that minute? What if the driver accelerated at the end? You decide that you can get a better estimate of the speed at the finish line by instead just looking at how far the car travelled in the single second before the car hit the finish line. This time the car travelled 30 metres, so you calculate

$$\frac{30 \text{ metres}}{1 \text{ second}} = \frac{108 \text{ kilometres}}{1 \text{ hour}}.$$

But still, this does not account for any change in acceleration which occurred in the last second. Your guess of 108 km/hr is probably close, but close is not good enough in mathematics! So you try again by measuring the distance after 0.1 seconds, then 0.01 seconds, and so on , but no matter how hard you try you cannot get the exact speed because there is always the chance that the car was not travelling at a constant speed during your measurements. Nonetheless, we know there must be an answer: the car was travelling at some speed, so what is it? Limits provide the solution.

### 2.1.1 Intuition

Limits are the mathematical device which allow us to infer information about a point by analyzing information about well-behaved points nearby. Let $f(x)$ be an arbitrary function and $c \in \mathbb{R}$. We say that "the limit of $f(x)$ as $x$ approaches $c$ is equal to $L$" if, whenever we let $x$ get arbitrarily close to $c$ then $f(x)$ gets arbitrarily close to $L$. This is written as

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

The best way to gain an intuitive understanding of limits is to see a few examples. We warn the student that this first example is rather nicely behaved and fails to capture why we use limits. Nonetheless, simple examples are often the best for getting a grasp as to how something works.

**Example 2.1**

Consider the function $f(x) = 4x + 2$. Determine the limits

$$\lim_{x \to 0} f(x), \qquad \lim_{x \to -4} f(x), \qquad \lim_{x \to 5} f(x).$$

Form a hypothesis as to what the limit is as $x \to c$ for any value of $c$.

*Solution.* We implore the student to keep in mind that this solution is purely heuristic and is only presented in a way to show the student how to think about these problems. The first example asks us to consider what happens when $x = 0$, so we would like to see what happens for values of $x$ which are close (but not equal to zero). Hopefully the student can guess that as $x$ gets close to zero, $4x + 2$ gets close to $4 \cdot 0 + 2 = 2$. Similarly, as $x \to -4$ then $4x + 2$ approaches $4 \cdot (-4) + 2 = -14$. The following table corroborates this idea:

| $x \to 0$ | | | | $x \to -4$ | | | |
|---|---|---|---|---|---|---|---|
| $x < 0$ | | $x > 0$ | | $x < -4$ | | $x > -4$ | |
| $x$ | $f(x)$ | $x$ | $f(x)$ | $x$ | $f(x)$ | $x$ | $f(x)$ |
| $-0.1$ | 1.6 | 0.1 | 2.4 | $-4.1$ | $-14.4$ | $-3.9$ | $-13.6$ |
| $-0.05$ | 1.8 | 0.05 | 2.2 | $-4.05$ | $-14.2$ | $-3.95$ | $-13.8$ |
| $-0.01$ | 1.96 | 0.01 | 2.04 | $-4.01$ | $-14.04$ | $-3.99$ | $-13.96$ |
| $-0.005$ | 1.98 | 0.005 | 2.02 | $-4.005$ | $-14.02$ | $-3.995$ | $-13.98$ |
| $-0.001$ | 1.9996 | 0.001 | 2.0004 | $-4.001$ | $-14.004$ | $-3.999$ | $-13.996$ |
| $-0.0005$ | 1.9998 | 0.0005 | 2.0002 | $-4.0005$ | $-14.002$ | $-3.9995$ | $-13.998$ |

As a matter of fact, it looks as though

$$\lim_{x \to 0} f(x) = f(0) = 2, \qquad \lim_{x \to -4} f(x) = f(-4) = -14$$

so we guess that in general,

$$\lim_{x \to c} f(x) = f(c) = 4c + 2.$$

■

In Example 2.1 we guessed that the limit as $x \to c$ could be determined by evaluating $f(c)$, and it turns out that in this example that is correct. However, one must be very careful about just freely plugging in numbers into equations as the function might not always be defined at that point.

**Example 2.2**

Let $f(x) = \dfrac{x^2 + x - 6}{x - 2}$. Determine the limit $\lim_{x \to 2} f(x)$.

*Solution.* Unlike the previous example, attempting the substitute $x = 2$ into $f(x)$ will result in division-by-zero, which we know is never permitted. However, we can evaluate $f(x)$ at any number other than 2 and the hope is that this will tell us what the function looks like at $x = 2$. Indeed, creating the following table we once again find that

| $x$ | $f(x)$ |
|---|---|
| 2.1 | 5.1 |
| 2.05 | 5.05 |
| 2.01 | 5.01 |
| 2.005 | 5.005 |
| 2.001 | 5.001 |
| 2.0005 | 5.0005 |

so it certainly appears as though $f(x)$ is approaching 5. Indeed, if $x \neq 2$ then we may factor $f(x)$ as

$$\frac{x^2+x-6}{x-2} = \frac{(x+3)(x-2)}{x-2} = x+3$$

and the behaviour of this function as $x \to 2$ agrees with our observations. ■

The previous example demonstrates that a function does not need to be defined at a point for the limit at that point to exist. In fact, this is an excellent opportunity to point out that the functions $f(x) = \frac{x^2+x-6}{x-2}$ and $g(x) = x+3$ are similar but *are not equal*: the distinction being that the domain of $f(x)$ is $\mathbb{R} \setminus \{2\}$ while the domain of $g(x)$ is $\mathbb{R}$. If two functions have different domains, they certainly cannot be equal! Of course, $x = 2$ is the only point where the functions do not agree, and their graphs are even identical with the exception that the graph of $f(x)$ will have a hole at $x = 2$. Luckily, this does not matter when we are taking limits, and we have the equality

$$\lim_{x\to 2} \frac{x^2+x-6}{x-2} = \lim_{x\to 2} x+3.$$

The reason is that while the functions differ at the point $x = 2$, the limit only looks at what the functions do at points close to *but not equal* to 2. Thus the limits see them as the same function (cf Figure 17).

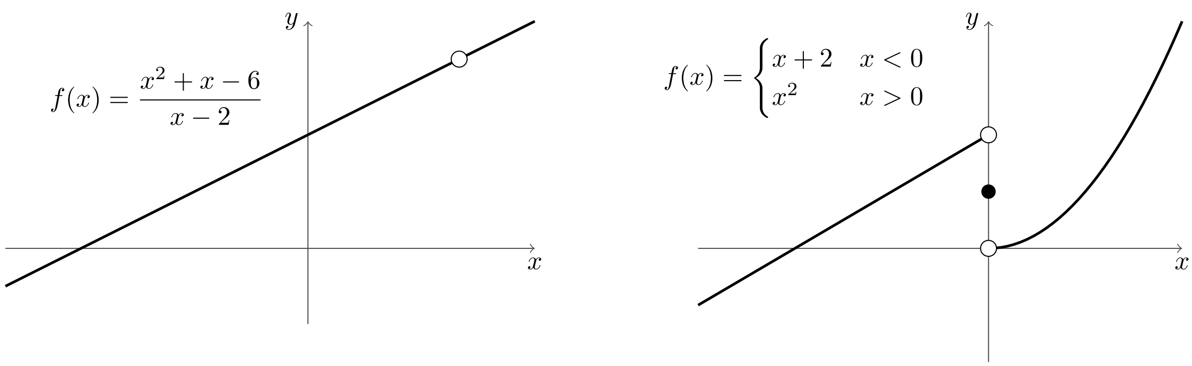



Figure 17: **Left:** The function $\frac{x^2+x-6}{x-2}$ is identical to the function $x+3$ except for the presence of a hole at $x = 2$. This does not affect the limit though, as the limit is only concerned with the behaviour of the function *near* $x = 2$. **Right:** A graph of a piecewise function whose limit at zero is dependent upon the direction of approach. Notice that in either case, the limit disagrees with the value of the function at zero.

**Example 2.3**

Compute the limit

$$\lim_{x\to 2} \frac{x-2}{\sqrt{x+7}-3}.$$

*Solution.* This is actually identical to the previous example, but it may not be obvious why. Instead, the usual thing to do in such situations where one summand is a square root is to multiply by the conjugate. In this case, $\sqrt{x+7}+3$. In that case we have

$$\begin{aligned}\lim_{x\to 2}\frac{x-2}{\sqrt{x+7}-3}\frac{\sqrt{x+7}+3}{\sqrt{x+7}+3} &= \lim_{x\to 2}\frac{(x-2)(\sqrt{x+7}+3)}{(x+7)-9}\\ &= \lim_{x\to 2}\frac{x-2}{x-2}\left[\sqrt{x+7}+3\right]\\ &= \lim_{x\to 2}\left[\sqrt{x+7}+3\right] = 6.\end{aligned}$$

The reason this is identical to Example 2.2 is that

$$x-2 = (x+7)-9 = \left[\sqrt{x+7}-3\right]\left[\sqrt{x+7}+3\right]$$

and so the steps we just iterated may be replaced by a much simpler cancellation argument. ■

## 2.2 One Sided Limits

Implicit in our previous discussion of limits is that when we take $x\to c$, we must get the same answer whether we are approaching from the left of $c$ or the right of $c$. It is possible that approaching from the left and right actually give different values of the limit, as can be seen in Figure 17. This naturally leads us to the idea of one-sided limits, where we restrict our attention to values of the function on only one side of the limiting point.

More generally, we will say that "the limit of $f(x)$ as $x$ approaches $c$ from the *right* is $L$" if whenever $x>c$ gets arbitrarily close to $c$, $f(x)$ gets arbitrarily close to $L$. This is written in symbols as

$$\lim_{x\to c^+} f(x) = L.$$

Similarly, we say that "the limit of $f(x)$ as $x$ approaches $c$ from the *left* is $L$" if whenever $x<c$ gets arbitrarily close to $c$, $f(x)$ gets arbitrarily close to $L$, and in this case we write

$$\lim_{x\to c^-} f(x) = L.$$

If both of the one-sided limits exist and are equal to the same value $L$, then the limit $x\to c$ exists and is also equal $L$. There are plenty of examples where the two-sided limit does not exist, as our following examples demonstrate.

**Example 2.4**

Consider the function

$$f(x) = \begin{cases} x+2 & x<0\\ 1 & x=0\\ x^2 & x>0\end{cases}.$$

Compute the limit of $f(x)$ as $x\to 0^-$ and as $x\to 0^+$. Does the two-sided limit exist?

*Solution.* We first look at the limit as $x\to 0^-$. In this case, we know that $x$ is always less than 0, so $f(x)$ effectively looks like the function $x+2$. As $x\to 0^-$ we see that $x+2\to 2$ and so we conclude that

$$\lim_{x\to 0^-} f(x) = 2.$$

On the other hand, the limit $x \to 0^+$ guarantees that $x$ is always positive. Here, $f(x)$ looks like the function $x^2$ and as $x$ approaches 0, $x^2$ approaches 0 as well, so

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

Each one-sided limit exists, but they are not equal. Hence the two-sided limit does not exist. The graph of $f(x)$ is given in Figure 17. ■

The other way that a two-sided limit can fail to exist is if the one-sided limits do not exist either. This can happen in one of two ways: The first is that it is impossible to find a number $L$ to which the function gets close. This can happen, for example, if our function oscillates infinitely in any small interval around a point. Alternatively, the limits can *diverge*, meaning that the function goes to either positive or negative infinity. This next example demonstrates both phenomena.

**Example 2.5**

Let $f(x)$ be the function

$$f(x) = \begin{cases} \sin\left(\dfrac{1}{x}\right) & x < 0 \\ \dfrac{1}{1-x} & 0 \leq x < 1 \\ \dfrac{x^2}{x-1} & x > 1 \end{cases}$$

Examine the limits as $x \to 0^\pm$ and $x \to 1^\pm$.

*Solution.* Take a look at Figure 18 which plots $f(x)$. In the limit as $x \to 0^-$ we want to examine how the function $\sin(1/x)$ behaves when $x$ is a small negative number. With any luck, the student will recognize that the function begins to oscillate infinitely often as $x$ gets small, making it impossible for the function to ever converge to a single point. This tells us that

$$\lim_{x \to 0^-} f(x) \qquad \text{does not exist.}$$

In the case $x \to 0^+$, we are only interested in what happens in a small positive neighbourhood of 0 and hence we can assume that $0 < x < 1$ so that $f(x)$ looks like $\frac{1}{1-x}$. We encounter no problem by substituting small positive numbers into this function, and one can quickly deduce that

$$\lim_{x \to 0^+} f(x) = 1.$$

Now $\frac{1}{1-x}$ is not defined at $x = 1$, so we must be careful in looking at the limit as $x \to 1^-$. As $x < 1$ gets arbitrarily close to 1, the number $1 - x$ is a very small *positive* number. This means that $\frac{1}{1-x}$ is a very large positive number. As we can make $1 - x$ arbitrarily small, $\frac{1}{1-x}$ can be arbitrarily large so

$$\lim_{x \to 1^-} f(x) = \infty.$$

Finally, as $x \to 1^+$ we again run into troubles with the function $\frac{x^2}{x-1}$. But we can see that the number is always positive and tends to the number 1, while the denominator $x - 1$ gets arbitrarily small, but positive since $x > 1$. Making $x$ arbitrarily close to 1 implies that $\frac{x^2}{x-1}$ can get arbitrarily large, so that

$$\lim_{x \to 1^+} f(x) = \infty.$$

We conclude that neither of the two sided limits $x \to 0$ nor $x \to 1$ exist. ■

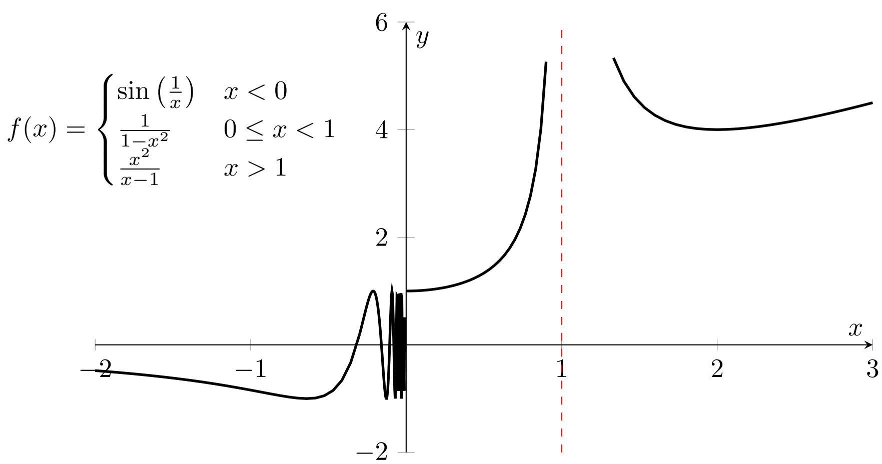



Figure 18: A pathological function for which no two sided limit exists at the points $x = 0$ and $x = 1$. Note that as $x \to 0^-$ the limit fails to converge, while as $x \to 1^\pm$ the function diverges to infinity.

## 2.3 Limit Laws

Mathematicians love to be lazy, in the sense that if we have already performed a calculation, why should we repeat it ever again? Similarly, we like to build complicated examples from simple examples. To this end, we formulate the following collection of limit laws, which are intended to dramatically simplify our life:

**Theorem 2.6: Limit Laws**

If $f(x)$ and $g(x)$ be two functions, and $c \in \mathbb{R}$ such that

$$\lim_{x\to c} f(x), \qquad \lim_{x\to c} g(x)$$

*both exist,* then

1. $\lim_{x\to c} [\alpha f(x)] = \alpha \lim_{x\to c} f(x)$,
2. $\lim_{x\to c} [f(x) \pm g(x)] = \left[\lim_{x\to c} f(x)\right] \pm \left[\lim_{x\to c} g(x)\right]$,
3. $\lim_{x\to c} [f(x)g(x)] = \left[\lim_{x\to c} f(x)\right] \left[\lim_{x\to c} g(x)\right]$,
4. $\lim_{x\to c} \left[\frac{f(x)}{g(x)}\right] = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}$ if $\lim_{x\to c} g(x) \neq 0$

It is crucial that both limits must exist. It is a common mistake for students to gleefully attempt to apply the above limits laws in instances in which it is not permitted. For example, the following **IS NOT CORRECT**:

$$\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right) = \left[\lim_{x\to 0} x^2\right] \left[\lim_{x\to 0} \sin\left(\frac{1}{x}\right)\right] = 0 \times \left[\lim_{x\to 0} \sin\left(\frac{1}{x}\right)\right] = 0.$$

Interestingly, this is the correct answer, but for all the wrong reasons. Hence this solution is absolutely incorrect.

**Example 2.7**

Compute the limit

$$\lim_{x\to 1} \frac{x^4 + 7x + 2}{x - 4}.$$

*Solution.* We would like to say that this is the limit of the quotient, and this will be the case so long as the limit of the numerator and denominator both exist, and the denominator is non-zero. For the numerator, we can again use the Limit Laws to break this into a sum of limits:

$$\begin{aligned}\lim_{x\to 1} \left[x^4 + 7x + 2\right] &= \lim_{x\to 1} x^4 + 7 \lim_{x\to 1} x + \lim_{x\to 1} 2 \\ &= \left[\lim_{x\to 1} x\right]^4 + 7\left[\lim_{x\to 1} x\right] + 2\left[\lim_{x\to 1} 1\right] \\ &= 1 + 7 + 2 = 10.\end{aligned}$$

Similarly, the denominator gives

$$\lim_{x\to 1} [x - 4] = 1 - 4 = -3.$$

Since both limits exist and the denominator is non-zero, we can apply the quotient Limit Law to get

$$\lim_{x\to 1} \frac{x^4+7x+2}{x-4} = \frac{\lim_{x\to 1} x^4+7x+2}{\lim_{x\to 1} x-4} = -\frac{10}{3}. \qquad \blacksquare$$

A similar argument to the previous example is the following theorem:

**Theorem 2.8**

If $f(x) = \frac{p(x)}{q(x)}$ is any rational functions (so that $p(x)$ and $q(x)$ are polynomials), and $c \in \mathbb{R}$ is such that $q(c) \neq 0$ then

$$\lim_{x\to c} \frac{p(x)}{q(x)} = \frac{p(c)}{q(c)}.$$

*Proof.* The key is to first show this for polynomials and apply the limit laws. We shall assume *a priori* that

$$\lim_{x\to c} x = c.$$

Now let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be an arbitrary polynomial. Certainly the limit of each $a_i x^i$ exists, and so the limit laws imply that

$$\begin{aligned}
\lim_{x\to c} p(x) &= \lim_{x\to c} [a_n x^n + \cdots + a_1 x + a_0] \\
&= a_n \left[\lim_{x\to c} x\right]^n + a_{n-1}\left[\lim_{x\to c} x\right]^{n-1} + \cdots + a_1 \left[\lim_{x\to c} x\right] + a_0 \\
&= a_n c^n + a_{n-1}c^{n-1} + \cdots + a_1 c + a_0 \\
&= p(c).
\end{aligned}$$

Thus the result holds for any polynomial. Now if $p(x)$ and $q(x)$ are two polynomials and $q(c) \neq 0$, then the limit laws for quotients implies

$$\lim_{x\to c} \frac{p(x)}{q(x)} = \frac{\lim_{x\to c} p(x)}{\lim_{x\to c} q(x)} = \frac{p(c)}{q(c)}. \qquad \square$$

**Example 2.9**

Determine the limit

$$\lim_{x\to 0} \frac{\sin^2(x)}{1-\cos(x)}.$$

*Solution.* On first glance, we are unable to substitute $x = 0$ into this equation since it would cause us to divide by zero. Instead, we recall that $\sin^2(x) = (1-\cos^2(x)) = (1-\cos(x))(1+\cos(x))$ so that

$$\begin{aligned}
\lim_{x\to 0} \frac{\sin^2(x)}{1-\cos(x)} &= \lim_{x\to 0} \frac{(1-\cos(x))(1+\cos(x))}{1-\cos(x)} \\
&= \lim_{x\to 0} [1+\cos(x)] = 2.
\end{aligned} \qquad \blacksquare$$

**Example 2.10**

Determine the limit

$$\lim_{t\to 0} \frac{\sqrt{1+t}-\sqrt{1-t}}{t}.$$

*Solution.* When one finds that an expression involving square roots is proving difficult, it is often a good idea to multiply by a conjugate form which places the square roots into a more amenable position. In this instance, notice that

$$\left(\sqrt{1+t}-\sqrt{1-t}\right)\frac{\sqrt{1+t}+\sqrt{1-t}}{\sqrt{1+t}+\sqrt{1-t}} = \frac{(1+t)-(1-t)}{\sqrt{1+t}-\sqrt{1-t}} = \frac{2t}{\sqrt{1+t}+\sqrt{1-t}}.$$

Hence our limit becomes

$$\begin{aligned}\lim_{t\to 0} \frac{\sqrt{1+t}-\sqrt{1-t}}{t} &= \lim_{t\to 0} \frac{2t}{t\left(\sqrt{1+t}+\sqrt{1-t}\right)} \\ &= \lim_{t\to 0} \frac{2}{\sqrt{1+t}+\sqrt{1-t}} \\ &= 1.\end{aligned}$$

■

**Example 2.11**

Compute the limit

$$\lim_{h\to 0} \frac{(x+h)^2-x^2}{h}$$

*Solution.* Notice here that the variable of the limit is $h$ rather than $x$! Since we cannot just substitute $h=0$ into this equation, we must manipulate the expression to see if we can derive a more meaningful representative. Expanding the denominator we get

$$\begin{aligned}\lim_{h\to 0} \frac{(x+h)^2-x^2}{h} &= \lim_{h\to 0} \frac{x^2+2xh+h^2-x^2}{h} \\ &= \lim_{h\to 0} \frac{2xh+h^2}{h} \\ &= \lim_{h\to 0} 2x+h \\ &= 2x.\end{aligned}$$

■

Expressions such as these will prove supremely useful when we cover derivatives.

## 2.4 The Squeeze Theorem

The Squeeze Theorem allows us to determine troublesome limits by bounded one function in terms of two other functions which converge to the same point.

**Theorem 2.12**

If $f(x), g(x)$, and $h(x)$ are all defined in an interval around the point $c$, satisfying $f(x) \leq g(x) \leq h(x)$ on that interval, and

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

then the limit of $g(x)$ as $x \to c$ also exists and is equal to $L$; that is,

$$\lim_{x\to c} g(x) = L.$$

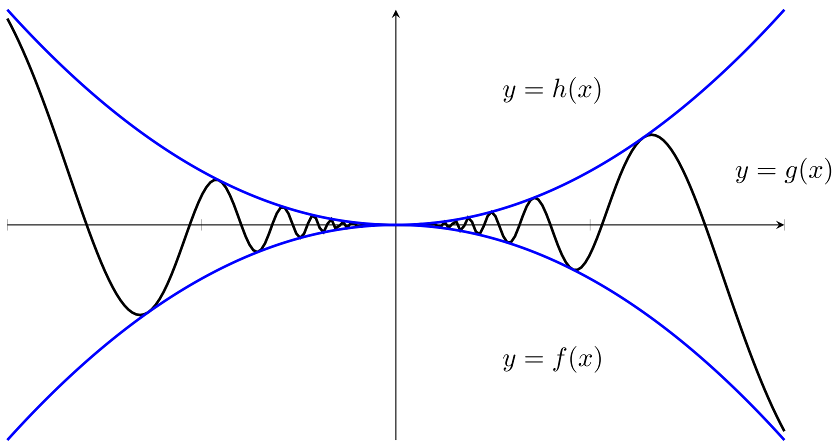



Figure 19: A visualization of a typical Squeeze Theorem example.

As is demonstrated by Figure 19, the idea is that the functions $f(x)$ and $h(x)$ squeeze $g(x)$ into having the same limit. In fact, Theorem 2.12 also holds when we take $x \to \pm\infty$.

**Example 2.13**

Show that

$$\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right) = 0.$$

*Solution.* Regardless of its argument, the functions $\sin(x)$ is always bounded above by $y = 1$ and below by $-1$; that is,

$$-1 \leq \sin\left(\frac{1}{x}\right) \leq 1.$$

As $x^2 > 0$ everywhere, we can multiply through to see that

$$-x^2 \leq x^2 \sin\left(\frac{1}{x}\right) \leq x^2.$$

We are interested in the behaviour as $x \to 0$, and we notice that

$$\lim_{x\to 0} x^2 = \lim_{x\to 0} -x^2 = 0.$$

The Squeeze Theorem thus applies and we get

$$\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right) = 0$$

as required. ■

**Example 2.14**

Determine $\lim_{x\to\infty} \frac{\cos(x^2)}{x^2+1}$.

*Solution.* We know that $-1 \le \cos(x^2) \le 1$ for all $x$, so

$$-\frac{1}{x^2+1} \le \frac{\cos(x^2)}{x^2+1} \le \frac{1}{x^2+1}.$$

It is straightforward to check that

$$\lim_{x\to\infty} \pm\frac{1}{x^2+1} = 0,$$

so by the Squeeze Theorem we conclude that

$$\lim_{x\to\infty} \frac{\cos(x^2)}{x^2+1} = 0.$$ ■

**Exercise**

1. If $\lim_{x\to 0} |f(x)| = 0$ then $\lim_{x\to 0} f(x) = 0$. [Hint: Convince yourself that $-|f(x)| \le f(x) \le |f(x)|$.]
2. If $f(x)$ is bounded (so that there is some $M > 0$ with $|f(x)| < M$ for all $x$) and $\lim_{x\to 0} g(x) = 0$ then $\lim_{x\to 0} f(x)g(x) = 0$.

## 2.5 Infinite Limits

There are two ways in which to consider "limits at infinity:" either the function itself can diverge to infinity, or we can take a limit as $x \to \pm\infty$. The following sections will discuss this behaviour.

### 2.5.1 Vertical Asymptotes

There are functions which are *singular* at a point $x = c$, and these can sometimes result in our limits being infinite. For example, the function $f(x) = 1/x$ becomes positively (resp. negatively) large when $x > 0$ (resp. $x < 0$) is small. So in this instance we have

$$\lim_{x\to 0^+} \frac{1}{x} = \infty, \qquad \lim_{x\to 0^-} \frac{1}{x} = -\infty. \tag{2.1}$$

Depending on the given function, the left and right limits might give the same signed infinity. For example,

$$\lim_{x\to 0} \frac{1}{x^2} = \infty.$$

In these instances we maintain that the limit *does not exist*, but will still write the equal sign to indicate that the one-sided limits agree.

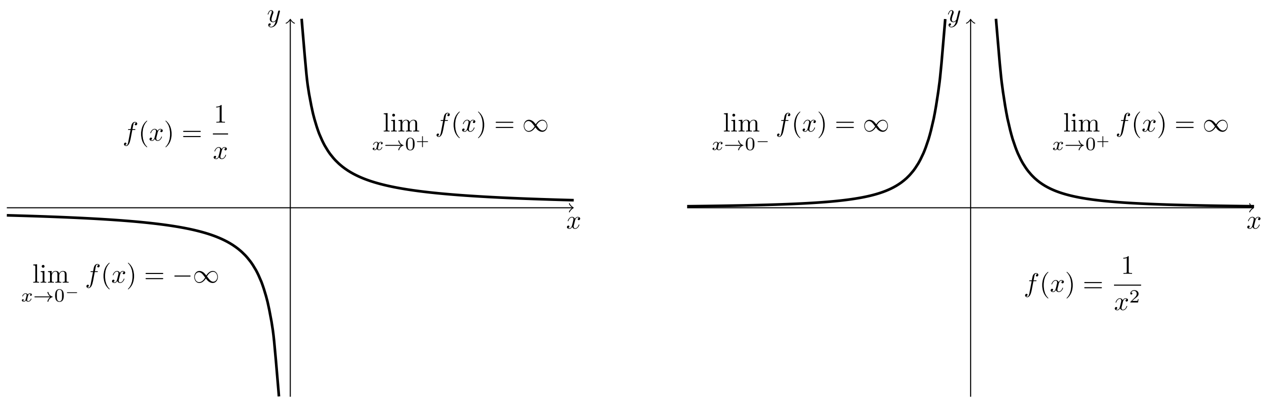



Figure 20: The functions $1/x$ and $1/x^2$ both have vertical asymptotes at $x = 0$.

**Definition 2.15**

The line $x = a$ is said to be a *vertical asymptote* for $f(x)$ if any of the one sided limits is infinite; that is, if

$$\lim_{x\to a^\pm} f(x) = \infty, \qquad \text{or} \qquad \lim_{x\to a^\pm} f(x) = -\infty.$$

**Example 2.16**

What are the vertical asymptotes of the function $f(x) = \csc(x)$?

*Solution.* The function $\csc(x)$ is really $f(x) = \frac{1}{\sin(x)}$. Since we are unable to divide by 0, our problem reduces to finding the points where $\sin(x) = 0$. Our keen knowledge of trigonometry tells us that this occurs at every integer multiple of $\pi$; that is,

$$n\pi, n = \ldots, -2, -1, 0, 1, 2, \ldots$$

In fact, we can go one step further. We know that $\sin(x) > 0$ on $[-2\pi, -\pi], [0, \pi], [2\pi, 3\pi], \ldots$, so that the endpoints of these intervals correspond to places where our $\csc(x)$ diverges to $+\infty$. In all other places, $\csc(x)$ goes to $-\infty$. See Figure 21. ■

**Example 2.17**

Determine the vertical asymptotes of the function $f(x) = \dfrac{1}{(x-3)(x-4)}$.

*Solution.* It will not take much to convince us that this limit is going to be infinite, since the numerator is constant but the denominator gets very small. The only question is whether or not this approaches positive or negative infinity. Notice that as $x \to 4$ we can always choose to limit ourselves to an arbitrarily small neighbourhood of $x = 4$, so in particular, let's assume that $3 < x < 4$. In this case $x - 3 > 0$ but $x - 4 < 0$, so that $(x-3)(x-4) < 0$. This tells us that

$$\lim_{x \to 4^-} \frac{1}{(x-3)(x-4)} = -\infty, \qquad \lim_{x \to 3^+} \frac{1}{(x-3)(x-4)} = -\infty.$$

On the other hand, when $x > 4$ or $x < 3$, the same argument above implies that $(x-3)(x-4) > 0$, so that

$$\lim_{x \to 4^+} \frac{1}{(x-3)(x-4)} = \infty, \qquad \lim_{x \to 3^-} \frac{1}{(x-3)(x-4)} = \infty.$$

Thus the limit diverges to infinity in both instances, but to *different* infinities. Hence the limit does not exist. ■

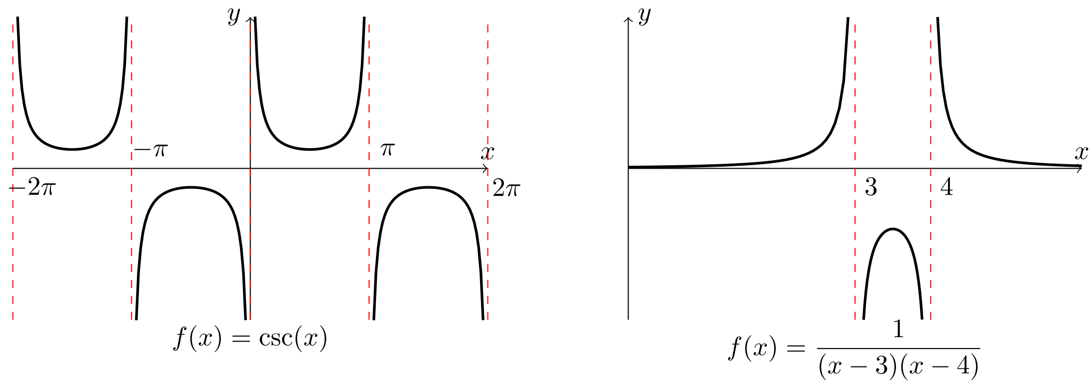



Figure 21: The vertical asymptotes of the functions found in Examples 2.16 and 2.17.

### 2.5.2 Horizontal Asymptotes

Many of the functions we have discussed so far fail to behave nicely as we tend to infinity. For example, the functions $x, x^2, e^x$ all get very large as $x$ gets very large, and other functions such as $\sin(x)$ and $\cos(x)$ just oscillate infinitely often. However, there are some functions which exhibit a finite behaviour as we head off towards infinity, and we shall dedicate ourselves in the short term to examining such asymptotic behaviour.

**Definition 2.18**

If $f(x)$ is defined on $(a, \infty)$ (resp. $(-\infty, a)$) then we say that

$$\lim_{x \to \infty} f(x) = L, \qquad \left(\text{resp. } \lim_{x \to -\infty} f(x) = L\right)$$

if whenever $x$ gets arbitrarily large and positive (resp. negative) then $f(x)$ gets arbitrarily close to $L$. In such instances, we say that $L$ is a *horizontal asymptote* of $f(x)$.

To re-iterate, polynomial and trigonometric functions are not going to have finite limits at infinity, since the polynomial functions diverge and the polynomials oscillate without bound. So what are examples of functions which do have finite? Well to start, for any $p > 0$, we have that

$$\lim_{x\to\pm\infty} \frac{1}{x^p} = 0.$$

This gives us a large number of unexciting functions with finite asymptotics. However, this result is of far more interest when used on other functions.

**Example 2.19**

Determine the limit $\displaystyle\lim_{x\to\infty} \frac{3x^2+6x-1}{4x^2-2}$.

*Solution.* It is rather unwieldy to deal with this function as written, since both the numerator and denominator become arbitrarily large and it is difficult to see what "cancellations" might occur. Instead, let us multiply and divide by the quantity $\frac{1}{x^2}$ to get

$$\begin{aligned}
\lim_{x\to\infty} \frac{3x^2+6x-1}{4x^2-2}\frac{1/x^2}{1/x^2} &= \lim_{x\to\infty} \frac{3x^2/x^2+6x/x^2-1/x^2}{4x^2/x^2-2/x^2} \\
&= \lim_{x\to\infty} \frac{3+6/x-1/x^2}{4-2/x^2} \\
&= \frac{3+\lim\limits_{x\to\infty}\frac{6}{x} - \lim\limits_{x\to\infty}\frac{1}{x^2}}{4-\lim\limits_{x\to\infty}\frac{2}{x^2}} \\
&= \frac{3}{4}
\end{aligned}$$

where in the last line we have used the fact that the quantities $\frac{1}{x}$ and $\frac{1}{x^2}$ go to zero as $x \to \infty$. ■

This suggests a general strategy for determining the limits of *rational functions*:

**Strategy for Rational Functions**

1. Determine the highest power $n$ which occurs in the functions,
2. Multiply and divide by $1/x^n$,
3. Take the limit as $x \to \infty$ using the fact that $1/x^p \to 0$ for all $p > 0$.

In the case of rational functions, this actually gives an incredibly convenient way of looking at a function and determining its asymptotic limit:

**Theorem 2.20**

Consider the rational function $f(x) = \dfrac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0}$.

1. If $n > m$ then $\lim_{x\to\infty} f(x) = \infty$.
2. If $n < m$ then $\lim_{x\to\infty} f(x) = 0$
3. If $n = m$ then $\lim_{x\to\infty} f(x) = \dfrac{a_n}{b_m}$.

As a matter of fact, the general strategy for determining the asymptotics of a function amounts to ascertaining which component of the function grows fastest. If the numerator grows fastest the function diverges, if the denominator grows fastest the function converges to 0, and if the numerator and denominator grow at the same rate, the function can attain a non-zero limit.

**Example 2.21**

Find the limit

$$\lim_{x\to\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}}.$$

*Solution.* Our goal is to divide by the "highest order term." In this case, that corresponds to the $e^x$ term. Can you see why this is true? In essence, we would like to get rid of the things the explode as $x \to \infty$, and keep the things that get small. This means we want to get rid of $e^x$ and keep $e^{-x}$ which is why we divide by $e^x$. We thus get

$$\begin{aligned} \lim_{x\to\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}} \frac{e^{-x}}{e^{-x}} &= \lim_{x\to\infty} \frac{1 + e^{-2x}}{1 - e^{-2x}} \\ &= \frac{1}{1} = 1. \end{aligned}$$

■

**Exercise**

What are the limits

$$\lim_{x\to 0^\pm} \frac{e^x + e^{-x}}{e^x - e^{-x}}, \qquad \lim_{x\to -\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}}.$$

**Example 2.22**

Determine the horizontal asymptotes of the function $f(x) = \dfrac{\sqrt{4x^2 + 5}}{x + 2}$.

*Solution.* The square root here presents many difficulties. In particular, while the numerator will always be positive, the denominator will change sign. We must ensure that we account for this. Furthermore, while the numerator might have an $x^2$ component, the square root means that the numerator effectively acts as $\sqrt{x^2} = |x|$ so only grows linearly rather than quadratically.

In the limit as $x \to \infty$ we get

$$\begin{aligned}
\lim_{x\to\infty} \frac{\sqrt{4x^2+5}}{x+2} &= \lim_{x\to\infty} \frac{\sqrt{4x^2+5}}{x+2}\frac{1/x}{1/x} \\
&= \lim_{x\to\infty} \frac{\sqrt{4+5/x^2}}{1+2/x} && \text{since } \frac{1}{x} = \frac{1}{\sqrt{x^2}} \text{ when } x > 0 \\
&= \frac{\sqrt{4 + \lim_{x\to\infty} \frac{5}{x^2}}}{1 + \lim_{x\to\infty} \frac{2}{x}} \\
&= \frac{\sqrt{4}}{1} = 2.
\end{aligned}$$

The tricky part above was that in order to pass the $1/x$ term into the square root, we needed to square it first. On the other hand, when we take the limit $x \to -\infty$, we will have that $\sqrt{x^2} = -x$ for $x < 0$, so that

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

Now a similar computation as that above yields

$$\begin{aligned}
\lim_{x\to-\infty} \frac{\sqrt{4x^2+5}}{x+2}\frac{1/x}{1/x} &= \lim_{x\to-\infty} -\frac{\sqrt{4+5/x^2}}{1+2/x} \\
&= -\frac{\sqrt{4}}{1} = -2.
\end{aligned}$$

Hence the horizontal asymptotes for $f(x)$ occur at $\pm 2$. ■

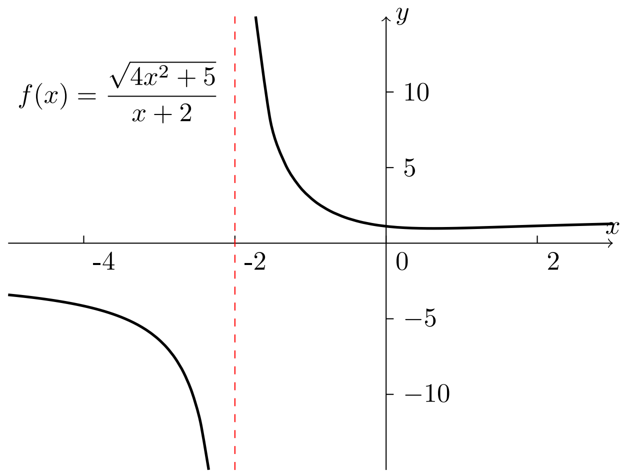



Figure 22: The plot of $f(x)$ for Example 2.22

**Example 2.23**

Find the limit

$$\lim_{x\to\infty} \sqrt{x^4+10} - x^2.$$

*Solution.* At first glance, this may look like it diverges, but be careful! The fact that there is an $x^4$ term under a square root means $\sqrt{x^4+10}$ behaves as $x^2$ as $x$ gets very big. This term will, in the long term, cancel the effect of the $x^2$ leaving a finite answer. In order to make this more precise, we multiply by the conjugate

$$\begin{aligned}\lim_{x\to\infty}\left[\left(\sqrt{x^4+10}-x^2\right)\left(\frac{\sqrt{x^4+10}+x^2}{\sqrt{x^4+10}+x^2}\right)\right] &= \lim_{x\to\infty}\frac{x^4+10-x^4}{\sqrt{x^4+10}+x^2}\\ &= \lim_{x\to\infty}\frac{10}{\sqrt{x^4+10}+x^2}\end{aligned}$$

Now the bottom term goes goes to $\infty$ as $x\to\infty$, so the whole limit goes to zero, and we conclude

$$\lim_{x\to\infty}\sqrt{x^4+10}-x^2=0. \qquad \blacksquare$$

## 2.6 Continuity

Of the examples seen thus far, we have seen instances in which the limit was easily computable by simply substituting the limiting value into the function, and other more pathological examples wherein this was not possible. The former examples are particularly special, not only because of the simplicity of evaluating limits, but because they tell us that the function is, in a sense, "well-behaved" at that limiting point.

**Definition 2.24**

We say that a function $f(x)$ is *continuous* at the point $c\in\mathbb{R}$ if

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

If the function $f(x)$ is continuous at all points in $\mathbb{R}$, we say that it is *continuous.*

In Example 2.7 we showed that

$$\lim_{x\to 1}\frac{x^4+7x+2}{x-4}=-10/3.$$

Substituting $x=1$ into this limiting function, we get

$$\frac{(1)^4+7(1)+2}{(1)-4}=\frac{10}{-3}.$$

so that the function $f(x)=(x^4+7x+2)/(x-4)$ is continuous at the point $x=1$.

In fact, we have already seen entire families of continuous functions. Recall from Theorem 2.8 that if $p(x)$ and $q(x)$ are polynomials and $q(c)\neq 0$ then

$$\lim_{x\to c}\frac{p(x)}{q(x)}=\frac{p(c)}{q(c)}.$$

This implies that all rational functions are continuous at points where their denominator does not vanish.

In general, proving that a function is continuous is a very complicated procedure which requires an in-depth knowledge of the properties of the function. For example, showing that $\sin(x)$ is continuous everywhere is rather technical and requires a large amount of geometry (perhaps unsurprisingly). Consequently, we shall omit the proofs that these functions are continuous. From here on, the student is allowed to assume that the following functions are continuous

1. All polynomials,
2. Root functions (example: $\sqrt{x}$, $\sqrt[3]{x}$),
3. The trigonometric functions $\sin(x)$ and $\cos(x)$,
4. Exponential functions $a^x$ for $a > 0$,
5. Logarithms $\log_a(x)$ for $a > 0$.

We can immediately deduce the following corollary from the Limit Laws

**Corollary 2.25**

If $f(x)$ and $g(x)$ are continuous at a point $c$, then

1. $\alpha f(x)$ is continuous at $c$, for any real number $\alpha$,
2. $f(x) \pm g(x)$ is continuous at $c$,
3. $f(x)g(x)$ is continuous at $c$,
4. $f(x)/g(x)$ is continuous at $c$, provided $g(c) \neq 0$.

*Proof.* I will give the proof for the sum $f(x) + g(x)$. The remaining proofs all follow precisely the same argument, and are essentially derivative of the Limit Laws.

To show that $f(x) + g(x)$ is continuous, our goal is to show that

$$\lim_{x \to c} [f(x) + g(x)] = f(c) + g(c).$$

Since both $f(x)$ and $g(x)$ are continuous at $c$, we know by hypothesis that the following limits exist:

$$\lim_{x \to c} f(x) = f(c), \qquad \lim_{x \to c} g(x) = g(c).$$

Since both limits exist, the Limit Laws imply that

$$\begin{aligned} \lim_{x \to c} [f(x) + g(x)] &= \left[\lim_{x \to c} f(x)\right] + \left[\lim_{x \to c} g(x)\right] \\ &= f(c) + g(c) \end{aligned}$$

which is precisely what we wanted to show. ☐

**Example 2.26**

Consider the function $f(x) = \dfrac{x^2+6x+2}{(x+4)(x^2-1)}$. Determine the points where $f(x)$ is not continuous.

*Solution.* This function fails to be continuous wherever its denominator vanishes. We may factor $(x+4)(x^2-1) = (x+4)(x-1)(x+1)$, which tells us that the points of discontinuity occur at $x = -4, -1, +1$. ■

Combining Corollary 2.25 with the list of functions assumed to be continuous, we can immediately generate very large families of continuous functions.

**Example 2.27**

Determine the limit $\displaystyle\lim_{x\to 4} \frac{x^2+\sqrt{x}}{2^x\cos(x)}$.

*Solution.* The numerator $x^2+\sqrt{x}$ is the sum of the continuous functions $x^2$ and $\sqrt{x}$, and hence is itself continuous. Similarly, the denominator $2^x\cos(x)$ is a product of continuous functions and hence is continuous. The quotient will be continuous so long as the denominator does not vanish at $x = 4$, and indeed

$$\lim_{x\to 4} 2^x\cos(x) = 2^4\cos(4) = 16\cos(4) \neq 0.$$

Thus the whole function $(x^2+\sqrt{x})/(2^x\cos(x))$ is continuous at $x = 4$ and the limit can be evaluated by substitution:

$$\lim_{x\to 4} \frac{x^2+\sqrt{x}}{2^x\cos(x)} = \frac{4^2+\sqrt{4}}{2^4\cos(4)} = \frac{18}{16\cos(4)} = \frac{9}{8\cos(4)}.$$ ■

**Composition of Functions:** Arguably, our most powerful operation on functions is that of composition, and it turns out that this will give us an incredible insight into what it means to be continuous. Recall that $f(x)$ is continuous at $c$ if the limit can be evaluated by simple substitution. This may be alternatively written as

$$\lim_{x\to c} f(x) = f(c) = f\left(\lim_{x\to c} x\right).$$

In a sense, it appears as though we are able to pass the limit inside of the function. In many ways, this is the true definition of continuity, so let's see how this might be useful to solving problems.

**Theorem 2.28**

If $g(x)$ has a limit as $c$, say $\lim_{x\to c} g(x) = L$ (note that $g(x)$ does not need to be continuous), and $f(x)$ is continuous at $L$, then

$$\lim_{x\to c} f(g(x)) = f\left(\lim_{x\to c} g(x)\right).$$

When $g(x) = x$ this reduces to the definition of continuity.

**Example 2.29**

Define the function

$$g(x) = \begin{cases} x & x \neq 0 \\ 2 & x = 0 \end{cases}.$$

Compute the limit $\lim_{x\to 0} e^{g(x)}$.

*Solution.* One might be tempted into thinking that $e^{g(x)}$ is continuous, but the fact that $g(x)$ fails to be continuous means that this is not the case. Indeed, if we attempted to substitution we would find that the limit would be $e^{g(0)} = e^2$ and this is not true (try graphing this and see for yourself). Instead, let us define $f(x) = e^x$ so that $e^{g(x)} = f(g(x))$. We know that $f(x)$ is then continuous, so by the previous theorem we have

$$\lim_{x\to c} f(g(x)) = f\left(\lim_{x\to c} g(x)\right) = e^{\lim_{x\to c} g(x)} = e^0 = 1. \qquad \blacksquare$$

An immediate result of Theorem 2.28 is the following:

**Corollary 2.30**

If $g(x)$ is continuous at $c$ and $f(x)$ is continuous at $g(c)$ then $f(g(x))$ is continuous at $c$.

*Proof.* Because both $f(x)$ and $g(x)$ are continuous at the necessary points, we just keep passing the limit into the arguments to see that

$$\lim_{x\to c} f(g(x)) = f\left(\lim_{x\to c} g(x)\right) = f\left(g\left(\lim_{x\to c} x\right)\right) = f(g(c)).$$

$\square$

Thus the composition of continuous functions is continuous. This can make complicated limits simple to solve.

**Example 2.31**

Compute the limit $\lim_{x\to 0} e^{\sin(x)}$.

*Solution.* Setting $f(x) = e^x$ and $g(x) = \sin(x)$ we get that $e^{\sin(x)} = f(g(x))$. Since both $f(x)$ and $g(x)$ are continuous, so too is $f(g(x))$ and the limit may be computed by simple substitution. Thus we get

$$\lim_{x\to 0} e^{\sin(x)} = e^{\sin(0)} = e^0 = 1. \qquad \blacksquare$$

### 2.6.1 One-Sided Continuity and Failures of Continuity

Just as there are one sided limits, we can consequently have one-sided continuity.

**Definition 2.32**

We say that $f(x)$ is continuous at $c$ *from the right* (resp. from the left) if

$$\lim_{x\to c^+} f(x) = f(c), \qquad \left(\text{resp. } \lim_{x\to c^-} f(x) = f(c)\right).$$

Certainly, any function which is continuous at $c$ will be continuous from both the left and the right at $c$. For an example of a function which is only continuous from a single side, consider the function

$$f(x) = \begin{cases} 4x+2 & x \leq 0 \\ -x & x > 0 \end{cases}.$$

At $x = 0$, $f(x)$ has the value $f(0) = 2$, and has one sided limits

$$\lim_{x\to 0^-} f(x) = 2, \qquad \lim_{x\to 0^+} f(x) = 0.$$

This implies that $f(x)$ is continuous from the left, as $\lim_{x\to 0^-} f(x) = f(0)$, but not from the right.

**Example 2.33**

Consider the function

$$f(x) = \begin{cases} \frac{x}{|x|} & \text{if } x \neq 0 \\ c & \text{if } x = 0. \end{cases}.$$

Determine the value of $c$ such that $f$ is continuous from the left at 0. What value of $c$ makes $f$ continuous from the right at 0?

*Solution.* All we need to do is determine the limit as $x \to 0^-$ and set $c$ to be this number. In this limit we may assume that $x < 0$, so that $|x| = -x$, and we get

$$\lim_{x\to 0^-} f(x) = \lim_{x\to 0^-} \frac{x}{-x} = \lim_{x\to 0^-} -1 = -1.$$

so we set $c = -1$. Doing the same analysis for the limit as $x \to 0^+$ we get $c = 1$. ■

If a limit fails to exist then it certainly has no chance of being continuous. On the other hand, the function $f(x) = (x-3)/(x^2 - x + 12)$ has a finite, two-sided limit at $x = 3$, but fails to be continuous there since $f(x)$ is not defined at $x = 3$. The fact that continuities seem inherently "worse" than other leads to a classification of discontinuities.

**Definition 2.34**

If $f(x)$ is a function, define the one-sided limits

$$L_+ = \lim_{x\to c^+} f(x), \qquad L_- = \lim_{x\to c^-} f(x).$$

If $f(x)$ fails to be continuous at $c$, we say that $c$ is

1. A *removable discontinuity* if both $L_+$ and $L_-$ exist and $L_+ = L_-$.
2. A *jump discontinuity* if $L_+, L_-$ exist but $L_+ \neq L_-$.
3. An *essential discontinuity* if one of $L_\pm$ does not exist or is infinite.

**Example:**

1. The function $f(x) = \frac{x-3}{x^2-x+12}$ has a removable discontinuity at $x = 3$.
2. The function $f(x) = x/|x|$ has a jump discontinuity at $x = 0$.
3. The function $f(x) = \sin(1/x)$ has an essential discontinuity at $x = 0$.

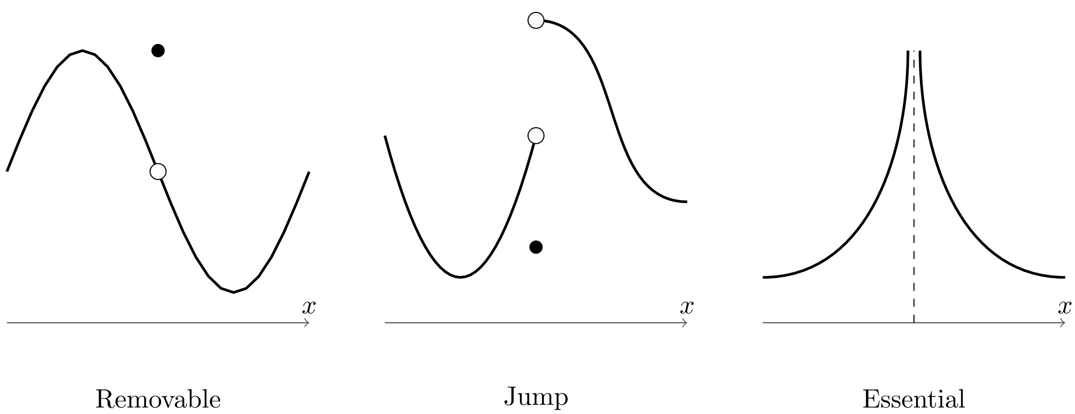



Figure 23: Examples of the types of discontinuity that can occur.

# 3 Derivatives

The power of calculus is that it gives us to tools to analyze how things change in time, or more precisely the instantaneous rate of change. This will be done by exploiting properties of limits.

## 3.1 First Principles

In Section 2 we considered the problem of measuring the speed of a race car. This was done by measuring the distance travelled by the car over successively smaller time intervals. While this was our motivation for introducing limits, we never finished the example and actually applied a limit to the problem. We now remedy that oversight.

Let $f(t)$ represent the position of the car at time $t$, and let $t_{\text{f}}$ denote the time at which the car passes the finish line. The average speed of the car in the second after passing the finish line is given by

$$\begin{matrix}\text{average speed}\\ \text{after 1 second}\end{matrix} = \frac{\text{distance travelled}}{\text{time elapsed}} = \frac{f(t_0 + 1\text{ sec}) - f(t_0)}{1\text{ sec}}.$$

More generally by letting $h$ denote an arbitrary quantity of time the average speed,

$$\begin{matrix}\text{average speed}\\ \text{after } h \text{ seconds}\end{matrix} = \frac{\text{distance}}{\text{time}} = \frac{f(t_0 + h) - f(t_0)}{h},$$

where if $h > 0$ then we are measuring after crossing the finish line, and if $h < 0$ we are measuring prior to crossing the finish line. The instantaneous speed of the car may then be determined by taking the limit as our time interval becomes arbitrarily small; that is, by taking $h \to 0$.

$$\begin{matrix}\text{instantaneous speed}\\ \text{at time } t_0\end{matrix} = \lim_{h\to 0} \frac{f(t_0 + h) - f(t_0)}{h}.$$

There is no reason why we cannot generalize this discussion to a more abstract setting or to other examples. The key point is that having precise knowledge about a function $f$ near a given point $t_0$ allows us to determine how quickly $f$ is changing at $t_0$.

**Definition 3.1**

Let $f(x)$ be a function and $x_0$ a point in the domain of $f(x)$. If the limit

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{3.1}$$

exists, then we say that $f(x)$ is *differentiable at* $x_0$ and $f'(x_0)$ is the *derivative of* $f(x)$ *at* $x_0$. If (3.1) exists for all $x$ in the domain of $f(x)$, we simply say that $f(x)$ is *differentiable*.

**Example 3.2**

Compute the derivative of $f(x) = x$ and $g(x) = x^2$ at the points $x = -1$ and $x = 5$.

*Solution.* With nothing more than Definition 3.1 to work with, we set to work. For the linear function $f(x) = x$, at $x = -1$ we get

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

while at $x = 5$ we find

$$f'(5) = \lim_{h\to 0} \frac{f(5 + h) - f(5)}{h} = \lim_{h\to 0} \frac{(5 + h) - 5}{h} = 1.$$

This shows that the derivative of $f(x) = x$ is the same at both $x = -1$ and $x = 5$. In fact, it is not too hard to convince ourselves that regardless of what value we substitute for $x_0$, we are always going to get $f'(x_0) = 1$. This is tantamount to the fact that $f'(x_0)$ describes the rate of change of the function, and $f(x) = x$ is function which is growing at a constant rate.

Before proceeding to the next example, let us try to anticipate the kind of solution we will see from the function $g(x) = x^2$. The shape of the parabola seems to suggest that $x^2$ is decreasing at $x = -1$ and so $g'(-1)$ should be negative. On the other hand, $g(x)$ is growing larger at $x = 5$ implying that $g'(5)$ should be positive. The necessary computations are slightly more complicated, but nothing beyond our capabilities:

$$\begin{aligned} g'(-1) &= \lim_{h\to 0} \frac{g(-1+h) - g(-1)}{h} = \lim_{h\to 0} \frac{(-1+h)^2 - (-1)^2}{h} \\ &= \lim_{h\to 0} \frac{(1-2h+h^2) - 1}{h} = \lim_{h\to 0} -2 + h \\ &= -2 \end{aligned}$$

$$\begin{aligned} g'(5) &= \lim_{h\to 0} \frac{g(5+h) - g(5)}{h} = \lim_{h\to 0} \frac{(5+h)^2 - 5^2}{h} \\ &= \lim_{h\to 0} \frac{h^2 + 10h + 25 - 25}{h} = \lim_{h\to 0} h + 10 \\ &= 10. \end{aligned}$$

The is precisely the behaviour we anticipated. Even further, the rate at which $g(x)$ is growing at $x = 5$ is much faster than the speed at which it is decreasing at $x = -1$. This is corroborated by the graph of $g(x) = x^2$: we expect the magnitude of the rate of change of $g(x)$ to increasing with the magnitude of $x$. ■

**Example 3.3**

Compute the derivative of the function $f(x) = \dfrac{3x+2}{x-4}$ at the point $x = 5$.

*Solution.* The behaviour of this function is not as transparent as that given in Example 3.2, so we must trust in our calculations. As before, we apply Definition 3.1 to $f(x)$ to find

$$\begin{aligned} f'(5) &= \lim_{h\to 0} \frac{f(5+h) - f(5)}{h} = \lim_{h\to 0} \frac{\frac{3(5+h)+2}{(5+h)-4} - 17}{h} \\ &= \lim_{h\to 0} \frac{\frac{3h+17}{h+1} - 17}{h} = \lim_{h\to 0} \frac{\frac{(3h+17)-(17h+17)}{h+1}}{h} && \text{common denominator} \\ &= \lim_{h\to 0} \frac{-14h}{h(h+1)} = \lim_{h\to 0} \frac{-14}{h+1} \\ &= -14. \end{aligned}$$

This tells us that the function is decreasing at $x = 5$. ■

The computations of the derivatives in Example 3.2 where redundant, requiring only small changes between the $x = -1$ and $x = 5$ cases. If instead of using a particular number, we let $x = a$ describe a general point, we could compute

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

This agrees with what we found when $a = -1$ and $a = 5$, but gives us the derivative at any point $a$. Hence one can view the derivative of a function as a function itself, with

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

In fact, this is where the word *derivative* comes from: the value of the function $f'(x)$ is derived from that of $f(x)$.

**Exercise:** How is the domain of $f'(x)$ related to $f(x)$? Must they always be the same? Or can one be larger than the other?

### 3.1.1 The Geometry of the Derivative

We can visualize the derivative geometrically by examining the graph of $f(x)$.

**Definition 3.4**

Consider the graph of a function $f(x)$ and let $a < b$ be distinct real numbers. The *secant line from a to b* is the unique straight line which passes through the points $(a, f(a))$ and $(b, f(b))$.

Of interest is the slope of this secant line, given by

$$m_{ab} = \frac{f(b) - f(a)}{b - a}.$$

In fact, this slope describes the average change in the value of $f(x)$ between $f(a)$ and $f(b)$. We can get successively better approximations to the instantaneous rate of change of a function by taking the distance between $a$ and $b$ to be successively smaller, and the instantaneous rate of change will be given by taking a limit as $a \to b$. See Figure 24. Notice that in this limit the secant line becomes a *tangent line*; that is,

If $f(x)$ is differentiable, the slope of the tangent line to the graph of $f(x)$ at $a$ is $f'(a)$.

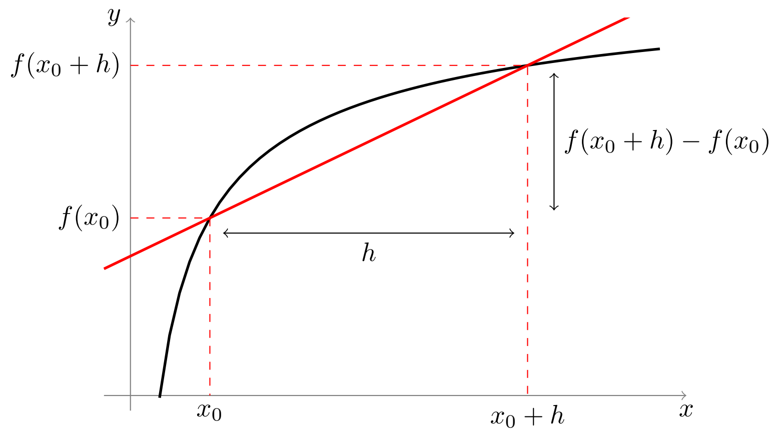



Figure 24

Since we have a slope and a point on in the plane $(a, f(a))$, we can form the equation of the tangent line through $f(x)$ at $a$ using the point-slope formula:

$$y - f(a) = f'(a)(x - a). \tag{3.2}$$

**Example 3.5**

Determine the equation of the tangent line through $f(x) = \sqrt{x}$ at the point $x = 1$.

*Solution.* To determine the slope we need the value $f'(1)$, though it is not extra work to compute the derivative in general.

