

My name is Leon Hostetler. I am currently a student at Florida State University majoring in physics as well as applied and computational mathematics. Feel free to download, print, and use these class notes. If you find them useful, consider [buying me a coffee](#).

All of my class notes can be found at www.leonhostetler.com/classnotes

Please bear in mind that these notes will contain errors. If you find one, please email me at leonhostetler@gmail.com with the name of the class notes, the page on which the error is found, and the nature of the error. If you include your name, I will probably list your name on the thank you page if I decide to compile and sell my notes.

This work is currently licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. That means you are free to copy and distribute this document in whole for noncommercial use, but you are not allowed to distribute derivatives of this document or to copy and distribute it for commercial reasons.

Last Updated: August 25, 2016

Chapter 1

Calculus

This covers the typical three semester (year and a half) calculus sequence taken by mathematics and science majors.

Calculus is a challenging and 3-semester long project. It is an endeavor not to be taken lightly. Because this is just the first of a three-semester learning project, extra care must be taken that every detail is understood and learned properly. Otherwise, your learning in the next two semesters will suffer.

1.1 Limits

1.1.1 Tangent

Calculus basically consists of the area problem (integral calculus) and the tangent problem (differential calculus). In a way, these two problem types are the inverse of each other. The tangent problem is the problem of getting a tangent to a function curve when only one point is known. It is the problem of finding the limit of a function as a point on that function is approached. The area problem seeks to find the limit of the area under a curve by summing the areas of increasingly small rectangles.

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

If position and time are graphed as a function, then average velocity is quite simply calculated as the slope of a secant line of the function. The secant line connects the starting time and the ending time and of course the starting position and the ending position.

Instantaneous velocity is the slope of the tangent line at a single point rather than the slope of a secant line between two points. Calculating this requires calculus.

The slope of the tangent line can be thought of as the limit of the slopes of the secant lines as the additional point of the secant line is brought closer and closer to the tangent point.

If you have a slope and one point, the equation for

The biggest downfall of calculus students is not a failure to grasp the calculus but an inability to fully utilize algebra.

Calculus requires a slightly different way of thinking, but the most difficult aspect remains the algebra. I think the key to becoming comfortable with the different way of thinking is to keep referring to the overall point of calculus—the mathematical study of change.

the tangent line is

$$y_1 - y = m(x_1 - x).$$

Many functions that occur in science aren't described by explicit equations, but rather, by experimental data points. When an explicit equation isn't given with a function (e.g. it is given as a table of points or just a curve), a tangent line can be approximated by taking the average of the two closest secant lines, or by drawing a tangent approximation and measuring the two sides (x and y) of the triangle of which the tangent line is the hypotenuse.

The **difference quotient** represents the average rate of change of $f(x)$ between $x = a$ and $x = a + h$. In other words, it gives the slope of the secant line for the points $(a, f(a))$ and $(a + h, f(a + h))$. The difference quotient giving the slope is

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

A tangent is the line that touches the curve only at that point. Any other line, would, simply by way of the line being a curve, touch the curve at more than one point.

1.1.2 The Limit

The **limit** of a function $f(x)$ at a point $x = a$ is the value L of the function $f(x)$ as it approaches $x = a$. It is

represented by

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

For example, the value of x^2 as x approaches 2 is 4, and so

$$\lim_{x \rightarrow 2} x^2 = 4.$$

In many cases, as above, we can calculate a limit simply by plugging the value being approached into the function. In the example above, we just plug $x = 2$ into x^2 .

A key fact is that a function does not have to be defined at the point that it is approaching. In fact, we never consider $x = a$. The only thing that matters is how $f(x)$ is defined near a .

Tip:

A limit can be thought of as the intended height of a function.

We cannot always figure out a limit by plugging in numbers closer and closer to a . Sometimes, calculators lie when you're trying to estimate a limit by trying x -values very close to a because the calculator isn't able to carry enough digits.

Some functions have one-sided limits in that L is different depending on which side a is being approached from.

The notation $x \rightarrow 0^-$ means that x is approaching 0 from the negative (left) side. The notation $x \rightarrow 0^+$ means that x is approaching 0 from the positive (right) side. So we say that the right-hand limit is

$$\lim_{x \rightarrow a^+} f(x) = L,$$

and it means that L is the limit of $f(x)$ as x approaches $x = a$ from the right.

The limit of a function exists if and only if both one-sided limits exist and agree with each other. Otherwise, the function may not have a limit at that point even though it has left-hand and right-hand limits at that point.

If a function value approaches infinity as x approaches a , the limit is said to NOT EXIST although the particular way in which such a limit does not exist is still represented by the positive or negative infinity symbol. For example, we say that

$$\lim_{x \rightarrow 0} \frac{1}{x^2} = \infty,$$

despite infinity not being a number.

There are multiple ways of finding or at least approximating limits. This gives us a way to check our work.

1. Analytically: using limit laws, deduction, etc.

- Graphically: using a calculator and estimating from a zoomed in graph
- Numerical experimentation: plugging in numbers that are closer and closer to a

Tip:

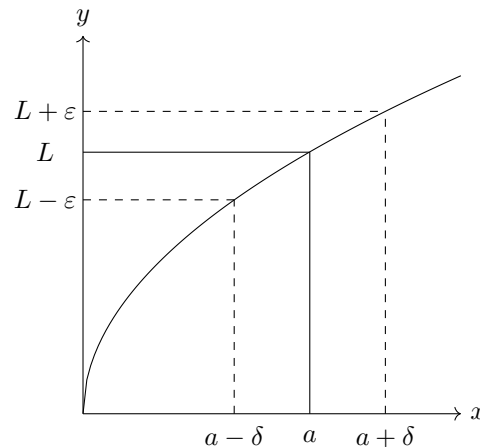
Always double check limits by graphing it on the calculator.

The precise definition of a limit is as follows. If f is a function defined on an open interval (except maybe at a) containing a , then the limit of $f(x)$ as x approaches a is L ,

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

if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$ then $|f(x) - L| < \varepsilon$. This is called the **epsilon-delta definition** of the limit.

Remember that $f(x)$ is the curve, a is the point on the x -axis that we're considering, and L is the intended height of $f(a)$. We can understand the precise definition of the limit as follows. The limit of $f(x)$ as x approaches a is L means that if the distance between x and a is less than δ then the distance between $f(x)$ and L is less than ε .



Delta and epsilon are two numbers such that whenever x is within delta units of a , then $f(x)$ is within epsilon units of L . However, delta and epsilon are not static numbers for any given limit. If delta is made smaller then epsilon must become smaller. The point is that for every possible value of epsilon there exists a corresponding value of delta such that both satisfy the if-then inequality. You can't just find two numbers that make the if-then inequality statement true and then substitute them for delta and epsilon because the definition requires that the if-then inequality statement holds true for all epsilons—not just the convenient value that you've chosen. If a smaller value of delta is chosen, then epsilon must be correspondingly smaller. This means that $f(x)$ can be made arbitrarily close to L by taking x sufficiently close to a .

This definition is used in several kinds of problems.

The first kind of problems involves finding deltas that correspond to given epsilons. A second kind of problem involves proving a limit equation.

To find deltas given epsilons is pretty easy. The goal is simply to find an x_1 and an x_2 such that $f(x_1) = L + \varepsilon$ and $f(x_2) = L - \varepsilon$. If this is unclear, simply graph $f(x)$, L , a , ε , and δ .

There are two ways of solving this. The first method is simply to solve the equations $f(x_1) = L + \varepsilon$ and $f(x_2) = L - \varepsilon$ for x since L and ε are given. The second method involves solving the inequality $L - \varepsilon < f(x) < L + \varepsilon$ for x . This inequality will reduce to $a - \delta < x < a + \delta$ and since a will be given, δ can easily be calculated. A third method, which is used when $f(x)$ is “ugly” or just when wanting to check the work found using other methods, is to use a calculator. You can either use tables or the graphing functionalities. I prefer the graphing method because it gives more significant figures. Graph $f(x)$, then also graph $y = L + \varepsilon$ and $y = L - \varepsilon$. Then use the intersect feature to find the x -value of the intersections of $f(x)$ and the two horizontal lines. The two x -values found will be $a - \delta$ and $a + \delta$.

Note: If the function is non-linear then two different values for δ may be found. Use the smaller one. Also not that the number found isn't equal to δ , it's greater than δ . In other words, $\delta < \min\{\delta_1, \delta_2\}$.

Sometimes, this type of problem is given in a form like

$$\text{If } 0 < |x - 4| < \delta \text{ then } |\sqrt{x} - 2| < 0.4, \text{ find } \delta.$$

To graph such a situation, it is easier to first convert it into a limit equation. The above example converted, gives us

$$\lim_{x \rightarrow 4} \sqrt{x} = 2.$$

Proving limits using the delta-epsilon definition of limits requires two steps

1. Guessing delta. This is the hard part. We use the precise definition of a limit to recast the limit equation in the form of absolute value inequality if-then statements.
2. Proving the limit using the guessed delta. This step involves the actual proof that our if-then absolute value inequalities hold true for the chosen value of delta. Showing this is true is now trivial since were basing the value of delta on the value of epsilon.

Example:

Use the epsilon-delta definition of the derivative to prove that

$$\lim_{x \rightarrow 1} (3x + 1) = 4.$$

We start by converting the limit to the inequality form using the epsilon-delta definition of the limit

$$\text{If } 0 < |x - 1| < \delta \text{ then } |(3x + 1) - 4| < \varepsilon,$$

where δ and ε are some small positive numbers.

Then we start on the conclusion and work backwards. Our goal is to express delta in terms of epsilon, and we start by simplifying the inequality containing epsilon.

$$\begin{aligned} |(3x + 1) - 4| &< \varepsilon \\ 3|x - 1| &< \varepsilon \\ |x - 1| &< \frac{1}{3}\varepsilon. \end{aligned}$$

This suggests that we should choose $\delta = \frac{1}{3}\varepsilon$ since both inequalities have the same $|x - 1|$ term.

In the second step, we use the above value of δ to prove the limit. We have that

$$|(3x + 1) - 4| = 3|x - 1| < 3\delta = 3\left(\frac{1}{3}\varepsilon\right) = \varepsilon.$$

Therefore,

$$\text{If } 0 < |x - 1| < \delta \text{ then } |(3x + 1) - 4| < \varepsilon.$$

1.1.3 Limit Laws

The following limit laws also apply to one-sided limits. Many of these laws can be derived by looking at generic graphs. With these rate laws, $f(x)$ and $g(x)$ are functions, c is a constant, and n is a positive integer.

$$\begin{aligned} \lim_{x \rightarrow a} [f(x) + g(x)] &= \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x) \\ \lim_{x \rightarrow a} [cf(x)] &= c \lim_{x \rightarrow a} f(x) \\ \lim_{x \rightarrow a} [f(x)g(x)] &= \left(\lim_{x \rightarrow a} f(x)\right) \left(\lim_{x \rightarrow a} g(x)\right) \\ \lim_{x \rightarrow a} \frac{f(x)}{g(x)} &= \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)} \\ \lim_{x \rightarrow a} c &= c \\ \lim_{x \rightarrow a} [f(x)]^n &= \left[\lim_{x \rightarrow a} f(x)\right]^n \\ \lim_{x \rightarrow a} \sqrt[n]{f(x)} &= \sqrt[n]{\lim_{x \rightarrow a} f(x)}. \end{aligned}$$

If f is a polynomial or a rational function and a is in the domain of f , simply substitute a into f to get the limit

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

This is the **direct substitution property**. In other words, the first step in finding a limit is to substitute a into the function you're trying to find the limit of. If a is in the domain of f , then this will directly give you the limit of the function at the point a . If you get a divide-by-zero error, then you know that a is not in the domain of the function, therefore, you must use another method to find the limit. When doing polynomial limits always use direct substitution. Since all real numbers are in the domain of polynomials, then this method is the fastest guaranteed method of finding the limit.

When doing rational limits try direct substitution first. If you end up with $\frac{0}{0}$, then you know you have a "hole" at the point a which can be canceled or reduced out of the rational function. Once that is done, you can go on to find the limit. If direct substitution results in some non-zero number over zero, then you know you have a vertical asymptote at the point a which cannot be cancelled out. At that point, you know that the limit is represented as DNE (Does Not Exist) or by positive or negative infinity.

For limits of functions including radicals, you must make sure that a is in the domain of the function and won't result in an imaginary number.

If $f(x) = g(x)$ when x does not equal a , then

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x),$$

provided that the limits exist. In other words, sometimes you can use a simpler function to find the limit of a more difficult one. An example is reducing (getting a simpler function) a rational function (from a more difficult one).

The **squeeze theorem** states that if $f(x) \leq g(x) \leq h(x)$ when x approaches a and

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = L,$$

then

$$\lim_{x \rightarrow a} g(x) = L.$$

Drawing this situation on a graph, makes it obvious that it's true. Sometimes, however, it's not this straightforward to solve a limit using the squeeze theorem.

There are two ways that squeeze theorem problems may be presented. The first is easy. In the first case, you are shown $f(x) \leq g(x) \leq h(x)$ and both $f(x)$ and $h(x)$ are given. Then you must simply find the limits of these two, and if they're the same, then you know that $g(x)$ has the same limit. The second case is when you're presented with a limit equation, and the only way to solve it is using the squeeze theorem. In such a case you can often use properties of the functions to find the limits.

Example:

Find the limit

$$\lim_{x \rightarrow a} x^2 \sin\left(\frac{1}{x}\right).$$

Notice that we can't use the direct substitution rule since that would require division by zero.

The key is to note that the sine function always stays between 1 and -1 . In particular,

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

Multiplying by x^2 gives us

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

Clearly the limits of both the left and right side are 0 as $x \rightarrow 0$. Therefore, by the squeeze theorem,

$$\lim_{x \rightarrow a} x^3 \sin\left(\frac{1}{x}\right) = 0.$$

It's important to know the definition of a polynomial. It's also important to know how to quickly work the difference of squares and the sum and difference of cubes. Must also be able to evaluate absolute values without resorting to graphs and must be able to perform binomial expansions without tediously FOILing them.

1.1.4 Continuity

A function f is continuous at a if

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

In other words, if function is at its *intended height* then it is continuous at that point.

For a function to be continuous at a it must fulfill all the following conditions

- It must be defined at $f(a)$ (if not, this causes a removable or an infinite discontinuity)
- The limit must exist, and (if not, this causes a gap discontinuity)
- $\lim_{x \rightarrow a} f(x) = f(a)$ (if not, this causes a jump discontinuity)

To prove continuity or discontinuity, you must prove that all three conditions hold true (continuity) or show which of them don't (discontinuity).

In other words, a hole or a gap means a function is discontinuous at that point. If asked why a function is discontinuous show which one of the three given conditions do not hold true.

Example:

Show that $f(x) = \sqrt{x-1}$ is continuous on $[1, \infty)$.

We know that $f(x)$ is defined on $[1, \infty)$. To prove that it is continuous, we have to show that its limit at any point in its domain is equal to its value. We begin by showing that the limit exists and is equal to its value in $(1, \infty)$.

$$\lim_{x \rightarrow a} \sqrt{x-1} = \sqrt{\lim_{x \rightarrow a} (x-1)} = \sqrt{a-1}.$$

Notice that $x = 1$ is not included. Now we must prove that the function is also continuous from the right.

$$\lim_{x \rightarrow 1^+} \sqrt{x-1} = 0 = f(1).$$

Since the above is true, we have now proven that the function is continuous in the interval $[1, \infty)$. The function is not continuous to the left of $x = 1$ because any x to the left results in imaginary numbers for the function.

A hole is a **removable discontinuity** since it could be fixed by the addition of a single point. A gap is a **jump discontinuity**, and the gap surrounding a vertical asymptote is an **infinite discontinuity**.

To remove a removable discontinuity, you have to re-define the function as a piecewise function with the condition that the hole is replaced by the limit of the function at that point.

Example:

Remove the discontinuity in the function

$$f(x) = \frac{x^2}{x}.$$

Notice that it has a discontinuity at $x = 0$. The limit of the function at the discontinuity is

$$\lim_{x \rightarrow 0} \frac{x^2}{x} = 0.$$

This is what the value of the function should be at $x = 0$, but instead it is undefined at that point. To remove this discontinuity, we simply redefine the function as

$$f(x) = \begin{cases} \frac{x^2}{x} & : x \neq 0 \\ 0 & : x = 0. \end{cases}$$

A function can also be said to be continuous from the left or the right, and a function can also be considered continuous along an entire interval. The limit as x approaches an integer on the floor function from the right is that integer. The limit as x approaches an integer on the floor function from the left is that integer minus one.

Both the left and right limits exist at that integer, but they're not the same. Overall, the limit does not exist, but the function is said to be continuous from the right.

To show that a function is continuous on an interval, you must show that it is continuous at every point on that interval.

If $f(x)$ and $g(x)$ are continuous at a and c is a constant, then all of the following functions are also continuous at a

$$\begin{aligned} f + g \\ f - g \\ fg \\ \frac{f}{g} \quad g(a) \neq 0 \\ cf \end{aligned}$$

For example, we know that x is continuous and 3 is continuous, therefore $x+3$ is also continuous. These all follow from the limit laws.

Using the above theorem, we can tell if a function is continuous simply by breaking it into little pieces to see if each of the pieces are continuous. If they're all continuous, then the entire function is continuous. When finding the domain of combined functions (e.g. rational function consisting of trigonometric and root functions), find the domain of each piece. Doing this gives the same domain as the overall function and it's way easier to deduce the domain of it piece by piece than to try and graph the overall function.

Example:

Determine where

$$f(x) = \frac{\tan x}{\sqrt{9-x^2}},$$

is continuous.

It would be difficult to manually graph this combination of functions in order to find its domain, but we don't have to—we simply find the domain of the pieces. We know that $9-x^2$ must be strictly greater than zero because the radical can't contain a negative number and it can't be zero. Therefore, the domain of the denominator is $(-3, 3)$. The domain of the numerator excludes every $\frac{\pi k}{2}$, but we only have to worry about the ones within our current $(-3, 3)$ domain. The domain for the entire function, then, is

$$\left(-3, -\frac{\pi}{2}\right) \cup \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \cup \left(\frac{\pi}{2}, 3\right).$$

Most familiar functions are continuous everywhere that they're defined. This includes

- polynomials
- rational functions
- root functions

- trigonometric functions
- inverse trigonometric functions
- exponential functions
- logarithmic functions

If you're taking the limit of a composite function and you know that the outer function is continuous at that point, then you can move that function outside of the limit. That is,

$$\lim_{x \rightarrow a} f(g(x)) = f\left(\lim_{x \rightarrow a} g(x)\right).$$

A related theorem can be stated as, *A continuous function of a continuous function is a continuous function.* For example, if two everywhere continuous functions are composed, then the result is also continuous.

The **intermediate value theorem** states that if f is continuous on $[a, b]$ and if k is any number between $f(a)$ and $f(b)$ (provided that $f(a) \neq f(b)$) then there exists a number c in $[a, b]$ such that $f(c) = k$. This becomes obvious if we graph an arbitrary function.

1.1.5 Limits at Infinity

For a limit

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

if L is $\pm\infty$, then we are dealing with a **vertical asymptote**. If a is $\pm\infty$, then we are dealing with a **horizontal asymptote**.

A function can have one or two horizontal asymptotes.

Most of the limit laws hold true when a is replaced by $\pm\infty$.

If r is a rational number and a is a constant, then

$$\lim_{x \rightarrow \infty} \frac{a}{x^r} = 0.$$

This is easy to understand. As $x \rightarrow \infty$, the numerator stays a constant value a , but the denominator larger and larger. Therefore, the whole fraction becomes smaller and smaller, approaching zero.

When dealing with these infinite limits it's often easy to deduce the limit simply by figuring out how the numerator compares with the denominator as x is made larger.

To evaluate the limit at infinity of any rational function, you must first divide both the numerator and denominator by the highest power of x that occurs in the denominator.

Example:

Evaluate

$$\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 + 3x - 5}}{4x + 2}.$$

We start by dividing the top and bottom by x and splitting the limit of the quotient into the quotient of limits

$$\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 + 3x - 5}}{4x + 2} = \frac{\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 + 3x - 5}}{x}}{\lim_{x \rightarrow \infty} \frac{4x + 2}{x}}.$$

Next, we simplify the denominator and rewrite the x in the top as $\sqrt{x^2}$. Since x is approaching ∞ , which is positive, we are allowed to do this—ignoring the negative square root.

$$\frac{\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 + 3x - 5}}{x}}{\lim_{x \rightarrow \infty} \frac{4x + 2}{x}} = \frac{\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 + 3x - 5}}{\sqrt{x^2}}}{\lim_{x \rightarrow \infty} \left(4 + \frac{2}{x}\right)}.$$

Now we can take the limit of the bottom and we can pull the $\sqrt{x^2}$ on the top inside the larger root.

$$\frac{\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 + 3x - 5}}{\sqrt{x^2}}}{\lim_{x \rightarrow \infty} \left(4 + \frac{2}{x}\right)} = \frac{\lim_{x \rightarrow \infty} \sqrt{\frac{x^2 + 3x - 5}{x^2}}}{4}.$$

Simplifying the top and pulling the limit inside the root gives us

$$\frac{\lim_{x \rightarrow \infty} \sqrt{\frac{x^2 + 3x - 5}{x^2}}}{4} = \frac{\sqrt{\lim_{x \rightarrow \infty} \left(1 + \frac{3}{x} - \frac{5}{x^2}\right)}}{4}.$$

Now we can take the limit in the top, and we just get 1. So

$$\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 + 3x - 5}}{4x + 2} = \frac{1}{4}.$$

Notice that we could have guessed this value from the beginning. The degree of the top is one (since it is the square root of x^2), and the leading coefficient of the top is 1. The the degree of the bottom is the same and the leading coefficient is 4. So overall, the limit at infinity is $\frac{1}{4}$.

The method used in the example above is useful even when the function is not rational, as in the next example.

Example:

Evaluate

$$\lim_{x \rightarrow \infty} (\sqrt{x^2 + 1} - x).$$

We start by multiplying by the radical conjugate

$$\lim_{x \rightarrow \infty} (\sqrt{x^2 + 1} - x) \cdot \frac{(\sqrt{x^2 + 1} + x)}{(\sqrt{x^2 + 1} + x)},$$

and simplifying to get

$$\lim_{x \rightarrow \infty} \frac{1}{\sqrt{x^2 + 1} + x}.$$

We now have a rational limit, and we can divide the top and bottom by $\sqrt{x^2}$ and proceed as in the example above. Alternatively, we can notice that we're taking the limit at positive infinity, that the denominator contains x 's so it's growing large, and that all the values in the denominator are positive. Therefore, as $x \rightarrow \infty$, the denominator becomes a larger and larger positive number, so overall, the expression is approaching zero. Therefore,

$$\lim_{x \rightarrow \infty} (\sqrt{x^2 + 1} - x) = 0.$$

To find all the asymptotes of a rational function $f(x)$ by calculus methods, follow the following procedure.

- **Horizontal Asymptotes:** Find L for positive and negative infinity to find one or both horizontal asymptotes

$$\lim_{x \rightarrow \infty} f(x) = L \quad \text{and} \quad \lim_{x \rightarrow -\infty} f(x) = L.$$

Remember to check both positive and negative infinity because the function may have two HA's.

- **Vertical Asymptotes:** Find the vertical asymptotes using algebra knowledge, but this won't tell you whether $f(x)$ goes to positive or negative infinity on either side of the asymptote. Use limit notation to show exactly what happens at the asymptote. Remember that the easy way to check whether an asymptote goes to positive or negative infinity is to evaluate $f(x)$ at an x -value very close to either side of the asymptote.

Note if finding the infinite limit of a rational function returns DNE then the function doesn't have a horizontal asymptote, but it may still have a slant asymptote. This will happen when the degree of the numerator is larger than the degree of the denominator. The function may also have a slant asymptote if it returns an infinite limit of positive or negative infinity.

A useful method when dealing with certain composed functions is to break it into two steps and substitute a new variable.

Example:

Evaluate

$$\lim_{x \rightarrow 0^+} \tan^{-1} \left(\frac{1}{x} \right).$$

Since the arctangent function is continuous on $(-\infty, \infty)$, we can move it outside the limit

$$\tan^{-1} \left(\lim_{x \rightarrow 0^+} \frac{1}{x} \right).$$

From the graph of $\frac{1}{x}$, we know that it goes to $+\infty$ as x approaches zero from the right. If we make the substitution $u = \frac{1}{x}$ then we know that as $x \rightarrow 0^+$, $u \rightarrow +\infty$, so we can rewrite the limit as

$$\lim_{u \rightarrow \infty} \tan^{-1}(u).$$

From the graph of the arctangent function, we know that its value approaches $\frac{\pi}{2}$ as you go toward positive infinity, so

$$\lim_{x \rightarrow 0^+} \tan^{-1} \left(\frac{1}{x} \right) = \lim_{u \rightarrow \infty} \tan^{-1}(u) = \frac{\pi}{2}.$$

Note that the infinite limits of certain periodic functions (e.g. sine function) do not exist because these periodic functions fluctuate to infinity.

Infinite limits at infinity. For example

$$\lim_{x \rightarrow \infty} f(x) = \infty,$$

means just that $f(x)$ gets larger as x gets larger. Note that the limit laws can't be applied to infinite limits because infinity is not a number.

There are usually several possible ways of solving infinite limits

- Deduce it by considering what happens to the function as x is taken to positive or negative infinity
- If it's a rational function or can be turned into a rational function (rationalizing the numerator or multiply the function by conjugate over conjugate) use the method where you divide every term by the largest degree in the denominator
- Graph the function and visualize it
- Make a table of values
- Break it into simpler functions
- If it's a composed function, substitute a new variable for the inner function so it can be broken two limits.
- If the limit argument contains something times a trig function try to remove the trig function using the squeeze theorem.

Whenever you're supposed to find the limit of something times a trig function, try to use the squeeze the-

orem. Set up in inequality like $-1 < \cos x < 1$, then multiply the whole inequality by the other factor. Now if the function on the left is always smaller or equal to the function on the right, and the limit of the left function

equals the limit of the right, then that is also the limit of the middle function, which is the one you're trying to find.

1.2 Derivatives and Differentiation

1.2.1 Derivative and Rate of change

Remember that the slope of a *secant* line or the average velocity of a function f between the points $(a, f(a))$ and $(x, f(x))$ are given by

$$m = \frac{f(x) - f(a)}{x - a}, \quad \text{or} \quad m = \frac{f(a+h) - f(a)}{h}.$$

The function on the right side is the one most commonly used in calculus. It is known as the **difference quotient**.

The slope of the *tangent* line or the instantaneous velocity of a function f at the point $(a, f(a))$ is given by

$$m = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}, \quad \text{or} \quad m = \lim_{x \rightarrow a} \frac{f(a+h) - f(a)}{h}.$$

This second limit is commonly used in science and engineering because it represents a rate of change. It is so commonly used that in mathematics it is given the special term “derivative”.

The slope of the secant line is the average rate of change of y with respect to x and the slope of the tangent line is the instantaneous rate of change of y with respect to x .

The **derivative** of f at a point a is

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

provided that this limit exists.

More generally, the **derivative** of the function $f(x)$ is defined as

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

provided that this limit exists.

Remember that the difference quotient gives the slope of the secant line between two points. The limit of that difference quotient as h approaches 0 gives the derivative or the slope of the tangent line at one point.

Example:

Calculate the derivative of $f(x) = x^2 + 3x - 5$.

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(x+h)^2 + 3(x+h) - 5 - x^2 - 3x + 5}{h} \\ &= \lim_{h \rightarrow 0} \frac{2hx + h^2 + 3h}{h} \\ &= \lim_{h \rightarrow 0} 2x + h + 3 \\ &= 2x + 3. \end{aligned}$$

Example:

Calculate the derivative of $f(x) = \sqrt{x}$ at the point a .

Using the standard limit definition of the derivative is tricky here because it requires us to expand $(x+h)^{\frac{1}{2}}$. Fortunately, we can use the other limit definition of the derivative.

$$\begin{aligned} f'(x) &= \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \\ &= \lim_{x \rightarrow a} \frac{\sqrt{x} - \sqrt{a}}{x - a} \\ &= \lim_{x \rightarrow a} \frac{x^{\frac{1}{2}} - a^{\frac{1}{2}}}{\left(x^{\frac{1}{2}}\right)^2 - \left(a^{\frac{1}{2}}\right)^2} \\ &= \lim_{x \rightarrow a} \frac{x^{\frac{1}{2}} - a^{\frac{1}{2}}}{\left(x^{\frac{1}{2}} - a^{\frac{1}{2}}\right)\left(x^{\frac{1}{2}} + a^{\frac{1}{2}}\right)} \\ &= \lim_{x \rightarrow a} \frac{1}{x^{\frac{1}{2}} + a^{\frac{1}{2}}} \\ &= \frac{1}{2\sqrt{a}}. \end{aligned}$$

Some practical interpretations of the derivative include

- If $f(t)$ is the position function of a particle with respect to time t (i.e. the y -axis represents height and the x -axis represents time), then $f'(a)$ is the velocity of the particle and $|f'(a)|$ is the speed of the particle.
- If $f(x)$ is the function describing the cost of pro-

ducing something with respect to amount of stuff produced, then $f'(x)$ is the rate of change of the production cost with respect to the number of units produced. In economics, this is known as marginal cost.