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \\
&= \lim_{h\to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \left[\frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}}\right] && \text{Multiply by conjugate} \\
&= \lim_{h\to 0} \frac{x + h - x}{h(\sqrt{x+h} + \sqrt{x})} \\
&= \lim_{h\to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} = \frac{1}{2\sqrt{x}}.
\end{aligned}$$

At $x = 1$ we have $f(1) = \sqrt{1} = 1$, so we know our line passes through the point $(1, 1)$. Our derivative formula tells us that $f'(1) = \frac{1}{2}$, so using our point slope formula, we thus deduce that the equation of the tangent line is $y - 1 = \frac{1}{2}(x - 1)$ or $y = \frac{1}{2}x + \frac{1}{2}$. ■

**Example 3.6**

Find the equation of the tangent line to the graph $f(x) = 1/x$ at the point $x = 1$.

*Solution.* While (3.2) tells us that it is sufficient to find $f'(1)$, it is not much extra work to find $f'(x)$ in general:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{\frac{1}{x+h}-\frac{1}{x}}{h} \\ &= \lim_{h\to 0} \frac{x-(x+h)}{hx(x+h)} = \lim_{h\to 0} \frac{-1}{x(x+h)} \\ &= -\frac{1}{x^2}, \end{aligned}$$

hence $f'(1) = -1$. To find the equation of the line, we need only determine a point through which the line passes. Since the line is tangent to $f(x)$ at $x = 1$, it must pass through the point $(1, f(1)) = (1, 1)$. In conclusion, (3.2) tells us that the tangent line is

$$y - 1 = -1(x-1), \qquad \text{or equivalently} \qquad y = -x + 2. \qquad \blacksquare$$

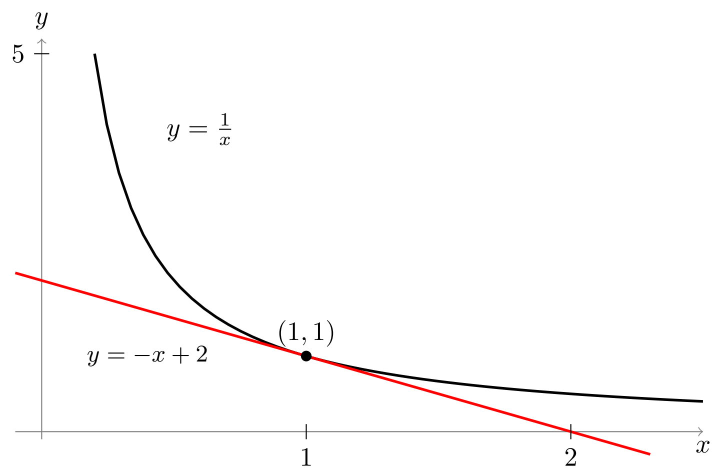



Figure 25: The derivative of $f(x) = 1/x$ at the point $x = 1$ gives us the slope of the tangent line to the graph at $x = 1$. Combining this with the fact that the line must pass through the point $(1, f(1)) = (1, 1)$ we know precisely the equation of the tangent line.

### 3.1.2 A Different Parameterization

An alternative way of writing the derivative comes from changing how we parameterize the limit. Recall from Definition 3.1 that

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

where the parameter $h$ denotes the distance from $a$. Taking the limit $h \to 0$ corresponding to letting our point $a+h$ get close to $a$. If we instead set $x = a+h$, the condition $h \to 0$ is equivalent to $x \to a$, giving an equivalent definition of the derivative:

$$f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} \tag{3.3}$$

**Example 3.7**

Compute the derivative of the function $f(x) = 9/x$ for any point $a \neq 0$ using (3.3).

*Solution.* Applying Equation (3.3) we get

$$\begin{aligned} f'(a) &= \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = \lim_{x\to a} \frac{9/x - 9/a}{x-a} \\ &= \lim_{x\to a} \frac{9(a-x)/xa}{x-a} = \lim_{x\to a} \frac{-9}{ax} \\ &= -\frac{9}{a^2}. \end{aligned}$$

■

### 3.1.3 Relating Variables and Leibniz Notation

Functions can be used to describe how one variable depends upon another. For example, the identity $y = x^2$ tells us precisely how the variable $y$ depends upon the function $x$. Given a general function $f(x)$, a functional relation between $y$ and $x$ can be written as $y = f(x)$. In this context, the derivative $f'(x)$ describes the instantaneous rate of change of $x$ as a function of $x$.

When the relationship is defined explicitly, say $y = x^2$, we need a way of talking about the derivatives without referring to the defining function $f(x) = x^2$. To do this, we write

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=c} = f'(c), \quad \text{or just } \frac{\mathrm{d}y}{\mathrm{d}x} \text{ when we are feeling lazy.}$$

The notation $\frac{\mathrm{d}y}{\mathrm{d}x}$ represents how the variable $y$ is changing with respect to $x$. The motivation for this notation comes from the definition of the derivative as the limit of secant lines. If $a < b$ are two distinct real numbers, the change in the $y$-value of the function $f(x)$ between $a$ and $b$ may be written as $\Delta y = f(b) - f(a)$ (this is still often used amongst physicists), with the change in the $x$-value being written as $\Delta x = b - a$. The secant line between $a$ and $b$ is then

$$m_{ab}^f = \frac{\Delta y}{\Delta x} \qquad \text{“}\xrightarrow{\Delta x\to 0}\text{”} \qquad \frac{\mathrm{d}y}{\mathrm{d}x}$$

and in the limit as $\Delta x \to 0$ these delta's “transform” into $d$'s.

**Example 3.8**

Determine $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ if $y = (x^2+1)/(2x)$.

*Solution.* Applying the definition of the derivative, one finds that

$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \lim_{h\to 0} \frac{\frac{(x+h)^2+1}{2(x+h)} - \frac{x^2+1}{2x}}{h} \\
&= \lim_{h\to 0} \frac{\frac{(x^2+2xh+h^2)(2x)-(x^2+1)(2x+2h)}{(2x)(2x+2h)}}{h} \\
&= \lim_{h\to 0} \frac{(2x^3+4x^2h+2x)-(2x^3-2x^2h-2x-2h)}{h(2x)(2x+2h)} \\
&= \lim_{h\to 0} \frac{2x^2h-2h}{h(2x)(2x+h)} = \frac{2x^2-2}{4x^2} \\
&= \frac{x^2-1}{2x^2}.
\end{aligned}$$

■

Leibniz notation let's us think of differentiation as an *operator*; that is, something which acts on functions to create new functions. If $f(x)$ is a function, we will use the notation

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x)$$

to represent the action of taking the derivative of $f$ with respect to $x$.

## 3.2 Some Derivative Results

Similar to how the limit laws simplified the process of taking limits, we wish to establish a collection of tools to simplify the process of differentiation. In this section, we give a few formulas to simplify some calculations, and examine how to differentiate polynomials and exponential functions.

### 3.2.1 Linearity and the Power Rule

Taking derivatives plays nicely with addition and multiplication by a constant, as the following proposition illustrates:

**Proposition 3.9**

If $f(x)$ and $g(x)$ are differentiable at $c$, then

1. For any constant $\alpha \in \mathbb{R}$ the function $(\alpha f)(x) = \alpha f(x)$ is differentiable at $c$, and moreover
$$\frac{\mathrm{d}}{\mathrm{d}x}[\alpha f(x)] = \alpha f'(x).$$
2. The function $(f+g)(x) = f(x) + g(x)$ is differentiable at $c$, and moreover
$$\frac{\mathrm{d}}{\mathrm{d}x}[f(x)+g(x)] = f'(x) + g'(x).$$

*Proof.* 1. By the definition of the derivative, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\alpha f(x)\right] = \lim_{h\to 0}\frac{\alpha f(x+h)-\alpha f(x)}{h} = \alpha\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \alpha f'(x).$$

2. Again, using the definition of the derivative we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)+g(x)\right] &= \lim_{h\to 0}\frac{[f(x+h)+g(x+h)]-[f(x)+g(x)]}{h} \\ &= \lim_{h\to 0}\frac{[f(x+h)-f(x)]+[g(x+h)-g(x)]}{h} && \text{re-arranging terms} \\ &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}+\lim_{h\to 0}\frac{g(x+h)-g(x)}{h} && \text{by the limit laws} \\ &= f'(x)+g'(x).\end{aligned}$$

□

The previous theorem tells us that $\frac{\mathrm{d}}{\mathrm{d}x}$ is what is called a *linear operator*, in that it preserves scalar multiplication and addition. As an immediate corollary, we see that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)-g(x)\right] = \frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)+(-1)\times g(x)\right] = f'(x)+(-1)\times g'(x) = f'(x)-g'(x)$$

so that the derivative of the difference is also the difference of the derivatives.

**Proposition 3.10**

For any positive integer $n$, $\frac{\mathrm{d}}{\mathrm{d}x}x^n = nx^{n-1}$.

*Proof.* Recall that one may always factor the $n^{th}$-powered difference of two elements as

$$(a^n-b^n) = (a-b)(a^{n-1}+a^{n-2}b^1+\cdots+a^1b^{n-2}+b^{n-1}).$$

Applying this to the definition of the derivative, we see that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}x^n &= \lim_{y\to x}\frac{y^n-x^n}{y-x} = \lim_{y\to x}\frac{(y-x)(y^{n-1}+y^{n-2}x+\cdots+yx^{n-2}+x^{n-1})}{y-x} \\ & \lim_{y\to x}\underbrace{y^{n-1}+y^{n-2}x+\cdots+yx^{n-2}+x^{n-1}}_{n\text{-times}} \\ &= nx^{n-1}.\end{aligned}$$

□

Combining Propositions 3.9 and 3.10 allow us to differentiate polynomials very quickly. Every polynomial function is built from scalar multiplication and addition of monomials of the form $a_nx^n$, so

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x}\left[a_nx^n+a_{n-1}x^{n-1}+\cdots+a_2x^2+a_1x+a+0\right] \\ &= a_n\left[\frac{\mathrm{d}}{\mathrm{d}x}x^n\right]+a_{n-1}\left[\frac{\mathrm{d}}{\mathrm{d}x}x^{n-1}\right]+\cdots+a_1\left[\frac{\mathrm{d}}{\mathrm{d}x}x\right]+0 \\ &= na_nx^{n-1}+(n-1)a_{n-1}x^{n-2}+\cdots+2a_2x+a_1.\end{aligned}$$

**Example 3.11**

Compute the derivative of $x^{654} + 13x^{45} - 20x^5 + 6$.

*Solution.* Using our template above, we see that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left[x^{654} + 13x^{45} - 20x^5 + 6\right] &= (654)x^{653} + 13(45)x^{44} - 20(5)x^4 \\ &= 654x^{653} + 585x^{44} - 100x^4. \end{aligned} \qquad \blacksquare$$

While the above proposition only showed that $\frac{\mathrm{d}}{\mathrm{d}x}x^n = nx^{n-1}$ for any *positive integer* $n$, it turns out that the formula works for all real numbers. We will see why this is the case in Section 3.5.1.

### 3.2.2 The Natural Exponent

One way of defining Euler's number $e$, defined in Section 1.6.3, is as the unique number which satisfies

$$1 = \lim_{h\to 0} \frac{e^h - 1}{h}; \tag{3.4}$$

that is, it is the unique number for which the slope of the tangent line at the point which crosses the $y$-axis is precisely 1. Using this, we may actually compute the derivative of $e^x$ as follows:

**Proposition 3.12**

The exponential function $e^x$ is its own derivative; that is, $\dfrac{\mathrm{d}}{\mathrm{d}x}e^x = e^x$

*Proof.* By definition of the derivative,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}e^x &= \lim_{h\to 0}\frac{e^{x+h} - e^x}{h} = \lim_{h\to 0}\frac{e^x e^h - e^x}{h} \\ &= e^x \lim_{h\to 0}\frac{e^h - 1}{h} && \text{by (3.4)} \\ &= e^x. \end{aligned} \qquad \square$$

A similar argument will actually show that for any $a > 0$, $\frac{\mathrm{d}}{\mathrm{d}x}a^x = Ka^x$ for some $K$, which satisfies $K = \frac{\mathrm{d}}{\mathrm{d}x}\big|_{x=0} a^x$. We will later see that $K = \log(a)$, but for the moment we do not have the tools to make this clear.

**Exercise:** Recall that if $n$ is a positive integer, we define $n! = n(n-1)(n-2)\cdots(3)(2)$. For example, $3! = 3 \times 2 = 6$ and $4! = 4 \times 3 \times 2 = 24$. Consider the function

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots.$$

Compute the derivative of $f(x)$. If you are given that derivatives are unique up to additive constants, what function must $f(x)$ be?

### 3.2.3 The Product and Quotient Rule

Computing the derivative of sums of functions was ultimately rather simple. However, it turns out that computing the derivative of a product is a far more complicated affair.

**Theorem 3.13**

If $f(x)$ and $g(x)$ are differentiable at $c$, then $f(x)g(x)$ is differentiable at $c$ and

$$\left.\frac{\mathrm{d}}{\mathrm{d}x}\right|_{x=c} f(x)g(x) = f'(c)g(c) + g'(c)f(c).$$

*Proof.* By the limit definition of the derivative, we have

$$\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}c}\right|_{c=c} f(c)g(c) &= \lim_{h\to 0} \frac{f(c+h)g(c+h) - f(c)g(c)}{h} \\
&= \lim_{h\to 0} \frac{f(c+h)g(c+h) + f(c+h)g(c) - f(c+h)g(c) - f(c)g(c)}{h} \\
&= \lim_{h\to 0} f(c+h)\frac{g(c+h)-g(c)}{h} + g(c)\frac{f(c+h)-f(c)}{h} \\
&= \left[\lim_{h\to 0} f(c+h)\right]\left[\lim_{h\to 0}\frac{g(c+h)-g(c)}{h}\right] + g(c)\left[\lim_{h\to 0}\frac{f(c+h)-f(c)}{h}\right] \\
&= f(c)g'(c) + g(c)f'(c).
\end{aligned}$$

□

**Example 3.14**

Let $f(x)$ be a differentiable function such that $f'(x) = 1/f(x)$. Compute $\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)\right]^2$.

*Solution.* Applying the product rule, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)\right]^2 = f'(x)f(x) + f'(x)f(x) = 2f'(x)f(x).$$

Now since we were told that $f'(x) = 1/f(x)$ we may substitute this to find that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[f(x)\right]^2 = 2f'(x)f(x) = 2\frac{f(x)}{f(x)} = 2.$$

■

**Exercise:** Find a differentiable function that satisfies $f'(x) = [f(x)]^{-1}$ as in Example 3.14.

**Example 3.15**

Let $f(x)$ be differentiable and satisfy $f(1) = 1$ and $f'(1) = 2$. Compute $g'(1)$ where $g(x) = f(x)/x$.

*Solution.* We have already seen that $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{x} = -\frac{1}{x^2}$, so the product rule tells us that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}g(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left[f(x) \times \frac{1}{x}\right] = f'(x)\left(\frac{1}{x}\right) + f(x)\left(\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{x}\right) \\ &= \frac{f'(x)}{x} + f(x)\left(-\frac{1}{x^2}\right) \\ &= \frac{f'(x)}{x} - \frac{f(x)}{x^2}.\end{aligned}$$

If we now substitute $x = 1$ into this equation we find

$$g'(1) = \frac{f'(1)}{1} - \frac{f(1)}{1^2} = \frac{2}{1} - \frac{1}{1} = 2 - 1 = 1.$$ ∎

In fact, there is no reason to limit ourselves to considering the product of only two functions. If $f(x), g(x)$, and $h(x)$ are all differentiable then

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x)g(x)h(x) = f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x).$$

The way to see this is to define a new function $n(x) = g(x)h(x)$ so that $n'(x) = g'(x)h(x)+g(x)h'(x)$ and $f(x)g(x)h(x) = f(x)n(x)$. Since the right-hand-side is a product of two functions, the product rule again gives us

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}f(x)n(x) &= f'(x)n(x) + f(x)n'(x) = f'(x)g(x)h(x) + f(x)\left[g'(x)h(x) + g(x)h'(x)\right] \\ &= f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x)\end{aligned}$$

and this process is easily generalized to any number of functions.

**Theorem 3.16: The Quotient Rule**

If $f(x)$ and $g(x)$ are differentiable at $c$ and $g(c) \neq 0$ then $f(x)/g(x)$ is differentiable at $c$ and

$$\left.\frac{\mathrm{d}}{\mathrm{d}c}\right|_{x=c}\frac{f(x)}{g(x)} = \frac{f'(c)g(c) - f(c)g'(c)}{[g(c)]^2}.$$

**Example 3.17**

Compute the derivative of $f(x) = \dfrac{x^2 - 2x + 1}{x^4 + 4}$.

*Solution.* Let us write $g(x) = x^2 - 2x + 1$ and $h(x) = x^4 + 4$ so that $f(x) = g(x)/h(x)$. We know that $g'(x) = 2x - 2$ and $h'(x) = 4x^3$ so

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\frac{x^2-2x+1}{x^4+4} &= \frac{g'(x)h(x) - g(x)h'(x)}{[h(x)]^2} \\ &= \frac{(2x-2)(x^4+4) - (4x^3)(x^2-2x+1)}{[x^4+4]^2} \\ &= \frac{-2x^5+6x^4-4x^3+8x-8}{[x^4+4]^2}\end{aligned}$$

■

**Example 3.18**

Confirm the computation of the derivative of $f(x)/x$ given in Example 3.15.

*Solution.* Applying the quotient rule to $f(x)/x$ we find that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\frac{f(x)}{x} &= \frac{f'(x)x - \left(\frac{\mathrm{d}}{\mathrm{d}x}x\right)f(x)}{[x^2]} \\ &= \frac{xf'(x) - f(x)}{x^2}.\end{aligned}$$

In Example 3.15 we found that

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{f(x)}{x} = \frac{f'(x)}{x} - \frac{f(x)}{x^2} = \frac{xf'(x) - f(x)}{x^2}$$

exactly as expected. ■

**Example 3.19**

If $g(x) = x/e^x$, find an expression for $g^{(n)}(x)$.

*Solution.* Once we have learned the chain rule, this example will be much more easily computed by using the product rule on $g(x) = xe^{-x}$. For the moment though we must content ourselves with using the quotient rule. Since we would like to find a general expression for the $n^{th}$ derivative of $g(x)$, we will start by computing the first few derivatives and see if we can find a pattern. The first several derivatives are as follows:

$$\begin{aligned}f'(x) &= \frac{e^x - xe^x}{e^{2x}} = \frac{1-x}{e^x} \\ f''(x) &= \frac{-e^x - (1-x)e^x}{e^{2x}} = \frac{-1-1+x}{e^x} = \frac{-(2-x)}{e^x} \\ f'''(x) &= \frac{e^x - (x-2)e^x}{e^{2x}} = \frac{1-x+2}{e^x} = \frac{3-x}{e^x}.\end{aligned}$$

The pattern would suggest that in general,

$$g^{(n)}(x) = \frac{(-1)^{n-1}(n-x)}{e^x}.$$

■

### 3.2.4 Higher Order Derivatives

Differentiating a function $f(x)$ resulted in another function $f'(x)$. If $f'(x)$ is itself differentiable then we can apply the derivative again to find the second derivative $f''(x)$. If $f''(x)$ is differentiable, we can differentiate a third time to get $f'''(x)$, and so on.

When using the prime notation becomes too cumbers on, we let $f^{(n)}(x)$ denote the $n^{th}$ derivative of $f(x)$. These have important interpretations in both mathematics and science which we shall explore later. In Leibniz notation, we use the operator $\frac{\mathrm{d}}{\mathrm{d}x}$ to take subsequent derivatives, hence if $y = f(x)$ then the first derivative is $\frac{\mathrm{d}y}{\mathrm{d}x}$, while the second, third, and fourth derivatives are

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}^2y}{\mathrm{d}x^2}, \qquad \frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\mathrm{d}^3y}{\mathrm{d}x^3}, \qquad \frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}^3y}{\mathrm{d}x^3} = \frac{\mathrm{d}^4y}{\mathrm{d}x^4}$$

respectively, with the pattern continuing *ad infinitum.* to

**Example 3.20**

Compute the second derivative of the function $f(x) = 1/x$. Determine a formula for the $n^{\text{th}}$ derivative $f^{(n)}(x)$.

*Solution.* In Example 3.6 we showed that $f'(x) = -1/x^2$. To compute $f''(x)$ we take the derivate of $f'(x)$ using the quotient rule, and find that

$$f''(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(-\frac{1}{x^2}\right) = \frac{2}{x^3}.$$

Were we to continue on in this fashion, the higher order derivatives would be computed to be

$$f'''(x) = \frac{-6}{x^4}, \quad f^{(4)}(x) = \frac{24}{x^5}, \quad f^{(5)}(x) = -\frac{120}{x^6}, \ldots.$$

In general, the $n$-th derivative of $f$ is given by

$$f^{(n)}(x) = \frac{(-1)^n n!}{x^{n+1}}. \qquad \blacksquare$$

## 3.3 Trigonometric Derivatives

We now explore the derivatives of the trigonometric functions $\sin(x)$ and $\cos(x)$. Since all other trigonometric functions, such as $\tan(x)$ and $\sec(x)$, are built from quotients of these two functions, our knowledge of the product and quotient rules will allow us to compute the derivatives of all the remaining fundamental trigonometric functions.

### 3.3.1 Two Important Limits

The proof of the following two limits requires some geometry and the Squeeze Theorem. They are good exercises in their own right, but their derivation is beyond the scope of this course.

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

This essentially tells us that $\sin(\theta)$ and $\theta$ converge to zero at approximately the same rate, while $\cos(\theta)-1$ goes to zero strictly faster than $\theta$. In fact, the above two results may be generalized as follows: If $k\neq 0$ is any real number, then

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

**Example 3.21**

Determine the limit $\displaystyle\lim_{x\to 0}\frac{\sin(3x)}{x}$.

*Solution.* The above limit is of the form $0/0$ so we may not simply substitute $x=0$ into the formula. On the other hand, we only have information on how $\sin(3x)/3x$ behaves, and not $\sin(3x)/x$. Our solution to this should be to manipulate $\sin(3x)/x$ into looking like $\sin(3x)/3x$ by multiplying and dividing by 3. This gives us

$$\lim_{x\to 0}\frac{\sin(3x)}{x}\frac{3}{3}=3\lim_{x\to 0}\frac{\sin(3x)}{3x}=3 \qquad \blacksquare$$

**Example 3.22**

Determine the limit $\displaystyle\lim_{x\to 0}\frac{\sin^2(4x)}{3x^2}$.

*Solution.* The presence of the $\sin^2(4x)$ term should not intimidate us as we know how to deal with the limit of a product. Indeed, our goal should be to rewrite this as

$$\frac{\sin^2(4x)}{16x^2}=\frac{\sin^2(4x)}{(4x^2)}=\frac{\sin(4x)}{4x}\frac{\sin(4x)}{4x}.$$

Since the denominator is only $3x^2$ as compared to the necessary $16x^2$, we multiply and divide everything by $16/3$, so that

$$\begin{aligned}\lim_{x\to 0}\frac{\sin^2(4x)}{3x^2}\frac{16/3}{16/3}&=\frac{16}{3}\lim_{x\to 0}\frac{\sin^2(4x)}{\frac{16}{3}3x^2}\\&=\frac{16}{3}\lim_{x\to 0}\frac{\sin^2(4x)}{16x^2}=\frac{16}{3}\left[\lim_{x\to 0}\frac{\sin(4x)}{4x}\right]\left[\lim_{x\to 0}\frac{\sin(4x)}{4x}\right]\\&=\frac{16}{3}.\end{aligned} \qquad \blacksquare$$

### 3.3.2 Differentiating Sine and Cosine

To differentiate sine and cosine, we will use the two limits above as well as the angle sum identities:

$$\sin(x \pm y) = \sin(x)\cos(y) \pm \cos(x)\sin(y), \quad \cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y).$$

**Theorem 3.23**

The functions $\sin(x)$ and $\cos(x)$ are everywhere differentiable, and in particular

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x), \qquad \frac{\mathrm{d}}{\mathrm{d}x}\cos(x) = -\sin(x).$$

*Proof.* We will demonstrate the proof for the derivative of $\sin(x)$ and leave the proof for $\cos(x)$ as an exercise for the student. Applying the definition of the derivative we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) &= \lim_{h\to 0}\frac{\sin(x+h)-\sin(x)}{h} \\
&= \lim_{h\to 0}\frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h} && \text{by the double angle identity} \\
&= \lim_{h\to 0}\sin(x)\frac{\cos(h)-1}{h}+\cos(x)\frac{\sin(h)}{h} \\
&= \sin(x)\underbrace{\left[\lim_{h\to 0}\frac{\cos(h)-1}{h}\right]}_{=0}+\cos(x)\underbrace{\left[\lim_{h\to 0}\frac{\sin(h)}{h}\right]}_{=1} && \text{by the limit laws} \\
&= \cos(x).
\end{aligned}$$

□

**Example 3.24**

Using the quotient rule, compute the derivative of $\tan(x)$.

*Solution.* Since $\tan(x) = \sin(x)/\cos(x)$ we can apply the quotient rule to find

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\tan(x) &= \frac{\left[\frac{\mathrm{d}}{\mathrm{d}x}\sin(x)\right]\cos(x)-\sin(x)\left[\frac{\mathrm{d}}{\mathrm{d}x}\cos(x)\right]}{\cos^2(x)} \\
&= \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)} \\
&= \sec^2(x).
\end{aligned}$$

■

Computing the remaining trigonometric derivatives is an excellent exercise in the use of the quotient rule, but we shall provide here a table of the trigonometric derivatives:

As a general rule, many of these identities are not worth remembering. Instead, the student should focus on how to *derive* these relationships using the quotient rule, so that the formulae can be called upon when needed.

| | |
|---|---|
| $\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x)$ | $\frac{\mathrm{d}}{\mathrm{d}x}\csc(x) = -\csc(x)\cot(x)$ |
| $\frac{\mathrm{d}}{\mathrm{d}x}\cos(x) = -\sin(x)$ | $\frac{\mathrm{d}}{\mathrm{d}x}\sec(x) = \sec(x)\tan(x)$ |
| $\frac{\mathrm{d}}{\mathrm{d}x}\tan(x) = \sec^2(x)$ | $\frac{\mathrm{d}}{\mathrm{d}x}\cot(x) = -\csc^2(x)$ |

Table 2: List of Trigonometric Derivatives.

It is not too hard to see that by iteratively differentiating sine/cosine, we eventually cycle back on ourselves:

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x), \quad \frac{\mathrm{d}^2}{\mathrm{d}x^2}\sin(x) = -\sin(x), \quad \frac{\mathrm{d}^3}{\mathrm{d}x^2} = -\cos(x), \quad \frac{\mathrm{d}^4}{\mathrm{d}x^4}\sin(x) = \sin(x),$$

so differentiating the sine function four times gives us back the function back (or alternatively, differentiate the function twice gives us back the negative of the function). The student should check that precisely the same argument works with the cosine function.

### 3.3.3 Degrees versus Radians

It was essential that we used radians in lieu of degrees when doing all of our trigonometric limits. The reason is precisely because the identity $\lim_{x\to 0}\frac{\sin(x)}{x} = 1$ would no longer be true, and would instead given a value of $\frac{180}{\pi}$. One can then see how this would affect our calculations of the derivative by either imitating the proof of Theorem 3.23 or using the fact that if $r$ is radians and $\theta$ is degrees, then $\theta = \frac{180}{\pi}r$. Indeed,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\sin(r) = \sin\left(\frac{180}{\pi}\theta\right) = \frac{180}{\pi}\cos\left(\frac{180}{\pi}\theta\right).$$

This destroys the nice symmetry between the derivatives of sine and cosine, and introduces a pesky scaling factor every time we differentiate.

Even more importantly, since we learn calculus from the point of view of radians, if the student is ever asked to do a computation, using degrees will actually give the wrong answer. The moral of the story: Always use radians.

## 3.4 Smoothness of Differentiable Functions

Differentiable functions are well behaved functions, in that their graphs appear to be smooth. This will be made more explicit when we discuss the different ways a function can fail to be differentiable, but we will start by discussion the relationship between continuity and differentiability.

### 3.4.1 Differentiable implies Continuous

**Proposition 3.25**

If $f(x)$ is differentiable at $a$, then $f(x)$ is continuous at $a$.

*Proof.* Let $f(x)$ be differentiable at the point $a$, so that

$$\lim_{x\to a} \frac{f(x)-f(a)}{x-a} = L \text{ exists and is finite.} \tag{3.5}$$

To show continuity, we wish to show that

$$\lim_{x\to a} f(x) = f(a), \qquad \text{or equivalently} \qquad \lim_{x\to a} [f(x)-f(a)] = 0.$$

At this point we say to ourselves, "what part of the hypothesis have we failed to use?" Well, differentiability! Since our function is differentiable, we know something about the limit $(f(x) - f(a))/(x-a)$ and now wish to say something about $f(x)-f(a)$. Multiplying and dividing by the quantity $(x-a)$ we find

$$\begin{aligned}
\lim_{x\to a} [f(x)-f(a)] &= \lim_{x\to a} \frac{f(x)-f(a)}{x-a}(x-a) \\
&= \left[\lim_{x\to a} \frac{f(x)-f(a)}{x-a}\right]\left[\lim_{x\to a} x-a\right] && \text{By the limit laws since both limits exist} \\
&= L \times 0 = 0
\end{aligned}$$

as desired. $\square$

It can be easy to confuse these two notions, so take a moment to get the direction of the implication correct. If a function is differentiable, then it is continuous. In the next section, we will see that the opposite direction is not true; namely, there are functions which are continuous but not differentiable.

### 3.4.2 Failures of Differentiability

There are three ways in which a function which fail to be differentiable.

1. The slope of the tangent line cannot be determined at a point.
2. The function may fail to be continuous.
3. The function may have a vertical tangent line.

The first condition can arise when each of the one-sided limits

$$\lim_{h\to 0^+} \frac{f(x+h)-f(x)}{h}, \qquad \lim_{h\to 0^-} \frac{f(x+h)-f(x)}{h}$$

exist, but are not equal. Since differentiability is defined in terms of the two sided limit, this would represent a failure of differentiability. This often presents in the graph of $f(x)$ as a 'corner' or jagged edge. Since all differentiable functions are continuous, if a function is not continuous then it certainly cannot be differentiable! Finally, a vertical tangent line has slope "infinity," meaning that the derivative diverges at that point.

**Example 3.26**

Show that the function

$$f(x) = \begin{cases} x^2 & x \geq -1 \\ x & x < -1 \end{cases}$$

is not differentiable at $x = -1$.

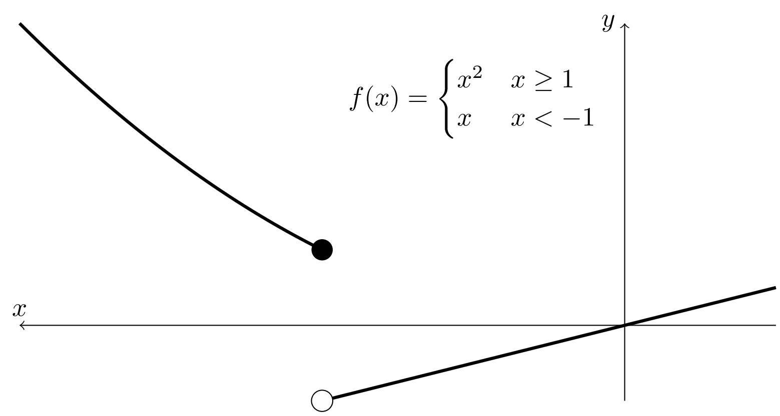



Figure 26: The graph of the function from Example 3.26. Since the function is not continuous at $x = -1$, it certainly cannot be differentiable there.

*Solution.* This function has a jump discontinuity at $x = -1$, and one cannot help but think this might affect its differentiability. Indeed, the one sided limits of the derivative yield

$$\begin{aligned} \lim_{h \to 0^+} \frac{f(1+h) - f(1)}{h} &= \lim_{h \to 0} \frac{(1+h)^2 - 1}{h} = \lim_{h \to 0} \frac{h^2 + 2h}{h} = 2 \\ \lim_{h \to 0^-} \frac{f(1+h) - f(1)}{h} &= \lim_{h \to 0} \frac{(1+h) - 1}{h} = \lim_{h \to 0} \frac{h}{h} = 1 \end{aligned}$$

and these certainly do not agree. We conclude non-differentiability at $x = -1$ as required. ∎

The discontinuity above feels (and is) contrived, so we offer a more enlightening example.

**Example 3.27**

Show that the function $f(x) = |x|$ is not differentiable at $x = 0$.

*Solution.* Unlike Example 3.26, the function $f(x) = |x|$ is continuous. Nonetheless, the left- and

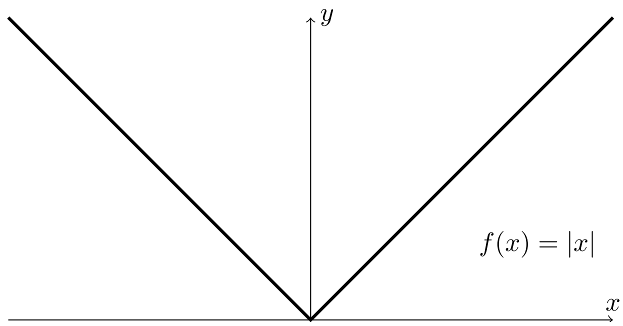



Figure 27: The graph of the function from Example 3.27. The corner at $x = 0$ keeps this function from being differentiable there.

right-sided limits disagree:

$$\lim_{h\to 0^+} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0^+} \frac{|h|}{h} = \lim_{h\to 0} \frac{h}{h} = 1$$
$$\lim_{h\to 0^-} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0^-} \frac{|h|}{h} = \lim_{h\to 0} \frac{-h}{h} = -1.$$

So continuous functions need not be differentiable! We saw in Proposition 3.25 that the converse is true: Every differentiable function is continuous. The proof of this fact is provided in Section (ref). ■

The pathology in the previous example is due to the corner which occurs at $x = 0$. Since the derivative measures the rate of change of a function, the jarring change of $|x|$ at the origin is accountable for the lack of differentiability.

**Example 3.28**

The function $f(x) = \sqrt[3]{x}$ is not differentiable at $x = 0$.

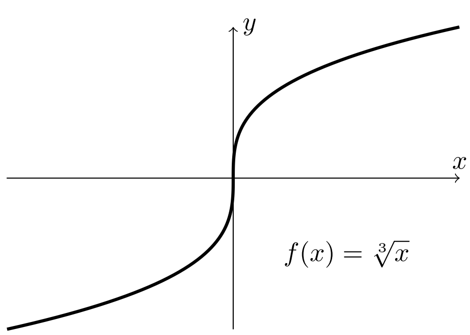



Figure 28: The graph of the function for Example 3.28. Note that the tangent line at $x = 0$ is vertical, and so has infinite slope.

*Solution.* This function does not have the abrupt changes exhibited in Example 3.27, but still fails to be differentiable at 0. Our one-sided limits give

$$\lim_{h\to 0^+} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0} \frac{\sqrt[3]{h}}{h} = \lim_{h\to 0} \frac{1}{h^{2/3}} = \infty.$$

This implies there is a sense in which the derivative exists, it is just that the tangent line has infinite slope; that is, the tangent line is vertical. As we would prefer to avoid having the incorporation of infinities into our formalism, we say that $f(x)$ is not differentiable at 0 all the same. ■

## 3.5 Chains and Inverses

We have seen how to take the derivatives of sums, products, and quotients of functions. The only major operation left is to look at function composition. Interestingly, determine how to differentiate a composition will also give us access to the derivative of inverse functions.

### 3.5.1 The Chain Rule

Why are we interested in differentiating compositions? Let's say that you're a weatherperson and you want to determine the outside temperature over the course of the day. You realize that the temperature $H$ depends on the amount of sunshine $s$, and they are related through a function $H = f(s)$. This makes perfect sense, and you can even use calculus to determine $\frac{\mathrm{d}H}{\mathrm{d}s}$, the instantaneous rate of change of temperature with respect to sunshine.

But maybe you find it difficult to measure the amount of sunshine directly. Instead, you realize that sunshine itself is a function of time $t$; that is, you can write $s = g(t)$ for some function $g$. This allows you to determine $\frac{\mathrm{d}s}{\mathrm{d}t}$, the instantaneous rate of sunshine with respect to time.

The composition $H = f(s) = f(g(t))$ now tells you how the temperature depends on time. However, to differentiate this function you need to be able to differentiate the composition $f \circ g$. It seems like you should be able to do this; after all, you know how temperature changes with sunshine, and how sunshine changes with time:

$$\text{time} \longrightarrow \text{sunshine} \longrightarrow \text{temperature.}$$

Here, sunshine just acts as an intermediary for getting from time to temperature, and we can write $T = f(s) = f(g(t))$. This is our goal.

**Theorem 3.29: Chain Rule**

If $f(x)$ and $g(x)$ are functions such that $g(x)$ is differentiable at $c$ and $f(x)$ is differentiable at $g(c)$, then the composition $f \circ g$ is differentiable at $c$ and $(f \circ g)'(c) = f'(g(c))g'(c)$.
In Leibniz notation, if $w = f(y)$ and $y = g(x)$ then

$$\left.\frac{\mathrm{d}w}{\mathrm{d}x}\right|_a = \left.\frac{\mathrm{d}w}{\mathrm{d}y}\right|_{g(a)} \left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_a . \tag{3.6}$$

**Example 3.30**

Compute the derivative $\frac{\mathrm{d}}{\mathrm{d}x}(x^2+2x-4)^{200}$.

*Solution.* If the only technique we know is the power rule, evaluation of this integral would require us to expand the 200-fold product of $x^2+2x-4$: what a mess! Instead, let us define the function $f(x) = x^{200}$ and $g(x) = x^2+2x-4$ so that $(x^2+2x-4)^{200} = f(g(x))$. Using the fact that we know $f'(x) = 200x^{199}$ and $g'(x) = 2x+2$ the chain rule then gives us

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(x^2+2x-4)^{200} &= \frac{\mathrm{d}}{\mathrm{d}x}f(g(x)) \\ &= f'(g(x))g'(x) \\ &= 200(x^2+2x-4)^{199}(2x+2).\end{aligned}$$

■

**Example 3.31**

Compute the derivative of $\sqrt{x+\sqrt{x}}$.

*Solution.* We need to realize $\sqrt{x+\sqrt{x}}$ as the composition of two functions. In particular, let $f(x) = \sqrt{x}$ and $g(x) = x+\sqrt{x}$ so that

$$f(g(x)) = \sqrt{g(x)} = \sqrt{x+\sqrt{x}}.$$

Now we know that $f'(x) = 1/(2\sqrt{x})$ and $g'(x) = 1+1/(2\sqrt{x})$ so using the chain rule we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{x+\sqrt{x}} &= \frac{\mathrm{d}}{\mathrm{d}x}f(g(x)) = f'(g(x))g'(x) \\ &= \frac{1}{2\sqrt{g(x)}}\left(1+\frac{1}{2\sqrt{x}}\right) \\ &= \frac{2\sqrt{x}+1}{4\sqrt{x^2+x\sqrt{x}}}.\end{aligned}$$

■

In Leibniz notation, set $y = f(g(x))$ and $u = g(x)$ so that $y = f(u)$. One could compute the derivative $\frac{\mathrm{d}y}{\mathrm{d}u} = f'(u)$ with no problem: This describes how the variable $y$ changes with respect to the variable $u$. However, if we want to know how $y$ changes with respect to the variable $x$, the chain rule is then written as

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}x}$$

which makes it look surprisingly like fraction-cancellation.

As with the product rule, we may also extend the chain rule to three or more functions. For example, given the function $f(g(h(x)))$ let us temporarily define a new function $n(x) = g(h(x))$, whose derivative is computed by the chain rule to be $n'(x) = g'(h(x))h'(x)$. To compute our three-fold composition, we then have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x} f(g(h(x))) &= \frac{\mathrm{d}}{\mathrm{d}x} f(n(x)) \\ &= f'(n(x))n'(x) \\ &= f'(g(h(x)))g'(h(x))h'(x).\end{aligned}$$

To me, this looks a lot like a collection of Matryoshka dolls!

**Example 3.32**

Compute the derivative $\dfrac{\mathrm{d}}{\mathrm{d}x} e^{\sin^2(x)}$.

*Solution.* There are really three function compositions occurring in this question. Let $f(x) = e^x, g(x) = x^2$, and $h(x) = \sin(x)$ so that $e^{\sin^2(x)} = f(g(h(x)))$. The corresponding derivatives are $f'(x) = e^x, g'(x) = 2x$ and $h'(x) = \cos(x)$ and the chain rule gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x} e^{\sin^2(x)} &= \frac{\mathrm{d}}{\mathrm{d}x} f(g(h(x))) = f'(g(h(x)))g'(h(x))h'(x) \\ &= \underbrace{e^{\sin^2(x)}}_{f'(g(h(x)))} \underbrace{2\sin(x)}_{g'(h(x))} \underbrace{\cos(x)}_{h'(x)}.\end{aligned}$$

■

**Example 3.33**

If $y = \sin(\pi w)$, $w = \sqrt{z}$ and $z = x^2 + 4x - 1$, compute $\left.\dfrac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=1}$.

*Solution.* There are two ways in which we may proceed. The first is to actually compose all of the functions and evaluate. In this case the composition yields $y = \sin(\pi\sqrt{x^2 + 4x - 1})$ and using the chain rule we can its derivative to be

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\pi}{2} \frac{\cos(\pi\sqrt{x^2 + 4x - 1})(2x + 4)}{\sqrt{x^2 + 4x - 1}}.$$

Evaluating at $x = 1$ gives the value $\frac{3\pi}{2}$. The alternative technique is to use Leibniz notation: notice that at $x = 1$ we have $z = x^2 + 4x - 1 = 4$, $w = \sqrt{z} = 2$ and $y = \sin(\pi w) = 0$. This means that

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=1} = \left[\frac{\mathrm{d}y}{\mathrm{d}w}\frac{\mathrm{d}w}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}x}\right]_{x=1} = \left.\frac{\mathrm{d}y}{\mathrm{d}w}\right|_{w=2} \left.\frac{\mathrm{d}w}{\mathrm{d}z}\right|_{z=4} \left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{x=1}.$$

These may all be computed separately. Indeed, notice that $\frac{\mathrm{d}y}{\mathrm{d}w} = \pi\cos(\pi w), \frac{\mathrm{d}w}{\mathrm{d}z} = \frac{1}{2\sqrt{z}}$, and $\frac{\mathrm{d}z}{\mathrm{d}x} = 2x + 4$. Hence

$$\left.\frac{\mathrm{d}y}{\mathrm{d}w}\right|_{w=2} = \pi, \qquad \left.\frac{\mathrm{d}w}{\mathrm{d}z}\right|_{z=4} = \frac{1}{4}, \qquad \left.\frac{\mathrm{d}z}{\mathrm{d}x}\right|_{x=1} = 6$$

which we may combine all together to find

$$\left.\frac{\mathrm{d}y}{\mathrm{d}w}\right|_{w=2} \left.\frac{\mathrm{d}w}{\mathrm{d}z}\right|_{z=4} \left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{x=1} = \pi \times \frac{1}{4} \times 6 = \frac{3\pi}{2}.$$ ■

**Example 3.34**

Compute the $26^{th}$ derivative of $\cos(2x)$.

*Solution.* As in all previous cases, we begin by simply computing the first few derivatives to see if we can find a pattern:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\cos(2x) &= -2\sin(2x) \\
\frac{\mathrm{d}^2}{\mathrm{d}x^2}\cos(2x) &= -4\cos(2x) \\
\frac{\mathrm{d}^3}{\mathrm{d}x^3}\cos(2x) &= 8\sin(2x) \\
\frac{\mathrm{d}^4}{\mathrm{d}x^4}\cos(2x) &= 16\cos(2x).
\end{aligned}$$

What we notice here is that after each 4 derivatives the function returns (even up to sign) to $\cos(2x)$, with the number in front being $2^n$ where $n$ is the value of the derivative. This means that the $24^{th}$ derivative will be

$$\frac{\mathrm{d}^{24}}{\mathrm{d}x^{24}}\cos(2x) = 2^{24}\cos(2x).$$

Since we are looking for the $26^{th}$ derivative, we differentiate two more times to find

$$\frac{\mathrm{d}^{26}}{\mathrm{d}x^{26}}\cos(2x) = -2^{26}\cos(2x).$$ ■

**Example 3.35**

Let $f(x)$ be a differentiable function. Show that for any positive integer $n$

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x)^n = nf(x)^{n-1}f'(x).$$

*Solution.* By setting $g(x) = x^n$ we have that $f(x)^n = g(f(x))$. Furthermore, $g'(x) = nx^{n-1}$, so applying the chain rule we have

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x)^n = g'(f(x))f'(x) = nf(x)^{n-1}f'(x)$$

as required. ■

**Example 3.36**

Denote by $f^{\circ n}(x)$ the $n$-fold composition of $f$. For example, $f^{\circ 3}(x) = f(f(f(x)))$. Assuming that $f(1) = 1$ and $f'(1) = 2$ find the derivative of $f^{\circ n}$ evaluated at $x = 1$.

*Solution.* This is just a repeated exercise of the chain rule requiring a small bit of trickery. First, we notice that since $f(1) = 1$ then no matter how many times we compose by $f$, we will always get 1. More explicitly, notice that $f(f(1)) = f(1) = 1$, and in general $f^{\circ n}(1) = 1$. For the sake of intuition, let us try this when $n = 3$. Notice in this case that

$$(f^{\circ 3})'(x) = f'(f(f(x))) \cdot f'(f(x)) \cdot f'(x)$$

so that

$$(f^{\circ 3})'(1) = f'(f(f(1))) \cdot f'(f(1) \cdot f'(1) = f'(1) \cdot f'(1) \cdot f'(1) = \left[f'(1)\right]^3 = 8.$$

In fact, precisely the same procedure will work for general $n$, and we get

$$(f^{\circ n})'(x) = f'(f^{\circ(n-1)}(x))f'(f^{\circ(n-2)}(x)) \cdots f'(f(x))f'(x) = \prod_{i=0}^{n-1} f'(f^{\circ i}(x))$$

where $f^{\circ 0} = x$. Hence we get

$$(f^{\circ n})'(1) = \underbrace{f'(1) \cdots f'(1)}_{n-\text{times}} = 2^n. \qquad \blacksquare$$

Just as we were able to use the product rule to extend the power rule from $n \in \mathbb{N}$ to $n \in \mathbb{Z}$, we can use the product rule to tell us something about inverse functions (we will discuss this more in a subsequent section).

### 3.5.2 Derivatives of Inverse Functions

**Theorem 3.37: Inverse Function Theorem**

Let $f(x)$ be differentiable at the point $c$ with $f'(x)$ continuous at $c$. If $f'(c) \neq 0$, then there is an interval $I$ containing $c$ on which $f$ is invertible. Moreover, the inverse $f^{-1}$ is differentiable with continuous derivative, and satisfies the formula

$$\left(f^{-1}\right)'(x) = \frac{1}{f'(f^{-1}(x))}. \tag{3.7}$$

The majority of this theorem is beyond our abilities, but deriving Equation (3.7) is not too difficult. Let us begin by assuming that we have been given $f^{-1}$ and we know that it is differentiable. By definition of the inverse function we have $f(f^{-1}(x)) = x$ for all $x$ in the range of $f$. Differentiating both sides (applying the chain rule to the composition) we then get

$$1 = \frac{\mathrm{d}}{\mathrm{d}x} f(f^{-1}(x)) = f'(f^{-1}(x))(f^{-1})'(x).$$

We can solve for $(f^{-1})'(x)$ by re-arranging to get $(f^{-1})'(x) = \left[f(f^{-1}(x))\right]^{-1}$ as required. Again, this is an instance in which the derivation of the formula is so easy that it would be wasteful to memorize (3.7). Instead, we emphasize that the student should focus on the derivation itself.

**Proposition 3.38**

The derivative of $\log(x)$ is
$$\frac{\mathrm{d}}{\mathrm{d}x}\log(x) = \frac{1}{x}.$$

*Proof.* It is possible to do this using first principles, but the proof turns out to be incredibly difficult. Instead, we use here the fact that $e^x$ and $\log(x)$ are inverses, so that $e^{\log(x)} = x$. Set $f(x) = e^x$ and $f^{-1}(x) = \log(x)$, so that by the Inverse Function Theorem we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\log(x) = (f^{-1})'(x) &= \frac{1}{f'(f^{-1}(x))} \\ &= \frac{1}{e^{\log(x)}} = \frac{1}{x}.\end{aligned}$$ □

Now that we have of $\log(x)$, we can generalize the Power Rule, and differentiate general exponential functions $f(x) = a^x$:

**Theorem 3.39: Generalized Power Rule**

If $n$ is *any* real number, then $f(x) = x^n$ is differentiable for all $x > 0$, and moreover
$$\frac{\mathrm{d}}{\mathrm{d}x}x^n = nx^{n-1}.$$

*Proof.* Let $n$ be any real number. By properties of the logarithm and exponential, we can write
$$x^n = e^{\log(x^n)} = e^{n\log(x)}.$$

Setting $f(x) = e^x$ and $g(x) = n\log(x)$, we can differentiate $e^{n\log(x)} = f(g(x))$ using the chain rule:
$$\frac{\mathrm{d}}{\mathrm{d}x}x^n = \frac{\mathrm{d}}{\mathrm{d}x}e^{n\log(x)} = f'(g(x))g'(x) = \frac{n}{x}e^{n\log(x)} = \frac{n}{x}x^n = nx^{n-1},$$
which is precisely what we wanted to show. □

**Theorem 3.40**

If $a > 0$ then the function $f(x) = a^x$ is differentiable for all $x$, and moreover
$$\frac{\mathrm{d}}{\mathrm{d}x}a^x = \log(a)a^x.$$

*Proof.* Using the properties of exponents and logarithms, we can write

$$a^x = e^{\log(a^x)} = e^{x\log(a)}.$$

Using the chain rule, we set $f(x) = e^x$ and $g(x) = x\log(a)$ so that $f'(x) = e^x$ and $g(x) = \log(a)$ giving

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}a^x = \frac{\mathrm{d}}{\mathrm{d}x}e^{x\log(a)} &= \frac{\mathrm{d}}{\mathrm{d}x}f(g(x)) \\ &= f'(g(x))g'(x) = e^{x\log(a)}\log(a) \\ &= \log(a)a^x. \end{aligned}$$

□

### 3.5.3 The Inverse Trigonometric Functions

The trigonometric functions are not invertible on their domains, but we can restrict those domains to create functions which, for all intents and purposes, are invertible and still trigonometric. In the case of $\sin(x)$, there are many possible intervals that we could choose. For the sake of simplicity, we choose the maximal interval at the origin around which the horizontal line test is satisfied; namely, $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. In the case of $\cos(x)$ we take $[0, \pi]$. On these intervals, $\sin(x)$ has an inverse $\arcsin(x)$ and $\cos(x)$ has an inverse $\arccos(x)$. We can perform a similar analysis of all the remaining trigonometric functions and deduce the following table:

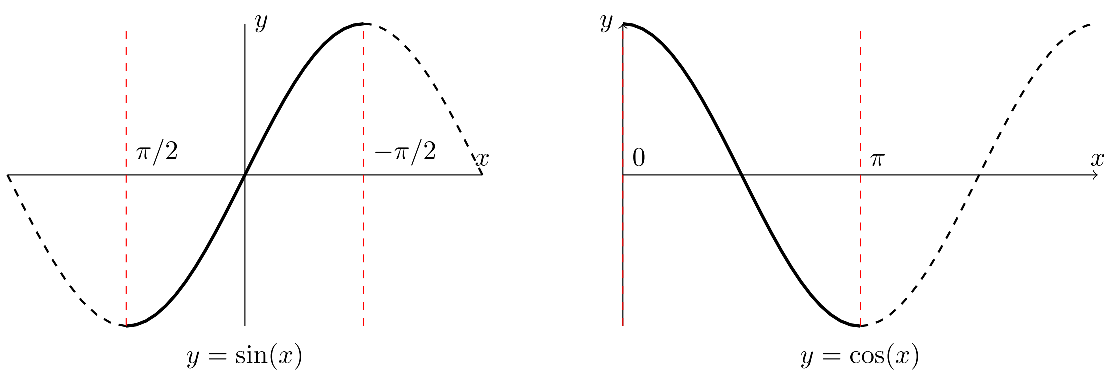



Figure 29: By restricting the domain of sine and cosine between the dashed lines, we can ensure that the horizontal line test holds. These restricted functions are thus invertible.

| Name | Definition | Notation | Maps | | |
|---|---|---|---|---|---|
| arcsine | $\theta = \arcsin(x)$ | $\sin(\theta) = x$ | $[-1, 1]$ | $\longrightarrow$ | $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ |
| arccosine | $\theta = \arccos(x)$ | $\cos(\theta) = x$ | $[-1, 1]$ | $\longrightarrow$ | $[0, \pi]$ |
| arctangent | $\theta = \arctan(x)$ | $\tan(\theta) = x$ | $\mathbb{R}$ | $\longrightarrow$ | $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ |
| arccosecant | $\theta = \operatorname{arccsc}(x)$ | $\csc(\theta) = x$ | $\mathbb{R} \setminus (-1, 1)$ | $\longrightarrow$ | $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \setminus \{0\}$ |
| arcsecant | $\theta = \operatorname{arcsec}(x)$ | $\sec(\theta) = x$ | $\mathbb{R} \setminus (-1, 1)$ | $\longrightarrow$ | $[0, \pi] \setminus \left\{\frac{\pi}{2}\right\}$ |
| arccotangent | $\theta = \operatorname{arccot}(x)$ | $\cot(\theta) = x$ | $\mathbb{R}$ | $\longrightarrow$ | $(0, \pi)$ |

We can use the Inverse Function Theorem (Theorem 3.37) to compute the derivatives of the inverse trigonometric functions.

**Theorem 3.41**

The function $\arcsin : [-1, 1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is differentiable on $(-1, 1)$ and its derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}x} \arcsin(x) = \frac{1}{\sqrt{1-x^2}}. \tag{3.8}$$

*Proof.* Set $f(x) = \sin(x)$ so that $f^{-1}(x) = \arcsin(x)$. Note that $\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x)$ is never zero on the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ and hence by the Inverse Function Theorem, we know that $f^{-1}(x) = \arcsin(x)$ is differentiable on $(-1, 1)$. Now we know that on $(-1, 1)$ we have $\sin(\arcsin(x)) = x$ and hence differentiating both sides we have[5]

$$1 = \cos(\arcsin(x)) \left[\frac{\mathrm{d}}{\mathrm{d}x} \arcsin(x)\right] \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}x} \arcsin(x) = \frac{1}{\cos(\arcsin(x))}.$$

This is the derivative, but does not agree with the statement given in (3.8). To deduce $\cos(\arcsin(x))$, set $y = \arcsin(x)$ so that $\sin(y) = x$ (see Figure 30). From the figure, we immediately read off that $\cos(y) = \cos(\arcsin(x)) = \sqrt{1-x^2}$ which gives us

$$\frac{\mathrm{d}}{\mathrm{d}x} \arcsin(x) = \frac{1}{\cos(\arcsin(x))} = \frac{1}{\sqrt{1-x^2}}$$

as required. □

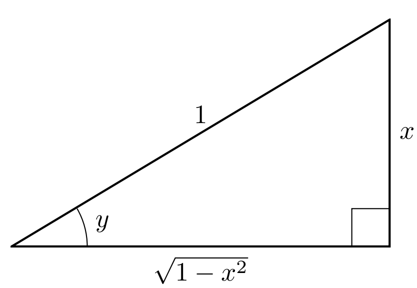



Figure 30: The right-triangle corresponding to setting $y = \arcsin(x)$.

The student is encouraged to attempt deriving the remaining functions, which I will provide below in an easily readable table[6]:

[5] This is of course equivalent to using (3.7).
[6] Be careful with $\operatorname{arccsc}(x)$, since it can be quite tricky depending on how you do it.

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\arcsin(x) = \frac{1}{\sqrt{1-x^2}} && \frac{\mathrm{d}}{\mathrm{d}x}\operatorname{arccsc}(x) = -\frac{1}{|x|\sqrt{x^2-1}} \\
&\frac{\mathrm{d}}{\mathrm{d}x}\arccos(x) = -\frac{1}{\sqrt{1-x^2}} && \frac{\mathrm{d}}{\mathrm{d}x}\operatorname{arcsec}(x) = \frac{1}{|x|\sqrt{x^2-1}} \\
&\frac{\mathrm{d}}{\mathrm{d}x}\arctan(x) = \frac{1}{1+x^2} && \frac{\mathrm{d}}{\mathrm{d}x}\operatorname{arccot}(x) = -\frac{1}{1+x^2}.
\end{aligned}
$$

### 3.5.4 Logarithmic Differentiation

One of the great things about logarithms is there is a sense in which they decrease the complexity of an operation. For example, we often think of addition as being easier than multiplication, and multiplication being easier than exponents:

$$\log(xy) = \log(x) + \log(y), \qquad \log(x^y) = y \log x.$$

At the cost of introducing a logarithm, we are able to convert product to sums, and powers to products! Since the logarithm is not very hard to differentiate, this does not seem like such a terrible cost.

This idea in general is known as *logarithmic differentiation*. Where it can be particularly useful is when we have a product/quotient of many objects which are individually simple to differentiate, but which will become complicated when nested with the product rule. For example, given a collection of functions $f_1(x), \ldots, f_n(x)$ and $g_1(x), \ldots, g_m(x)$, notice that we can write

$$\log\left[\frac{f_1(x)\cdots f_n(x)}{g_1(x)\cdots g_m(x)}\right] = \log f_1(x) + \cdots + \log f_n(x) - \log g_1(x) - \cdots - \log g_m(x).$$

Hence implicit differentiation yields

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\frac{f_1(x)\cdots f_n(x)}{g_1(x)\cdots g_m(x)} &= \frac{f_1(x)\cdots f_n(x)}{g_1(x)\cdots g_m(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left[\log f_1(x) + \cdots + \log f_n(x) - \log g_1(x) - \cdots - \log g_m(x)\right] \\
&= \frac{f_1(x)\cdots f_n(x)}{g_1(x)\cdots g_m(x)}\left[\frac{f_1'(x)}{f_1(x)} + \cdots + \frac{f_n'(x)}{f_n(x)} - \frac{g_1'(x)}{g_1(x)} - \cdots - \frac{g_m'(x)}{g_m(x)}\right]
\end{aligned}
$$

**Example 3.42**

Compute the derivative of $f(x) = \dfrac{(x-1)^2(x^2+2)\sqrt{x}}{x^4+5}$.

*Solution.* This would be an absolute nightmare to compute using the quotient rule, so instead we use logarithm differentiation. Taking the logarithm of both sides yields:

$$\log f(x) = 2\log(x-1) + \log(x^2+2) + \frac{1}{2}\log(x) - \log(x^4+5).$$

Differentiating implicitly gives

$$\begin{aligned} f'(x) &= f(x)\frac{\mathrm{d}}{\mathrm{d}x}\left[2\log(x-1)+\log(x^2+2)+\frac{1}{2}\log(x)-\log(x^4+5)\right] \\ &= \frac{(x-1)^2(x^3+2)\sqrt{x}}{x^4+5}\left[\frac{2}{x-1}+\frac{2x}{x^2+2}+\frac{1}{2x}-\frac{4x^3}{x^4+5}\right]. \end{aligned}$$ ■

This takes care of converting products to sums, but now what about powers to products? Given two functions $f(x)$ and $g(x)$, let's try to differentiate $f(x)^{g(x)}$. The problem here is that neither the power rule, nor the rules for differentiating exponents can apply (in both of those cases, the function should only occur in the power or the base, but not both). To deal with this, we set $y = f(x)^{g(x)}$ so that $\log y = g(x)\log f(x)$. We can now differentiate implicitly:

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = g'(x)\log f(x) + g(x)\frac{f'(x)}{f(x)}$$

which we may then solve for $\frac{\mathrm{d}y}{\mathrm{d}x}$ to get

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x)^{g(x)}\left[g'(x)\log f(x) + g(x)\frac{f'(x)}{f(x)}\right]. \tag{3.9}$$

Like many of the formulae that we've derived, Equation (3.9) is not worth remembering on its own. Rather, what is important is remembering how this derivation was performed so that it can be repeated when necessary.

**Example 3.43**

Compute the derivative of $x^{\sin(x)}$.

*Solution.* Setting $y = x^{\sin(x)}$ we have $\log y = \sin(x)\log x$. Differentiating implicitly we get

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \cos(x)\log(x) + \frac{\sin(x)}{x}$$

which we may solve for $\frac{\mathrm{d}y}{\mathrm{d}x}$ to get

$$\frac{\mathrm{d}y}{\mathrm{d}x} = x^{\sin(x)}\left[\cos(x)\log(x) + \frac{\sin(x)}{x}\right].$$ ■

# 4 Applications of Derivatives

We have developed numerous hammers, and the time has come to start looking for nails. The next several section will discuss how we can apply derivatives to solve applied problems.

## 4.1 Rates of Change

We already saw at the beginning of this section how derivatives can be used to deduce the *instantaneous rate of change* of one quantity with respect to another. The power of this is that it allows us to take boring, static relationships, and differentiate them to discover the dynamic interplay between variables.

**Example 4.1**

The volume of a sphere of radius $r$ is $V = \frac{4}{3}\pi r^3$. Determine how the volume changes as $r$ is allowed to vary.

*Solution.* All we need to do is differentiate the given expression with respect to $r$ to find that

$$\frac{\mathrm{d}V}{\mathrm{d}r} = 4\pi r^2.$$

This says that the rate of change of volume $V$ with respect to the radius $r$ is $4\pi r^2$. As an example, this means that doubling the radius of a sphere will quadruple its volume, while tripling the radius will increase the volume 9-fold. ■

One of the more utilized relationships is that of position, velocity, acceleration, jerk, etc. If $p(t)$ describes the position of an object with respect to time, then $p'(t)$ is its velocity and $p''(t)$ is its acceleration with respect to time. This can be used to model physical situations which can then be solved by mathematical methods:

### 4.1.1 Gravity

**Example 4.2**

Consider a child standing on the surface of the earth. If the child has a ball of mass $m$ and, at time $t = 0$, throws the ball from an initial height $y_0$ with an initial velocity $v_0$, describe the trajectory of the ball as a function of time.

*Solution.* Here we exploit Newton's Second Law, which say that the net force $F_{\text{net}}$ exerted on a body is equal to the object's inertial mass times its acceleration; that is, $F_{\text{net}} = ma$. In our case, we shall ignore air-friction and assume that the force of gravity is a constant, denoted by $g$. A quick look at Figure 31 shows us that the net force acting on the ball is $F = -mg$, where the negative sign is included to indicate that the force of gravity is acting in the opposite direction of initial motion. Newton's second law then implies that

$$F = ma = -mg.$$

Since the non-zero mass is present on either side of the equation, we may cancel it to deduce that $a = -g$. But we mentioned earlier that if $y$ is the ball's vertical displacement, then its acceleration

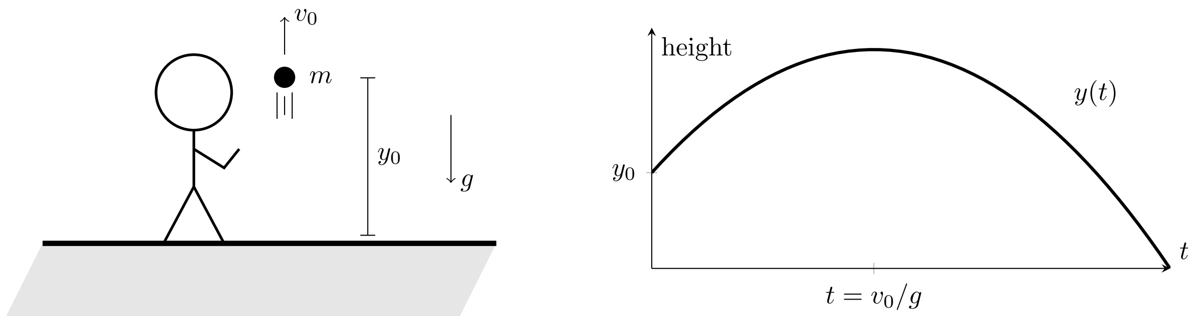



Figure 31: **Left:** A child throws a ball of mass $m$ into the air from an initial height of $y_0$. It is pulled back to Earth via the force of gravity. **Right:** A plot of the height of the ball as a function of time.

$a$ satisfies $a = \frac{\mathrm{d}^2 y}{\mathrm{d}t^2}$, yielding the ordinary differential equation

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = -g.$$

We do not know how to solve such differential questions in general; however, this is not too difficult to handle since $g$ is a constant. In particular, we ask ourselves "What kind of function may we differentiate twice to get a constant?" The immediate answer is that quadratic functions will suffice. To see this, recall that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left(at^2 + bt + c\right) = 2a \tag{4.1}$$

which is a constant. We thus make the *ansatz* (educated guess) that we may describe the equation of motion as $y(t) = at^2 + bt + c$, for which we then need to determine the coefficients $a, b, c \in \mathbb{R}$. Equation (4.1) implies that $y''(t) = 2a = -g$ so that $a = -g/2$. The other coefficients are determined by the two pieces of information that we have yet to use: the initial height and initial velocity. The initial velocity, at time $t = 0$, is $v_0$ which means that

$$v_0 = \frac{\mathrm{d}y}{\mathrm{d}t}(0) = \left[2at + b\right]_{t=0} = b.$$

The initial height is

$$y_0 = y(0) = \left[at^2 + bt + c\right]_{t=0} = c.$$

Putting this all together we conclude that the equation of motion is given by

$$y(t) = -\frac{g}{2}t^2 + v_0 t + y_0. \tag{4.2}$$

The leading term of $t^2$ should be negative since the ball is being projected upwards and then is returning to the ground.

The derivative yields $y'(t) = -gt + v_0$ which has a zero at $t = v_0/g$. This is clearly a maximum since the second derivative test is $y''(t) = -g < 0$. The height at this time is given by

$$y\left(\frac{v_0}{g}\right) = -\frac{g}{2}\left(\frac{v_0}{g}\right)^2 + v_0\left(\frac{v_0}{g}\right) + y_0 = \frac{v_0^2}{2g} + y_0 \tag{4.3}$$

■

A good way of checking the feasibility of your answers in physics is to ensure that things are dimensionally correct. For example, the dimension of $y_0$, denoted $[y_0]$, is length. We write this as $[y_0] = L$. Since $v_0$ is velocity its units are length per time, written as $[v_0] = LT^{-1}$. Acceleration is length per time per time, so $[g] = LT^{-2}$. Finally, time is just $[t] = T$. We can multiply any collection of dimensions, but we may only add dimensions when they are the same. Looking at equation (4.2), the dimensions of each component are

$$\left[\frac{g}{2}t^2\right] = [g][t]^2 = (LT^{-2})T^2 = L, \quad [v_0 t] = [v_0][t] = (LT^{-1})T = L, \quad [y_0] = L$$

so that each component has the correct dimensions. Similarly, in (4.3) we have

$$\left[\frac{v_0^2}{g}\right] = \frac{L^2T^{-2}}{LT^{-2}} = L$$

so again our solution makes sense. All of this implies that $\left[\frac{\mathrm{d}}{\mathrm{d}t}\right] = T^{-1}$. In general, differentiating by a quantity with units $U$ results in an object whose units are affected by $U^{-1}$.

### 4.1.2 The Simple Harmonic Oscillator

The simple harmonic oscillator is easily one of the most important physical systems that can be described. It appears constantly throughout all of physics and much of mathematics, and so warrants some of our attention.

**Example 4.3**

Consider an object of mass $m$ attached to a spring with spring constant $k$. If the object is initially a distance of $x_0$ from its equilibrium and is given an initial velocity of $v_0$ at time $t = 0$, what is the equation of motion of the spring?

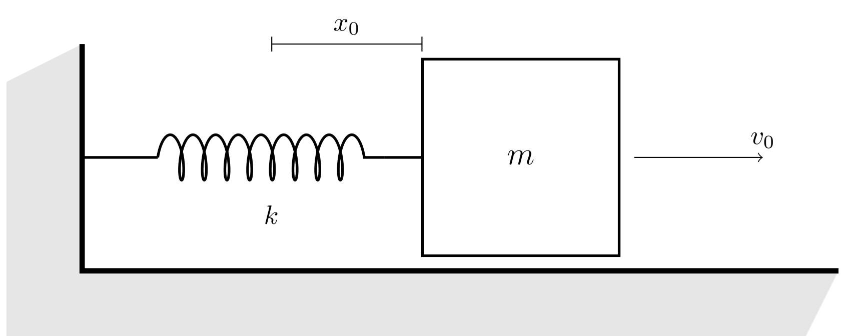



Figure 32: A spring-mass system, wherein the spring coefficient is given by the quantity $k$.

*Solution.* Let the equilibrium point of the spring be given by $x = 0$, in which case Hooke's law tells us that the net force experienced by the object is given by $F = -kx$. Again a simple application of Newton's second law tells us that

$$F = ma = -kx. \tag{4.4}$$

Writing $a = \frac{d^2x}{dt^2}$, dividing by $m$, and defining a new variable $\omega = \sqrt{k/m}$, equation (4.4) is equivalent to

$$\frac{d^2x}{dt^2} + \omega^2 x = 0. \tag{4.5}$$

Again, a proper treatment of this question warrants the use of some ODE theory, but we can formulate another ansatz as to the solutions of this equation. Essentially, this equation is asking us "What function, when you differentiate it twice, gives you back the negative of the function?" The only such functions, of which we are currently aware, are the sine and cosine functions. In fact, the student should convince themselves that equation (4.5) is satisfied by

$$x(t) = A\sin(\omega t) + B\cos(\omega t)$$

for as-of-yet determined coefficients $A, B \in \mathbb{R}$. Just as we did before, we may calculate $A, B$ using the unused information given to us in the problem. The initial displacement is $y_0$ implying that

$$y_0 = x(0) = A\sin(0) + B\cos(0) = B$$

and the initial velocity is

$$v_0 = x'(0) = A\omega\cos(0) - B\omega\sin(0) = A\omega.$$

Putting this all together, the equation of motion for the spring is

$$x(t) = \frac{v_0}{\omega}\sin(\omega t) + y_0\cos(\omega t). \qquad \blacksquare$$

**Remark 4.4** The $\omega$ introduced above is known as the *resonance frequency* and is the natural frequency of the system. Many systems have natural frequencies and much technology has been developed by exploiting this frequency (example: Magnetic *Resonance* Imaging). On the other hand, resonance is also responsible for some catastrophes. Look for videos of the original Tacoma Narrows Bridge. When engineers designed this bridge, they failed to consider what the bridge's natural frequency would be. There would be some days wherein high-winds would match this frequency causing the bridge to oscillate. The bridge eventually collapsed.

### 4.1.3 The Pendulum

The previous two problems admitted explicit solutions; however, the reality is that while we can often formulate differential equations to describe a system, it is very rare that these equations can be properly solved. Nonetheless, there is an entire field of mathematical study dedicated to solving differential equations which do not have closed form solutions. The moral of the story: we can use calculus to describe the motion of a great deal many things.

Here is an example which does not have a closed form solution, but whose equation of motion is simple to derive.

**Example 4.5**

Consider a bob of mass $m$ hanging from a rigid rod of length $L$. If the rod is not in an equilibrium state, describe its equation of motion.

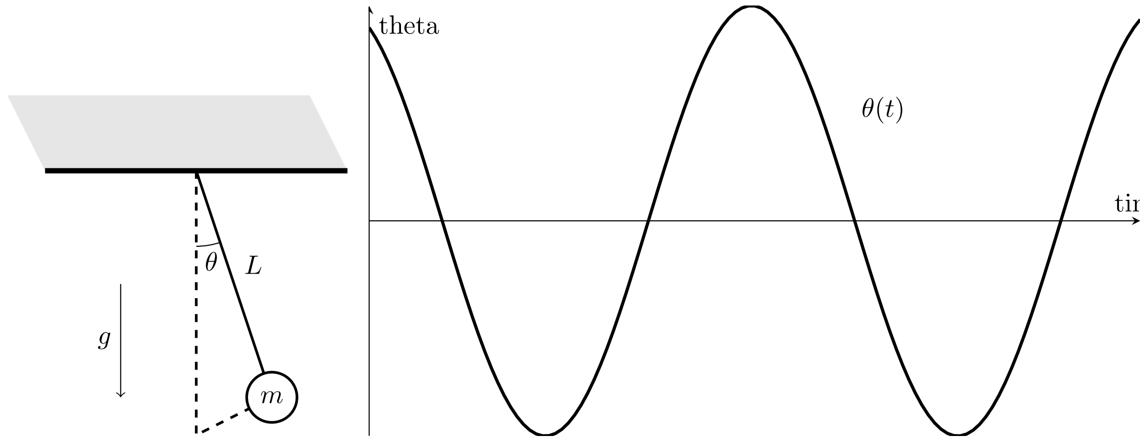



Figure 33: A pendulum system.

*Solution.* This is best done by considering Figure 33. The net force acting on the bob will correspond to $F = mg\sin\theta$ and so Newton's second law tells us that $ma = mg\sin\theta$. If we try our similar trick of writing $a = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}$ we will have an equation in both the variable $x$ and $\theta$, which is not what we desire. Instead, we must find a way to relate $x$ and $\theta$. This is not too hard once we realize that $x$ describes the arclength swept out by the rod, meaning that $x = L\theta$ (this is just the arclength formula). Twice differentiating tells us that $\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = L\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}$ so substituting we get

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} - \frac{g}{L}\sin\theta = 0.$$

It may not be obvious at this point that this equation cannot by solved analytically, but let me assure you that it cannot. However, there is one extenuating circumstance in which we may find *approximate* solutions. If $\theta$ is a very small angle then $\sin\theta \approx \theta$ in which case this becomes

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} - \frac{g}{L}\theta = 0$$

which we recognize as the equation for the simple harmonic oscillator. ■

#### 4.1.4 Exponential Growth

Let $P(t)$ be an object which grows in proportion to its size. For example, consider a species of bacteria in which each bacterium splits and doubles after a period of 5-minutes. A colony of 100 bacteria will grow to 200 after 5-minutes, resulting in a growth of 100 bacteria, while a colony of 1000 bacteria will double to 2000 bacteria in 5 minutes, resulting in a growth of 1000 new bacteria! It should be clear then that the more bacteria present in the colony, the faster the colony will grow.

If we think about this example in more detail, let us say we start with a colony of a single bacteria. The colony size (specified over 5-minute intervals) will look like

$$1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536, \ldots$$

In general, this growth appears to be exponential, corresponding to a discrete function $2^x$.

Modelling exponential growth is often simplified by assuming continuous rather than discrete growth, and it turns out that this is actually one of the motivating example for defining Euler's number $e$. Perhaps the easiest way to see this is in the language of investment. Assume that you are given an initial investment $I_0$ which grows with an interest rate $r$. If we compound the interest annually (once per year) then after one year we have $I_0(1+r)$ dollars. If we compound the interest semi-annually (twice per year) then we take half the interest rate $r/2$ and compound twice to get

$$I_0\left(1+\frac{r}{2}\right)\left(1+\frac{r}{2}\right) = I_0\left(1+\frac{r}{2}\right)^2.$$

Similarly, if we compounded the interest $n$ times in a year, we would be left with the equation $I_0\left(1+\frac{r}{n}\right)^n$. The idea of continuously compounding interest will then occur as we let $n \to \infty$. Namely, the amount of money earned after one year will be

$$I_0 \lim_{n\to\infty}\left(1+\frac{r}{n}\right)^n.$$

Now what happens if we make the change of variable $x = \frac{1}{n}$ so that as $n \to \infty$ we have $x \to 0$. Our equation then becomes

$$I_0 \lim_{n\to\infty}\left(1+\frac{r}{n}\right)^n = I_0 \lim_{x\to 0}(1+rx)^{1/x} = I_0 e^r;$$

that is, the number $e$ represents the proportion by which your money will increase if compounded continuously over the course of a year. If we had started with \$1 then after one year of continuous interest at 1% we would finish the year with \$$e$. This is the sense to which $e$ is *natural*, it represents continuous growth.

Now if we return to our bacteria example, let us assume that our bacteria population grows at a continuous rate which is proportional to its population. That is, there is some constant of proportionality $k$ such that $\frac{\mathrm{d}P}{\mathrm{d}t} = kP$. What kind of function satisfies this differential equation? A bit of inspection actually reveals that if $I_0$ is the initial population, $P(t) = I_0 e^{kt}$ satisfies the equation, since

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} I_0 e^{kt} = I_0 k e^{kt} = kP(t).$$

Hence any quantity which grows (or shrinks) continuously and proportional to its population size is modelled with an exponential function.

**Example 4.6**

If an original investment of \$100 is invested at a rate of 6% and compounded continuously, how long will it take for the investment to triple in size?

*Solution.* Our model is given by $I(t) = 100e^{0.06t}$ and we would like to find the $t$ such that $I(t) = 300$ (since 300 is triple the number 100). Indeed, we may solve to find that $300 = 100e^{0.6t}$ implies

$$3 = e^{0.06t} \qquad \Rightarrow \qquad 0.06t = \ln(3) \qquad \Rightarrow \qquad t = \frac{\ln(3)}{0.06} \approx 18.3 \text{ years.}$$ ■

**Example 4.7**

Assume that a culture of bacteria has an initial population of 100 bacteria, which becomes 350 bacteria after 10-hours. Determine the population of the bacteria after 2-days.

*Solution.* We know that our growth curve is modelled by the formula $P(t) = 100e^{kt}$, for $t$ measured in hours. We are also told that

$$P(10) = 100e^{10k} = 350,$$

which will allow us to solve for the growth rate $k$. Indeed,

$$100e^{10k} = 350 \quad \Rightarrow \quad e^{10k} = 3.5 \quad \Rightarrow \quad 10k = \ln(3.5) \quad \Rightarrow \quad k = \frac{\ln(3.5)}{10}.$$

To be consistent with our choice of units, we note that two days is 48-hours, so the population after two days is

$$P(48) = 100e^{48k} = 100e^{48\frac{\ln(3.5)}{10}} = 100e^{\ln(3.5^{4.8})} = 100 \cdot 3.5^{4.8} \approx 40881 \text{ bacteria.} \qquad \blacksquare$$

**Example 4.8**

Caffeine in the blood stream has a biological half-life of 5-hours. If I drink a venti café Americano from Starbucks, consisting of 300mg of caffeine, at 8am in the morning. How much caffeine will be in my system by midnight?

*Solution.* Let us that $t = 0$ corresponds to 8am, so that midnight is $t = 16$. The amount of caffeine is modelled by the equation $c(t) = 300e^{kt}$ and we know that $c(5) = 300e^{5k} = 150$ (since 150 is half of 300). We may solve for $k$ to find that $k = -\frac{\ln(2)}{30}$. Hence at time $t = 16$ we have

$$c(16) = 300e^{-16\frac{\ln(2)}{5}} = 300 \cdot 2^{-16/5} \approx 32.64 \text{ mg.} \qquad \blacksquare$$

One can show that if the half life of a substance is a time $t_0$, then the growth/decay rate $k$ is always given by $k = -\frac{\ln(2)}{t_0}$.

**Logistic Growth:** While this model of exponential growth can be quite useful for modelling short term growth, it quickly becomes unrealistic. Exponential functions grow incredibly quickly: in fact, if a species of bacteria grow at the same rate as in Example 4.7, then a single bacterium would grow into a colony with the same mass as the entire Earth in about one month!

Our exponential model breaks down in the long term as it fails to incorporate things like competition with other bacteria, or the limited number of resources available to the bacteria. If we know *a priori* that the bacteria only have enough resources to sustain a certain size colony, we can adapt our model to take that into consideration. The maximum number $M$ of the species is called the *carrying capacity*, and changes our model to be

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{M}\right). \tag{4.6}$$

Saying that there is no maximum species is akin to setting "$M = \infty$," in which case (4.6) just becomes the usual $\frac{dP}{dt} = kP$. On the other hand, in the limit as $P \to M^-$ we get

$$\lim_{P \to M^-} \frac{dP}{dt} = \lim_{P \to M^-} \left[ kP \left( 1 - \frac{P}{M} \right) \right] = 0,$$

showing that our growth slows as our population approaches carrying capacity. The solution to (4.6) is not easy to guess, but is given by

$$P(t) = \frac{P_0 M e^{kt}}{P_0(e^{kt} - 1) + M}. \tag{4.7}$$

The asymptotic nature of this function can be seen in Figure 34.

> **Exercise:** Show that (4.7) satisfies (4.6).

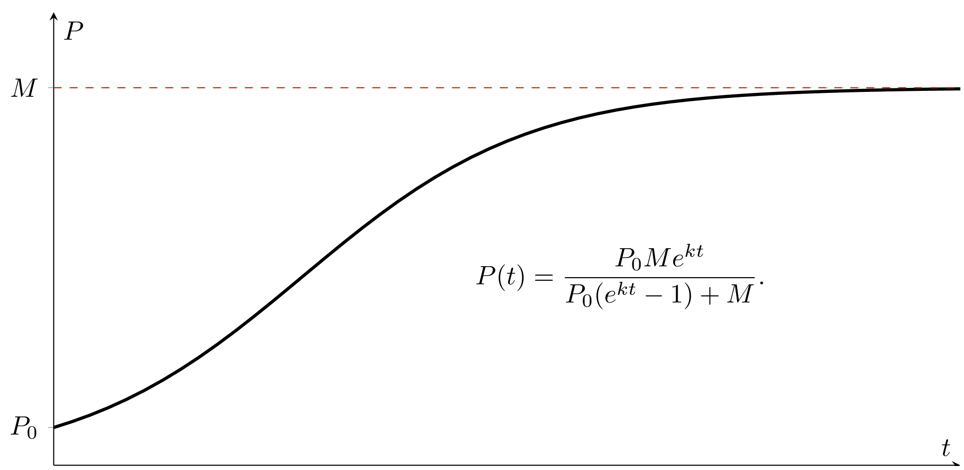



Figure 34: Logistic growth incorporates a carrying capacity $M$ so that the population cannot run off to infinity.

**Example 4.9**

As a biologist for a famous national park, your job is to ensure that the park maintains a stable ecology. In particular, you are tracking a group of red squirrels, which you know reproduce in an exponential fashion with growth coefficient $k = 0.32$, over a period of one year. The park has a theoretical capacity of 15000 squirrels, but you know that if the population reaches 80% of this number, it will begin to destabilize the surrounding populations. If the current population sits at 1300 squirrels, how long do you have before the squirrel population begins to pose a problem?

*Solution.* We are given a growth rate of $k = 0.32$, an initial population of $P_0 = 1300$, and a carrying capacity of $M = 15000$. The squirrels will become troublesome when the reach 80% of our carrying capacity, which is 12000 squirrels. Solving

$$12000 = \frac{1300 \times 15000 e^{0.32t}}{1300(e^{0.32t} - 1) + 15000}$$

we get $t = 11.7$ years. ■

## 4.2 Implicit Differentiation

### 4.2.1 The Idea of Implicit Functions

The idea of implicit differentiation is that we may be given variables in which it is *implied* that those variables depend on other variables, though we may not be able to explicitly write that relationship down. For example, to this point we have typically seen examples where we might write $y = f(x)$, in which case it is clear that that changes in $x$ affect changes in $y$, as prescribed by the function $f(x)$. We were then able to determine the rate of change of $y$ with respect to $x$ by computing $\frac{dy}{dx}$.

However, we can sometimes write relationships down without being able to solve for one variable as a function of the other; for example

$$\begin{aligned} x^2 + y^2 &= 25 \\ e^x + x\cos(y) + y &= 1 \\ f(x,y) &= k \quad \text{for some constant } k \end{aligned}$$

We can convince ourselves that the variables $x$ and $y$ above depend on one another. For example, consider the equation of the circle $x^2 + y^2 = 25$. If we set $y = 5$ then $x$ is forced to be 0, while if we were to set $x = 3$ then $y$ would have to be one of $y = \pm 4$. However, there is no function which makes the relationship between $x$ and $y$ explicit, since as our above example indicates, a single $x$-value may correspond to two possible $y$-values, and hence the relationship is not one given by a function.[7]

As an alternative example, consider the volume $V$ of a cylinder as a function of its radius $r$ and height $h$:

$$V = \pi r^2 h.$$

This equation defines an explicit relationship between the three entities $V, r,$ and $h$. However, what if our cylinder were made of metal, and we were told that as temperature $T$ increases, both the radius and the height increase, and conversely when temperature decreases, the radius and the height decrease[8]? In that case, we are *implicitly* assuming that $r$ and $h$ are functions of temperature $T$. This means that $V$ is also implicitly a function of temperature, and we have

$$V(T) = \pi r(T)^2 h(T). \tag{4.8}$$

This implicit understanding that all the variables now depend on temperature allows us to determine how the volume of our cylinder is changing with temperature.

[7] Some students might be distressed at the fact that I have written $x^2 + y^2 = 25$ as an implicit equation, since certainly we could solve to find

$$y = \sqrt{25 - x^2},$$

but I claim that this actually not an explicit representation of this function. The reason is that, for example, both $(0, 5)$ and $(0, -5)$ are solutions to this equation, but we are unable to recover $(0, -5)$ from the expression $y = \sqrt{25 - x^2}$.

[8] Metals typically expand with increased heat

In both of the aforementioned cases, it seems as though we should still be able to discuss the rate of change of one variable with respect to another, even if we are unable to explicitly describe the relationship between the variables using a function. This leads us to a process known as implicit differentiation.

### 4.2.2 How Implicit Differentiation Works

Let's say that we know a variable $y$ implicitly depends upon another variable $x$. The idea of implicit differentiation is to differentiate as though the exact nature of the relationship were known. The best way to understand this is to see an example.

**Example 4.10**

Consider the equation of the circle centered at the origin with radius 1, $x^2+y^2=1$. Determine the rate of change of $y$ with respect to $x$.

*Solution.* Our goal is to compute $\frac{\mathrm{d}y}{\mathrm{d}x}$. We saw earlier that the equation of a circle is an implicit relationship as there is no function which describes how $y$ changes with respect to $x$ or vice versa. Nonetheless, we are going to differentiate the equation $x^2+y^2=1$, but we keep in mind always that we are assuming that $y$ is a function of $x$. To make this more clear, let's actually write $y=f(x)$, so that our equation of the circle is

$$1 = x^2 + y^2 = x^2 + f(x)^2.$$

Now differentiating both sides, we have

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}x}\left(x^2 + f(x)^2\right) \\ &= 2x + 2f(x)f'(x). \end{aligned}$$

Remember that we are trying to solve for $\frac{\mathrm{d}y}{\mathrm{d}x}$, which under our choice of $y=f(x)$ is just $\frac{\mathrm{d}y}{\mathrm{d}x}=f'(x)$, hence

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f'(x) = -\frac{x}{y}. \tag{4.9}$$

**Remark 4.11** Many people do not like to use this $y=f(x)$ notation, and instead will just write down

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^2+y^2\right) = 2x + 2y\frac{\mathrm{d}y}{\mathrm{d}x}.$$

This is acceptable and if you are comfortable using it, then you should feel free to do that. However, writing this down often hides what is really happening, so we have used the $y=f(x)$ notation just to be clear.

On the other hand, let me write $\tilde{y}=\sqrt{1-x^2}$, where the fact that I have used the tilde is to indicate that $\tilde{y}$ is not actually the same thing as $y$. When we differentiate we get

$$\frac{\mathrm{d}\tilde{y}}{\mathrm{d}x} = \frac{-x}{\sqrt{1-x^2}} = -\frac{x}{\tilde{y}}. \tag{4.10}$$

The inquisitive student may realize looks very similar to (4.9), but I claim is not quite the same.

Indeed, let us try to find the slope of the tangent line to the circle at the point $x = \frac{1}{\sqrt{2}}$. Notice on the circle that there are two possible $y$ values, corresponding to $y_+ = +\frac{1}{\sqrt{2}}$ and $y_- = -\frac{1}{\sqrt{2}}$. Using (4.9) we find that the slope of the tangent lines at $y_\pm$ are

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)} = -\frac{1/\sqrt{2}}{1/\sqrt{2}} = -1, \qquad \left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right)} = -\frac{1/\sqrt{2}}{-1/\sqrt{2}} = 1$$

which is in fact what we would expect. On the other hand, using (4.10) we find that at $x = \frac{1}{\sqrt{2}}$ there is only one possible $\tilde{y}$ value, corresponding to $\tilde{y} = \frac{1}{\sqrt{2}}$ in which case the slope of the tangent line is the same as that found above, namely

$$\left.\frac{\mathrm{d}\tilde{y}}{\mathrm{d}x}\right|_{\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)} = -\frac{1/\sqrt{2}}{1/\sqrt{2}} = -1.$$

We have in fact lost the other tangent line! This is because when we took the square root of $y^2 = 1 - x^2$ we needed to make a choice as to whether to take the positive or negative root. In doing so, we actually lost information. ∎

**Example 4.12**

Consider the volume of a cylinder as a function of temperature, as given in (4.8). Determine the rate of change of $V$ with respect to temperature, written in terms of how $r$ and $h$ vary with respect to temperature.

*Solution.* We already know that $V = \pi r^2 h$. Although (4.8) has the temperature dependence written in, we will ignore it for this exercise to show the student how the notation is typically conveyed. Differentiating, we get

$$\begin{aligned}\frac{\mathrm{d}V}{\mathrm{d}T} &= \frac{\mathrm{d}}{\mathrm{d}T}\left[\pi r^2 h\right] \\ &= \pi\left[2r\frac{\mathrm{d}r}{\mathrm{d}T}h + r^2\frac{\mathrm{d}h}{\mathrm{d}T}\right].\end{aligned}$$ ∎

The power of implicit differentiation can be even greater if one is given a transcendental equation such as $e^x + x\cos(y) + y = 1$, wherein it is impossible to isolate and solve for $y$. In such instances, one is indeed *forced* to use implicit differentiation.

**Example 4.13**

Compute the derivative $\frac{\mathrm{d}y}{\mathrm{d}x}$ of $y$ in the equation $e^x + x\cos(y) + y = 1$.

*Solution.* Keeping in mind that $y$ is a function of $x$, we apply $\frac{\mathrm{d}}{\mathrm{d}x}$ to both sides of our equation to find:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(e^x + x\cos(y) + y\right) &= \frac{\mathrm{d}}{\mathrm{d}x}1 \\ e^x + \cos(y) - x\sin(y)\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\mathrm{d}y}{\mathrm{d}x} &= 0 \\ \frac{e^x + \cos(y)}{x\sin(y) - 1} &= \frac{\mathrm{d}y}{\mathrm{d}x}.\end{aligned}$$

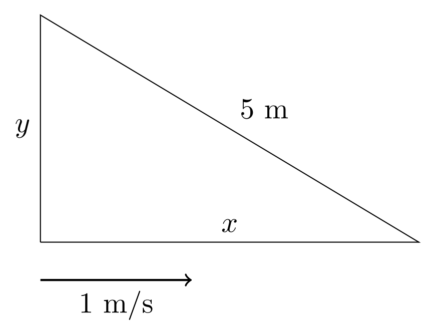
■

## 4.3 Related Rates

Much as we saw in our example relating the volume of a cylinder to its radius and height, related rates are way of using the relationship between objects to infer how they change with respect to an implicit variable. The trick when doing related rates questions is to find an equation that involves the quantities you are interested in, assume everything varies with time, then differentiate implicitly. The best way to get the hang of related rates questions is to do examples, so let's get started.

**Example 4.14**

Assume that a custodian is standing on a latter 5 metres tall and a pesky student pulls that ladder out from the wall at a rate of 1 metre/second. Determine how quickly the custodian is falling when he is 3 metres from the ground.




*Solution.* If one begins by drawing a simple diagram of the situation, we see that we get a triangle whose side lengths are $(x, y, 5)$, where $x$ is the distance of the ladder from the wall, $y$ is the height of the ladder above the ground, and 5 is length of the ladder (the hypotenuse). The obvious relationship between the variables is to use the Pythagorean theorem, $x^2 + y^2 = 25$. Since $x$ and $y$ are both changing as a function of time, we differentiate both implicitly to find

$$2x\frac{\mathrm{d}x}{\mathrm{d}t} + 2y\frac{\mathrm{d}y}{\mathrm{d}t} = 0.$$

Since we want to know how fast the custodian is falling we would like to evaluate $\frac{\mathrm{d}y}{\mathrm{d}t}$, which is solved as $\frac{\mathrm{d}y}{\mathrm{d}t} = -\frac{x}{y}\frac{\mathrm{d}x}{\mathrm{d}t}$. We already know that $\frac{\mathrm{d}x}{\mathrm{d}t} = 1$ (as this is the rate at which the latter is being pulled from the wall), and we are told to evaluate when $y = 3$. Hence we need only find the $x$-value which corresponds to a $y$-value of 3. The Pythagorean theorem again implies that if $y = 3$ then $x = \sqrt{5^2 - 3^2} = 4$, so that

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\frac{4}{3}(1) = -\frac{4}{3}.$$

Thus the custodian is falling at 4/3 metres per second. ■

**Example 4.15**

A sphere is shrinking in such a way that its surface area changes at a constant rate of 1 unit$^2$ per time. Find the rate at which the diameter of the sphere is decreasing when the diameter is 5 units.

*Solution.* First, we look at the quantities involved. We are given information about the surface area $A$ of the sphere, and its diameter $\delta$. We know that the formula for the surface area of a sphere is given by $A = 4\pi r^2$ where $r$ is the radius of the sphere. This is not quite what we want though, since we are interested in the diameter. Luckily, we also know that the radius is half the diameter, so that $r = \delta/2$. Substituting this we find that

$$A = 4\pi \left(\frac{\delta}{2}\right)^2 = \pi\delta^2.$$

Both the area and the diameter are changing with time, so when we differentiate we will have to do so implicitly with respect to time:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = 2\pi\delta\frac{\mathrm{d}\delta}{\mathrm{d}t}.$$

We want to know $\frac{\mathrm{d}\delta}{\mathrm{d}t}$ as this represents the change in the diameter with respect to time, so we solve to find that

$$\frac{\mathrm{d}\delta}{\mathrm{d}t} = \frac{1}{2\pi\delta}\frac{\mathrm{d}A}{\mathrm{d}t}.$$

Since $A$ is decreasing at a rate of 1 unit$^2$ per time, we set $\frac{\mathrm{d}A}{\mathrm{d}t} = -1$, where the negative sign indicates that the quantity is decreasing. Thus we know every quantity on the right hand side, so we may substitute to find that

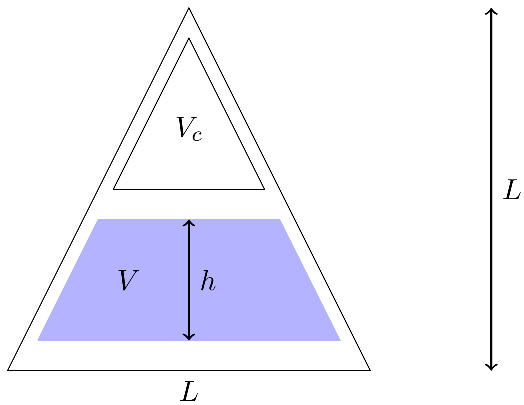
$$\frac{\mathrm{d}\delta}{\mathrm{d}t} = \frac{1}{2\pi \cdot 5}(-1) = -\frac{1}{10\pi}.$$ ■

**Example 4.16**

Consider a square pyramid which is being filled with water. The base square has length and width both equal to $L > 0$ and the height of the apex of the pyramid is also a distance $L$ from the base. As a function of the current height of the liquid, how must we poor the mysterious quantity into the pyramid such that the rate of change of the height of the material does not depend on the size of the base square. Is this rate dependent on the height of the pyramid?

*Solution.* Let $V$ be the volume of the liquid contained in the pyramid, which we know is an increasing function in time. We have the capability of adjusting the flow rate, which corresponds to the quantity $\frac{\mathrm{d}V}{\mathrm{d}t}$. We are interested in the determining how the height of the water $h$ is changing as a function of time, so that we might choose an appropriate $\frac{\mathrm{d}V}{\mathrm{d}t}$ to ensure that $\frac{\mathrm{d}h}{\mathrm{d}t}$ has no dependence upon $L$.

If we project the pyramid into two dimensions, we have the following figure:




Where $V_c$ is the volume not occupied by water. Notice that the height of the pyramid bounding $V_c$ is $L-h$, so if its base has dimension $\ell$, its volume is $V_c = \frac{1}{3}\ell^2(L-h)$. In order to relate this to the volume of the water, notice that the total volume of the pyramid is

$$V_{\mathrm{tot}} = \frac{1}{3}Ah = \frac{1}{3}L^3 = V + V_c \tag{4.11}$$

so we can write $V = \frac{1}{3}L^3 - V_c$. As a side comment, note that (4.11) implies that $\frac{\mathrm{d}V}{\mathrm{d}t} = -\frac{\mathrm{d}V_c}{\mathrm{d}t}$, which makes sense. Now the triangle which bounds the projected pyramid is similar to the triangle which bounds $V_c$, and this implies that the ratio of their side lengths will be equal. In particular, if we let $\ell$ be the length of the base of the triangle bounding $V_c$, then

$$\frac{\ell}{L-h} = \frac{L}{L} = 1$$

so that $\ell = L - h$. Hence we have

$$V = \frac{1}{3}L^3 - V_c = \frac{1}{3}L^3 - \frac{1}{3}(L-h)^3.$$

Differentiating with respect to time $t$ gives

$$\frac{\mathrm{d}V}{\mathrm{d}t} = (L-h)^2\frac{\mathrm{d}h}{\mathrm{d}t}, \qquad \text{or equivalently} \qquad \frac{\mathrm{d}h}{\mathrm{d}t} = -\frac{1}{(L-h)^2}\frac{\mathrm{d}V}{\mathrm{d}t}.$$

Thus by setting $\frac{\mathrm{d}V}{\mathrm{d}t} = C(L-h)^2$ for some constant $C$, we can guarantee that $\frac{\mathrm{d}h}{\mathrm{d}t}$ does not depend on $L$, and changes at a constant rate $C$. ■

**Example 4.17**

A cat is sitting on a ledge a distance $d$ from a door when suddenly a frantic math professor bursts through, chased by a mob of angry students. The cat, having watched many Olympic races, figures that the teacher is running at a rate of $s$ metres per second. Assuming that the angry mob of students is unable to catch the spry professor, at what rate must the cat's head move in order to watch the chase once the professor has run $r$ metres. Your answer should be expressed purely in terms of $r, s$ and $d$.

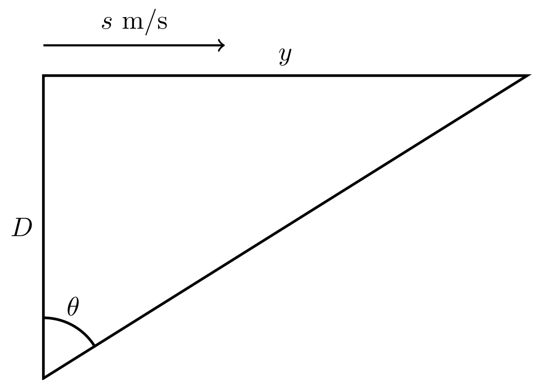



Figure 35: The professor running, while the cat patiently observes.

*Solution.* We begin by drawing an appropriate diagram, labeling the angle the cat's head makes with the door $\theta$ and the distance from the professor to the door by $y$. Notice we will be interested in the case when $y = r$, which occurs when the professor has run a distance $r$, and when $\frac{\mathrm{d}y}{\mathrm{d}t} = s$, as this is the speed of the professor. This is illustrated in Figure 35.

The given triangle implies that $\tan(\theta) = \frac{y}{d}$, so implicit differentiation yields

$$\sec^2(\theta)\theta' = \frac{y'}{d}.$$

All units are consistent and the professor is running at $s$ metres per second, so $y' = s$ giving

$$\theta' = \frac{s}{d\sec^2(\theta)}.$$

We need only remove the dependency on $\theta$ to be finished. The triangle illustrated in Figure 35 tells us that the hypotenuse is given by $\sqrt{r^2+d^2}$ so that $\cos(\theta) = d/\sqrt{r^2+d^2}$. Since $\sec^2(\theta) = 1/\cos^2(\theta) = (r^2+d^2)/d^2$ we conclude that the cat's head is moving at a rate of

$$\theta' = \frac{sd}{r^2+d^2} \quad \text{radians per second.}$$

■

## 4.4 L'Hôpital's Rule

Just when we thought we had left the world of limits behind, we find ourselves inexplicably talking about them once again. The reason for this late treatment of the subject is due to the fact that we had initially sheltered the student from the cruel world of limits, choosing to show you only the nicest limits which could be computed via some algebraic magic. Now, with a much more developed toolbox, we venture into the treatment of these remaining limits.

### 4.4.1 Standard Indeterminate Type

We have seen that in the case of many functions formed from quotients, if both the numerator and denominator independently gave finite limits and the denominator was not zero, then we could simply apply the limit law for quotients. Similarly, if only one of the numerator or denominator gave a limit of 0 then one could often convince him/herself that the limit was either 0 or infinity, respectively. The troublesome point comes when both the numerator and the denominator both approach zero. How then can we treat such problems in general? We begin with a classification of the types of problems which can occur:

**Definition 4.18**

Let $f(x)$ and $g(x)$ be two functions,

1. If $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ then we say that $\lim_{x\to c} \dfrac{f(x)}{g(x)}$ is *indeterminate of type* $0/0$.
2. If $\lim_{x\to c} |f(x)| = \lim_{x\to c} |g(x)| = \infty$ then we say that $\lim_{x\to c} \dfrac{f(x)}{g(x)}$ is *indeterminate of type* $\infty/\infty$.

Note: If we take the limit as $x \to \infty$ or $x \to -\infty$, or either of the one sided limits $x \to c^{\pm}$ we will still say that the limits are of the above indeterminate types.

Indeterminate forms are precisely the aforementioned pesky functions. Here is how we attack them.

**Theorem 4.19: L'Hôpital's Rule**

Suppose that $f(x)$ and $g(x)$ are differentiable in an interval containing $c$, and that the limit $\lim_{x\to c} \dfrac{f(x)}{g(x)}$ is indeterminate of either type $0/0$ or type $\infty/\infty$. If one of the following conditions holds:

1. The limit $\lim_{x\to c} \dfrac{f'(x)}{g'(x)}$ exists,
2. There is an interval containing $c$ on which $g'$ is never zero,
3. There is an interval containing $c$ in which $g'$ does not switch signs infinitely often,

then $\lim_{x\to c} \dfrac{f(x)}{g(x)}$ exists and $\lim_{x\to c} \dfrac{f(x)}{g(x)} = \lim_{x\to c} \dfrac{f'(x)}{g'(x)}$.

**Example 4.20**

Compute the limit $\lim_{x\to 0} \frac{e^{2x}-1}{\sin(x)}$.

*Solution.* A quick check shows that this limit is of indeterminate type 0/0, since as $x \to 0$ we have both $e^2 - 1 \to 0$ and $\sin(x) \to 0$. Hence we may apply L'Hôpital's rule to find that

$$\lim_{x\to 0} \frac{e^{2x}-1}{\sin(x)} = \lim_{x\to 0} \frac{\frac{\mathrm{d}}{\mathrm{d}x}\left[e^{2x}-1\right]}{\frac{\mathrm{d}}{\mathrm{d}x}\sin(x)} = \lim_{x\to 0} \frac{2e^{2x}}{\cos(x)} = 2. \qquad \blacksquare$$

**Remark 4.21**

1. Note that we take the derivative of both the top and the bottom separately. Many students accidentally take the derivative of the quantity $f(x)/g(x)$, which by the quotient rule is *very* different than $f'(x)/g'(x)$.

2. It is absolutely essential that the student first check that the limit is of indeterminate type. **L'Hôpital's Rule does not hold if the limit is not of indeterminate type.** Indeed, we recall that

$$\lim_{x\to 0} \frac{\sin(x)}{\cos(x)} = \frac{0}{1} = 0,$$

so this limit is not of indeterminate type and can be evaluated quite simply. On the other hand, if we had tried to apply L'Hôpital's Rule we would have gotten

$$\lim_{x\to 0} \frac{\sin(x)}{\cos(x)} = \lim_{x\to 0} \frac{\cos(x)}{-\sin(x)} = -\infty$$

which is clearly not equal to 0. We thus re-emphasize that the student MUST check that the limit is indeterminate.

There may be instances in which a single application of L'Hôpital's Rule is not sufficient and the student will be required to apply the rule multiple times, as the following example demonstrates.

**Example 4.22**

Compute the limit $\lim_{x\to\infty} \frac{e^{x^2}}{x^4}$.

*Solution.* It is easy to convince ourselves that this is indeterminate of type $\infty/\infty$, so that we may apply L'Hôpital's Rule. Indeed, we find that

$$\begin{aligned}
\lim_{x\to\infty} \frac{e^{x^2}}{x^4} &= \lim_{x\to\infty} \frac{2xe^{x^2}}{4x^3} \\
&= \frac{1}{2}\lim_{x\to\infty} \frac{e^{x^2}}{x^2} && \text{again type } \infty/\infty \\
&= \frac{1}{2}\lim_{x\to\infty} \frac{2xe^{x^2}}{2x} \\
&= \frac{1}{2}\lim_{x\to\infty} e^{x^2} = \infty.
\end{aligned}$$

■

In the previous example, we were required to apply L'Hôpital's Rule twice. It is also important to note that the solution does not have to be finite: We can still have L'Hôpital give us infinite solutions.

**Exercise:** We say that a function $f(x)$ "grows" faster than $g(x)$ if

$$\lim_{x\to 0} \frac{g(x)}{f(x)} = 0, \qquad \text{or equivalently} \qquad \lim_{x\to 0} \frac{f(x)}{g(x)} = \infty.$$

Using L'Hôpital's Rule, show that $e^x$ grows faster than *any* polynomial; that is, show that $e^x$ is faster than $x^n$ for any $n$.

**Example 4.23**

Compute the limit $\lim_{x\to 0} \dfrac{e^x - 1 - x}{x^2}$.

*Solution.* This is again indeterminate of type 0/0, so we use L'Hôpital and compute

$$\begin{aligned}
\lim_{x\to 0} \frac{e^x - 1 - x}{x^2} &= \lim_{x\to 0} \frac{e^x - 1}{2x} && \text{indeterminate of type } 0/0 \\
&= \lim_{x\to 0} \frac{e^x}{2} = \frac{1}{2}.
\end{aligned}$$

■

**Example 4.24**

Determine the limit $\lim_{x\to 0} \dfrac{xe^{nx} - x}{1 - \cos(nx)}$ for $n \neq 0$.

*Solution.* It is easy to check that this is of type 0/0, so we can apply L'Hôpital to get

$$\begin{aligned}
\lim_{x\to 0} \frac{xe^{nx} - x}{1 - \cos(nx)} &\overset{\langle H\rangle}{=} \lim_{x\to 0} \frac{e^{nx} + nxe^{nx} - 1}{n\sin(nx)} \\
&\overset{\langle H\rangle}{=} \lim_{x\to 0} \frac{ne^{nx} + ne^{nx} + n^2xe^{nx}}{n^2\cos(nx)} \\
&= \frac{2n}{n^2} = \frac{2}{n}.
\end{aligned}$$

■

**Example 4.25**

If $f(x)$ is twice differentiable at $c$ and $f''(x)$ is continuous at $c$, compute the limits

$$\lim_{h\to 0} \frac{f(c+h)-f(c-h)}{2h}, \qquad \lim_{h\to 0} \frac{f(c+h)-2f(c)+f(c-h)}{h^2}. \tag{4.12}$$

*Solution.* We recognize that the limit in question is indeterminate of type 0/0 and so we have

$$\lim_{h\to 0} \frac{f(c+h)-f(c-h)}{2h} \overset{\langle H\rangle}{=} \lim_{h\to 0} \frac{f'(c+h)-(-1)f'(c-h)}{2} = f'(c) \tag{4.13}$$

wherein again we stress that we must *differentiate with respect to h*.

The second limit in (4.12) is a little bit harder to see without using L'Hôpital, so let's just go ahead and compute to find

$$\begin{aligned}
\lim_{h\to 0} \frac{f(c+h)-2f(c)+f(c-h)}{h^2} &\overset{\langle H\rangle}{=} \lim_{h\to 0} \frac{f'(c+h)-f'(c-h)}{2h} && \text{by L'Hôpital with type } 0/0 \\
&= (f')'(c) && \text{by (4.13)} \\
&= f''(c).
\end{aligned}$$

■

### 4.4.2 Other Indeterminate Types

Having formulated a strategy for computing limits of indeterminate quotients, we can adapt the strategy above to other types of indeterminate forms. In particular, the remaining indeterminate forms are of type $0\times\infty, \infty-\infty, 1^\infty, 0^0$ and $\infty^0$. The strategy for dealing with each of these cases is to turn it into an indeterminate of type 0/0 or $\infty/\infty$, as follows:

1. $\underline{0\times\infty}$: Possibly the easiest case to deal with, assume that we are given a limit of the form

$$\lim_{x\to a} f(x)g(x), \qquad \text{where } \lim_{x\to a} f(x) = 0, \qquad \text{and} \qquad \lim_{x\to a} g(x) = \infty.$$

   We can thus turn this into 0/0 or $\infty/\infty$ as

$$\lim_{x\to a} \frac{f(x)}{1/g(x)} \quad \text{type } 0/0, \qquad \lim_{x\to a} \frac{g(x)}{1/f(x)} \quad \text{type } \infty/\infty.$$

2. $\underline{\infty-\infty}$: In contrast to the previous cases, these are possibly the hardest to transform into one of 0/0 or $\infty/\infty$. This case arises when we have a limit of the form:

$$\lim_{x\to a} [f(x)-g(x)], \qquad \text{where } \lim_{x\to a} f(x) = \infty, \qquad \text{and} \qquad \lim_{x\to a} g(x) = \infty.$$

   There is no algorithmic way to do this, so the student is required to use their math-skills to manipulate such a limit into the desired form. Indeed, the student already has some practice with these types of limits, such as $\lim_{x\to a} \sqrt{x^2+1}-x$.

3. $1^\infty, 0^0, \infty^0$: These cases are all handled identically, so we shall treat them together. For the sake of concreteness, assume we are given a limit of the form

$$\lim_{x\to a} f(x)^g(x), \qquad \text{where } \lim_{x\to a} f(x) = \infty, \qquad \text{and} \qquad \lim_{x\to a} g(x) = 0$$

This case is handled by calling upon our old friend, implicit differentiation. Let $y = f(x)^{g(x)}$ so that $\ln(y) = g(x)\ln f(x)$. As we have seen previously, continuity of $e^x$ means that it is sufficient to compute

$$\lim_{x\to a} \ln y, \qquad \text{since} \qquad \lim_{x\to a} y = \lim_{x\to a} e^{\ln(y)} = e^{\lim\limits_{x\to a} \ln y}.$$

Now in the limit, $\ln(y) = g(x)\ln f(x)$ is of type $0\times\infty$, so we resort to case (1) which tells us that we compute this limit by either

$$\lim_{x\to a} \frac{g(x)}{1/\ln f(x)}, \qquad \text{or} \qquad \lim_{x\to a} \frac{\ln f(x)}{1/g(x)}.$$

In the case where $f(x)\to 1$ and $g(x)\to\infty$, or $f(x)\to 0$ and $g(x)\to 0$ we also get $0\times\infty$ and so these cases are treated the same as above.

Note that there is no type $0^\infty$ despite the fact that this seems like it could potentially be indeterminate. The intuitive reason for this is that as soon as the base becomes less than the number 1, taking powers will actually drive the limit closer to 0. Hence $0^\infty = 0$ and is not an indeterminate form.

**Example 4.26**

Compute the limit $\lim\limits_{x\to 0^+} x\ln(x)$.

*Solution.* In the limit as $x\to 0$ the function $\ln(x)\to -\infty$, so this limit is indeterminate of type $0\times\infty$. We now have to make a choice as to which component to invert. Let us choose to first invert the $x$, giving

$$\begin{aligned}\lim_{x\to 0^+} x\ln(x) &= \lim_{x\to 0} \frac{\ln(x)}{1/x} \overset{\langle H\rangle}{=} \lim_{x\to 0} \frac{1/x}{-1/x^2}\\ &= \lim_{x\to 0} -x = 0.\end{aligned}$$

■

**Example 4.27**

Determine the limit $\lim\limits_{x\to 0} [\csc(x) - \cot(x)]$

*Solution.* It is easy to see that this limit is indeterminate of type $\infty-\infty$, but as everything is trigonometric, there is likely some simplification that can be performed. Writing everything in terms of $\sin(x)$ and $\cos(x)$ we get

$$\lim_{x\to 0}[\csc(x) - \cot(x)] = \lim_{x\to 0}\left[\frac{1}{\sin(x)} - \frac{\cos(x)}{\sin(x)}\right] = \lim_{x\to 0}\frac{1-\cos(x)}{\sin(x)}$$
$$\stackrel{\langle H\rangle}{=} \lim_{x\to 0}\frac{\sin(x)}{\cos(x)} = 0.$$

Note that alternatively in this last step, we could have avoided using L'Hôpital's Rule by writing

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

**Example 4.28**

Determine the limit $\lim_{x\to\infty}\left(1 + \frac{a}{x}\right)^x$.

*Solution.* It is not too hard to see that this is indeterminate of type $1^\infty$, so we must take logarithms in order to simplify our lives. Indeed, set $y = \left(1 + \frac{a}{x}\right)^x$ so that $\ln y = x\ln\left(1 + \frac{a}{x}\right)$. In the limit as $x \to \infty$, this is of type $0 \times \infty$, so we take the reciprocal of one of the elements and get

$$\lim_{x\to\infty} x\ln\left(1 + \frac{a}{x}\right) = \lim_{x\to\infty}\frac{\ln\left(1 + \frac{a}{x}\right)}{1/x}$$
$$\stackrel{\langle H\rangle}{=} \lim_{x\to\infty}\frac{(-a/x^2)/\left(1 + \frac{a}{x}\right)}{-1/x^2}$$
$$= \lim_{x\to\infty}\frac{a}{1 + \frac{a}{x}} = a.$$

Hence re-exponentiating, we get

$$\lim_{x\to\infty}\left(1 + \frac{a}{x}\right)^x = e^{\lim_{x\to a}\ln y} = e^a. \qquad \blacksquare$$

## 4.5 Derivatives and the Shape of a Graph

## 4.6 First Derivative Information

Knowing something about the derivative of a function can often lead us to insights about what it's graph looks like.

**Theorem 4.29**

Let $f(x)$ be a function with domain $(a, b)$. If $f'(x) = 0$ for all $a < x < b$ then $f(x)$ is a constant function.

This is a very useful theorem, as it effectively tells us that the process of reversing differentiation results in unique functions up to ad additive constant (Section 5.3). We can also use it to help us prove identities:

**Example 4.30**

Show that for all $x \geq 0$,
$$2\arcsin(x) = \arccos(1-2x^2).$$

*Solution.* Consider the function $f(x) = 2\arcsin(x) - \arccos(1-2x^2)$. Differentiating this function gives
$$f'(x) = \frac{2}{\sqrt{1-x^2}} - \frac{4x}{\sqrt{1-(1-2x^2)^2}}.$$
By expanding the $1-(1-2x^2)^2$ term we get
$$1-(1-2x^2)^2 = 1-(1-4x^2+4x^4) = 4x^2-4x^4 = 4x^2(1-x^2)$$
and our above equation becomes
$$f'(x) = \frac{2}{\sqrt{1-x^2}} - \frac{4x}{2x\sqrt{1-x^2}} = \frac{2}{\sqrt{1-x^2}} - \frac{2}{\sqrt{1-x^2}} = 0$$
Hence $f(x)$ is a constant function, say $f(x) = C$. We need only determine the value of $C$, which can be done by substituting any value of $x$ into our function. For example, substituting $x=0$ gives
$$C = f(0) = 2\arcsin(0) - \arccos(1) = 0 - 0 = 0$$
so that $C = 0$. Thus $f(x) = 0$ for all $x \geq 0$, and by re-arranging this we get the desired equality. ■

**Definition 4.31**

We say that a function $f(x)$ is *increasing* if whenever $a < b$ then $f(a) \leq f(b)$, and *strictly increasing* if whenever $a < b$ then $f(a) < f(b)$.
In the same vein, we say that $f(x)$ is *decreasing* if whenever $a < b$ then $f(a) \geq f(b)$, and *strictly decreasing* if whenever $a < b$ then $f(a) > f(b)$.

**Theorem 4.32**

If $f(x)$ is a function with domain $(a,b)$ and $f'(x) > 0$ for all $a < x < b$ then $f$ is strictly increasing.

The above theorem can also be used to determine where a function $f(x)$ is increasing and decreasing, by determining the intervals on which $f'(x) > 0$ and $f'(x) < 0$ respectively.

**Example 4.33**

Show that the function $f(x) = \arctan(x)$ is everywhere strictly increasing.

*Solution.* Differentiating $f(x)$ we get
$$f'(x) = \frac{1}{1+x^2}$$
which is positive for every real number $x$. Hence by Theorem 4.32 we know that $f(x) = \arctan(x)$ is everywhere strictly increasing. ■

If a function is going to change from increasing to decreasing, its derivative will pass through zero or a singularity. For this reason, we define the following:

**Definition 4.34**

If $f(x)$ is a function and $c$ is in the domain of $f(x)$, then we say that $c$ is a *critical point* of $f(x)$ if $f'(c) = 0$ or $f'(c)$ does not exist.

Critical points therefore represent the possible points where a function could change from increasing to decreasing.

**Example 4.35**

Compute the critical points of the functions $f(x) = x^2$ and $g(x) = \cos(x)$.

*Solution.* Differentiating our functions yields $f'(x) = 2x$ and $g'(x) = -\sin(x)$. Solving $f'(x) = 0$ gives $x = 0$ (the location of the global min) and $g'(x) = 0$ gives the points $x = n\pi$ for $n$ an integer. ■

**Example 4.36**

Determine the intervals on which the function $f(x) = 2x^3 + 3x^2 - 12x + 15$ is increasing and decreasing.

*Solution.* Differentiating our function gives

$$f'(x) = 6x^2 + 6x - 12 = 6(x^2 + x - 2) = 6(x + 2)(x - 1).$$

The derivative is zero at $x = -2$ and $x = 1$, so we check the surrounding intervals to determine the sign of $f'(x)$:

| | $x < -2$ | $-2 < x < 1$ | $1 < x$ |
|---|---|---|---|
| $x + 2$ | $-$ | $+$ | $+$ |
| $x - 1$ | $-$ | $-$ | $+$ |
| $f'(x)$ | $+$ | $-$ | $+$ |

So we conclude that our function is strictly increasing on $(-\infty, -2)$ and $(1, \infty)$. ■

### 4.6.1 Second Derivative Information

**Definition 4.37**

A function $f(x)$ is defined to be *concave down* on an interval $[a, b]$ if for every $c \in [a, b]$ the graph of the function $f(x)$ as restricted to $[a, b]$ lies beneath the tangent line to $f(x)$ at the point $c$. Similarly, $f(x)$ is said to be *concave up* on $[a, b]$ if for every $c \in [a, b]$ the graph of the function $f(x)$ restricted to $[a, b]$ lies above the tangent line to $f(x)$ at the point $c$.

This is a rather cumbersome definition: to determine whether a function is concave up/down one would need to find the equation of a tangent line at each point in an interval and then manipulate complex inequalities to show that the graph of the function lies below/above that tangent line. As with most concepts in mathematics, the introductory idea is complicated and cumbersome, but often gives way to a much more simple characterization with the use of more sophisticated tools.

**Definition 4.38**

We say that a point $c$ is an *inflection point* of $f(x)$ is $f(x)$ changes from being convex up to convex down (or vice-versa) at the point $c$.

**Proposition 4.39**

Let $f(x)$ be a function and $c$ be some point in the domain of $f(x)$.

1. If $f''(c) > 0$ then $f(x)$ is concave up at $c$,
2. If $f''(c) < 0$ then $f(x)$ is concave down at $c$,
3. If $f''(c) = 0$ then let $k$ be the smallest positive integer such that $f^{(k)}(c) \neq 0$. If $k$ is odd then $c$ is an inflection point. Otherwise, let $k$ be even.
   (a) If $f^{(k)}(c) > 0$ then $f(x)$ is concave up at $c$,
   (b) If $f^{(k)}(c) < 0$ then $f(x)$ is concave down at $c$.

Proposition 4.39 thus tells us that points where $f''(x) = 0$ are candidates for inflection points.

**Example 4.40**

Let $f(x) = \ln(x^2+1) - x$. Determine and classify the critical points of $f(x)$, find where $f(x)$ is increasing/decreasing, and determine the intervals of concavity for $f(x)$.