- In physics, the rate of change of work with respect to time is called power.
- In chemistry, the rate of change in the concentration of a reactant with respect to time is called the rate of reaction.

To estimate $f'(x)$ from experimental data, first compute the values of the difference quotient (average rate of change), then take the average of the two values closest on either side of the target value.

When given the velocity function of a thrown object moving in a parabolic curve, you may be asked when the rock will hit the ground. Note that the velocity function is the first derivative of the height function. You can set the velocity function equal to 0 and solve for t (time), but remember to double that value. The velocity is equal to zero at the top of the parabolic curve, and then it goes negative as it starts to fall. So when the velocity is zero, the thrown object is halfway (in time) to hitting the ground.

To graph the derivative function of a function that is not differentiable at all places (i.e. it includes corners, kinks, or discontinuities) make sure you graph open intervals (i.e. endpoints not included) in the derivative function. If the graph of the first function has a corner at $x = 3$, then the function is not differentiable at 3, which means the derivative function has to be undefined at that point.

There are a number of different notations used to represent the derivative of y with respect to x including

$$f'(x) = y' = \frac{dy}{dx} = \frac{d}{dx}f(x) = Df(x) = D_x f.$$

A function f is differentiable on an open interval if it is differentiable at every number in the interval.

Geometrically, the derivative $f'(x)$ is the slope of the tangent line to the function $f(x)$ at the point x .

The domain of $f'(x)$ is every number at which $f(x)$ is differentiable.

The derivative $f'(x)$ crosses the x -axis wherever the graph of $f(x)$ turns (i.e. where the slope of the tangent line is 0). Elsewhere, estimate the slope of the tangent line at various x -values of $f(x)$. The derivative $f'(x)$ is positive wherever the slope of $f(x)$ is positive.

A function is not differentiable where the graph makes a kink (or corner), where it's not continuous, and where the slope is infinite. For example, $f(x) = |x|$ is not differentiable at $f(0)$ where the graph makes an abrupt turn because the limit of the difference quotient at $x = 0$

as it is approached from the left is -1 but it is 1 when approached from the right. For the derivative to exist at a certain a it needs to be the same whether approached from the left or the right.

If $f(x)$ is differentiable at a , then it is necessarily continuous at a . The converse of this is not true. A function that is continuous is not necessarily differentiable. For example $|x|$ is continuous, but it's not differentiable at $x = 0$ because it has a sharp corner at that point.

If f is a differentiable function then its derivative is a function which means that it may have its own derivative. The derivative of a derivative function is called the **second derivative** and is usually denoted as:

$$f''(x), \quad \text{or} \quad \frac{d^2y}{dx^2}.$$

Derivatives of order higher than three are usually written in the form

$$f^{(n)}(x).$$

If $s = s(t)$ is the position function of an object that moves in a straight line, then the first derivative is the velocity, the second derivative is the acceleration, and the third derivative is the jerk.

1.2.2 Polynomials and Exponentials

The derivative of a constant is always equal to zero. In other words, a constant function has a slope of zero everywhere.

$$\frac{d}{dx}c = 0.$$

The **power rule** tells us how to differentiate a variable raised to a power

$$\frac{d}{dx}x^n = nx^{n-1}.$$

Note: The power rule also works with negative exponents, which makes it a lot easier when dealing with radicals.

The **normal line** to a curve at a point P is the line that runs through P that is perpendicular to the tangent line at P . To find the normal line of a curve at point P , first find the derivative of the curve at P to get the slope of the tangent line, then using the x and y values of the given point, construct an equation for the tangent line. The normal line runs through the same point but is perpendicular to the tangent line. Remember, that the slope of a perpendicular line is the negative inverse. So if the slope of one line is $\frac{1}{2}$, then the line perpendicular to that will have a slope of -2 . The normal line is used in physics. For example, the normal force is the line of force perpendicular to the surface. In optics, one needs to consider the angle between a light ray and the normal line to the lens.

The derivative of a constant times a function is the constant times the derivative of the function

$$\frac{d}{dx} [cf(x)] = c \frac{d}{dx} f(x).$$

Interpreted geometrically, a function multiplied by a constant c is stretched upward by a factor of c . The slope (rise) at every point has been multiplied by a factor of c but the run has not been affected.

The derivative of a sum of functions is the sum of the derivatives provided that they are both differentiable

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

Similarly, the derivative of a difference of functions is the difference of the derivatives provided that they are both differentiable

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

The above rules can be used in combination to differentiate any polynomial.

To find the **turning points** of a function (points where the tangent line to the graph has a slope of zero), differentiate the function. The derivative function will cross the x -axis at every point where the tangent line to the first point is equal to zero. So to find the horizontal tangents of a function, simply find the zeros of the first derivative.

The rate of change of any exponential function is proportional to the function itself. In other words, the slope at any point is proportional to the height at that point.

Tip:

Keep in mind that the derivative of a function like e^5 is 0, not $5e^4$. This is because e is a constant (not a variable), and so e^5 is also a constant.

The derivative of the natural exponential function is the natural exponential function

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

Geometrically, this means that the slope for the function e^x at point (x, y) is equal to y .

Tip:

Be careful when dealing with negative exponents. Don't lose track of the signs.

The derivative of a function can tell us a lot about its graph. For example, the first derivative is 0 wherever the original function turns, and the first derivative is positive wherever the original function is increasing.

1.2.3 Product and Quotient Rules

The **product rule** tells us how to differentiate the product of two functions $f(x)$ and $g(x)$

$$(fg)' = fg' + gf'.$$

The derivative of a product is the derivative of one times the other plus the derivative of the other times the one.

Because of the properties of multiplication, this process of finding the derivative of a product is similar to binomial expansion. For example, if $F(x) = f(x)g(x)$ then

$$F''' = f'''g + 3f''g' + 3f'g'' + fg'''.$$

This same binomial expansion process works for F^n .

Depending on the problem given, it may be easier to solve using the product rule immediately or by first simplifying the function (by foiling for example).

The **quotient rule** tells us how to differentiate the quotient of two functions $f(x)$ and $g(x)$.

$$\left(\frac{f}{g}\right)' = \frac{gf' - fg'}{g^2}.$$

Note, the order of operations must be performed exactly as shown—you can't shuffle any of the pieces around. Therefore, it's important to memorize the formula.

Several different ways to memorize the formula are

$$\left(\frac{t}{b}\right)' = \frac{bt' - tb'}{b^2}.$$

Here, we just replaced f and g by t (i.e. top) and b (i.e. bottom). Another way to remember it is as

$$\left(\frac{Hi}{Ho}\right)' = \frac{HoDHi - HiDHo}{HoHo}.$$

Here the D denotes that the next item is differentiated.

Note: When working with these, always foil the numerator, but there's never a reason to foil the denominator. Besides, leaving the denominator factored may allow you to simplify the final result.

Tip:

Concave up means the opening of the concavity is pointed upwards—not the hump.

With the above rules and the ones in the preceding section, we can now differentiate any rational function.

Before using the quotient rule, check that the rational function cannot easily be simplified into a simpler polynomial that would be a lot faster to differentiate.

Notice that multiple applications of the product rule looks like this

$$(fgh)' = f'gh + fg'h + fgh'.$$

1.2.4 Trigonometric Functions

Note: The left and right hand limits of an even function at 0 are always equal to each other.

The derivative of the sine function is

$$\frac{d}{dx} \sin x = \cos x.$$

D $\sin x = \cos x$ You can guess this by looking at the graph of $\sin x$ and drawing its derivative function.

The derivative of the cosine function is

$$\frac{d}{dx} \cos x = -\sin x.$$

You can guess this by looking at the graph of $\cos x$ and drawing its derivative function.

The derivatives of the rest of the trigonometric functions can be derived from the derivatives of the sine and cosine functions by using the quotient rule. However, they are listed here

$$\begin{aligned} \frac{d}{dx} \tan x &= \sec^2 x \\ \frac{d}{dx} \csc x &= -\csc x \cot x \\ \frac{d}{dx} \sec x &= \sec x \tan x \\ \frac{d}{dx} \cot x &= -\csc^2 x. \end{aligned}$$

An important limit to remember is

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

Notice that this also implies that the reciprocal limit is the same

$$\lim_{x \rightarrow 0} \frac{x}{\sin x} = 1.$$

Another important limit to remember is

$$\lim_{x \rightarrow 0} \frac{\cos x - 1}{x} = 0.$$

When finding limits of trig functions, if direct substitution leads to $\frac{0}{0}$, we must algebraically manipulate the function so that we can either simplify it or use one of the above special limits.

Example:

Evaluate

$$\lim_{x \rightarrow 0} \frac{\sin(7x)}{3x}.$$

Notice that it's not quite in the form of the special limit given above, but we can manipulate it into that form by using a clever substitution $u = 7x$.

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\sin(7x)}{3x} &= \lim_{x \rightarrow 0} \frac{\sin(7x)}{7x} \cdot \frac{7}{3} \\ &= \frac{7}{3} \lim_{u \rightarrow 0} \frac{\sin u}{u} \\ &= \frac{7}{3} \cdot 1 = \frac{7}{3}. \end{aligned}$$

Example:

Evaluate

$$\lim_{x \rightarrow 0} \frac{\tan(7x)}{\sin(2x)}.$$

This example is a little more difficult than the previous example. Note that first we write it in terms of sine and cosine. Next we break the complex fraction into the product of three simpler fractions. Then we do the multiplication trick used in the previous example to turn two of the limits into the special sine limit. The cosine limit is just 1 since we know that $\cos 7x$ approaches 1 as x approaches zero.

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\tan(7x)}{\sin(2x)} &= \lim_{x \rightarrow 0} \frac{\frac{\sin(7x)}{\cos(7x)}}{\sin(2x)} \\ &= \lim_{x \rightarrow 0} \left(\frac{\sin(7x)}{\cos(7x)} \frac{1}{\sin(2x)} \right) \\ &= \lim_{x \rightarrow 0} \left(\frac{\sin(7x)}{1} \right) \lim_{x \rightarrow 0} \left(\frac{1}{\cos(7x)} \right) \lim_{x \rightarrow 0} \left(\frac{1}{\sin(2x)} \right) \\ &= \lim_{x \rightarrow 0} \left(\frac{\sin(7x)}{7x} \cdot \frac{7x}{1} \right) \lim_{x \rightarrow 0} \left(\frac{1}{\cos(7x)} \right) \lim_{x \rightarrow 0} \left(\frac{2x}{\sin(2x)} \cdot \frac{1}{2x} \right) \\ &= \frac{7}{2} \lim_{x \rightarrow 0} \left(\frac{\sin(7x)}{7x} \right) \lim_{x \rightarrow 0} \left(\frac{1}{\cos(7x)} \right) \lim_{x \rightarrow 0} \left(\frac{2x}{\sin(2x)} \right) \\ &= \frac{7}{2} (1)(1)(1) \\ &= \frac{7}{2}. \end{aligned}$$

Tip:

Memorize the trig identities because they may be necessary when it comes time to simplify something.

Remember how to solve trig equations, for example, to find the zeros of a function containing multiple trig functions. To do so, you have to trig use identities. For example, to find the zeros of the following function, we rearrange, square both sides, then use a trig identity to solve it.

$$\begin{aligned} \cos x - \sin x &= 0 \\ \cos^2 x &= \sin^2 x \\ \cos^2 x &= 1 - \cos^2 x \\ 2 \cos^2 x &= 1 \\ \cos x &= \pm \frac{\sqrt{2}}{2} \\ x &= \frac{\pi}{4} + 2\pi k, \quad \frac{5\pi}{4} + 2\pi k. \end{aligned}$$

Sometimes, it saves a lot of time if you can simplify a trig function using identities before trying to differentiate it.

1.2.5 Chain Rule

The **chain rule** tells us how to differentiate a composed function $f(g(x))$

$$[f(g(x))] = f'(g(x)) \cdot g'(x),$$

provided that g is differentiable at x and f is differentiable at $g(x)$. If $y = f(u)$ and $u = g(x)$, we can also write this in Leibniz form as

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

Note that to differentiate a composed function you take the derivative of the outer function at the inner function and multiply that by the derivative of the inner function.

Example:

Differentiate $f(x) = \sin(2x^2 + 3x)$.

Notice that this is the function $\sin x$ composed with the function $2x^2 + 3x$. The chain rule tells us that to differentiate such a function, we take the derivative of the outer function evaluated at the inner function and multiply it by the derivative of the inner function.

$$f'(x) = \cos(2x^2 + 3x) \cdot (4x + 3) = (4x + 3) \sin(2x^2 + 3x).$$

Notice that we implicitly use the chain rule when we differentiate very simple functions. The 1 in the case below is the derivative of the inside function x

$$\frac{d}{dx} x^2 = 2x \cdot 1 = 2x.$$

The power rule combined with the chain rule gives

us

$$\frac{d}{dx}[g(x)]^n = n[g(x)]^{n-1} \cdot g'(x).$$

Example:Differentiate $\sin^2 x$.

Notice that this calls for the combination of the power rule and the chain rule.

$$\frac{d}{dx} \sin^2 x = \frac{d}{dx} (\sin x)^2 = 2 \sin x \cos x.$$

The chain rule must often be used in combination with other rules like the power rule and quotient rule to differentiate functions.

The chain rule can also be used repeatedly to differentiate nested functions. For such functions, simply work your way inwards, applying the chain rule at each step.

Example:Differentiate $f(x) = \sin(\cos(\tan x))$.

This is a case of three nested functions, so we have to apply the chain rule twice.

$$\begin{aligned} f'(x) &= \frac{d}{dx} \sin(\cos(\tan x)) \\ &= \cos(\cos(\tan x)) \frac{d}{dx} \cos(\tan x) \\ &= \cos(\cos(\tan x)) (-\sin(\tan x)) \frac{d}{dx} \tan x \\ &= \cos(\cos(\tan x)) (-\sin(\tan x)) \sec^2 x \\ &= -\sin(\tan x) \sec^2 x \cos(\cos(\tan x)). \end{aligned}$$

To differentiate exponential functions with bases other than e remember that $a = e^{\ln a}$. This tells us that

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

So by the chain rule,

$$\begin{aligned} \frac{d}{dx} a^x &= \frac{d}{dx} e^{x \ln a} \\ &= e^{x \ln a} \cdot \ln a \\ &= e^{\ln a^x} \ln a \\ &= a^x \ln a. \end{aligned}$$

Example:

Calculate

$$\frac{d}{dx} (4^{5^{x^2}}).$$

$$\begin{aligned} \frac{d}{dx} (4^{5^{x^2}}) &= 4^{5^{x^2}} \ln 4 \frac{d}{dx} (5^{x^2}) \\ &= 4^{5^{x^2}} \ln 4 \frac{d}{dx} \cdot 5^{x^2} \cdot \frac{d}{dx} (x^2) \\ &= 2x \ln(20) \cdot 5^{x^2} \cdot 4^{5^{x^2}}. \end{aligned}$$

Notice that the above is just repeated application of the chain rule for an exponential function of base other than e . The function is a nested function with 3 levels: 4^u , 5^v , and x^2 .

When differentiating complicated functions requiring multiple rules used in the proper sequence, it may be helpful to write down the relevant formulas (such as for the product rule or quotient rule) before you start. Then, you're less likely to lose where you're at in the process.

Often a function is given that looks like it requires using the quotient and the chain rule, but sometimes it can be rewritten so that it requires only the chain rule—making it a lot easier. For example, the two functions below are equivalent, but the second one is a lot faster to differentiate because it requires only the chain rule.

$$\frac{d}{dx} \frac{1}{\sin x} = \frac{d}{dx} (\sin x)^{-1}.$$

Dont mistake the tangent line at P on a curve C for the equation of the derivative function of C! I made that mistake on the last test.

Remember that $D(\sin(2x))$ requires the chain rule because the $2x$ is a function inside of another function, in this case the sine function. In the same way, $D(e^{-u})$ requires the chain rule because the $-u$ is a function inside of the function e^x .

1.2.6 Implicit Differentiation

Remember the chain rule in Leibniz notation is

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

So far, the functions we've differentiated have been defined explicitly in terms of y . Now we look at how to differentiate functions that are not explicitly defined as functions of x . Here are two examples of such equations

$$\begin{aligned} x^2 + y^2 &= 4 \\ x^3 + y^3 &= 6xy. \end{aligned}$$

Sometimes it is possible to solve these kinds of equations for y , but usually it gives multiple functions.

To differentiate these implicit functions, we differentiate both sides of the function with respect to x and then solve the resulting equation for y' .

The process of differentiating these is essentially the same... we just show the derivative of y as $\frac{dy}{dx}$. For example

$$\begin{aligned}\frac{d}{dx}(x^2 + y^2) &= \frac{d}{dx}4 \\ 2x + 2y\frac{dy}{dx} &= 0 \\ \frac{dy}{dx} &= -\frac{x}{y}.\end{aligned}$$

Notice that to differentiate y^2 with respect to x , we use the chain rule $\frac{d}{dx}(y^2) = 2y \cdot \frac{dy}{dx}$. The final result $\frac{dy}{dx} = -\frac{x}{y}$ is the slope of the tangent line at any point (x, y) on that circle.

Example:

Differentiate the following equation implicitly with respect to x

$$x^3 + y^3 = 6xy.$$

We treat the derivative with respect to x as an *operator* that can be applied to both sides of the equation to leave it still the same

$$\begin{aligned}\frac{d}{dx}(x^3 + y^3) &= \frac{d}{dx}6xy \\ 3x^2 + 3y^2\frac{dy}{dx} &= 6y + 6x \cdot \frac{dy}{dx} \\ 3y^2\frac{dy}{dx} - 6x \cdot \frac{dy}{dx} &= 6y - 3x^2 \\ \frac{dy}{dx}(3y^2 - 6x) &= 6y - 3x^2 \\ \frac{dy}{dx} &= \frac{6y - 3x^2}{3y^2 - 6x} \\ \frac{dy}{dx} &= \frac{2y - x^2}{y^2 - 2x}.\end{aligned}$$

Notice that on line two, we had to use the product rule to differentiate the righthand side.

Example:

Find the slope at $(2, 3)$ on the ellipse

$$3x^2 + 4y^2 = 48,$$

first using regular differentiation and then using implicit differentiation.

Before we can differentiate using our former methods, we need to solve for y :

$$y = \pm\sqrt{12 - \frac{3}{4}x^2}.$$

Since our point of interest $(2, 3)$ is on the upper side of the ellipse, we differentiate the positive solution to get

$$\begin{aligned}y' &= \frac{1}{2}\left(12 - \frac{3}{4}x^2\right)^{-\frac{1}{2}}\left(-\frac{6}{4}x\right) \\ &= -\frac{3x}{3\sqrt{12 - \frac{3}{4}x^2}}.\end{aligned}$$

Plugging in $x = 2$ gives us our slope on the ellipse at $(2, 3)$ as

$$y' = -\frac{3(2)}{3\sqrt{12 - \frac{3}{4}(2)^2}} = -\frac{1}{2}.$$

Differentiating it implicitly and plugging in $(x, y) = (2, 3)$ is a lot faster and cleaner

$$\begin{aligned}\frac{d}{dx}(3x^2 + 4y^2) &= \frac{d}{dx}48 \\ 6x + 8y\frac{dy}{dx} &= 0 \\ \frac{dy}{dx} &= -\frac{3x}{4y} \\ \frac{dy}{dx} &= -\frac{3(2)}{4(3)} = -\frac{1}{2}.\end{aligned}$$

Example:

Differentiate the following equation implicitly with respect to x

$$\cos xy = 1 + \sin y.$$

Applying the derivative operator to both sides then simplifying, gives us

$$\begin{aligned} \frac{d}{dx}(\cos xy) &= \frac{d}{dx}(1 + \sin y) \\ -\sin xy \left(y + x \frac{dy}{dx} \right) &= 0 + \cos y \frac{d}{dx} \\ \frac{dy}{dx} &= -\frac{y \sin xy}{\cos y + x \sin xy}. \end{aligned}$$

Notice that on the left side of line two, we used the chain rule and the product rule to differentiate $\cos xy$ with respect to x .

To find the second derivative of an implicit function, first find the first derivative and solve for $\frac{dy}{dx}$. Then find the derivative of that function. Even in the second function, the derivative of y is represented by $\frac{dy}{dx}$ not by $\frac{d^2y}{dx^2}$.

When doing second derivatives of implicit functions, try to reduce a portion of your second derivative to match one side of the original equation. This allows you to simplify even further by replacing the value in the second derivative with the value in the original function.

Remember that $\sin^{-1} x = y$ means $\sin y = x$ and y is between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Therefore, to get the derivative of the inverse sine function, we can do an implicit differentiation of $\sin y = x$, which yields $y' = \frac{1}{\cos y}$. Then we can write

$$\cos y = \sqrt{1 - \sin^2 y} = \sqrt{1 - x^2}.$$

The first part is true because y is being considered only between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. The contents of the radical come from the trig identity $\cos^2 x + \sin^2 x = 1$. The second part is true because of our original equation $\sin y = x$. Therefore, we have that the derivative of the inverse sine function is

$$y' = \frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1 - x^2}}.$$

The other inverse trig functions can be differentiated the same way.

The derivative of the inverse trig functions are as

follows

$$\begin{aligned} \frac{d}{dx} \sin^{-1} x &= \frac{1}{\sqrt{1 - x^2}} \\ \frac{d}{dx} \cos^{-1} x &= -\frac{1}{\sqrt{1 - x^2}} \\ \frac{d}{dx} \tan^{-1} x &= \frac{1}{1 + x^2} \\ \frac{d}{dx} \csc^{-1} x &= -\frac{1}{x\sqrt{x^2 - 1}} \\ \frac{d}{dx} \sec^{-1} x &= \frac{1}{x\sqrt{x^2 - 1}} \\ \frac{d}{dx} \cot^{-1} x &= -\frac{1}{\sqrt{1 + x^2}}. \end{aligned}$$

An identity involving the inverse trig functions is

$$\cos^{-1} x + \sin^{-1} x = \frac{\pi}{2}.$$

Tip:

When taking the derivative of an inverse trig function, you can use the derivatives listed above. However, don't forget to use the chain rule if the argument of the trig function contains more than just x .

Example:

Find the derivative of $\tan^{-1} x^3$.

We have to use the chain rule here

$$\begin{aligned} \frac{d}{dx} \tan^{-1}(x^3) &= \frac{1}{1 + (x^3)^2} \frac{d}{dx}(x^3) \\ &= \frac{2x^2}{1 + x^6}. \end{aligned}$$

1.2.7 Logarithmic Functions

The following equations need to be remembered

$$\begin{aligned} \frac{d}{dx} \log_a x &= \frac{1}{x \ln a} \\ \frac{d}{dx} \ln x &= \frac{1}{x} \\ \frac{d}{dx} \ln u &= \frac{1}{u} \cdot \frac{du}{dx} \\ \frac{d}{dx} \ln g(x) &= \frac{g'(x)}{g(x)}. \end{aligned}$$

Note, the last two are just adding the chain rule.

To differentiate a logarithmic problem, simply add the argument (and base if applicable) into the appropriate formula and simplify. Remember, if the argument of the log is not just x then the chain rule must be applied.

Example:

Calculate the derivative of $\ln(7x)$. Using the chain rule, we get

$$\frac{d}{dx} \ln(7x) = \frac{1}{7x} \cdot 7 = \frac{1}{x}.$$

There are always at least two ways of differentiating logarithms. One method is the straightforward method involving product rule and quotient rule as necessary. Another method is to simplify the logarithms using log rules before trying to differentiate. When differentiating logs the answer must be given as one function—not the sum or difference of functions. This means we have to combine complex fractions as necessary to get one rational function.

Tip:

It is often much faster to expand a logarithm using log rules before differentiating it. Doing this often allows you to not have to use the product or quotient rules when differentiating.

Logarithmic differentiation allows us to simplify complicated functions prior to differentiating them by using logarithms. Remember that taking the logarithm of both sides of an equation, doesn't change the equation. We use logarithmic differentiation to solve complicated equations involving products, quotients, or powers. To use logarithmic differentiation, we follow the following steps

1. Take the natural log of both sides of the equation.
2. Simplify (i.e. expand) the equations using the properties of logarithms
3. Differentiate the equation using implicit differentiation
4. Solve for y'
5. Substitute any y 's in the final equation with the value given for y in the original equation

Be vigilant when dealing with exponential equations. Any term with an x (or other independent variable) in the exponent, cannot be differentiated simply with the power rule. In general, there are four cases involving the derivatives of exponents and bases

1. Constant base, constant exponent

$$\frac{d}{dx} a^b = 0.$$

This is just the derivative of a constant, so the derivative is zero.

2. Variable base, constant exponent

$$\frac{d}{dx} [f(x)]^b = b [f(x)]^{b-1} f'(x),$$

The derivative is found by using the power and chain rules as shown above.

3. Constant base, variable exponent

$$\frac{d}{dx} [a^{g(x)}] = a^{g(x)} \ln(a) g'(x).$$

4. Variable base, variable exponent

$$\frac{d}{dx} [f(x)]^{g(x)}.$$

This kind of derivative must be found by using logarithmic differentiation. This makes it easier because you can use the properties of logarithms to bring the exponent down.

1.2.8 Rates of Change and Related Rates

If $s = f(t)$ is the position function with respect to time, then $\frac{\Delta s}{\Delta t}$, is the **average velocity**, $s'(t)$ is the **instantaneous velocity**, $s''(t)$ is the **instantaneous acceleration**, and $s'''(t)$ is the **instantaneous jerk**.

If you are given the equation of motion for an object, then the zeros of the first derivative of that equation are the places where the tangents to the original equation have a slope of zero. In other words, the zeros of the derivative function can give the time t when the object hit maximum height and/or where velocity is zero.

Whenever the function $y = f(x)$ has a specific interpretation in one of the sciences, its derivative will have a specific interpretation as a **rate of change**. Important rates of change in physics include velocity, density, current, power (the rate at which work is done), the rate of heat flow, temperature gradient (the rate of change of temperature with respect to position), and the rate of decay of a radioactive substance.

What does the graph of a position function mean? The position function gives the position of the object with respect to time. Specifically, it gives the distance of the object along one line (left/right, up/down, etc.) from a certain spot (the entire x -axis). The “along one line” specification is usually clarified by saying that the object moves along a straight line. For example, if you're standing at a certain spot and an object is moving along a straight line left/right or up/down from you, then the position function gives the distance from you to the object at any time.

Take any point on the graph of the position function and the first coordinate gives the time and the second coordinate gives the distance. If the coordinate is positive, it usually means the particle is to your right or above. In other words, considering two points on the graph—one is above and the other is below the x -axis, then it means the object at those two times is on opposite sides of you (i.e. you're the x -axis and as time goes along and the graph goes up and down, the object is moving in a straight line left and right or up and down through your position). If the graph of the position function curves down, it means the object is coming towards you. The steeper the slope

of the downward curve, the faster it is coming towards you.

Think of an object hanging from a spring. If you pull the object down and release it, it is just going to bob up and down. However, if you graph its height with respect to time, you get a sine function. The horizontal (time) axis is simply the natural resting position of the object when it isn't moving. When the graph of the function is positive, it means the object is above its natural resting position, and when the graph is negative, it means the object is below its reference point (i.e. natural resting position). When the graph curves up or down (i.e. the tangent is horizontal) it means the object is changing direction.

The first derivative of the position function gives the instantaneous velocity of the object. Whenever the velocity function is positive, it means the velocity is positive. Keep in mind that negative velocity is just velocity in the opposite direction. In our spring case, a positive velocity means the object is going up. A negative velocity means the object is going down. The further a point is from the x -axis, the faster the object is moving at that point in time. The closer to the x -axis, the slower the object is moving at that point in time. At the point where the velocity function crosses the x -axis, the object is not moving, meaning the spring is at its high or low and is ready to change direction. At a horizontal tangent of the velocity function, the object is moving at a constant unchanging speed, that is, its acceleration is zero.

The second derivative of the position function gives the acceleration of the object. Acceleration is negative whenever the velocity is going down. Note that velocity going down doesn't necessarily mean the object is slowing down because an increasing velocity in the negative direction can still be interpreted as the velocity going down.

Acceleration is zero whenever velocity is at its greatest (in either the positive or negative directions). Note that deceleration is simply acceleration in the opposite direction.

Wherever the acceleration function crosses the x -axis, the acceleration is zero, which means the velocity is constant at that point. For periodic movement (like the spring) wherever the acceleration function crosses the x -axis, the object is at its greatest (local) velocity—it is done speeding up and hasn't started slowing down.

When is the object moving in the positive direction? Whenever the slope of the position function is positive and/or whenever the velocity is positive.

What is the total distance traveled during the first n seconds? Break the position function into sections wherever it turns. The total distance traveled between x_1 and x_2 is the absolute value of the height difference between x_1 and the first turning point, plus the total height distance between the first and second turning points, and so on.

When is the object speeding up? When the acceleration and velocity have the same signs. In other words, for the intervals in which the sign of the velocity function equals the sign of its slope, the object is speeding up. So the object is speeding up wherever the velocity function has a positive slope above the x -axis and wherever it has a negative slope below the x -axis.

When is the object slowing down? When the acceleration and velocity have opposite signs. In other words, for the intervals in which the sign of the velocity function is opposite the sign of its slope, the object is slowing down. So the object is slowing wherever the velocity function has a positive slope beneath the x -axis and wherever it has a negative slope above the x -axis.

When you're given a function of something and you're asked to find the highest or lowest value of the function, you're being asked to find the maximum or minimum. Both of these occur when the slope of the tangent to the function is zero. To find those points, find the zeros of the derivative function.

When you're given a quantity with a unit of something over something, for example: C/s (coulombs per second), you know you're potentially dealing with a rate of change. The unit for electrical current is C/s or ampere. Current is a rate of change. Whenever something increases or decreases nonlinearly with respect to something else (e.g. x or time), then you have a non-constant rate of change.

Whenever you're asked for the value of something at a specific point rather than up to that point, then you probably have to differentiate and plug your independent variables into the derivative function. In other words, if you're given a function with respect to something and the output of the function has a simple unit (such as kg) but you're asked to find a value with more complex units (such as kg/m), then you probably have to differentiate the original function in order to get a new function that outputs the right units. As an example, the distance function gives you units of distance, but its derivative function gives units of distance per time.

When doing rates of change, think about whether the answer should be positive or negative. In a word problem it's easy to miss whether a derivative should be positive or negative.

Remember that average rate of change is rise/run and instantaneous change is the limit of rise/run as the run approaches zero.

A **related rates** problem is where you are given two related rates (i.e. derivatives) and have to figure out the relationships between the two rates in order to answer some question. Or we might be given one rate of change and have to figure out the related rate with the given information. For example, if you blow up a balloon, the radius, the surface area, and the volume of the balloon is changing with time. However, all of these rates of change

are different. They are also all related to each other by the relationships between radius, surface area, and volume.

To solve related rate problems, use the following procedure

1. Read the problem very carefully
2. Identify the given rate and the unknown rate
3. Draw a picture
4. Assign notation (i.e. variables) to quantities that are changing
5. Express information and required rates in terms of derivatives
6. Write an equation that relates the quantities in the problem. If necessary, use the geometry of the situation to eliminate one of the variables.
7. Differentiate both sides of the equation with respect to time
8. Substitute givens into the equation and solve for the unknown rate

Things to keep in mind are

- Read the problem carefully.
- Don't input your variables too early—the equation that relates everything must be differentiated first.
- Don't forget when differentiating the equation that relates everything—that the derivative of a lone constant is zero.
- Make sure you know whether or not a variable varies with time.

When trying to create an equation that relates everything, if you have two equations in terms of a third variable, you can solve both for the third variable and then set the results equal to each other so you have a single equation in two variables.

How do you identify rates of change? If something can be defined as something per something else, then it is a rate of change. Often a rate of change is something per unit time, but it doesn't have to be. As long as the first something is entirely dependent on the second something, and as long as the second something is completely independent, then we have a pure rate of change in the mathematical sense. General examples include

- distance per time (velocity)
- velocity per time (acceleration)
- volume per time
- energy per time (power)
- electric charge per time (electric current)
- momentum per time (force)
- anything per time
- height per distance (slope)
- heat loss per time (i.e. rate of cooling)

In physics, something per unit time is called the time derivative, and it is used extensively. Specific examples include

- Gallons of water per minute
- Miles per hour
- Words per minute

1.2.9 Linear Approximation

For difficult to calculate functions, we sometimes use approximations of a lesser degree. For example the linear approximation or tangent line approximation of $f(x)$ at the point $x = a$ is

$$f(x) \approx f(a) + f'(a)(x - a).$$

The linear function whose graph is this tangent line is called the **linearization** of f at a

$$L(x) = f(a) + f'(a)(x - a).$$

To estimate a value using a linear approximation, find the linearization of the function at a convenient a close to the given value, then insert the given value in place of x .

Example:

Approximate the value of 1.9999^5 using linearization.

The function we are linearizing is x^5 . A convenient value close to the value we are trying to approximate is $a = 2$. The linearization of this function at $a = 2$ is

$$L(x) = 2^5 + 5(2)^4(x - 2).$$

Then our approximation is

$$f(1.9999) \approx 2^5 + 5(2)^4(1.9999 - 2) = 31.992.$$

Using a calculator, we find that the actual value of 1.9999^5 is 31.9920008, so our approximation is very close.

We've learned about linear approximations but there are also polynomial approximations of higher degrees. The n th degree Taylor polynomial of f centered at a , is an equation that constructs a polynomial approximation of desired degree.

1.2.10 Differentials

Remember the different notations for the derivative

$$\frac{dy}{dx}y' = f'(x).$$

Previously, we've treated $\frac{dy}{dx}$ as a single entity, but now we'll begin to treat them as two separate entities, dy and dx . These two are called **differentials**. The differential dx can have an actual numerical value (it approximates change in x) and dy is a real dependent variable

(it depends on dx). The differential dy approximates the change in y . In fact,

$$dy = f'(x) dx.$$

Keep in mind that dy is not the entire derivative y' .

When asked to solve for the differential, solve for dy . When asked to evaluate dy you'll need the values for both x and dx .

Remember that $\frac{dx}{dy}$ is a derivative, but dy and dx are differentials that represent instantaneous rates of change.

1.2.11 Hyperbolic Functions

Certain even and odd combinations of e^x and e^{-x} arise so frequently in the study of decay and change that they have been given special names. The hyperbolic functions are not trig functions, but they display some of the same behavior as their equivalent trig functions.

The hyperbolic functions are defined as follows

$$\begin{aligned}\sinh x &= \frac{e^x - e^{-x}}{2} \\ \cosh x &= \frac{e^x + e^{-x}}{2} \\ \tanh x &= \frac{\sinh x}{\cosh x} \\ \operatorname{csch} x &= \frac{1}{\sinh x} \\ \operatorname{sech} x &= \frac{1}{\cosh x} \\ \operatorname{coth} x &= \frac{\cosh x}{\sinh x}.\end{aligned}$$

An important identity of the hyperbolic functions is

$$\cosh^2 x - \sinh^2 x = 1.$$

1.2.12 Maximums and Minimums

The **absolute maximum** of a function is the highest point on its graph. The **absolute minimum** of a function is the lowest point on its graph. A **local maximum** is the highest point on a graph in some open interval. Technically, $f(c)$ is a local maximum if $f(c)$ is greater than or equal to $f(x)$ when x is near c . A **local minimum** is the lowest point on a graph in some open interval. Technically, $f(c)$ is a local minimum if $f(c)$ is less than or equal to $f(x)$ when x is near c . The first two of these are known as **extreme values**.

Some functions, for example, $\sin x$, take on their absolute minimum and absolute maximum infinitely many times. Other functions, for example, x^3 have neither absolute minimums or absolute maximums. A function that increases to positive or negative infinity on both sides, has

neither an absolute maximum or an absolute minimum. Note that an endpoint of a function cannot be a local maximum, although it can be an absolute maximum.

The **extreme value theorem** (EV) states that if a function is continuous on a closed interval $[a, b]$, then it is guaranteed to have both an absolute maximum and an absolute minimum value. First of all, for the EVT to be valid, the function must be continuous. Continuity and closed intervals are the first things to check for when doing EVT problems. If both of these aren't true, then EVT does not apply to the function.

If a function is either discontinuous or isn't being evaluated in a closed interval, then it does not necessarily have extreme values. If the graph has a hole where an extreme value would be, then the extreme value does not exist. If the graph is undefined at the extreme (for example a vertical asymptote) then it does not have an extreme value there.

Fermat's theorem states that if a function has a local maximum or minimum at a point c , and if $f'(c)$ exists, then $f'(c) = 0$. In other words, if a function has a local maximum or minimum at some point, then the tangent line to the function at that point will have a slope of 0.

Fermat's theorem can help us to start looking for maximums and minimums, but we have to be careful. The equation $f'(x) = 0$ does not necessarily mean that f has a maximum at x . For example, the function x^3 has a derivative of 0 at $x = 0$. However, it does not have a maximum or a minimum at that point. The graph just happens to have a horizontal tangent at 0. Also, if a derivative at a certain point does not exist, it does not necessarily mean that there is no maximum or minimum at that point. For example, the function $|x|$ has an absolute minimum at 0, but it is not differentiable at that point.

The x -values where the derivative is 0 or nonexistent aren't necessarily locations of maximums or minimums, but they are a good place to start looking. These x -values are called **critical values**.

Critical values are found by finding where the derivative is zero and where the derivative does not exist. To find the first, differentiate the function and find the zeros of the derivative. To find the second, find where the derivative does not exist—where the derivative function is undefined (such as in division by 0). For the last part, you can also graph the function to find the asymptotes where the derivative doesn't exist.

When finding critical numbers and your solutions are imaginary then the critical values don't exist. Remember, they have to be in the domain of the function.

To find the extreme values of a continuous function on a closed interval, use the following method

1. Find the values of f at the critical numbers of f in the interval.
2. Find the values of f at the endpoints of the interval.
3. The largest of the values from steps 1 and 2 is the absolute maximum of the function in that interval and the smallest is the absolute minimum.

Careful! In step one, your goal is not just to find the critical values but to evaluate the original function at its critical values.

Remember, when asked for extrema, give the y -values, not the x -values.

When finding critical values, always graph the original function on the calculator to show that your critical values are reasonable.

1.2.13 Mean Value Theorem

Rolle's theorem states that if f is continuous on $[a, b]$, f is differentiable on (a, b) , and $f(a) = f(b)$, then there is a number c in (a, b) such that $f'(c) = 0$. In other words, if two points on a continuous graph are at the same height, then at at least one point between those points, the tangent to the curve will have a slope of zero.

Rolle's theorem is often used to prove by contradiction. For example, for there to be a horizontal tangent between two points on a graph then there must be two points on the graph that are at equal height. If no such points exist, then there's no horizontal tangent. The theorem can also be used to prove that a function has less than 2 roots. If it had two roots, then there must exist a c such that $f'(c) = 0$. If it can be shown that the derivative is never zero, then there cannot be two roots. When using Rolle's theorem, make it explicit that you checked the three conditions.

The **mean value theorem** states that if f is continuous on $[a, b]$ and f is differentiable on (a, b) then there is a number c in (a, b) such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}, \quad \text{or} \quad f(b) - f(a) = f'(c)(b - a).$$

In other words, if you pick any two points on the graph and connect them via a secant line, then there must exist a point between the two points where the tangent line to the curve equals the slope of that secant line.

When using the mean value theorem make it explicit that you checked the continuity and differentiability. In general, the mean value theorem can be interpreted as saying that there is a number at which the instantaneous rate of change equals the average rate of change over an interval.

If the derivative of a function is zero for all values of x in some interval, then the function must be a straight line on that interval. This is obvious, but it leads to

the useful corollary that if $f'(x) = g'(x)$ for all x in (a, b) , then $f - g$ is constant on (a, b) . That is, $f(x) = g(x) + C$.

To show that an equation has exactly one real root, use the following process

1. Deduce from the equation an interval on one side of which the function will always yield one sign and on the other side the function yields the opposite sign.
2. Using that interval, demonstrate the function is continuous and differentiable within the interval, then use Rolle's Theorem in a proof by contradiction to show that two or more roots do not exist in the interval.
3. Then use the Intermediate Value Theorem to show that at least 1 root exists in that interval.

1.2.14 L'Hopital's Rule

L'Hopital's rule tells us that if f and g are differentiable and $g'(x)$ does not equal zero on an open interval I that contains a , then assuming that the left side (see below) evaluates to $\frac{0}{0}$ or $\frac{\pm\infty}{\pm\infty}$, then the limit of the rational function $\frac{f}{g}$ is given by

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)},$$

provided that the limit on the right side exists or evaluates to positive or negative infinity. In other words, if your rational limit evaluates to one of the first two indeterminate forms, then the limit of that quotient is equal to the limit of the quotient of their derivatives.

If a limit evaluates to one of the following results, which are known as **indeterminate forms**, then the derivative is unknown and may or may not exist

$$\begin{array}{c} \frac{0}{0} \\ \frac{\pm\infty}{\pm\infty} \\ 0 \cdot \infty \\ \infty - \infty \\ 0^0 \\ \infty^0 \\ 1^\infty \end{array} .$$

If it's one of the first two forms, then L'Hopital's rule can immediately be applied. If it's one of the other ones, then a manipulation of the limit will usually allow us to put it in one of the forms solvable by L'Hopital's rule.

Note: A limit that evaluates to $\frac{0}{0}$ means a "hole" exists in the graph. Sometimes, the hole can be removed by factoring and cancelling, but other times it cant so then we have to use l'Hospitals Rule.

This method is so much faster and easier than factoring and canceling, but remember, LHR can only be used when the limit evaluates to one of the first two indeterminate forms. The first thing we always do is use

direct substitution to see if the limit evaluates to one of the accepted forms.

Remember, when using LHR, we differentiate the numerator and denominator separately. We do not use the quotient rule. LHR is also valid for one-sided limits and for limits at positive or negative infinity.

LHR can be used repeatedly for the same function—each time it simplifies the function a little further. Provided, of course, that the function still evaluates to one of the two accepted indeterminate forms.

Indeterminate Products

Indeterminate products are limits that evaluate to 0 times positive or negative infinity. With a little algebra, we can manipulate these limit arguments so that they yield one of the first two indeterminate forms allowing us to use LHR.

We manipulate the argument by converting it from a product to a ratio simply by dividing the one product by the reciprocal of the other product. Choose to invert the part that will be the easiest of the two to work with in inverted form.

$$\lim_{x \rightarrow a} fg = \lim_{x \rightarrow a} \frac{f}{\left(\frac{1}{g}\right)}.$$

Indeterminate Differences

These are limits that evaluate to infinity minus infinity. To solve these, convert the difference into a quotient using a common denominator, rationalization, or by factoring out a common factor so that we have an indeterminate type of the first two kinds which can be solved with LHR.

For example, you can rewrite the following limit as

$$\begin{aligned} \lim_{x \rightarrow 0} (\sec x - \tan x) &= \lim_{x \rightarrow 0} \left(\frac{1}{\cos x} - \frac{\sin x}{\cos x} \right) \\ &= \lim_{x \rightarrow 0} \frac{1 - \sin x}{\cos x} = 1. \end{aligned}$$

Note that limits that evaluate to infinity plus infinity are not indeterminate. Their limit is infinity.

Indeterminate Power

The three indeterminate forms 0^0 , ∞^0 , and 1^∞ arise from limits of the form

$$\lim_{x \rightarrow a} [f(x)]^{g(x)}.$$

We need to decompose these with logarithms before we can solve them.

The general process to solve this kind of limit is to make an equation of $y =$ the argument of the function, then take the natural log of both sides. The limit of this new function should yield an indeterminate form that can be worked with. Once the limit of the natural log of the original argument is found, we know that $y = e^{\ln y}$ and so,

$$\lim_{x \rightarrow a} y = \lim_{x \rightarrow a} e^{\ln y}.$$

Example:

Compute the limit

$$\lim_{x \rightarrow 0^+} x^x.$$

We start by letting $y = x^x$. Taking the natural log of both sides gives us $\ln y = x \ln x$. Now, we consider another limit

$$\lim_{x \rightarrow 0^+} \ln y = \lim_{x \rightarrow 0^+} x \ln x = 0 \cdot (-\infty).$$

We now have an indeterminate product form, which we can rewrite as

$$\lim_{x \rightarrow 0^+} \frac{\ln x}{1/x} = -\frac{\infty}{\infty}.$$

Now it is an indeterminate form that can be handled by l'Hopital's rule. Applying l'Hopital's rule gives us

$$\begin{aligned} \lim_{x \rightarrow 0^+} \frac{\ln x}{1/x} &= \lim_{x \rightarrow 0^+} -\frac{1/x}{1/x^2} \\ &= \lim_{x \rightarrow 0^+} -x \\ &= 0. \end{aligned}$$

We know that $y = e^{\ln y}$, so returning to the original limit, we get

$$\begin{aligned} \lim_{x \rightarrow 0^+} x^x &= \lim_{x \rightarrow 0^+} y \\ &= \lim_{x \rightarrow 0^+} e^{\ln y} \\ &= e^{(\lim_{x \rightarrow 0^+} \ln y)} \\ &= e^0 \\ &= 1. \end{aligned}$$

The crucial step

$$\lim_{x \rightarrow 0^+} e^{\ln y} = e^{(\lim_{x \rightarrow 0^+} \ln y)},$$

follows from the **composition law** of limits.

1.2.15 Curve Sketching

Increasing/decreasing test

- If $f'(x) > 0$ on an interval, then f is increasing on

that interval.

- If $f'(x) < 0$ on an interval, then f is decreasing on that interval.

When asked to find the intervals where a function is increasing or decreasing, find the zeros and the undefined points of the derivative function. Mark these on a number line and check the sign of the function in each interval. A positive sign means the original function is increasing in that interval, and a negative sign means the original function is decreasing in that interval.

The **first derivative test** tells us that if c is a critical value of a continuous function f , then

- If f' changes from positive to negative at c , then f has a local maximum at c .
- If f' changes from negative to positive at c , then f has a local minimum at c .
- If f' does not change sign at c , then there is no local minimum or maximum at that point.

If the graph of a function lies above all of its tangent lines on an interval, then it is called **concave up** on that interval. If the graph lies below all of its tangent lines on an interval, then it is called **concave down** on that interval. For concave up, think of a cup—it is the opening, not the hump, that points up. An inflection point is the exact point on a graph where its concavity changes orientation.

The **concavity test** tells us that

- if $f''(x) > 0$ for all x on an interval, then f is concave up on that interval.
- if $f''(x) < 0$ for all x on an interval, then f is concave down on that interval.

To find inflection points, find the zeros of the second derivative function.

The **second derivative test** tells us that assuming that the second derivative of f is continuous near a point c , then

- If $f'(c) = 0$ and $f''(c) > 0$ then f has a local minimum at c .
- If $f'(c) = 0$ and $f''(c) < 0$ then f has a local maximum at c .

The second derivative test is inconclusive when it equals 0 or when the derivative doesn't exist at the point.

Analyzing Graphs Given a function, we will be asked to analyze the graph by finding the following

- Intervals of increase/decrease: Use the increasing/decreasing test. Find the zeros of the first

derivative, mark them on a number line, and evaluate (with the derivative function) a sample from each interval to find the intervals where the derivative is positive and negative. Wherever the derivative is positive, the original function is increasing.

- Local minimums and maximums: These can be deduced from the marked and signed number line from the previous item in this list. Wherever the sign of the first derivative changes, there's a local minimum or maximum. Remember to give these as y -values, not x -values! You have to evaluate the original function at the x -values where the first derivative function changes sign.
- Inflection points: The zeros of the second derivative mark possible inflection points. They're not actual inflection points unless the concavity actually changes. When answering this, give the whole (x, y) coordinates rather than just the x -values. Remember that to evaluate these in the original function to get the y -values rather than in the second derivative.
- Concavity: Wherever the second derivative is greater than zero, the original function is concave up. Find the zeros of the second derivative, mark them on a new number line, and evaluate a sample from each interval.

Doing all of this requires bringing together a lot of prior knowledge. For logs, for example, don't forget to find their domains (where the arguments are zero), and to use limits to find the vertical asymptotes.

To sketch the graph of a function, follow the following steps

1. Find the domain of the function.
2. Find the x and y intercepts. To find the x -intercepts, find where $f(x) = 0$ and to find the y -intercepts, find $f(0)$.
3. Check for symmetry. There are three kinds of symmetry:
 - Even (y -axis symmetry), where $f(-x) = f(x)$
 - Odd (rotational symmetry), where $f(-x) = -f(x)$
 - Periodic symmetry, where $f(x) = f$
4. Find the asymptotes:
 - Horizontal: If

$$\lim_{x \rightarrow \pm\infty} f(x) = L,$$

then $y = L$ is a horizontal asymptote. For rational functions, the horizontal asymptotes can be deduced by looking at the degrees of the function. If the degree of the numerator is less than the degree of the denominator, then $y = 0$ is the horizontal asymptote. If the degree of the numerator is equal to the degree of the denominator, then the horizontal asymptote is equal to the leading coefficient of the numerator divided by the leading coefficient of

the denominator. Check for the infinite limits to find HAs. Remember how to manipulate rational functions by dividing each term by the highest power of x . Also remember how to bring this power of x in under a radical. Also keep your domain in mind. If x is bounded on both sides then obviously it cannot go off to infinity to make a horizontal asymptote.

- Vertical: If

$$\lim_{x \rightarrow \pm a} f(x) = \pm\infty,$$

then $x = a$ is a vertical asymptote. For rational functions, the vertical asymptotes are simply the zeros of the denominator. Mark any vertical asymptotes on your interval line so you don't mistake it for a local maximum or minimum later on. Be sure not to misinterpret holes as asymptotes. If its a rational function be sure to factor the top and bottom to find any holes in the graph.

- Slant: If

$$\lim_{x \rightarrow \pm\infty} [f(x) - (mx + b)] = 0,$$

then $y = mx + b$ is the slant asymptote. For rational functions, you'll have a slant asymptote when the degree of the numerator is exactly one more than the degree of the denominator. In those cases, the slant asymptote can be found simply by doing long division of the rational function. If the long division yields a remainder, create a new function with the remainder included then deduce whether the function will approach the remainder goes to nothing as a approached positive and negative infinity.

5. Intervals of Increase and Decrease. Find the critical values (zeros and undefined of the first derivative), and evaluate the intervals to find where the first derivative is positive and negative.
6. Local minimums and maximums. See #5. Find where the first derivative crosses the x -axis.
7. Concavity and inflection points. Find the zeros of the second derivative and evaluate the intervals where $f''(x)$ is positive and negative. Remember to get the y -value of the inflection points by evaluating the x -value in the original function. Note that a vertical asymptote is not an inflection point even if the concavity changes direction there.
8. Sketch the function.

Check the graph on a calculator before starting the problem. This way you know the kind of graph that you're aiming for, and you can catch your mistakes before wasting too much time on them.

Remember how to evaluate limits at infinity of rational functions—divide top and bottom by the highest power of x .

Remember that there are 13 values to find to sketch the graph of a function.

The most difficult part of this is solving the trigonometric equations. Here are some tips

- Sometimes, the equations are too difficult to solve, in those cases, you may be able to draw a good enough graph without them.
- Sometimes, you can deduce the answer. For example, if you're trying to find the zeros but you can deduce that the function will always be positive or negative then you know there aren't zeros. Alternatively, if one function is always less than another function then you can tell whether or not the difference of the two will ever be positive/negative.

1.2.16 Optimization

Follow the following procedure to solve optimization problems:

1. Understand the problem
2. Draw a diagram
3. Introduce notation by assigning variables
4. Express the quantity to be maximized or minimized in terms of the variables
5. Put the equation in terms of only one variable
6. Find the absolute minimum or maximum. This is done by finding the critical points of the first derivative and evaluating these as well as the endpoints in the original equation.

The most creative part is #5. Keep the unit circle in mind, sometimes you can use it to reduce an equation to one in a single variable.

Note: The area of almost all the common geometric shapes can be derived from the area of a rectangle. For example, for a parallelogram, if you cut a triangle off of one end and add it to the other end, you have a rectangle with area $A = bh$. For a trapezoid, if you copy the shape, rotate one of them and join them together, you get a parallelogram. From this you can deduce the formula for the area of a trapezoid

$$A = \frac{b_1 + b_2}{2}h.$$

When dealing with problems involving circles, keep the following in mind

- Think about triangles connecting the center of the circle to points on the rim.
- Remember, a point on the rim of a circle can be represented as $(r \cos \theta, r \sin \theta)$ where r is the radius of the circle.

When dealing with problems involving triangles, keep the following in mind

- Use the Pythagorean theorem to simplify triangles (i.e. remove a variable)

- Think about similar/congruent triangles, i.e. ratios
- Think in terms of angles and trig functions

When working with revenue, remember that revenue equals unit price times units sold. However, when you want to maximize the revenue function let x be the dollars discounted rather than the units sold. Build a revenue function from that.

Tip:

When you're done, refer to the original question to make sure you're answering the question and not something else.

My Procedure:

1. Draw a diagram and label everything with variables or given information
2. Identify the variable being maximized or minimized
3. Write the equation in terms of a single variable
4. Graph it on the calculator to ensure that the graph is reasonable, based on the approximate min/max you can estimate from looking at the drawing.

Tip:

Don't forget the units!

1.2.17 Newton's Method

Newton's method is a procedure that allows us to approximate the roots of an equation. You can get within arbitrary distance of a root by repeated application of

Newton's method.

The method works like this: If you guess the root, and then take the derivative of the function at that root, then the root of the derivative will be closer to the root of the function than your original guess was.

Use the equation

$$x_2 = x - \frac{f(x)}{f'(x)},$$

where x is your guess for the root of the equation. Plug the answer back into the equation as your next guess, and so on. In most cases, the result will be that you converge quickly to the root of the original function.

If you want to be accurate to a specific number of decimal places, repeat the procedure until you get to the point where a guess and the subsequent result agree to that many decimal places.

When using the procedure with a calculator, follow this procedure:

1. Take the derivative of the function
2. Enter the right side of the above equation into the calculator
3. Try a guess and then input the results repeatedly

Observe that:

$$\begin{aligned} a^b &= x \text{ means } x^{\frac{1}{b}} - a = 0 \\ \sqrt[b]{a} &= x \text{ means } x^b - a = 0. \end{aligned}$$

You can use these relationships to approximate the value of something, for example $\sqrt[5]{5}$, by treating that value as the root of an equation and then using Newton's method on the equation.

1.3 Integrals

1.3.1 Antiderivatives

Function A is an **antiderivative** of function B if function B is the derivative of function A. For example, x^2 is the antiderivative of $2x$ since $2x$ is the derivative of x^2 .

The antiderivative of the velocity function gives us the position function and the antiderivative of the acceleration function gives us the velocity function.

To find the antiderivative of a function you take the rules of differentiation into reverse.

Tip:

Remember, for the general antiderivative you need to add the constant C, sometimes called *Charlie*.

Tip:

When doing these, check your answer by differentiating.

When finding the antiderivative of $\frac{1}{x}$ you have to use $\ln|x|$, but when you find the antiderivatives of higher powers of inverse x such as $\frac{1}{x^2}$, then you can use the normal power rule for derivatives in reverse. If you try the normal (i.e. power) method, you're left with a division by 0, which is not allowed.

1.3.2 Area under a Curve

How do you calculate the area under a curve? Specifically, how do you calculate the area between a function and the x -axis within a specified interval? First of all, why would we want to calculate the area? Well, the area under a velocity curve gives the total distance traveled. The area under the curve of a function is the output of the function whose derivative is the original function.

One way to estimate the area under a curve is to separate the interval into rectangles and sum the areas of the rectangles. However, we can choose the heights of the rectangles to be either the values of the function at the top-left corners of the rectangles or the top right corners of the rectangles. If the function is either increasing or decreasing through the overall interval, then using left points will give either the upper bound of the area or the lower bound of the area.

The true area lies somewhere between the area calculated by using the left endpoints of the rectangles for their heights and the area calculated by using the right endpoints of the rectangles for their heights.

There are several ways we can increase the accuracy

of our area estimation

- Dividing the interval into more rectangles
- Averaging the values received by using the left and right endpoints
- Using the midpoints of the rectangles instead of the endpoints
- Using trapezoids instead of rectangles

The width of rectangles for an interval $[a, b]$ is given by

$$\Delta x = \frac{b - a}{n} = \frac{\text{length of interval}}{\text{number of rectangles}}.$$

The more rectangles we divide the interval into, the more accurate our result is, then the area under the curve is the limit of the sums of the areas of the approximating rectangles as the number of rectangles is taken to infinity.

Instead of using the left or right endpoints of the rectangle to define their height, we can actually take the height of the rectangle to be the value of f at any number x_k^* within the small intervals. These numbers are called the sample points. We find the lower sum of the area by using the shortest possible rectangles and we find the upper sum by using the tallest possible rectangles.

An equivalent definition of the area under a curve, then, is the unique number that is smaller than all possible upper sums and larger than all possible lower sums.

The approximate area under the curve $f(x)$ using rectangles is given by

$$\text{Area} \approx \sum_{k=1}^n \frac{b-a}{n} f(x_k^*).$$

Note that $\frac{b-a}{n}$ is the width of each rectangle and $f(x_k^*)$ is the height of the rectangles. Overall, the summation gives the approximate area under the curve of a function that lies above the x -axis. The whole thing is called a **Riemann sum**.