*Solution.* To solve for the critical points as well as increasing/decreasing, we must compute the first derivative:

$$f'(x) = \frac{2x}{x^2+1} - 1 = \frac{2x - x^2 - 1}{x^2+1} = -\frac{(x-1)^2}{x^2+1}.$$

The only critical point of this function thus corresponds to $x = 1$ (since $f'(1) = 0$). Furthermore, we can see that $f'(x) \leq 0$ for all $x$, since both $(x-1)^2$ and $x^2+1$ are always non-negative, so $f(x)$ is decreasing on all of $\mathbb{R}$. To classify the critical point and compute concavity, we use the second derivative:

$$f''(x) = -\left[\frac{2(x-1)(x^2+1) - 2x(x-1)^2}{(x^2+1)^2}\right] = -\frac{2(x-1)(x+1)}{(x^2+1)^2}.$$

Let us check concavity first, which corresponds to determining the sign of $f''(x)$. Since the denominator is always positive, this reduces to determining the sign of $-(x-1)(x+1)$, which we do with the following table:

| | $x<-1$ | $-1<x<1$ | $x>1$ |
|---|---|---|---|
| $x+1$ | $-$ | $+$ | $+$ |
| $x-1$ | $-$ | $-$ | $+$ |
| $-(x-1)(x+1)$ | $-$ | $+$ | $-$ |

Hence the function is concave down on $(-\infty,-1)\cup(1,\infty)$ and concave up on $(-1,1)$. This tells us that $-1$ and $1$ are the inflection points, which is actually sufficient to tell us that the critical point $x=1$ is neither a max nor a min. Of course, we could also compute the third derivative

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

for which $f^{(3)}(1)\neq 0$. Our concavity criterion hence corroborates the fact that $x=1$ is an inflection point. ■

**Example 4.41**

Let $f(x)=\cos^2(x)-2\sin(x)$ be defined on $[0,2\pi]$. Find and classify the critical points of $f(x)$, and determine the intervals of concavity.

*Solution.* The first derivative of $f(x)$ is computed as

$$f'(x) = -2\cos(x)\sin(x) - 2\cos(x) = -2\cos(x)(\sin(x)+1).$$

Hence the critical points will occur when $\cos(x)=0$ and $\sin(x)=1$ on the interval $[0,2\pi]$. For $\cos(x)=0$ this corresponds to $x=\pi/2$ and $x=3\pi/2$ while for $\sin(x)=1$ we also get $x=\pi/2$. For the second derivative, we get

$$\begin{aligned} f''(x) &= 2\sin^2(x) - 2\cos^2(x) + 2\sin(x) = 2\sin^2(x) - 2(1-\sin^2(x)) + 2\sin(x) \\ &= 4\sin^2(x) + 2\sin(x) - 2. \end{aligned}$$

At the critical point $x=\pi/2$ we get $f''(\pi/2) = 4(1)^2+2(1)-2 = 4 > 0$ implying that $x=\pi/2$ is a minimum. On the other hand, the critical point $x=3\pi/2$ gives $f''(3\pi/2) = 4(-1)^2+2(-1)-2=0$ so might be an inflection point: we will have the further investigate.

One way of continuing our investigation would be to continue to take derivatives. Before doing that, let us see if we can determine the intervals of convexity. Our equation for $f''(x)$ actually looks like a quadratic where instead of $x$ we have $\sin(x)$. Nonetheless, we can still solve using the quadratic formula and get

$$\sin(x) = \frac{-1\pm\sqrt{1-4(2)(-1)}}{4} = \frac{-1\pm 3}{4} = -1, \frac{1}{2}.$$

Solving $\sin(x)=-1$ gives $x=3\pi/2$ and solving $\sin(x)=\frac{1}{2}$ gives $x=\pi/6$. The sign of $f''(x)$ is given by

| | $x<\pi/6$ | $\pi/6<x<3\pi/2$ | $x>3\pi/2$ |
|---|---|---|---|
| $f''(x)$ | $-$ | $+$ | $+$ |

.

So $f''(x)$ is concave down on $[0,\pi/6]$ and concave up on $[\pi/6,2\pi]$. ■

**Example 4.42**

Consider the function $f(x) = \dfrac{1+x}{1+x^2}$. Determine all inflection points of this curve.

*Solution.* We plug along until we get to the second derivative of $f(x)$. The first derivative is given by

$$f'(x) = \frac{(1+x^2) - 2x(1+x)}{(1+x^2)} = -\frac{x^2+2x-1}{x^2+1},$$

while the second derivative is

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

The inflection points come from finding the zeroes of the numerator. It is not too hard to see that $x = 1$ is a zero, so we can factor this to get

$$x^3 + 3x^2 - 3x - 1 = (x-1)(x^2+4x+1)$$

and this second term has roots $x = -2 \pm \sqrt{3}$. It is not too hard to convince ourselves that these are inflection points (an easy argument comes from the fact that there are three distinct linear roots with no multiplicities, hence the sign of the third derivative will change after passing any one root, implying each root is a proper inflection point). ■

## 4.7 Maxima and Minima

This section describes one of the most important applications of calculus: the ability to find necessary conditions for a point to be an extreme point of a function. Solving such problems is of exceptional importance. For example, all of classical mechanics works on the principle of *least* action: minimizing the difference between potential and kinetic energy. Relativity theory implies that gravity operates by moving particles along geodesics: paths of *minimal* length. When we sent people to the moon it was important to do it in the *quickest* amount of time while using the *least amount of fuel*. In business, we often like to *maximize* profits while *minimizing* waste. So how does calculus give us the ability to find extreme points? In order to make sense of this, we should first define what it means to be a max/min!

**Definition 4.43**

Let $f(x)$ be a function with domain $D$. We say that $c$ is an *absolute maximum* if $f(x) \leq f(c)$ for all $x \in D$, and a *local maximum* if $f(x) \leq f(c)$ for all $x$ in a interval around $c$. Similarly, we say that $c$ is an *absolute minimum* if $f(x) \geq f(c)$ for all $x \in D$, and a local minimum if $f(x) \geq f(c)$ for all $x$ in a interval around $c$.

**Example 4.44**

Find (in a hand-waving fashion) the absolute and local maxima and minima of the following functions:

$$f(x) = x^2, \qquad g(x) = \cos(x), \qquad h(x) = x\sin(x).$$

*Solution.* Our intuition tells us that we should look at the plots of each of these functions. Notice that $f(x) = x^2$ has a global minimum at the point $x = 0$, but increases to infinity otherwise. Hence $f(x)$ has a global min at $x = 0$ and no other extreme points.

The function $g(x) = \cos(x)$ has a global max wherever $\cos(x) = 1$; that is, the points $2n\pi$ for $n$ an integer. Similarly, $g(x)$ has a global min wherever $\cos(x) = -1$, corresponding to the set

$$(2n + 1)\pi \text{ where } n \text{ is an integer.}$$

Hence $g(x)$ has multiple global maxima and minima.

Finally, if we look at the graph of the function $h(x) = x\sin(x)$ we see that there are many local maxima and minima, but as the amplitude of the function grows as $|x|$ gets quite large, there cannot be any global max or min. ■

Specifying the domain is very important. For example, the function $f(x) = x$ has no global/local maximum/minimum on $\mathbb{R}$ or $(0, 1)$, but over the interval $[0, 1]$ it has a minimum at $x = 0$ and a maximum at $x = 1$. The following useful theorem tells us that intervals of the form $[a, b]$ ensure that maxima and minima always exist:

**Theorem 4.45: Extreme Value Theorem**

If $f(x)$ is a continuous function on a closed interval $[a, b]$, then $f(x)$ will attain its maximum and minimum on $[a, b]$.

**Example 4.46**

Determine on which of the following intervals the function $1/x$ is guaranteed to attain its global maximum and minimum.

$$[-1, 1], \qquad (0, 1), \qquad [1, 2].$$

*Solution.* According to the Extreme Value Theorem, we can be guaranteed that $1/x$ attains its max and min if it is continuous on a closed interval. The first interval $[-1, 1]$ is closed but $1/x$ is not continuous at 0 which is a point in $[-1, 1]$. Hence we cannot guarantee that $1/x$ attains a max and min on $[-1, 1]$, though note that it does attain global minima at $x = \pm 1$. Similarly, the interval $(0, 1)$ is not closed so we cannot guarantee that the maximum or minimum is attained. In fact, there is no max or min of $1/x$ in $(0, 1)$. Finally, $1/x$ *is* continuous on $[1, 2]$ and so by the Extreme Value Theorem the max and min are attained. ■

This is all fine and dandy, but this is an existential theorem, meaning that it tells us when extrema exist but fails to provide any information on how to find them. In fact, we notice that the Extreme Value Theorem only requires the concept of continuity and so does not really fall within the regime of calculus. The real power of calculus is to provide a necessary condition for a point to be an extreme point.

**Theorem 4.47**

Let $f(x)$ be a differentiable function on the interval $[a,b]$. If $a < c < b$ is a local max or min of the function $f(x)$, then it is a critical point of $f(x)$.

The idea is that if a max/min occurs on the interior of the domain, then the function must curve back on itself. Think again of the function $f(x) = x^2$ which we noticed had a minimum at $x = 0$. We could ensure it was a local min (and in fact a global on) because the function decreased until it hit $x = 0$, then started to increase again. This means that at some point, the slope $f'(x)$ of the function must have been zero. Similarly, if the max/min occurs at the endpoint of a domain then the derivative there did not exist.

*Proof.* If the domain of $f(x)$ has endpoints and the max/min occur on the endpoints, then $f'(c)$ does not exist and we are done. Hence let us assume that $c$ occurs inside the domain of $f(x)$. We shall proceed by showing the proof when $c$ is a maximum value, and we claim that the proof when $c$ is a minimum value follows almost precisely the same logic.

Since $c$ is a maximum value, we know that for all $x$ in a neighbourhood of $c$ that $f(x) \leq f(c)$. In particular, for sufficiently small $h > 0$ we have $f(c+h) \leq f(c)$ or equivalently, $f(c+h) - f(c) \leq 0$. Dividing by $h$ we still get that $\frac{f(c+h)-f(c)}{h} \leq 0$ and taking the limit as $h \to 0^+$, we get

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

If we now let $h < 0$, then again $f(c+h) - f(c) \leq 0$, but this time dividing by $h$ changes the sign. Taking the limit as $h \to 0^-$ gives

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

Since the both one sided limits exist, the two sided limit also exists and we have

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

As mentioned previously, the proof for when $c$ is a local minimum proceeds in exactly the same manner but with the inequalities reversed. $\square$

The idea behind why maxima and minima on the interior of the domain correspond to critical points gives us the following test for maximality/minimality:

**Theorem 4.48: First Derivative Test**

Let $f : [a,b] \to \mathbb{R}$ be differentiable with critical point $c \in (a,b)$.

1. If there exists $\delta > 0$ such that $f'(x) > 0$ on $(c-\delta, c)$ and $f'(x) < 0$ on $(c, c+\delta)$ then $c$ is a maximum.
2. If there exists $\delta > 0$ such that $f'(x) < 0$ on $(c-\delta, c)$ and $f'(x) > 0$ on $(c, c+\delta)$ then $c$ is a minimum.

**Example 4.49**

Find the (local) maxima and minima of the function $f(x) = 6x^4 - 3x^2 + 2$.

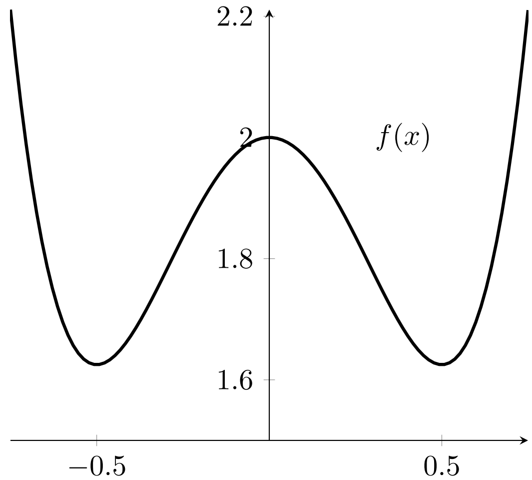



Figure 36: The function $f(x) = 6x^4 - 3x^2 + 2$ from Example 4.49

.

*Solution.* We begin by finding the critical points, so we differentiate to get $f'(x) = 24x^3 - 6x = 6x(4x^2 - 1) = 6x(2x-1)(2x+1)$. Setting $f'(x) = 0$ and solving for $x$, we get $x = 0, \pm\frac{1}{2}$. Now since $f'(x)$ splits into linear factors, and each linear factor can only switch sign once, we can determine whether $f$ is increasing or decreasing on each interval by creating the following chart:

| | $x < -\frac{1}{2}$ | $-\frac{1}{2} < x < 0$ | $0 < x < \frac{1}{2}$ | $x > \frac{1}{2}$ |
|---|---|---|---|---|
| $2x+1$ | $-$ | $+$ | $+$ | $+$ |
| $x$ | $-$ | $-$ | $+$ | $+$ |
| $2x-1$ | $-$ | $-$ | $-$ | $+$ |
| $f(x)$ | $-$ | $+$ | $-$ | $+$ |

From our chart we see that $-1/2$ is a min, 0 is a max, and $\frac{1}{2}$ is a min. This is confirmed by the graph of our function. ■

Note however that not all critical points are a max or a min.

**Example 4.50**

What are the critical points of the function $f(x) = x^3$? What are the max/min of $f(x)$?

*Solution.* The derivative of $f(x)$ is $f'(x) = 3x^2$ which has a single zero at $x = 0$ and hence has a single critical point at 0. However, the function $f(x)$ has no local maximum or minimum anywhere, so $x = 0$ is a critical point which is neither a maximum nor a minimum. ■

Theorem 4.47 combined with Example 4.50 imply that while every max/min is a critical point, not all critical points are max/mins. Regardless, critical points provide a powerful tools for finding absolute maxima and minima. As we must also check whether the function achieves its max/min on the boundary, our strategy is as follows:

**Finding Extreme Points**

1. Determine the value of $f(x)$ at the boundary points.
2. Determine the critical points of $f(x)$ by computing the points where $f'(x) = 0$.
3. Evaluate the function $f(x)$ at its critical points.
4. The absolute max and min will be the largest and smallest values from steps (1) and (3).

**Example 4.51**

Determine the global maximum and minimum of the function $f(x) = xe^{-x^2}$ on the interval $[0, 1]$.

*Solution.* Following our algorithm above, we first evaluate $f(x)$ on its endpoints to find that

$$f(0) = 0, \qquad f(1) = \frac{1}{e}.$$

Next we determine the critical points: the derivative of $f(x)$ is given by $f'(x) = e^{-x^2}(1-2x^2)$. The component $e^{-x^2}$ is never zero, so the zeroes of $f'(x)$ will occur precisely when $1 - 2x^2 = 0$ which corresponds to $x = \pm 1/\sqrt{2}$. However, we notice that $-1/\sqrt{2}$ is not in the interval $[0, 1]$ so we throw it away and just plug $x = 1/\sqrt{2}$ into $f(x)$ to get

$$f\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}e^{-\frac{1}{2}} = \frac{1}{\sqrt{2e}}.$$

A quick comparison tells us that $f(0) < f(1) < f(1/\sqrt{2})$ so that 0 is the global minimum and $1/\sqrt{2}$ is the global maximum. ■

**Example 4.52**

Compute the global maximum and minimum of the function $f(x) = x - \arctan(x)$ on the interval $[-1, 1]$.

*Solution.* We recall that $\arctan(\pm 1) = \pm\frac{\pi}{4}$, so that

$$f(-1) = -1 + \frac{\pi}{4}, \qquad f(1) = 1 - \frac{\pi}{4}.$$

Next we compute the critical points. The derivative of $f(x)$ is

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

which has a single zero at $x = 0$. Plugging this back into our function we get $f(0) = 0 - \arctan(0) = 0$. A quick comparison reveals that

$$f(-1) < f(0) < f(1)$$

so that the global min is at $-1$ and the global max is at $+1$. ■

There is another way to tell whether the endpoints of of function will be a max and min. In the example above, we computed the derivative $f'(x) = \frac{x^2}{1+x^2}$ which is always non-negative, and away from $x = 0$ is actually always positive. This means that the function is always increasing, and a function which is always increasing cannot have any max/min on the interior of its domain. With increasing functions, the minimum must occur at the left-most endpoint, while the maximum must occur at the right-most endpoint.

The algorithm given above is excellent for determining global maxima and minima, but what if we want to determine which critical points are local extreme points?

**Theorem 4.53: Second Derivative Test**

Let $f(x)$ be a function which is twice continuously differentiable in a neighbourhood of a critical point $c$.

1. If $f''(c) > 0$ then $c$ is a local minimum.
2. If $f''(c) < 0$ then $c$ is a local maximum.
3. If $f''(c) = 0$ let $k$ be the smallest positive integer such that $f^{(k)}(c) \neq 0$. If $k$ is odd then $c$ is neither a max nor a min. If $k$ is odd and $f^{(k)}(c) > 0$ then $c$ is a local max, while if $f^{(k)}(c) < 0$ then $c$ is a local min.

Note the similarities between this and Proposition 4.39. Indeed, the Second Derivative Test simply states that if your function is concave up at a critical point, it must have been a minimum; while if your function is concave down, your critical point must have been a maximum.

**Example 4.54**

Classify all maxima/minima of the function $f(x) = 2x^3 - 3x^2 - 12x + 10$.

*Solution.* We determine the critical points via the derivative, which we compute to be $f'(x) = 6x^2 - 6x - 12 = 6(x^2 - x - 2) = 6(x-2)(x+1) = 0$. This has roots at $x = 2$ and $x = -1$. In order to determine whether these are local maxima or minima, we use the second derivative test. We compute $f''(x) = 12x - 6 = 6(2x-1)$, which evaluated at the critical points gives

$$f''(2) = 18 > 0, \qquad f''(-1) = -18 < 0$$

which implies that $x = 2$ is a minimum while $x = -1$ is a maximum. ■

**Example 4.55**

Classify the maxima and minima of the function $f(x) = x - \sin(x)$.

*Solution.* The derivative $f'(x)$ is given by $f'(x) = 1 - \cos(x)$. Setting $f'(x) = 0$ amounts to solving $\cos(x) = 1$ which occurs when $x = 2n\pi$ for $n \in \mathbb{Z}$. In order to determine which of these are maxima and which are minima, we compute the second derivative $f''(x) = \sin(x)$. Our goal is thus to determine the sign of $\sin(2n\pi)$ for different integers $n$. However, it is easy to see that this is always zero, so our second derivative test is at first indeterminate. The third derivative is given by $f^{(3)}(x) = -\cos(x)$ and this is non-zero for every value of $x = 2n\pi$. Since this occurred in the third derivative, our criterion above indicates that none of these points are maxima or minima. ■

### 4.7.1 Optimization

Here we consider constrained optimization, where our goal is to maximize/minimize a function subject to constraints. There are entire fields of mathematics and engineering dedicated entirely to the study of optimization. The principal difference between an optimization problem and a max/min problem is that optimization problems have *constraints*; that is, we are asked to find the maximize/minimize a quantity, but only so long as some other condition is satisfied. The addition of constraints typically makes the problem much more difficult (or if you're a mathematician, much more interesting).

Keeping with the mathematical philosophy that one should reduce new problems to problems which we have already solved, our goal will be to take constrained optimization problems and turn them into simple max/min problems. Typically, such questions will be posed as word-problems amounting to the following

"Maximize $f(x_1, \ldots, x_n)$ subject to the constraint that $g(x_1, \ldots, x_n) = c$."

Written abstractly, this is guaranteed to terrify the student; how does one go about solving such a system? Since this is only a single section in a course in single-variable calculus, it turns out that the constraint will almost always allow us rewrite $f(x_1, \ldots, x_n)$ as a function of a single variable, to which we can then apply our usual techniques. Let's look at an example:

**Example 4.56**

[a] An exiled queen and her entourage washed up on the coast of North Africa, and pleaded with a local king to be given a small plot of land that she might catch her breath and rebuild. The king agreed, and they settled on however so much land could be encompassed by a piece of oxhide. The clever queen cut the oxhide into many small strips, so that they stretched a distance of 800 metres. She then also settled beside a straight river so that the river formed one of the barriers of her new kingdom. Given that her primitive people can only lay the oxhide strips in straight lines, find the dimensions of the corresponding rectangle that maximizes the area of her new queendom.

[a] This is a modification of the problem of Queen Dido and the founding of Carthage. Dido's problem is also called an *isoperimetric inequality.* I have made modifications here so that we can solve it using the tools we have developed, but a proper treatment requires the *calculus of variations*.

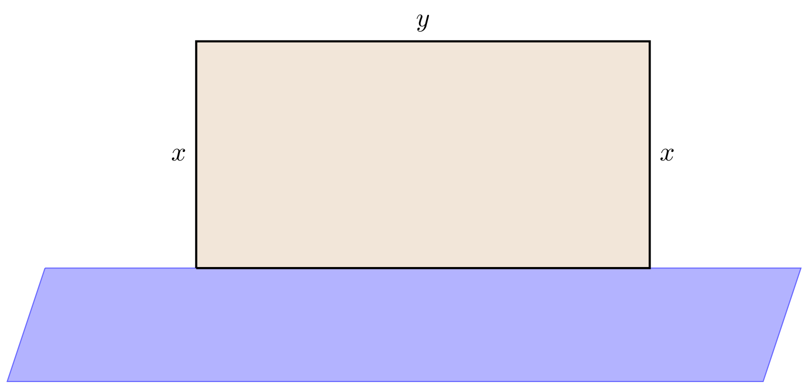



Figure 37: Establishing Queen Dido's kingdom, using the river as one of the edges.

*Solution.* Stripping away the extra information, we are told to build three sides of a rectangle which maximizes the area of the rectangle, ensuring that the perimeter is 800 metres. Let $y$ be the length of the side opposite the rivers, and $x$ be the length of the two sides perpendicular to the river. Our perimeter is thus $2x + y = 800$ while our area is $A = xy$; that is,

"Maximize the function $f(x, y) = xy$ subject to the constraint $2x + y = 800$."

As I mentioned earlier, the constraint can be manipulated so that $f(x, y) = xy$ becomes an equation of only one variable. Indeed, $2x + y = 800$ implies that $y = 800 - 2x$. Substituting this into $f(x, y)$ we get

$$\hat{f}(x) = f(x, y) = xy = x(800 - 2x) = 800x - 2x^2.$$

This new function already has the information about the constraint encoded into it, so we now content ourselves to simply find the maximum of $\hat{f}(x)$. Finding the critical points we get $\hat{f}'(x) = 800 - 4x = 0$ implies that $x = 200$. We can now use our constraint $2x + y = 800$ to see that $y = 400$. To see that this is indeed a maximum, we note that $\hat{f}''(x) = -4 < 0$, so the second derivative test verifies maximality. ∎

If we think about this solution, it turns out that this is actually somewhat counter intuitive for the following reason: We found that the rectangle we should build is twice as long as it is wide, but if we had not built along the side of the river then solving the optimization problem would have revealed that the optimal area is given by a square (try this on your own, to see that $x = y = 200$). The fact that we built along the river means that our constraint equation changed, and so we see that even the slightest change to the constraint equation can make a significant difference to the solution.

Note that sometimes the constraints can be avoided by being clever, as the following example demonstrates.

**Example 4.57**

Find the area of the largest rectangle which can be inscribed in a circle of radius $R$.

*Solution.* The trick to doing this question is to choose an intelligent parameterization of the rectangles. Assume that the circle is centred at the origin and notice that every inscribed rectangle can be uniquely prescribed by the angle one of its vertices makes with the positive $x$ axis (see Figure 38). In this way one may construct the area function $A(x) = 4R^2 \sin\theta\cos\theta = 2R^2\sin(2\theta)$ for[9] $\theta \in [0, \pi/2]$. We differentiate to find the critical points, giving

$$0 = A'(\theta) = 4R^2\cos(2\theta)$$

which has solution $\theta = \pi/4$. It is not too hard to see that this is a max, since $\sin(2\theta)$ has a max at $\theta = \pi/4$. Furthermore, the two endpoints actually have zero area, so this is the only such maximum.

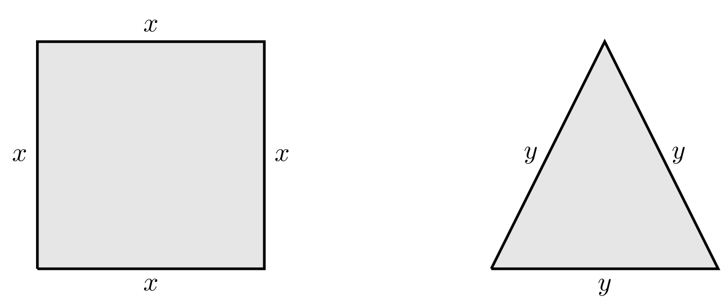



Figure 38: The construction of the parameterization of inscribed rectangles of a circle of radius $R$.

■

**Example 4.58**

Consider a segment of string of length 20 centimetres. If we cut this string into two segments and from from each segment a square and an equilateral triangle, find the cuts which will both maximize and minimize the sum of the areas of each shape.

[9] Strictly speaking, we need only take $\theta \in [0, \pi/4]$ but this requires arguing about the symmetry of the problem which will just serve to complicate the solution.

*Solution.* Let the side length of our square be given by $x$ and the side length of the triangle by $y$. Our constraint implies that the sum of the perimeters of these shapes must be 20 centimetres, so $4x + 3y = 20$. Now one can calculate that the area of an equilateral triangle of side-length $y$ is $A_\Delta = \frac{\sqrt{3}}{4}y^2$ and the area of the square is obviously $A_\square = x^2$. Hence our problem is

Maximize/minimize $f(x, y) = A_\Delta + A_\square = x^2 + \frac{\sqrt{3}}{4}y^2$, subject to the constraint $4x + 3y = 20$.

As before, our constraint allows us to rewrite $f(x, y)$ as a function of just one variable. The equation $4x + 3y = 20$ implies that $y = (20 - 4x)/3$, and substituting this into $f(x, y)$ we get

$$f(x) = x^2 + \frac{\sqrt{3}}{4}\left(\frac{20-4x}{3}\right)^2 = x^2 + \frac{1}{12\sqrt{3}}\left(400 - 160x + 16x^2\right).$$

Differentiating to find the critical points, we get

$$0 = f'(x) = 2x + \frac{1}{12\sqrt{3}}(-160 + 32x)$$

which we can solve to get $x = 20/(3\sqrt{3} + 4)$. This implies that

$$y = \frac{20-4x}{3} = \frac{20\sqrt{3}}{3\sqrt{3}+4}.$$

Hence the side lengths of the square and triangle are, respectively:

$$\frac{80}{3\sqrt{3}+4}, \qquad \frac{60}{3\sqrt{3}+4}. \tag{4.14}$$

One can use a calculator to see that this corresponds to an approximate area of 10.87 cm$^2$. This is also a local minimum since it is easy to see that the second derivative of this function is always positive. By our previous treatment of max/min, we know that we must also check the endpoints for a solution. If $x = 0$ then $y = 20/3$ gives an area of 19.25 while if $y = 0$ then $x = 5$ gives an area of 25. Hence our area is maximized when we use the string to only make the square, and minimized with the side lengths given in (4.14). ■

### 4.7.2 Curve Sketching

The grand-total of all the tools hitherto developed give us that ability to analyze functions and determine the behaviour of their graphs. One of the important applications of this is that while we may implement Computer Algebra Systems to help us analyze functions, it is still essential for the operator to understand the fundamentals in order to find things that a computer would otherwise miss. A simple but important example is as follows: Consider the function $f(x) = \frac{1}{300}(x^4 - 2x^2 + 1)$. If we were to graph this using software, we might get the graph in Figure 39: (Left).

Now let us assume that this figure describes the potential energy of a system. Any state will try to minimize its potential energy, and so it is tempting to assume that the point at $x = 0$ describes a global (stable) minimum and so would be an excellent place to initialize a state. However, the use of calculus actually shows that the point $x = 0$ is a local maximum and hence is an unstable equilibrium. If implemented as an engineering solution, this could quickly lead to disaster.

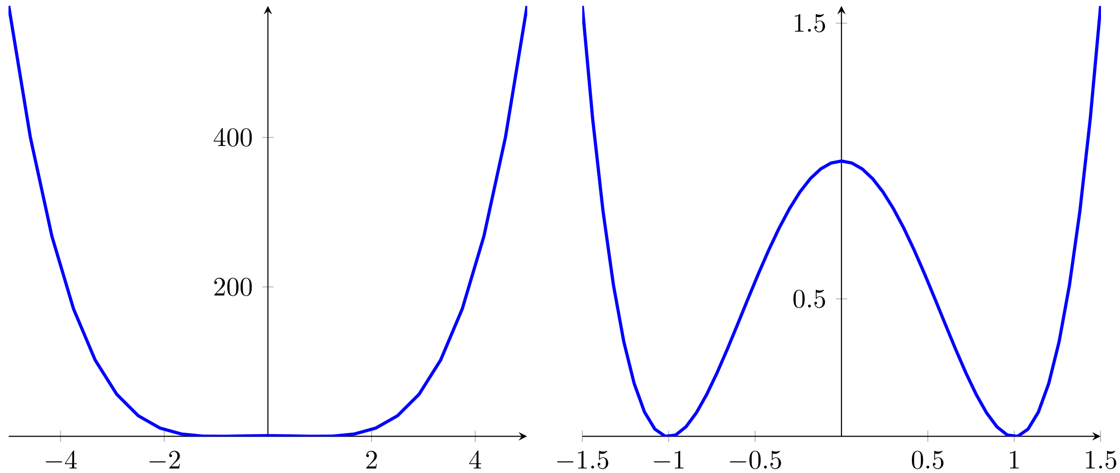



Figure 39: The graph of a misleading function. **Left:** The graph of our function as given by a computer. **Right:** Zooming in, we see the presence of a local maximum at $x = 0$.

Our goal is thus to combine all of our information into a system which allows us to analyze the qualitative behaviour of a function without knowing the nitty-gritty details of its exact value at every point. There are approximately seven pieces of information that we need to compute to ascertain the general behaviour.

1. Domain (and range of if possible),
2. Intercepts ($x$- and $y$-),
3. Symmetry (even/odd/none),
4. Asymptotes (horizontal/vertical/oblique),
5. Increasing (and obviously decreasing),
6. Maxima (and minima),
7. Concavity.

Of course, there is one piece of information above that I have not included; namely, what is an oblique (aka slant) asymptote?

**Definition 4.59**

Let $f(x)$ and $g(x)$ be continuous functions. We say that $f(x)$ *behaves like* $g(x)$ *asymptotically* if

$$\lim_{x\to\pm\infty} [f(x) - g(x)] = 0.$$

We say that $f(x)$ has an *oblique asymptote* if $f(x)$ behaves asymptotically like $g(x) = mx+b$ for some $m \neq 0$.

**Example 4.60**

Compute any asymptotics of the function $f(x) = \dfrac{2x^3 - x^2 + 2x}{x^2+1}$.

*Solution.* The easiest way to proceed is to try to write $f(x)$ as an improper rational function. Performing long division, we see that

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

The idea is that in the limit as $x \to \infty$, then $1/(x^2+1)$ term will die off and contribute very little to the behaviour of $f(x)$, so that $f(x)$ looks like the function $2x-1$. To see that this satisfies the definition above, set $g(x) = 2x - 1$ so that

$$\begin{aligned}\lim_{x\to\infty}[f(x) - g(x)] &= \lim_{x\to\infty}\left[2x - 1 + \frac{1}{x^2+1} - (2x-1)\right] \\ &= \lim_{x\to\infty}\frac{1}{x^2+1} = 0,\end{aligned}$$

which is precisely what we wanted to show. Similarly, the limit as $x \to -\infty$ shows that $f(x)$ is also asymptotically like $g(x)$ in that limit as well. Thus $2x-1$ is an oblique asymptote for $f(x)$ at both $\pm\infty$. ■

Now that we know how to compute all terms involved in this computation, we shall proceed with some examples.

**Example 4.61**

Plot the function $f(x) = \dfrac{x^3}{(x+1)^2}$.

*Solution.* **Domain:** The only point which could possibly give us trouble is $x = -1$. Hence our domain is simply $\mathbb{R} \setminus \{-1\}$.

**Intercepts:** The $y$-intercept occurs when $x = 0$, so namely $f(0) = 0$. Similarly the $x$-intercept comes when $y = 0$, for which we see that

$$\frac{x^3}{(x+1)^2} = 0, \qquad \Leftrightarrow \qquad x = 0.$$

Thus the $x$- and $y$-intercepts both occur at the origin.

**Symmetry:** There is no symmetry involved: Since the functions are polynomial they have no periodicity. The student may check that $f(-x)$ has no relation to $f(x)$, so that the function is neither even nor odd.

**Asymptotes:** We begin with the horizontal asymptotes. It is easy to see that since the numerator dominates the denominator, the limit will go to infinity (check this by dividing top and bottom by $1/x^3$). Further, since the denominator is always positive, the sign is determined entirely by the $x^3$ factor, so

$$\lim_{x\to\infty}\frac{x^3}{(x+1)^2}=\infty,\qquad \lim_{x\to-\infty}\frac{x^3}{(x+1)^2}=-\infty.$$

We conclude there are no horizontal asymptotes.

The only candidate for a vertical asymptote occurs at $x=-1$. Again the denominator $(x+1)^2$ is always positive, so the sign of the "infinity" is entirely determined by the behaviour of $x^3$ around $x=-1$, which is negative. It is then clear that

$$\lim_{x\to 1^\pm}\frac{x^3}{(x+1)^2}=-\infty.$$

Finally, we want to check for oblique asymptotes. Using long polynomial division we may easily find that

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

so we claim that $y=x-2$ is an oblique asymptote. Indeed, notice that

$$\begin{aligned}
\lim_{x\to\pm\infty}[f(x)-(x-2)] &= \lim_{x\to\pm\infty}\left[\left((x-2)-\frac{3x+2}{x^2+2x+1}\right)-(x-2)\right]\\
&= -\lim_{x\to\pm\infty}\frac{3x+2}{x^2+2x+1}\\
&= -\lim_{x\to\pm\infty}\frac{3/x+2/x^2}{1+2/x+1/x^2} && \text{multiply and divide by } 1/x^2.\\
&= 0.
\end{aligned}$$

**First Derivative:** Computing the first derivative can be a chore, but we find that

$$\begin{aligned}
f'(x) &= \frac{3x^2(x+1)^2-2(x+1)x^3}{(x+1)^4}\\
&= \frac{3x^4+6x^3+3x^2-2x^4-2x^3}{(x+1)^4}\\
&= \frac{x^2(x+3)(x+1)}{(x+1)^4}\\
&= \frac{x^2(x+3)}{(x+1)^3}
\end{aligned}$$

so that the critical points correspond to $x=-1$, $x=0$ and $x=-3$. The $y$-values for these points will be useful when we plot, so we substitute to find that $f(0)=0$ and $f(-3)=27/16$. To determine where the function is increasing and decreasing, we consider the following table:

| | $x<-3$ | $-3<x<-1$ | $-1<x<0$ | $0<x$ |
|---|---|---|---|---|
| $x+3$ | $-$ | $+$ | $+$ | $+$ |
| $(x+1)^3$ | $-$ | $-$ | $+$ | $+$ |
| $x^2$ | $+$ | $+$ | $+$ | $+$ |
| $f(x)$ | $+$ | $-$ | $+$ | $+$ |

**Second Derivative:** The second derivative is a little messy, but simplifies if done correctly.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\frac{x^2(x+3)}{(x+1)^3} &= \frac{(3x^2+6x)(x+1)^3 - 3(x+1)^2(x^3+3x^2)}{(x+1)^6} \\ &= \frac{3x^3+3x^2+6x^2+6x-3x^3-9x^2}{(x+1)^4} \\ &= \frac{6x}{(x+1)^4}\end{aligned}$$

so there is an inflection point at $(0,0)$ (telling us that one of the critical points is an inflection point. Furthermore, $f''(-3) = -9/4 < 0$ so the point $(3, 27/16)$ is a max. Since the denominator is a quartic it is always positive, and we can see that concavity is entirely determined by the numerator $6x$. Hence $f(x)$ is concave up when $f''(x) > 0$, corresponding to $x > 0$; and $f(x)$ is concave down when $f''(x) < 0$, corresponding to $x < 0$.

**Plotting:** Putting all of this information together, the student should get the following plot:

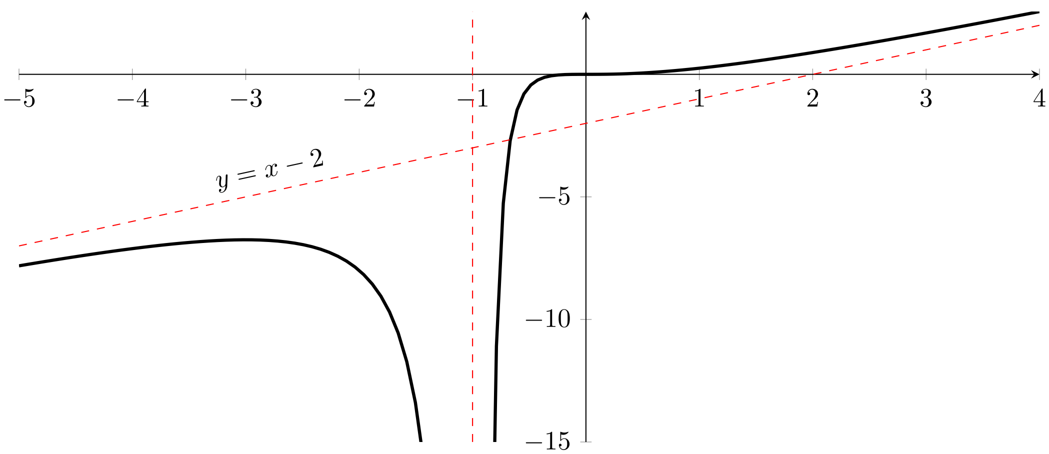



Figure 40: A plot of the curve $f(x) = \frac{x^3}{(1+x)^2}$.

■

**Example 4.62**

Plot the function $f(x) = \dfrac{x^3}{1-x^2}$.

*Solution.* Following our new six-step program we set to work:

**Domain:** By this point, this should not be too hard to see. In particular, our function will not be defined whenever the denominator is zero. This happens at the points $x = \pm 1$ and so our domain is $\mathbb{R} \setminus \{\pm 1\}$.

**Intercepts:** The $y$-intercept occurs when $x = 0$, so namely $f(0) = 0$. Similarly the $x$-intercept comes when $y = 0$, for which we see that

$$\frac{x^3}{1-x^2} = 0, \qquad \Leftrightarrow \qquad x = 0.$$

Thus the $x$- and $y$-intercepts both occur at the origin.

**Symmetry:** Since we are dealing with polynomials, there is no obvious periodicity to worry about. It's not too hard to see that this is actually an odd function, since

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

**Asymptotes:** The vertical asymptotes will clearly occur at $x = \pm 1$. Typically, one would calculate the limits

$$\lim_{x\to 1^\pm} \frac{x^3}{1-x^2}, \quad \lim_{x\to -1^\pm} \frac{x^3}{1-x^2}$$

but this is laborious and is redundant once we have information on the first derivative. For the interested student who would like to see how to do this all the same, we have the following table

| | $x^3$ | $1-x^2$ | $x^3/1-x^2$ |
|---|---|---|---|
| $x \to 1^+$ | $+$ | $-$ | $-$ |
| $x \to 1^-$ | $+$ | $+$ | $+$ |
| $x \to -1^+$ | $-$ | $+$ | $-$ |
| $x \to -1^-$ | $-$ | $-$ | $+$ |

so that

$$\lim_{x\to 1^-} \frac{x^3}{1-x^2} = \lim_{x\to -1^-} \frac{x^3}{1-x^2} = \infty, \qquad \lim_{x\to 1^+} \frac{x^3}{1-x^2} = \lim_{x\to -1^+} \frac{x^3}{1-x^2} = \infty$$

Because the degree of the numerator is strictly greater than the degree of the denominator, there are no horizontal asymptotes:

$$\lim_{x\to\pm\infty} \frac{x^3}{1-x^2} = \mp\infty.$$

Finally, we want to check for oblique asymptotes. Using long polynomial division we may easily find that

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

so we claim that $y = -x$ is an oblique asymptote. Indeed, notice that

$$\begin{aligned}
\lim_{x\to\pm\infty} [f(x) - (-x)] &= \lim_{x\to\pm\infty} \left[\frac{x^3}{1-x^2} + x\right] \\
&= \lim_{x\to\pm\infty} \frac{x}{1-x^2} \\
&= \lim_{x\to\pm\infty} \frac{1/x}{1/x^2 - 1} \\
&= 0.
\end{aligned}$$

**First Derivative:** This step allows us to determine where the function is increasing, decreasing, the critical points, and when combined with the second derivative, maxima and minima. The first derivative is computed to be

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\frac{x^3}{1-x^2} &= \frac{(3x^2)(1-x^2)-(-2x)(x^3)}{(1-x^2)^2} \\ &= \frac{x^2(3-x^2)}{(1-x^2)^2}.\end{aligned}$$

We may simply read off the critical points as $\left\{\pm 1, \pm\sqrt{3}, 0\right\}$ with potential extrema at $\left\{0, \pm\sqrt{3}\right\}$. Setting up a quick table for increasing and decreasing we have

| | $x < -\sqrt{3}$ | $-\sqrt{3} < x < -1$ | $-1 < x < 0$ | $0 < x < -1$ | $1 < x < \sqrt{3}$ | $x > \sqrt{3}$ |
|---|---|---|---|---|---|---|
| $f(x)$ | $-$ | $+$ | $+$ | $+$ | $+$ | $-$ |

This table is easy to deduce once we realize that the $\frac{x^2}{(1-x^2)^2}$ portion is always positive, so the sign of $f'(x)$ is entirely determined by the sign of $3-x^2$, which is negative whenever $|x| > \sqrt{3}$. It will likely be useful to know the function values corresponding to our critical points. We already know that $f(0) = 0$ and we find that

$$f(\pm\sqrt{3}) = \frac{\pm 3\sqrt{3}}{-2} = \frac{\mp 3\sqrt{3}}{2}.$$

**Second Derivative:** The second derivative is a little messy, but simplifies if done correctly.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\frac{x^2(3-x^2)}{(1-x^2)^2} &= \frac{(6x-4x^3)(1-x)^2 - 2(1-x^2)(-2x)(3x^3-x^4)}{(1-x^2)^4} \\ &= \frac{2x(x^2+3)}{(1-x^2)^3}.\end{aligned}$$

The inflection points will occur when $f''(x) = 0$ or does not exist, which we can again read off as being $\{0, \pm 1\}$. We form a table to check for concavity and find

| | $x < -1$ | $-1 < x < 0$ | $0 < x < 1$ | $x > 1$ |
|---|---|---|---|---|
| $f(x)$ | $+$ | $-$ | $+$ | $-$ |

Finally, recalling that we have extrema candidate at $\pm 3$ we check to find that

$$f(\pm\sqrt{3}) = \frac{\pm 2\sqrt{3}(3+3)}{(1-3)^2} = \pm\frac{12\sqrt{3}}{-8} = \mp\frac{3\sqrt{3}}{2}.$$

Thus $\left(\sqrt{3}, -\frac{3\sqrt{3}}{2}\right)$ is a local maximum and $\left(-\sqrt{3}, \frac{3\sqrt{3}}{2}\right)$ is a local minimum. Since $f''(0) = 0$ we cannot infer any information about this critical point. If we continue to take derivatives, we will find that $f^{(3)}(0) = 6$ and so by the generalized second derivative test, 0 is an inflection point.

**Plotting:** Putting all of this information together, the student should get the following plot:

■

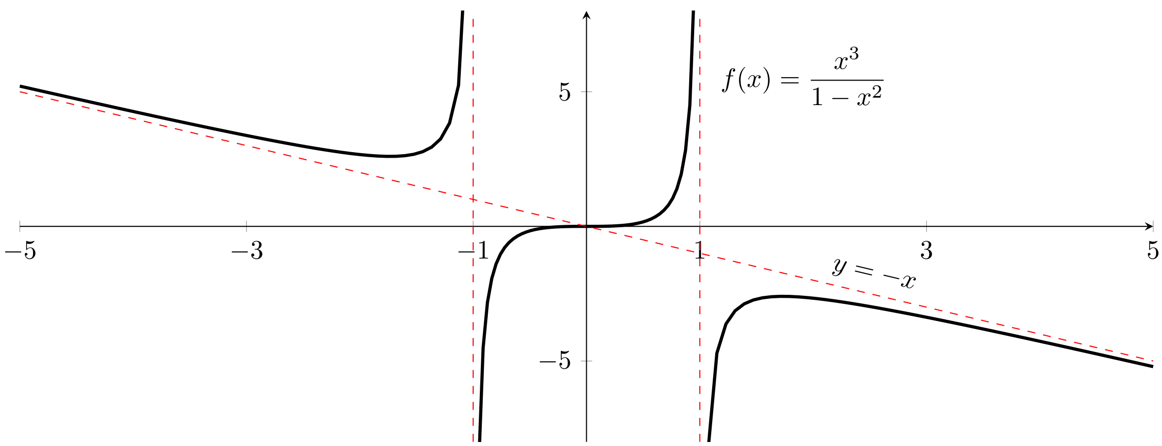



Figure 41: A plot of the curve $f(x) = \frac{x^3}{1-x^2}$.

# 5 Integration

The over-arching goal of integration is to add things together in a continuous fashion. This manifests in applications such as finding the area under a curve, the volume of an object, or even as calculating physical quantities such as work, flux, or voltage potentials. The fact that this is even remotely related to the process of differentiation is not at all obvious, though we will see shortly that there is in fact an intimate relationship.

## 5.1 Sigma Notation

As mentioned, integration will amount to continuous summation. To do this, we must first come to terms with finite summations, and the notation used therein. Sigma notation is used to make complicated sums much easier to write down. In particular, we use a summation index to iterate through elements of a list and then sum them together.

Consider the expression

$$\boxed{\sum_{i=n}^{m} r_i \qquad (5.1)}$$

which is read as "the sum from $i = n$ to $m$ of $r_i$." The element $i$ is known as the *dummy* or *summation* index, $n$ and $m$ are known as the *summation bounds*, and $r_i$ is the *summand*. In order to decipher this cryptic notation, we adhere to the following algorithm:

1. Set $i = n$ and write down $r_i$;

2. Add 1 to the index $i$ and add $r_i$ to the current sum;

3. If $i$ is equal to $m$ then stop, otherwise go to step 2 and repeat.

For those computer savvy students out there, this is nothing more than a for-loop. Interpreting (5.1) we thus have

$$\sum_{i=n}^{m} r_i = r_n + r_{n+1} + r_{n+2} + \cdots r_m.$$

**Example 5.1**

Set $r_1 = 5, r_2 = -8, r_3 = 4$. Compute $\displaystyle\sum_{i=1}^{3} r_i$.

*Solution.* Via our discussion above, we may write the summation explicitly as

$$\sum_{i=1}^{3} r_i = r_1 + r_2 + r_3 = 5 + (-8) + 4 = 1. \qquad \blacksquare$$

The $r_i$ could be a collection of unrelated numbers as in Exercise 5.1, but they could be a "function" of the index variable as follows:

**Example 5.2**

Compute $\displaystyle\sum_{i=1}^{4}(2i+1)$.

*Solution.* Following our algorithm, we start by setting $i = 1$ and then evaluating the summand. I will write out the steps in slightly more detail than usual to illustrate the process:

$$\begin{aligned}
\sum_{i=1}^{4}(2i+1) &= (2i+1)_{i=1} + (2i+1)_{i=2} + (2i+1)_{i=3} + (2i+1)_{i=4} \\
&= (2 \cdot 1 + 1) + (2 \cdot 2 + 1) + (2 \cdot 3 + 1) + (2 \cdot 4 + 1) \\
&= 3 + 5 + 7 + 9 \\
&= 24. \qquad \blacksquare
\end{aligned}$$

Sometimes we can find closed form expressions for summations. The student will not be expected to memorize the following, but they are nonetheless important identities:

$$\begin{aligned}
&\sum_{j=1}^{n} 1 = n && \sum_{j=1}^{n} j = \frac{n(n+1)}{2} \\
&\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6} && \sum_{j=1}^{n} j^3 = \left[\frac{n(n+1)}{2}\right]^2
\end{aligned} \tag{5.2}$$

Additionally, summations are linear, in that

$$\left[\sum_{i=n}^{m} a_i\right] + \left[\sum_{i=n}^{m} b_i\right] = \sum_{i=n}^{m}(a_i + b_i), \qquad c\left[\sum_{i=n}^{m} a_i\right] = \sum_{i=n}^{m}(ca_i).$$

Note that the upper and lower bounds of the summation are the same in every summation.

Using linearity and the identities in (5.2) we can redo Example 5.2 with a general upper bound, to find

$$\begin{aligned}\sum_{i=1}^n (2i+1) &= 2\left(\sum_{i=1}^n i\right) + \left(\sum_{i=1}^n 1\right)\\ &= 2\frac{n(n+1)}{2} + n \\ &= n(n+2).\end{aligned}$$

Plugging in $n=4$ we get 24, just as we found in Example 5.2.

**Remark 5.3** For any positive integer $p$, there is a closed form expression for

$$\sum_{i=1}^n i^p$$

but these expressions become more difficult as $p$ becomes larger. Luckily, there is a standard way of deriving the closed form for any $p$ using the *Bernoulli polynomials*, which are popular objects in the study of number theory but are tricky to define.

There are other summations which also admit closed form expressions, which are not evaluated as easily as the examples above.

**Example 5.4**

Guess a closed form expression for the summation $\displaystyle\sum_{i=1}^n \frac{1}{i^2+i}$.

*Solution.* We will try a few values of $n$, such as $n = 1, 2, 3, 4$; to see if we can spot a pattern. Indeed,

$$\begin{aligned}
n=1: &\qquad \sum_{i=1}^1 \frac{1}{i^2+i} = \frac{1}{2}\\
n=2: &\qquad \sum_{i=1}^2 \frac{1}{i^2+i} = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}\\
n=3: &\qquad \sum_{i=1}^3 \frac{1}{i^2+i} = \frac{2}{3} + \frac{1}{12} = \frac{3}{4}\\
n=4: &\qquad \sum_{i=1}^4 \frac{1}{i^2+i} = \frac{3}{4} + \frac{1}{20} = \frac{4}{5}.
\end{aligned}$$

Were we to guess, it looks as though

$$\sum_{i=1}^n \frac{1}{i^2+i} = \frac{n}{n+1}, \tag{5.3}$$

and indeed this is true. Can you prove it? ■

**Exercise:** Show that Equation (5.3) is correct.

## 5.2 The Definite Integral

### 5.2.1 The Intuition

To illustrate what we are try to accomplish, we revisit our car example from Section 2. Let $v(t)$ be a function which describes the instantaneous velocity of a car between time 0 and time 1. If our car is travelling at a constant velocity $v(t) = v_0$, some physics intuition tells us that

$$\text{distance} = \text{velocity} \times \text{time} = (1-0) \cdot v_0.$$

Travelling at constant velocity represents an exceptional case, so now assume that $v(t)$ is no longer constant. As in the case where we compute limits, we will try to approximate how far the car has travelled by looking at smaller and smaller time lengths. For example, if we divide $[0,1]$ into five subintervals of uniform length

$$[0,0.2], \quad [0.2,0.4], \quad [0.4,0.6], \quad [0.6,0.8], \quad [0.8,1],$$

it seems reasonable to assume that the velocity has not changed too dramatically over the course of 0.2 time units. Let's use the speed of the car at the beginning of each interval to approximate our speed:

| Interval | Approximate Speed | Approximate Distance Travelled |
|---|---|---|
| $[0,0.2]$ | $v(0)$ | $0.2\ v(0)$ |
| $[0.2,0.4]$ | $v(0.2)$ | $0.2\ v(0.2)$ |
| $[0.4,0.6]$ | $v(0.4)$ | $0.2\ v(0.4)$ |
| $[0.6,0.8]$ | $v(0.6)$ | $0.2\ v(0.6)$ |
| $[0.8,1]$ | $v(0.8)$ | $0.2\ v(0.8)$ |

The approximate total distance travelled is the sum of the approximations on each subinterval

$$0.2\ v(0) + 0.2\ v(0.2) + 0.2\ v(0.4) + 0.2\ v(0.6) + 0.2\ v(0.8).$$

This is a great starting point, but there are several things we should consider before proceeding:

1. One will compute a different approximation if, instead of the left endpoint, we use the right endpoint, the midpoint, or any other point in the interval to make the approximation. For example, instead of $v(0.2)$ on $[0.2,0.4]$ we might have used $v(0.3), v(0.4)$ or even something exotic like $v(\pi/10)$.
2. There is no reason to use uniform intervals. If the car has an initially slow acceleration, it might make more sense to use finer approximations as time progresses. For example, breaking our approximations into

$$[0,0.6], \quad [0.6,0.8], \quad [0.8,0.9], \quad [0.9,0.95], \quad [0.95,1]$$

might yield an even better approximation.

3. It seems natural that taking smaller intervals will lead to better approximations. This leads us to suspect that a limit will be at play.

We must take the three aforementioned points into consideration when trying to find the true distance travelled.

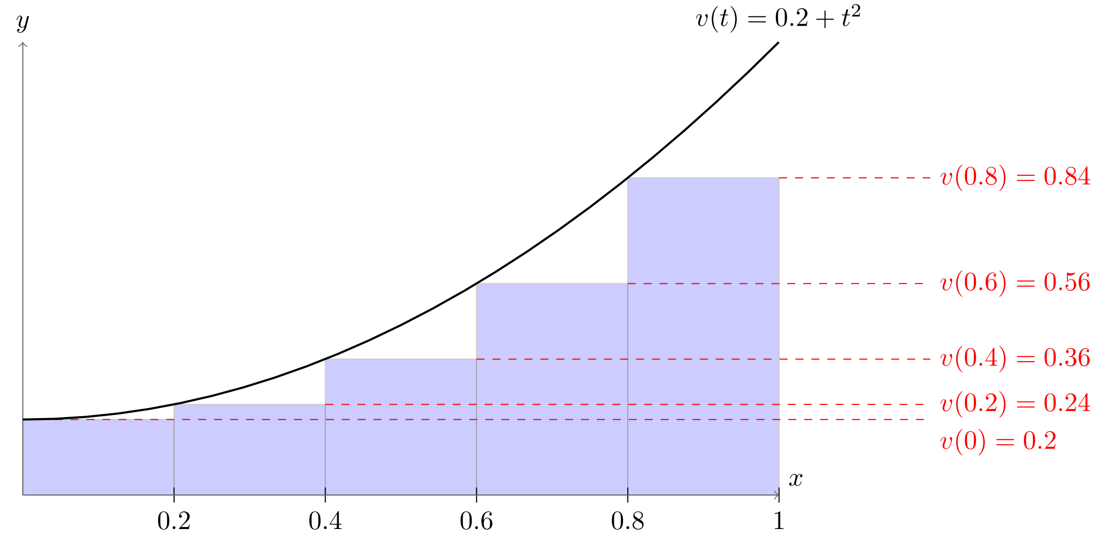



Figure 42: Our process is akin to computing the area of the blue rectangles. This estimate yields an approximate area of 0.44.

So far our conversation has been dominated by a discussion that lends little geometric insight as to what is happening. We can fix this by considering an actual function $v(t)$ and ascribing a geometric interpretation to the distance travelled. Let $v(t) = 0.2 + t^2$ for $t \in [0, 1]$, and consider the graph of $v(t)$. Breaking $[0, 1]$ into subintervals amounts to partitioning the $x$-axis into five even length segments; choosing to approximate our velocity by the left endpoint of each subinterval amounts to fixing a height within each subinterval. We then approximate the distance by multiplying the width of the subinterval by the chosen height; that is, we compute the area of a rectangle! In this particular example, we get

$$\begin{aligned}\text{Area} &= 0.2\,v(0) + 0.2\,v(0.2) + 0.2\,v(0.4) + 0.2\,v(0.6) + 0.2\,v(0.8)\\ &= 0.2\,(0.2 + 0.24 + 0.36 + 0.56 + 0.84) = 0.44.\end{aligned}$$

We can convince ourselves that in the limit, **the total distance travelled will be the area under the graph** (which we will later show is $8/15$).

### 5.2.2 Estimating Areas

Geometrically we want to estimate the area under the graph of a function. We have already seen one application in Section 5.2.1, and many more will follow when we get to Section 7. For now, we endeavor to establish the framework with which to do these computations.

Consider the case of a function $f(x)$ defined on an interval $[a, b]$. To estimate the area under the graph of $f(x)$ on $[a, b]$, we will need to do three things:

1. Break the interval $[a, b]$ into smaller subintervals,
2. On each subinterval, approximate the height of the function $f(x)$, creating a rectangle.
3. Sum together the area of the rectangles on each subinterval.

Strictly speaking, we are not required to use subintervals of uniform length, but to simplify our presentation we will enforce equal length partitions. For example, if we break the interval $[0, 2]$ into 5 subintervals, we get

$$[0, 0.4], \quad [0.4, 0.8], \quad [0.8, 1.2], \quad [1.2, 1.6], \quad [1.6, 2].$$

We will use the notation $\Delta x$ to indicate the length of each subinterval, which in this case is $\Delta x = 2/5 = 0.4$. In general, dividing the interval $[a, b]$ into $n$ equal length partitions results in

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

with subintervals

$$[a + i\Delta x, a + (i+1)\Delta x], \quad i = 0, \ldots, n-1. \tag{5.4}$$

There are also many ways to sample the height of $f(x)$ on each subinterval, but for our purposes we will limit our attention to three: the left-endpoint, the right-endpoint, and the mid-point, of each interval.

For the subinterval given in Equation (5.4), the left-endpoint is $a + i\Delta x$, the right-endpoint is $a+(i+1)\Delta x$, and the midpoint is $a+(i+1/2)\Delta x$. This gives the left, right, and midpoint Riemann sums:

Left Riemman Sum: $$L(f) = \sum_{i=0}^{n-1} f(a + i\Delta x)\, \Delta x$$

Right Riemman Sum: $$R(f) = \sum_{i=0}^{n-1} f(a + (i+1)\Delta x)\, \Delta x$$

Midpoint Riemman Sum: $$M(f) = \sum_{i=0}^{n-1} f\left(a + \left(i + \frac{1}{2}\right)\Delta x\right)\Delta x$$

**Example 5.5**

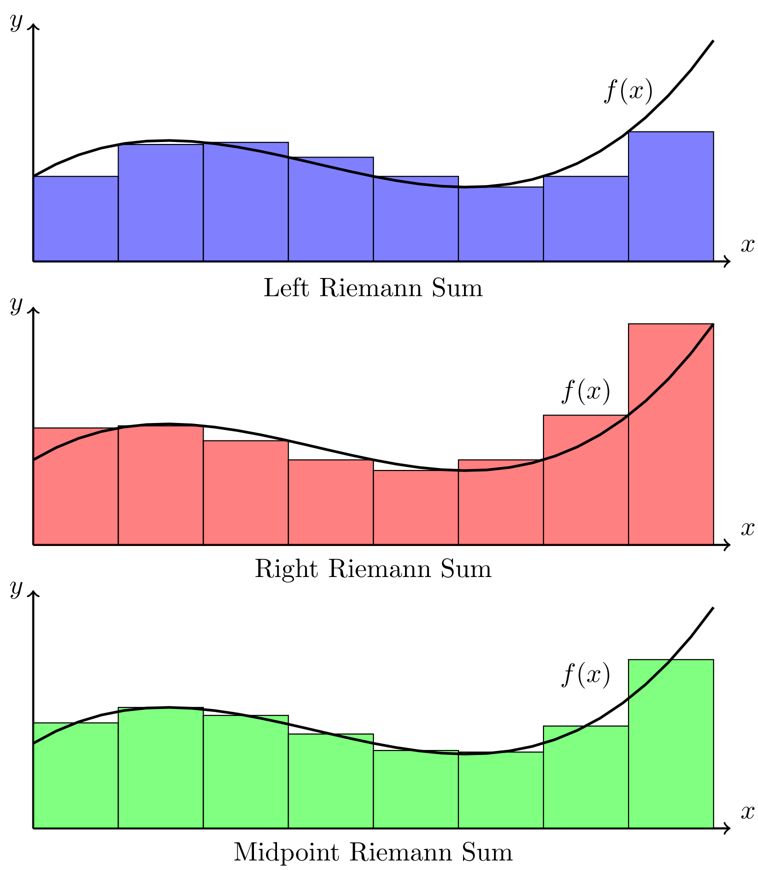
Let $f(x) = x$ on the interval $[0, 1]$. Determine the left Riemann sum for $f(x)$ using 4 subintervals. Generalize this to $n$ subintervals.