As the number of rectangles is taken to infinity, the approximate area becomes the exact area under the curve.

$$\text{Area} = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{b-a}{n} f(x_k^*) = \int_a^b f(x) dx.$$

The thing on the right is known as the **definite integral** of $f(x)$ on the interval $[a, b]$.

The area under a curve has other practical uses. The area under the curve of a variable force function is the work performed. The area under the curve of a velocity function is the displacement.

1.3.3 Definite Integrals

The definite integral arises when we compute area, distance, lengths of curves, volumes of solids, centers of mass, force due to water pressure, and work as well as other quantities.

If $f(x)$ is a function defined on the interval $[a, b]$, then we can divide that interval into n subintervals of equal width Δx . We then choose sample x -values x_k^* in each of those subintervals. Then the definite integral of $f(x)$ from a to b is

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k^*) \Delta x,$$

where $\Delta x = \frac{b-a}{n}$ and $x_k = a + k\Delta x$.

What we are doing here can easily be visualized with a graph of an arbitrary function $f(x)$. We choose an interval $[a, b]$ over which to integrate $f(x)$. Next, we divide the

interval into a certain number of equal-width subintervals and choose some x -value from each of these subintervals. Next, we evaluate $f(x)$ at each of these x -values and multiply it by the width of the subinterval. This is effectively the area of the rectangle by our subinterval bounded at the bottom by the x -axis and the top by the graph of the function. Finally, we add up these values for each of the subintervals. What we end up with is a crude approximation of the area between the graph and the x -axis in the interval $[a, b]$. If we increase the number of subintervals between a and b to infinity, that is, we make the rectangles infinitely narrow, then our approximation becomes the exact area under the curve—which is the definite integral.

The **Riemann sum** is an approximation of the net area under a curve, and its limit is the definite integral as shown in the formula above. The Riemann sum approximates the area under a curve by dividing the area into rectangles and summing the areas of the rectangles.

Example:

Integrate

$$\int_0^4 x^2 dx,$$

using the Riemann sum definition of the definite integral.

First, we have to determine our Δx , which is

$$\Delta x = \frac{b-a}{n} = \frac{4-0}{n} = \frac{4}{n}.$$

Next, we have to figure out what our x_k are. This is the hardest and the most important part. Note that x_k is the k 'th value of x . In other words, it's the sample points. Note that the first sample point will appear at the left end of the interval... at 0 in this case, and every sample point after that will have an additional Δx or width of a rectangle added to it. Think of it like this

$$\begin{aligned} x_1 &= a \\ x_2 &= a + \Delta x \\ x_3 &= a + 2\Delta x \\ &\vdots \\ x_k &= a + k\Delta x. \end{aligned}$$

In our case, the left side of the interval is $a = 0$. So

$$x_k = a + k\Delta x = \frac{4k}{n}.$$

So our integral becomes

$$\begin{aligned} \int_a^b f(x) dx &= \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k^*) \frac{b-a}{n} \\ \int_0^4 x^2 dx &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \left(\frac{4k}{n} \right)^2 \frac{4}{n} \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{64}{n^3} k^2 \\ &= \lim_{n \rightarrow \infty} \frac{64}{n^3} \sum_{k=1}^n k^2 \\ &= \lim_{n \rightarrow \infty} \frac{64}{n^3} \frac{n(n+1)(2n+1)}{6} \\ &= \lim_{n \rightarrow \infty} \frac{64}{6} \left(2 + \frac{3}{n} + \frac{1}{n^2} \right) \\ &= \frac{64}{3}. \end{aligned}$$

Note: Review the rules for manipulating sums and for solving certain sums in the precalculus chapter.

Tip:

Remember, the definite integral is just the net area between a curve and the x -axis.

number of jump discontinuities, then it is integrable on that interval.

Not every function can be integrated. If a function is continuous on an interval or it has only a finite num-

ber of jump discontinuities, then it is integrable on that interval. Although the definition of the definite integral defines the subintervals as being of equal widths, they don't necessarily have to be. As long as the width of the widest interval approaches zero in the limit, then the integral exists. If the width of the widest rectangle approaches 0,

this implies that the number of rectangles in the interval is approaching infinity.

To convert a limit of sums into an integral, replace $\lim \sum$ with the stylized integral \int ; replace x_k^* with x and replace Δx with dx .

The connection between derivatives and integrals is the connection between the slope of a curve and the area under the curve. More surprising than this connection is the sheer application of these concepts in the real world. For example, the slope of the velocity function gives acceleration and the area under the curve of the velocity function gives distance traveled. Looking at a single function you can see three different levels—the function (e.g. velocity) its slope (i.e. acceleration) and the area under it (i.e. distance traveled). The graph of a function takes on a whole new meaning. Even physical phenomena take on a new meaning—distance, velocity, and acceleration can be visualized with a single mathematical function. Know one and you know all three of them.

Following are the properties of definite integrals that need to be remembered

$$\begin{aligned}\int_a^b f(x) dx &= -\int_b^a f(x) dx \\ \int_a^a f(x) dx &= 0 \\ \int_a^b c dx &= c(b-a) \\ \int_a^b [f(x) + g(x)] dx &= \int_a^b f(x) dx + \int_a^b g(x) dx \\ \int_a^b cf(x) dx &= c \int_a^b f(x) dx.\end{aligned}$$

These properties can all be visualized by remembering that the integral gives the net area.

Note: An integral containing the product of two functions cannot be split into a product of two integrals.

When doing definite integration, always graph your integrand to make sure that your solution makes sense. This way you'll catch yourself if the integrand is not actually continuous there, and you'll be able to estimate the area under the curve visually to ensure that your solution makes sense.

Note: Definite integration on the TI-84 is not exactly accurate. For the function

$$\int_0^1 \frac{\sqrt{x}}{1+x^3} dx,$$

the TI gives the right answer to only the 3rd decimal place. The digits after the 3rd decimal place are not accurate. When using this method to check indefinite integrals, if the TI gives the same answer to the first couple of decimal places, then it's probably the right solution. This is when using the integration feature under the calc menu (after entering an equation in Y=). Using fnInt

from the Math method yields an answer that is correct to 6 decimal places (with 4 erroneous digits after that point).

1.3.4 Fundamental Theorem of Calculus

The first part of the **fundamental theorem of calculus** states that if f is continuous on $[a, b]$, then

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

is also continuous on $[a, b]$ and is differentiable on (a, b) with $g'(x) = f(x)$.

We can think of $g(x) = \int_a^x f(t) dt$ as the *area so far* function, since on some interval $[a, b]$, $g(x)$ gives the area under the curve from a up to x .

If a function is continuous on an interval then its integral is continuous and differentiable on that interval.

The FTC establishes a connection between differential and integral calculus—between instantaneous slope and area under curves—between the instantaneous rate of change of a function and its antiderivative.

Differentiation and integration are inverse processes much like addition and subtraction.

The FTC allows us to compute areas and integrals without having to compute them as limits of sums.

If x is a number, then so is the entire integral. If x is a variable, then so is the entire integral—defining a function of x . Such an integral (as given above on the right side of the equation) is an antiderivative of its integrand. If you graph the integral and the integrand, you'll see that the integrand is the derivative of the integral. That is to say that $g'(x) = f(x)$.

In other words, the derivative of a definite integral with respect to its upper limit is the integrand evaluated at the upper limit.

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

Note: When finding the derivative of definite integrals, there's no need to actually differentiate the integrand. The derivative is simply the integrand (substituting the variable for x , of course).

If the upper limit is not just x then you may have to use the chain rule when differentiating a definite integral. Essentially, you replace b with u , do what you normally do, and then differentiate u as in the chain rule.

Example:

Differentiate

$$\int_a^{x^2} f(t) dt,$$

with respect to x .

Since the upper variable is not just x , we have to use the chain rule

$$\frac{d}{dx} \int_a^{x^2} f(t) dt = f(x^2) \cdot 2x.$$

Example:

Differentiate

$$\int_a^{\sin x} t^2 dt,$$

with respect to x .

Making the substitution $u = \sin x$ and using the chain rule and the fundamental theorem of calculus, we get

$$\begin{aligned} \frac{d}{dx} \int_a^{\sin x} t^2 dt &= \frac{d}{du} \left[\int_a^u t^2 dt \right] \frac{du}{dx} \\ &= u^2 \frac{du}{dx} \\ &= \sin^2 x \cos x. \end{aligned}$$

Keep in mind that when differentiating an integral with this method, the derivative must be with respect to the variable in the top position. For example

$$\frac{d}{dx} \int_a^y f(t) dt = 0,$$

because the derivative is with respect to x , but the integral does not depend on x , it depends on y .

If the variable is on the bottom instead of on the top, simply switch them and put a negative sign front of the integral.

If the variable appears in both the top and the bottom limits, split the integral into multiple integrals. For example

$$\int_{ax}^{bx} f(t) dt = \int_{ax}^c f(t) dt + \int_c^{bx} f(t) dt.$$

The second part of the **fundamental theorem of calculus** states that if $f(x)$ is continuous on $[a, b]$, then

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

where $F(x)$ is any **antiderivative** of f .

This allows us to evaluate integrals much more easily than computing a limit of sums.

Since the definition says *any* antiderivative, we don't need to add the constant Charlie to our answer.

The FTC requires that the function is continuous on the interval being evaluated. For example, using the FTC on

$$\int_{-1}^1 \frac{1}{x^2} dx,$$

will yield a false answer because the integrand is not continuous on that interval.

As we have seen, differentiation and integration are inverse processes.

To get a better grasp of the concept, always visualize the integral as the net area of the integrand as you evaluate the integrals.

To double check your integrations, use your TI-84. First enter the integrand as an equation, then go into **CALC > 7:**. After hitting enter, key in your lower bound and then your upper bound.

1.3.5 Indefinite Integrals

Recall that a definite integral gives the net area between a lower and an upper limit and it is evaluated by subtracting the antiderivative at the upper limit from the antiderivative at the lower limit

$$\int_a^b f(x) dx = F(b) - F(a) = F(x) \Big|_a^b.$$

An indefinite integral is the function of the net area—that is, it has no upper or lower limits, and as such, it's just another way of representing an antiderivative

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

To solve an indefinite integral, you can find the antiderivative of the function using a table of antiderivatives/indefinite integrals.

Tips for evaluating integrals:

- If the integrand contains radicals, convert them to inverse powers so the power antiderivative rule can be applied.
- If the integrand contains quotients (i.e. x in the denominator) convert them to inverse powers so the power antiderivative can be applied. Also, you may be able to cancel stuff if you expand a quotient into multiple quotients.
- If the integrand contains a product or a number of terms all raised to a power (e.g. a binomial squared) which we can't solve yet, expand it out and then apply the antiderivative rules.
- Keep the inverse trig functions in mind.

- Review trig identities so that you can simplify integrands with trig functions. For example, we have no antiderivative rules for $\sin(2x)$, but $\sin(2x) = 2 \sin x \cos x$ which might help simplify the integrand by allowing one or more of the functions to cancel out.
- Memorize the derivatives and integrals of all the trig functions.
- Check your answers by differentiating them.
- Don't forget Charlie!!!

The fundamental theorem of calculus can be rewritten as

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

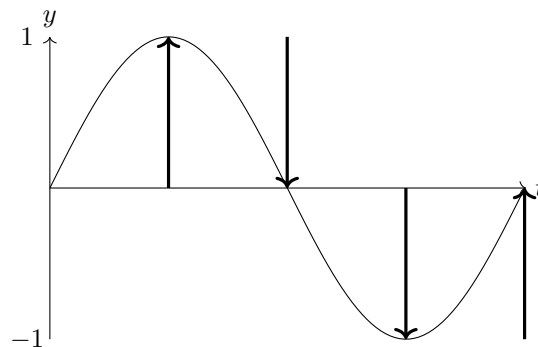
In words, the integral of a rate of change is the net change. This is the **net change theorem**. In other words, the integral of a derivative is the net change of the original function.

Here are some real-world examples

- The integral of the rate at which water flows into a tank, gives the net change in volume.
- The integral of the rate of a chemical reaction gives the net change in concentration of the product.
- The integral of the linear density of a rod gives the net change in mass.
- The integral of the rate of growth of a population gives the net change in population.
- The integral of the marginal cost of a commodity gives the net change in cost of production.
- The integral of the velocity of an object moving in a straight line gives the net change in position—or displacement of the object.
- The integral of the rate of change of charge (i.e. current) gives the net change of charge.

If we want to calculate the total change (e.g. distance traveled) rather than net change (e.g. displacement), then we need to take into account the sum positives and negatives rather than just the difference between the first and the last. You do this by integrating the absolute value of the rate of change instead of just integrating the rate of change.

Consider a function such as $f(x) = \sin x$ that gives the vertical displacement of some object. Below is a graph of the first cycle of this function. At time zero, the object is at position zero. Then it begins moving upward (the first arrow) for a little time. Once it reaches the top it starts going down (the second arrow) passing through the zero-height (third arrow) until it reaches its low point. Finally, it heads back up (fourth arrow).



The **net change** (the integral of the derivative of the function) is the sum of the up arrows minus the sum of the lengths of the down arrows, which is $1+1-(1+1) = 0$. If we integrate the derivative ($\cos x$) on the same interval ($[0, 2\pi]$), we get the same value. This corresponds to the **displacement** of the object. In this case, the displacement is zero, because the object ends up at the same height that it started from. On the other hand, the **total change** (the integral of the absolute value of the function) is the same as the lengths of all the arrows. In this case, $1 + 1 + 1 + 1 = 4$. In this case, the total change corresponds to the total distance traveled by the object.

In other words, the sum of the absolute values of the extreme positions (i.e. total length of the arrows) of a function equals the area under the curve of the absolute value of the derivative of the function.

To integrate the absolute value of a function, graph the function. Graph function without the absolute value and then reflect all negative parts across the x -axis. Note where the zeros of the original function are, because this is where you break your integral into multiple integrals to calculate the net area of the absolute value function. For example,

$$\int_{-1}^1 |x| dx = \int_{-1}^0 -x dx + \int_0^1 x dx.$$

The only zero of the function $y = x$, is at 0, so we break the integral into two integrals, one for the interval left of zero and the other for the integral to the right of zero. Notice that we reflect the part left of zero about the x -axis by negating the entire function. When finding total change, you'll usually have to evaluate multiple integrals.

When trying to understand the relationships between real world functions and their derivatives and integrals, think in terms of units. Consider the velocity function, for example. The units of the velocity function $v(t)$ are meters per second or meters/second, and the independent variable is time, measured in seconds. The units of the derivative of the velocity function can easily be determined as follows

$$\frac{dv}{dt} = \frac{\text{meters/second}}{\text{second}} = \frac{\text{meters}}{(\text{second})^2}.$$

So the units of the derivative of the velocity function are meters per second squared, which is acceleration. When

evaluating units, we just replace a differential such as dv with the units of v . What about the integral of the velocity function? Using the same method of replacing the variables by their units and then simplifying, we find that the integral of the velocity function

$$\int v dt = \int \frac{\text{meters}}{\text{second}} \text{second} = \text{meters},$$

gives us something with units of meters. That is, it gives us a measure of distance or displacement.

The hardest part about all this is translating the real world problems into the language of math, but it doesn't have to be. The words give context, but the process of solving such problems is always the same.

How to check integrations

- For definite integrals, graph the integrand and visually estimate the area under the curve between the limits to see if it matches your result. If you have a TI-84 graphing calculator, enter the integrand as an equation then go into the Calc menu and select the 7th option before keying in your lower and upper limits.
- For indefinite integrals, differentiate your answer to see if it matches the original integrand.

Not all functions are integrable. For example,

$$\int e^{-x^2} dx,$$

is not integrable, but

$$\int_0^{\infty} e^{-x^2} dx = \frac{\pi}{2}.$$

The following interesting function was created by Dirichlet. It cannot be integrated using Riemann's method of a limit of sums, but it can be integrated—demonstrating that Riemann's method is not sufficient to integrate all functions

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

The function would look like an infinitely dotted line at $y = 1$ and a dotted line at $y = 0$. There are infinite rational numbers between any two numbers and there are also infinite irrational numbers between any two numbers. However, the infinity of irrational numbers is much larger. If the top function is subtracted from the bottom function, you'd be left with a line at $y = 0$. If the bottom function is subtracted from the top function, you'd be left with nothing visible.

Be aware of equivalent integral solutions. The answer you get may be different from the answer the textbook gives, but that doesn't necessarily mean you're wrong. For example,

$$1 - \ln(\cos x) + C = \ln(\sec x) + C.$$

The leading 1 in the first equation disappeared because the C is already an arbitrary constant. In other words $1 + C = C$. The argument of the natural log changed from cosine to secant because of a rule of logarithms. The negative in front of the log can be taken to mean the argument of the log raised to the power of negative one, and one over cosine is secant. In much the same way, the following equations are equivalent

$$-\ln(e^x - x + 1) + C = x - \ln(e^x + 1) + C.$$

One way to check to see if your solution is equivalent to the textbook solution is to graph both of them. If they're the same or if they only differ by a constant, then they're equivalent. Another way to check is to plug random numbers into both. If the results are the same, chances are the equations are equivalent. If they're not the same, check another random number and if the results differ only by a constant, the equations may be equivalent.

To do a quick calculator check of an indefinite integral when it's difficult or time consuming to check it exactly by differentiation, pretend that it is a definite integral. Pick a top and bottom number such as 1 to 3.5 and do a definite integration in your calculator. Then enter your solution equation $F(x) + C$ and test to see if $F(3.5) - F(1)$ yields the same answer.

1.4 Integration Techniques

1.4.1 Substitution

The substitution rule allows us to integrate things that previous knowledge didn't allow us to. It states that if $u = g(x)$ is a differentiable function whose range is some interval and f is continuous on the same interval, then

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

Here, we treat dy/dx as differentials (or like variables in their own rights), meaning they can be treated as separable variables. The goal is to substitute u and du for the original integrand and differential.

Example:

Evaluate

$$I = \int x^3 \sqrt{x^2 + 5} dx.$$

If we make the substitution $u = x^2 + 5$, then the derivative of this is $\frac{du}{dx} = 2x$. Solving for dx gives us $dx = \frac{1}{2x} du$. Making these substitutions gives us

$$I = \int x^3 \sqrt{u} \frac{1}{2x} du = \frac{1}{2} \int x^2 \sqrt{u} du.$$

We still have an x^2 that we have to get rid of. From our substitution, we know that $x^2 = u - 5$. Substituting this in gives us

$$\begin{aligned} \frac{1}{2} \int (u - 5) \sqrt{u} du &= \frac{1}{2} \int u \sqrt{u} du - \frac{5}{2} \int \sqrt{u} du \\ &= \frac{1}{2} \int u^{\frac{3}{2}} du - \frac{5}{2} \int u^{\frac{1}{2}} du. \end{aligned}$$

This is now an integral that we can easily solve.

Don't forget to take the u out in the end and put back in what you substituted for earlier. Also, don't forget Charlie!

Finding the right substitution is not always straightforward. First try to substitute something whose differential (less a constant) already appears, or substitute a complicated internal function. Be able to find the outside function. Make sure the derivative of the inside is part of the function.

Remember that an integral of sums can be split into a sum of integrals.

For definite integrals, the only change is the limits

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

When using this rule with definite integrals, everything including the limits must be in terms of u . To convert

them, plug the a and b into the equation for u that you plan to use.

In the end, we don't put x back into the equation, instead, we simply evaluate the integral at the limits ending up with a definite number.

If $f(x)$ is an even function, then

$$\int_{-a}^a f(x) dx = 2 \int_0^a f(x) dx.$$

If $f(x)$ is an odd function, then

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

If you visualize the symmetry and think of the definite integral as the area under the curve, then the above two facts become obvious. Even functions centered at 0 will have an area two times the area of the right half, and odd functions centered at 0 will cancel each other out.

Whenever you do a substitution, don't forget to back-substitute at the end if necessary.

1.4.2 Integration by Parts

Integration by parts is an integration rule that corresponds to the product rule of differentiation. The formula for integration by parts is given as

$$\int f(x)g'(x) dx = f(x)g(x) - \int g(x)f'(x) dx.$$

A good way of remembering the formula is as

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

Tip:

Choose for $f(x)$ (i.e. u) the function that becomes simpler when integrated because on the other side the integral will contain $f'(x)$ instead of $f(x)$.

Typically, integration by parts is useful when the integrand is a product of a function of x and a transcendental function such as the exponential, a trigonometric, or a logarithmic function.

You can derive this equation by putting the product rule in integral notation and rearranging it. You could use any permutation of the above three terms, which means it actually doesn't matter whether $\int u dv$ comes first or $\int v du$ as long as a function is with the differential of the other function.

Tip:

You may have to do successive integrations by part to solve a single problem. In such a case, integration by parts often acts as a reduction formula as it may be used to reduce the exponent on a variable by one each time that integration by parts is performed.

Overall, there are three cases in which integration by parts may be useful:

1. It leads to an integral on the right side that you already know the value of.
2. It reduces an exponent that is holding you back, that is, it acts as a reduction formula. In this case, we use integration by parts repeatedly until the exponent becomes 0.
3. On the right side you get the negative of or some multiple of the original integral. In this case, you can solve simply by combining the left and right in-

tegrals. With this method, remember to add the constant of integration on the right side!

Tip:

Although integration by parts is the integration of a product of two functions, it can also be used to integrate single functions such as $f(x) = \ln x$ if we take the second function to be $g(x) = 1$.

To evaluate definite integrals by part, use the form

$$\int_a^b u dv = uv \Big|_a^b - \int_a^b v du.$$

To check your work, you can differentiate the results for indefinite integrals to see if you arrive back at the original integrand. For definite integrals, check the integration with your calculator.

Example:

Integrate

$$\int (\ln x)^2 dx.$$

We start by making the substitution $z = \ln x$ then $dz = \frac{1}{x} dx$ or $dx = x dz$. To replace the x , we solve $z = \ln x$ for x to get $x = e^z$, so our integral with the substitutions is

$$\int z^2 e^z dz.$$

Now, we can use integration by parts. We let $u = z^2$ because we want to reduce the power on z , then $dv = e^z$, $v = e^z$ and $du = 2z$, which gives us

$$\int z^2 e^z dz = z^2 e^z - 2 \int z e^z dz,$$

when we apply the formula for integration by parts. Now we integrate the remaining integral on the right

side by parts again letting $u = z$, $dv = e^z$, $v = e^z$, and $u = 1$, to get

$$\int z^2 e^z dz = z^2 e^z - 2 \left[z e^z - \int e^z dz \right].$$

We've reduced our integral to a known integral so

$$\int z^2 e^z dz = z^2 e^z - 2z e^z + 2e^z + C.$$

Plugging our substitution $z = \ln x$ back in gives us our solution

$$\begin{aligned} \int (\ln x)^2 dx &= (\ln x)^2 e^{\ln x} - 2e^{\ln x} \ln x + 2e^{\ln x} + C \\ &= x(\ln x)^2 - 2x \ln x + 2x + C. \end{aligned}$$

1.4.3 Integrating Logarithms

The rules for manipulating logarithms allow us to make some nice simplifications when integrating them. Here are some examples.

The integral of the natural logarithm is

$$\int \ln x dx = x \ln x - x + C.$$

Logarithm of a Root

$$\begin{aligned} \int \ln \sqrt[n]{x} dx &= \int \ln x^{\frac{1}{n}} dx \\ &= \int \frac{1}{n} \ln x dx \\ &= \frac{1}{n} \int \ln x dx \end{aligned}$$

This same technique can be used to simplify the integral of something like $\ln x^n$.

Logarithm of a Product

If k is a constant:

$$\begin{aligned}\int \ln kx \, dx &= \int (\ln k + \ln x) \, dx \\ &= x \ln k + \int \ln x \, dx\end{aligned}$$

Logarithms Raised to a Power

For integrals of the form

$$\int (\ln x)^n \, dx,$$

you can start by making the substitution $z = \ln x$, $dz = \frac{1}{x} dx$, and $x = e^z$, then it becomes

$$\int (\ln x)^n \, dx = \int z^n e^z \, dz,$$

after which you can use integration by parts to reduce the n on the z .

1.4.4 Integrating Trigonometric Functions

To integrate trig functions, we often need to use one or more of the following identities

$$\begin{aligned}\sin^2 x + \cos^2 x &= 1 \\ \sin^2 x &= \frac{1 - \cos 2x}{2} \\ \cos^2 x &= \frac{1 + \cos 2x}{2} \\ 2 \sin x \cos x &= \sin 2x \\ \sin A \cos B &= \frac{1}{2} \sin(A - B) + \frac{1}{2} \sin(A + B) \\ \sin A \sin B &= \frac{1}{2} \cos(A - B) - \frac{1}{2} \cos(A + B) \\ \cos A \cos B &= \frac{1}{2} \cos(A - B) + \frac{1}{2} \cos(A + B).\end{aligned}$$

Below we consider different types of trigonometric integrals and give strategies for evaluating each of them.

For integrals of the form

$$\int \sin^m x \, dx \quad \text{or} \quad \int \cos^m x \, dx,$$

- If m is an odd integer, use the identity $\sin^2 x + \cos^2 x = 1$ to simplify the integrand into something that can be solved. You want to be left with the original function to the single power and the pairs of powers to be converted to the other function. For example, $\int \sin^5 x \, dx = \int (1 - \cos^2 x)^2 \sin x \, dx$. This allows you to use the substitution $z = \cos x$ after expanding it all out.

- If m is an even integer, use a half angle identity to manipulate the integrand into something that can be solved. For example, $\int \cos^4 x \, dx = \int \left(\frac{1 + \cos 2x}{2}\right)^2 dx$. After expanding it out, you're left with multiple simpler integrals. If m is larger than 2, you'll have to reapply this procedure on one or more of the simpler integrals, since this is a "reduction" method. If you have to apply the procedure again, remember if the argument of the cosine squared function is not simply x , you must adjust the argument of the resulting $\cos 2x$ identity.

For integrals of the form

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

- If either m or n are odd integers, use the identity $\sin^2 x + \cos^2 x = 1$ to manipulate the integrand. The goal here is to manipulate the integrand so you're left with one function to the power of 1. Then you can apply simple z -substitution. For example, if the integral is $\int \sin^4 x \cos^5 x \, dx$ convert it to $\int \sin^4 x (1 - \sin^2 x)^2 \cos x \, dx$, which simplifies to $\int \sin^4 x \cos x \, dx - \int \sin^6 x \cos x \, dx$ both of which can easily be solved with the simple substitution $z = \sin x$.
- If both m and n are even integers, the problem is a little more difficult. In this case, you have to use half angle identities to manipulate the integrand. Typically, you'll have to simplify both functions with their half-angle counterparts and if the powers are higher than 2, you'll have to repeat the process until you reach the point where neither function is raised to a power.
- If m and n are even integers and equal to each other, you can also use the identity $\sin 2x = 2 \sin x \cos x$ to simplify the integral. You'll probably have to simplify again using one of the other half angle identities.

For integrals of the form

$$\int \tan^m x \sec^n x \, dx,$$

- If n is an even integer, save one factor of $\sec^2 x$ and convert any remaining pairs to functions of tangent by using the identity $1 + \tan^2 x = \sec^2 x$. Then use simple z -substitution with $z = \tan x$ and $dz = \sec^2 x \, dx$.
- If m is an odd integer, save one factor of $\sec x \tan x$ and convert any remaining pairs of tangent to secant by using the variant of the identity $\tan^2 x = \sec^2 x - 1$. Then use simple z -substitution with $z = \sec x$ and $dz = \sec x \tan x \, dx$.
- If m is an even integer and n is an odd integer, convert everything to sine and cosine and cross your fingers, hoping you can solve it using integration by parts.

For integrals of the form

$$\int \cot^m x \csc^n x dx,$$

you use the same process as solving for $\int \tan^m x \sec^n x dx$ except you use the identity $\cot^2 x + 1 = \csc^2 x$.

For integrals of the form

$$\int \sec^m x dx,$$

- For $m = 1$, it must be remembered that $\int \sec x dx = \ln |\sec x + \tan x| + C$. In reality, this is done by multiplying the integrand by $(\sec x + \tan x)/(\sec x + \tan x)$ and then using z -substitution, setting $z = \sec x + \tan x$.
- For $m = 3$, use integration by parts, using $u = \sec x$ and $dv = \sec^2 x dx$. In the resulting integral, substitute $\sec^2 x - 1$ for $\tan^2 x$, and you're left with a simpler expression in terms of twice the original integral.
- If m does not equal 1 or 3, there's no easy way to solve the problem.

For integrals of the form

$$\int \csc^m x dx,$$

- For $m = 1$, use the same procedure as for integrating $\sec x$. Multiply the integrand by $(\csc x + \cot x)/(\csc x + \cot x)$ and then using z -substitution, setting $z = \csc x + \cot x$.