*Solution.* Breaking $[0, 1]$ into four subintervals gives

$$[0, 0.25], \quad [0.25, 0.5], \quad [0.5, 0.75], \quad [0.75, 1],$$

with $\Delta x = 0.25 = 1/4$. Taking the left-endpoints as sample points gives

$$f(0)\Delta x + f(0.25)\Delta x + f(0.5)\Delta x + f(0.75)\Delta x = 0.25 \times (0 + 0.25 + 0.5 + 0.75) = 0.375.$$




$$L(f) = \sum_{i=1}^{n-1} f\left(a + i\Delta x\right)\Delta x$$

$$R(f) = \sum_{i=1}^{n-1} f\left(a + (i+1)\Delta x\right)\Delta x$$

$$M(f) = \sum_{i=1}^{n-1} f\left(a + \left(i + \frac{1}{2}\right)\Delta x\right)\Delta x$$

Figure 43: For a function $f(x)$, an illustration of the left-, right-, and midpoint-Riemann sums.

More generally, when we have $n$ subintervals the length of each partition is $\Delta x = 1/n$ and our subintervals are of the form

$$\left[\frac{i}{n}, \frac{(i+1)}{n}\right].$$

The left Riemann sum is thus

$$L(f) = \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right)\Delta x = \frac{1}{n}\sum_{i=0}^{n-1}\frac{i}{n} = \frac{1}{n^2}\frac{n(n-1)}{2} = \frac{n-1}{2n}.$$

■

**Example 5.6**

Let $f(x) = x^2$ on $[1,4]$. Determine the right Riemann sum for $f(x)$ using three subintervals. Generalize this to $n$ subintervals.

*Solution.* Breaking $[1,4]$ into three subintervals gives

$$[1,2], \quad [2,3], \quad [3,4],$$

with $\Delta x = 1$. Thus the right Riemann sum is

$$f(2)\Delta x + f(3)\Delta x + f(4)\Delta x = 1 \times (4+9+16) = 29.$$

More generally, when we have $n$ subinterval the length of each partition is $\Delta x = (4-1)/n = 3/n$ and our subintervals are of the form

$$\left[1 + \frac{3i}{n}, 1 + \frac{3(i+1)}{n}\right].$$

Thus the right Riemann sum is

$$\begin{aligned} R(f) &= \sum_{i=0}^{n-1} f\left(1 + \frac{3(i+1)}{n}\right)\Delta x = \frac{3}{n}\sum_{i=0}^{n}\left(1 + \frac{3(i+1)}{n}\right)^2 \\ &= \frac{3}{n}\sum_{i=0}^{n-1}\left[1 + \frac{6(i+1)}{n} + \frac{9(i+1)^2}{n^2}\right]. \end{aligned} \tag{5.5}$$

This expression can be simplified further, but we won't worry about that. ■

**Exercise:** Simplify (5.5) using (5.2).

### 5.2.3 Defining the Definite Integral

We really want to know the exact area under a curve, not just an approximation. With the idea that taking a larger number of subintervals in our estimate should result in a more accurate approximation to the area, we are lead to the definition of the definite integral.

**Definition 5.7**

Let $f(x)$ be a function defined on $[a, b]$. For a positive integer $n$, let $\Delta x = (b-a)/n$ and $\hat{x}_i$ be any point in the interval $[a + i\Delta x, a + (i+1)\Delta x]$. We say that $f$ is *integrable* if

$$\int_a^b f(x)\,\mathrm{d}x := \lim_{n\to\infty} \sum_{i=0}^{n-1} f(\hat{x}_i)\Delta x$$

exists. When this limit exists, we say that $\int_a^b f(x)\,\mathrm{d}x$ is the *definite integral* of $f(x)$.

Important in this definition is that if $f(x)$ is integrable, the choice of the points $\hat{x}_i$ do not matter. Hence we may choose the left, right, midpoint, or any other Riemann sum we like to evaluate the definite integral.

While we will not prove it, any function which is continuous is integrable. In fact, any function that has finitely many discontinuities is integrable. We will assume this result henceforth.

**Example 5.8**

Determine the value of $\displaystyle\int_0^1 x\,\mathrm{d}x$. More generally, what is the value of $\displaystyle\int_a^b x\,\mathrm{d}x$?

*Solution.* We showed in Example 5.5 that the left Riemann sum for $f(x) = x$ over the interval $[0, 1]$ is given by

$$L = \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right)\Delta x = \frac{n-1}{2n}.$$

Taking the limit as $n \to \infty$ gives

$$\int_0^1 x\,\mathrm{d}x = \lim_{n\to\infty} \frac{n-1}{2n} = \frac{1}{2}.$$

More generally, since $f$ is continuous it is integrable. The left Riemann sum is

$$\begin{aligned}
L(f) &= \frac{b-a}{n}\sum_{i=0}^{n-1} f\left(a + i\left(\frac{b-a}{n}\right)\right) \\
&= \frac{b-a}{n}\sum_{i=0}^{n-1} \left[a + i\left(\frac{b-a}{n}\right)\right] \\
&= \frac{b-a}{n}\left[a\left(\sum_{i=0}^{n-1} 1\right) + \frac{b-a}{n}\left(\sum_{i=0}^{n-1} i\right)\right] \\
&= \frac{b-a}{n}\left[an + (b-a)\frac{n-1}{2}\right] \\
&= (ab - a^2) + (b-a)^2\left(\frac{n-1}{2n}\right).
\end{aligned}$$

It is easy to check that $\lim_{n\to\infty} \frac{n-1}{2n} = \frac{1}{2}$, so

$$\begin{aligned}\int_a^b x\,\mathrm{d}x &= \lim_{n\to\infty} L(f) = (ab-a^2) + (b-a)^2 \lim_{n\to\infty} \frac{2^n+1}{2^n} \\ &= ab - a^2 + \frac{1}{2}(b-a)^2 = ab - a^2 + \frac{1}{2}(b^2 - 2ab + a^2) \\ &= \frac{1}{2}(b^2 - a^2).\end{aligned}$$

This is precisely the answer we expect, since the area of a triangle is $\frac{1}{2}bh$ where $b$ is the base length and $h$ is the height (See Figure 44). ■

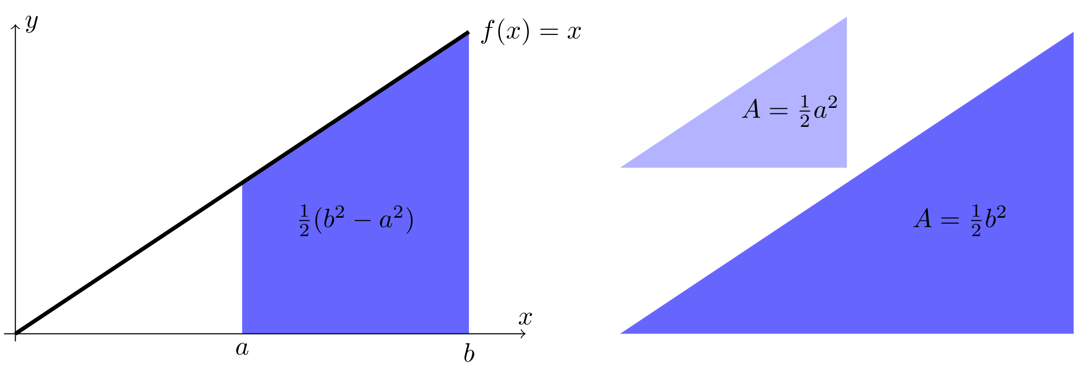



Figure 44: One can determine $\int_a^b f(x)\,\mathrm{d}x$ without Riemann sums, by recognizing that the area is just a difference of triangles.

**Theorem 5.9: Properties of the Definite Integral**

1. **Additivity of Domain:** If $f(x)$ is integrable on $[a,b]$ and $[b,c]$ then $f(x)$ is integrable on $[a,c]$ and
$$\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x.$$
2. **Additivity of Integral:** If $f(x), g(x)$ are integrable on $[a,b]$ then $f(x)+g(x)$ is integrable on $[a,b]$ and
$$\int_a^b [f(x)+g(x)]\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x.$$
3. **Scalar Multiplication:** If $f(x)$ is integrable on $[a,b]$ and $c \in \mathbb{R}$, then $cf(x)$ is integrable on $[a,b]$ and
$$\int_a^b cf(x)\,\mathrm{d}x = c\int_a^b f(x)\,\mathrm{d}x.$$
4. **Inherited Integrability:** If $f(x)$ is integrable on $[a,b]$ then $f(x)$ is integrable on any subinterval $[c,d] \subseteq [a,b]$.
5. **Monotonicity of Integral:** If $f(x), g(x)$ are integrable on $[a,b]$ and $f(x) \leq g(x)$ for all $x \in [a,b]$ then
$$\int_a^b f(x)\,\mathrm{d}x \leq \int_a^b g(x)\,\mathrm{d}x.$$
6. **Subnormality:** If $f(x)$ is integrable on $[a,b]$ then $|f(x)|$ is integrable on $[a,b]$ and satisfies
$$\left|\int_a^b f(x)\,\mathrm{d}x\right| \leq \int_a^b |f(x)|\,\mathrm{d}x.$$

## 5.3 Anti-Derivatives

Before continuing, we must take a slight detour and examine the topic of anti-differentiation. Anti-differentiation is the reverse process of differentiation; that is, if I give you a function $f(x)$ then our goal is to find a function $F(x)$ such that $F'(x) = f(x)$. To this end, we have the formal definition:

**Definition 5.10**

Given a function $f(x)$ on $[a,b]$, we say that a function $F(x)$ is an *anti-derivative* of $f$ if $F'(x) = f(x)$ for all $x \in [a,b]$.

**Example 5.11**

Compute the anti-derivative of $f(x) = 5x^4$.

*Solution.* We know that polynomials differentiate to give polynomials, so let's assume that $F(x) = x^n$ for some $n$. For $F(x)$ to be an anti-derivative of $f(x)$ it must be that $F'(x) = nx^{n-1} = f(x) = 5x^4$. It is not too hard to see that $n = 5$ works, so that the anti-derivative of $f(x) = 5x^4$ is $F(x) = x^5$. ∎

The previous example was exceptionally easy to solve because of the coefficient 5 in the monomial term. If that term had not been there, then we would just artificially add it. For example, the anti-derivative of $3x^4$ may be computed by realizing that

$$3x^4 = 3 \cdot \frac{5}{5}x^4 = \frac{3}{5}5x^4.$$

Since scalar multiples pass through derivatives, we hypothesize that the anti-derivative of $3x^4$ is $\frac{3}{5}x^5$ and a quick computation confirms this.

In fact, just using the properties of differentiation, we can immediately infer a few results about anti-derivatives. Since the derivative is linear, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}cf(x) = cf'(x), \qquad \frac{\mathrm{d}}{\mathrm{d}x}\left[f(x) + g(x)\right] = f'(x) + g'(x)$$

and this tells us that anti-differentiation will also be linear. To see this, let $F(x)$ and $G(x)$ be anti-derivatives of $f(x)$ and $g(x)$, so that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left[cF(x)\right] &= cF'(x) = cf(x) = c\frac{\mathrm{d}}{\mathrm{d}x}F(x) \\ \frac{\mathrm{d}}{\mathrm{d}x}\left[F(x) + G(x)\right] &= F'(x) + G'(x) = f(x) + g(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) + \frac{\mathrm{d}}{\mathrm{d}x}G(x). \end{aligned}$$

Note that the anti-derivative of a function is not unique, since we may add any constant to a function to find a new anti-derivative. For example, assume that $F(x)$ is an anti-derivative for $f(x)$ so that $F'(x) = f(x)$. Define a new function $F_c(x) = F(x) + c$ for any constant $c \in \mathbb{R}$. We then have that

$$\frac{\mathrm{d}}{\mathrm{d}x}F_c(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[F(x) + c\right] = F'(x) = f(x).$$

so that $F_c(x)$ is also an anti-derivative. This implies that there are an entire real number's worth of functions which are the anti-derivative of a function. More concretely, Example 5.11 shows that $x^5$ is the anti-derivative of $5x^4$, but a quick computation easily shows that $x^5 + c$ also differentiates to $5x^4$ for any constant $c$.

**Corollary 5.12**

If $f(x)$ is a function with an anti-derivative $F(x)$, then $F(x)$ is unique up to an additive constant; that is, if $\tilde{F}(x)$ is any other anti-derivative of $f(x)$, then there exists some constant $c$ such that $F(x) = \tilde{F}(x) + c$.

For reference sake, the following is a list of simple anti-derivatives where the additive constant is taken to be zero:

| Function | Anti-derivative |
|---|---|
| $x^n (n \neq -1)$ | $\dfrac{x^{n+1}}{n+1}$ |
| $\dfrac{1}{x}$ | $\log\lvert x\rvert$ |
| $e^x$ | $e^x$ |
| $\cos(x)$ | $\sin(x)$ |
| $\sin(x)$ | $-\cos(x)$ |

| Function | Anti-Derivative |
|---|---|
| $\sec^2(x)$ | $\tan(x)$ |
| $\sec(x)\tan(x)$ | $\sec(x)$ |
| $\dfrac{1}{\sqrt{1-x^2}}$ | $\arcsin(x)$ |
| $\dfrac{1}{1+x^2}$ | $\arctan(x)$ |

**Example 5.13**

Compute $f(x)$ if $f''(x) = \sqrt{x} + \sin(x) + e^x$.

*Solution.* Notice that the second derivative is given, so we will have to compute the anti-derivative twice. Here in particular it is essential to recall that anti-derivatives are only defined up to additive constants. According to our table above, we have the following derivative - anti-derivative pairs:

$$\sqrt{x} = \frac{\mathrm{d}}{\mathrm{d}x}\frac{2}{3}x^{3/2}, \qquad \sin(x) = \frac{\mathrm{d}}{\mathrm{d}x}(-\cos(x)), \qquad e^x = \frac{\mathrm{d}}{\mathrm{d}x}e^x$$

so that the first derivative (given by the anti-derivative of $f''(x)$ is

$$f'(x) = \frac{2}{3}x^{3/2} - \cos(x) + e^x + c$$

for some constant $c$. It is important to include the $c$ here since when we take another anti-derivative, it will contribute to the solution. Once again, the anti-derivatives are given by

$$x^{3/2} = \frac{\mathrm{d}}{\mathrm{d}x}\frac{2}{5}x^{5/2}, \qquad \cos(x) = \frac{\mathrm{d}}{\mathrm{d}x}\sin(x), \qquad e^x = \frac{\mathrm{d}}{\mathrm{d}x}e^x, \qquad c = \frac{\mathrm{d}}{\mathrm{d}x}cx,$$

so that $f(x)$ is

$$f(x) = \frac{4}{15}x^{5/2} - \sin(x) + e^x + cx + d$$

where $c, d$ are constants. ∎

If additional criteria are supplied, such as the value of $f(x)$ (or its derivatives) at particular points, then a truly unique solution may be identified.

**Example 5.14**

Using your solution to Example 5.13, compute the unique anti-derivative which satisfies $f(0) = 10$ and $f'(0) = 0$.

*Solution.* Our above example showed that $f'(x) = \frac{2}{3}x^{3/2} - \cos(x) + e^x + c$. By substituting $x = 0$ into this we get

$$0 = f'(0) = \frac{2}{3}0^{3/2} - \cos(0) + e^0 + c = c$$

so that $c = 0$. Thus $f(x) = \frac{4}{15}x^{5/2} - \sin(x) + e^x + d$. Substituting $x = 0$ into this gives

$$10 = f(0) = \frac{4}{15}0^{5/2} - \sin(0) + e^0 + d = 1 + d$$

so that $d = 9$. In conclusion, the corresponding $f(x)$ is

$$f(x) = \frac{4}{15}x^{5/2} - \sin(x) + e^x + 9. \qquad \blacksquare$$

Notice that Example 5.14 required two conditions to specify the number of constants. In general, if one is given the $n^{\text{th}}$ derivative of a function, one needs to specify $n$-conditions to uniquely determine the function.

## 5.4 The Fundamental Theorem of Calculus

In this section, we will make the connection between the theory of integration and the theory of differentiation, by means of the *Fundamental Theorem of Calculus*. Let $f(x)$ be an integrable function on $[a, b]$. By Theorem 5.9 we know that for any subinterval $[c, d] \subseteq [a, b]$, $f(x)$ is also integrable on $[c, d]$. In particular, let's fix the left endpoint at $a$. Now for each $x \in [a, b]$, we have an integrable function on $[a, x]$ and hence the definite integral exists and produces a number. Thus we have a function

$$F(x) = \int_a^x f(s)ds$$

which assigns to each point $x$ the value of the definite integral on $[a, x]$. Analogous to differentiation, wherein we had a function $f$ on $[a, b]$ and created a function $f'$ on $[a, b]$ with interesting properties, we now have the function $F$ on $[a, b]$, and we are interested in its properties.

**Theorem 5.15: Fundamental Theorem of Calculus**

1. If $f$ is integrable on $[a, b]$ then $F(x) = \int_a^x f(s)ds$ is continuous on $[a, b]$. Moreover, $F$ is differentiable at any point where $f$ is continuous, and in this case $F$ is an anti-derivative of $f$.

2. If $f$ is integrable on $[a, b]$, and $F$ be a continuous anti-derivative of $f$ which is differentiable at all but finitely many points, then

$$\int_a^b f(s)\,\mathrm{d}s = F(b) - F(a). \tag{5.6}$$

**Remark 5.16**

1. Effectively, the fundamental theorem of calculus indicates that differentiation and integration are 'inverses' of one another. This is not exactly true, as Example 5.20 demonstrates:

2. The choice of anti-derivative $F$ in Theorem 5.15(2) does not matter. If $\tilde{F}$ is another anti-derivative $f$ then by Corollary 5.12 there exists some real number $C$ such that $F(x) = \tilde{F}(x) + C$. Substituting this into (5.6) yields

$$\int_a^b f(s)\,\mathrm{d}s = F(b) - F(a) = \left[\tilde{F}(b) + C\right] - \left[\tilde{F}(a) + C\right] = \tilde{F}(b) - \tilde{F}(a).$$

3. The lower bound of integration does not matter, so long as the function stays integrable on $[a, x]$. Indeed, if $c$ is any other point such that $f(x)$ is integrable on $[c, x]$ then

$$\int_c^x f(s)\,\mathrm{d}s = \int_a^c f(s)\,\mathrm{d}s + \int_a^x f(s)\,\mathrm{d}s = \int_a^x f(s)\,\mathrm{d}s + C$$

where $C$ is the value of the integral on $[a, c]$. Hence $\int_c^x f(s)\,\mathrm{d}s$ only differs from $F(x)$ by an additive constant, and hence is an anti-derivative as well.

**Example 5.17**

Verify Example 5.8; that is, show that $\displaystyle\int_a^b x\,\mathrm{d}x = \frac{1}{2}(b^2 - a^2)$.

*Solution.* It suffices to find an anti-derivative of the function $f(x) = x$. The reader can quickly check that $F(x) = \frac{1}{2}x^2$ satisfies this requirement, so by the Fundamental Theorem of Calculus:

$$\int_a^b x\,\mathrm{d}x = F(b) - F(a) = \frac{1}{2}b^2 - \frac{1}{2}a^2 = \frac{1}{2}\left(b^2 - a^2\right),$$

precisely as shown in Example 5.8. ■

**Example 5.18**

Compute $\displaystyle\int_{-\pi}^{\pi} \sin(t)\,\mathrm{d}t$, and $\displaystyle\int_0^1 \frac{1}{1+x^2}\,\mathrm{d}x$.

*Solution.* We know that $F(x) = -\cos(x)$ is an anti-derivative of $\sin(t)$, so by the Fundamental Theorem of Calculus we have

$$\int_{-\pi}^{\pi} \sin(t)\,\mathrm{d}t = F(\pi) - F(-\pi) = -\cos(\pi) - (-\cos(-\pi)) = 0.$$

Similarly, we know that $G(x) = \arctan(x)$ is an anti-derivative of $\frac{1}{1+x^2}$, and hence

$$\int_{-1}^{0} \frac{1}{1+x^2}\,\mathrm{d}x = \arctan(0) - \arctan(-1) = -\frac{\pi}{4}.$$ ■

**Example 5.19**

Determine the value $\displaystyle\int_1^4 \frac{x+x^3}{x^4}\,\mathrm{d}x$.

*Solution.* By expanding the integrand and using linearity of the definite integral, we get

$$\int_1^4 \frac{x+x^3}{x^4}\,\mathrm{d}x = \int_1^4 \left[\frac{1}{x^3} + \frac{1}{x}\right]\mathrm{d}x = \int_1^4 \frac{1}{x^3}\,\mathrm{d}x + \int_1^4 \frac{1}{x}\,\mathrm{d}x.$$

The function $f(x) = x^{-3}$ has an anti-derivative $F(x) = -\frac{1}{2}x^{-2}$, while $g(x) = x^{-1}$ has an anti-derivative $G(x) = \log(x)$. By the Fundamental Theorem of Calculus, we thus have

$$\begin{aligned}
\int_1^4 \frac{x+x^3}{x^4}\,\mathrm{d}x &= \int_1^4 \frac{1}{x^3}\,\mathrm{d}x + \int_1^4 \frac{1}{x}\,\mathrm{d}x \\
&= [F(4) - F(1)] + [G(4) - G(1)] \\
&= -\frac{1}{2}\left[\frac{1}{16} - 1\right] + [\log(4) - \log(1)] \\
&= \frac{15}{32} + \log(4).
\end{aligned}$$ ■

**Example 5.20**

Let $f$ be a continuous function on $\mathbb{R}$. Evaluate

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_0^x f(t)\,\mathrm{d}t - \int_0^x \frac{\mathrm{d}}{\mathrm{d}t} f(t)\,\mathrm{d}t.$$

*Solution.* If integration and differentiation were truly inverses, then this would simply evaluate to zero. However, let us be a bit more prudent in our evaluation. By the Fundamental Theorem of Calculus, $F(x) = \int_0^x f(t)\,\mathrm{d}t$ is an anti-derivative of $f(x)$, and hence

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_0^x f(t)\,\mathrm{d}t = f(x).$$

On the other hand, $f(x)$ is clearly an anti-derivative of $f'(x)$, and so

$$\int_0^x \frac{\mathrm{d}}{\mathrm{d}t} f(t)\,\mathrm{d}t = \int_0^x f'(t)\,\mathrm{d}t = f(x) - f(0).$$

Hence the difference between these two terms comes out to $f(0)$; that is, they differ up to a constant. ■

**Example 5.21**

Determine $G'(x)$ if $G(x) = \int_{-1}^{e^x} \cos(t^2)\,\mathrm{d}t$.

*Solution.* The reader should stare at this equation until he/she is convinced that this is indeed a function of $x$. To proceed, define a function

$$F(x) = \int_{-1}^{x} \cos(t^2)\,\mathrm{d}t$$

which, according to the Fundamental Theorem of Calculus, is an anti-derivative of the function $f(x) = \cos(x^2)$. We can write $G(x)$ in terms of $F(x)$, since

$$G(x) = \int_{-1}^{e^x} \cos(t^2)\,\mathrm{d}t = F(e^x).$$

We can thus differentiate $G(x)$ using the Chain Rule:

$$G'(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(e^x) = F'(e^x)e^x = f(e^x)e^x = \cos(e^{2x})e^x. \qquad \blacksquare$$

### 5.4.1 Indefinite Integrals

We have seen that integration and differentiation are spiritual inverses of one another, up to an additive constant. In particular, for a function $f(x)$ on $[a, b]$ we saw that $F(x) = \int_a^x f(x)dx$ is an anti-derivative of $f$, and for any anti-derivative $G$ of $f$ we have

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

Anti-derivatives are unique up to constants, so there exists some constant $C$ such that $F = G + C$, with $F$ playing a particularly nice representative. However, the constant seems rather artificial: we know that the anti-derivative of $x^3$ is $\frac{1}{4}x^4 + C$, but the meat-and-bones lies with the $\frac{1}{4}x^4$ term, not the constant. Hence our goal for this section is to represent the entire class of anti-derivatives, something called the *indefinite integral*.

The indefinite integral does not concern itself with upper and lower bounds of integration: our goal is to represent an entire class of functions; imposing bounds forces us to look at particular representatives. Consequently, we denote the indefinite integral with the usual integral sign, albeit with the bounds omitted:

$$\int f(x)\,\mathrm{d}x.$$

Remember, this notation means *the entire set of anti-derivatives*.

**Example 5.22**

Determine the following indefinite integrals:

1. $\displaystyle\int \left(\frac{x^4+2x^2+1}{x^3}\right) dx$,
2. $\displaystyle\int \sin(2x)\cos(2x) dx$,
3. $\displaystyle\int f(x)f'(x) dx$, where $f(x)$ is differentiable.

*Solution.* In time, we will learn more systematic ways of determining these integrals, but for now we will need to use the clever part of our brains to find appropriate classes of anti-derivatives.

1. Notice that we can re-write the integrand as
$$\frac{x^4+2x^2+1}{x^3} = x + \frac{2}{x} + \frac{1}{x^3}.$$
We are well acquainted with the functions which yield these as derivatives, and we get
$$\int \frac{x^4+2x^2+1}{x^3} = \int \left(x + \frac{2}{x} + \frac{1}{x^3}\right) dx = \frac{1}{2}x^2 + 2\log(x) - \frac{1}{2x^2} + C.$$

2. The important step here is to realize that $\frac{1}{2}\sin(4x) = \sin(2x)\cos(2x)$, hence our integral becomes
$$\int \sin(2x)\cos(2x) dx = \frac{1}{2}\int \sin(4x) dx = -\frac{1}{8}\cos(4x) + C.$$

3. This problem is a little more abstract: We need to find a function which differentiates to $f(x)f'(x)$. If we think hard, we see that $\frac{\mathrm{d}}{\mathrm{d}x}[f(x)]^2 = 2f(x)f'(x)$, so by dividing by 2 we will get the desired integrand. Applying the Fundamental Theorem of Calculus, we thus get
$$\int f(x)f'(x) dx = \int \frac{\mathrm{d}}{\mathrm{d}x} f(x)^2 dx = f(x)^2 + C. \tag{5.7}$$
We will see more on how to solve integral like this in Section 6.1. ■

### 5.4.2 Integral Notation

We have been dramatically overloading our use of the integral sign. We have thus far seen three different objects: If $f$ is integrable on $[a,b]$ then

$$\int_a^b f(s)\,\mathrm{d}s, \qquad F(x) = \int_a^x f(s)\,\mathrm{d}s, \qquad \int f(s)\,\mathrm{d}s.$$

The first is simply a *number* which represents the signed area under the function $f$; the second is a *function* which assigns to each $x$ the area under the function from $a$ to $x$; the third is an infinite *family of functions*, all representing anti-derivatives of $f$. They are all intimately related to be certain, but each has a very different lifestyle. One must be careful not to confuse the relationships.

Additionally, some authors prefer to the use the notation

$$\int \mathrm{d}s f(s) \quad \text{instead of} \quad \int f(s)\,\mathrm{d}s.$$

There are occasions when this is useful, but we will never use it in this course.

# 6 Integration Techniques

In the following sections, we will develop a plethora of tools to help us compute integrals.

## 6.1 Integration by Substitution

Having seen that integration and differentiation are essentially inverses, we would like to develop some techniques and rules form computing integrals. It should be unsurprising that those rules will arise as the "inverse" operations of the rules obtained from differential calculus. We recall the chain rule of differential calculus tells us that

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

hence by applying the fundamental theorem of calculus, we see that

$$\int f'(g(x))g'(x)\,\mathrm{d}x = f(g(x)) + C. \tag{6.1}$$

Unfortunately, the majority of times nature will conspire against us and not write our integrand so plainly as $f'(g(x))g'(x)$. Hence we develop some techniques to make life simpler. Our strategy is as follows:

1. Look to see if we can find the occurrence of a function and its derivative. In (6.1) above, we are looking for the function $g(x)$, since it occurs in the argument of $f(x)$ and its derivative appears as $g'(x)$.
2. Define a new variable, $u = g(x)$ so that $\frac{\mathrm{d}u}{\mathrm{d}x} = g'(x)$. We often write this second equation as $\mathrm{d}u = g'(x)\,\mathrm{d}x$.
3. Replace all $x$ dependencies with $u$ dependencies. Namely, recognize that

$$\underbrace{f'(g(x))}_{f(u)}\underbrace{g'(x)\,\mathrm{d}x}_{\mathrm{d}u} = f'(u)\,\mathrm{d}u$$

4. Using the fundamental theorem of calculus, evaluate our new integral:

$$\int f'(u)\,\mathrm{d}u = f(u) + C.$$

5. We now have our solution, but it is in terms of the variable $u$. This is not a problem since we know that $u = g(x)$, so we just make this substitution to get our final solution

$$\int f'(g(x))g'(x)\,\mathrm{d}x = f(g(x)) + C.$$

Let's try a simple example:

**Example 6.1**

Determine $\int \sin(2x)\cos(2x)\,\mathrm{d}x$.

*Solution.* We can choose either $u = \sin(2x)$ or $u = \cos(2x)$. If $u = \sin(2x)$ then $\mathrm{d}u = 2\cos(2x)\,\mathrm{d}x$ and our integral becomes

$$\int \sin(2x)\cos(2x)\,\mathrm{d}x = \frac{1}{2}\int u\,\mathrm{d}u = \frac{1}{4}u^2 + C = \frac{1}{2}\sin^2(2x) + C.$$

On the other hand, if we choose $u = \cos(2x)$ then $\mathrm{d}u = -\frac{1}{2}\sin(2x)$ and we get

$$\int \sin(2x)\cos(2x)\,\mathrm{d}x = -\frac{1}{2}\int u\,\mathrm{d}u = -\frac{1}{4}\cos^2(2x) + C.$$

Of course, we know that $\sin^2(2x)$ and $\cos^2(2x)$ are very different functions, and here we see the importance of the constants.

To add to this bit of noise, notice that in Example 5.22 we calculated $\int \sin(x)\cos(x) = -\frac{1}{8}\cos(4x) + C$. ■

**Example 6.2**

For $a, b \neq 0$, compute $\int \frac{x^{n-1}}{\sqrt{a + bx^n}}\,\mathrm{d}x$.

*Solution.* Following the above program, our first step should be to identify a function and its derivative. The fact that there is an $x^n$ and an $x^{n-1}$ is a pretty good sign. Since constants do not affect the integration, we can make our lives *even easier* if we define $u = a + bx^n$ so that $\mathrm{d}u = bnx^{n-1}\,\mathrm{d}x$. Unfortunately, there is no $bnx^{n-1}\,\mathrm{d}x$ in the integrand, but there is an $x^{n-1}\,\mathrm{d}x$. Since these are related only up to a constant, we can divide both sides to find that $x^{n-1}\,\mathrm{d}x = \frac{1}{bn}\,\mathrm{d}u$. Adding our substitutions we then get

$$\int \frac{x^{n-1}}{\sqrt{a + bx^n}}\,\mathrm{d}x = \int \frac{\frac{1}{bn}\,\mathrm{d}u}{\sqrt{u}} = \frac{1}{bn}\int \frac{1}{\sqrt{u}}\,\mathrm{d}u.$$

This is now a very simple integral to calculate, and indeed we find that

$$\frac{1}{bn}\int \frac{1}{\sqrt{u}}\,\mathrm{d}u = \frac{2}{bn}\sqrt{u} + C.$$

We need this to be in terms of $x$ rather than $u$, so we recall that $u = a + bx^n$ to finally find that

$$\int \frac{x^{n-1}}{\sqrt{a+bx^n}}\,\mathrm{d}x = \frac{2}{bn}\sqrt{a+bx^n} + C. \qquad \blacksquare$$

Substitution is not just handy for applying the chain rule. It also allows us to "change variables."

**Example 6.3**

Compute $\displaystyle\int x\sqrt{x+1}\,\mathrm{d}x$.

*Solution.* Notice that if we could somehow switch the $x$ and the $x+1$, this integral would be much simpler, since then $(x+1)\sqrt{x} = x^{3/2} + x$. Normally in mathematics, if we want to do such a thing, we just define a new variable $u = x+1$ so that $x = u-1$ and then a similar trick to the one above will work.

Since we are working with an integral though, we must be a bit more careful. We shall still define $u = x+1$ with $x = u-1$, but we must also track the differentials. Luckily, in this case $\mathrm{d}u = \mathrm{d}x$ and there is nothing to do. We thus get

$$\begin{aligned}\int x\sqrt{x+1}\,\mathrm{d}x &= \int (u-1)\sqrt{u}\,\mathrm{d}u \\ &= \int (u^{3/2} - \sqrt{u})\,\mathrm{d}u \\ &= \frac{2}{5}u^{5/2} - \frac{2}{3}u^{3/2} + C.\end{aligned}$$

Converting back to a function of $x$ yields

$$\int x\sqrt{x+1}\,\mathrm{d}x = \frac{2}{5}(x+1)^{5/2} - \frac{1}{2}(x+1)^2 + C. \qquad \blacksquare$$

### 6.1.1 Definite Integrals

When dealing with definite integrals, we adhere to the same process as indefinite integrals, but we must also accommodate the lower and upper bounds of integration. Let $f$ be a continuous function with anti-derivative $F$, while $g$ is continuously differentiable. Once again consider the case when we are integrating the function

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x.$$

We know that $F(g(x))$ is an integrable anti-derivative, so the Fundamental Theorem of Calculus implies that

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x = [F(g(x))]_a^b = F(g(b)) - F(g(a)).$$

This implies that the correct lower and upper bounds of integration are $g(a)$ and $g(b)$ respectively, since

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x = \int_{g(a)}^{g(b)} f(u)\,\mathrm{d}u = [F(u)]_{g(a)}^{g(b)} = F(g(b)) - F(g(a))$$

gives the correct solution.

For a different perspective, the lower and upper bounds of integration say that our variable $x$ is moving between the values $x = a$ and $x = b$. If we make the substitution $u = g(x)$, then our new variable is $u$. As $x$ goes from $a$ to $b$, then $u$ goes between $g(a)$ and $g(b)$, and so our integral becomes

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

As an example, consider what happens if we **fail to change the bounds of integration.** Consider the integral

$$\int_0^1 \cos\left(\frac{\pi}{2}x\right)\,\mathrm{d}x.$$

One can easily see compute this as

$$\int_0^1 \cos\left(\frac{\pi}{2}x\right)\,\mathrm{d}x = \frac{2}{\pi}\sin\left(\frac{\pi}{2}x\right)\Big|_0^1 = \frac{2}{\pi}.$$

Alternatively though, one should recognize that integrating $\cos\left(\frac{\pi}{2}\right)$ from 0 to 1 is the same as integrating $\cos(x)$ from 0 to $\frac{\pi}{2}$. How does this reflect in the integral if we make such a substitution? Let $u = \frac{\pi}{2}x$ so that $\mathrm{d}u = \frac{\pi}{2}\,\mathrm{d}x$. Were we to naively carry though with the integration without changing the bounds, we would get

$$\int_0^1 \cos\left(\frac{\pi}{2}x\right)\,\mathrm{d}x = \frac{2}{\pi}\int_0^1 \cos(x)\,\mathrm{d}x = \frac{2}{\pi}\sin(x)\Big|_0^1 = \frac{2}{\pi}\sin(1).$$

This is a different answer! The reason can be seen in the first equality above. We said that integrating $\cos\left(\frac{\pi}{2}x\right)$ should be the same as integrating $\cos(x)$ from 0 to $\frac{\pi}{2}$, but we failed to change the upper and lower limits of integration. To do this correct, we think of the upper and lower limits as corresponding to $x_L = 0$ and $x_U = 1$. Having made the substitution to $u$, we need to find the corresponding $u_L$ and $u_U$. Since $u$ is just a function of $x$, we then get that $u_L = \frac{\pi}{2}x_L = 0$ and $u_U = \frac{\pi}{2}x_U = \frac{\pi}{2}$. Now if we do our work, we get the correct answer.

**Example 6.4**

Determine the value of $\displaystyle\int_0^1 x\sqrt{2+x^2}\,\mathrm{d}x$.

*Solution.* We will proceed using the substitution $u = 2 + x^2$ so that $\mathrm{d}u = 2x\,\mathrm{d}x$. When $x = 0$ we have $u = 2$, while when $x = 1$ we get $u = 3$, so

$$\begin{aligned}\int_0^1 x\sqrt{2+x^2}\,\mathrm{d}x &= \frac{1}{2}\int_2^3 \sqrt{u}\,\mathrm{d}u \\ &= \frac{1}{3}u^{3/2}\Big|_{u=2}^{u=3} \\ &= \frac{3\sqrt{3}-2\sqrt{2}}{3}.\end{aligned}$$ ■

Note that when performing the definite integral, we do not need to convert back to the $x$-representation, since our upper and lower bounds have already accommodated for that change.

**Example 6.5**

Determine the integral $\displaystyle\int_2^4 \frac{1}{x\log(2x)}\,\mathrm{d}x$.

*Solution.* We will proceed using the substitution $u = \log(2x)$ so that $\mathrm{d}u = x^{-1}\,\mathrm{d}x$. When $x = 2$ we have $u = \log(4)$, while when $x = 4$ we have $u = \log(8)$, so that

$$\begin{aligned}\int_2^4 \frac{1}{x\log(2x)}\,\mathrm{d}x &= \int_{\log(4)}^{\log(8)} \frac{1}{u}\,\mathrm{d}u \\ &= \log(u)\Big|_{\log(4)}^{\log(8)} \\ &= \log(\log(8)) - \log(\log(4)).\end{aligned}$$

■

## 6.2 Integration by Parts

Just as Integration by Substitution was the inverse of the chain rule, Integration by Parts is the analog of the product rule. Namely, we know that if $u(x)$ and $v(x)$ are functions, then

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[u(x)\cdot v(x)\right] = u'(x)v(x) + u(x)v'(x).$$

Integrating and applying the Fundamental Theorem of Calculus, we then find that

$$\begin{aligned}\int \frac{\mathrm{d}}{\mathrm{d}x}(u\cdot v)\,\mathrm{d}x &= u(x)\cdot v(x) \\ &= \int v\frac{\mathrm{d}u}{\mathrm{d}x}\,\mathrm{d}x + \int u\frac{\mathrm{d}v}{\mathrm{d}x}\,\mathrm{d}x \\ &= \int v\,\mathrm{d}u + \int u\,\mathrm{d}v\end{aligned}$$

which we may re-arrange to find

$$\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u. \tag{6.2}$$

In the event of the definite integral, the bounds of integration can be carried throughout; that is,

$$\int_a^b u\,\mathrm{d}v = uv\Big|_a^b - \int_a^b v\,\mathrm{d}u.$$

Our strategy should be as follows: Assume we are integrating a product $\int f(x)g(x)\,\mathrm{d}x$. We want to choose a candidate for $\mathrm{d}v$ and for $u$. Since we will need to integrate $\mathrm{d}v$, it is often best to choose a function which is easy to integrate:

1. First look at the integrand and see if we can apply substitution. If so, do not worry about integration by parts.
2. Choose $\mathrm{d}v$ and $u$ (I often choose $\mathrm{d}v$ to be whichever function is easiest to integrate),
3. Compute $v$ by integrating $\int \mathrm{d}v = \int f(x)\,\mathrm{d}x$. Compute $\mathrm{d}u$ by differentiating $u$.
4. Substitute all appropriate variables into (6.2).

This is just a general idea of how you should proceed. To give some insight as to what is happening, consider Equation (6.2) by omitting the $uv$-term:

$$\int u(x)v'(x)\,\mathrm{d}x = -\int u'(x)v(x)\,\mathrm{d}x.$$

The is the power of integration by parts: It effectively allows us to transfer the derivative from one function to another!

**Example 6.6**

Evaluate the integral $\displaystyle\int x\sin(x)\,\mathrm{d}x$.

*Solution.* Following our program, I personally find that $\sin(x)$ is easier to integrate than just $x$ so we set $\mathrm{d}v = \sin(x)\,\mathrm{d}x$. Furthermore, my note above suggests that we should set $u = x$ so everything works out. Computing $\mathrm{d}u$ and $v$ we find that

$$\begin{aligned} u &= x & \mathrm{d}v &= \sin(x)\,\mathrm{d}x \\ \mathrm{d}u &= \mathrm{d}x & v &= -\cos(x). \end{aligned}$$

Plugging these into (6.2) we find that

$$\begin{aligned} \int x\sin(x)\,\mathrm{d}x &= -x\cos(x) + \int \cos(x)\,\mathrm{d}x \\ &= -x\cos(x) + \sin(x) + C. \end{aligned}$$

The best thing about integration is that you can always check your answers by differentiating. Try it! ∎

**Example 6.7**

Determine $\displaystyle\int_0^3 xe^x\,\mathrm{d}x$.

*Solution.* We will take as our integration by parts:

$$\begin{aligned} u &= x & \mathrm{d}v &= e^x\,\mathrm{d}x \\ \mathrm{d}u &= \mathrm{d}x & v &= e^x \end{aligned}$$

which gives us the integral

$$\begin{aligned}\int_0^3 xe^x \,\mathrm{d}x &= xe^x\Big|_0^3 - \int_0^3 e^x \,\mathrm{d}x \\ &= 3e^3 - \Big[e^x\Big]_0^3 \\ &= 2e^3 + 1.\end{aligned}$$

■

Alternatively, there are time when the integrand does not look like a product, but we may still apply integration by parts.

**Example 6.8**

Compute $\displaystyle\int \log(x)\,\mathrm{d}x$.

*Solution.* Looking at the integrand, there do not immediately appear to be two functions, so how can we apply integration by parts? The solution is to realize that $\log(x) = \log(x)\cdot 1$, so that the constant function $1(x) = 1$ is actually our second function. I think that 1 is really easy to integrate, so let us set $\mathrm{d}v = 1\cdot \mathrm{d}x$ and $u = \log(x)$ to find that

$$\begin{array}{ll} u = \log(x) & \mathrm{d}v = 1\cdot \mathrm{d}x \\ \mathrm{d}u = \frac{\mathrm{d}x}{x} & v = x \end{array}$$

Substituting these values into (6.2) we find

$$\begin{aligned}\int \log(x)\,\mathrm{d}x &= x\log(x) - \int \frac{x}{x}\,\mathrm{d}x \\ &= x\log x - x + C.\end{aligned}$$

Again, try differentiating this to ensure that it works! ■

**Remark 6.9** Why can we use Integration by Parts on $\int \log(x)\,\mathrm{d}x$, and to what other functions does this same trick apply? As an exercise, the student can use Integration by Parts to show that if $f$ is an invertible, integrable function with anti-derivative $F$ and inverse $f^{-1}$, then

$$\int f^{-1}(x)\,\mathrm{d}x = xf^{-1}(x) - F(f^{-1}(x)) + C.$$

Thus invertible, integrable functions can be integrated using integration by parts.

Let's conclude with an example that will require all of our skills thus far.

**Example 6.10**

Compute the integral $\displaystyle\int \cos(\log x)\,\mathrm{d}x$.

*Solution.* It is not clear that either integration by parts or substitution will actually work. Nonetheless, a change of variables is the appropriate thing to do, as it will let us "pull" the function $\log(x)$ outside of the $\cos(x)$ as follows.

Let $y = \log(x)$ so that $\mathrm{d}y = \frac{\mathrm{d}x}{x}$. Alternatively, we can write $\mathrm{d}x = x\,\mathrm{d}y$. This is silly though, since if we were to substitute this into our integral we would have variables in both $x$ and $y$ and that would be horrible. Instead, realize that since $y = \log x$ then $x = e^y$ and so $\mathrm{d}x = e^y\,\mathrm{d}y$. Putting this all together, we have

$$\int \cos(\log x)\,\mathrm{d}x = \int e^y \cos(y)\,\mathrm{d}y.$$

Now we can proceed via integration by parts. It seems to be that $e^y$ is easiest to integrate, so let us set $\mathrm{d}v = e^y\,\mathrm{d}y$ and $u = \cos(y)$ yielding

$$\begin{array}{ll} u = \cos(y) & \mathrm{d}v = e^y\,\mathrm{d}y \\ \mathrm{d}u = -\sin(y)\,\mathrm{d}y & v = e^y. \end{array}$$

Plugging these into (6.2) we get

$$\int e^y \cos(y)\,\mathrm{d}y = e^y \cos(y) + \int e^y \sin(y)\,\mathrm{d}y. \tag{6.3}$$

Looking at the integral $\int e^y \sin(y)\,\mathrm{d}y$, we have failed to simplify the integral. However, let's just see what happens if we integrate by parts once more. Set $u = \sin(y)$ and $\mathrm{d}v = e^y$ so that

$$\begin{array}{ll} u = \sin(y) & \mathrm{d}v = e^y\,\mathrm{d}y \\ \mathrm{d}u = \cos(y)\,\mathrm{d}y & v = e^y. \end{array}$$

Equation (6.2) then tells us that

$$\int e^y \sin(y)\,\mathrm{d}y = e^y \sin(y) - \int e^y \cos(y)\,\mathrm{d}y. \tag{6.4}$$

Putting (6.3) and (6.4) together we have

$$\begin{aligned} \int e^y \cos(y)\,\mathrm{d}y &= e^y \cos(y) + \int e^y \sin(y) \\ &= e^y \cos(y) + e^y \sin(y) - \int e^y \cos(y)\,\mathrm{d}y. \\ 2\int e^y \cos(y)\,\mathrm{d}y &= e^y \Big(\cos(y) + \sin(y)\Big) && \text{adding } \int e^y \cos(y) \text{ to both sides} \\ \int e^y \cos(y)\,\mathrm{d}y &= \frac{1}{2} e^y \Big(\cos(y) + \sin(y)\Big) && \text{dividing by 2.} \end{aligned}$$

Now we have our solution in terms of $y$, but need to revert to a solution in terms of $x$. Since $y = \log(x)$ we conclude

$$\int \cos(\log x)\,\mathrm{d}x = \frac{1}{2} e^{\log x} \Big(\cos(\log x) + \sin(\log x)\Big) + C = \frac{x}{2}\Big(\cos(\log x) + \sin(\log x)\Big) + C. \qquad \blacksquare$$

## 6.3 Integrating Trigonometric Functions

Integrating trigonometric functions often comes down to a combination of using the Pythagorean identities, or angle sum identities. Notice very conveniently that the three Pythagorean identities are

- $\sin^2(x) + \cos^2(x) = 1$, and $\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x)$, $\frac{\mathrm{d}}{\mathrm{d}x}\cos(x) = -\sin(x)$,
- $\tan^2(x) + 1 = \sec^2(x)$, and $\frac{\mathrm{d}}{\mathrm{d}x}\tan(x) = \sec^2(x)$, $\frac{\mathrm{d}}{\mathrm{d}x}\sec(x) = \sec(x)\tan(x)$
- $1 + \cot^2(x) = \csc^2(x)$, and $\frac{\mathrm{d}}{\mathrm{d}x}\cot(x) = -\csc^2(x)$, $\frac{\mathrm{d}}{\mathrm{d}x}\csc(x) = -\csc(x)\cot(x)$.

Hence each identity only contains the components of a function and its derivatives. This will prove to be an exceptionally useful tool shortly.

The other trick to remember is the following two identities:

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

We can immediately use the above to equations to see that

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

Now what happens if our integrand is made up a several different trigonometric functions? For example, how would we deal with something of the form

$$\int \sin^2(x)\cos^3(x)\,\mathrm{d}x?$$

The idea is to use the Pythagorean identities listed above to reduce the integrand to an expression of the form $g(f(x))f'(x)\,\mathrm{d}x$, since from here one can easily apply the substitution method: indeed,

$$\begin{aligned}
\int g(f(x))f'(x)\,\mathrm{d}x &= \int g(u)\,\mathrm{d}u \qquad\qquad \begin{array}{c} u = f(x) \\ \mathrm{d}u = f'(x)\,\mathrm{d}x \end{array} \\
&= G(f(x)) + C
\end{aligned}$$

for some anti-derivative $G$ of $g$.

Returning to our example, $\int \sin^2(x)\cos^3(x)\,\mathrm{d}x$, we see that by writing

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

our integral becomes

$$\int \sin^2(x)\cos^3(x)\,\mathrm{d}x = \int \sin^2(x)(1-\sin^2(x))\cos(x)\,\mathrm{d}x.$$

Our integrand now consists entire of sine functions, and a single cosine function which will serve as a substitution. Setting $u = \sin(x)$ so that $\mathrm{d}u = \cos(x)\,\mathrm{d}x$ we have

$$\begin{aligned}\int \sin^2(x)\cos^3(x)\,\mathrm{d}x &= \int \sin^2(x)[1-\sin^2(x)]\cos(x)\,\mathrm{d}x \\ &= \int u^2(1-u^2)\,\mathrm{d}u \\ &= \frac{1}{3}u^3 - \frac{1}{5}u^5 + C \\ &= \frac{1}{3}\sin^3(x) - \frac{1}{5}\sin^5(x) + C.\end{aligned}$$

**Example 6.11**

Determine the integral $\int \sec^4(x)\tan^5(x)\,\mathrm{d}x$.

*Solution.* One again we conveniently have only secant and tangent functions, which are intimately related to one another through their derivatives. We have to now determine which function we are going to use as a substitution. If we make the substitution $u = \tan(x)$ then $\mathrm{d}u = \sec^2(x)\,\mathrm{d}x$ and our integral becomes

$$\int \sec^4(x)\tan^5(x)\,\mathrm{d}x = \int \sec^3(x)\tan^4(x)\,\mathrm{d}u = \int \sec^3(x)u^4\,\mathrm{d}u.$$

But now we are in trouble, since there is no obvious way to turn $\sec^3(x)$ into a function involving only $\tan(x)$.

Let's try the other substitution. If we make the $u = \sec(x)$ substitution, we will have $\mathrm{d}u = \sec(x)\tan(x)$ and so our integral will become

$$\int \sec^4(x)\tan^5(x)\,\mathrm{d}x = \int \sec^3(x)\tan^4(x)\,\mathrm{d}u = \int \tan^4(x)u^3\,\mathrm{d}u.$$

This is perfect, since $\tan^4(x) = [\tan^2(x)]^2 = [\sec^2(x)-1]^2$. Hence we can compute our integral as

$$\begin{aligned}\int \sec^4(x)\tan^5(x)\,\mathrm{d}x &= \int \left[u^2-1\right]^2 u^3\,\mathrm{d}u \\ &= \frac{1}{8}u^8 - \frac{1}{3}u^6 + \frac{1}{4}u^4 + C \\ &= \frac{1}{8}\sec^8(x) - \frac{1}{3}\sec^6(x) + \frac{1}{4}\sec^4(x) + C.\end{aligned}$$

■

The previous two example really seem to indicate the importance that the powers of one of the functions have odd degree. In the event that everything is of even degree, we once again have to use our identities on relating $\sin^2(x)$ and $\cos^2(x)$ to $\cos(2x)$.

**Example 6.12**

Determine the integral $\int \cos^4(x)\,\mathrm{d}x$.

*Solution.* We know that $\cos^2(x) = \frac{1}{2}(1+\cos(2x))$ and so our integral becomes

$$\begin{aligned}
\int \cos^4(x)\,\mathrm{d}x &= \frac{1}{4}\int [1+\cos(2x)]\,[1+\cos(2x)]\,\mathrm{d}x \\
&= \frac{1}{4}\int \left[1 + 2\cos(2x) + \cos^2(2x)\right]\,\mathrm{d}x \\
&= \frac{1}{4}[x + \sin(2x)] + \frac{1}{4}\int \cos^2(2x)\,\mathrm{d}x \\
&= \frac{1}{4}x + \frac{1}{4}\sin(2x) + \frac{1}{8}\int [1+\cos(4x)]\,\mathrm{d}x \\
&= \frac{3}{8}x + \frac{1}{4}\sin(2x) + \frac{1}{32}\sin(4x) + C.
\end{aligned}$$

■

**Example 6.13**

Determine the integral $\int_0^{\pi/2} \sin^3(x)\,\mathrm{d}x$.

*Solution.* We have three sines, so we will keep one and turn to the other two into cosines.

$$\sin^3(x) = \sin(x)\sin^2(x) = \sin(x)\left[1-\cos^2(x)\right].$$

Our integral is thus

$$\begin{aligned}
\int_0^{\pi/2} \sin^3(x)\,\mathrm{d}x &= \int_0^{\pi/2} \sin(x)\left[1-\cos^2(x)\right]\,\mathrm{d}x \\
&= -\int_1^0 (1-u^2)\,\mathrm{d}u && \begin{aligned} u &= \cos(x) \\ \mathrm{d}u &= -\sin(x)\,\mathrm{d}x \end{aligned} \\
&= \left[u - \frac{1}{3}u^3\right]_0^1 \\
&= \frac{2}{3}.
\end{aligned}$$

■

One could enumerate a rather large list of all the possible cases and the tricks to use, but this is a rather large waste of time. Instead, the student should just realize that the fundamental idea is to apply trigonometric identities in a way that isolates a single occurrence of a function's derivative and then apply substitution.

## 6.4 Trigonometric Substitution

The observant student will have recognized something very curious happened when we were computing the derivatives of the inverse trigonometric functions. In particular, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\arctan(x) = \frac{1}{1+x^2}, \qquad \frac{\mathrm{d}}{\mathrm{d}x}\arcsin(x) = \frac{1}{\sqrt{1-x^2}}$$

so that the derivative of trigonometric functions yielded quantities which were effectively rational in description. If one recollects how these derivatives were computed, one of the main reasons that rational functions appear is the Pythagorean identities:

$$\sin^2(x) + \cos^2(x) = 1, \qquad \tan^2(x) + 1 = \sec^2(x).$$

We will use this, but in reverse, to develop a new integration technique. The idea is that when we see something that looks like $\sqrt{1-x^2}$, we should think of the how this relates to the sine function, and we will make a substitution $x = \sin(\theta)$. When we see something of the form $1 + x^2$ we think of the tangent function and will make the substitution $x = \tan(\theta)$.

The way I have written these substitutions is important: up to now many of our substitutions have involved writing something of the form

$$\text{new variable} = \text{ some function of the old variable}$$

and that is effectively what we are doing again, by defining say $\theta = \arctan(x)$. However, this is rather messy to work with as written, so we often write $\tan(\theta) = x$. Computing differentials will give $\sec^2(\theta)\,\mathrm{d}\theta = \mathrm{d}x$.

Let's first try this on an integrand for which we know what the result should be:

**Example 6.14**

Determine the integral $\displaystyle\int \frac{1}{1+x^2}\,\mathrm{d}x$.

*Solution.* We make the substitution $x = \tan(\theta)$ so that $\mathrm{d}x = \sec^2(\theta)\,\mathrm{d}\theta$. Our integral thus becomes

$$\begin{aligned}\int \frac{1}{1+x^2}\,\mathrm{d}x &= \int \frac{\sec^2(\theta)}{1+\tan^2(\theta)}\,\mathrm{d}\theta \\ &= \int \frac{\sec^2(\theta)}{\sec^2(\theta)}\,\mathrm{d}\theta \\ &= \theta + C.\end{aligned}$$

Since $x = \tan(\theta)$ we know that $\theta = \arctan(x)$, so that

$$\int \frac{\mathrm{d}x}{1+x^2} = \arctan(x) + C$$

exactly as we expected. ■

Not all things will be so simple, so let us try a slightly harder question:

**Example 6.15**

Determine $\int \frac{x^2}{\sqrt{16-x^2}}\,\mathrm{d}x$.

*Solution.* This time we are tempted to make the substitution $x = \sin(\theta)$ and morally this is correct. However, if we think about it a bit more we can make our lives a lot easier by setting $x = 4\sin(\theta)$. The reason is that

$$16 - x^2 = 16 - (4\sin(x))^2 = 16\left[1 - \sin^2(x)\right] = 16\cos^2(x).$$

This will make our lives much easier, as removing the addition sign will allow us to remove the square root. If $x = 4\sin(\theta)$ then $\mathrm{d}x = 4\cos(\theta)\,\mathrm{d}\theta$ and we get

$$\begin{aligned}\int \frac{x^2}{\sqrt{16-x^2}}\,\mathrm{d}x &= \int \frac{\left[16\sin^2(\theta)\right]\left[4\cos(\theta)\,\mathrm{d}\theta\right]}{\sqrt{16\cos^2(\theta)}}\,\mathrm{d}\theta \\ &= 16\int \sin^2(\theta)\,\mathrm{d}\theta \\ &= 8\left[\theta - \frac{1}{2}\sin(2\theta)\right] + C\end{aligned}$$

The tricky part comes in changing this back into a function of $x$. Since $x = 4\sin(\theta)$ we know that $\theta = \arcsin(x/4)$, but the pesky presence of the $2\theta$ means that we cannot just simply plug this back in. Instead, we must make the following substitution:

$$\begin{aligned}\sin(2\theta) &= 2\sin(\theta)\cos(\theta) \\ &= 2\frac{x}{4}\frac{\sqrt{16-x^2}}{4} \\ &= \frac{x\sqrt{16-x^2}}{8}.\end{aligned}$$

We conclude that

$$\int \frac{x^2}{\sqrt{16-x^2}}\,\mathrm{d}x = 8\arcsin\left(\frac{x}{4}\right) - \frac{x\sqrt{16-x^2}}{2} + C. \qquad \blacksquare$$

**Example 6.16**

Compute the integral

$$\int \frac{x}{(x^2+6x+18)^{3/2}}\,\mathrm{d}x.$$

*Solution.* My hope would be that each student would *try* substitution first, though it is unlikely to succeed given that I am trying to demonstrate the technique of trigonometric substitution. Nonetheless, it should always be your starting point when you recognize a function and its derivative. Unfortunately, we do not have any of the recommended forms above so we must endeavour to manipulate the integrand until it looks amenable to our techniques. Indeed, we may quickly complete the square of the denominator to find that

$$x^2 + 6x + 18 = (x+3)^2 + 9.$$

Referencing our table above, we decide that we should make the substitution $x + 3 = 3\tan\theta$ so that $\mathrm{d}x = 3\sec^2(\theta)\,\mathrm{d}\theta$ and

$$(x+3)^2 + 9 = 9(\tan^2(\theta) + 1) = 9\sec^2(\theta).$$

Substituting into our integral we find

$$\begin{aligned}\int \frac{x}{(x^2+6x+18)^{3/2}}\,\mathrm{d}x &= \int \frac{(3\tan(\theta) - 3)}{(9\sec^2(\theta))^{3/2}}(3\sec^2(\theta))\,\mathrm{d}\theta \\ &= \int \frac{9(\tan(\theta)-1)\sec^2(\theta)}{27\sec^3(\theta)}\,\mathrm{d}\theta \\ &= \frac{1}{3}\int \frac{\tan(\theta)-1}{\sec(\theta)}\,\mathrm{d}\theta \\ &= \frac{1}{3}\int (\sin(\theta) - \cos(\theta))\,\mathrm{d}\theta \\ &= -\frac{1}{3}[\cos(\theta) + \sin(\theta)] + C.\end{aligned}$$

Now $\tan\theta = (x+3)/3$ which implies (by drawing our triangle) that

$$\cos(\theta) = \frac{3}{\sqrt{x^2+6x+18}}, \qquad \sin(\theta) = \frac{x+3}{\sqrt{x^2+6x+18}}$$

and we conclude that

$$\int \frac{x}{(x^2+6x+18)^{3/2}}\,\mathrm{d}x = -\frac{1}{3}\frac{x+6}{\sqrt{x^2+6x+18}} + C. \qquad \blacksquare$$

In the case of the definite integral, we perform our substitution just as in Section 6.1.

**Example 6.17**

Determine the integral $\displaystyle\int_1^2 \frac{1}{x^2\sqrt{x^2+4}}\,\mathrm{d}x$.