For integrals of the form

$$\int \tan^m x dx,$$

- For $m = 1$, substitute $\frac{\sin x}{\cos x}$ for $\tan x$ and solve using z -substitution with $z = \cos x$.
- For $m = 2$, substitute $\frac{\sin x}{\cos x}$ for $\tan x$ and then use integration by parts with $u = \frac{1}{\cos^2 x}$ and $dv = \sin^2 x dx$. Your "simplified" integral will end up involving the third power of $\tan x$ so you'll have to apply one of the above strategies to solve the resulting integral.
- For $m = 3$, use the identity $\tan^2 x = \sec^2 x - 1$ to make a substitution and then use z -substitution with $z = \tan x$.

For integrals of the form

$$\begin{aligned} &\int \sin A \cos B dx \\ &\int \sin A \sin B dx \\ &\int \cos A \cos B dx \end{aligned}$$

use the trigonometric product identities to simplify the integrand.

1.4.5 Trigonometric Substitution

Trigonometric substitution is a kind of backwards substitution that allows us to use trig identities to simplify an integrand that doesn't even contain trig functions. This is essentially using the substitution rule in reverse.

Since this technique allows us to factor certain kinds of terms, it's very helpful in getting rid of radical signs, although the method could also be used in the absence of radical signs. Here are the three kinds of expressions and the three substitutions that are most helpful

$$\begin{array}{ll} \sqrt{a^2 - x^2} & \text{Let } x = a \sin \theta \\ \sqrt{a^2 + x^2} & \text{Let } x = a \tan \theta \\ \sqrt{x^2 - a^2} & \text{Let } x = a \sec \theta. \end{array}$$

Note that this inverse substitution can also be made with hyperbolic functions and their identities. This is sometimes easier than using the trig functions.

Example:

Evaluate

$$I = \int \sqrt{1 + e^{-y}} dy.$$

We have that

$$\begin{aligned} I &= \int \sqrt{1 + \frac{1}{e^y}} dy \\ &= \int \sqrt{\frac{e^y + 1}{e^y}} dy \\ &= \int e^{-\frac{y}{2}} \sqrt{e^y + 1} dy \\ &= 2 \int e^{-x} \sqrt{e^{2x} + 1} dx \quad \text{let } y = 2x \\ &= 2 \int \frac{\sqrt{z^2 + 1}}{z^2} dz \quad \text{let } z = e^x \end{aligned}$$

Now, we can solve it by making the substitution $z = \tan \theta$.

Another substitution that we should know is hyperbolic substitution. In this case, we could have used $z = \sinh \theta$ and then used the identity $\cosh^2 \theta - \sinh^2 \theta = 1$ to simplify it.

Example:

Evaluate

$$\int \frac{\sqrt{16-x^2}}{x^2} dx.$$

We start by making the substitution $x = 4 \sin \theta$, which means $x^2 = 16 \sin^2 \theta$ and $dx = 4 \cos \theta d\theta$. Then we use the identities $\cos^2 \theta + \sin^2 \theta = 1$ and $\cot^2 \theta = \csc^2 \theta - 1$ to simplify along the way.

$$\begin{aligned} \int \frac{\sqrt{16-x^2}}{x^2} dx &= \int \frac{\sqrt{16-16\sin^2\theta}}{16\sin^2\theta} 4\cos\theta d\theta \\ &= \int \frac{\sqrt{16\cos^2\theta}}{16\sin^2\theta} 4\cos\theta d\theta \\ &= \int \frac{\cos^2\theta}{\sin^2\theta} d\theta \\ &= \int \cot^2\theta d\theta \\ &= \int \csc^2\theta - 1 d\theta \\ &= -\cot\theta - \theta + C. \end{aligned}$$

We are not done yet. Our solution is in terms of θ , but the problem is given in terms of x , so we must back substitute.

Since $x = 4 \sin \theta$, we know that $\theta = \sin^{-1}\left(\frac{x}{4}\right)$. To find $\cot \theta$ in terms of x , we draw a right triangle with angle θ . We know that $\sin \theta$ which is opposite over hypotenuse is $\frac{x}{4}$, so we can interpret the opposite side as length x and the hypotenuse as length 4. From the Pythagorean theorem, we find that the adjacent side is then $\sqrt{16-x^2}$. Since cotangent is adjacent over opposite, we find that $\cot \theta = \frac{\sqrt{16-x^2}}{x}$. So our solution in terms of x is

$$\int \frac{\sqrt{16-x^2}}{x^2} dx = -\frac{\sqrt{16-x^2}}{x} - \sin^{-1}\left(\frac{x}{4}\right) + C.$$

1.4.6 Integrating Rational Functions

Rational functions can often be broken up into a sum of simpler functions by the method of **partial fraction decomposition** which makes it a lot easier to integrate them.¹

Tip:

Before you use partial fraction decomposition to integrate a rational function, check to see if the problem can be solved with a simple z -substitution.

After you've decomposed your rational function, you simply integrate the sum of terms individually, which should be relatively easy. Check your expanded result immediately by differentiating to make sure your answer matches the integrand.

The last term is often the trickiest to integrate. Try one of these techniques:

1. If the numerator has multiple terms, split the fraction into multiple fractions each having the same denominator.
2. If the last term is 1 over an irreducible quadratic, complete the square in the quadratic in order to convert the entire term into the form $\frac{1}{x^2+a^2}$ then use the integration formula

$$\int \frac{1}{a^2+x^2} dx = \frac{1}{a} \tan^{-1}\left(\frac{x}{a}\right) + C.$$

Completing the square will often simplify earlier terms containing irreducible quadratics in the de-

nominator because you can do a z -substitution for anything inside the brackets.²

Rationalizing Substitutions

Sometimes, nonrational functions can be converted to rational functions by first making a substitution. For example, when an integrand contains a radical, it's sometimes effective to make a z -substitution of the entire radical as in this example

$$\int \frac{\sqrt{x+4}}{x} dx = \int \frac{z}{z^2-4} 2z dz = \frac{1}{2} \int \frac{z^2}{z^2-4} dz.$$

Because $z = \sqrt{x+4}$, we have that $x = z^2 - 4$, and $dx = 2z dz$. The integral can then be evaluated by factoring the denominator and using partial fraction integration.

Any rational function of $\sin x$ and $\cos x$ can be converted into a regular rational function by using the following substitutions

$$\begin{aligned} z &= \tan\left(\frac{x}{2}\right), & \cos z &= \frac{1-z^2}{1+z^2}, \\ \sin x &= \frac{2z}{1+z^2}, & dx &= \frac{2}{1+z^2} dz. \end{aligned}$$

1.4.7 Improper Integrals

For an integral with an infinite interval, we use a limit

$$\int_a^\infty f(x) dx = \lim_{n \rightarrow \infty} \int_a^n f(x) dx.$$

¹See the section in the pre-calculus chapter on partial fraction decomposition.

²See the section in the pre-calculus chapter on parabolas for instructions on how to complete the square for any quadratic.

The same process can be used for integrals going to negative infinity. The process for solving these is to do the integration and then take the limit at infinity of the integral. If the corresponding limit exists, the improper integral is said to be convergent. If it does not exist, it is said to be divergent.

An integral like this going from negative infinity to positive infinity has to be split into two limit integrals—one going from negative infinity to some convenient number and the other going from the same convenient number to positive infinity.

A **p -integral** is an integral of the form

$$\int_1^{\infty} \frac{1}{x^p} dx.$$

The integral is convergent if $p > 1$ and divergent if $p \leq 1$.

The second type of improper integral occurs when the integrand is discontinuous (i.e. it contains a gap or a vertical asymptote). These are solved by finding the limit of the integral as it approaches the discontinuity. If the discontinuity is in the middle, then you have to approach it from both sides with a limit integral.

When doing these, it is helpful to graph the integrand to locate the asymptotes.

The **comparison theorem** is a simple theorem that

allows us to show that certain integrals are divergent or convergent without having to actually evaluate them. To understand the theorem, draw two functions—one always under the other. Now it is easy to see if the top function has finite area, then so will the bottom and that if the bottom function has infinite area then so will the top one.

To check integrals with infinite intervals, use the calculator's **fnInt** feature and replace the infinity sign with a large number like 100 or 1000. If the integrand approaches the x -axis quickly, then this should yield an answer close to the real answer. Be careful with this because rounding errors can quickly add up. For example, a certain infinite integral evaluated from $[2,100]$ yields an answer very close to the real answer. However, when evaluated from $[2,1000]$ even though the answer should be even closer, it's not even remotely close anymore.

To check the second kind of improper integral (i.e. vertical asymptote) you can also use the calculator. For example, if you're integrating from $(6,8]$ just set your limits at $[6.0001,8]$.

For all problems, make sure the integrand doesn't include a vertical asymptote. Make sure that every number being integrated across is in the domain of the integrand. If it isn't then it's an improper integral and you have to use the limit to approach the infinity.

1.5 Integration Applications

1.5.1 Area Between Curves

Remember that the area *under* a curve is a special case of the area between two curves. The one function is simply $y = 0$. To find the area between two curves $f(x)$ and $g(x)$ when $f(x)$ is always greater than $g(x)$, that is, it's always above it, then the area between the two curves is simply the integral of the first minus the integral of the second

$$\text{Area} = \int_a^b [f(x) - g(x)] dx.$$

When asked to find the areas *bounded* by two curves, you must solve for the intersections before integrating. Sometimes it's difficult or impossible to find the points of intersection algebraically. When asked to approximate the area between the curves use the trace function on your calculator or the root finding function to find the approximate intersection points.

Tip:

If you're having a hard time finding the intersection of two curves, guess values by looking at the graphs on your calculator and then plug them into the equation to see if they work. Then try to figure out how to solve it algebraically.

Note, when integrating functions using substitution, you don't have to use the substitution for everything. As long as you can break it into a sum of functions, you could integrate one part with respect to u and the other part with respect to x .

The general way to find the area between curves when one function is not always greater than the other is to split the problem into several integrals by integrating the absolute value of the difference of the functions.

$$\text{Area} = \int_a^b |f(x) - g(x)| dx.$$

To solve these, you must first find the intersection points and break the problem into separate integrals (each integrating the top function minus the bottom function for that specific interval). Watch for symmetry that would allow you to easily extrapolate by solving the integrals for part of the problem.

Remember, all of these can be approximated on the calculator, provided that you know how to construct the integral in the first place.

When asked to find the area between two curves when there's not a clear top and bottom, but there's a clear left and right function, then it's probably easier to integrate with respect to y

$$\text{Area} = \int_c^d [X_R - X_L] dy.$$

In this case, d is the top bound, c is the bottom bound, X_R is the right function, and X_L is the left function. Remember that both equations have to be solved for x . If you can't remember what goes where, just remember that this is simply the previous method rotated counterclockwise 90° (however, the bounds are switched to turn what would be a negative result into a positive result). Also, remember that d and c are y -values instead of x -values.

To remember how to find the area between vertical curves (with respect to the y -axis), think about it this way. To find the area between curves, you subtract the far (with respect to the axis of integration) curve from the near curve and integrate from smallest number to the largest number.

Following are some techniques for checking your work:

- The most important part of this is knowing how to set up your integral. After it's been setup, you can always check your answer by doing the integration on a calculator.
- When finding areas and volumes, always make sure your answers are positive numbers. If they contain symbolic numbers such as e or π , plug them into your calculator to ensure that the final answer is positive.
- When integrating with respect to y you can also sometimes integrate it with respect to x , although it may be more difficult. It's a way to double-check your work.
- Double-check your graphs and your intersections because if you get these wrong you won't have the right integral and so your error won't be revealed by doing the integral on the calculator.
- To check your answer on the calculator for an integral taken with respect to y -axis graph the inverse functions, and find the area that way. Note that you must deduce what your new upper and lower bounds are.

When given the three points of a triangle and you are asked to find the area, find the line equations for the lines connecting the points and then set up a pair of integrals to find the area between all three lines.

Tip:

Remember that area is always positive. If you get a negative value, you must have made a mistake.

1.5.2 Volume

Volumes by Slices

The volume of a cylinder (we don't care what the cylinder is shaped like as long as it has parallel bases and its cross-sectional area is constant) is simply the area of the base times its height. The area of pretty much any volume can be approximated as the volume of a bunch of cylinders of various base areas all stacked together. In other words, slice a volume like a loaf of bread and each slice is a tiny cylinder. Add up the volumes of all the cylinders and you have an approximate area of the volume. Increase the number of slices toward infinity and you have the exact volume of the shape.

Consider an arbitrary solid that lies between $x = a$ and $x = b$. If the cross-sectional area of the solid in the plane P_x , through x and perpendicular to the x -axis is $A(x)$, where $A(x)$ is a continuous function, then the volume of the solid is

$$\text{Volume} = \int_a^b A(x) dx.$$

In this case, $A(x)$ is the area of the base of a cylinder and dx is the height of the cylinder. It doesn't actually matter how the volume is situated with respect to the x -axis since the answer will be the same regardless of the orientation of our parallel slices. In the integral, $A(x)$ is the area of a moving cross-section obtained by slicing through x perpendicular to the x -axis.

If the solid is formed by rotating a region about the y -axis then it would probably be easier to integrate with respect to y . This involves the same formula but with all the x 's replaced by y 's. But you can't just substitute, of course, you must reinterpret the integral with respect to x —instead of y —taking horizontal slices centered on the y -axis instead of vertical slices centered on the x -axis.

If the area of a slice is a ring, just subtract the area of the inner circle from the area of the outer circle to obtain the area of the ring.

A solid created by revolving a region about a line is called a **solid of revolution**. It doesn't actually matter exactly where the line of revolution is as long as it is parallel either to the x -axis or the y -axis. In other words, to find the volume of a region rotated about the line $y = 3$, just pretend the region was rotated around the line $y = 0$. The same principle applies but your radii must be updated.

Just remember that the cross-section is perpendicular to the axis being integrated. If the cross-section is horizontal, then you must integrate with dy instead of dx . In general, the volume of such a solid is calculated by using the basic formula:

$$\text{Volume} = \int_a^b A(x) dx, \quad \text{or} \quad \int_c^d A(y) dy.$$

For volumes that are not solids of revolution, the process is the same. Simply find the left/right or top/bottom bounds of the shape and determine the area of each slice using geometry and trigonometry. The hardest parts are figuring out how to position the solid in the Cartesian plane and figuring out how to find the area of a slice.

When deducing the radius of the cylinder, remember that distance between two functions (a line is a function) is given by $|f(x) - g(x)|$.

To check your work, double check every detail of your integral setup because this aspect can't be checked by the calculator. Triple check your axis of rotation. Don't rotate about $y = 1$ if it's really asking you to rotate about $x = 1$.

Volumes by Cylinders

Some volumes are not easy to calculate using the previous method. Another method involves thinking of a volume as infinitely thin cylinders stacked inside of each other. This works particularly well for volumes that are symmetric about one axis.

The volume of a cylindrical shell can be calculated by subtracting the volume of the inner (empty) shell from the outer shell.

$$V = 2\pi r h \Delta r = \text{circumference} * \text{height} * \text{thickness}.$$

In this equation, r is the average radius of the cylinder, that is, the radius of the outside plus the radius of the inside all divided by 2. Δr is the thickness of the shell.

Note: $[\text{circumference}] * [\text{height}]$ is actually the area of the cylinder if it was cut open and flattened. So like volume by slices, this is also simply the integral of an area function.

By forming a summation involving cylindrical shells, we can get an integral to find the volume of a solid obtained by rotating about the y -axis the region under the curve $y = f(x)$ from a to b as

$$V = \int_a^b 2\pi x f(x) dx, \quad 0 \leq a < b.$$

We can think of this integral in terms of circumference ($2\pi r$), times height ($f(x)$), times thickness (dx).

If you're asked to find the volume obtained by rotating a region bounded by two functions around the y -axis, and one function is always above the other, simply subtract the other function from this one

$$V = \int_a^b 2\pi x [f(x) - g(x)] dx.$$

Note you have to find the bounds of the integral. In the case of the two equations, set them equal to each other and solve for x .

For solids rotated about an axis other than $y = 0$, draw the situation and figure out how the shells will be stacked. Then figure out the shell radius and height and build an integral.

For equation given in x , remember if they include roots and you solve for y in order to graph it on a calculator, you may see only half of the graph.

If the region is rotated about a vertical line other than the y -axis, about the only thing that changes is the radius. The radius is x only when the volume is centered at the y -axis.

1.5.3 Work

Newton's second law states that the force on an object is equal to its mass times its acceleration

$$F = ma.$$

If an object moves along a straight line with position $s(t)$, then force is

$$F = m \frac{d^2 s}{dt^2}.$$

If acceleration is constant, force is also constant and the work done is defined to be the product of the force and the distance that the object moves

$$W = Fd.$$

Note, the equations for force and work can be combined to form $W = mad$ which is helpful when calculating the force required to lift something against the acceleration of gravity.

Note: When given a work/gravity problem, if you're given the object's weight in pounds, know that pounds is a force, not a mass. When given pounds, you can calculate the work as $W = \text{weight} \times \text{distance}$.

The above equation defines work as long as the force is constant, but what if the force is variable? If an object moves along the x -axis in the positive direction from $x = a$ to $x = b$, and at each point a force $f(x)$ acts on the object, then this interval can be sliced and the forces summed with a Riemann sum. Therefore, the work done in moving the object from a to b is

$$W = \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k) \Delta x = \int_a^b f(x) dx.$$

In other words, WORK is the time-integral of FORCE.

Hooke's law states that the force required to maintain a spring stretched x units beyond its natural length is proportional to x

$$f(x) = kx,$$

where k is a constant specific to the spring.

Tips:

- Think each problem through. Is it asking for force or work? Force should be easy to calculate. Work from force requires integration.
- Always include units when setting up your problems and dimensional analysis to ensure that you'll end up with the right answer.
- Remember, to solve integration applications, break something into many little pieces so you can calculate an individual piece, then you add them all up.
- When working against gravity, the constant force equation becomes $F = mg$

1.5.4 Average Value of a Function

The average value of a function f on the interval $[a, b]$ is

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

If the function is positive then the average value can be thought of as its area divided by its width.

If a is the beginning of the day, b is the end of the day, and $f(x)$ is a continuous function showing the temperature with respect to time of day, then the average value of the function is the average temperature for the day.

The **mean value theorem** for integrals tells us that if f is continuous on $[a, b]$, then there exists a number c in that interval such that

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

In other words,

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

In other words, for a continuous function, there is at least one place within an interval where the value of the function equals its average value over that interval.

To find c note that $f(c) = f_{ave} = \frac{1}{b-a} \int_a^b f(x) dx$ so just set the original function equal to its average over the interval and solve for c (or x).

To check your work, do the integration on the calculator and divide that value by the length of the interval $b - a$.

1.5.5 Arc Length

To measure the length of a curve, we break it into an infinite sum of tiny lengths so that we can find it via an integral.

Arc length is given by the formulas

$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx, \quad \text{for } y = f(x)$$

$$L = \int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy, \quad \text{for } x = g(y).$$

We can also represent a function in terms of another parameter, and solve it that way (i.e. parametric equation). This may greatly simplify the problem if we're dealing with a semicircle, but it'll rarely help otherwise:

$$\int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt.$$

Tip:

Remember that arc lengths are positive so if you get a negative value, you have made a mistake.

Be very careful when doing the differentiation because you have to do that before you can plug the integral into the calculator to check your solution, so if you make an error in the differentiation your solution may check correctly in the calculator without actually being correct.

If the integral becomes very difficult or impossible to solve, then you probably made a mistake in the earlier algebra.

1.5.6 Area of a Surface

When finding the area of a **surface of revolution**, you can find it either by integrating with respect to x or by integrating with respect to y .

When the function is rotated around the x -axis, the

formula for surface area is

$$S = \int 2\pi y ds,$$

where ds is one of

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx \quad \text{or} \quad \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy.$$

Putting it all together, if a function $y = f(x)$ from (x_1, y_1) to (x_2, y_2) is rotated about the x -axis, the overall formula is one of the following

$$S = \int_{x_1}^{x_2} 2\pi f(x) \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx$$

$$= \int_{y_1}^{y_2} 2\pi y \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy.$$

For the surface area of a segment of a function $x = g(y)$ rotated about the y -axis, the surface area is given by one of the following

$$S = \int_{x_1}^{x_2} 2\pi x \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx$$

$$= \int_{y_1}^{y_2} 2\pi g(y) \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy.$$

Be extra careful when finding the surface area of a solid rotated about the y -axis. If the equation is given as $y = \text{something}$, you must solve for $x = \text{something}$. If the limits of integration are given in terms of y , you're good, but if they're given in terms of x , then you must solve for them in terms of y . When doing these, it's very helpful to draw a graph to keep track of what's going on.

Notice that the formula for arc length is just

$$\int ds,$$

and the formula for surface area is

$$\int 2\pi y ds, \quad \text{or} \quad \int 2\pi x ds,$$

depending on the axis of rotation.

1.6 Sequences and Series

1.6.1 Sequences

A **sequence** is an ordered set of numbers that can be thought of as a function with a domain of only the natural numbers.

$$\{a_n\} = \{a_1, a_2, a_3, \dots\}.$$

Some functions can be represented with a formula for the n th term, but others (such as the Fibonacci sequence) can only be defined recursively—by referencing previous terms.

Sequences can **converge** (limit to a specific number) or they can **diverge** to positive or negative infinity, or-

they can do neither (such as oscillating sequences). A **convergent sequence** can be visualized on a number line (the dots get more crowded as you approach the limit, but the dots never occur after the limit) or on a Cartesian plane—like regular functions that converge asymptotically. Remember however, on the Cartesian plane, a sequence is a bunch of dots rather than a line.

To test whether or not a sequence converges, we can use limits on the defining formula for the sequence since the dots of the sequence all lie on the line of the real function. In other words, if a sequence has a generating function $f(n) = a_n$, then we can determine whether or not it converges to some number by taking the limit of $f(x)$.

For example, the sequence $a_n = \frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \dots$ has the associated real function $f(x) = \frac{1}{x^2}$. Since

$$\lim_{x \rightarrow \infty} \frac{1}{x^2} = 0,$$

we say that the sequence $a_n = \frac{1}{n^2}$ converges to zero.

When using limits to check whether or not a sequence converges, we can use the same limit laws that we learned earlier. We can also use the squeeze theorem.

Theorem:

$$\text{If } \lim_{n \rightarrow \infty} |a_n| = 0, \text{ then } \lim_{n \rightarrow \infty} (a_n) = 0.$$

Example:

Determine whether the sequence $a_n = \frac{(-1)^n}{n^2}$ converges, and if so, to what value.

Notice that the numerator of the terms oscillates between 1 and -1 . By taking the absolute value of a_n , we see that

$$\left| \frac{(-1)^n}{n^2} \right| = \frac{|(-1)^n|}{n^2} = \frac{1}{n^2},$$

so by the theorem given above, since

$$\lim_{n \rightarrow \infty} \left| \frac{(-1)^n}{n^2} \right| = \lim_{n \rightarrow \infty} \frac{1}{n^2} = 0,$$

we know that $a_n = \frac{(-1)^n}{n^2}$ converges to 0.

Tip:

Sequences can be sort of graphed on the calculator using parametric equations. Set the mode to **PAR** and **DOT**, then enter the a_n in terms of T at Y_1T and enter simple “T” at X_1T . Then from **WINDOW**, set **TSETP** to 1 and set **TMAX** to a large number like 100.

A sequence that goes up along its entire domain is an **increasing sequence**. If it decreases over its entire

domain, it’s a **decreasing sequence**. A **monotonic sequence** is one that is either a decreasing or an increasing sequence.

A sequence is bounded above if there is a number that the sequence will never surpass. A sequence can also be bounded below. In general, a sequence is **bounded** if it has both upper and lower bounds.

The **monotonic sequence theorem** states that every bounded, monotonic sequence is convergent.

The above theorem makes sense if you think about it, but we can actually expand on the definition. An increasing sequence that has an upper bound is convergent as is a decreasing sequence that is bounded below.

Tip:

You can often plot the formula for the sequence and see if the graph appears to converge or diverge. You can also evaluate the formula at several progressively larger numbers to see if it appears to be converging.

There’s no simple calculator test for checking your answers. The best approach is to develop an intuitive understanding of sequences and then see if your intuition matches your calculated results. If not, then you must investigate and analyze to determine whether it’s your intuition or your calculations that went wrong and why.

1.6.2 Mathematical Induction

Mathematical induction is used to prove a hypothesis about all natural numbers. Since the set of natural numbers is sequential, it’s like a row of infinite dominoes—if you can prove that if a hypothesis is true for a number then it is true for the number after it, and the hypothesis is true for at least one number, then it must be true for every number—you’ve made the whole row of dominoes fall.

In other words, to prove by mathematical induction that some rule is true for all natural numbers, we have to show that *if* it is true for some arbitrary number n , then it is necessarily true for the next number $n + 1$. Then if the rule works for a specific number such as $n = 1$, then by our earlier implication, we know that it works for $n + 1$, which means it works for $n + 2$, which means it works for $n + 3$, and so on. It follows that it works for all natural numbers.

Example:

Given that $\frac{d}{dx}x^2 = 2x$, $\frac{dx}{dx} = 1$, and the product rule, prove by mathematical induction that

$$\frac{d}{dx}x^n = nx^{n-1}.$$

If

$$\frac{d}{dx}x^n = nx^{n-1},$$

then by the properties of exponents, we have that

$$\frac{d}{dx}x^{n+1} = \frac{d}{dx}x^n x.$$

By the product rule, we can expand the right side as

$$\frac{d}{dx}x^{n+1} = x^n \frac{dx}{dx} + x \frac{d}{dx}x^n.$$

Since $\frac{dx}{dx} = 1$, and from the assumption that $\frac{d}{dx}x^n = nx^{n-1}$, we get

$$\begin{aligned} \frac{d}{dx}x^{n+1} &= x^n(1) + x(nx^{n-1}) \\ &= (n+1)x^n. \end{aligned}$$

Notice that our final result is the same as the original rule but with n replaced by $n+1$.

We have shown that *if* the rule holds for n , then it necessarily holds for $n+1$. Since we know that it holds for a particular n , specifically, $n=2$, it follows that it holds for all n , that is, for all natural numbers.

1.6.3 Series

An **infinite series** is an infinite sum of sequence terms denoted

$$\sum_{n=1}^{\infty} a_n.$$

A **partial sum** s_n is the sum of the first n terms of a series.

$$s_n = a_1 + a_2 + a_3 + \cdots + a_n.$$

The limit of the partial sum s_n as n approaches infinity is the sum of the infinite series if that limit exists. If the limit does not exist, then the series has no finite sum. The way to check is to write out the first several partial sums and see if you can figure out a formula for the n th partial sum. Then take the limit of that formula at infinity, which gives the sum of the original series. In general, it is not easy to find the formula for the n th partial sum.

If a series has a finite sum (i.e. the limit of the partial sum formula exists), the series is said to be **convergent**. Otherwise, the series is **divergent**.

A **geometric series** is a special series of the form

$$a_n = a + ar + ar^2 + ar^3 + \cdots,$$

where r is the **common ratio**. Any term in the series is the previous term times r . For a geometric series, it is easy to find the partial sum. Consider the partial sum

$$s_n = a + ar + ar^2 + \cdots + ar^{n-1}.$$

Multiplying both sides by r gives us

$$rs_n = ar + ar^2 + ar^3 + \cdots + ar^n.$$

Subtracting the second partial sum from first results in most of the terms canceling out, and we get

$$\begin{aligned} s_n - rs_n &= a - ar^n \\ s_n(1-r) &= a - ar^n \\ s_n &= \frac{a - ar^n}{1-r} \\ &= \frac{a}{1-r} - \frac{ar^n}{1-r}. \end{aligned}$$

Notice that if $-1 < r < 1$, then the limit of s_n as n goes to infinite is $s_n = \frac{a}{1-r}$ since the second term on the right cancels out, which means if $-1 < r < 1$, the geometric series a_n is convergent. The limit of the partial sum is $\frac{a}{1-r}$, which is also the sum of the series. Thus, if a geometric series is convergent, then its sum is

$$\frac{\text{first term}}{1 - \text{common ratio}}.$$

There are several ways of finding the exact sum of an infinite series.

1. Finding a formula for the n th partial sums and then taking the limit of this equation as per the definition of the sum of a series. The most difficult aspect of this is finding the equation for the partial sums.
2. If it's a geometric series, then the sum is the first term divided by 1 minus the common ratio.
3. You may be able to group or reorder the terms so that all but a few at the beginning and end cancel. This happens with **telescoping series** and may be accomplished by using partial fractions to decompose a rational function into two functions where one is negative. Sometimes (i.e. when there's a single term in the denominator), you can decompose a rational function into a positive and negative function without having to use partial fraction decomposition.

Checking your work is easy here—just plug the sum into your calculator, substituting progressively larger numbers for infinity.

If a series is convergent, then the limit of the formula for the series is 0. However, the converse is not true—if the limit of the formula for the series is 0, the series may or may not be convergent. A better way to put it is, if the limit of the formula for a series is not 0, then the series is divergent. This is called the **test for divergence**.