*Solution.* Let $x = 2\tan(\theta)$ so that $\mathrm{d}x = 2\sec^2(\theta)\,\mathrm{d}\theta$. Moreover, when $x = 1$ we have $\theta = \arctan(1/2)$, while when $x = 2$ we have $\theta = \arctan(1/2) = \pi/4$. Thus our integral becomes

$$\begin{aligned}\int_1^2 \frac{1}{x^2\sqrt{x^2+4}}\,\mathrm{d}x &= \int_{\arctan(1/2)}^{\pi/4} \frac{2\sec^2(\theta)}{4\tan^2(\theta)\sqrt{4\tan^2(\theta)+4}}\,\mathrm{d}\theta \\ &= \frac{1}{4}\int_{\arctan(1/2)}^{\pi/4} \frac{\sec(\theta)}{\tan^2(\theta)}\,\mathrm{d}\theta \\ &= \int_{\arctan(1/2)}^{\pi/4} \csc(\theta)\cot(\theta)\,\mathrm{d}\theta \\ &= -\frac{1}{4}\csc(\theta)\Big|_{\arctan(1/2)}^{\pi/4}.\end{aligned}$$

To compute csc(arctan(1/2)), note that

$$\sin(\arctan(1/2)) = \frac{1}{\sqrt{5}}$$

so that $\csc(\arctan(1/2)) = \sqrt{5}$, giving

$$\int_1^2 \frac{1}{x^2\sqrt{x^2+4}}\,\mathrm{d}x = \frac{\sqrt{5}-\sqrt{2}}{4}. \qquad \blacksquare$$

**Strategy**

| Integrand | Substitution |
|---|---|
| $x^2 + a^2$ | $x = a\tan(\theta)$ |
| $a^2 - x^2$ | $x = a\sin(\theta)$ |
| $x^2 - a^2$ | $x = a\sec(\theta)$ |

## 6.5 Partial Fractions

The method of partial fractions is a means to turn integrands which are rational functions (that is, the quotient of polynomials) into simpler constituent pieces which may be more simply integrated. Let $\frac{p(x)}{q(x)}$ be a rational function, so that $p$ and $q$ are polynomials. If we can write $q(x)$ as in product of linear factors, say $q(x) = (x-r_1)(x-r_2)\cdots(x-r_n)$ then our goal is to write

$$\frac{p(x)}{q(x)} = \frac{C_1}{x-r_1} + \frac{C_2}{x-r_2} + \cdots + \frac{C_n}{x-r_n},$$

for constants $C_i, i = 1,\ldots,n$. If $q$ contains an irreducible quadratic factor, say $q(x) = (x-r_1)(ax^2+bx+c)$ then we try to write[10]

$$\frac{p(x)}{q(x)} = \frac{A}{x-r_1} + \frac{Bx+C}{ax^2+bx+c}.$$

If in the factorization a factor occurs with multiplicity greater than 1, say $q(x) = (x-r_1)(x-r_2)^2$ then we seek an expression of the form

$$\frac{p(x)}{q(x)} = \frac{A}{x-r_1} + \frac{B}{x-r_2} + \frac{C}{(x-r_2)^2}.$$

More generally, if the factor $(x-r)$ occurs with multiplicity $m$, then we must account for a factor $(x-r)^j$ for all $j = 1,\ldots,m$. All possible combinations of the above rules also hold.

**Example 6.18**

Find the partial fractions decomposition of the rational function

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

[10] There are no irreducible polynomials of degree 3 or higher, so you only need to worry about quadratics.

*Solution.* We can factor the denominator as $x^2+x-2=(x+2)(x-1)$ so we look for a decomposition of the form

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

By cross multiplying, we must thus have $5x+1=A(x-1)+B(x+2)$. If one so wishes, we could expand this out to find that

$$5x+1=(A+B)x+(-A+2B), \qquad \begin{aligned} A+B&=5 \\ -A+2B&=1 \end{aligned}$$

which is a system of equations that can be solved. On the other hand, realizing that our equation must hold for all values of $x$, we could substitute convenient values of $x=1, x=-2$ into our expression to find that

$$6=3B, \qquad -9=-3A.$$

This tells us that $A=3$ and $B=2$, so that

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

The process of partial fractions allows us to decompose our integrand into bite size pieces. From here we can exploit the linearity of the integral.

**Example 6.19**

Determine the integral $\displaystyle\int \frac{5x+1}{x^2+x-2}\,\mathrm{d}x$.

*Solution.* One might be tempted to try substitution here, but there is no obvious way to make it work. Instead, we exploit the partial fraction decomposition that we computed in Example 6.18 to write

$$\begin{aligned}\int \frac{5x+1}{x^2+x-2}\,\mathrm{d}x &= \int\left[\frac{3}{x+2}+\frac{2}{x-1}\right]\mathrm{d}x \\ &= 3\log|x+2|+2\log|x-1|+C. \end{aligned} \qquad \blacksquare$$

The problem with partial fractions, as written, is that it is impossible for us to consider the case when the degree of the numerator is greater than or equal to that of the numerator. In such cases, we have to exploit the power polynomial long division:

**Example 6.20**

Compute $\displaystyle\int \frac{x^3-4x-10}{x^2-x-6}\,\mathrm{d}x$.

*Solution.* We would like to apply partial fractions, but first need to perform polynomial long division. Some quick calculations show us that

$$\frac{x^3-4x-10}{x^2-x-6}=x+1+\frac{3x-4}{x^2-x-6}.$$

We may now perform partial fractions on the latter part of this expression; namely

$$\frac{3x-4}{x^2-x-6} = \frac{A}{x-3} + \frac{B}{x-2}.$$

One may find that $A = 1$ and $B = 2$. Substituting this into the integral we get

$$\begin{aligned}\int \frac{x^3-4x-10}{x^2-x-6}\,\mathrm{d}x &= \int \left[x+1+\frac{1}{x-3}+\frac{2}{x+2}\right]\mathrm{d}x \\ &= \frac{1}{2}x^2 + x + \log|x-3| + 2\log|x-2| + C.\end{aligned}$$ ∎

As an unnecessarily difficult example, consider the final following problem:

**Example 6.21**

Compute $\displaystyle\int \frac{2x^3-4x^2+2x-2}{x^4-2x^3+2x^2-2x+1}\,\mathrm{d}x.$

*Solution.* By inspection, $x = 1$ is a root of the denominator, meaning that we can remove a factor of $x - 1$. Performing long division, we get that

$$x^4-2x^3+2x^2-2x+1 = (x-1)(x^3-x^2+x+1).$$

Once again the only possible rational roots for $x^3 - x^2 + x + 1$ are $x = \pm 1$, and in fact we already know that $x = -1$ cannot be a root. Checking $x = 1$ we again see that 1 is a root, so we may remove another factor of $x - 1$ to get

$$x^4-2x^3+2x^2-2x+1 = (x-1)^2(x^2+1).$$

As $x^2 + 1$ has no real roots, we cannot factor any further. Our partial fraction decomposition is thus of the form

$$\frac{2x^3-4x^2+2x-2}{x^4-2x^3+2x^2-2x+1} = \frac{A}{x-1} + \frac{B}{(x-1)^2} + \frac{Cx+D}{x^2+1},$$

which if we cross multiply gives us

$$2x^3-4x^2+2x-2 = A(x-1)(x^2+1) + B(x^2+1) + (Cx+D)(x-1)^2.$$

Substituting $x = 1$ we get $-2 = 2B$ so that $B = -1$. Now there are no other immediately nice $x$'s to substitution, so let's try $x = 0$:

$$-2 = -A - 1 + D, \qquad \Rightarrow \qquad D - A = -1.$$

Substituting $x = 2$ we get

$$2 = 5A - 5 + 2C + D \qquad \Rightarrow \qquad 5A + 2C + D = 7.$$

Substituting $x = -1$ we get

$$-10 = -4A - 2 + 4(D - C), \qquad \Rightarrow \qquad -4A + 4D - 4C = -8.$$

One can solve this to find $A = 1, D = 0, C = 1$, so that

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

Integrating we thus have

$$\begin{aligned}\int \frac{2x^3 - 4x^2 + 2x - 2}{x^4 - 2x^3 + 2x^2 - 2x + 1}\,\mathrm{d}x &= \int \left[\frac{1}{x-1} - \frac{1}{(x-1)^2} + \frac{x}{x^2+1}\right]\mathrm{d}x \\ &= \log|x-1| + \frac{1}{x-1} + \frac{1}{2}\log|x^2+1| + C.\end{aligned}$$

■

# 7 Applications of Integration

## 7.1 Area Computations

One of the primary motivations for developing the theory of integral calculus is to actually compute areas. The area of some objects are easy to compute; such as rectangles, triangles, parallelograms, and even trapezoids. Our ability to find formulas for the area of these shapes hinges upon the fact that they are constructed with straight lines, and so may be related to rectangles.

A quick glance around any room, let alone the free expanse of nature, very quickly confirms that there are very few naturally occurring rectangles, triangles, or trapezoids. Mother Nature, it seems, is not a fan of straight lines. You might cry, "But we know the area of a circle is $\pi r^2$!" Ah, but this formula was actually determined by Archimedes, effectively emulating integration.

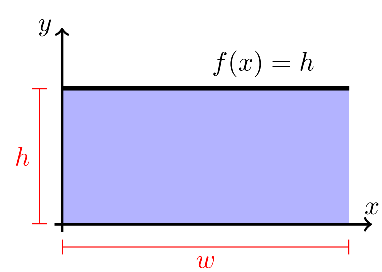
### 7.1.1 What We Already Know

Let us recall that integration does give us our classical formulas:

**Rectangles:**

Consider a rectangle with height $h$ and width $w$. Let $f : [0, w] \to \mathbb{R}$ be the constant function $f(x) = h$. The area under the graph of $f$ is precisely the area of the rectangle, and indeed we have

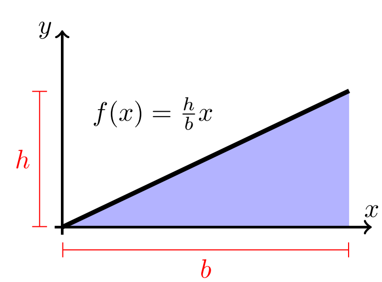
$$\int_0^w f(x)\,\mathrm{d}x = \int_0^w h\,\mathrm{d}x = hw$$




**Triangles:**

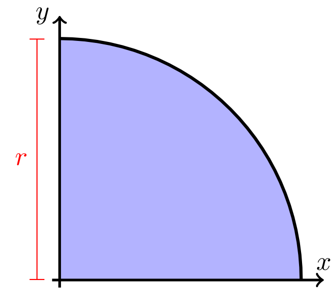
Consider a triangle with base $b$ and height $h$. Define the function $f : [0, h] \to \mathbb{R}$ by $f(x) = \frac{b}{h}x$, which is a straight line with height $f(h) = b$. The area under $f$ is the area of the desired rectangle, and integrating yields

$$\int_0^h \left[\frac{h}{b}x\right]\mathrm{d}x = \frac{b}{h}\frac{x^2}{2}\bigg|_0^h = \frac{1}{2}bh.$$



**Circles:**

Let $r > 0$ be the radius of our circle. We know that the formula of a circle is given by $x^2+y^2 = r^2$. We cannot write the circle as a function though, but by writing $y = \sqrt{r^2 - x^2}$ and integrating on $[0, r]$ we can determine a quarter of the area of the circle. If we multiply by 4 at the end we will get the full area of the circle.




$f(x) = \sqrt{r^2 - x^2}$

This kind of equation lends itself to trigonometric substitution. We set $x = r\sin(\theta)$ so $\mathrm{d}x = r\cos(\theta)\,\mathrm{d}\theta$,

$$\begin{aligned}
\int_0^r \sqrt{r^2 - x^2}\,\mathrm{d}x &= \int_0^{\pi/2} \sqrt{r^2 - r^2\sin^2(\theta)}\,r\cos(\theta)\,\mathrm{d}\theta && \text{substitution}\\
&= r^2 \int_0^{\pi/2} \cos^2(\theta)\,\mathrm{d}\theta\\
&= \frac{r^2}{2}\int_0^{\pi/2} [1+\cos(2\theta)]\,\mathrm{d}\theta && \text{half-angle identity}\\
&= \frac{r^2}{2}\left[\theta + \frac{1}{2}\sin(2\theta)\right]_0^{\pi/2} = \frac{\pi r^2}{4}.
\end{aligned}$$

As this is the quarter area, the total area of the circle is $\pi r^2$ as suspected.

### 7.1.2 More Complicated Shapes

Most other shapes require that we use integration to determine their area. The next few examples are straightforward applications of the things we have learned thus far, but we really emphasize that without integration, the corresponding areas would be impossible to compute.

**Example 7.1**

Determine the area under the graph of $f(x) = xe^{x/4} + 1$ on $[0, 3]$.

*Solution.* The plot of the corresponding area is given in Figure 45. Using integration by parts ($u = x, dv = e^{x/4}\,\mathrm{d}x$), one finds that

$$\begin{aligned}
\int_0^3 \left[xe^{x/4} + 1\right]\,\mathrm{d}x &= \left[4xe^{x/4} + x\right]_0^3 - 4\int_0^3 e^{x/4}\,\mathrm{d}x\\
&= \left[12e^{3/4} + 3\right] - 16\left[e^{x/4}\right]_0^3\\
&= 19 - 4e^{3/4}.
\end{aligned}$$

■

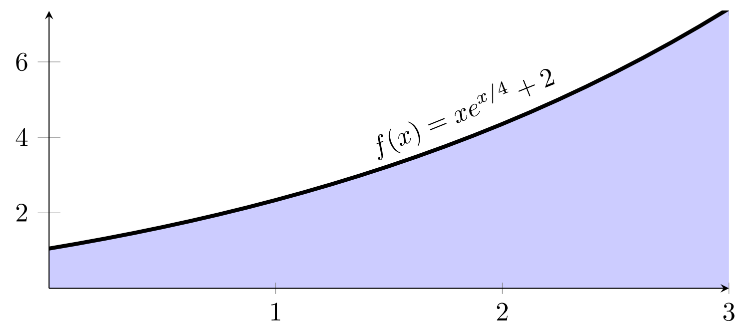



Figure 45: A relatively simple shape whose area is difficult to compute.

**Example 7.2**

Determine the area under the graph of $f(x) = \dfrac{\sin\left(\frac{3\pi}{x}\right)}{x^2}$ on the interval $[0.15, 0.5]$.

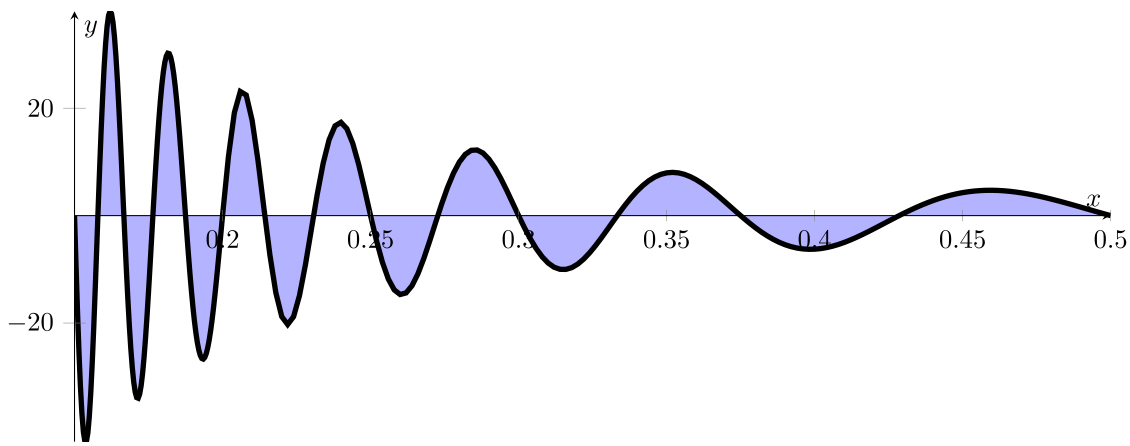



Figure 46: A function which is rapidly oscillating.

*Solution.* The corresponding plot is drawn in Figure 46. We proceed by integrating with a simple substitution $u = \frac{3\pi}{x}$ to get

$$\begin{aligned}\int_{0.15}^{0.5} \frac{\sin\left(\frac{3\pi}{x}\right)}{x^2}\,\mathrm{d}x &= -\frac{1}{3\pi}\int_{20\pi}^{6\pi} \sin(u)\,\mathrm{d}u \\ &= \frac{1}{3\pi}\cos(u)\big|_{20\pi}^{6\pi} = 0.\end{aligned}$$

This is rather interesting. It is clear that adjacent areas are very *similar* in shape, but not equal. Nonetheless, the areas managed to completely cancel one another out. ■

### 7.1.3 Unsigned (Absolute) Area

By now we are quite familiar with the fact that integrals compute *signed* areas; that is, areas above the $x$-axis are given positive sign, while those beneath the $x$-axis carry a negative sign. As such, it is possible for areas to be negative or even zero, as in Example 7.2.

To compute the unsigned, absolute area, we modify our function so that formerly negative areas now lie above the $x$-axis. This is done by taking the absolute value of the integrand: If $f : [a,b] \to \mathbb{R}$ is integrable, then

$$\text{Absolute Area} = \int_a^b |f(x)| \, dx.$$

**Example 7.3**

Determine the unsigned (that is, total) area under the graph of $\sin(x)$ on the interval $[-\pi, \pi]$.

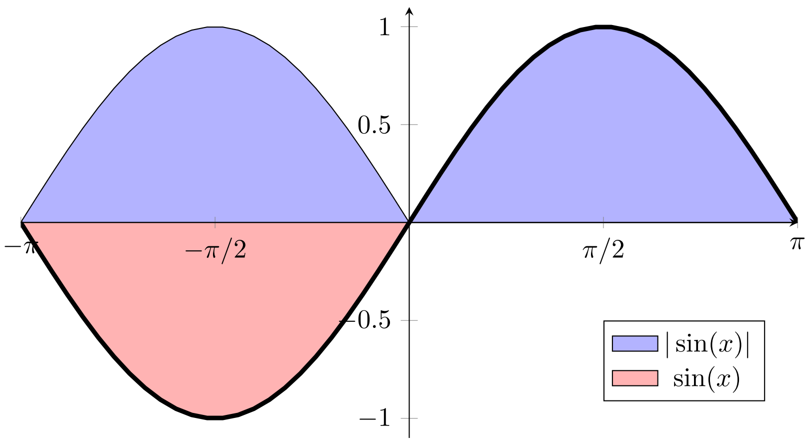



Figure 47: When we take the absolute value of a function, the areas which lie beneath the $x$-axis are flipped to lie above, so the integral will compute the total area. Here, the blue area is given by $|\sin(x)|$. The red area on the interval $[-\pi, 0]$ is the same as the blue area directly above it, but carries a negative sign.

*Solution.* Example 5.18 told us that the signed area was exactly zero despite the fact that $\sin(x)$ is not identically the zero function. To compute the total area, we shall compute the integral $\int_{-\pi}^{\pi} |\sin(x)| \, dx$. We deal with the absolute value in precisely the same manner that we always deal with absolute values: We break it into cases:

$$|\sin(x)| = \begin{cases} \sin(x) & 0 \leq x \leq \pi \\ -\sin(x) & -\pi \leq x < 0 \end{cases}.$$

By additivity of domain, we thus get

$$\begin{aligned} \int_{-\pi}^{\pi} |\sin(x)| \, dx &= \int_{-\pi}^{0} |\sin(x)| \, dx + \int_0^{\pi} |\sin(x)| \, dx \\ &= -\int_{-\pi}^{0} \sin(x) \, dx + \int_0^{\pi} \sin(x) \, dx \\ &= -\left[-\cos(x)\right]_{x=-\pi}^{0} + \left[-\cos(x)\right]_{x=0}^{\pi} \\ &= \cos(0) - \cos(\pi) - \cos(\pi) + \cos(0) = 4. \end{aligned}$$ ■

**Example 7.4**

Determine the total area beneath the graph of the function $f(x) = x^2 - 1$ on the interval $[-2, 2]$.

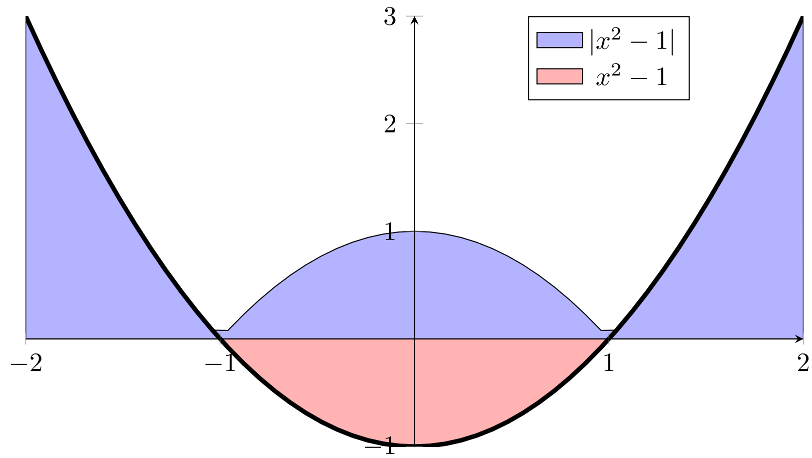



Figure 48: The blue area is that which is integrated by taking the absolute value. The red area on $[-1, 1]$ is the same as the blue area directly above it, except the blue carries a sign of $+1$ while the red carries a sign of $-1$.

*Solution.* We may determine the absolute area by integrating $|f(x)| = |x^2 - 1|$, but in practice this requires that we determine where $f(x) < 0$. The student can quickly verify that the roots of $f(x)$ lie at $x = \pm 1$, and that $f(x) < 0$ on $[-1, 1]$. Furthermore, $|f(x)|$ is an even function, allowing us to perform the integral on just $[0, 2]$:

$$\begin{aligned}\int_{-2}^{2} |x^2 - 1| \, \mathrm{d}x &= 2 \int_{0}^{2} |x^2 - 1| \, \mathrm{d}x \\ &= 2 \left[ \int_0^1 \left[1 - x^2\right] \, \mathrm{d}x + \int_1^2 \left[x^2 - 1\right] \, \mathrm{d}x \right] \\ &= 2 \left[ x - \frac{x^3}{3} \right]_0^1 + 2 \left[ \frac{x^3}{3} - x \right]_1^2 \\ &= 4.\end{aligned}$$

■

### 7.1.4 Integrating along the $y$-axis

There may be occasions where one is interested in the area under a curve which cannot necessarily be represented by a function, or in which it is merely inconvenient to write as a function. Examples like this will manifest in Section 7.1.5. As such, it may be more useful to integrate along the $y$-axis rather than the $x$-axis.

For example, consider the curve described by $y^2 - x - 2 = 0$, which is plotted in Figure 49. The curve does not describe a function in $x$, though we can solve for the two function components

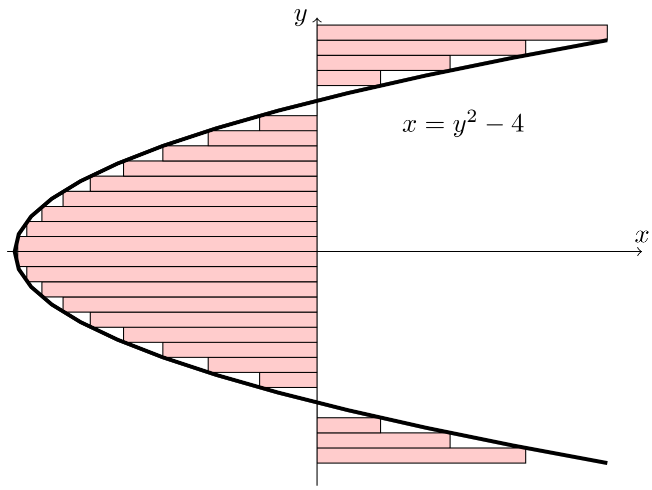



Figure 49: A Riemann sum for $x = y^2 - 4$, treated as a function of $y$.

$y = \pm\sqrt{x+2}$. On the other hand, the curve is a function in $y$ as we can write $x = y^2 - 2$. Were we set up a Riemann sum for this function, it would look like Figure-49. Notice that this is not the same area one would find if we were to integrate with respect to $x$ (Figure-50).

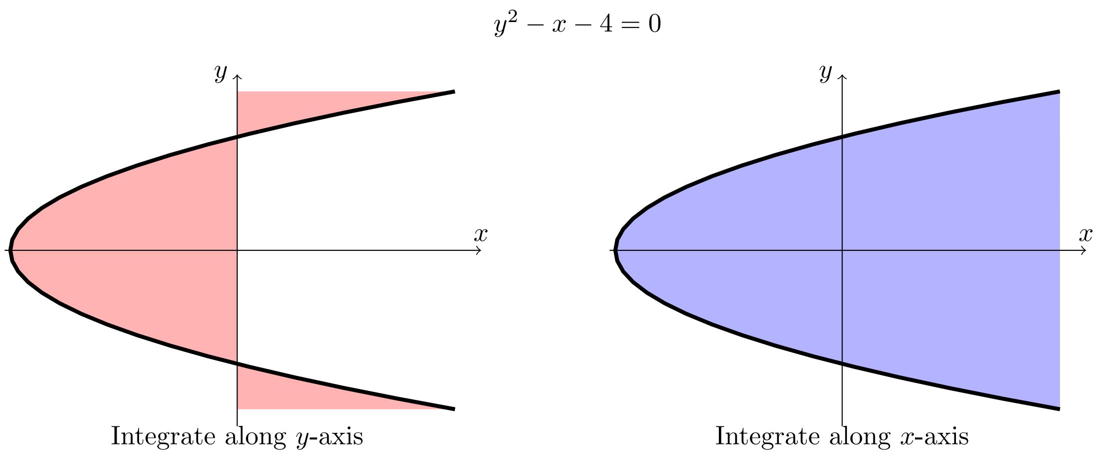



Figure 50: The computed area depends upon whether we are integrating with respect to $x$ or $y$.

**Example 7.5**

Compare the areas under the curve $y^2 - x - 4 = 0$ when integrated along the $x$- and $y$-direction, on the interval $x \in [-4, 5]$.

*Solution.* We first integrate along the $x$-direction as usual. While the curve is not described by a function, we can just integrate $y = \sqrt{x+4}$ and double the final answer.

$$2\int_{-4}^{5} \sqrt{x+4}\,\mathrm{d}x = 3\,(x+4)^{3/2}\Big|_{-4}^{5} = 81.$$

On the other hand, if we set $x = y^2 - 4$ then $x \in [-4, 5]$ implies that $y \in [-3, 3]$, so

$$\begin{aligned}\int_{-3}^{3} \left(y^2 - 4\right) \, \mathrm{d}y &= 2 \int_0^3 \left(y^2 - 4\right) \, \mathrm{d}y \\ &= 2 \left[\frac{y^3}{3} - 4y\right]_0^3 \\ &= -6.\end{aligned}$$

A very significant difference. ■

### 7.1.5 The Area Between Curves

Every example thus far measured the area between the graph of a function and the $x$-axis. We can increase our flexibility with a bit of creativity. For example, let's say we want to find the area bounded by the curves $y = x^2$ and $y^2 = x$ (Figure 51).

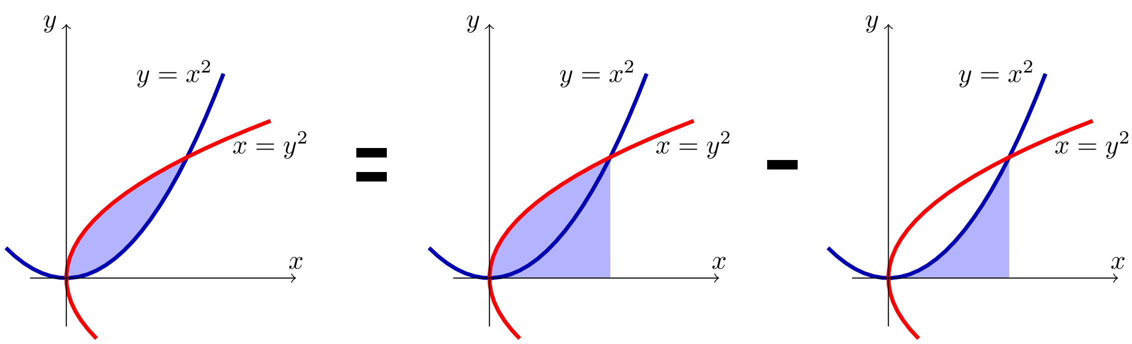



Figure 51: To compute the area bounded between the curves $y = x^2$ and $x = y^2$ requires that we be more creative.

The idea is fairly simple: By computing the area under $y = x^2$ and subtracting the area under $x = y^2$, we should get the area bounded between the two curves. More generally, given two curves $f, g$ on $[a, b]$, the area between $f$ and $g$ can be computed as

$$\int_a^b [f(x) - g(x)] \, \mathrm{d}x. \tag{7.1}$$

We note though that this is still a *signed area*. In particular, any interval where $f(x) > (x)g$ will be assigned a positive area, while area where $f < g$ will be given a negative area. Of course, the total area can be computed by $\int_a^b |f(x) - g(x)| \, \mathrm{d}x$.

Equation (7.1) can also be interpreted as the limit of a Riemann sum, or as integrating the actual function given by $f(x) - g(x)$. In the case where $f(x) = \sqrt{x}$ and $g(x) = x^2$, these two pictures are given by Figure 52.

**Example 7.6**

Find the area bounded between the functions $f(x) = \sqrt{x}$ and $g(x) = x^2$.

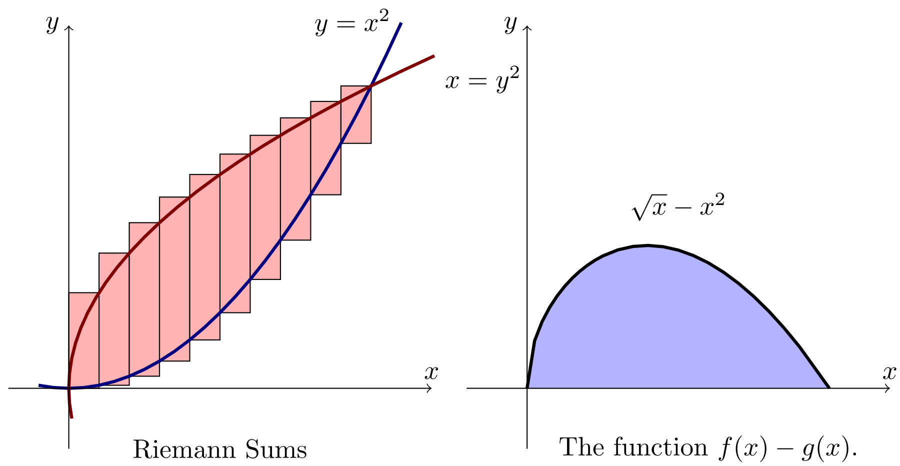



Figure 52: Two different interpretations of $\int_a^b [f(x) - g(x)]\, \mathrm{d}x$. Left: As a Riemann sum. Right: As the integral of the function $f(x) - g(x)$.

*Solution.* Notice that we were not explicitly given an interval over which to integrate. The reason is that if one sketches the graphs of $f$ and $g$, there is only one area that can be said to be enclosed by the two functions. Consequently, we need to determine where the appropriate intersections occur. This can be done by equation $f(x) = g(x)$.

Setting $x^2 = \sqrt{x}$, one can easily to solve to find that the intercepts occur at $x = 0$ and $x = 1$. Furthermore, on $[0, 1]$ we have that $\sqrt{x} > x^2$, and hence our area is given as

$$\begin{aligned} \int_0^1 \left[\sqrt{x} - x^2\right] \mathrm{d}x &= \left[\frac{2}{3}x^{3/2} - \frac{1}{3}x^3\right]_0^1 \\ &= \frac{2}{3} - \frac{1}{3} = \frac{1}{3}. \end{aligned}$$

We can also integrate in the $y$-direction. Here our interval is still $[0, 1]$ but $y^2 < \sqrt{y}$, so

$$\int_0^1 \left[\sqrt{y} - y^2\right] \mathrm{d}y = \frac{1}{3}$$

gives exactly the same answer. ■

**Example 7.7**

Determine the area bounded by the curves $y^2 = 2x + 6$ and $y = x - 1$.

*Solution.* To integrate with respect to $x$ we would need to break the curves down into the regions where the difference can be written as a function. This is doable, but is fairly complicated. Instead, if we integrate with respect to $y$ this becomes rather simple.

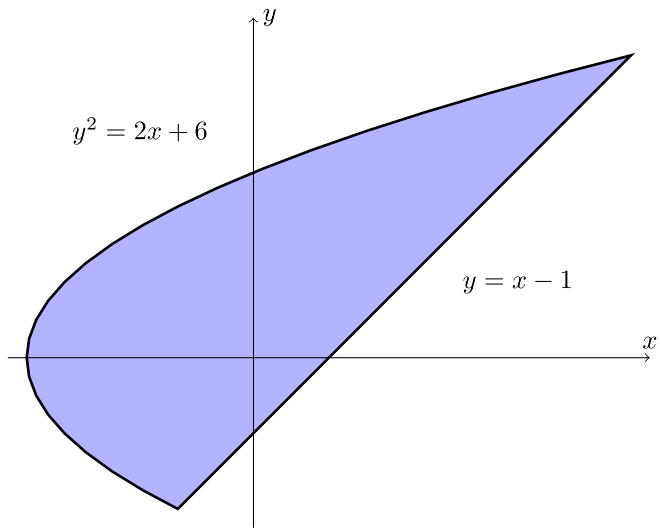



Figure 53: The area between the curves $y^2 = 2x + 6$ and $y = x - 1$.

We must first determine where the two curves intersect. Substituting $y = x - 1$ into $y^2 = 2x + 6$ we get

$$(x-1)^2 = 2x + 6 \quad \Leftrightarrow \quad x^2 - 4x - 5 = 0 \quad \Leftrightarrow \quad x = 5, -1.$$

Computing the corresponding $y$-values, we get that the intercept points occur at $(-1, -2)$ and $(5, 4)$. Hence $y$ ranges over the interval $[-2, 4]$. Rewriting our curves as functions of $y$, we get $x = y + 1$ and $x = y^2/3 - 3$. The area can be computed to be

$$\begin{aligned}
\int_{-2}^{4} \left[ (y+1) - \left( \frac{y^2}{3} - 3 \right) \right] \mathrm{d}y &= \int_{-2}^{4} \left[ -\frac{y^2}{3} + y + 4 \right] \mathrm{d}y \\
&= \left[ -\frac{y^3}{9} + \frac{y^2}{2} + 4y \right]_{-2}^{4} \\
&= \left( \frac{64}{9} + 8 + 16 \right) - \left( \frac{8}{9} + 2 - 9 \right) \\
&= 22.
\end{aligned}$$

■

## 7.2 Work and Mass

Science is abound with physical laws, often described by equations whose parameters remain constant in time. For example:

1. The distance $d$ traversed by an object travelling with constant speed $s$ for a time $t$ is given by $d = st$,
2. The work $W$ done by a constant force $F$ to move a distance $d$ is given by $W = Fd$.
3. The center of mass of a one-dimensional rod with uniform density $\rho$ and length $\ell$ is $\ell/2$,

These equations are very nice, but represent idealized situations. For example, what happens to our distance formula $d = st$ if our speed is allowed to vary with time? Similarly, our force $F$ may vary with distance or our density $\rho$ might not be uniform. When we allow objects to behave dynamically, we are forced to employ calculus.

We have already addressed the problem of calculating distance with variable speed; in fact, this was our motivational problem presented in Section 5.2.1. If $s(t)$ describes the speed of a particle as a function of time $t$, we can estimate the total distance travelled by breaking the interval $[t_0, t_1]$ into smaller time intervals, then approximating the speed by choosing a sampling of $s$ on each subinterval. For example, if we choose to take $n$-uniform intervals of length $\Delta t$, and approximate $s$ on each subinterval by the speed at the beginning of the interval, then

$$\text{approximate distance} = \sum_{k=0}^{n-1} s\left(\frac{t_1 - t_0}{n} k\right) \Delta t,$$

which we recognize as the left Riemann sum. The true distance travelled manifests as we take the limit $\Delta t \to 0$, yielding a total distance

$$d = \int_{t_0}^{t_1} s(t)\, \mathrm{d}t.$$

We will implement the same strategy to determine the resolution for our other hypothesized scenarios, when density and force are allowed to vary with time.

> **Remark 7.8** We emphasize that we have chosen work and mass as representative examples only. The strategies developed in the following sections are meant to be generalized by the reader to arbitrary situations, whenever needed.

### 7.2.1 Work

Deriving the formula for work dependent upon a variable force will almost be identical to the derivation of distance subject to variable speed. Nonetheless, it is important to see how such derivations are done mathematically, so we begin with this simple example to reinforce the logical process.

Consider a particle, acted upon by a force $F$. If the force is constant in time, and the particle is displaced by an amount $s$, then the work $W$ done on the particle by the force $F$ is given by $W = Fs$. The SI-units for force are Newtons, while work is given by Joules = Newtons · seconds. More realistically, force often varies either with respect to distance $s$, time $t$, or both. For example, the force of gravity exerted by the sun is significantly greater on Mercury than on Jupiter as a result of the distances of these planets to the sun. On the other hand, an electric motor on an alternating current experiences a time-dependent force, arising from the change in polarity of a magnetic field.

Our goal now is to consider the correct equation for $W$ when our force is allowed to vary. Let's assume that $F(x)$ is a continuous function of the displacement $s$ on $[0, s_f]$, where $s_f > 0$ is the final displacement of the particle. We can approximate the work exerted by the force on the particle by breaking $[0, s_f]$ into sub-intervals, approximating $F$ by a constant on each such interval, and using the equation $W = Fs$ on each of these intervals.

To this effect, we partition $[0, s_f]$ into $n$ equal length subintervals, with $\Delta s = s_f/n$, and choose a sampling $t_i \in [i\Delta s, (i+1)\Delta s]$. The approximate work done on the interval $[i\Delta s, (i+1)\Delta s]$ is then given by $W_i = F(t_i)\Delta s$, so the total work is approximated by

$$\text{approximate work} = \sum_{k=1}^{n} W_i = \sum_{k=1}^{n} F(t_i)\Delta s.$$

This is just a Riemann sum! The total work is then found by taking the limit as our partition intervals become infinitesimally small. Since $F$ is continuous it is integrable, implying that the limit will converge, so we conclude that

$$W = \int_0^{s_f} F(s)\,\mathrm{d}s. \tag{7.2}$$

**Displacement versus Distance:** Our integral above is in terms of the *displacement* of the particle; that is, the total distance the particle has travelled from its starting position. However, displacement is relative to the particle, while forces are typically more absolute, making displacement an inconvenient unit in which to express force. For example, a particle at the point $x$ on the real line might experience a force of $F(x) = 2/x^2$ Newtons. If the particle starts at $x = 5$ and finds itself at $x = 7$, then its displacement is $s = 2$. The force it feels at $s = 2$ corresponds to $x = 7$ though, so clearly our units $s$ and $x$ do not agree.

The relationship between displacement and distance is straightforward: If $x_0$ is the initial location of the particle, then $s = x - x_0$. One solution is therefore to write the force $F$ in terms of $s$, giving $F(x) = F(s + x_0)$, which can be substituted into (7.2) and integrated to find the answer. Alternatively, we can perform a change of variables! Since $s = x - x_0$ we know that $\mathrm{d}s = \mathrm{d}x$, and so

$$W = \int_0^{s_f} F(s)\,\mathrm{d}s = \int_{x_0}^{x_0+s_f} F(x)\,\mathrm{d}x.$$

**Example 7.9**

A particle at position $x \in \mathbb{R}$ is acted upon by a force of $F(x) = 2/x^2$ N. If the particle begins at $x_0 = 5$ and finishes at $x = 7$, determine the work done by the force on the particle.

*Solution.* Our particle experiences a displacement of $s_f = 2$. As mentioned above, we know that the relationship between $s$ and $x$ is given by $s = x - x_0 = x - 5$. Rewriting $F$ in terms of the displacement, we have $F(x) = F(s+5) = 2/(x+5)^2$, so that

$$W = \int_0^2 F(s+5)\,\mathrm{d}s = \int_0^2 \frac{2}{(x+5)^2}\,\mathrm{d}s = -\left.\frac{2}{x+5}\right|_0^2 = \frac{4}{35} \text{ Joules.}$$

Alternatively, we can integrate using the change of variable method, to find

$$W = \int_5^7 \frac{2}{x^2}\,\mathrm{d}x = -\left.\frac{2}{x}\right|_5^7 = \frac{4}{35} \text{ Joules.} \qquad \blacksquare$$

There are many variations one could employ to change the dynamics of our problem. For example, the force $F$ could vary upon time $t$ but not upon distance. In this instance, our integral formula changes. Alternatively, consider an instance where a particle is moving in two dimensions. If $F$ is a force which acts only in the $x$-direction, then the effective force felt by the particle is $F\cos(\theta)$, where $\theta$ is the angle made by the tangent vector to the curve and the $x$-axis. A full treatment of this scenario requires multi-variable calculus, but one could imagine being given $\theta$ as a function of time, and being told to determine the work. These are examples that will be taken up in the exercises.

### 7.2.2 Mass

The study of mass will yield several important results. Most obviously, mass is an important quantity in physics and warrants discussion on those grounds alone. However, more pertinent to our study will be the technique used to compute the center of mass for one- and two-dimensional spaces.

We begin with the problem of determining the mass of a one-dimensional rod of length $\ell$, composed of a material with density $\rho$. In a single dimension, density measures how much mass occurs per unit length; therefore, if the rod has a uniform density $\rho$ the total mass will be $m = \rho\ell$. If the density is allowed to vary with the length of the rod, the total mass may be computed in a manner almost identical to what was considered in section 7.2.1. In the exercise below, the student will show that if a rod is placed on the $x$-axis with one endpoint at $x = 0$ and the other at $x = \ell$, then the total mass of the rod will be

$$m = \int_0^\ell \rho(x)\,\mathrm{d}x, \tag{7.3}$$

where the density $\rho$ now varies as a function of the length of the rod.

**Exercise:** (Mass of a Rod) Show that if a rod of length $\ell$ is composed of a material with density $\rho(x)$, $0 \leq x \leq \ell$, then the mass of the rod is given by (7.3).

Let's try something new: The center of mass of an object is the unique point of an object where the mass is equally distributed in all possible directions. In effect, one can always balance an object by positioning the fulcrum at the center of mass. One can often utilize the center of mass to reduce difficult computations involving a large object to that of a single point.

For a one-dimensional rod of uniform mass-density $\rho$, spanning the interval $[a, b]$, the center of mass will be located at the center of the interval; namely, $\text{CoM} = (a + b)/2$. Notice that in this case the density plays no role in determining the center of mass. How do we extend this notion to multiple different rods? Consider a collection of rods, who have centers of mass located at $\{x_1, \ldots, x_n\}$ and mass $\{m_1, \ldots, m_n\}$. The center of mass of the combined system will be given by

$$\text{CoM} = \frac{1}{m_1 + \cdots + m_n} \sum_{i=1}^n x_i m_i. \tag{7.4}$$

Take a moment and convince yourself that this makes sense. For example, if a marble with mass 1 gram is located at $x = -1$ and a marble of mass 1 gram is located at $x = +1$, then the center of mass should be at $x = 0$. However, if we increase the mass of the second marble to $x = 2$ grams, then the mass should be at $x = 1/3$, owing to the fact that the second marble is heavier.

We can use (7.4) to build an integral expression for the center of mass of a rod with continuously varying mass-density $\rho$. Assume the rod is placed so that one end is at $x = 0$ and the other is at $x = \ell$. We will approximate the center of mass by breaking the rod into smaller rods, choosing a sample of $\rho$ on each smaller rod, and then summing according to (7.4). Partition $[0, \ell]$ into $n$ equal subintervals, with $\Delta x = \ell/n$ and tags $t_i \in [i\Delta x, (i+1)\Delta x]$. Each smaller rod on $[i\Delta x, (i+1)\Delta x]$ has an approximate total mass of $m_i = \rho(t_i)\Delta x$, so that the center of mass of the $n$ rods is

$$\begin{aligned}\text{Approximate CoM} &= \frac{1}{m_1 + \cdots + m_n} \sum_{i=1}^n \left(\frac{x_{i-1} + x_i}{2}\right) \rho(t_i)\Delta x. \\ &= \underbrace{\frac{1}{M}\sum_{i=1}^n \frac{x_{i-1}}{2}\rho(t_i)\Delta x}_{\text{(I)}} + \underbrace{\frac{1}{M}\sum_{i=1}^n \frac{x_i}{2}\rho(t_i)\Delta x_i}_{\text{(II)}}.\end{aligned}$$

Term (II) is a proper Riemann sum for the function $\frac{x}{2}\rho(x)$, and will converge if $\rho$ is integrable. Term (I) is *not* a Riemman sum, but we know that it converges to the integral for $\frac{x}{2}\rho(x)$ by Bliss' Theorem. We conclude that the center of mass is given by

$$\text{Center of Mass} = \left[\int_0^\ell \rho(x)\,\mathrm{d}x\right]^{-1} \left[\int_0^\ell x\rho(x)\,\mathrm{d}x\right],$$

which is the continuous version of the sum given in (7.4).

**Example 7.10**

Calculate the center of mass for a rod spanning $[-1, 1]$ with a mass-density given by $\rho(x) = 1 - x^3$.

*Solution.* We could choose to move our bar so as to be position along the interval $[0, 2]$, but as the mass density is in terms of $x \in [-1, 1]$, it is easiest just to leave the system as it is. We begin by computing the mass of the rod,

$$M = \int_{-1}^1 \rho(x)\,\mathrm{d}x = \int_{-1}^1 [1 - x^3]\,\mathrm{d}x = 2.$$

Next, we compute the integral over $x\rho(x)$, which gives

$$\int_{-1}^1 x\rho(x)\,\mathrm{d}x = \int_{-1}^1 [x - x^4]\,\mathrm{d}x = -\frac{2}{5}.$$

Thus the center of mass is located at $x = -1/5$. ■

## 7.3 Cross Sections

Beyond two dimensional area lies three dimensional volume. In general, we will need to develop the theory of multivariate integration to compute volumes. However, if one is presented with a sufficiently simple shape, the characterization of which we will make clear shortly, our current theory of integration will be sufficient to determine volume.

The major theme of this entire section is that integrating 'adds a dimension,' so that volume can be computed by integrating area. One way of thinking about computing the area under the graph of a function $f(x)$ is by continuously summing vertical lines that sweep out the area (Figure-54).

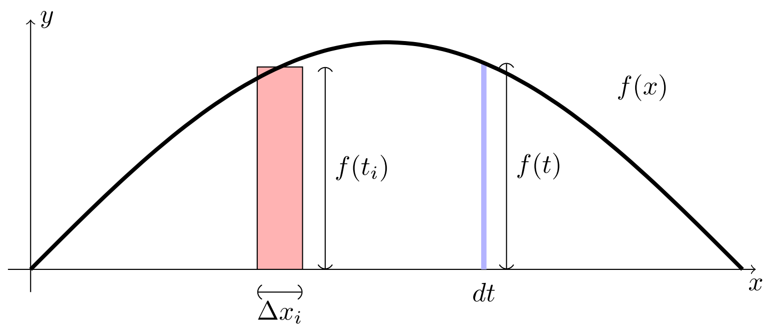



Figure 54: To be rigorous, we use rectangles of positive width (red) to ensure that the area approximation could be done with a finite sum. In the limit as $\Delta x_i \to 0$, we may think of this as adding the heights of infinitely many lines with infinitesimal width $\mathrm{d}t$ (blue).

Integrating length - a one dimensional thing - yielded area. To compute three dimensional volumes, we will thus want to integrate areas. To do this, we need to find a way of "sweeping out the volume;" that is, using a continuous collection of two-dimensional objects re-create the three-dimensional object. When computing areas, there was effectively only one way of doing this using lines. In computing volumes, there will be several more methods.

### 7.3.1 The Cross-Section Formula

Our first approach is to use cross-sections. If $S$ is a three-dimensional object, a cross-section of $S$ is the intersection of $S$ with a plane. Cross-sections are thus slices of $S$. If we allow the plane to move across $S$, we can piece together cross-sections to reconstruct $S$. This is similar to how a Magnetic Resonance Imaging (MRI) machine represents data: by piecing together two dimensional pictures, one can reconstruct the three dimensional object.

In the case of areas, slices of the area were given by lines. Slicing along the line $x = t$ gives a line of height $f(t)$ with infinitesimal width $\mathrm{d}t$. By continuously summing these lines, we derived the area.

To make this somewhat more precise, recall that we did not really integrate lines, but rather, we looked at rectangles with height $f(x)$ and length $\Delta x$. It is in the limit that $\Delta x$ becomes a $\mathrm{d}x$ and we perceive the infinitesimal picture. For volume, we will do the same thing. By partitioning our interval, the elements $A(x)\,\mathrm{d}x$ will represent infinitesimal volume elements, which we will integrate to get volume.

Consider a three dimensional object $S$ sitting above the interval $[a, b]$ on the $x$-axis. Partition $[a, b]$ into $n$ equal length subintervals, with $\Delta x = (b - a)/n$. Choose a point $t_i \in [a + i\Delta x, a + (i + 1)\Delta x]$ so that $A(t_i)$ denotes the area of the cross section which given by the intersection of $S$ with

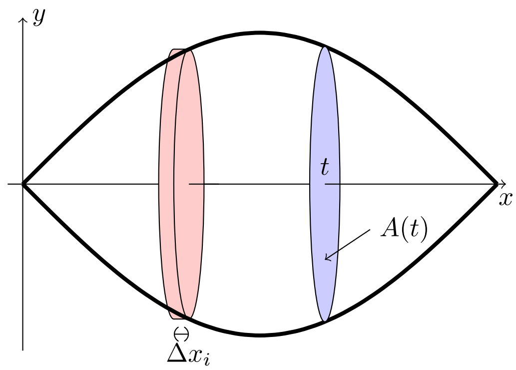



Figure 55: We can approximate the volume by summing together thick cross sections. Each thick cross-section has a volume of $A(t_i)\Delta x_i$ for some $t_i \in [x_{i-1}, x_i]$. In the limit as $\Delta x_i \to 0$, we find that the volume is the integral of the cross-sectional area function.

the plane $x = t_i$. The quantity

$$S(A, P) = \sum_{i=1}^{n} A(t_i)\Delta x$$

represents an approximation of the volume. If $A$ is an integrable function, the quantity above is a Riemann sum, and so

$$\text{Volume} = \int_a^b A(x)\,\mathrm{d}x. \tag{7.5}$$

Naturally, we could choose to take cross sections with the planes $y = s$, so that the cross-sectional area will be a function of $y$, $A(y)$. The volume will then be given by

$$\text{Volume} = \int_c^d A(y)\,\mathrm{d}y.$$

### 7.3.2 Simple Cross-Sections

We can use our cross-section formula to verify several well-known volume formulas, and determine new formulas for less orthodox shapes.

**Example 7.11**

Consider the cylinder with radius $r$ and height $h$. Verify that the volume of the cylinder is given by $V = \pi r^2 h$.

*Solution.* Let's draw the cylinder so that it is lying with its radial center coinciding with the $x$-axis. With this setup, it is most natural to take $x$ based cross sections (try to image what $y$ based cross sections would look like). The cross-section at the point $x$ is a circle of radius $\pi r^2$, and in fact is constant with respect to $x$. As such, we integrate on $[0, h]$ to get

$$\text{Volume of cylinder} = \int_0^h \pi r^2\,\mathrm{d}x = \pi r^2 h. \qquad \blacksquare$$

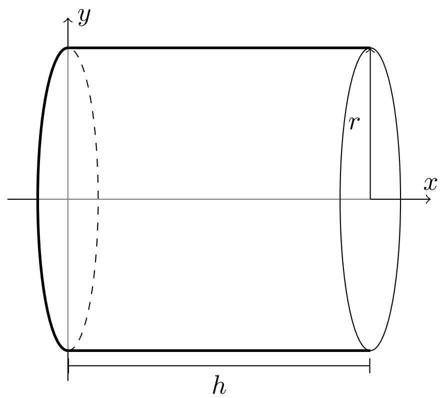



Figure 56: The cylinder is the solid of revolution for the constant function $f(x) = r$ on $[0, h]$, rotated about the $x$-axis.

**Example 7.12**

Consider the rectangular pyramid with base dimensions $\ell, w$ and height $h$. Verify that the volume of this pyramid is $V = \frac{1}{3}w\ell h$.

*Solution.* Let the pyramid be drawn as in Figure 57. The cross-sections with $x$-planes yields trapezoids while $y$-planes yield squares, so let's integrate in the $y$-direction. For an arbitrary $y$, we need to determine the area of the corresponding square with height $y$. Let $\hat{\ell}(y)$ and $\hat{w}(y)$ be the corresponding length and width of the square at height $y$. By projecting into the $xy$-plane, we get the triangles shown in Figure 57, which are similar. As such,

$$\frac{\hat{\ell}(y)/2}{h-y} = \frac{\ell/2}{h} \quad \Rightarrow \quad \hat{\ell}(y) = \frac{(h-y)\ell}{h}.$$

Projecting into the $yz$-plane yields precisely the same argument but with $\hat{w}(y)$ instead of $\hat{\ell}(y)$, so

$$\hat{w}(y) = \frac{(h-y)w}{h}.$$

The cross-sectional area at $y$ is thus

$$A(y) = \hat{\ell}(y)\hat{w}(y) = \frac{(h-y)^2\ell w}{h^2}.$$

Integrating the cross-sections gives us the volume

$$\begin{aligned} V &= \int_0^h A(y)\,\mathrm{d}y = \int_0^h \frac{(h-y)^2\ell w}{h^2}\,\mathrm{d}y \\ &= \frac{\ell w}{h^2}\int_0^h (h-y)^2\,\mathrm{d}y = -\frac{\ell w}{3h^2}\left[(h-y)^3\right]_0^h \\ &= \frac{\ell w h}{3}. \end{aligned}$$

■

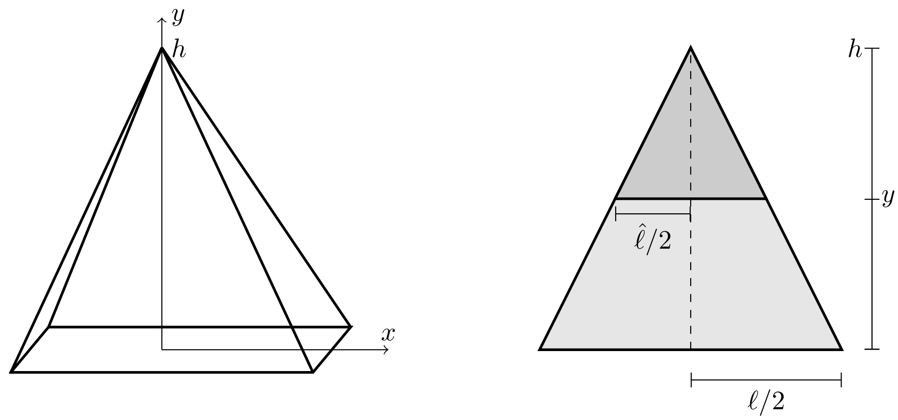



Figure 57: Left: Orientation of our pyramid. Right: Projecting into the $xy$-plane gives us similar triangles.

**Remark 7.13** The argument for the volume of a cone is almost identical to that of Example 7.12 and is left as an exercise. The observant reader may have noticed that the volume for the cone and pyramid are 1/3 the volume of the cylinder and cuboid, which are the unpointed versions of the cone and pyramid. Our derivation above shows why the factor of 1/3 appears.

We can extend the techniques illustrated above to determine the volumes of more exotic shapes, as the next two examples illustrate:

**Example 7.14**

Consider the solid object, having as its base the ellipse $4x^2 + 9y^2 = 36$, and having $x$ cross-sections given by squares. Determine the volume of the object.

*Solution.* For a fixed $x \in [-3, 3]$, we know that the corresponding cross-section is a square, whose half-length is given by $y = \sqrt{4 - \frac{4}{9}x^2}$. As such, the area of the cross-sections is $A(x) = 16 - \frac{16}{9}x^2$ and we get

$$\begin{aligned} V &= \int_{-3}^{3} \left[16 - \frac{16}{9}x^2\right] \, \mathrm{d}x \\ &= 32 \left[x - \frac{x^3}{27}\right]_0^3 \\ &= 64. \end{aligned}$$

■

**Example 7.15**

Consider the solid object whose base is prescribed by the region bounded by the curves $y = x$ and $y = x^2$. If the $x$ cross sections are equilateral triangles, determine the volume of the solid.

*Solution.* The student can quickly check that $y = x$ and $y = x^2$ intersect at $x = 0$ and $x = 1$, so fix an $x \in [0, 1]$. We are told that the cross section is an equilateral triangle, so at $x$ this triangle has a base of length $\ell = x - \sqrt{x}$. The area of an equilateral triangle with base $\ell$ is $A = \frac{\sqrt{3}}{4}\ell^2$, so

$$\begin{aligned} V &= \int_0^1 \frac{\sqrt{3}}{4}(x - \sqrt{x})^2 \, \mathrm{d}x \\ &= \frac{\sqrt{3}}{4} \int_0^1 \left[x^2 - 2x^{3/2} + x\right] \mathrm{d}x \\ &= \frac{\sqrt{3}}{4} \left[\frac{1}{3}x^3 - \frac{4}{5}x^{5/2} + \frac{1}{2}x^2\right]_0^1 \\ &= \frac{\sqrt{3}}{120}. \end{aligned}$$

■

### 7.3.3 Cross-Sections for Solids of Revolution

Symmetry has played an important role in our conversation thus far. The motif of the previous section was to specify a shape with a given base and highly regular cross-sections. Here we will explore a different kind of symmetry, wherein our cross-sections are universally given by circles of varying radii.

The idea is as follows: Suppose that one specifies a function $f$ with some domain $[a, b]$. By rotating that function around the $x$-axis one generates a "solid of revolution." Our goal will be to compute the volumes of such solids.

By virtue of having revolved the function $f(x)$ around the $x$-axis, cross-sections with respect to the axis of revolution will yield circles whose radius will be precisely $f(x)$. Thus the cross-sectional area as a function of $x$ is $A(x) = \pi f(x)^2$, and our cross-section formula reads

$$\text{Volume} = \int_a^b \pi f(x)^2 \, \mathrm{d}x.$$

**Example 7.16**

Let $f(x) = x^2 - 2x + 2$ with domain $[0, 3]$ and revolve the graph of $f$ about the $x$-axis. Determine the corresponding volume of the solid of revolution.

*Solution.* Taking a cross section at a typical point $x$, we get a circle whose radius is $f(x) =$

$x^2 - 2x + 2$. As such our cross-section formula reveals that

$$\begin{aligned} V &= \int_0^3 \pi \left(x^2 - 2x + 2\right)^2 \, \mathrm{d}x \\ &= \pi \int_0^3 \left[x^4 - 4x^3 + 8x^2 - 8x + 4\right] \mathrm{d}x \\ &= \pi \left[\frac{1}{5}x^5 - x^4 + \frac{8}{3}x^3 - 4x^2 + 4x\right]_0^3 \\ &= \frac{78\pi}{5}. \end{aligned}$$

■

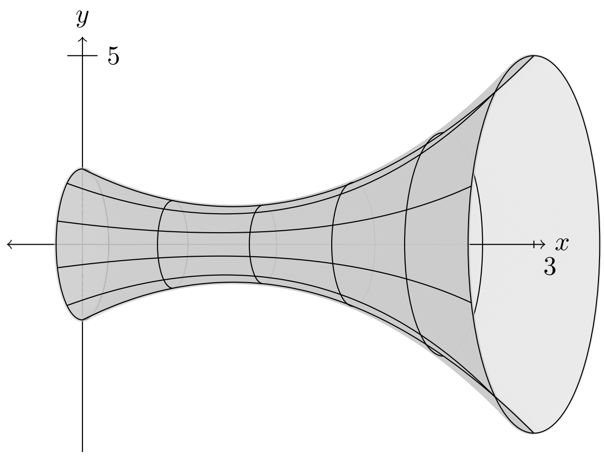



Figure 58: The surface of revolution for $f(x) = x^2 - 2x + 2$ on $[0, 3]$ for Example 7.16.

Of course, there is nothing preventing us from revolving around the $y$-axis, so long as we take $y$ cross sections.

**Example 7.17**

Consider the curve $y = \sqrt{x}$ for $x \in [0, 1]$. If this curve is revolved around the $y$-axis, determine the volume of the corresponding solid of revolution.