A constant times a convergent series yields a convergent series. The sum and the difference of two convergent series are both convergent series.

Keep in mind the notation differences between series and sequences. If enclosed by $\{\}$, then it is a sequence rather than a series, which is the sum of a sequence.

Integral Test

If $a_n = f(n)$ is continuous, positive, and ultimately decreasing, that is it is decreasing on $[k, \infty)$, then

$$\text{If } \int_k^{\infty} f(x) dx \text{ converges, then so does } \sum_{n=k}^{\infty} a_n.$$

Conversely,

$$\text{If } \int_k^{\infty} f(x) dx \text{ diverges, then so does } \sum_{n=k}^{\infty} a_n.$$

The condition that a function is decreasing to infinity is sometimes hard to prove. Use the derivative test—if the derivative of a function is negative over the domain in question, then the function is decreasing over the domain in question.

It doesn't matter if a series is initially increasing, provided that it is ultimately decreasing. If the sum $\sum_{n=k}^{\infty} a_n$, where k is a finite number, is convergent, then so is $\sum_{n=1}^{\infty} a_n$. In other words, a finite sum added to the beginning of a series does not affect its convergence or divergence.

Note: The value of the integral is not equal to the sum of the series. The integral test merely tells us whether or not the series fitting the given conditions is convergent or not.

Note: Be sure the function is continuous, positive, and ultimately decreasing before using the integral test.

Tip:

Make full use of the summation function under the MATH menu of the TI-84. You can approximate infinite series by replacing the infinity symbol with some large number in the upper limit.

A series of the form

$$\sum_{n=1}^{\infty} \frac{1}{n^p},$$

where p is a number, is called a **p-series**. A p-series is convergent if $p > 1$ and divergent if $p \leq 1$.

Comparison Tests

The comparison tests are intuitively straightforward. If you have a series in which all the terms are smaller than the equivalent terms in a convergent series, then your series must be convergent (i.e. its graph lies entirely under the graph of a convergent series). If you have a series in which all the terms are larger than the equivalent terms in a divergent series, then your series must be divergent (i.e. the graph lies entirely above the graph of a divergent series).

Beware of erroneously comparing a smaller series to a known divergent series or a larger series to a known convergent series.

The most common series to compare against are the p-series (if p is less than or equal to 1 the series diverges, and otherwise it converges) and geometric series (if r is inclusively between -1 and 1 , the series converges, otherwise it diverges).

This still works if you have to exclude a finite number of terms at the beginning of your series since a finite number of terms doesn't change the convergence or divergence of the series.

When doing this comparison, the terms of both series must be positive, so be sure to indicate this as in $0 < a_n \leq b_n$.

For two series with positive terms $\sum a_n$ and $\sum b_n$, if

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = c,$$

where $0 < c < \infty$, then the two series behave the same, that is, they either both converge or they both diverge. This is the **limit comparison test**.

In testing many series, we find a useful comparison series by keeping only the largest powers in the numerator and denominator of our series to be tested.

Alternating Series

An alternating series has the form

$$\sum_{n=1}^{\infty} (-1)^{n+1} b_n, \quad \text{or} \quad \sum_{n=1}^{\infty} (-1)^n b_n.$$

An alternating series converges if both of the following conditions are true:

1. $b_{n+1} \leq b_n$. In other words, if the absolute values of the terms are demonstrably decreasing either by comparing the formulas for b_{n+1} and b_n or by taking the derivative to prove that the absolute values of the terms are decreasing. Note: To double-check, graph the b function and make sure it appears to be decreasing.
2. $\lim_{n \rightarrow \infty} b_n = 0$. Note: This excludes the portion of the formula that produces the \pm signs.

If the alternating series test is inconclusive, you may be able to use the test for divergence.

In this example,

$$\lim_{n \rightarrow \infty} \frac{(-1)^n}{n^2} = 0,$$

we have a definite limit, 0, because although the numerator oscillates between -1 and 1 , the denominator grows to infinity, so the oscillations will grow smaller and smaller.

Here

$$\lim_{n \rightarrow \infty} \frac{n(-1)^n}{n} = \text{DNE},$$

the limit does not exist since the numerator and denominator have the same degree, so the oscillations are not growing smaller and smaller.

Ratio and Root Test

A series $\sum a_n$ is called **absolutely convergent** if the series of the absolute values, $\sum |a_n|$ is convergent.

If a series, is absolutely convergent, then it is convergent.

If a series is convergent but not absolutely convergent, then it is **conditionally convergent**. An example of this case is the alternating harmonic series which is convergent whereas the absolute value series (i.e. the harmonic series) is not.

The idea of absolute convergence is helpful when dealing with series in which the signs of the terms switch back and forth irregularly.

The **ratio test** involves taking the limit

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L.$$

The **root test** involves taking the limit of a root

$$\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = L.$$

In both of these cases,

- If $L < 1$, then the series is convergent
- If $L > 1$, then the series is divergent (infinity is considered greater than 1)
- If $L = 1$, then the test is inconclusive.

Note: If any one of these tests gives the failure $L = 1$, then so will the other one.

Any rearrangement of an absolutely convergent series will yield the same result. A rearrangement of a conditionally convergent series will not necessarily yield the same result.

Summary of Series Testing

1. **Geometric Series:** A geometric series will converge if the common ratio is strictly between -1 and 1 .
2. **Test for divergence:** If the limit of the formula does not equal zero, then the series is divergent. If you graph a function and it appears to approach a nonzero limit, try this test.
3. **P-series:** A p-series is convergent if p is strictly greater than 1 .
4. **Integral Test:** If the series is continuous, positive, and ultimately decreasing, then the integral of the function behaves the same as the series.
5. **Comparison Tests:** If you can show that a series is smaller than a convergent series, then your series is convergent. If you can show that a series is larger than a divergent series, then your series is divergent.
6. **Limit Comparison Test:** If the limit of the ratio of the formulas for two positive series is a number strictly between 0 and infinity, then both series behave the same.
7. **Alternating Series Test:** An alternating series with a formula of the form $a_n = (-1)^n b_n$ is convergent if $b_{n+1} \leq b_n$ for all n and the limit of b_n as n approaches infinity is 0 . In other words, if the absolute value of the terms are getting demonstrably smaller and they're approaching zero, then the alternating series is convergent.
8. **Absolute Convergence:** If a series is absolutely convergent then it is convergent. In other words, if a series is convergent even if you replace all minus signs with plus signs, then it must be convergent.
9. **Ratio Test:** $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = L$
 - If $L < 1$, then the series is convergent
 - If $L > 1$, then the series is divergent (infinity is considered greater than 1)
 - If $L = 1$, then the test is inconclusive.
10. **Root Test:** $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = L$
 - If $L < 1$, then the series is convergent
 - If $L > 1$, then the series is divergent (infinity is considered greater than 1)
 - If $L = 1$, then the test is inconclusive.

1.6.4 Power Series

A power series is a series of the form

$$\sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots = f(x).$$

where x is a variable and the c 's are coefficients. For each fixed x , the series is a series of constants that we can test for convergence or divergence. The domain of the function $f(x)$ is the set of all x for which the series converges. The only difference between $f(x)$ and a polynomial is the fact that this function has infinite terms. Notice that the index n starts at 0 .

More generally, a power series centered at a (instead of 0) has the form

$$\begin{aligned}\sum_{n=0}^{\infty} c_n(x-a)^n &= c_0 + c_1(x-a) + c_2(x-a)^2 + \dots \\ &= f(x-a).\end{aligned}$$

By testing the series, you can figure out for which values of x the series converges. There are only 3 possibilities:

1. The series converges only when $x = a$
2. The series converges for all x
3. There is a positive number R , called the **radius of convergence**, such that the series converges if $|x - a| < R$ and diverges if $|x - a| > R$

The **interval of convergence** is the finite or infinite interval on the number line for which a power series converges.

The power series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots = \sum_{n=0}^{\infty} x^n,$$

is an important one and needs to be remembered. We can use it to find the power series for similar functions. For example, to find the power series for $\frac{1}{1+x^2}$, we substitute $-x^2$ for the x to get

$$\begin{aligned}\frac{1}{1+x^2} &= \frac{1}{1-(-x^2)} \\ &= \sum_{n=0}^{\infty} (-x^2)^n \\ &= \sum_{n=0}^{\infty} (-1)^n x^{2n}.\end{aligned}$$

Similarly, to get the power series for $\frac{1}{5+x}$, we can factor 5 out of the denominator, then substitute $-\frac{x}{5}$ for x

$$\begin{aligned}\frac{1}{5+x} &= \frac{1}{5\left(1-\left(-\frac{x}{5}\right)\right)} \\ &= \frac{1}{5} \sum_{n=0}^{\infty} \left(-\frac{x}{5}\right)^n \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{5^{n+1}} x^n.\end{aligned}$$

Infinite series can also be integrated and differentiated in what is known as **term-by-term integration** or **term-by-term differentiation**. For example, term by term differentiation of the first power series gives us

$$\begin{aligned}\frac{d}{dx} \left(\frac{1}{1-x} \right) &= \frac{d}{dx} 1 + \frac{d}{dx} x + \frac{d}{dx} x^2 + \frac{d}{dx} x^3 + \dots \\ &= 0 + 1 + 2x + 3x^2 + \dots,\end{aligned}$$

and similarly, term by term integration of the original series gives us

$$\int \frac{1}{1-x} dx = C + x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{4}x^4 + \dots.$$

1.6.5 Taylor and Maclaurin Series

To find the power series for some random function $f(x)$ we start by realizing that it will have the form

$$f(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \dots$$

The a will be a given constant, but the c_k 's are the coefficients of the terms.

In order to express the function as a power series, we must figure out what the coefficients will be. This is actually fairly easy. To find the first one c_0 , we simply let $x = a$, then every term except the first one will zero out and we find that $c_0 = f(a)$. To find the next coefficient we first differentiate the function, so that setting $x = a$ will once again zero everything but the new first term and we find that $c_1 = f'(a)$. We continue this process and we find that $c_n = \frac{f^{(n)}(a)}{n!}$.

Therefore, if a function has a power series representation at a , called a **Taylor series**, then it will have the form

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n.$$

The Taylor series for $a = 0$ is called the **Maclaurin Series**

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n.$$

The above shows that if a function can be represented as a power series about a point a then it is equal to the sum of its Taylor series. However, there exist functions that are not equal to the sum of their Taylor series.

Typically, you'll be given a function of x and asked to find its

1. **Taylor or Maclaurin series expansion:** Use the sum definition of Taylor or Maclaurin series to construct the power series. Identify the summation equation and the first several terms of the expansion. Another way to find the power series expansion is to construct it from known series by multiplying or dividing.
2. **Radius of convergence:** Use the ratio test to find out for which values of x that the series will be less than 1 and thus convergent. Since the ratio test is inconclusive at 1, you must test the both endpoints separately. The radius of convergence is equal to half of the interval of convergence. If there is only a single value of x for which the series is convergent, then the radius is 0.

3. Interval of convergence: If this is a Maclaurin series, then it is centered at 0 and the interval of convergence is simply $(-\text{radius}, \text{radius})$. If it is a Taylor series, then you must figure out where it is centered at by referring to the summation formula. For example, if the formula contains $(x - 3)^n$, then it's centered at $x = 3$.

Once we know the Taylor series for a function, we can easily find the infinite sum expression for a specific

value of the function. For example, knowing the Taylor series for e^x , to find the infinite sum expression for e , we simply substitute 1 for all x 's in the Taylor series.

To multiply two power series together, you do the same thing as with multiplying polynomials except you do only the first couple terms of two infinite sums. Use the same idea as FOILING and then collect the like terms. To divide two power series, use long division of two infinite series.

1.7 Vectors

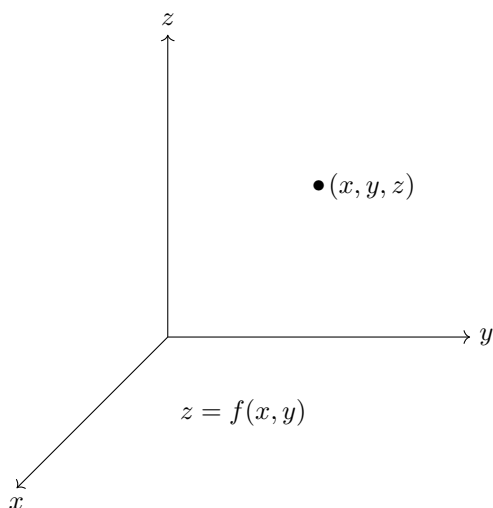
1.7.1 Three-Dimensional Coordinate System

In a two-dimensional coordinate system, a point is represented by two coordinates as in $P = (x, y)$. A two-dimensional coordinate system is the set of all coordinates (x, y) such that x and y are in \mathbb{R} . We denote this as $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$.

In a three-dimensional coordinate system, a point is represented by three coordinates as in $P = (x, y, z)$. A three-dimensional coordinate system is the set of all coordinates (x, y, z) such that x , y , and z are in \mathbb{R} . We denote this as $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$.

The concepts of limits, continuity, derivatives, and integrals are all carried over, but, they are a little more complicated. Consider the limit for example. In single variable calculus, for a function to have a limit at some point it must approach that point from both directions. The limit notion is quite different in multivariable calculus because in 3D, a function can approach a value from an infinite number of directions.

In calculus of two dependent variables x and y , that is, $z = f(x, y)$, functions can be represented as surfaces in a 3-D coordinate system. For more dependent variables, the functions can't even be visualized.



The 2-D coordinate system contains four quadrants.

In contrast, the 3-D coordinate system contains 8 octants. A quadrant has area and an octant has volume. In \mathbb{R}^2 , $y = 4$ is a line. In \mathbb{R}^3 , $y = 4$ is a plane.

To calculate the Euclidean distance in \mathbb{R}^2 between two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, we use the formula $d(P_1P_2) = |P_1P_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. To calculate the Euclidean distance in \mathbb{R}^3 between two points $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$, we use the formula

$$d(P_1P_2) = |P_1P_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

Similarly, a sphere in \mathbb{R}^3 centered at (h, k, l) , has the equation

$$(x - h)^2 + (y - k)^2 + (z - l)^2 = r^2.$$

Recall that when given the equation of a circle in standard form, you have to complete the square. When given the equation of a sphere in standard form, you use the same process.

The general equation for a conic section is

$$Ax^2 + Bxy + Cy^2 + Ex + Fy + G = 0.$$

If the coefficients A , and C , are not 1, the result will be an ellipse. If the Bxy term is present, the result will be a counterclockwise rotation of the shape.

The midpoint of a line segment in a 3D coordinate system is given by

$$\left(\frac{x_1 + x_2}{2} + \frac{y_1 + y_2}{2} + \frac{z_1 + z_2}{2} \right)$$

We rationalize rational functions containing a radical in the denominator by multiplying the top and bottom of the function by the conjugate of the denominator. For example

$$\frac{1}{\sqrt{x} - 2} \cdot \frac{\sqrt{x} + 2}{\sqrt{x} + 2} = \frac{\sqrt{x} + 2}{x - 4}.$$

What about functions containing higher degree radicals such as $\sqrt[3]{x}$? The principle of conjugation is the same,

but it gets more complicated for higher degrees. To conjugate ones containing cube roots, recall the formula for the sum and difference of cubes

$$a^3 - b^3 = (a - b)(a^2 + ab + b^3),$$

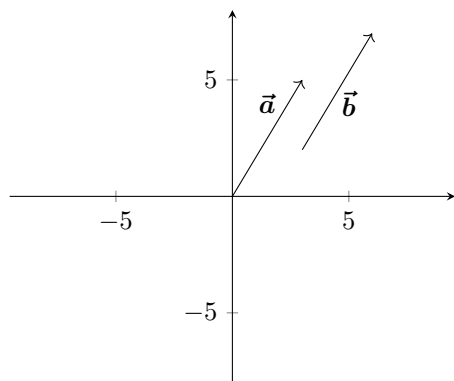
1.7.2 Vectors

In \mathbb{R}^2 , a vector is a directed line segment, unlike a ray, which extends indefinitely in one direction. A vector consists of a line with an initial point and a terminal point (where the arrow is). A vector quantity is any quantity which includes a magnitude and a direction. Examples of real-world vector quantities include

- Displacement
- Position
- Velocity
- Acceleration

The magnitude of a vector is its length. The **standard form** of a vector is when the vector's initial point is at the origin. In the example in the graph, \vec{a} is the standard form of \vec{b} .

Two vectors are equal if they have the same magnitude and direction.



We can use a vector to denote the position of a point P . The nice thing about vectors is that we can manipulate them algebraically without regard to the underlying coordinate system.

There are several ways to represent a vector in two dimensions. Here are some examples:

$$\begin{aligned}\vec{u} &= a\hat{i} + b\hat{j} \\ \vec{u} &= (a, b) \\ \vec{u} &= \langle a, b \rangle\end{aligned}$$

The first example is called **component form**. In three dimensions, the same method is used, the vectors just have an added c constant for the length in the z direction. Here, we primarily represent a vector by giving the coordinates of its terminal point in standard form bracketed by angle brackets, as in $\vec{u} = \langle a, b \rangle$.

$$a^3 + b^3 = (a + b)(a^2 - ab + b^2).$$

So,

$$\frac{1}{\sqrt[3]{x-1}} = \frac{1}{\sqrt[3]{x-1}} \cdot \frac{x^{\frac{2}{3}} + x^{\frac{1}{3}} + 1}{x^{\frac{2}{3}} + x^{\frac{1}{3}} + 1} = \frac{x^{\frac{2}{3}} + x^{\frac{1}{3}} + 1}{x-1}.$$

The magnitude (i.e. length) of a vector $\vec{u} = \langle a, b \rangle$ is calculated as

$$|\vec{u}| = \sqrt{a^2 + b^2}.$$

Similarly, the magnitude of a vector $\vec{u} = \langle a, b, c \rangle$ in \mathbb{R}^3 is calculated as

$$|\vec{u}| = \sqrt{a^2 + b^2 + c^2}.$$

This is simply an application of the Pythagorean theorem to calculate the length of the hypotenuse formed by the triangle in which the hypotenuse is formed by the vector in its standard form. Notice that the magnitude of a vector is always positive.

Properties of Real Numbers For scalars, there are two binary operations—addition and multiplication. Subtraction and division are just inverses and other operations such as exponentiation are built from addition/multiplication.

Here are the properties of the real numbers:

1. $a + b \in \mathbb{R}$, $ab \in \mathbb{R}$, (closure property)
1. $a + b = b + a$, $ab = ba$, (commutativity)
2. $a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$, (associativity)
3. $a + 0 = a$, $1a = a$, (additive/multiplicative identity)
4. $a + (-a) = 0$, $a \cdot \frac{1}{a} = 1$, (additive/multiplicative inverse)
5. $a(b + c) = ab + ac$, (distributivity)

In mathematics, the concepts of groups, fields, and rings are separated by which of the above properties that hold true in them. The above properties hold true for addition and multiplication of real numbers, but they don't necessarily hold true for other operations or other sets of elements. For example, for matrix elements, the multiplication operation is not commutative.

Vector Addition

Vector addition observes the properties

1. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
2. $\vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$
3. $\vec{u} + \vec{0} = \vec{u}$
4. $\vec{u} + (-\vec{u}) = \vec{0}$

Vectors can be added graphically by putting the tail of one vector at the head of the first vector. When moving a vector to do this, you have to make sure you maintain its original direction. The **resultant vector** or the sum

of the two vectors, then, is the vector from the tail of the first to the head of the second. It doesn't matter which vector is the one that is moved—no matter which one is moved first, the resultant vector is the same, so vector addition is commutative. The same principle applies to subtraction except that you add a negative vector, which has the same direction but points in the opposite direction. The graphical method for adding vectors is not very accurate, but its useful for quick results in the field.

Tip:

When adding vectors using the graphical method make sure that you do head-to-tail addition. It makes a difference. You'll get the wrong answer if two heads or two tails touch.

To add two vectors algebraically, you simply add the corresponding components. For example, for vectors $\vec{u} = \langle u_1, u_2 \rangle$ and $\vec{v} = \langle v_1, v_2 \rangle$,

$$\vec{u} + \vec{v} = \langle u_1 + v_1, u_2 + v_2 \rangle.$$

Note, the coordinates within the angle brackets are the Cartesian coordinates of the terminal point of the vector in standard form. We call them the **components** of the vector. Before just adding coordinates, make sure your vectors are in standard form.

Example:

Add the vectors $\vec{A} = \langle 2, 4, -1 \rangle$, $\vec{B} = \langle 2, 2, 1 \rangle$, and $\vec{C} = \langle 8, -4, 3 \rangle$.

$$\begin{aligned} \vec{A} + \vec{B} + \vec{C} &= \langle 2 + 2 + 8, 4 + 2 - 4, -1 + 1 + 3 \rangle \\ &= \langle 12, 2, 3 \rangle. \end{aligned}$$

In physics, if each force acting on some object is represented by a vector, then the net force on the object is just the resultant vector. That is, the net force vector is the sum of the individual force vectors.

To add vectors, they must have the same dimensions. You can't add a 2D vector to a 3D vector.

What about the additive identity? Is there a zero vector? Yes, it is defined simply as the origin $\vec{0} = \langle 0, 0 \rangle$. Any vector plus the zero vector equals itself, so the additive identity holds for vectors.

The additive inverse also holds for vectors. A negative vector $-\vec{w}$ is defined as a vector with the same magnitude and opposite direction as the original vector \vec{w} .

Polar Form

Converting between polar and rectangular coordinates is easy if you construct a right triangle with one of the acute

vertexes at the origin. The other acute vertex has coordinates (x, y) and (r, θ) depending on which coordinate system we're using. Label every side on the triangle and it becomes easy to convert between rectangular and polar coordinates using the transformation equations

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \\ r &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1} \left(\frac{y}{x} \right) = \cos^{-1} \left(\frac{x}{r} \right). \end{aligned}$$

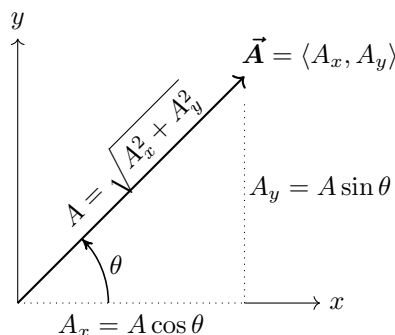
where r is the distance to the point and θ is the counterclockwise angle it makes with the positive x -axis.

The most important thing to remember is that a vector \vec{A} can be thought of as the sum of two components. The x -component of \vec{A} is denoted by A_x , and the y -component is denoted by A_y . The magnitude, or length of the vector, is denoted simply A . If we draw the vector \vec{A} in standard position (with its tail at the origin), then we can convert between Cartesian and polar components using the following equations. The hypotenuse of the triangle is now the vector $\langle A_x, A_y \rangle$.

$$\begin{aligned} A_x &= A \cos \theta \\ A_y &= A \sin \theta \\ A &= \sqrt{A_x^2 + A_y^2} \\ \theta &= \tan^{-1} \left(\frac{A_y}{A_x} \right) = \cos^{-1} \left(\frac{A_x}{A} \right). \end{aligned}$$

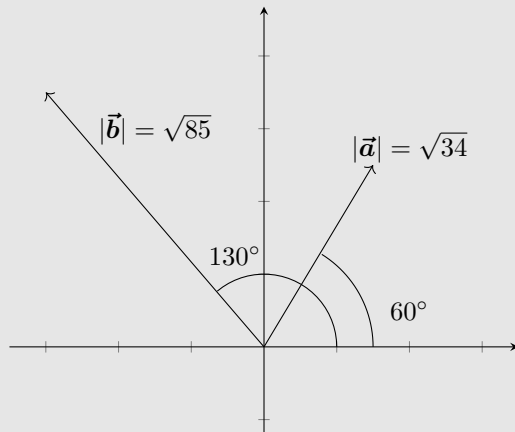
These equations are only valid when θ is measured counterclockwise from the positive x -axis. Just remember that the arctangent function gives you the reference angle and you must interpret it to determine the absolute angle. Make a habit of always drawing a right triangle to deduce the rectangular components.

Note that the signs of the components depend on the quadrant in which theta is. Draw it if unsure.



Example:

Add the two vectors shown in the graph, and give the resultant vector in polar form.



Using basic trigonometry, we can break the two vectors into components.

$$\begin{aligned} a_x &= \sqrt{34} \cos 60^\circ = 2.92 \\ a_y &= \sqrt{34} \sin 60^\circ = 5.05 \\ b_x &= \sqrt{85} \cos 130^\circ = -5.93 \\ b_y &= \sqrt{85} \sin 130^\circ = 7.06. \end{aligned}$$

So the two vectors can be written as

$$\begin{aligned} \vec{a} &= \langle 2.92, 5.05 \rangle \\ \vec{b} &= \langle -5.93, 7.06 \rangle. \end{aligned}$$

Adding the components, we find that

$$\vec{a} + \vec{b} = \langle -3.01, 12.11 \rangle.$$

Then for polar form, we need the magnitude and the angle it makes with the positive x -axis, which can be found as

$$\begin{aligned} |\vec{a} + \vec{b}| &= \sqrt{(-3.01)^2 + (12.11)^2} = 12.5 \\ \tan^{-1} \left(\frac{12.11}{-3.01} \right) &= -76^\circ \rightarrow \theta = 104^\circ. \end{aligned}$$

Scalar Multiplication

There are three different kinds of vector multiplication

- Scalar multiplication: $a\vec{u}$
- Dot product: $\vec{u} \cdot \vec{v}$
- Cross product: $\vec{u} \times \vec{v}$

Scalar multiplication of vectors observes the following properties:

1. $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$

2. $(a + b)\vec{u} = a\vec{u} + b\vec{u}$

3. $(ab)\vec{u} = a(b\vec{u})$

4. $1\vec{u} = \vec{u}$

Scalar multiplication occurs when a vector is multiplied by a real number scalar. To multiply a vector by a scalar, simply multiply each component by the scalar. The result is a “scaling” of the vector. For example, $2\vec{u}$ is twice as long as \vec{u} and $\frac{1}{2}\vec{u}$ is half as long. The direction of the vector is not affected by scalar multiplication except when the scalar is a negative in which case the direction of the vector is reversed.

A unit vector, denoted with a hat instead of an arrow, like \hat{i} vector with magnitude 1. Vectors can be represented as a sum of unit vectors—breaking them up into components. This is called **unit vector notation** or **component form**. For example, the vector $\vec{u} = \langle a, b, c \rangle$ can be represented in unit vector notation by

$$\vec{u} = a\hat{i} + b\hat{j} + c\hat{k},$$

where

$$\hat{i} = \langle 1, 0, 0 \rangle$$

$$\hat{j} = \langle 0, 1, 0 \rangle$$

$$\hat{k} = \langle 0, 0, 1 \rangle$$

The unit vectors \hat{i} , \hat{j} , and \hat{k} are called the **standard basis**.

Unit vectors have a magnitude of 1. No matter what direction a vector is pointing, if it has a magnitude of 1, it is called a **unit vector**. To test whether a vector is a unit vector, simply calculate its magnitude and see if it's equal to one.

To find the unit vector in the direction of \vec{u} , simply divide \vec{u} by its magnitude $|\vec{u}|$.

Note that for a vector in standard position, the end point is $(|\vec{u}| \cos \theta, |\vec{u}| \sin \theta)$, where θ is the angle between the vector and the positive x -axis. This means that

$$\begin{aligned} \vec{u} &= \langle |\vec{u}| \cos \theta, |\vec{u}| \sin \theta \rangle \\ &= |\vec{u}| \langle \cos \theta, \sin \theta \rangle, \end{aligned}$$

and so,

$$u_1 = |\vec{u}| \cos \theta,$$

$$u_2 = |\vec{u}| \sin \theta.$$

This concept can be extended to \mathbb{R}^3 where

$$\vec{w} = \langle w_1, w_2, w_3 \rangle = |\vec{w}| \langle \cos \alpha, \cos \beta, \cos \gamma \rangle,$$

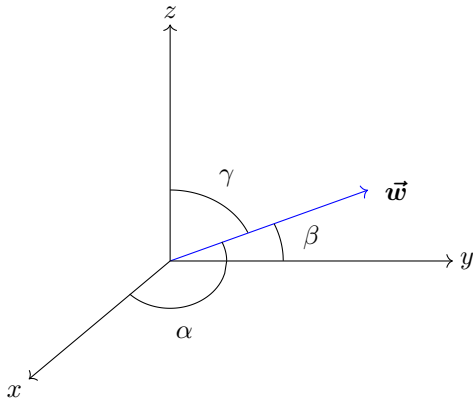
where α is the angle between the vector and the positive x -axis, β is the angle between the vector and the positive y -axis, and γ is the angle between the vector and the positive z -axis.

$$\cos \alpha = \frac{w_1}{|\vec{w}|}$$

$$\cos \beta = \frac{w_2}{|\vec{w}|}$$

$$\cos \gamma = \frac{w_3}{|\vec{w}|}.$$

The angles α , β , and γ all go from 0° to 180° .



It is not an accident that the components of an \mathbb{R}^3 vector can be given in terms of the cosines of these angles. It is possible because in \mathbb{R}^3

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1,$$

although we apparently don't have the tools yet to prove this. Likewise, in \mathbb{R}^2 ,

$$\cos^2 \theta + \cos^2 \psi = 1,$$

where ψ is the angle between the vector and the positive y -axis. This occurs because θ and ψ are complimentary angles.

When expressing a vector in 3D space, we can also project the vector onto the xy -plane, then it can be specified by a distance from the origin and two angles—the angle θ in the xy -plane measured counterclockwise from the positive x axis and the angle γ measured from the positive z -axis. This process of projection is often used by engineers who consider the floor or the ground to be the xy -plane.

To calculate the components of a 3D vector when given the angles γ and θ , we can use the equations

$$\begin{aligned} A_z &= A \cos \gamma \\ A_y &= A \sin \gamma \sin \theta \\ A_x &= A \sin \gamma \cos \theta. \end{aligned}$$

The magnitude of a 3D vector is calculated in the same way as for a 2D vector

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}.$$

1.7.3 Dot Product

The dot product of two vectors is also sometimes called the “inner” product or the “scalar” product. The dot product of two vectors \vec{u} and \vec{v} is

$$\vec{u} \bullet \vec{v} = u_1 v_1 + u_2 v_2 + u_3 v_3.$$

Notice that the dot product results in a scalar (number).

The dot product has the properties

1. $\vec{u} \bullet \vec{v} = \vec{v} \bullet \vec{u}$ (commutativity)
2. $\vec{u} \bullet \vec{u} = |\vec{u}|^2$
3. $\vec{u} \bullet (\vec{v} + \vec{w}) = \vec{u} \bullet \vec{v} + \vec{u} \bullet \vec{w}$
4. $c(\vec{u} \bullet \vec{v}) = (c\vec{u}) \bullet \vec{v}$
5. $\vec{u} \bullet \vec{0} = 0$

Notice that the dot product is not associative when all three multiplicands are vectors. This is because if you do the dot product inside of parentheses, as in $\vec{u} \bullet (\vec{v} \bullet \vec{w})$, your result will be a scalar, and it doesn't make sense to take the dot product of a scalar and a vector.

Tip:

If the dot product between two vectors is zero, then the vectors are orthogonal. Conversely, if two vectors are orthogonal, then their dot product will be zero.

Note that $|\vec{u}|$ can be called the magnitude, the norm, or the modulus of the vector.

The law of cosines states that

$$c^2 = a^2 + b^2 - 2ab \cos C.$$

Notice that this statement yields the Pythagorean theorem when the angle C is 90° .

If we construct a triangle using the vectors \vec{u} and \vec{v} , then the line closing the triangle is given by $\vec{u} - \vec{v}$, so using the law of cosines, we can write this as

$$|\vec{u} - \vec{v}|^2 = |\vec{u}|^2 + |\vec{v}|^2 - 2|\vec{u}||\vec{v}| \cos \theta.$$

Using several of the properties of the dot product, we can rewrite the left side as

$$\begin{aligned} |\vec{u} - \vec{v}|^2 &= (\vec{u} - \vec{v}) \bullet (\vec{u} - \vec{v}) \\ &= \vec{u} \bullet \vec{u} - 2\vec{u} \bullet \vec{v} + \vec{v} \bullet \vec{v} \\ &= |\vec{u}|^2 - 2\vec{u} \bullet \vec{v} + |\vec{v}|^2. \end{aligned}$$

Putting this back into the equation and simplifying, we get

$$\begin{aligned} |\vec{u}|^2 - 2\vec{u} \bullet \vec{v} + |\vec{v}|^2 &= |\vec{u}|^2 + |\vec{v}|^2 - 2|\vec{u}||\vec{v}| \cos \theta \\ \vec{u} \bullet \vec{v} &= |\vec{u}||\vec{v}| \cos \theta, \end{aligned}$$

which is the geometric definition of the dot product. So for two vectors \vec{A} and \vec{B} , their dot product can be given as

$$\vec{A} \bullet \vec{B} = AB \cos \theta,$$

where θ is the angle between them. Notice, from this definition of the dot product, we can calculate the angle between two vectors in any dimension, provided that the angle is between 0 and 180° :

$$\theta = \cos^{-1} \left(\frac{\vec{u} \bullet \vec{v}}{|\vec{u}||\vec{v}|} \right).$$

Tip:

To find the angle between two vectors \vec{a} and \vec{b} , use the $\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos \theta$ definition of the dot product and solve for θ .

Notice that

$$\begin{aligned}\hat{i} \cdot \hat{i} &= 1 \\ \hat{i} \cdot \hat{j} &= 0 \\ \hat{i} \cdot \hat{k} &= 0.\end{aligned}$$

The dot product of a unit vector with itself is 1 and the dot product of a unit vector with another is 0 since they are at right angles to each other.

If we take the dot product of a vector with a standard unit vector, we get the component along that axis:

$$\begin{aligned}\vec{u} \cdot \hat{i} &= \langle u_1, u_2, u_3 \rangle \cdot \langle 1, 0, 0 \rangle \\ &= 1u_1 + 0u_2 + 0u_3 \\ &= u_1\end{aligned}$$

Returning to our geometric definition of the dot product, we can rearrange to get the angle between a vector and an axis

$$\cos \alpha = \frac{\vec{u} \cdot \hat{i}}{|\vec{u}||\hat{i}|} = \frac{u_1}{|\vec{u}|}.$$

When asked to give the **direction cosines** for an angle in 3-D space, we give the following:

$$\cos \alpha = \frac{u_1}{|\vec{u}|}$$

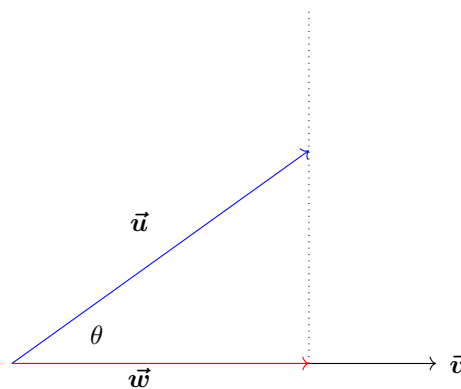
$$\cos \beta = \frac{u_2}{|\vec{u}|}$$

$$\cos \gamma = \frac{u_3}{|\vec{u}|}$$

Now we have the tools to prove that $\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1$. We simply square both sides of the three direction cosines, add them all together, and simplify.

Orthogonal Projection

A vector projection can be understood by looking at the image below where \vec{w} is the projection of \vec{u} onto \vec{v} . You can think of the projection as the shadow cast by a vector from a light shining perpendicularly onto the vector being projected upon.



There are two different projections. The **scalar projection**

$$\text{comp}_{\vec{v}} \vec{u} = \frac{\vec{v} \cdot \vec{u}}{|\vec{v}|} = |\vec{w}|,$$

where \vec{w} is the red vector (or the shadow). Notice that the scalar projection gives the magnitude of the projection. Think of it this way. The length of \vec{w} in the image above is simply $u \cos \theta$ and from the definition of the dot product as $\vec{u} \cdot \vec{v} = uv \cos \theta$, you can find the scalar projection w simply by solving for $u \cos \theta$.

The **vector projection** gives the vector \vec{w} rather than just its magnitude. It is given by

$$\text{proj}_{\vec{v}} \vec{u} = \frac{\vec{v} \cdot \vec{u}}{|\vec{v}|^2} \vec{v} = \vec{w}.$$

The formula for the scalar projection can be derived from the geometric definition of the dot product

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta.$$

Notice that $|\vec{u}| \cos \theta$, can be interpreted as the scalar projection of \vec{u} onto \vec{v} . If this isn't clear, consider the image above. Notice the right triangle formed by \vec{u} , the projection of \vec{u} onto \vec{v} , and the dotted line. Taking the cosine of theta and solving for the bottom line gives us $\vec{w} = |\vec{u}| \cos \theta$.

The vector projection of \vec{u} onto \vec{v} can be obtained simply by multiplying the scalar projection by the unit vector in the direction of \vec{v} , which is $\frac{\vec{v}}{|\vec{v}|}$.

The **Cauchy-Schwarz inequality** states that the magnitude of the dot product of two vectors is less than the product of the magnitudes of the two vectors:

$$|\vec{u} \cdot \vec{v}| \leq |\vec{u}||\vec{v}|.$$

To prove this, we start with the geometric definition of the dot product

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta.$$

Taking the absolute values of both sides gives us

$$|\vec{u} \cdot \vec{v}| = |\vec{u}||\vec{v}| |\cos \theta|.$$

Since, $|\vec{u}||\vec{v}|$ are always positive, we can pull them outside of the outer absolute value bars on the right side to get

$$|\vec{u} \cdot \vec{v}| = |\vec{u}||\vec{v}| |\cos \theta|.$$

Since the absolute value of cosine is always less than or equal to 1, we can rewrite this as

$$|\vec{u} \bullet \vec{v}| \leq |\vec{u}||\vec{v}|.$$

Prove that

$$|\vec{u} + \vec{v}| \leq |\vec{u}| + |\vec{v}|.$$

Recall that $|\vec{u}|^2 = \vec{u} \bullet \vec{u}$. Using this property and then expanding the right side using other properties of vectors, we get

$$\begin{aligned} |\vec{u} + \vec{v}|^2 &= (\vec{u} + \vec{v}) \bullet (\vec{u} + \vec{v}) \\ &= \vec{u} \bullet \vec{u} + \vec{u} \bullet \vec{v} + \vec{v} \bullet \vec{u} + \vec{v} \bullet \vec{v} \\ &= |\vec{u}|^2 + 2(\vec{u} \bullet \vec{v}) + |\vec{v}|^2. \end{aligned}$$

Utilizing the properties of the absolute value, we can rewrite this as

$$|\vec{u} + \vec{v}|^2 \leq |\vec{u}|^2 + 2|\vec{u} \bullet \vec{v}| + |\vec{v}|^2.$$

Using the Cauchy-Schwarz inequality allows us to rewrite this as

$$|\vec{u} + \vec{v}|^2 \leq (|\vec{u}|^2 + |\vec{v}|^2)^2.$$

Since both sides are positive, we can take the square roots to get

$$|\vec{u} + \vec{v}| \leq |\vec{u}| + |\vec{v}|.$$

1.7.4 Cross Product

The cross product of two 3D vectors $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$, is given by the 3×3 determinant

$$\begin{aligned} \vec{u} \times \vec{v} &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \hat{i} \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} - \hat{j} \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \hat{k} \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \\ &= (u_2v_3 - u_3v_2)\hat{i} - (u_1v_3 - u_3v_1)\hat{j} + (u_1v_2 - u_2v_1)\hat{k} \\ &= \langle u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1 \rangle. \end{aligned}$$

Notice the rearrangement done on the middle term in the last step. This was done in order to remove the leading negative sign.

Notice that the cross product is a vector, so it has a direction. To find the direction of the cross product, use the **right-hand rule**. Point the fingers of your right hand in the direction of the first vector then curl the fingers in the direction of the second vector. Your raised thumb will point in the direction of their cross product.

The cross product is defined only in \mathbb{R}^3 , at least at this level of mathematics.

The magnitude of the cross product of two vectors is

$$|\vec{a} \times \vec{b}| = |\vec{a}||\vec{b}| \sin \theta.$$

This can be verified by setting \vec{a} along the x -axis, and \vec{b} in the xy -plane. Then the components are

$$\begin{aligned} \vec{a} &= \langle a, 0, 0 \rangle \\ \vec{b} &= \langle b \cos \theta, b \sin \theta, 0 \rangle \end{aligned}$$

Then the cross product is

$$\vec{a} \times \vec{b} = \langle 0, 0, ab \sin \theta \rangle.$$

Since the magnitude is the square root of the sum of the squared components,

$$|\vec{a} \times \vec{b}| = \sqrt{(ab \sin \theta)^2} = ab \sin \theta.$$

This also demonstrates that the cross product of two vectors is orthogonal to both vectors since the vectors are in the xy -plane, and their cross product has only a z component.

Tip:

If the cross product of two vectors is 0, then one of the inputs is the zero vector or the two vectors are parallel.

Recall the properties of matrices. These properties also occur with the cross product since the cross product is defined by matrices.

1. $\vec{u} \times \vec{v} = -(\vec{v} \times \vec{u})$
2. $\vec{u} \times \vec{u} = \vec{0} = \langle 0, 0, 0 \rangle$
3. $\vec{u} \times 2\vec{u} = \vec{0} = \langle 0, 0, 0 \rangle$
4. $\vec{u} \bullet (\vec{u} \times \vec{v}) = \vec{v} \bullet (\vec{u} \times \vec{v}) = 0$
5. $(c\vec{u}) \times \vec{v} = \vec{u} \times (c\vec{v}) = c(\vec{u} \times \vec{v})$
6. $\vec{u} \times (\vec{v} + \vec{w}) = (\vec{u} \times \vec{v}) + (\vec{u} \times \vec{w})$ (distributive)

7. $\vec{u} \bullet (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \bullet \vec{w}$
 8. $\vec{u} \times (\vec{v} \times \vec{w}) \neq (\vec{u} \times \vec{v}) \times \vec{w}$ (not associative)

Notice from the first property given that the order that you apply the cross product matters. If you reverse the order of the vectors, then the resulting cross product will point in the opposite direction.

The fourth property can be verified by expanding $\vec{u} \times \vec{v}$ and $\vec{u} \times \vec{v}$ and then taking the relevant dot products. Recall that $\vec{u} \bullet \vec{v} = 0$ implies orthogonality. So what the fourth property means, geometrically, is that $\vec{u} \times \vec{v}$ is a vector that is orthogonal to both \vec{u} and \vec{v} .

Tip:

The cross product of two vectors is orthogonal to both inputs and its direction can be determined using the right-hand rule.

The fourth property may make you wonder if the cross product is associative. That is, is $\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \times \vec{w}$? If we plug in the unit vectors and work it out, it works because both results will equal the zero vector. However, this will not work in general—the cross product is not associative.

Property number four allows us to pull a constant outside of a cross product due to the fact that

$$\begin{vmatrix} a & b \\ 2c & 2d \end{vmatrix} = \begin{vmatrix} 2a & 2b \\ c & d \end{vmatrix} = 2 \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

where 2 could be any constant.

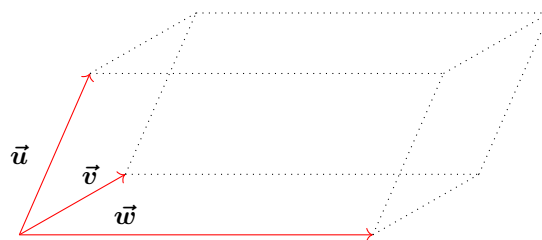
To justify number six, we could do it the hard way and expand everything out manually, or we can use determinant addition to show that

$$\begin{aligned} \vec{u} \times (\vec{v} + \vec{w}) &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1 & u_2 & u_3 \\ v_1 + w_1 & v_1 + w_2 & v_3 + w_3 \end{vmatrix} \\ &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_1 & v_3 \end{vmatrix} + \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \\ &= (\vec{u} \times \vec{v}) + (\vec{u} \times \vec{w}). \end{aligned}$$

When we expand $\hat{i} \times \hat{j}$, we find that it equals \hat{k} . By the first property of the cross product, this implies that $\hat{j} \times \hat{i} = -\hat{k}$. Similarly, $\hat{k} \times \hat{i} = \hat{j}$ and $\hat{i} \times \hat{k} = -\hat{j}$. The cross product of any unit vector with itself is the zero vector.

The seventh property has its own name. It's called the **triple product**. It can be justified by expanding the

cross product on both sides and then doing some rearranging. The triple product gives the volume of a parallelepiped defined by three vectors \vec{u} , \vec{v} , and \vec{w} at one corner.



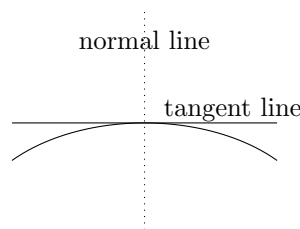
The volume of a parallelepiped is given by

$$V = |\vec{u} \bullet (\vec{v} \times \vec{w})| = |(\vec{u} \times \vec{v}) \bullet \vec{w}|.$$

Parallel and Normal Vectors

If \vec{u} and \vec{v} are parallel, then $\vec{u} = c_1 \vec{v} \implies c_2 \vec{u} = \vec{v}$. In other words, if one vector is parallel to the other, then they are scalar multiples of each other. Also, recall that the cross product of two parallel vectors is the zero vector: $\vec{u} \times c\vec{u} = \vec{0}$. To find a vector parallel to a given vector, you simply find the vector or set of vectors that results in the zero vector when the cross product is taken.

Recall from earlier courses, that a line tangent to a circle touches the circle at a single point. The **normal** line is a line running from the radius to the point of tangency. The normal line is perpendicular to the tangent line.



In three dimensions, there are infinite vectors that are orthogonal to a given vector. For a given vector the set of all vectors that are orthogonal to that vector forms a disk around the given vector. There are also an infinite number of vectors that are orthogonal to a pair of given vectors, which describe a plane.

Recalling that the cross product of two vectors gives a vector that is orthogonal to both vectors allows us to find vectors that are orthogonal to a plane. Because two vectors can describe a plane, their cross product yields a vector that is orthogonal to that plane.

Given three points, P , Q , and R , we can find a vector orthogonal to the plane containing those three points, by first finding the vectors \vec{PQ} and \vec{PR} . The cross product of those two vectors yields a vector orthogonal to that plane. Recall that the vector described by the initial point $P = (x_1, y_1, z_1)$ and final point $Q = (x_2, y_2, z_2)$ is given by

$$\vec{PQ} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1 \rangle.$$

The magnitude of the cross product of two vectors gives the area of the parallelogram defined by those two vectors. This implies that the area of a triangle formed by the points P , Q , and R , is given by

$$A = \frac{1}{2} \left| \vec{PQ} \times \vec{PR} \right|.$$

Example:

Find all vectors \vec{u} such that

$$\langle 1, 2, 3 \rangle \times \vec{u} = \langle 2, -13, 8 \rangle.$$

Expanding \vec{u} as $\langle u_1, u_2, u_3 \rangle$ and doing the cross product gives us

$$\langle 1, 2, 3 \rangle \times \vec{u} = \langle 2u_3 - 3u_2, 3u_1 - u_3, u_2 - 2u_1 \rangle.$$

Now we have the system of 3 linear equations

$$\begin{aligned} 2u_3 - 3u_2 &= 2 \\ 3u_1 - u_3 &= -13 \\ u_2 - 2u_1 &= 8 \end{aligned}$$

The system of equations has infinite solutions. That means there are infinite vectors \vec{u} which satisfy the equation. The solution can be written in the form

$$\vec{u} = \langle u_1, 8 + 2u_1, 13 + 3u_1 \rangle.$$

Example:

Show that

$$|\vec{u} \times \vec{v}| = |\vec{u}||\vec{v}| \sin \theta.$$

Expanding the cross product on the left side of the equation gives us

$$\vec{u} \times \vec{v} = \langle u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1 \rangle$$

Between 0 and π , $\sin \theta \geq 0$, which demonstrates that both sides of the above equation is positive. This allows us to square both sides.

$$|\vec{u} \times \vec{v}|^2 = |\vec{u}|^2|\vec{v}|^2 \sin^2 \theta$$

Rewriting, gives us

$$|\vec{u} \times \vec{v}|^2 = (u_2v_3 - u_3v_2)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2.$$

Expanding the right side gives us

$$|\vec{u} \times \vec{v}|^2 = u_2^2v_3^2 - 2u_2v_3u_3v_2 + u_3^2v_2^2 + u_3^2v_1^2 - 2u_3v_1u_1v_3 + u_1^2v_3^2 + u_1^2v_2^2 - 2u_1v_2u_2v_1 + u_2^2v_1^2.$$

Notice that $(u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2)$ gives us most of the above terms. However, this expression gives some extra terms and also doesn't give all of the terms in the above expansion, so we can't say that this expression equals the right side without taking care of the extra terms. Subtracting $(u_1v_1 + u_2v_2 + u_3v_3)$ takes care of them, so

$$|\vec{u} \times \vec{v}|^2 = (u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_1 + u_2v_2 + u_3v_3).$$

This can be rewritten as

$$|\vec{u} \times \vec{v}|^2 = |\vec{u}|^2|\vec{v}|^2 - (\vec{u} \cdot \vec{v})^2.$$

Because $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$, we can rewrite this as

$$|\vec{u} \times \vec{v}|^2 = |\vec{u}|^2|\vec{v}|^2 - (|\vec{u}||\vec{v}| \cos \theta)^2.$$

Factoring out the norms squared from the right side gives us

$$\begin{aligned} |\vec{u} \times \vec{v}|^2 &= |\vec{u}|^2|\vec{v}|^2(1 - \cos^2 \theta) \\ &= |\vec{u}|^2|\vec{v}|^2 \sin^2 \theta \end{aligned}$$

Taking the square root of both sides gives us our answer

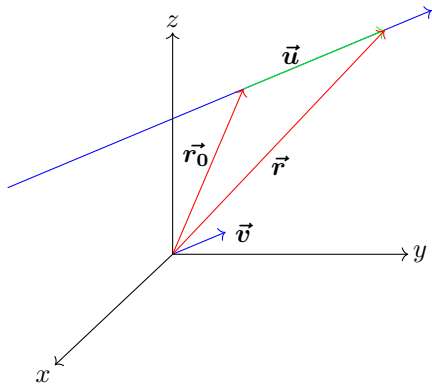
$$|\vec{u} \times \vec{v}| = |\vec{u}||\vec{v}| \sin \theta$$

1.7.5 Equations of Lines and Planes

Recall that in \mathbb{R}^2 , a straight line can be represented by $y = mx + b$, and a point on that line can be represented by (x, y) or $(x, mx + b)$. We can substitute $x = t$, then $y = mt + b$ and the point on the line is $(t, mt + b)$. This

$$\begin{cases} x = t \\ y = mt + b \end{cases} \quad \begin{cases} x = f(t) \\ y = g(t) \end{cases}$$

is known as the parametric form of the line, where t is the "parameter". A generic function $y = f(x)$ in parametric form is $(x, f(x))$. In \mathbb{R}^3 , lines are a little trickier.



Consider the image above. Our line of interest is the long blue line. We know that for any straight line, there will always be a line through the origin that is parallel. This allows us to assume the existence of a fixed unit vector, \vec{v} , that has the same direction as our line. The vector \vec{r}_0 is a fixed vector that points to a fixed point, P_0 , on the line. The vector \vec{r} is a movable vector pointing to a “free” point, P , on the line. The vector \vec{u} is a vector pointing from the fixed point to the free point.

What we want is an equation for the line L so that L is in the direction of \vec{v} . Since \vec{u} and \vec{v} are parallel, they are scalar multiples of each other. Writing down the relevant vectors, we have

$$\begin{aligned}\vec{r}_0 &= \langle x_0, y_0, z_0 \rangle \\ \vec{r} &= \langle x, y, z \rangle \\ \vec{v} &= \langle a, b, c \rangle \\ \vec{u} &= \langle ta, tb, tc \rangle\end{aligned}$$

Notice that \vec{r} is the sum $\vec{r}_0 + \vec{u}$. The equation

$$\vec{r} = \vec{r}_0 + t\vec{v}$$

is a **vector equation** of the line, since the tip of the vector traces out the line as t varies. This sum can be written as

$$\begin{aligned}\langle x, y, z \rangle &= \langle x_0, y_0, z_0 \rangle + \langle ta, tb, tc \rangle \\ &= \langle x_0 + ta, y_0 + tb, z_0 + tc \rangle,\end{aligned}$$

which implies the parametric form of the equation of the line through P_0 and parallel to \vec{v} :

$$\begin{cases} x = x_0 + ta \\ y = y_0 + tb \\ z = z_0 + tc \end{cases}.$$

The components, a , b , and c are called the **direction numbers** because they come from the fixed vector \vec{v} . Notice that each of the three parametric equations is in the form of an equation of a line. As such, a can be thought of as the slope of the line in \mathbb{R}^3 that is parallel to the x -axis, and likewise with the others.

Notice that with these parametric equations, we can easily tell where our line intersects with a plane. For

example, the line will intersect with the xy -plane when $z = 0$.

We can rewrite the parametric equations in symmetric form where there is no dependence on t :

$$t = \frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

Example:

Given the fixed point $\vec{r}_0 = \langle 1, 2, 3 \rangle$ and the direction vector $\vec{v} = \langle 4, 5, 6 \rangle$, find the equation of the line in parametric form. Find where the line passes through the xy -plane.

To do this, we simply plug the numbers in, so the parametric equation of the line is

$$\begin{cases} x = 1 + 4t \\ y = 2 + 5t \\ z = 3 + 6t \end{cases}.$$

To find where the line passes through the xy -plane, we set $z = 0$, so $3 + 6t = 0$ implying that $t = -\frac{1}{2}$. To find the exact point, we plug this value for t back into the equations for x , y , and z , which gives us $(-1, -\frac{1}{2}, 0)$.

Example:

Given the points $A = (1, 2, 3)$ and $B = (2, -1, 4)$ find the equation of the straight line connecting them.

We are not explicitly given a direction vector \vec{v} , however, the vector \vec{AB} is parallel to the line, so we can use

$$\vec{AB} = \vec{v} = \langle 2 - 1, -1 - 2, 4 - 3 \rangle = \langle 1, -3, 1 \rangle,$$

thus our direction numbers are $a = 1$, $b = -3$, and $c = 1$ and our line is given by

$$\begin{cases} x = 1 + t \\ y = 2 - 3t \\ z = 3 + 1t \end{cases}.$$

Notice in the example above that our parametric equations give the first point when we plug in $t = 0$ and they give the second point when we plug in $t = 1$. This means we can describe a **line segment** defined by an initial point and a final point using the same process with the added condition that $0 \leq t \leq 1$.

This implies that the line segment from the tip of \vec{r}_0 to the tip of \vec{r}_1 , is given by the vector equation

$$\vec{r}(t) = (1 - t)\vec{r}_0 + t\vec{r}_1 \quad 0 \leq t \leq 1.$$

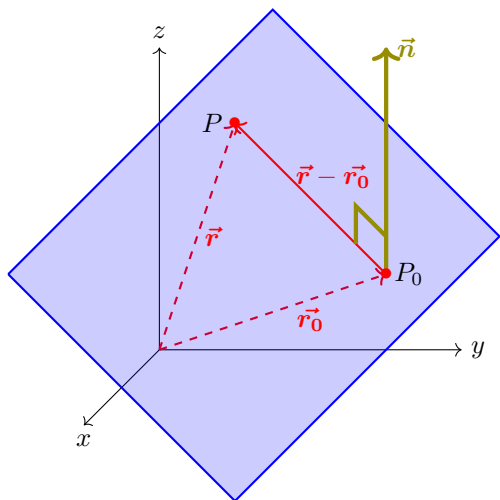
If we're asked to find the intersection of two lines L_1 and L_2 , we set up two different sets of parametric equations. **Note that we can't use t for both parameters.**

$$L_1 = \begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = z_0 + ct \end{cases} \quad L_2 = \begin{cases} x = x_1 + as \\ y = y_1 + bs \\ z = z_1 + cs \end{cases}$$

Now we have two unknowns— t and s and three equations, so we simply solve it as a system of equations—setting the x equations equal to each other and so on. In addition to getting a single solution, we could get a redundancy if one or more of the equations are the same, implying that one or more of the lines are the same and there are infinite solutions. We could also get a contradiction implying that there is no solution. When a pair of lines does not intersect, they are called **skew lines**.

Planes

A line is determined by 2 points. A plane is determined by 3, non-collinear points. A single vector parallel to a plane is not sufficient to determine the plane, however, a single vector orthogonal to the plane, is.



In the image, the red vector \vec{r}_0 is a fixed vector pointing to the fixed point P_0 that lies in the blue plane. The red vector \vec{r} is a free vector that points to the point P , which is also in the plane. Notice that the vectors \vec{r} and \vec{r}_0 are not in the plane, however, the vector connecting the two endpoints $\vec{r} - \vec{r}_0$ is in the plane. Since \vec{r} is a free vector whose endpoint is always in the plane, the vector that is normal to $\vec{r} - \vec{r}_0$ must be normal to the plane. Since $\vec{n} \perp \vec{r} - \vec{r}_0$, it must be the case that

$$\begin{aligned} \vec{n} \cdot (\vec{r} - \vec{r}_0) &= 0 \\ \vec{n} \cdot \vec{r} &= \vec{n} \cdot \vec{r}_0. \end{aligned}$$

We can convert these vector equations for the plane into a scalar equation of the plane using the normal vector $\vec{n} = \langle a, b, c \rangle$, the free point $P = \langle x, y, z \rangle$, and the fixed

point $P_0 = \langle x_0, y_0, z_0 \rangle$. Then

$$\begin{aligned} \vec{n} \cdot (\vec{r} - \vec{r}_0) &= 0 \\ \langle a, b, c \rangle \cdot \langle x - x_0, y - y_0, z - z_0 \rangle &= 0 \\ a(x - x_0) + b(y - y_0) + c(z - z_0) &= 0. \end{aligned}$$

This last equation is the **scalar equation of the plane** through the fixed point $P_0 = \langle x_0, y_0, z_0 \rangle$ and with normal vector $\vec{n} = \langle a, b, c \rangle$.