*Solution.* Notice that when $0 \leq x \leq 1$ we have that $0 \leq y \leq 1$, so we will be integrating over the same range. For a fixed $y \in [0, 1]$ the cross section is a circle with radius $x$. Of course, the relationship between $x$ and $y$ is given by $y = \sqrt{x}$, or more conveniently written, $x = y^2$. Thus the cross-section at $y$ is a circle with radius $y^2$, yielding $A(y) = \pi y^2$ and

$$V = \int_0^1 \pi y^2 \, \mathrm{d}y = \frac{\pi}{3}.$$

■

### 7.3.4 Area bounded by two curves

We can step up our difficulty level by rotating an area enclosed by two curves. Let $f, g$ be two functions on $[a, b]$ with $f(x) \leq g(x)$, and consider the volume between them. If we rotate this

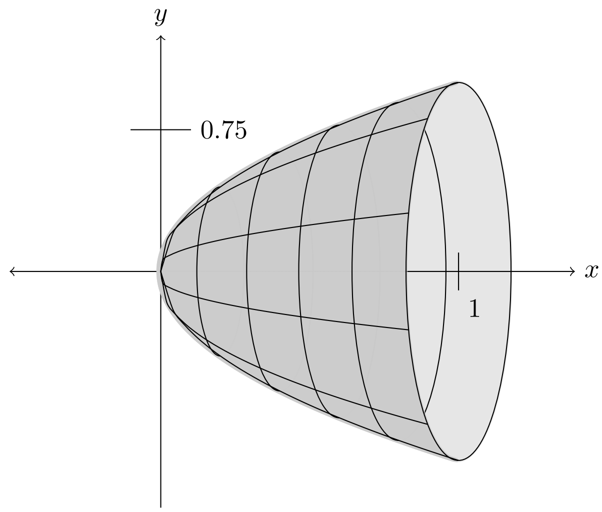



Figure 59: The surface of revolution of $f(x) = \sqrt{x}$ on $[0, 1]$, for Example 7.17.

area about the $x$-axis we still get a solid of revolution, but one must be more careful regarding the cross-sections. Indeed, examine Figure-60, and notice that cross sections no longer looks like a circle, but rather like an annulus. The area of this annulus is just the difference in area between the outer circle (of radius $f(x)$) and the inner (of radius $g(x)$). We can write the cross-sectional area as $A(x) = \pi \left[f(x)^2 - g(x)^2\right]$, so that

$$V = \int_a^b \pi \left[g(x)^2 - f(x)^2\right] \, \mathrm{d}x. \tag{7.6}$$

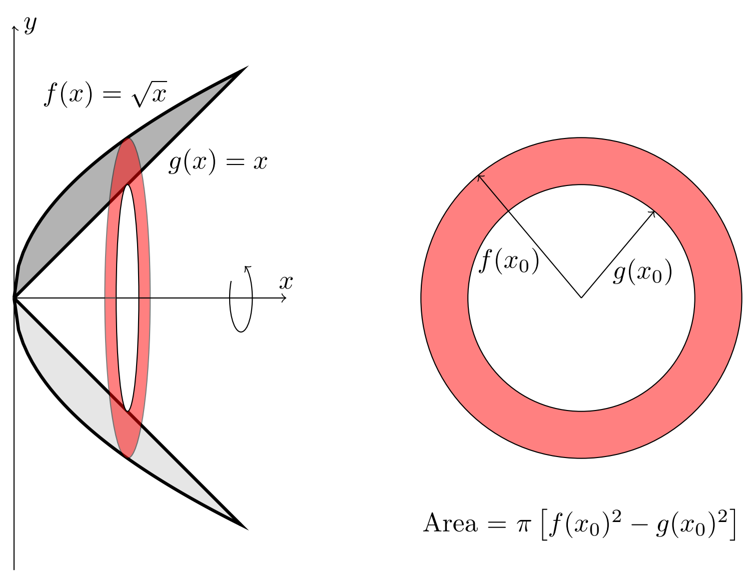



Figure 60: Rotating the area bounded by two curves $f(x)$ and $g(x)$ about the $x$-axis.

**Example 7.18**

Consider the area enclosed by the functions $f(x) = \sqrt{x}$ and $g(x) = x$. Determine the volume of the object given by rotating that area about the $x$-axis.

*Solution.* Once again, we consider a typical cross-section at the point $x_0$ (See Figure 60). The object we get by taking out this slice looks like a washer, whose outer radius is $f(x_0)$ and whose inner radius is $g(x_0)$. The cross-sectional area of this element is thus

$$A(x) = \pi f(x_0)^2 - \pi g(x_0)^2 = \pi[f(x_0)^2 - g(x_0)^2].$$

Integrating we thus get

$$\begin{aligned} V &= \int_0^1 A(x)\,\mathrm{d}x = \int_0^1 \pi\left[\sqrt{x}^2 - x^2\right]\,\mathrm{d}x \\ &= \pi\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1 = \frac{\pi}{6}. \end{aligned}$$

■

Let's revolve the same shape about the $y$-axis and see what volume we find.

**Example 7.19**

Determine the volume of the object given by rotating the area enclosed by $f(x) = \sqrt{x}$ and $g(x) = x$ about the $y$-axis.

*Solution.* This time, taking cross sections about a point $x$ yields really odd shapes whose area is difficult to compute. Instead, let's try taking cross-sections in the $y$-direction. Notice that the function $y = \sqrt{x}$ can be written as $y^2 = x$, so that the cross section in the $y$-direction again looks like a washer, but this time with outer radius $(x =)y$ and inner radius $(x =)y^2$. The cross-sectional area is thus $A(y) = \pi[y^2 - y^4]$. Again, our interval is going from 0 to 1, and hence the volume is given by

$$\begin{aligned} V &= \int_0^1 \pi\left[y^2 - y^4\right]\,\mathrm{d}y \\ &= \left[\frac{1}{3}y^3 - \frac{1}{5}y^5\right]_0^1 = \frac{2\pi}{15}. \end{aligned}$$

■

### 7.3.5 Rotating about different axes

We start with lines which are parallel to the coordinate axes; namely, lines of the form $x = c$ or $y = d$. In these cases the argument for determining the volume of a solid of revolution is a simple adaptation of what we have discussed above. In essence, the only thing which changes is radius of the cross-sections.

Consider a function $f$ defined on $[a, b]$ and the horizontal line $y = d$. Rotate the area between $f$ and this line about the axis $y = d$. For a fixed $x \in [a, b]$, the cross-section at $x$ is a circle of radius

$|f(x) - d|$. The cross-sectional area is thus $A(x) = \pi(f(x) - d)^2$ and our volume is given by

$$V = \int_a^b \pi \left(f(x) - d\right)^2 \, dx.$$

**Example 7.20**

Consider the curve $y = \sin(x)$ for $x \in [0, 2\pi]$, rotated about the line $y = 1$. Determine the volume of the solid of revolution.

*Solution.* Fix a point $x \in [0, 2\pi]$ and notice that the cross-section at $x$ is a circle of radius $1 - \sin(x)$. Thus the cross-sectional area at the point is given by $A(x) = \pi(1 - \sin(x))^2$ and so

$$\begin{aligned}
V &= \int_0^{2\pi} \pi \left(1 - \sin(x)\right)^2 \, dx \\
&= \pi \int_0^{2\pi} 1 - 2\sin(x) + \sin^2(x) \, dx \\
&= \pi \left[ x + 2\cos(x) + \frac{1}{2}\left[ x - \frac{1}{2}\sin(2x) \right] \right]_0^{2\pi} && \sin^2(x) = \frac{1}{2}\left[1 - \sin(2x)\right] \\
&= \pi \left[ (2\pi + 2 + \pi) - (2) \right] \\
&= 3\pi^2.
\end{aligned}$$

■

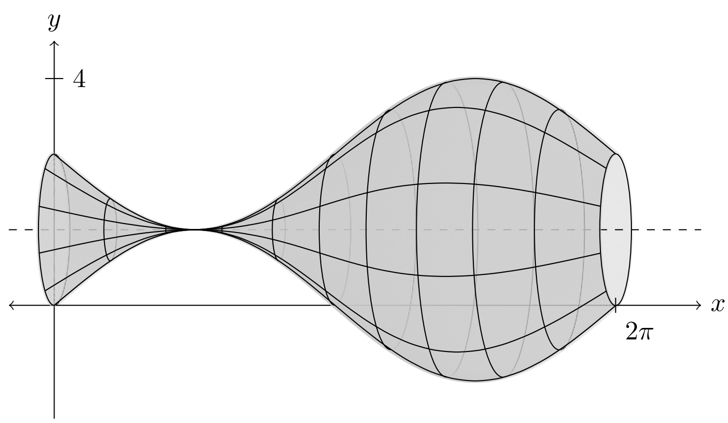



Figure 61: The surface of revolution for $f(x) = \sin(x)$ on $[0, 2\pi]$, revolved around the line $y = 1$.

**Example 7.21**

Consider the curve $x = y - y^2$ for $y \in [0, 1]$ rotated about the axis $x = -2$. Determine the volume of the solid of revolution.

*Solution.* For a fixed $y \in [0, 1]$, the cross-section at $y$ is a circle of radius $x + 2 = y - y^2 - 2$. Thus our cross-sectional area is $A(y) = \pi(y - y^2 - 2)^2$ and the volume can be computed as

$$\begin{aligned} V &= \int_0^1 \pi(y^2 - y + 2)^2 \, \mathrm{d}y \\ &= \pi \int_0^1 y^4 - 2y^3 + 5y^2 - 4y + 4 \, \mathrm{d}y \\ &= \pi \left[ \frac{1}{5}y^5 - \frac{1}{2}y^4 + \frac{5}{3}y^3 - 2y^2 + 4y \right]_0^1 \\ &= \frac{101\pi}{30}. \end{aligned}$$

■

As in every case before, we can also look at what happens when we revolve the area between two curves along a non-standard axis.

**Example 7.22**

Determine the volume of the object given by rotating the area enclosed by $f(x) = \sqrt{x}$ and $g(x) = x$ about the line $x = 1$.

*Solution.* As always, we take a typical cross-section, again in the $y$-direction this time. We again have a annulus whose outer radius is $1 - y^2$ and whose inner radius is $1 - y$. The cross sectional area is thus

$$A(y) = \pi[(1 - y^2)^2 - (1 - y)^2] = \pi(2y - 3y^2 + y^4),$$

and the volume can be computed as

$$V = \pi \int_0^1 [2y - 3y^2 + y^4] \, \mathrm{d}y = \pi \left[ y^2 - y^3 + \frac{1}{5}y^5 \right]_0^1 = \frac{\pi}{5}.$$

■

## 7.4 Other Volumes

The method of cross-sections is very powerful, but does not represent the only way to get at the volume of an object. Here we will introduce another method for determining the volume of objects, based off an alternative scheme for sweeping out volume.

### 7.4.1 Sweeping by Shells

Cross-sections may not always be convenient, resulting in difficult or impossible integrals. Here we develop a shell-technique, which involves using successively nested cylinder to approximate the volume. As in the previous sections, the trick is to think about an area, representing an infinitesimal piece of volume, and sweep out the shape with which we are concerned. For our first example, let $f$ be a function with domain $[a, b]$, and assume that we rotate the area under the graph of $f$ about the $y$-axis. Choose a point $x \in [a, b]$ and consider the cylinder which is centred at the $y$-axis, with radius $x$ and height $f(x)$. The surface area of this cylinder is given by

$$S(x) = 2\pi x f(x).$$

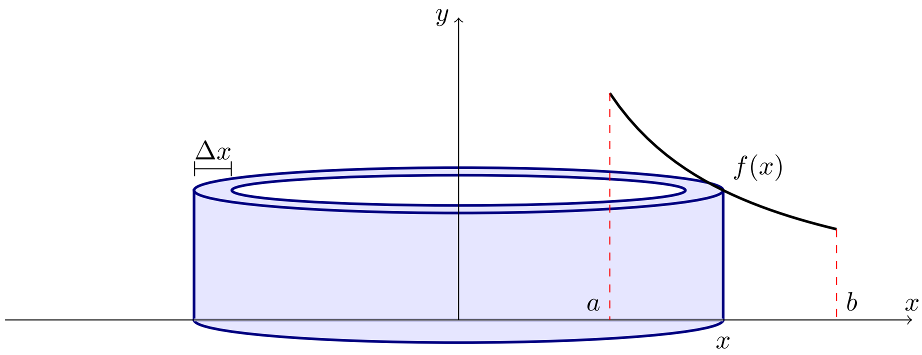



Figure 62: A fattened cylinder with height $f(x)$, radius $x$, and thickness $\Delta x$. If $\Delta x$ is very small, the volume of this shell is approximately $2\pi x f(x)\Delta x$.

We could think about fattening up this cylinder by adding $\Delta x$ to the radius, as in Figure-62. If $\Delta x$ is very small, we can unroll the cylinder to get a rectangular cube, whose volume is *approximately* given by $2\pi x f(x)\Delta x$. The volume of the object can be computed by sweeping these cylinders from $a$ to $b$, to give us the volume

$$V = \int_a^b 2\pi x f(x)\,\mathrm{d}x. \tag{7.7}$$

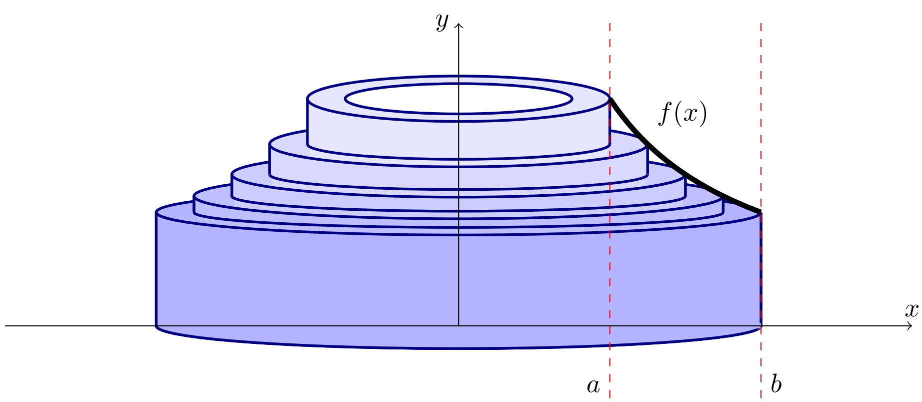



Figure 63: An illustration of how one can use cylindrical shells to approximate a volume.

**Example 7.23**

Determine the volume of the object created by rotating the function $\sqrt{x}$ on $[0,4]$ about the $y$-axis.

*Solution.* If we use our shell-method above, the radius of each infinitesimally small shell is $x$ and the height is $f(x)$, so the volume can be computed to be

$$V = \int_0^4 2\pi x\sqrt{x}\,\mathrm{d}x = 2\pi\left[\frac{2}{5}x^{5/2}\right]_0^4 = \frac{128}{5}\pi.$$

Our function is sufficiently simple that we can verify our result using the cross-sections. Since we are revolving around the $y$-axis, the $y$ cross-sections are annuli, with an outer radius of 4 and an of radius $x = y^2$. As $0 \leq x \leq 4$ we know that $0 \leq y \leq 2$ and so

$$V = \int_0^2 \pi \left[16 - y^4\right] \, dy = \pi \left[16y - \frac{1}{5}y^5\right]_0^2 = \frac{128}{5}\pi. \qquad \blacksquare$$

**Example 7.24**

Determine the volume of the sphere of radius $r$.

*Solution.* Consider the curve $y = \sqrt{r^2 - x^2}$ from $[0, r]$. By rotating this about the $y$-axis, we will get the upper hemisphere of the sphere, and hence half the area. Using the method of shells, we thus get

$$\begin{aligned} V &= 2\pi \int_0^r x\sqrt{r^2 - x^2} \, dx && \text{Substitution } u = r^2 - x^2 \\ &= -\pi \int_{r^2}^0 \sqrt{u} \, du \\ &= \pi \left[\frac{2}{3}u^{3/2}\right]_0^{r^2} = \frac{2}{3}\pi r^3. \end{aligned}$$

Of course, the total area of the sphere is then twice this amount, giving $\frac{4}{3}\pi r^3$, which is exactly what we expected. $\blacksquare$

Naturally, we can also rotate about the $y$-axis, wherein our radius is now given by $y$ and the height by $x = g(y)$.

**Example 7.25**

Consider the curve $y = \arcsin(x)$ for $x \in [0, 1]$. If the area between the curve and the $y$-axis is rotated about the $x$-axis, determine the volume of the corresponding object.

*Solution.* Revolving about the $x$-axis, our shells look like cylinders with radius $y \in [0, \pi/2]$ and height $x = \sin(y)$. Integrating over $y$ yields

$$\begin{aligned} \int_0^{\pi/2} 2\pi y \sin(y) \, dy &= 2\pi \left[ -y\cos(y)\Big|_0^{\pi/2} + \int_0^{\pi/2} \cos(y) \, dy\right] && \text{by parts} \\ &= 2\pi \left[\sin(y)\right]_0^{\pi/2} = 2\pi. \end{aligned}$$

As an exercise, the student should verify that this result agrees with the cross-section method. $\blacksquare$

Depending on the complexity of the corresponding curve, either the cross-section or shell method is often easier. However, in some cases one method is completely intractable and we must rely completely on the other.

**Example 7.26**

Determine the volume when $e^{x^2}$ on $[0,2]$ is rotated about the $y$-axis.

*Solution.* If we were to do this using cross-sections, one would need to solve $x$ as a function of $y$. Since $x \geq 0$ this can be done to yield $x = \sqrt{\log(y)}$, but this is not a function which is easily integrated. Instead, if we use the shell method we have cylinders of radius $x$ and height $e^{x^2}$, so that

$$\begin{aligned} V &= \int_0^2 2\pi x e^{x^2}\,\mathrm{d}x && u = x^2, du = 2x\,\mathrm{d}x \\ &= \pi \int_0^4 e^u\,\mathrm{d}u = \pi\, e^u\big|_0^4 = \pi(e^4 - 1) \end{aligned}$$

■

We can easily generalize this process to the case when we have a difference of functions. For example, consider the area bounded between two function $f(x)$ and $g(x)$ with $f(x) \leq g(x)$ on $[a,b]$. If we look at the cylinder in this case, the radius of the cylinder remains constant, but the height is given by the difference of the two functions. Hence the surface area is given by $S(x) = 2\pi x[g(x) - f(x)]$ and the volume becomes

$$V = \int_a^b 2\pi x[g(x) - f(x)]\,\mathrm{d}x. \tag{7.8}$$

**Example 7.27**

Determine the volume of the object given by rotating the area enclosed between $\sqrt{x}$ and $x^2$ about the $y$-axis.

*Solution.* We did this exact problem in Example 7.19 where we got a solution of $\frac{2\pi}{15}$. We certainly expect that we should get the same solution this time. Again, using the method of shells we get

$$\begin{aligned} V &= \int_0^1 2\pi x\left[\sqrt{x} - x\right]\,\mathrm{d}x = 2\pi \int_0^1 \left[x^{3/2} - x^2\right]\,\mathrm{d}x \\ &= 2\pi\left[\frac{2}{5}x^{5/2} - \frac{1}{3}x^4\right]_0^1 = \frac{2\pi}{15}. \end{aligned}$$

■

If we rotate about an axis which is not one of the coordinate axes, then the problem does not become too much more complicated. We keep in our mind this incessant picture that we are integrating using cylinders. Revolving around a different axis now just means we must be more careful in establishing the correct radii.

For example, let $y = f(x)$ be a curve for $x \in [a,b]$ and assume that we rotate this function about an axis $x = c$. For simplicity, let's assume for the moment that $c \geq b$. For each fixed $x \in [a,b]$ we are still using cylinders of height $f(x)$, but the radius of this cylinder is now $(c-x)$. This gives the volume formula

$$V = \int_a^b 2\pi(c-x)f(x)\,\mathrm{d}x.$$

If $c \leq a$ then the same argument holds, but the radius is now given by $x - c$. The corresponding changes can be made to adapt this for functions rotated about the $x$-axis, and for rotating areas bounded by two functions.

**Example 7.28**

Determine the volume of the object given by rotating the area enclosed by $\sqrt{x}$ and $x$ about the $x = 1$ axis.

*Solution.* We did this problem in Example 7.22, and found a solution of $\frac{\pi}{5}$. Setting up our cylinder, we see that for a typical slice our radius is now $1 - x$ while our height is $\sqrt{x} - x$; thus our surface area is given by

$$S(x) = 2\pi(1 - x)[\sqrt{x} - x].$$

Integrating to find the volume, we have

$$\begin{aligned} V &= \int_0^1 2\pi(1-x)[\sqrt{x} - x]\,\mathrm{d}x = 2\pi \int_0^1 \left[\sqrt{x} - x - x^{3/2} + x^2\right]\,\mathrm{d}x \\ &= 2\pi \left[\frac{2}{3}x^{3/2} - \frac{1}{2}x^2 - \frac{2}{5}x^{5/2} + \frac{1}{3}x^3\right]_0^1 = \frac{\pi}{5}. \end{aligned}$$ ■

### 7.4.2 Shells with Other Bases

When we examined how to compute volumes using cross-sections, we began with objects whose cross-sections were arbitrary shapes, not necessarily circles. Afterwards, we introduced solids of revolution as a special case of computing volumes where the cross-sections were always guaranteed to be circles. Our shell method discussed above works for solids of revolution, but what about other shapes with arbitrary cross-sections?

For example, what if we given the curve $y = f(x)$ for $x \in [a, b]$, and told to generate the solid about the $y$-axis whose base is a square rather than a circle? Is there a shell method for such a shape? The answer is very much yes, and the derivation is almost identical to the shell method for solids of revolution. We will leave the details as an exercise for the student, but the volume formula is given by

$$V = \int_a^b 8x f(x)\,\mathrm{d}x. \tag{7.9}$$

**Example 7.29**

Verify the volume formula for a square pyramid using the shell method.

*Solution.* Let $P$ by a pyramid with a square base of width $w$ and height $h$. This pyramid may be described by a square-rotation of the line $y = h - \frac{2h}{w}x$ about the $y$-axis. Using our shell method, we thus have

$$\begin{aligned} V &= \int_0^{w/2} 8x \left[ h - \frac{2h}{w} x \right] \mathrm{d}x \\ &= \left[ 4hx^2 - \frac{16h}{3w} x^2 \right]_0^{w/2} \\ &= hw^2 - \frac{2}{3} hw^2 = \frac{hw^2}{3}. \end{aligned}$$

This is precisely the equation for the volume of a square pyramid, as required. ■

There are other more complicated shapes that can be used, but in these cases the corresponding formula for the shell method becomes so convoluted that it only worthwhile as an exercise rather than in practice.

## 7.5 Improper Integrals

It was essential in defining the integral that we examined *bounded* functions on *bounded* intervals $[a, b]$, so that everything under consideration could be finite. In this section we examine how to now extend the idea of integrals to cover unbounded functions, and functions defined on unbounded intervals.

### 7.5.1 Infinite Intervals

Our goal is to extend the notion of integration from a finite interval $[a, b]$ to an infinite interval $[a, \infty)$ or $(-\infty, b]$. We will develop the idea for the interval $[a, \infty)$ and leave the details for $(-\infty, b]$ to the student.

**Definition 7.30**

Let $f$ be a bounded function on the interval $[a, \infty)$ such that $f$ is integrable on $[a, x]$ for every $x > a$. We define the *improper integral* of $f$ on $[a, \infty)$ as

$$\int_a^{\infty} f(t)\,\mathrm{d}t = \lim_{x \to \infty} \int_a^x f(t)\,\mathrm{d}t.$$

We say that that the improper integral *converges* if this limit is finite, and *diverges* otherwise.

Let us take a moment to think about what this is saying: We are defining a new function

$$F(x) = \int_a^x f(t)\,\mathrm{d}t,$$

which is precisely the anti-derivative of $f$. The improper integral then converges if the function $F(x)$ has a horizontal asymptote; that is, the area under the graph of $f$ asymptotically stabilizes to a single, finite number.

Naturally, one then defines the improper integral on $(-\infty, b]$ as

$$\int_{-\infty}^{b} f(t)\,\mathrm{d}t = \lim_{x\to-\infty}\int_{x}^{b} f(t)\,\mathrm{d}t.$$

**Example 7.31**

Determine $\displaystyle\int_0^\infty e^{-t}\,\mathrm{d}t$, if it exists.

*Solution.* By definition, we know that $\displaystyle\int_0^\infty e^{-t}\,\mathrm{d}t = \lim_{x\to\infty}\int_0^x e^{-t}\,\mathrm{d}t$. We are very familiar with computing the integral on the right hand side, and know that $\int_0^x e^{-t}\,\mathrm{d}t = -e^{-t}\big|_0^x = 1-e^{-x}$. Thus

$$\int_0^\infty e^{-t}\,\mathrm{d}t = \lim_{x\to\infty}\int_0^x e^{-t}\,\mathrm{d}t = \lim_{x\to\infty}\left[1-e^{-t}\right] = 1. \qquad \blacksquare$$

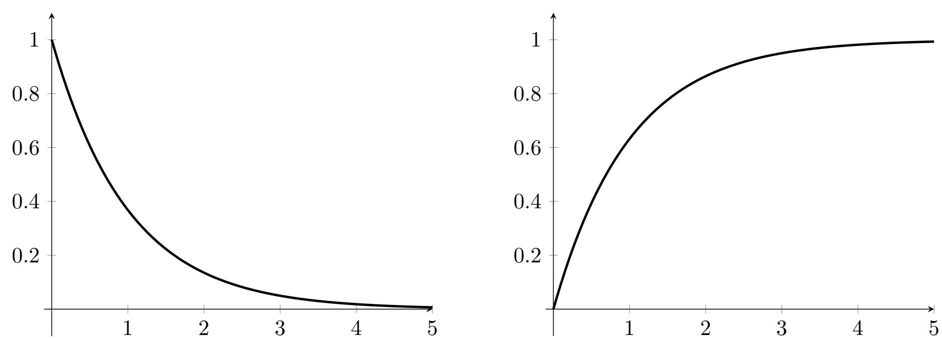



Figure 64: The function $f(x) = e^{-x}$ and an anti-derivative $F(x) = 1-e^{-x}$. We see that $F(x)$ tends to the number 1 as $x\to\infty$. This occurs because as $x\to\infty$ the graph under the function $f(x)$ becomes very small.

**Example 7.32**

Determine $\displaystyle\int_0^\infty \sin(x)\,\mathrm{d}x$, if it exists.

*Solution.* Proceeding by definition, we have that

$$\int_0^\infty \sin(t)\,\mathrm{d}t = \lim_{x\to\infty}\int_0^x \sin(t)\,\mathrm{d}t = \lim_{x\to\infty}\left[-\cos(t)\right]$$

which does not exist. If we think about this for a while, we can convince ourselves why it is true. The area under the graph of $\sin(x)$ is constantly oscillating between $+1$ and $-1$, and this oscillation never stops. Hence even though there are many points where the area is arbitrarily small, it's not enough to ensure that the area ever converges. $\blacksquare$

Intuitively, it seems like functions which tend to zero may have a convergent improper integral. However, this is not always the case. It is necessary that function goes to zero 'fast enough' to ensure that the area is positive.

**Proposition 7.33**

If $a > 0$ is an arbitrary positive number, then

$$\int_a^\infty \frac{1}{x^p}\,\mathrm{d}x \quad \text{converges if and only if} \quad p > 1.$$

*Proof.* If $p = 1$ then

$$\int_a^\infty \frac{1}{t}\,\mathrm{d}t = \lim_{x\to\infty} \log(x/a) = \infty$$

so the integral diverges. Thus assume that $p \neq 1$, for which we have

$$\int_a^\infty \frac{1}{t^p}\,\mathrm{d}t = \lim_{x\to\infty} \left[\frac{1}{1-p}\frac{1}{x^{p-1}}\right]_a^x.$$

We know that $1/x^{p-1}$ converges only if and if the power is non-negative; that is, $p - 1 \geq 0$. Combining this with $p \neq 1$ tells us that the improper integral converges if and only if $p > 1$. $\square$

As was computed explicitly, this implies that $\displaystyle\int_1^\infty \frac{1}{x}$ does not converge, which is a result that often confuses students.

What happens if we want to define the improper integral on $(-\infty, \infty)$?

**Definition 7.34**

If $f$ is integrable on every interval $[a, b] \subseteq \mathbb{R}$, then we say that $\displaystyle\int_{-\infty}^\infty f(t)\,\mathrm{d}t$ converges if, for any $c \in \mathbb{R}$ we have both

$$\int_{-\infty}^c f(t)\,\mathrm{d}t \text{ converges, and } \int_c^\infty f(t)\,\mathrm{d}t \text{ converges.}$$

In this case, we set[a]

$$\int_{-\infty}^\infty f(t)\,\mathrm{d}t = \int_{-\infty}^c f(t)\,\mathrm{d}t + \int_c^\infty f(t)\,\mathrm{d}t.$$

[a]The student should convince him/herself that the value of the improper integral does not depend on the value of $c$.

This is very different than simply demanding that

$$\int_{-\infty}^\infty f(t)\,\mathrm{d}t = \lim_{x\to\infty}\int_{-x}^x f(t)\,\mathrm{d}t \text{ exists.}$$

The reason is that the value of the integral should not depend on 'how quickly' we take our limits. For example

$$\lim_{x\to\infty}\int_{-x}^{x}\sin(t)\,\mathrm{d}t = \lim_{x\to\infty}\left[\cos(-x)-\cos(x)\right] = 0$$

since cosine is an even function. On the other hand,

$$\lim_{x\to\infty}\int_{-x}^{2x}\sin(t)\,\mathrm{d}t = \lim_{x\to\infty}\left[\cos(-x)-\cos(2x)\right] \text{ does not exist.}$$

**Example 7.35**

Determine $\int_{-\infty}^{\infty} e^{-|t|}\,\mathrm{d}t$.

*Solution.* A natural place to split our interval will be at 0. Now

$$\begin{aligned}\int_0^\infty e^{-|t|}\,\mathrm{d}t &= \int_0^\infty e^{-t}\,\mathrm{d}t && \text{since } t>0\\ &= 1 && \text{by Example 7.31.}\end{aligned}$$

Similarly, $\int_{-\infty}^{0} e^{-|t|}\,\mathrm{d}t = 1$, thus

$$\int_{-\infty}^{\infty} e^{-|t|}\,\mathrm{d}t = \int_{-\infty}^{0} e^{-|t|}\,\mathrm{d}t + \int_{0}^{\infty} e^{-|t|}\,\mathrm{d}t = 2.$$

■

### 7.5.2 Unbounded Functions

The case of unbounded functions often poses even more difficulty, since it is very tempting to just blindly apply the Fundamental Theorem of Calculus without paying attention.

**Example 7.36**

Compute the integral $\int_{-1}^{1}\frac{1}{x^2}\,\mathrm{d}x$.

*Solution.* We know that the anti-derivative of $\frac{1}{x^2}$ is $-\frac{1}{x}$, so if we were to just blindly apply the FTC we would get

$$\int_{-1}^{1}\frac{1}{x^2}\,\mathrm{d}x = -\frac{1}{x}\Big|_{-1}^{1} = -2.$$

Unfortunately, **this is completely and totally wrong.** Our first hint at a miscalculation is probably the fact that $\frac{1}{x^2}$ is everywhere positive, yet we somehow ended up with a negative integral.

In fact, the integral is infinite. To see this, note that since $\frac{1}{x^2} > 0$ then

$$\begin{aligned}\int_{-1}^{1} \frac{1}{x^2}\,\mathrm{d}x &\geq \int_{\epsilon}^{1} \frac{1}{x^2}\,\mathrm{d}x \\ &= -\frac{1}{x}\bigg|_{\epsilon}^{1} \\ &= \frac{1}{\epsilon} - 1,\end{aligned}$$

and that by choosing $\epsilon$ to be small enough we can make the integral arbitrarily large. The reason is that the function $\frac{1}{x^2}$ is not integrable on the interval $[-1,1]$ (all integrable functions are bounded!), and hence we could not apply the FTC. ■

**Exercise:** Compare the following two expressions:

$$\lim_{\epsilon\to 0^+}\left[\int_{-1}^{-\epsilon} \frac{1}{x}\,\mathrm{d}x + \int_{\epsilon}^{1} \frac{1}{x}\,\mathrm{d}x\right] \qquad \lim_{\epsilon\to 0^+}\left[\int_{-1}^{-\epsilon} \frac{1}{x}\,\mathrm{d}x + \int_{2\epsilon}^{1} \frac{1}{x}\,\mathrm{d}x\right]$$

The way to deal with unbounded functions is precisely the same way that we deal with unbounded intervals: we take a limit.

**Definition 7.37**

If $f$ is a function on $[a,b]$, unbounded at $a$, but integrable on $[x,b]$ for every $x > a$ then we define the improper integral

$$\int_a^b f(t)\,\mathrm{d}t = \lim_{x\to a^+} \int_x^b f(t)\,\mathrm{d}t.$$

If this limit is finite we say that the improper integral *converges*; otherwise, we say that the improper integral *diverges*.

Similarly, if $f$ were unbounded at $b$, we would define the improper integral as

$$\int_a^b f(t)\,\mathrm{d}t = \lim_{x\to b^-} \int_a^x f(t)\,\mathrm{d}t.$$

**Example 7.38**

Evaluate $\displaystyle\int_3^5 \frac{t}{\sqrt{t^2-9}}\,\mathrm{d}t$, if it exists.

*Solution.* We immediately recognize that there could be a problem at $t = 3$. Using the substitution $u = t^2 - 9$ we get

$$\begin{aligned}\int_3^5 \frac{t}{\sqrt{t^2-9}}\,\mathrm{d}t &= \lim_{x\to 3^+}\left[\sqrt{t^2-9}\right]_x^5 \\ &= \lim_{x\to 3^+}\left[4 - \sqrt{x^2-9}\right] = 4.\end{aligned}$$ ■

Just as in the case of integrating from $(-\infty, \infty)$ we must also be careful about integrating on both sides of an unbounded function.

**Definition 7.39**

If $f$ is a function on $[a,b]$ which is unbounded at $c \in [a,b]$ then we say that the improper integral

$$\int_a^b f(t)\,\mathrm{d}t \text{ converges, if and only if } \int_a^c f(t)\,\mathrm{d}t \text{ and } \int_c^b f(t)\,\mathrm{d}t \text{ both exist.}$$

In this case, we set

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t.$$

**Proposition 7.40**

For any $a \neq 0$ we have that

$$\int_0^a \frac{1}{t^p}\,\mathrm{d}t \text{ converges, if and only if } p < 1.$$

*Proof.* For simplicity, let's assume that $a > 0$. If $p = 1$ then

$$\int_0^a \frac{1}{t}\,\mathrm{d}t = \lim_{x\to 0^+} \int_x^a \frac{1}{t}\,\mathrm{d}t = \lim_{x\to 0^+} \log(t)\bigg|_x^a = \infty.$$

If $p \neq 1$ then

$$\int_0^1 \frac{1}{t^p}\,\mathrm{d}t = \lim_{x\to 0^+} \int_x^a \frac{1}{t}\,\mathrm{d}t = \frac{1}{1-p} \lim_{x\to 0^+} \frac{1}{x^{p-1}}\bigg|_x^a.$$

The limit converges if and only if $p - 1 \leq 0$ so that $p \leq 1$. Combined with the fact that $p \neq 1$ we get $p < 1$ as required. □

### 7.5.3 The Basic Comparison Test

In this section, we develop some techniques to make our lives simpler in terms of dealing improper integrals. The idea is something like the following: Say that you were asked to determine whether the following improper integral converged:

$$\int_0^\infty \left[\frac{2+\sin(x)}{x}\right]\mathrm{d}x,$$

or how about

$$\int_1^\infty \frac{x^2+1}{x^4+3x^2-4x+1}\,\mathrm{d}x?$$

Neither of these integrals are easy to compute explicitly; in fact, $\frac{\sin(x)}{x}$ cannot be solved in terms of elementary functions, but appears often enough to have a name: The Sine Integral. The goal, as is often the case with mathematics, is to reformulate these problems into ones that are much easier to solve, or that have already been solved. The Basic Comparison Test is the simplest of the basic tests, and exploits the "monotonicity" of the integral. In particular, the following lemma is rather intuitive:

**Theorem 7.41: The Basic Comparison Test for Improper Integrals**

Let $f(x), g(x)$ be functions on an interval $[a, \infty)$ such that $0 \leq f(x) \leq g(x)$ for all $x \in [a, \infty)$.

1. If $\displaystyle\int_a^\infty g(t)\,\mathrm{d}t$ converges then $\displaystyle\int_a^\infty f(t)\,\mathrm{d}t$ converges.
2. If $\displaystyle\int_a^\infty f(t)\,\mathrm{d}t$ diverges then $\displaystyle\int_a^\infty g(t)\,\mathrm{d}t$ diverges.

The idea is again a type of 'Squeeze Theorem' argument. If the integral of the bigger function $g$ becomes finite, the monotonicity of the integral cannot allow $f$ to go off to infinity. Similarly, if the integral of the smaller function $f$ goes off to infinity, the larger function's integral must also diverge. A good question at this point is to ask whether any of the integrals could oscillate and hence not converge. The condition that $0 \leq f(x) \leq g(x)$ guarantees that the integrands are positive: this means that the corresponding integrals are increasing functions.

**Example 7.42**

Show that $\displaystyle\int_1^\infty \frac{2+\sin(x)}{x}\,\mathrm{d}x$ diverges.

*Solution.* We want to compare the function $\frac{2+\sin(x)}{x}$ to some function which we know diverges. The idea to keep in mind is that in the limit as $x \to \infty$, the $\sin(x)$ term does not play a significant role; rather, the $x$-term in the denominator dominates. This suggests that we make the comparison against $\frac{1}{x}$. Indeed, notice that since $-1 \leq \sin(x) \leq 1$ for all $x$, we have

$$\frac{2+\sin(x)}{x} \geq \frac{1}{x}.$$

By Proposition 7.40, we know that $\int_1^\infty \frac{1}{x}$ diverges, so by the comparison test it follows that

$$\int_1^\infty \frac{2+\sin(x)}{x}\,\mathrm{d}x \text{ diverges.}$$ ■

**Example 7.43**

Determine whether $\displaystyle\int_0^\infty \frac{x}{\sqrt{x^6+1}}\,\mathrm{d}x$ converges or diverges.

*Solution.* Once again, we just want to look at which terms in the numerator and denominator dominate in the limit $x \to \infty$. Certainly, we do not expect $x^6 + 1$ to be too much different than $x^6$ for very large $x$, so we will compare our integrand to the function

$$\frac{x}{\sqrt{x^6}} = \frac{1}{x^2}.$$

Indeed, since $1 + x^6 \geq x^6$ we have that $\frac{1}{\sqrt{x^6}} \geq \frac{1}{\sqrt{1+x^6}}$, which in turn implies that

$$\frac{1}{x^2} = \frac{1}{\sqrt{x^6}} \geq \frac{x}{\sqrt{x^6+1}}.$$

Now the integral of the left-hand-side converges by 7.40, so by the Comparison Test we know that

$$\int_0^\infty \frac{x}{\sqrt{x^6+1}}\, dx \text{ converges.}$$

■

### 7.5.4 The Limit Comparison Test

The Basic Comparison Test is just that, basic. Often times the obvious inequality that you want actually ends up going in the wrong direction, yet the integrals are so similar that you feel like you should still be able to compare them. Example 7.45 below will demonstrate precisely this.

The Limit Comparison Test will fix this by asking the question: "Do $f$ and $g$ grow at roughly the same rate?"

**Theorem 7.44: The Limit Comparison Test**

Let $f, g$ be integrable functions on all subintervals of $[a, \infty)$ and satisfy $0 \leq f(x) \leq g(x)$. If

$$0 < \lim_{x\to\infty} \frac{f(x)}{g(x)} < \infty$$

then $\int_a^\infty f(x)\, dx$ converges if and only if $\int_a^\infty g(x)\, dx$ converges.

The statement that $\frac{f(x)}{g(x)}$ converges to some finite, non-zero number, means that $f$ and $g$ grow asymptotically at the same speed, up to some multiplicative constant (which is precisely the value of the limit). Note that if the limit is 0, eventually one must have $g(x) \geq f(x)$ and can use the Basic Comparison Test appropriately. Similarly, if the limit is $\infty$, then eventually $f(x) \geq g(x)$ and again the Basic Comparison Test can be used.

Also notice that the ratio does not matter, for if $\frac{f(x)}{g(x)} \to L$ which is finite and positive, then $\frac{g(x)}{f(x)} \to \frac{1}{L}$ which is also finite and positive.

**Example 7.45**

Determine whether $\int_1^\infty \frac{1}{\sqrt{1+x}}\, dx$ converges or diverges.

*Solution.* If one were to try to use the Basic Comparison Test, the obvious inequality is that $\frac{1}{\sqrt{1+x}} \leq \frac{1}{\sqrt{x}}$. But this does not tell us anything! The right-hand-side diverges, and so does not impose its will on the left-hand-side. Instead, we recognize that

$$\lim_{x\to\infty} \frac{\sqrt{x}}{\sqrt{1+x}} = 1.$$

Thus by the Limit Comparison Test, since $\int_1^\infty \frac{1}{\sqrt{x}}\,\mathrm{d}x$ diverges, we necessarily have that $\int_1^\infty \frac{1}{\sqrt{1+x}}\,\mathrm{d}x$ diverges as well. ■

**Example 7.46**

Determine whether $\displaystyle\int_1^\infty \frac{x^2+1}{x^4+3x^2-4x+1}\,\mathrm{d}x$ converges or diverges.

*Solution.* With a strong enough argument, one might be able to argue this example using the Basic Comparison Test, but the Limit Comparison Test proves much simpler. Again the idea is to look at how the numerator and denominator grow asymptotically. The numerator grows as $x^2$, while the denominator grows as $x^4$, meaning that the combined system grows as $1/x^2$. To invoke the Limit comparison Test, we must compute the limit of the ratio of these functions:

$$\lim_{x\to\infty} \frac{x^2+1}{x^4+3x^2-4x+1}\frac{1}{1/x^2} = \lim_{x\to\infty} \frac{x^4+x^2}{x^4+3x^2-4x+1} = 1.$$

Since $\int_1^\infty \frac{1}{x^2}\,\mathrm{d}x$ converges (by Proposition 7.40), we conclude by the Limit Comparison Test that

$$\int_1^\infty \frac{x^2+1}{x^4+3x^2-4x+1}\,\mathrm{d}x \text{ also converges.}$$ ■

# 8 Sequences and Series

The remainder of this text will seem tangential to that which came before, but exists in the same vein. In particular, our focus digresses from objects on the continuum, like functions of real numbers, and instead concentrates on discrete objects. We will draw the appropriate analogies between differential/integral calculus as we progress.

## 8.1 Sequences

### 8.1.1 Definition

An *infinite sequence* in $\mathbb{R}$ is any ordered collection of real numbers. Alternatively, a sequence is any function $a(n)$ which has domain $\mathbb{N}$ and outputs a real number. For example, the function $a(n) = 1/n$ is such a sequence, and we have

$$a(1) = 1, \quad a(2) = \frac{1}{2}, \quad a(3) = \frac{1}{3}, \quad a(4) = \frac{1}{4}, \ldots .$$

Notationally, we often write

$$a_1 = 1, \quad a_2 = \frac{1}{2}, \quad a_3 = \frac{1}{3}, \quad a_4 = \frac{1}{4}, \ldots ,$$

with a general element of the sequence written as $a_n = a(n)$. When we wish to refer to the entire sequence, and not just a particular element, we write $(a_n)_{n=1}^\infty$. On occasion we will abuse this notation, and conflate $a_n$ with $(a_n)_{n=1}^\infty$.

Some sequences look like functions on $\mathbb{R}$, except they are restricted to the natural numbers. For example,

1. $a_n = 2^n$,
$$a_1 = 2, \quad a_2 = 4, \quad a_3 = 8, \quad a_4 = 16, \ldots$$
2. $a_n = \sqrt{n}$,
$$a_1 = 1, \quad a_2 = \sqrt{2}, \quad a_3 = \sqrt{3}, \quad a_4 = 2, \ldots$$
3. $a_n = \log(n)$
$$a_1 = 0, \quad a_2 = \log(2), \quad a_3 = \log(3), \quad a_4 = \log(4), \ldots$$
4. $a_n = \sin\left(\frac{\pi}{2}n\right)$,
$$a_1 = 1, \quad a_2 = 0, \quad a_3 = -1, \quad a_4 = 0, \ldots$$
5. $a_n = \frac{e^n}{n+1}$,
$$a_1 = \frac{e}{2}, \quad a_2 = \frac{1}{3}e^2, \quad a_3 = \frac{1}{4}e^3, \quad a_4 = \frac{1}{5}e^4, \ldots$$

are all sequences which correspond to real functions. Since sequence are effectively functions,

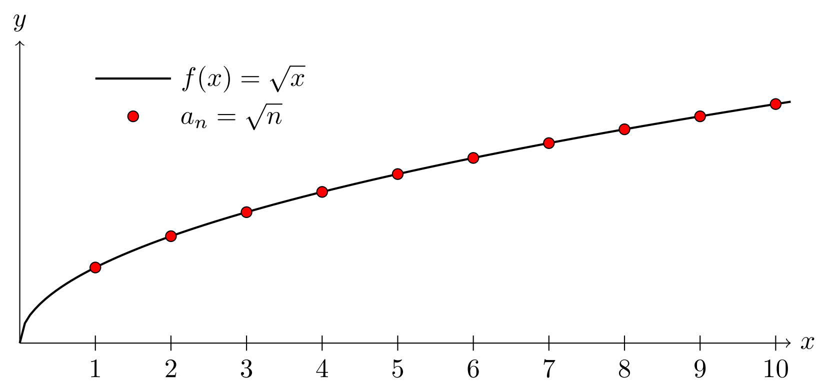



Figure 65: Some sequences arise as the restriction of our usual functions to the natural numbers. In this case, the function $f(x) = \sqrt{x}$ restrict to the sequence $a_n = \sqrt{n}$.

**Definition 8.1**

If $(a_n)_{n=1}^{\infty}$ is a sequence then we say that

1. $a_n$ is *increasing* if $a_n < a_{n+1}$ and *non-decreasing* if $a_n \leq a_{n+1}$,
2. $a_n$ is *decreasing* if $a_n > a_{n+1}$ and *non-increasing* if $a_n \geq a_{n+1}$,
3. $a_n$ is *bounded above* if there exists $M \in \mathbb{R}$ such that $a_n \leq M$ for all $n \in \mathbb{N}$,
4. $a_n$ is *bounded below* if there exists $m \in \mathbb{R}$ such that $a_n \geq m$ for all $n \in \mathbb{N}$,
5. $a_n$ is *bounded* if it is bounded above and below.

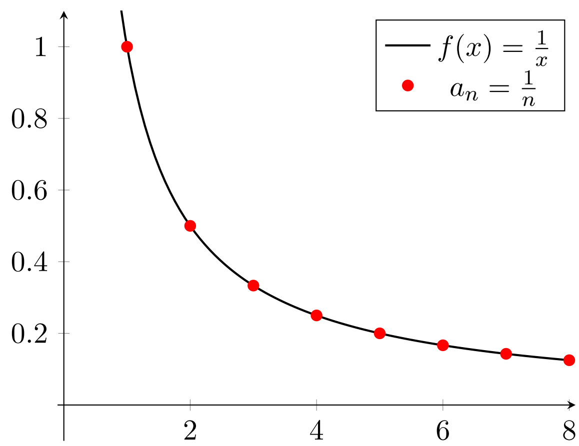



Figure 66: The sequence $a_n = 1/n$ can be recognized as the restriction of the function $f(x) = 1/x$ to the positive integers. This sequence is decreasing and is bounded.

**Example 8.2**

Show that the sequence $a_n = n + \sin(n)$ is a non-decreasing sequence, bounded from below but not from above.

*Solution.* Evaluating $\sin(n)$ for an arbitrary integer is not a simple task; however, since we can think of $a_n$ as the restriction of the function $f(x) = x + \sin(x)$, we can use differential calculus to help solve our problem. Notice that $f'(x) = 1 + \cos(x) \geq 0$, which means that $f$ is a non-decreasing function. In turn, the sequence $a_n$ must then also be non-decreasing (convince yourself of this).

Since the function is non-decreasing, it is bounded from below by its first term: $a_1 = 1 + \sin(1)$. To see that it is not bounded above, we have that for any $M > 0$ the element $a_{M+1} = M + \sin(M) \geq M$, showing that $a_n$ grows unbounded. ∎

**Exercise:** One can convince his/herself that if $f(x)$ is an increasing function, then the sequence defined by $a_n = f(n)$ is also increasing. A similar result holds for decreasing and bounded. However, the converse is certainly not true. Give an example of a sequence $a_n$ and a real function $f$ such that $a_n = f(n)$, $a_n$ is increasing, but $f$ is not increasing. Give a similar example in the case when $a_n$ is bounded but $f$ is not.

**Recursively Defined Sequences:** There are sequences which have no (obvious) relationship to functions on $\mathbb{R}$, and hence must really be thought of purely in terms of a function on $\mathbb{N}$. The prototypical example of a sequence with no $\mathbb{R}$-function equivalent[11]is the sequence $a_n = n!$, whose

[11]There *is* a function-equivalent of the factorial sequence, though it is not quite amenable to the type of analysis we would like to apply. The function is given by improper integral

$$a_n = \Gamma(n+1) = \int_0^\infty x^n e^{-x} dx.$$

first few terms are given by

$$a_1 = 1,\ a_2 = 2,\ a_3 = 6,\ a_4 = 24,\ a_5 = 120,\ a_6 = 720,\ a_7 = 5040, \ldots.$$

Another class of example are those functions which are defined *recursively*. For example, defining $a_1 = 1$ and setting $a_n = \sqrt{a_{n-1} + 1}$ yields the sequence[12]

$$a_1 = 1,\ a_2 = \sqrt{2},\ a_3 = \sqrt{\sqrt{2} + 1},\ a_4 = \sqrt{\sqrt{\sqrt{2}+1}+1}, \ldots$$

or equivalently

$$a_1 = \sqrt{1},\ a_2 = \sqrt{1+\sqrt{1}},\ a_3 = \sqrt{1+\sqrt{1+\sqrt{1}}}, a_4 = \sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1}}}}, \ldots$$

A rather famous example is given by the *Fibonacci sequence*. Define $a_0 = a_1 = 1$ and set $a_n = a_{n-1} + a_{n-2}$ for $n \geq 2$. The first few terms of this sequence are given by[13]

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots.$$

We will not concern ourselves too much with recursively defined sequences, but it is worth noting that they exist and are an interesting area of study in their own right.

### 8.1.2 Limits of Sequences

Treating sequences as functions, and especially realizing the analogy of a sequence to a function on $\mathbb{R}$, allows us to talk about the notion of convergence of a sequence. Roughly speaking, we will say that a sequence converges if it behaves as though the function $a(n)$ has a horizontal asymptote. If we have this idea in our mind, we can immediately adapt the notion of a horizontal asymptote for real functions to sequences:

**Definition 8.3**

A sequence $a_n$ is said to *converge to a limit* $L \in \mathbb{R}$ if, by taking $n$ sufficiently large, we can make $a_n$ arbitrarily close to $L$. In this case we write

$$\lim_{n\to\infty} a_n = L, \qquad \text{or} \qquad (a_n) \xrightarrow{n\to\infty} L.$$

If no such $L$ exists, we say that the sequence *diverges*.

For example, the sequence $a_n = 1/n$ converges. As $n$ becomes very large, $1/n$ becomes very small, so the limit is

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{n} = 0.$$

[12] This is a very important example of a recursive sequence. To see why, consider Exercise 8.1.3

[13] Again, there is a real function which defines this sequence: If $\varphi_\pm = \frac{1}{2}\left(1 \pm \sqrt{5}\right)$ then the function $F$ defined on $[1, \infty)$ given by $F(x) = \frac{1}{\sqrt{5}}\left[\varphi_+^x - \varphi_-^x\right]$ restricts to the Fibonacci sequence.

On the other hand, the sequence $a_n = \sin(\frac{\pi}{2}n)$ diverges, since

$$a_1 = 1, \quad a_2 = 0, \quad a_3 = -1, \quad a_4 = 0, \quad a_5 = 1, \ldots$$

continues in this pattern forever, failing to approach any one particular number. The sequence $a_n = n^2$ also does not converge, since

$$a_1 = 1, \quad a_2 = 4, \quad a_3 = 9, \quad a_4 = 16, \quad a_5 = 25, \ldots$$

grows unbounded.

The previous two sequences demonstrate that a sequence can fail to converge in two different ways. The sequence $a_n = \sin\left(\frac{\pi}{2}n\right)$ failed to converge because because it simply did not approach any number, while $a_n = n^2$ went off to infinity. In this latter case, we will sometimes write

$$\lim_{n\to\infty} a_n = \infty \quad \text{or} \quad (a_n) \to \infty.$$

This does *not* mean that the sequence converges, we are merely specifying the precise nature with which the sequence diverges.

Notice that $a_n = 1/n$ converges to zero, just as the function $f(x) = 1/x$ did. This behaviour generalizes quite nicely.

**Theorem 8.4**

Let $f(x)$ be a function on $\mathbb{R}$ and define a sequence $a_n = f(n)$.

$$\text{If } \lim_{n\to\infty} f(x) = L \qquad \text{then} \qquad \lim_{n\to\infty} a_n = L.$$

Many of the Limit Laws we saw in Theorem 2.6 hold for sequences as well:

**Theorem 8.5: Limit Laws for Sequences**

Let $(a_n) \to L$ and $(b_n) \to M$ be convergent sequences and $c \in \mathbb{R}$ be a constant.

1. The sequence $(a_n + b_n)$ is convergent and $(a_n + b_n) \to L + M$,
2. The sequence $(a_n b_n)$ is convergent and $(a_n b_n) \to LM$,
3. The sequence $(ca_n)$ is convergent and $(ca_n) \to cL$,
4. If $M \neq 0$ then the sequence $(a_n/b_n)$ converges and $(a_n/b_n) \to L/M$.

**Example 8.6**

Show that the sequence $a_n = \dfrac{n^2+1}{2n^2 - 7n + 3}$ converges, and determine its limit.

*Solution.* We can proceed in precisely the same manner we would for functions on $\mathbb{R}$. As the $n^2$ term grows fastest in the limit $n \to \infty$, we renormalize by dividing the numerator and denominator by this factor.

$$\begin{aligned}\lim_{n\to\infty} \frac{n^2+1}{2n^2-7n+3} &= \lim_{n\to\infty} \frac{n^2+1}{2n^2-7n+3}\frac{1/n^2}{1/n^2} \\ &= \lim_{n\to\infty} \frac{1+1/n^2}{2+7/n+3/n^2} \\ &= \frac{1}{2}.\end{aligned}$$

■

**Example 8.7**

Show that the sequence $a_n = \dfrac{e^n + e^{-n}}{4e^n - 3n^2}$ converges and determine its corresponding limit.

*Solution.* The term which grows fastest is $e^n$, so we normalize by this term to get

$$\begin{aligned}\lim_{n\to\infty} \frac{e^n+e^{-n}}{4e^n-3n^2} &= \lim_{n\to\infty} \frac{e^n+e^{-n}}{4e^n-3n^2}\frac{e^{-n}}{e^{-n}} \\ &= \lim_{n\to\infty} \frac{1+e^{-2n}}{4-3e^{-n}n^2} \\ &= \frac{1}{4}.\end{aligned}$$

Here the student should check that $e^{-n}n^2 \xrightarrow{n\to\infty} 0$. ■

**Example 8.8**

Determine whether the sequence $a_n = \dfrac{1-n^3}{n^2+3n+1}$ converges.

*Solution.* The term which grows fastest is $n^3$, so normalizing gives

$$\begin{aligned}\lim_{n\to\infty} \frac{1-n^3}{n^2+3n+1} &= \lim_{n\to\infty} \frac{1-n^3}{n^2+3n+1}\frac{1/n^3}{1/n^3} \\ &= \lim_{n\to\infty} \frac{1/n^3-1}{1/n+3/n^2+1/n^3}.\end{aligned}$$

In the limit as $n \to \infty$ our numerator converges to 1, while our denominator converges to 0. The combination of these two things tells us that our sequence diverges. In fact, for large $n$ notice that the numerator is negative and the denominator is positive, so

$$\lim_{n\to\infty} \frac{1-n^3}{n^2+3n+1} = -\infty.$$

■

**Example 8.9**

Determine whether the sequence $a_n = n[2+\cos(n)]$ converges.

*Solution.* We claim that this sequence does not converge. Indeed, since $2 + \cos(n) \geq 1$ for all $n$ we have that

$$|a_n| = |n[2 + \cos(n)]| \geq |n|.$$

which means $a_n$ cannot possibly be bounded, either above or below. Hence $(a_n)_{n=1}^{\infty}$ is divergent. ■

### 8.1.3 Theorems on Convergent Sequences

**Theorem 8.10: Squeeze Theorem for Sequences**

Assume that for sufficiently large $k$ we have $a_n \leq b_n \leq c_n$. If $(a_n) \to L$ and $(c_n) \to L$, then $(b_n)$ is also convergent with limit $L$.

**Example 8.11**

Show that the sequence $a_n = (-1)^n/n^2$ converges and determine its limit.

*Solution.* In absolute value, we have

$$0 \leq a_n \leq |a_n| = \left|\frac{(-1)^n}{n^2}\right| = \frac{1}{n^2}.$$

Setting $a_n = 0$ and $c_n = 1/n^2$, both sequences converge to 0. By the Squeeze Theorem, we conclude that

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{(-1)^n}{n^2} = 0.$$ ■

**Example 8.12**

Show that the sequence $a_n = \sin(n)/n$ converges.

*Solution.* Since $|\sin(n)| \leq 1$ for all $n$, we have

$$-\frac{1}{n} \leq \frac{\sin(n)}{n} \leq \frac{1}{n}.$$

Setting $a_n = -1/n$ and $c_n = 1/n$, both limits converge to 0. By the Squeeze Theorem,

$$\lim_{n\to\infty} \frac{\sin(n)}{n} = 0.$$ ■

There are a plethora of ways of defining continuous functions, and now we have the tools to introduce yet another.

**Theorem 8.13**

A real function $f(x)$ is continuous if and only if whenever $a_n \to L$ then $f(a_n) \to f(L)$.

One often says that 'continuous functions preserve convergent sequences' or that 'continuous functions map convergent sequences to convergent sequences.' Effectively, one can take limits inside of continuous functions.

**Example 8.14**

Show that if $a > 0$ then $\lim_{n\to\infty} \sqrt[n]{a} = 1$.

*Solution.* Certainly the function $f(x) = a^x$ is continuous, so by Theorem 8.13 we know that

$$\lim_{n\to\infty} a^{1/n} = \lim_{n\to\infty} f(1/n) = f\left(\lim_{n\to\infty} \frac{1}{n}\right) = f(0).$$

But $f(0) = a^0 = 1$, so

$$\lim_{n\to\infty} \sqrt[n]{a} = f(0) = 1$$

as required. ■

Using Theorem 8.4 we can adapt L'Hôpital's rule to sequences. If $a_n = f(n)/g(n)$ where $f(x)$ and $g(x)$ are differentiable functions such that either

- $f(x), g(x) \xrightarrow{x\to\infty} \infty$, or
- $f(x), g(x) \xrightarrow{x\to\infty} 0$,

then

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{f'(n)}{g'(n)}.$$

**Example 8.15**

Determine the limit of the sequence $a_n = \dfrac{n}{e^n}$.

*Solution.* Our limit is indeterminate of type $\infty/\infty$. By L'Hôpital's rule, we have

$$\lim_{n\to\infty} \frac{n}{e^n} \overset{\langle \text{L'H} \rangle}{=} \lim_{n\to\infty} \frac{1}{e^n} = 0.$$ ■

**Example 8.16**

Determine the limit of the sequence $a_n = n^{1/n}$.

*Solution.* Taking logarithms, we have $\log(a_n) = \log(n)/n$, which is indeterminate of type $\infty/\infty$. By L'Hôpital's rule we have

$$\lim_{n\to\infty} \log(a_n) = \lim_{n\to\infty} \frac{\log(n)}{n} \overset{\langle \text{L'H} \rangle}{=} \lim_{n\to\infty} \frac{1/n}{1} = 0.$$

This tells us that $\log(a_n) \to 0$, but we want the limit of $a_n$. Appealing to Theorem 8.13 we have

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \exp(\log(a_n)) = \exp\left(\lim_{n\to\infty} \log(a_n)\right) = \exp(0) = 1. \qquad \blacksquare$$

**Theorem 8.17**

Every convergent sequence is bounded.

The converse of Theorem 8.17 is not true, for the function $a_n = (-1)^n$ is bounded but not convergent. The problem is that $a_n$ is allowed to oscillate. We can get a partial converse by adding the additional criterion that the sequence is either increasing or decreasing.

**Theorem 8.18: Monotone Convergence Theorem**

If $(a_n)$ is bounded from above and non-decreasing, then $(a_n)$ is convergent.

**Exercise:** Using the Monotone Convergence Theorem, can you show that if $(a_n)$ is bounded from below and non-increasing then $(a_n)$ is convergent?

**Example 8.19**

Determine whether the sequence $a_n = \dfrac{2^n}{n!}$ is convergent.

*Solution.* We have

$$\frac{a_{n+1}}{a_n} = \frac{2^{n+1}}{(n+1)!}\frac{n!}{2^n} = \frac{2}{n+1},$$

so that if $n \geq 2$ we have $a_{n+1} < a_n$ and the sequence is decreasing. It is easy to see that $a_n$ is always positive, and so by the Monotone Convergence Theorem we know that $(a_n)$ converges. $\blacksquare$

**Example 8.20**

Consider the sequence defined by $a_1 = 1$ and $a_{n+1} = \dfrac{n}{2n+1}a_n$. Determine whether this limit converges and if so, find the limit.

*Solution.* We begin by noticing that this sequence is decreasing, since $\frac{a_{n+1}}{a_n} = \frac{n}{2n+1} < 1$ which implies that $a_{n+1} < a_n$. Furthermore, the sequence is bounded below by 0; that is, we claim that every $a_n$ is positive. This follows quite quickly by induction, for clearly $a_1 = 1 > 0$ and $a_{n+1} = \frac{n}{2n+1}a_n > 0$ by the induction hypothesis. By (Corollary to) the Monotone Convergence Theorem, $(a_n)$ converges, say to $L$. Now taking the limit as $n \to \infty$ of the equation $a_{n+1} = \frac{n}{2n+1}a_n$ we get

$$L = \frac{1}{2}L$$

for which the only possible real value of $L$ is $L = 0$. $\blacksquare$

**Exercise:** Consider the sequence $a_1 = 1$ and $a_n = \sqrt{1 + a_{n-1}}$. Show that this sequence is increasing and bounded from above. Conclude that it converges. Show that the limit of this sequence is $L = \frac{1+\sqrt{5}}{2}$: the golden ratio.

## 8.2 Infinite Series

### 8.2.1 Definition

Section 5.1 introduced sigma notation as a convenient way of writing down finite sums; namely,

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + \cdots + a_{n-1} + a_n.$$

The summand here conveniently looks like a sequence, though with finitely many terms. In fact, we will be extending the notion of the finite sum above to an infinite sum, called an *infinite series*.

**Definition 8.21**

Let $(a_n)_{n=1}^{\infty}$ be a sequence. Define the $n^{\text{th}}$ *partial sum* of $(a_n)$ to be

$$S_n = \sum_{k=1}^{n} a_k.$$

We say that the infinite series $\displaystyle\sum_{n=1}^{\infty} a_n$ *converges* if the sequence $(S_n)_{n=1}^{\infty}$ converges as a sequence, and we set

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} S_n.$$

We say that the infinite series *diverges* otherwise.

Notice that $S_n = S_{n-1} + a_n$. For example, if we let $a_n = \frac{1}{n}$, then our first few partial sums are given by

$$\begin{aligned} S_1 &= 1 & S_2 &= S_1 + \frac{1}{2} = 1 + \frac{1}{2} = \frac{3}{2} \\ S_3 &= S_2 + \frac{1}{3} = \frac{11}{6} & S_4 &= S_3 + \frac{1}{4} = \frac{25}{12} \\ S_5 &= S_4 + \frac{1}{5} = \frac{137}{60} & S_6 &= S_5 + \frac{1}{6} = \frac{49}{20}. \end{aligned}$$

It is not at all clear from looking at the partial sums whether or not this sequence converges. Shortly, we will study techniques that will allow us to determine whether series converge (though we will rarely be able to determine the limit to which the series converges).

Here is an important distinction to make, as students often confuse sequences and series:

- A *sequence* is a collection of elements $(a_n)_{n=1}^\infty$. Note that $a_n$ need not be in any way related to $a_{n-1}$. The analogy to keep in mind that is sequences behave like functions.
- A *series* is a sum of elements of a sequence, with the partial sums being related via $S_n = S_{n-1} + a_n$. The analogy to keep in mind is that a sum is like an improper integral of a function.

This analogy between sequences/series and functions/integrals is an important one to keep in mind.

Since the convergence of a series is defined in terms of the limits of its partial sums (a sequence), the limit laws for sequences immediately give us the following result:

**Theorem 8.22**

Let $\displaystyle\sum_{k=1}^\infty a_n$ and $\displaystyle\sum_{k=1}^\infty b_n$ be convergent series.

1. The sum of the series is convergent and
$$\sum_{k=1}^\infty (a_n + b_n) = \sum_{k=1}^\infty a_n + \sum_{k=1}^\infty b_n.$$
2. For any $c > 0$, we have
$$\sum_{k=1}^\infty (ca_n) = c\sum_{k=0}^\infty a_n.$$

*Proof.* Let $(s_n) \to L$ be the partial sums of $\sum_{k=1}^\infty a_n$ and $(t_n) \to M$ be the partial sums for $\sum_{k=1}^\infty b_n$. The partial sum of the sum is given by

$$u_n = \sum_{k=1}^\infty (a_n + b_n) = \sum_{k=1} a_n + \sum_{k=1} b_n = s_n + t_n.$$

By Theorem 8.5, we have $(u_n) = (s_n + t_n) \to L + M$, and so

$$\sum_{k=1}^\infty (a_n + b_n) = \sum_{k=1}^\infty a_n + \sum_{k=1}^\infty b_n.$$

The proof for (2) is similar. □

### 8.2.2 Some Special Series

**Geometric Series:** A geometric series is a series defined by a sequence $(a_n)$ where $a_n = ra_{n-1}$ for some $r \in \mathbb{R}$. For example,

$$a_0 = 1,\ a_1 = \frac{1}{2},\ a_2 = \frac{1}{4},\ a_3 = \frac{1}{8},\ a_4 = \frac{1}{16}, \ldots$$

satisfies the relation $a_n = \frac{1}{2}a_{n-1}$. We can write such series as

$$\sum_{k=0}^{\infty} ar^k,$$

where $a = a_1$ and $r = a_n/a_{n-1}$ for any $n$.

**Theorem 8.23**

For any $|r| < 1$ we have

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r} \tag{8.1}$$

and the series diverges otherwise.

*Proof.* The partial sums are given by $S_n = 1 + r + r^2 + \cdots + r^n$. Multiplying by $1 - r$ we get

$$\begin{aligned}(1-r)S_n &= 1 + r + r^2 + r^3 + r^4 + \cdots + r^n \\ &\qquad - r - r^2 - r^3 - r^4 + \cdots - r^n - r^{n+1} \\ &= 1 - r^{n+1}.\end{aligned}$$

Solving for $S_n$ gives $s_n = \dfrac{1 - r^{n+1}}{1-r}$. Since $|r| < 1$, we have that $r^{n+1} \xrightarrow{n\to\infty} 0$, so

$$\sum_{k=0}^{\infty} r^k = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{1 - r^{n-1}}{1-r} = \frac{1}{1-r}.$$ □

An interesting if non-rigorous argument that proves the same result is to see that

$$\begin{aligned}(1-r)(1 + r + r^2 + r^3 + \cdots) &= 1 + r + r^2 + r^3 + \cdots \\ &\qquad - r - r^2 - r^3 - \cdots \\ &= 1.\end{aligned}$$

In an entirely formal sense (that is, treating $r$ purely as a symbol without assigning it any value) we see that $(1 - r)$ is the multiplicative inverse of $(1 + r + r^2 + \cdots)$, giving the desired results as well. Determining from this which values of $r$ actually make sense is something that we will see in a later chapter.

**Example 8.24**

Determine the limit $\displaystyle\sum_{k=2}^{\infty} \frac{1}{2^k}$.

*Solution.* Notice that our summation index begins at 2, and not 0 as in (8.1). We can fix this by realizing that

$$\sum_{k=2}^{\infty} \frac{1}{2^k} = \sum_{k=0}^{\infty} \frac{1}{2^k} - 1 - \frac{1}{2} = \frac{1}{1 - 1/2} - 1 - \frac{1}{2} = 2 - 1 - \frac{1}{2} = \frac{1}{2}.$$ ■

**Example 8.25**

Determine the value of the series $\displaystyle\sum_{n=0}^{\infty} \frac{(-2)^n}{e^n}$.

*Solution.* Here our ratio is $r = -2/e$. Since $e \approx 2.7182$ we know $|r| < 1$. By (8.1) we have

$$\sum_{n=0}^{\infty} \frac{(-2)^n}{e^n} = \frac{1}{1-(-2/e)} = \frac{e}{e+2}. \qquad \blacksquare$$

**Example 8.26**

Determine the value of the series $\displaystyle\sum_{k=10}^{\infty} \frac{-3}{4^k}$.

*Solution.* The $-3$ can be pulled outside of the sum. As our index does not start at 0 we could subtract the first 10 terms, but this is rather onerous. Instead, realize that by writing

$$\begin{aligned}
\sum_{k=10}^{\infty} \frac{1}{4^k} &= \frac{1}{4^{10}} + \frac{1}{4^{11}} + \frac{1}{4^{12}} + \frac{1}{4^{13}} + \cdots \\
&= \frac{1}{4^{10}}\left(\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots\right) \\
&= \frac{1}{4^{10}} \sum_{k=0}^{\infty} \frac{1}{4^k}.
\end{aligned}$$

Putting this all together we get

$$\sum_{k=10}^{\infty} \frac{-3}{4^k} = \frac{-3}{4^{10}} \sum_{k=0}^{\infty} \frac{1}{4^k} = \frac{-3}{4^{10}} \frac{1}{1-1/3} = -\frac{9}{2 \cdot 4^{10}} = -\frac{9}{2^{21}}. \qquad \blacksquare$$

**Telescoping Series:** Telescoping series are ones in which many of the internal summands cancel one another out. For example, a telescoping sum arises when we consider the Riemann sum of a constant function. If $f(x) = c$ and we partition $[a, b]$ into $n$ equal subintervals $[x_i, x_{i+1}]$, then the Riemann sum is

$$L(f) = \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1}) = c \sum_{i=1}^{n} (x_i - x_{i-1}).$$

Notice what happens if we expand out a few terms of the sum: we get

$$(x_1 - x_0) + (x_2 - x_1) + (x_3 - x_2) + \cdots + (x_{n-1} - x_{n-2}) + (x_n - x_{n-1}).$$

For that majority of terms, both $x_i$ and $-x_i$ appear in the summation and hence cancel one another out. The only terms which survive are the first and last, so that the sum reduces to $x_n - x_0$.

$$U_f(P) = c(x_n - x_0) = c(b-a)$$

exactly as expected.

**Example 8.27**

Determine the value of the series $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$ if it converges.

*Solution.* The trick is to realize our sum as a telescoping series. Using partial fractions, we get that

$$\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$$

so that the $n^{\text{th}}$ partial fraction is given by

$$\begin{aligned} S_n &= \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{\infty} \left[\frac{1}{k} - \frac{1}{k+1}\right] \\ &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) + \left(\frac{1}{n} - \frac{1}{n+1}\right) \\ &= 1 - \frac{1}{n+1} = \frac{n}{n+1}. \end{aligned}$$

We now take the limit as $n \to \infty$ and we find that

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k(k+1)} = \lim_{k\to\infty} \frac{k}{k+1} = 1. \qquad \blacksquare$$

## 8.3 Convergence Tests

The analogy between infinite series and improper integrals continues in this section with the introduction of the basic and limit comparison test. Later, we will develop several more tests which allow us to infer convergence of infinite series.

Our first result really tests divergence rather than convergence. Intuitively, one should expect that we require $(a_n) \to 0$ to even have a chance for the infinite series to converge. Otherwise, if $(a_n) \to L \neq 0$ then the infinite series would effectively be 'adding $L$ to itself infinitely many times,' which would certainly yield a non-finite number. The following is the more precise way of saying this:

**Theorem 8.28**

If $\displaystyle\sum_{k=1}^{\infty} a_k$ converges, then $(a_k) \to 0$.

The contrapositive of this is exactly what we mentioned above: "If $(a_k) \not\to 0$ then $\sum_{k=1}^{\infty} a_k$ does not converge." However, the converse of Theorem 8.28 is not true. We do not yet have the techniques to show that this is the case, but the series

$$\sum_{k=1}^{\infty} \frac{1}{k} \quad \text{does not converge,}$$

despite the fact that $1/k \xrightarrow{k\to\infty} 0$.

**Example 8.29**

Show that the series $\displaystyle\sum_{k=0}^{\infty} \sin\left(\frac{\pi k}{2}\right)$ diverges.

*Solution.* Setting $a_k = \sin\left(\frac{\pi k}{2}\right)$, we know that $a_k$ alternates between the values $\{0, \pm 1\}$ and fails to converge at all, let alone 0. By Theorem 8.28 we conclude that

$$\sum_{k=0}^{\infty} \sin\left(\frac{\pi k}{2}\right) \quad \text{diverges.} \qquad \blacksquare$$

**Example 8.30**

Determine whether the sequence $\displaystyle\sum_{k=0}^{\infty} \frac{k^2}{k^2+1}$ converges.

*Solution.* Setting $a_k = k^2/(k^2+1)$ we see that

$$\lim_{k\to\infty} a_k = \lim_{k\to\infty} \frac{k^2}{k^2+1} = 1.$$

Since this limit is not zero, the series cannot possibly converge, by Theorem 8.28. $\blacksquare$

### 8.3.1 The Integral Test

This tests takes our series/integral analogy to the next level, but allowing us to directly compare them.

**Theorem 8.31: The Integral Test**

If $f$ is a continuous, non-negative, decreasing function on $[1, \infty)$ then

$$\sum_{k=1}^{\infty} f(k) \text{ converges} \quad \Leftrightarrow \quad \int_1^{\infty} f(x)\,\mathrm{d}x \text{ converges.}$$

The idea here is by appropriate choosing our Riemann sums, we can bound the series by the integrals. This is shown in Figure 67.

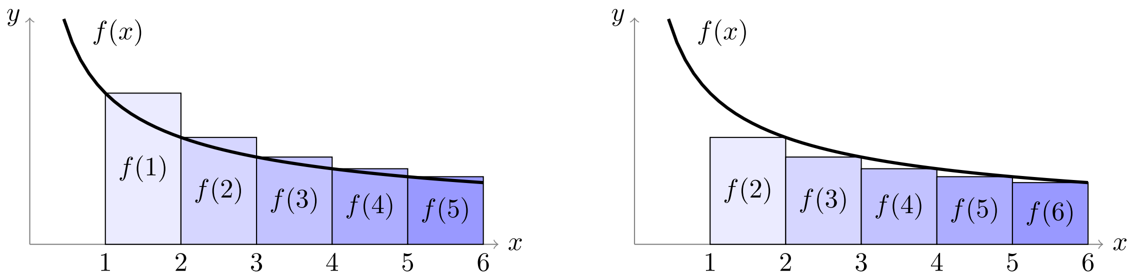



Figure 67: Using Riemann sums, we can over- and under-estimate the values of the series.

**Example 8.32**

Consider the series whose terms are given by $a_n = \dfrac{1}{n}$. Determine the convergence/divergence of this series.

*Solution.* This is what is called the *Harmonic Series* and is the classic example of how to trick first year students into making mistakes. It is clear that $a_n \to 0$ as $n \to \infty$, but $\sum a_n$ diverges. Indeed, the integral test tells us that $\sum \frac{1}{n}$ converges if and only if $\int_1^n \frac{1}{x}\, dx$ converges, but

$$\int_1^n \frac{1}{x} dx = \log(x)\Big|_1^\infty = \infty.$$

However, regardless of how many times I emphasize the point that "Just because the sequence converges to zero, does not mean the series converges" there are countless students who get this wrong! Do not make the same mistake, and learn this classic example immediately. ■

**Example 8.33**

Determine whether the following series converges: $\displaystyle\sum_{n=0}^\infty e^{-n}$.

*Solution.* Here, the terms of the series are given by the function $f(x) = e^{-x}$. Applying the integral test, we have

$$\int_1^\infty e^{-x} dx = -e^{-x}\Big|_1^\infty = \frac{1}{e}$$

so $\sum e^{-n}$ converges. Note that since the index occurs as a power, we can also apply the root test. Indeed, setting $a_n = e^{-n}$ we have that $\sqrt[n]{e^{-n}} = e^{-1}$. Since $e^{-1} < 1$ it then follows that the series converges. ■

**Corollary 8.34: The $p$-test**

The series $\displaystyle\sum_{n=1}^\infty \frac{1}{n^p}$ converges if and only if $p > 1$.

*Proof.* By the Integral Test (Theorem 8.31) and the $p$-test (Propostion 7.33), we know that

$$\sum_{k=1}^{\infty} \frac{1}{n^p} \text{ converges} \quad \overset{\text{Theorem 8.31}}{\Leftrightarrow} \quad \int_1^{\infty} \frac{1}{x^p}\,\mathrm{d}x \text{ converges} \quad \overset{\text{Proposition 7.33}}{\Leftrightarrow} \quad p > 1. \qquad \square$$

From the $p$-test we can immediately conclude, for example, that $\sum_{k=1}^{\infty} n^{-2}$ converges. Note however that this does not tell us how to determine the value of this series. In fact,

$$\sum_{k=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6},$$

but this lies well beyond the scope of these notes.

### 8.3.2 Comparison Tests

In Section 7.5.3 we saw the Basic Comparison Test (Theorem 7.41) and Limit Comparison Test (Theorem 7.44) for improper integrals. The statement of these tests for series are almost identical.

**Theorem 8.35: Basic Comparison Test for Series**

Suppose that $(a_n)$ and $(b_n)$ are sequences such that $0 \leq a_k \leq b_k$ for sufficiently large $k$.

1. If $\sum_k b_k$ converges, then $\sum_k a_k$ converges.
2. If $\sum_k a_k$ diverges, then $\sum_k b_k$ diverges.

The idea here is precisely the same as that of Theorem 7.41: If the larger of the two series converges, it forces the smaller sequence to also converge. In the opposite direction, if the smaller sequence diverges, then the larger sequence must also diverge.

**Example 8.36**

Determine whether the series $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n-1}$ converges or diverges.

*Solution.* Notice that $n > n-1$ so that

$$\frac{1}{n} \leq \frac{1}{n-1}.$$

The series over the left hand side diverges, as it is the Harmonic Series. By the Basic Comparison Test (2), we know that

$$\sum_{n=2}^{\infty} \frac{1}{n-1} \text{ diverges.}$$

$\blacksquare$

**Example 8.37**

Determine whether $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^3+k^2+1}$ converges.

*Solution.* Note that $k^2+1 \geq 0$ for all $k$, so $k^3+k^2+1 \geq k^3$. Taking reciprocals reverses the inequality, giving

$$\frac{1}{k^3+k^2+1} \leq \frac{1}{k^3}.$$

The series over the right hand side converges by the $p$-test, so by the Basic Comparison Test (1),

$$\sum_{k=1}^{\infty} \frac{1}{k^3+k^2+1} \text{ converges.}$$

■

**Example 8.38**

Determine whether the sequence $\displaystyle\sum_{k=1}^{\infty} \frac{7^n}{8^n+2}$ converges.

*Solution.* By considering the terms which grow fastest in this expression, we suspect that the 2 in the denominator should not affect the long term behaviour of the function. We thus have the comparison

$$\frac{7^n}{8^n+2} \leq \frac{7^n}{8^n} = \left(\frac{7}{8}\right)^n.$$

Now the series $\sum_k \left(\frac{7}{8}\right)^n$ is a geometric series with ratio less than 1, and hence converges. We conclude by the Basic Comparison Test the $\sum_k \frac{7^n}{8^n+2}$ converges as well. ■

There are very simple circumstances where the Basic Comparison Test fails, when it morally feels like it should succeed. For example, consider the simple series

$$\sum_{n=1}^{\infty} \frac{1}{2n^2-1}.$$

It feels like we should be able to compare this to the series $\sum 1/n^2$, but when we try to do our analysis we get

$$\frac{1}{n^2} \leq \frac{1}{n^2-1}.$$

Since the left hand side converges, the Basic Comparison Test tells us nothing. The key to resolving this is the Limit Comparison Test.

**Theorem 8.39: Limit Comparison Test for Series**

If $(a_n)$ and $(b_n)$ are sequences with positive terms, and $\lim_{n\to\infty} \frac{a_k}{b_k} = L$ for some $0 < L < \infty$ then

$$\sum_k a_k \text{ converges} \quad \Leftrightarrow \quad \sum_k b_k \text{ converges.}$$

The fact that $a_k/b_k$ converges to some positive, non-zero number means that $a_k$ and $b_k$ grow at 'approximately the same speed.' Hence convergence or divergence of one will immediately imply convergence/divergence of the other.

**Example 8.40**

Determine whether the sequence $\displaystyle\sum_{k=0}^{\infty} \frac{k^{10} + 25k^7 + 1}{k^{12} - 20}$ converges or diverges.

*Solution.* We only care about the terms which grow the fastest, so to form our comparison function we just consider the biggest factors in each of the numerator and the denominator. Those terms are $k^{10}/k^{12} = 1/k^2$ and since we know that the series $\sum 1/k^2$ converges, it seems likely that so too will our series above. Indeed, this actually gives us the other series we should use in our Limit Comparison Test. Notice that

$$\begin{aligned}
\lim_{k\to\infty} \frac{\frac{1}{k^2}}{\frac{k^{10}+25k^7+1}{k^{12}-20}} &= \lim_{k\to\infty} \frac{k^{12} - 20}{k^{12} + 25k^9 + k^2} \frac{1/k^{12}}{1/k^{12}} \\
&= \lim_{k\to\infty} \frac{1 - 20/k^{12}}{1 + 25 1/k^2 + 1/k^{10}} \\
&= 1.
\end{aligned}$$

Thus the Limit Comparison Test tells us that the series converges. ∎

**Example 8.41**

Determine whether the series $\displaystyle\sum_{k=1}^{\infty} \frac{n + \sqrt{n}}{n^2 - 3}$ converges or diverges.

*Solution.* Taking the largest terms in the numerator and denominator, we compare to $n/n^2 = 1/n$.

Indeed, this is a good comparison function since

$$\begin{aligned}\lim_{n\to\infty} \frac{1/n}{\frac{n+\sqrt{n}}{n^2-2}} &= \lim_{n\to\infty} \frac{n^2-2}{n^2-n\sqrt{n}} \\ &= \lim_{n\to\infty} \frac{1-2/n^2}{1-1/\sqrt{n}} \\ &= 1.\end{aligned}$$

Since $\sum_n 1/n$ diverges, by the Limit Comparison Test we conclude that

$$\sum_{k=1}^\infty \frac{n+\sqrt{n}}{n^2-3} \text{ diverges.} \qquad \blacksquare$$

**Example 8.42**

If $a_n$ are such that $\displaystyle\sum_{k=0}^\infty \frac{1}{a_n}$ converges, show that $\displaystyle\sum_{k=0}^\infty \frac{1}{a_n+M}$ converges for any $M>0$.

*Solution.* Since constants neither shrink nor grow, we do not expect them to affect convergence of the limit, and we can use both the Basic Comparison Test.

We realize that $a_n < a_n + M$ so that $\frac{1}{a_n} > \frac{1}{a_n+M}$. Thus $\sum \frac{1}{a_n+M} < \sum \frac{1}{a_n}$ and $\sum \frac{1}{a_n}$ converges, so $\sum \frac{1}{a_n+M}$ will also converge.

On the other hand, by using the Limit Comparison Test we can actually extend this result to work for any $M \in \mathbb{R}$. Recall that since $\sum \frac{1}{a_n}$ converges, it must be the case that $\frac{1}{a_n} \to 0$ as $n \to \infty$. The Limit Comparison Test then tells us that

$$\begin{aligned}\lim_{n\to\infty} \frac{1/a_n}{1/a_n+M} &= \lim_{n\to\infty} \frac{a_n+M}{a_n} \\ &= \lim_{n\to\infty} 1 + \frac{M}{a_n} = 1\end{aligned}$$

and so again we conclude that $\sum \frac{1}{a_n+M}$ converges. $\blacksquare$

### 8.3.3 Alternating Series and Absolute Convergence

**Definition 8.43**

If $(a_n)$ is a sequence of non-negative numbers, an *alternating series* is any series of the form

$$\sum_k (-1)^k a_k.$$

For example, the following are alternating series:

- $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^k}{k} = -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \cdots$
- $\displaystyle\sum_{k=1}^{\infty} (-1)^{k-1} k^2 = 1 - 4 + 9 - 16 + 25 - 36 + \cdots$

Luckily, there is a very simple test to determine whether an alternating series converges:

**Theorem 8.44: Alternating Series Test**

If $(a_k)$ is a positive decreasing sequence, the alternating series $\sum_k (-1)^k a_k$ converges if and only if $\lim_{k\to\infty} a_k = 0$.

**Example 8.45**

Determine whether $\displaystyle\sum_{k=2}^{\infty} (-1)^k \frac{k^2}{3k^3 - 7}$ converges.

*Solution.* This is an alternating series. Setting $a_k = k^2/(3k^3 - 7)$ we have

$$\lim_{k\to\infty} \frac{k^2}{3k^3 - 7} \frac{1/k^3}{1/k^3} = \lim_{k\to\infty} \frac{1/k}{3 - 7/k^3} = 0,$$

so by the Alternating Series Test, the series converges. ■

**Example 8.46**

Determine whether $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$ converges.

*Solution.* By taking a negative sign out of the summation, our series is equivalent to $-\sum_k \frac{(-1)^k}{k}$, and this is an alternating series with $a_k = \frac{1}{k}$. These $a_k$ are clearly limiting to zero, and hence by the alternating series test, the limit converges. ■

Notice that the alternating series given in Example 8.46 is closely related to the Harmonic series which we actually know does not converge, and is often called the *alternating harmonic series*. In fact, we will (very luckily) be able to determine the exact value of the alternating harmonic series after we have seen Taylor series.

**Absolute Convergence:** We begin with a definition:

**Definition 8.47**

A series $\sum_k a_k$ is said to *converge absolutely* if $\sum_k |a_k|$ converges. If $\sum_k a_k$ converges but $\sum_k |a_k|$ does not, then the series is said to *converge conditionally.*

For example, the Alternating Harmonic Series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$$

converges conditionally. Indeed, Example 8.46 showed that the series converged, but it's absolute value is

$$\sum_{k=1}^{\infty} \left| \frac{(-1)^{k+1}}{k} \right| = \sum_{k=1}^{\infty} \frac{1}{k}$$

is the Harmonic Series, which we know does not converge.

On the other hand, the series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k}}{k^2}$$

converges absolutely. It satisfies the Alternating Series Test and so converges, and its absolute value

$$\sum_{k=1}^{\infty} \left| \frac{(-1)^{k}}{k^2} \right| = \sum_{k=1}^{\infty} \frac{1}{k^2}$$

also converges, by the $p$-test.

There is an important relationship between series which converge, and which converge absolutely.

**Theorem 8.48: Absolute Convergence**

If $\sum_k |a_k|$ converges then $\sum a_k$ converges; that is, all absolutely convergent series are convergent.

**Example 8.49**

Determine whether the series $\displaystyle\sum_k \frac{\cos(\pi k)}{k\sqrt{k}}$ converges or diverges.

*Solution.* Notice that $\cos(\pi k) = (-1)^{k+1}$, so that the terms of our series are actually given by $a_k = \frac{(-1)^{k+1}}{k\sqrt{k}}$. In absolute value, we have

$$|a_k| = \frac{1}{k\sqrt{k}}.$$

By the $p$-test, we know that $\sum_k |a_k| = \sum_k k^{-3/2}$ converges. Since our series is absolutely convergent, by our theorem we know that it is actually convergent. ■

### 8.3.4 Non-Comparison Tests

The following tests are intrinsic to the given series. There are many ways in which this is more appealing. For example, the following tests do not require us to formulate a comparison series.

**Ratio Test:** One of the most powerful tests, the ratio test is somewhat similar to the Limit Comparison Test at first glance, but only uses sequential terms of the sequence itself.

**Theorem 8.50: Ratio Test**

Let $(a_n)$ be a sequence such that $\left|\frac{a_{k+1}}{a_k}\right| \xrightarrow{n\to\infty} L$.

1. If $L < 1$ then $\displaystyle\sum_{k=1}^\infty a_k$ converges absolutely (and hence converges),
2. If $L > 1$ then $\displaystyle\sum_{k=1}^\infty a_k$ diverges,
3. If $L = 1$ then the test is inconclusive.

**Example 8.51**

Determine whether the series $\displaystyle\sum_{k=1}^\infty \frac{(2k)!}{k!}$ converges or diverges.

*Solution.* Set $a_k = (2k)!/k!$ so that

$$\begin{aligned}\lim_{k\to\infty}\frac{a_{k+1}}{a_k} &= \lim_{k\to\infty}\frac{(2k+2)!}{(k+1)!}\frac{k!}{(2k)!}\\ &= \lim_{k\to\infty}\frac{(2k+2)(2k+1)}{k+1}\\ &= \infty.\end{aligned}$$

By the Ratio Test, the limit diverges. ■

**Example 8.52**

Determine whether the series $\displaystyle\sum \frac{2^k k!}{k^k}$ converges or diverges.

*Solution.* We might be tempted to use the root test here but the presence of the factorial term suggests that we should use the ratio test instead. Let $a_k = \frac{2^k k!}{k^k}$ so that

$$\begin{aligned}
\lim_{k\to\infty} \frac{a_{k+1}}{a_k} &= \lim_{k\to\infty} \frac{2^{k+1}(k+1)!}{(k+1)^{k+1}} \frac{k^k}{2^k k!} \\
&= \lim_{k\to\infty} \frac{2^{k+1}}{2^k} \frac{(k+1)!}{k!} \frac{k^k}{(k+1)^{k+1}} \\
&= 2 \lim_{k\to\infty} \frac{k^k(k+1)}{(k+1)^{k+1}} \\
&= 2 \lim_{k\to\infty} \left(\frac{k}{k+1}\right)^k \\
&= \frac{2}{e}
\end{aligned}$$

where the last line is an exercise in L'Hôpital's rule. By the Ratio Test, we then know that the series converges. ■

**Example 8.53**

Determine the values of $c > 0$ such that the series $\displaystyle\sum_{k=2}^{\infty} \frac{c^k}{\log(k)}$ converges.

*Solution.* Set $a_k = c^k / \log(k)$ so that

$$\begin{aligned}
\lim_{k\to\infty} \frac{a_{k+1}}{a_k} &= \lim_{k\to\infty} \frac{c^{k+1}}{\log(k+1)} \frac{\log(k)}{c^k} \\
&= c \lim_{k\to\infty} \frac{\log(k)}{\log(k+1)} \\
&= c
\end{aligned}$$

where we have used the fact that $\log(x)/\log(x+1) \xrightarrow{x\to\infty} 1$ (use L'Hôpital). By the Ratio Test, if $c > 1$ then the series diverges, while if $c < 1$ the series converges. When $c = 1$ the Ratio Test is inconclusive. It turns out that the series diverges, but we leave this as an exercise. ■

**Root Test:** The root test is similar in both statement and proof to the ratio test, but is harder to use in all but a few exceptional scenarios.

**Theorem 8.54: The Root Test**

Let $(a_n)$ be a sequence with non-negative terms, such that $|a_k|^{1/k} \xrightarrow{n\to\infty} L$.

1. If $L < 1$ then $\sum a_k$ converges absolutely (and hence converges),
2. If $L > 1$ then $\sum a_k$ diverges,
3. If $L = 1$ then the test is inconclusive.

**Example 8.55**

Determine whether the series $\sum_{k=1}^{\infty} \frac{k^k}{3^{k^2}}$ converges or diverges.

*Solution.* Since everything involves the index in the power, we should probably apply the root test. Indeed, set $a_k = \frac{k^k}{3^{k^2}}$ so that $\sqrt[k]{a_k} = \frac{k}{3^k}$. It is then clear (using L'Hôpital's rule) that $\sqrt[k]{a_k} \to 0$ as $k \to \infty$. By the root test, we then have that the series converges. ∎

**Example 8.56**

Determine whether the series $\sum_{k=1}^{\infty} \frac{3^k + k^9}{k^k}$ converges or diverges.

*Solution.* This one is a bit tricky, as it will require us to use two tests. First of all, we notice that the terms which grow that fastest are given by $\frac{3^k}{k^k}$, and so we use the Limit Comparison Test to see that

$$\lim_{k\to\infty} \frac{3^k + k^9}{k^k}\frac{k^k}{3^k} = \lim_{k\to\infty} 1 + \frac{k^9}{3^k} = 1$$

where the last equality is an exercise in L'Hôpital's rule. Thus $\sum \frac{3^k+k^9}{k^k}$ converges if and only if $\sum \frac{3^k}{k^k}$ converges. Setting $a_k = \frac{3^k}{k^k}$ the root test tells us that

$$\begin{aligned} \lim_{k\to\infty} \sqrt[k]{\frac{3^k}{k^k}} &= \lim_{k\to\infty} \sqrt[k]{\left(\frac{3}{k}\right)^k} \\ &= \lim_{k\to\infty} \frac{3}{k} = 0 \end{aligned}$$

and since this is certainly less than 1, we know that $\sum \frac{3^k}{k^k}$ converges. We conclude that the original series converges as required. ∎

# 9 Power Series

Power series are, in a sense, infinite polynomials of the form

$$\sum_{k=0}^{\infty} c_k (x-a)^k,$$

and have important applications in statistics, combinatorics, algebra, and analysis. To an extent they represent functions: given a value of $x$ we substitute that value of $x$ into the above expression and attempt to evaluate the series. Examples of power series which we will see later include

- $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^{k+1} x^k}{k}$,
- $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k (x-1)^{2k}}{(2k)!}$,
- $\displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k!}$,
- $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k x^k}{2k+1}$.

In this section, we will see how to describe functions in terms of power series.

## 9.1 Radius of Convergence

Whether or not a power series converges depends heavily upon the choice of the coefficient $c_k$, and it is often too much to ask that the series always converges. For example, when $c_k = 1$ for all $k$ the corresponding power series is that of a geometric series $\sum_k x^k$, which only converges when $|x| < 1$. Therefore, we need a way of discussing when a series converges.

**Proposition 9.1**

Consider the power series $\sum_k c_k (x-a)^k$. There exists a unique number $0 \leq R \leq \infty$ such that the power series converges absolutely for all $a - R < x < a + R$ and diverges for all other $x$. This $R$ is called the *radius of convergence*.

This proposition gives us three are three distinct scenarios:

1. $R = 0$ means that our power series only converges at a single point. This is not an enlightening case,
2. $0 < R < \infty$ means that our power series converges, but only within some specific range for the value of $x$,
3. $R = \infty$ means that our power series always converges, regardless of what value of $x$ we use.

We can use the convergence tests that we learned about in Section 8.3.4 to determine radii of convergence, just like in the following example:

**Example 9.2**

Find the radius of convergence for the power series $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k^2 2^k}$.

*Solution.* We would like to determine the values of $x$ such that the series $\sum b_k$ defined by setting

$b_k = \frac{|x|^k}{k^2 2^k}$ converges. Using the ratio test, we have

$$\begin{aligned}\lim_{k\to\infty} \frac{b_{k+1}}{b_k} &= \lim_{k\to\infty} \frac{x^{k+1}}{(k+1)^2 2^{k+1}} \frac{2^k k^2}{x^k} \\ &= \frac{|x|}{2} \lim_{k\to\infty} \left(\frac{k}{k+1}\right)^2 \\ &= \frac{|x|}{2}.\end{aligned}$$

We know that we get convergence when $|x|/2 < 1$ and so $|x| < 2$. ■

Proposition 9.1 only tells us what happens on the open interval $(a-R, a+R)$ but not what happens at the endpoints. If we include the endpoints where our series also converges, we have the *interval of convergence*. Unfortunately, the endpoints must often be computed directly after finding the radius of convergence.

**Example 9.3**

Determine the interval of convergence for the power series

$$\sum_{k=0}^{\infty} \frac{3^k}{k}(2x-4)^k.$$

*Solution.* Set $a_k = \dfrac{3^k}{k}|2x-4|^k$ and apply the ratio test to get

$$\begin{aligned}\lim_{k\to\infty} \frac{a_{k+1}}{a_k} &= \lim_{k\to\infty} \frac{3^{k+1}|2x-4|^{k+1}}{k+1} \frac{k}{3^k|2x-4|^k} \\ &= 3|2x-4| \lim_{k\to\infty} \frac{k}{k+1} \\ &= 3|2x-4|.\end{aligned}$$

Setting $3|2x-4| < 1$ we equivalently get $|x-2| < \frac{1}{6}$, so our radius of convergence if $\frac{1}{6}$, and our interval so far is $\left(\frac{11}{6}, \frac{13}{6}\right)$.

To determine endpoint convergence, we plug these number directly into our series. At $x = \frac{11}{6}$ we get

$$\sum_{k=0}^{\infty} \frac{3^k}{k}\left(2\frac{11}{6} - 4\right)^k = \sum_{k=0}^{\infty} \frac{(-1)^k 3^{k-1}}{k!}.$$

Since $\frac{3^{k-1}}{k!} \xrightarrow{x\to\infty} 0$, this series converges by the Alternating Series Test. On the other hand, at $x = \frac{13}{6}$ we have

$$\sum_{k=0}^{\infty} \frac{3^k}{k}\left(2\frac{13}{6} - 4\right)^k = \sum_{k=0}^{\infty} \frac{3^{k-1}}{k!}$$

which converges by the Ratio Test. Hence the interval of convergence included both endpoints and is $\left[\frac{11}{6}, \frac{13}{6}\right]$. ■

**Example 9.4**

Determine the interval of convergence for the power series $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k!}$.

*Solution.* Set $a_k = x^k/k!$ and using the ratio test we have

$$\begin{aligned}\lim_{k\to\infty} \frac{a_{k+1}}{a_k} &= \lim_{k\to\infty} \frac{|x|^{k+1}}{(k+1)!}\frac{k!}{|x|^k} \\ &= |x| \lim_{k\to\infty} \frac{1}{k} = 0.\end{aligned}$$

This limit is zero regardless of the value of $x$, implying that $R = \infty$ is the radius of convergence. There is no need to check endpoints, since the interval of convergence is all of $\mathbb{R}$. ■

**Example 9.5**

Determine the radius of convergence of the power series $\displaystyle\sum_{k=0}^{\infty} k!x^k$.

*Solution.* Again using the ratio test on $a_k = k!x^k$ we have

$$\begin{aligned}\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| &= \lim_{k\to\infty} \frac{(k+1)!|x|^{k+1}}{k!|x|^k} \\ &= |x| \lim_{k\to\infty} (k+1).\end{aligned}$$

The only way this number is finite is if $|x| = 0$, which implies that $R = 0$ is the radius of convergence, and the power series converges at precisely one point. ■

On their interval of convergence, power series define functions. For example, in Example 9.4 we have a function

$$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

whose domain is the interval of convergence $\mathbb{R}$, while from Example 9.5 we have a function

$$g(x) = \sum_{k=0}^{\infty} k!x^k$$

whose domain is the singleton set $\{0\}$.

## 9.2 Differentiation and Integration of Power Series

Since power series are very similar to polynomials, it seems natural that we should try to differentiate and integrate them. A question which naturally arises is whether the radius of convergence can possibly change via the process of integration or differentiation.

**Proposition 9.6**

If $\sum_{k=0}^\infty c_k(x-a)^k$ is a power series with radius of convergence $0 \leq R \leq \infty$, then the differentiated and integrated power series

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^\infty c_k(x-a)^k = \sum_{k=0}^\infty \frac{\mathrm{d}}{\mathrm{d}x}\left[c_k(x-a)^k\right] = \sum_{k=1}^\infty kc_k(x-a)^{k-1}, \qquad \text{and}$$

$$\int_a^x \sum_{k=0}^\infty c_k(t-a)^k\,\mathrm{d}t = \sum_{k=0}^\infty \int_a^x c_k(t-a)^k\,\mathrm{d}t = \sum_{k=0}^\infty \frac{c_k}{k+1}(x-a)^{k+1}$$

both have radii of convergence $R$ as well.

We note however that this theorem *does not* say that the interval of convergence is preserved. Indeed, it is sometimes the case that we lose an endpoint in the process of integrating and differentiating.

**Example 9.7**

Show that the power series $\displaystyle\sum_{k=1}^\infty \frac{(-1)^k x^k}{k}$ has a different interval of convergence than its differentiated power series.

*Solution.* If $c_k = (-1)^k x^k/k$ then the Ratio Test quickly tells us that

$$\lim_{n\to\infty}\left|\frac{c_{n+1}}{c_n}\right| = |x|\lim_{n\to\infty}\frac{k}{k+1} = |x|,$$

so that the series has radius of convergence one. Checking the endpoints $x = \pm 1$ we have

$$\sum_{k=1}^\infty \frac{(-1)^k x^k}{k}\bigg|_{x=-1} = \sum_{k=1}^\infty \frac{1}{k}, \qquad \sum_{k=1}^\infty \frac{(-1)^k x^k}{k}\bigg|_{x=1} = \sum_{k=1}^\infty \frac{(-1)^k}{k}$$

the first of which does not converge, while the latter does converge. We conclude that the interval of convergence is $(-1, 1]$.

On the other hand, the differentiated power series has the form $\sum_k (-1)^k x^{k-1}$. By Proposition 9.6 this series also has radius of convergence 1, which the reader can quickly verify using the Ratio Test. Evaluating at the endpoints $x = \pm 1$ yields

$$\sum_{k=1}^\infty (-1)^k, \qquad \sum_{k=1}^\infty (-1)^{k+1}$$

neither of which converge. Thus the interval of convergence for the differentiated power series is $(-1,1)$, which is a subset of $(-1,1]$. ■

## 9.3 Taylor Series

It seems odd to define a function using a power series as above. One might wonder if any of these power series functions are in fact functions that we have seen before. In this section, we will see how to take functions with which we are familiar and turn them into power series.

### 9.3.1 Taylor Polynomials

As a prelude to discussing an infinite series which depends on a variable $x$, we should start by discussing the finite dimensional case: a polynomial. Our goal then is as follows: given a function $f$, find a polynomial $p_n$ of degree $n$ such that $p_n$ is a good approximation to $f$ at the point $a = 0$.

To proceed, we will demand that the error in our approximation vanishes faster than the order of the polynomial we use. For example, if we use an order $n$ polynomial $p(x)$ to approximate the function $f(x)$, then we should demand that error $\epsilon(x) = f(x) - p(x)$ of our approximation vanish to order $n+1$. This can be written as

$$\lim_{x\to 0} \frac{f(x)-p(x)}{x^n} = 0.$$

If $f(x)$ can be differentiated sufficiently often, applying L'Hôpital's rule turns this statement into

$$f^{(k)}(0) = p^{(k)}(0), \quad k = 1,\ldots,n.$$

Taking this as our template, we should try to specify our polynomial $p(x)$ to have its first $n$ derivatives agree with $f(x)$.

To see how to formulate the theory, write

$$p_n(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0 = \sum_{k=0}^n a_k x^k.$$

Our goal is to choose the $a_k$ such that $p_n^{(k)}(0) = f^{(k)}(0)$. Starting this process, notice that

$$\begin{aligned} p_n(0) &= a_0 \\ p_n'(0) &= a_1 \\ p_n''(0) &= 2a_2 \\ p_n^{(3)}(0) &= 3!a_3 \\ &\vdots \\ p_n^{(k)}(0) &= k!a_k. \end{aligned}$$

Hence if we want $f^{(k)}(0) = p_n^{(k)}(0) = k!a_k$ we should set $a_k = \dfrac{f^{(k)}(0)}{k!}$, and our polynomial is thus

$$p_n(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n = \sum_{k=0}^n \frac{f^{(k)}(0)}{k!}x^k.$$

Now what happens if we want to perform this expansion around a more general point $a$ instead of just 0? Instead of doing more work, we just translate our $p_n$ above by the term $x-a$. More precisely, consider the polynomial

$$p_{n,a}(x) = a_n(x-a)^n + a_{n-1}(x-a)^{n-1} + \cdots + a_2(x-a)^2 + a_1(x-a) + a_0$$

for which we have $p_n^{(k)}(a) = k! a_k$. We would like to set this to $f^{(k)}(a)$, meaning that we should take $a_k = \frac{f^{(k)}(a)}{k!}$, and our polynomial becomes

$$p_{n,a}(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

This is called the $n^{th}$ *order Taylor polynomial at $a$*.

**Example 9.8**

Determine the $n^{\text{th}}$ order Taylor polynomial of $f(x) = e^x$ at $x=0$.

*Solution.* We are well familiar with the fact that $f^{(k)}(x) = e^x$ and so $f^{(k)}(0) = e^0 = 1$. Thus the Taylor polynomial is

$$p_{n,0}(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^{n} \frac{x^k}{k!}. \qquad \blacksquare$$

**Example 9.9**

Determine the $n^{\text{th}}$ order Taylor polynomial of $f(x) = \log(1+x)$ at $x=0$.

*Solution.* This requires that we determine a general form for the derivative $f^{(k)}(x)$. Checking our first few derivatives, we find that

$$\begin{aligned}
f(x) &= \log(1+x) & f(0) &= 0 \\
f'(x) &= \frac{1}{1+x} & f'(0) &= 1 \\
f''(x) &= -\frac{1}{(1+x)^2} & f''(0) &= -1 \\
f^{(3)}(x) &= \frac{2!}{(1+x)^3} & f^{(3)}(0) &= 2! \\
&\vdots \\
f^{(k)}(x) &= \frac{(-1)^{k-1}(k-1)!}{(1+x)^k} & f^{(k)}(0) &= (-1)^{k-1}(k-1)!
\end{aligned}$$

The veracity of this most general form can be checked by induction, and is left as an exercise for the student. Hence the $n^{\text{th}}$-order Taylor polynomial is

$$\begin{aligned} p_{n,0}(x) &= \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^{n} \frac{(-1)^{k-1}(k-1)!}{k!} x^k \\ &= \sum_{k=0}^{n} \frac{(-1)^{k-1}}{k} x^k. \end{aligned}$$

■

**Example 9.10**

Compute the Taylor polynomial of $x\sin(x)$ about 0.

*Solution.* It will quickly become laborious for us to differentiate the function $f(x) = x\sin(x)$, and in fact it will be effectively impossible to derive a closed form expression for a generic $n$-th derivative. Instead, the Taylor polynomial of $x$ is just $x$ itself, so we need only compute the Taylor polynomial of $\sin(x)$.

We know that

$$\begin{array}{ll} \sin(0) = 0 & \left.\frac{\mathrm{d}^2}{\mathrm{d}x^2}\right|_{x=0} \sin(x) = -\sin(x)|_{x=0} = 0 \\ \left.\frac{\mathrm{d}}{\mathrm{d}x}\sin(x)\right|_{x=0} = \cos(x)|_{x=0} = 1 & \left.\frac{\mathrm{d}^3}{\mathrm{d}x^3}\right|_{x=0} = -\cos(x)|_{x=0} = -1 \end{array}.$$

Since the derivatives of sine repeat with periodicity 4, this is all we need to compute. In particular though, we notice that only the odd terms will survive, and they will switch sign. By writing our odd numbers as $2n+1$, the sign of the $2n+1$-st derivative is $(-1)^n$ and we get the Taylor polynomial (or order $2n+1$)

$$\sin(x) = \sum_{k=0}^{n} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.$$

The Taylor polynomial of $x\sin(x)$ is thus

$$x\sin(x) = x\sum_{k=0}^{n} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = \sum_{k=0}^{n} \frac{(-1)^n}{(2n+1)!} x^{2n+2}.$$

■

### 9.3.2 Taylor Series

By extending the idea of Taylor polynomials to power series, we get the definition of the *Taylor series* of the function $f$ about the point $x = a$. In particular, consider the following result:

**Theorem 9.11**

Let $f(x)$ be an infinitely differentiable function and consider the power series defined by

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

If $I$ is the interval of convergence of this power series, then for every $x \in I$ we have

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

That is, on its interval of convergence, a function is equal to the value of its Taylor series.

We say that the function $f$ is *analytic* at the point $c$ if the Taylor series at $f$ has a non-zero radius of convergence at $c$. If $f$ is analytic at every $c \in \mathbb{R}$, we just say that $f$ is analytic.

One must be careful to realize that infinitely differentiable is not the same thing as analytic. Define the function

$$f(x) = \begin{cases} e^{-\frac{1}{x}} & x > 0 \\ 0 & x \leq 0 \end{cases}.$$

This function is infinitely differentiable. However, it is not analytic at $x = 0$. With a bit of work, one can show that the Taylor series of $f(x)$ at 0 has radius of convergence 0.

1. $e^x = \displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k!}$, interval of convergence $\mathbb{R}$.
2. $\sin(x) = \displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$, $\quad \cos(x) = \displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$, interval of convergence $\mathbb{R}$.
3. $\dfrac{1}{1-x} = \displaystyle\sum_{k=0}^{\infty} x^k$, interval of convergence $(-1, 1)$.
4. $\log(1+x) = \displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^k}{k}$, interval of convergence $(-1, 1]$.
5. $\arctan(x) = \displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2k+1}$, interval of convergence $(-1, 1]$.

An easy way to remember property (2): we know $\sin(x)$ is odd as a function, and indeed its Taylor series only has powers in odd order. Similarly, $\cos(x)$ is even as a function and its Taylor series only has powers of even order. We are not yet in a position to compute (5), but we provide it here as an interesting and motivating example which we will use (and prove) later.

It is worth noting that I am cheating here. Theorem 9.11 only tells us that the function is equal to its Taylor series on the interior of the interval of convergence and not at the endpoints.

It is another theorem to show that if convergence holds at the endpoints then the function still converges there, but let's sweep that under the rug for now.

**Exercise:** Notice how the Taylor series for $e^x$ looks very similar to the Taylor series for $\sin(x)$ and $\cos(x)$. Indeed, if we ignore the $(-1)^k$ terms we see that $\sin(x)$ contains the odd terms of $e^x$ while $\cos(x)$ contains the even terms. If we can account for these minus signs, maybe we can get a relationship between *a priori* unrelated exponential function and the trigonometric functions.

1. Recall that in the complex numbers, we can define a number $i$ such that $i^2 = -1$. Using their Taylor series expansions, show that $e^{ix} = \cos(x) + i\sin(x)$.

2. We know that every point on the unit circle can be written as $(\cos(\theta), \sin(\theta))$ where $\theta$ is the angle between the point and the positive $x$-axis. If we think about $\cos(x) + i\sin(x)$ as $(\cos(x), \sin(x))$ convince yourself that $e^{ix}$ sweeps out the unit circle as $x$ traverses through $[0, 2\pi)$.

## 9.4 Differentiation and Integration of Taylor Series

So far we have seen that functions have power series representations on their radius of convergence. Moreover, Theorem 9.6 told us that we could integrate and differentiate power series while preserver their radius of convergence.

Combining these two results means that we we can interchange the summation sign with the derivative/integral. Indeed,

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x) = \frac{\mathrm{d}}{\mathrm{d}x} \sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{\infty} a_k \left[ \frac{\mathrm{d}}{\mathrm{d}x} x^k \right] = \sum_{k=1}^{\infty} k a_k x^{k-1},$$

and similarly for integration. This is not the same thing as linearity, and we should not assume that this result follows immediately. We encourage the student to accept the theorem as stated, but not to underestimate the subtlety of the situation.

**Example 9.12**

Using the fact that $\dfrac{1}{1-x} = \displaystyle\sum_{k=0}^{\infty} x^k$ on $(-1, 1)$, show that $\log(1+x) = \displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k-1} x^k}{k}$ on $(-1, 1)$.

*Solution.* Note that the Taylor series for $1/(1+x)$ can be derived from $1/(1-x)$ by substituting $-x$ into the Taylor series:

$$\frac{1}{1+x} = \frac{1}{1-(-x)} = \sum_{k=0}^{\infty} (-x)^k = \sum_{k=0}^{\infty} (-1)^k x^k.$$

Now we know that

$$\int_0^x \frac{1}{1+t} \, \mathrm{d}t = \log(1+x)$$

so applying a term-by-term integration, we get

$$\begin{aligned}\log(1+x) &= \int_0^x \frac{1}{1+t}\,\mathrm{d}t = \int_0^x \sum_{k=0}^{\infty}(-1)^k t^k\,\mathrm{d}t \\ &= \sum_{k=0}^{\infty}(-1)^k \int_0^x t^k\,\mathrm{d}t = \sum_{k=0}^{\infty}(-1)^k \left.\frac{t^{k+1}}{k+1}\right|_{t=0}^{t=x} \\ &= \sum_{k=1}^{\infty}\frac{(-1)^{k-1}x^k}{k}\end{aligned}$$

where in the last line we have re-indexed the sum. ■

**Example 9.13**

Confirm that the Taylor series of $\arctan(x)$ is given by $\displaystyle\sum_{k=0}^{\infty}\frac{(-1)^k x^{2k+1}}{2k+1}$.

*Solution.* We know that $\arctan(x) = \displaystyle\int_0^x \frac{1}{1+t^2}dx$, and we know the Taylor series for $\frac{1}{1+t^2}$ since it is given by

$$\frac{1}{1+t^2} = \frac{1}{1-(-t^2)} = \sum_{k=0}^{\infty}(-t^2)^k = \sum_{k=0}^{\infty}(-1)^k t^{2k}.$$

Integrating term by term, we get

$$\arctan(x) = \sum_{k=0}^{\infty}\int_0^x (-1)^k t^{2k}dt = \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{2k+1}$$

as required. ■

## 9.5 Applications of Taylor Series

**Calculators and calculating:** One of the historic applications of power series is the ability to compute the value of a function to arbitrary precision in terms of polynomials. In the case of transcendental functions especially (such as $\log, \sin, \cos, \exp$), this is how modern day computers and calculators often compute values.

**Computing Limits and Asymptotics:** Mathematicians often cheat when looking at limits, in the sense that we can often guess that value of the limit if we know a little something about the Taylor series of the terms involved.

Before proceeding, let us introduce a little notation. If $f(x)$ and $g(x)$ are functions, we say that that $f(x) = o(g(x))$ if $f(x)$ and $g(x)$ grow at the same rate; that is, $f(x)$ and $g(x)$ shared the same asymptotics in the limit as $f(x)$ and $g(x)$ go to infinity. In a similar vein, we may define 'big-O' notation as describing the asymptotics of $f(x)$ and $g(x)$ when they go to zero [14]. In effect, we will write $f(x) = O(g(x))$ if $f(x)$ is at worst $g(x)$ in the limit as $x \to 0$. For example, the Taylor series of $e^x$ is

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

If we are playing with numbers very near zero, and are only interested in what is happening up to the quadratic term, we might write this as

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

where here $O(x^3)$ indicates that the dominant factor in the error is of order $x^3$ (for $x$ near 0, the term $x^4$ will be much smaller than $x^3$, and so on). When taking limits as $x \to 0$, anything with $O(x^k)$ $k \geq 1$ will vanish, meaning that we are only interested in the constant terms.

Now we often used Taylor series to create L'Hôpital problems, such as the following:

**Example 9.14**

Consider the limit $\displaystyle\lim_{x\to 0} \frac{e^x - 1 - x}{x^2}$.

*Solution.* When we made this example, we used the fact that we knew what the Taylor series for $e^x$. Indeed, the numerator becomes

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

Hence applying this to the limit, we have

$$\begin{aligned}
\lim_{x\to 0} \frac{e^x - 1 - x}{x^2} &= \lim_{x\to 0} \frac{\frac{x^2}{2!} + O(x^3)}{x^2} \\
&= \lim_{x\to 0} \left[\frac{1}{2!} + O(x)\right] \\
&= \frac{1}{2}.
\end{aligned}$$

One can check this is the same answer we get by applying L'Hôpital. ■

**Example 9.15**

(Example 4.24) Determine the limit $\displaystyle\lim_{x\to 0} \frac{xe^{nx} - x}{1 - \cos(nx)}$ for $n \neq 0$.

[14] There is also a version of big-O for infinite asymptotics.

*Solution.* Using Taylor series, we have

$$\begin{aligned} xe^{nx} - x &= x\left[1 + (nx) + \frac{(nx)^2}{2} + O(x^3)\right] - x \\ &= nx^2 + \frac{n^2x^3}{2} + O(x^4). \end{aligned}$$

On the other hand, the denominator is

$$1 - \cos(nx) = 1 - \left[1 - \frac{(nx)^2}{2} + O(x^4)\right] = \frac{n^2x^2}{2} + O(x^4).$$

Taking the limit, we have

$$\begin{aligned} \lim_{x\to 0} \frac{xe^{nx} - x}{1 - \cos(nx)} &= \lim_{x\to 0} \frac{nx^2 + O(x^3)}{\frac{n^2x^2}{2} + O(x^4)} \\ &= \lim_{x\to 0} \frac{n + O(x)}{\frac{n^2}{2} + O(x^2)} \\ &= \frac{2}{n}, \end{aligned}$$

and this is exactly what we found in Example 4.24. ∎

**Difficult Integrals:** We have seen a few examples of functions that cannot be integrated, and with some more advanced theory (called differential Galois theory), one can actually prove that some of these integrals do not have anti-derivatives that can be expressed with elementary functions. However, the simplicity of integrating polynomials means that, so long as we are content with a power series representation, one can still get a closed form expression for these integrals.

**Example 9.16**

Using power series, integrate $\displaystyle\int \frac{\sin(x)}{x} dx$.

*Solution.* The power series for $\frac{\sin(x)}{x}$ is given by

$$\begin{aligned} \frac{\sin(x)}{x} &= \frac{1}{x}\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k+1)!}. \end{aligned}$$

Since we can integrate term-by-term, we get

$$\begin{aligned} \int \frac{\sin(x)}{x} dx &= \sum_{k=0}^{\infty} \int \frac{(-1)^k x^{2k}}{(2k+1)!} dx \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)^2(2k)!}. \end{aligned}$$ ∎

**Example 9.17**

Using power series, integrate $\int_0^x e^{-x^2} dx$.

*Solution.* The power series for $e^{-x^2}$ is given by

$$e^{-x^2} = \sum_{k=0}^{\infty} \frac{(-x^2)^k}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{k!}.$$

Integrating term-by-term, we get

$$\begin{aligned} \int_0^x e^{-t^2} dt &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^x t^{2k} dt \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)k!}. \end{aligned}$$

■

**Differential Equations:** In a similar vein to integrals, one can solve differential equations using Taylor series. One begins by assuming that the solution can be written as a power series, then uses the differential equation to determine a relationship amongst the coefficients.

**Example 9.18**

Use a power series to find a solution to $y'' + y = 0$.

*Solution.* We have already seen that the most general solution to this equation is $y(x) = A\cos(x) + B\sin(x)$ for variables $A$ and $B$. We expect that we should arrive at the same solution using power series. Indeed, suppose that we can write $y = \sum_k a_k x^k$, so that

$$y = \sum_{k=0}^{\infty} a_k x^k \qquad y' = \sum_{k=1}^{\infty} k a_k x^{k-1}$$

$$y'' = \sum_{k=2}^{\infty} k(k-1) a_k x^{k-2}$$

Plugging these into our differential equation, we get

$$\begin{aligned} 0 = y'' + y &= \left[\sum_{k=2}^{\infty} k(k-1)a_k x^{k-2}\right] + \left[\sum_{k=0}^{\infty} a_k x^k\right] \\ &= \left[\sum_{k=0}^{\infty} (k+2)(k+1)a_{k+2} x^k\right] + \left[\sum_{k=0}^{\infty} a_k x^k\right] && \text{re-labelling the first sum to start at 0} \\ &= \sum_{k=0}^{\infty} \left[(k+2)(k+1)a_{k+2} + a_k\right] x^k \end{aligned}$$

Now the power series is zero if and only if each coefficient is identically zero. This means that we can solve for $a_{k+2}$ in terms of $a_k$ as

$$a_{k+2} = -\frac{a_k}{(k+2)(k+1)}.$$

Hence once we have determine $a_0$ and $a_1$, we can find all the remaining coefficients. Set $a_0 = A$ and $a_1 = B$ and notice that

$$\begin{array}{ll} a_2 = -\frac{A}{2\cdot 1} & a_3 = -\frac{B}{3\cdot 2} \\ a_4 = \frac{-a_2}{4\cdot 3} = \frac{A}{4!} & a_5 = -\frac{a_3}{5\cdot 4} = \frac{B}{5!} \\ \quad\vdots & \quad\vdots \\ a_{2n} = (-1)^n \frac{A}{(2n)!} & a_{2n+1} = (-1)^n \frac{B}{(2n+1)!} \end{array}$$

By splitting the even and odd powers of the power series, our solution thus looks like

$$y = A\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} + B\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} = A\cos(x) + B\sin(x),$$

exactly as we expected. ■