If we expand this scalar equation of the plane we get

$$ax + by + cz = ax_0 + by_0 + cz_0.$$

Recall that a , b , and c as well as x_0 , y_0 , and z_0 are just regular numbers, so we can combine them all into a single arbitrary constant

$$ax + by + cz + d = 0.$$

This simple equation represents a plane with a normal vector $\langle a, b, c \rangle$.

To find the equation of a plane specified by three non-collinear points, we simply construct two different vectors between the three points and then take the cross product of the two vectors in order to obtain $\vec{n} = \langle a, b, c \rangle$. Then we simply plug these numbers and one of the fixed points into

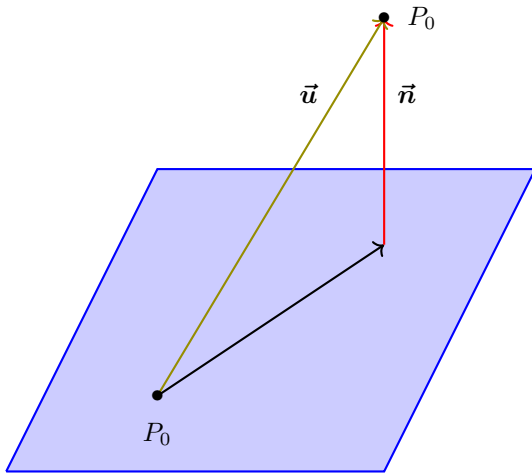
$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

Remember to find the vector between two points it is **final point minus initial point**. So for $P(1, 3, 2)$ and $Q(3, -1, 6)$, the vector \vec{PQ} starting at P and ending at Q is $\langle 3 - 1, -1 - 3, 6 - 2 \rangle = \langle 2, -4, 4 \rangle$.

When asked to give the angle between two intersecting angles, give the small one—the acute angle. The angle between two intersecting planes is the same as the angle between their normals. Let \vec{n}_1 and \vec{n}_2 be their respective normals, then use the definition of the dot product to find the angle between them:

$$\vec{n}_1 \cdot \vec{n}_2 = |\vec{n}_1| |\vec{n}_2| \cos \theta.$$

To find the distance between a point and a line in \mathbb{R}^2 , we find the equation of the line that is perpendicular to the first line and goes through the point. Then you solve the system of equations created by setting the two lines equal to each other to get the point of intersection. Then find the distance between the intersection point and the given point.



To find the distance from a plane, $ax+by+cz+d=0$, to a point, P_1 , off of a plane is a little more complicated. First, pick any point P_0 that is in the plane. Since the coordinates of this point must satisfy the equation of the plane, we have that $ax_0+by_0+cz_0+d=0$. Now, find the vector going from P_0 to P_1 , which is

$$\vec{u} = \vec{P_0P_1} = \langle x_1 - x_0, y_1 - y_0, z_1 - z_0 \rangle.$$

Now the distance from the point to the plane is the same as the scalar projection of \vec{u} onto the normal vector $\vec{n} = \langle a, b, c \rangle$, so we have that

$$D = \text{comp}_{\vec{n}} \vec{u} = \frac{\vec{u} \cdot \vec{n}}{|\vec{n}|} = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

1.7.6 Quadric Surfaces

A **quadric surface** is the graph of a second-degree equation in three variables. The most general such equation is

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0.$$

By translation and rotation, any such equation can

be brought into one of the following two standard forms:

$$Ax^2 + By^2 + Cz^2 + J = 0,$$

$$Ax^2 + By^2 + Iz = 0.$$

To sketch the shape of a quadric surface use **traces**, which involves setting one variable equal to 0 in order to see the shape drawn in that plane.

1.8 Vector Functions

1.8.1 Vector Functions and Space Curves

Up to this point, we have been used to dealing with scalar, real-valued functions. Now we will be using vector-valued functions. A scalar function goes from a domain of real numbers to a range of real numbers. A vector function goes from a domain of real numbers to a range of vectors.

A vector function is given as

$$\vec{r}(t) = \langle f(t), g(t), h(t) \rangle,$$

where $f(t)$, $g(t)$, and $h(t)$ are the regular scalar functions we are used to.

The biggest difference when dealing with these functions as opposed to scalar functions is that we can't graph them. A vector function is strongly related to a curve in \mathbb{R}^3 because the tip of the vector \vec{r} traces out the curve as t goes from negative infinity to infinity.

To find the domain of a vector function \vec{r} , we look at the domains of the component functions f , g , and h . Then the domain of the vector function is the intersection (or most restrictive) of the domains of the component functions.

The limit of a vector function is the vector of the limit of the component functions provided that each limit exists:

$$\lim_{t \rightarrow a} \vec{r}(t) = \langle \lim_{t \rightarrow a} f(t), \lim_{t \rightarrow a} g(t), \lim_{t \rightarrow a} h(t) \rangle.$$

Notice that a vector function has component functions, whereas the limit of a vector function will be a constant vector—with scalar components.

A vector function \vec{r} is continuous at a if

$$\lim_{t \rightarrow a} \vec{r}(t) = \vec{r}(a).$$

Note that for a vector function to be continuous at a point, each component function has to be defined at that point, the limit of each component function has to exist at that point, and the limit of each function at that point has to each the value of the function at that point.

Is the limit of the dot product of two functions, the dot product of the limits of the two functions? Here we assume that the component functions are nice functions with shared domains, and we prove it for two dimensional vector functions although the same reasoning applies to three dimensional vector functions. Given the two vector functions

$$\begin{aligned} \vec{r}_1 &= \langle f_1(t), f_2(t) \rangle \\ \vec{r}_2 &= \langle g_1(t), g_2(t) \rangle \end{aligned}$$

Taking the dot product of the two vector functions, gives us

$$\vec{r}_1 \bullet \vec{r}_2 = f_1(t)g_1(t) + f_2(t)g_2(t)$$

Taking the limit of the dot product gives us

$$\begin{aligned} \lim_{t \rightarrow a} (\vec{r}_1 \bullet \vec{r}_2) &= \lim_{t \rightarrow a} (f_1(t)g_1(t)) + \lim_{t \rightarrow a} (f_2(t)g_2(t)) \\ &= \lim_{t \rightarrow a} f_1(t) \cdot \lim_{t \rightarrow a} g_1(t) + \lim_{t \rightarrow a} f_2(t) \cdot \lim_{t \rightarrow a} g_2(t) \\ &= \langle \lim_{t \rightarrow a} f_1(t), \lim_{t \rightarrow a} f_2(t) \rangle \bullet \langle \lim_{t \rightarrow a} g_1(t), \lim_{t \rightarrow a} g_2(t) \rangle \\ &= \lim_{t \rightarrow a} \langle f_1(t), f_2(t) \rangle \bullet \lim_{t \rightarrow a} \langle g_1(t), g_2(t) \rangle \\ &= \lim_{t \rightarrow a} \vec{r}_1 \bullet \lim_{t \rightarrow a} \vec{r}_2 \end{aligned}$$

Doing the same for the cross product is a bit more tedious and requires the use of the matrix definition of the cross product.

To find the vector function of the line segment connecting two points, recall that the vector equation for the line segment that joins the tip of \vec{r}_0 and the tip of \vec{r}_1 is given by

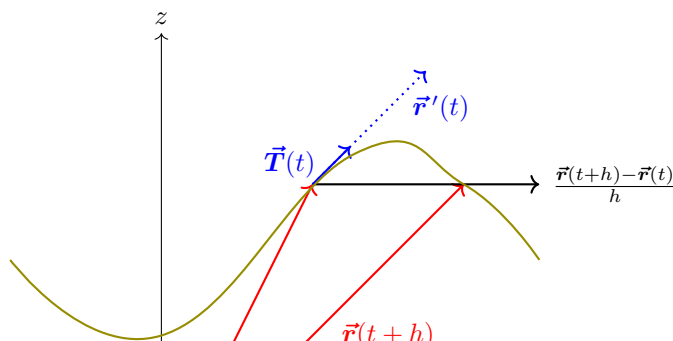
$$\vec{r}(t) = (1-t)\vec{r}_0 + t\vec{r}_1 \quad 0 \leq t \leq 1.$$

If given the points are $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$, we let P be the tip of \vec{r}_0 and let Q be the tip of \vec{r}_1 , so that

$$\vec{r}(t) = (1-t)\langle x_1, y_1, z_1 \rangle + t\langle x_2, y_2, z_2 \rangle \quad 0 \leq t \leq 1.$$

Then just expand everything out and combine into a single vector with components that are scalar functions of t .

1.8.2 Derivatives and Integrals of Vector Functions



Notice from the image, that if $\vec{r}(t)$ is our vector corresponding to the green curve going through \mathbb{R}^3 , then $\vec{r}(t+h)$ corresponds to the vector pointing to a point further along that curve. The vector from the endpoint of the first vector to the endpoint of the second vector is given by $\vec{r}(t+h) - \vec{r}(t)$. Multiplying this vector by the scalar $\frac{1}{h}$ gives the vector extending through the final point. Notice that as h gets smaller and smaller, $\vec{r}(t+h) - \vec{r}(t)$ comes closer and closer to being the tangent vector $\vec{r}'(t)$. This implies that

$$\vec{r}'(t) = \lim_{h \rightarrow 0} \frac{\vec{r}(t+h) - \vec{r}(t)}{h},$$

provided that the limit exists.

This looks the same as the definition of the derivative of a scalar function, and it is. It implies that for a vector function $\vec{r}(t) = \langle f(t), g(t), h(t) \rangle$

$$\vec{r}'(t) = \langle f'(t), g'(t), h'(t) \rangle,$$

so differentiating a vector function is a straightforward matter of differentiating each of the component functions.

Recall that differentiability implies continuity, but continuity does not imply differentiability. The Weierstrass function is a good example. It is continuous everywhere, but differentiable nowhere.

Also of use will be the unit vector of the derivative of \vec{r} , which is given by

$$\vec{T}(t) = \frac{\vec{r}'(t)}{|\vec{r}'(t)|}.$$

Here are the properties of vector function derivatives where f is a scalar function and c is a scalar number:

1. The derivative of a sum of vector functions is the sum of the derivatives of the vector functions:

$$[\vec{u}(t) + \vec{v}(t)]' = \vec{u}'(t) + \vec{v}'(t)$$

It follows from the sum property of scalar function derivatives.

2. The derivative of a constant times a vector function is the constant times the derivative of the vector function.

$$[c\vec{u}(t)]' = c\vec{u}'(t)$$

3. The derivative of a scalar function times a vector function gives our first analogue of the regular product formula for derivatives.

$$[f(t)\vec{u}(t)]' = f'(t)\vec{u}(t) + f(t)\vec{u}'(t)$$

4. Here's the dot product version:

$$[\vec{u}(t) \bullet \vec{v}(t)]' = \vec{u}'(t) \bullet \vec{v}(t) + \vec{u}(t) \bullet \vec{v}'(t)$$

5. And here's the cross product version

$$[\vec{u}(t) \times \vec{v}(t)]' = \vec{u}'(t) \times \vec{v}(t) + \vec{u}(t) \times \vec{v}'(t)$$

6. Lastly, here's the chain rule analogue:

$$[\vec{u}(f(t))]' = \vec{u}'(f(t)) f'(t).$$

Notice that there are three product properties corresponding to the two new types of products that we have with vectors. Also, there is no quotient rule because dividing one vector by another is not a defined operation.

A relatively short proof of property # 5 can be obtained using matrices. One property of matrices is that the derivative of a matrix containing functions is the matrix containing the derivatives of the function:

$$\frac{d}{dt} \begin{bmatrix} f(t) & g(t) & h(t) \\ i(t) & j(t) & k(t) \\ l(t) & m(t) & n(t) \end{bmatrix} = \begin{bmatrix} f'(t) & g'(t) & h'(t) \\ i'(t) & j'(t) & k'(t) \\ l'(t) & m'(t) & n'(t) \end{bmatrix}$$

To find the derivative of a determinant containing functions, we take the derivative of one row at a time like this:

$$\frac{d}{dt} \begin{vmatrix} f(t) & g(t) & h(t) \\ i(t) & j(t) & k(t) \\ l(t) & m(t) & n(t) \end{vmatrix} = \begin{vmatrix} f'(t) & g'(t) & h'(t) \\ i(t) & j(t) & k(t) \\ l(t) & m(t) & n(t) \end{vmatrix} + \begin{vmatrix} f(t) & g(t) & h(t) \\ i'(t) & j'(t) & k'(t) \\ l(t) & m(t) & n(t) \end{vmatrix} + \begin{vmatrix} f(t) & g(t) & h(t) \\ i(t) & j(t) & k(t) \\ l'(t) & m'(t) & n'(t) \end{vmatrix}$$

This means that $\frac{d}{dt}[\vec{u}(t) \times \vec{v}(t)]$ is

$$\begin{aligned} \frac{d}{dt} \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1(t) & u_2(t) & u_3(t) \\ v_1(t) & v_2(t) & v_3(t) \end{vmatrix} &= \begin{vmatrix} \vec{0} & \vec{0} & \vec{0} \\ u_1(t) & u_2(t) & u_3(t) \\ v_1(t) & v_2(t) & v_3(t) \end{vmatrix} + \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u'_1(t) & u'_2(t) & u'_3(t) \\ v_1(t) & v_2(t) & v_3(t) \end{vmatrix} + \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ u_1(t) & u_2(t) & u_3(t) \\ v'_1(t) & v'_2(t) & v'_3(t) \end{vmatrix} \\ &= \vec{0} + \vec{u}'(t) \times \vec{v}(t) + \vec{u}(t) \times \vec{v}'(t) \\ &= \vec{u}'(t) \times \vec{v}(t) + \vec{u}(t) \times \vec{v}'(t) \end{aligned}$$

Integrating vector functions is as easy as differentiating them—you just integrate the component functions.

$$\int \vec{r}(t) dt = \left\langle \int f(t) dt, \int g(t) dt, \int h(t) dt \right\rangle$$

When doing indefinite integration, make sure that each component integration includes a constant of integration. For definite integration, the same process is used:

$$\int_a^b \vec{r}(t) dt = \left\langle \int_a^b f(t) dt, \int_a^b g(t) dt, \int_a^b h(t) dt \right\rangle$$

Note that indefinite integration of a vector function will result in a vector function, and definite integration of a vector function results in a constant vector containing scalar components.

To find the derivative of a vector function containing vectors, just use the rules of derivatives. For example, if $\vec{r}(t) = \vec{a} + t\vec{b} + t^2\vec{c}$ then $\vec{r}'(t) = \vec{b} + 2t\vec{c}$.

Example:

Find the parametric equations for the tangent line to the curve

$$x = \sqrt{t}, \quad y = 3 + t^2, \quad z = t^2 + 3t$$

at the point $(1, 4, 4)$.

First, we write the parametric form of the curve as the vector equation

$$\vec{r}(t) = \langle \sqrt{t}, 3 + t^2, t^2 + 3t \rangle.$$

Next, we solve for the parameter t such that the function puts out $(1, 4, 4)$. We find that $t = 1$. Now, the question is, what is the tangent vector when $t = 1$? The tangent vector can be found by differentiating the vector function to get

$$\vec{r}'(t) = \left\langle \frac{1}{2\sqrt{t}}, 2t, 2t + 3 \right\rangle,$$

and then finding the tangent vector when $t = 1$, which is given by

$$\vec{r}'(1) = \left\langle \frac{1}{2}, 2, 5 \right\rangle.$$

What we are looking for is the line that goes through $(1, 4, 4)$ and is parallel to the vector $\vec{r}'(1) = \langle \frac{1}{2}, 2, 5 \rangle$. To find this, recall that a vector parallel to $\langle a, b, c \rangle$ and going through the point (x_0, y_0, z_0) is given by

$$\begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = z_0 + ct \end{cases}$$

so in our case the tangent line at the point $(1, 4, 4)$, in parametric form, is

$$\begin{cases} x = 1 + \frac{1}{2}t \\ y = 4 + 2t \\ z = 4 + 5t \end{cases}$$

1.8.3 Arc Length and Curvature

A curve is called **rectifiable** if you can compute its arc length. The formulas for calculating the arc length L

of a curve $y = f(x)$ from point a to point b in \mathbb{R}^2 are

$$L = \int_a^b \sqrt{1 + (y')^2} dx = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt.$$

The second formula is the parametric form where $x = x(t)$ and $y = y(t)$.

When doing these, be careful that you're not dealing with an improper integral. If you are, you must take the limit of the integral as you approach the asymptote.

The parametric formula can also be used to compute arc length for curves in \mathbb{R}^3 with the addition of the z component. For the curve described by the vector function $\vec{r}(t) = \langle f(t), g(t), h(t) \rangle$ for $a \leq t \leq b$, the arc length is given by

$$\begin{aligned} L &= \int_a^b \sqrt{(f'(t))^2 + (g'(t))^2 + (h'(t))^2} dt \\ &= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt. \end{aligned}$$

We'll primarily use this parametric form since we'll be given a vector function containing the component functions f , g , and h . We just need the three component functions to be differentiable at least once.

Notice that the radical is simply the magnitude or norm of the vector $\vec{r}'(t)$. So our formula becomes

$$L = \int_a^b |\vec{r}'(t)| dt$$

For the above formulas, a and b are fixed constants. What happens if we let b be the variable t ? Then the result—the arc length—won't be a number anymore. It will be a function of t , so

$$s(t) = \int_a^t |\vec{r}'(t')| dt'.$$

Recall that the fundamental theorem of calculus states that

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

and

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

Applying the theorem of calculus to our arc length function gives us

$$\frac{ds}{dt} = |\vec{r}'(t)|.$$

Example:

Find the arc length function for $\vec{r} = \langle \cos t, \sin t, t \rangle$ with respect to the point $(1, 0, 0)$.

The first thing we do is find the derivative of the vector and its norm.

$$\vec{r}'(t) = \langle -\sin t, \cos t, 1 \rangle$$

$$|\vec{r}'(t)| = \sqrt{(-\sin t)^2 + (\cos t)^2 + (1)^2} = \sqrt{2}$$

Setting this equal to the derivative of the arc length function, which we derived earlier, gives us

$$\frac{ds}{dt} = \sqrt{2}.$$

Integrating both sides with respect to t gives us

$$s = \sqrt{2}t + c.$$

However, do we really need the constant of integration in there? Recall that our arc length function is starting at $(1, 0, 0)$, which corresponds to $t = 0$. Really, what we are doing when we integrate the derivative of s is integrating from 0 to t :

$$\int_0^t \sqrt{2} du = \sqrt{2}u \Big|_0^t = \sqrt{2}t,$$

so $c = 0$, since the arc length function is starting at 0. Now, if we solve for t , we get

$$t = \frac{1}{\sqrt{2}}s,$$

so our arc length function becomes

$$\vec{r}(s) = \left\langle \cos\left(\frac{s}{\sqrt{2}}\right), \sin\left(\frac{s}{\sqrt{2}}\right), \left(\frac{s}{\sqrt{2}}\right) \right\rangle$$

Some curves turn more sharply than others. We can quantify the curvature of a curve by looking at how the tangent vector is changing. Recall that the unit tangent of the derivative of \vec{r} , is given by

$$\vec{T}(t) = \frac{\vec{r}'(t)}{|\vec{r}'(t)|}.$$

So, what we want to know is how fast is $\vec{T}(t)$ changing with respect to the arc length s . We'll call this *curvature* and represent it with κ .

$$\kappa = \left| \frac{d\vec{T}}{ds} \right|$$

For us, κ is positive, though it could be negative in higher math courses. **Note: The absolute value bars refer to absolute value—not magnitude.**

Recall that the chain rule states that

$$\frac{dT}{dt} = \frac{dT}{ds} \cdot \frac{ds}{dt}.$$

This means we can rewrite our equation for curvature as

$$\kappa = \left| \frac{dT}{ds} \right| = \left| \frac{\frac{dT}{dt}}{\frac{ds}{dt}} \right|.$$

However, recall that $\frac{ds}{dt} = |\mathbf{r}'(t)|$ and $\frac{dT}{dt} = \left| \frac{\mathbf{T}'(t)}{|\mathbf{r}'(t)|} \right|$. This allows us to write

$$\kappa = \left| \frac{dT}{ds} \right| = \frac{\left| \mathbf{T}'(t) \right|}{|\mathbf{r}'(t)|}$$

Note that the bars on $\left| \frac{dT}{ds} \right|$ mean absolute value and the bars on $\frac{|\mathbf{T}'(t)|}{|\mathbf{r}'(t)|}$ mean magnitude or norm. The disadvantage of using the first one is that we have to find the arc length function s .

For a circle, the curvature is a fixed constant. Smaller circles have greater curvature than larger circles. The vector function for a circle with radius a is $\mathbf{r}(t) = \langle a \cos t, a \sin t \rangle$. If you find the curvature of this vector function it turns out that the curvature of a circle is inversely proportional to its radius:

$$\kappa_{\text{circle}} = \frac{1}{a}.$$

An **osculating circle** is the circle tangent to a curve at point P such that the radius of the circle is the reciprocal of the curvature of the curve. The osculating circle at point P is the circle that best describes how the curve behaves near P . It can be thought of as the largest circle on the inside of the curve, tangent to the curve at P , and not extending past the curve at any point.

The radius of the osculating circle changes as you move along the curve—it is a function of the point on the line. If we let ρ be the radius of the osculating circle, then

$$\rho(t) = \frac{1}{\kappa(t)}.$$

Here is another formula for curvature:

$$\kappa(t) = \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|}{|\mathbf{r}'(t)|^3}.$$

This formula is good for space curves, but not for planar curves since the cross product is not defined for two dimensions.

Here is a formula for the curvature of a planar curve $y = f(x)$.

$$\kappa(x) = \frac{|f''(x)|}{|1 + (f'(x))^2|^{\frac{3}{2}}}.$$

This is a special case of the formula before this. We just let $t = x$ and there's one less dimension to deal with, so it simplifies to this equation.

The parametric form of the last equation is

$$\kappa(x) = \frac{|x'y'' - y'x''|}{(x'^2 + y'^2)^{\frac{3}{2}}}.$$

Use this to find the curvature when given $x(t)$ and $y(t)$. Instead of using this equation, you may also be able to eliminate the parameter to make the calculation of the

curvature easier. For example, if given $y = \cos t$ and $x = \sin t$, you could square both and add them together to get the relation $x^2 + y^2 = 1$, which doesn't contain t .

Recall that the unit tangent vector is given by

$$\vec{T}(t) = \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|}.$$

The derivative of this vector is an orthogonal vector that points toward the center of curvature. We call this the **normal vector**. The unit normal vector is given by

$$\vec{N}(t) = \frac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|}.$$

The **binormal vector** is orthogonal to both the tangent and the normal vectors. It can be obtained by taking the cross product of the two. The unit binormal vector is given by

$$\vec{B}(t) = \vec{T}(t) \times \vec{N}(t).$$

We know this is a unit vector because the other two vectors are unit vectors, and the cross product of two orthogonal unit vectors is a unit vector. Recall that $|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}|\sin\theta$. Plug in the unit tangent and unit normal vectors and you see that their cross product must be unity too.

When finding the binormal vector, double-check your work by finding its magnitude. It should be one.

Jacobi gives us the identity

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}.$$

Using that identity, can we get the relation $\vec{B}(t) = \vec{T}(t) \times \vec{N}(t)$ from the relation $\vec{N}(t) = \vec{B}(t) \times \vec{T}(t)$? Using the identity $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$, we can write

$$\vec{N}(t) = -\left(\vec{T}(t) \times \vec{B}(t)\right).$$

If we cross both sides by $\vec{T}(t)$, we get

$$\vec{T}(t) \times \vec{N}(t) = \vec{T}(t) \times \left(-\vec{T}(t) \times \vec{B}(t)\right).$$

On the right side, we can pull the negative out to the front since it is a constant and $\mathbf{u} \times (c\mathbf{v}) = c(\mathbf{u} \times \mathbf{v})$:

$$\vec{T}(t) \times \vec{N}(t) = -\vec{T}(t) \times \left(\vec{T}(t) \times \vec{B}(t)\right).$$

Now, using the Jacobi identity given earlier, we can rewrite the right side and simplify

$$\vec{T}(t) \times \vec{N}(t) = \left(\vec{T}(t) \cdot \vec{T}(t)\right) \vec{B}(t) - \left(\vec{T}(t) \cdot \vec{B}(t)\right) \vec{T}(t).$$

We know that $\vec{T}(t) \cdot \vec{T}(t) = |\vec{T}|^2$ and since $\vec{T}(t)$ is a unit vector, we know that $|\vec{T}|^2 = 1$. We also know that $\vec{T}(t) \cdot \vec{B}(t) = 0$ because they are orthogonal. So our equation simplifies to

$$\vec{T}(t) \times \vec{N}(t) = \vec{B}(t).$$

Curvature is used a lot in differential geometry—a field of mathematics that is very important in modern physics.

1.8.4 Velocity and Acceleration

If $\vec{r}(t) = \langle f(t), g(t), h(t) \rangle$ is the position vector function of a particle with respect to time, t , then

$$\vec{r}'(t) = \langle f'(t), g'(t), h'(t) \rangle = \vec{v}(t)$$

is the velocity function for the particle and

$$\vec{r}''(t) = \langle f''(t), g''(t), h''(t) \rangle = \vec{v}'(t) = \vec{a}(t),$$

is the acceleration function for the particle.

The magnitude of the velocity function $|\vec{v}(t)|$ gives the scalar speed of the particle.

To find the velocity of position vector given the acceleration vector requires integration. This will necessitate the introduction of a constant C unless it is an initial value problem with the initial value specified.

Centripetal Force

What is the centripetal force acting on a particle moving in a circle at constant speed? An object with mass m and angular speed ω has a position vector $\vec{r}(t) = \langle a \cos \omega t, a \sin \omega t \rangle$. Newton's second law gives force as $\vec{F}(t) = m\vec{a}(t)$. Differentiating the position vector twice gives us $\vec{a}(t) = \langle -a\omega^2 \cos \omega t, -a\omega^2 \sin \omega t \rangle = -\omega^2 \vec{r}$, so the force acting on the particle is $\vec{F}(t) = -m\omega^2 \vec{r}$. The negative sign indicates that the force operates in the opposite direction of the position vector. In other words, the centripetal force points toward the center.

Projectile Motion

What is the position vector and horizontal distance traveled of a projectile fired with initial velocity \vec{v}_0 at an angle of elevation α ?

We can assume that the only force acting on the particle is gravity, and that it is constant throughout the projectile's path. The acceleration vector, then, is given by $\vec{a}(t) = \langle 0, -g \rangle$ since it acts only in the negative y -direction. Integrating the acceleration vector gives us the velocity vector

$$\begin{aligned} \vec{v}(t) &= \langle c_1, -gt + c_2 \rangle \\ &= \langle 0, -gt \rangle + C \\ &= \langle 0, -gt \rangle + \vec{v}_0 \end{aligned}$$

We know that the C is equal to \vec{v}_0 simply by solving $\vec{v}(0)$. If we integrate the velocity function, we get the position function

$$\vec{r}(t) = \langle c_1, -\frac{1}{2}gt^2 + c_2 \rangle + \vec{v}_0 t + c_3.$$

Solving $\vec{r}(0)$ and knowing that we're starting from the point $(0, 0)$, we find that all of the constants are zero, so we have

$$\vec{r}(t) = \langle 0, -\frac{1}{2}gt^2 \rangle + \vec{v}_0 t.$$

By looking at the right triangle formed with \vec{v}_0 as the hypotenuse, we know that

$$\vec{v}_0 = \langle |v_0| \cos \alpha, |v_0| \sin \alpha \rangle,$$

and our function of \vec{v}_0 becomes

$$\vec{v}_0(t) = \langle (|v_0| \cos \alpha)t, (|v_0| \sin \alpha)t \rangle.$$

Putting it all together, our position function becomes

$$\vec{r}(t) = \langle (|v_0| \cos \alpha)t, -\frac{1}{2}gt^2 + (|v_0| \sin \alpha)t \rangle.$$

The max distance traveled by the projectile will be given by the first component, $(|v_0| \cos \alpha)t$, when the second component $-\frac{1}{2}gt^2 + (|v_0| \sin \alpha)t$ is zero. Setting the second component equal to zero and plugging the resulting value for t into the first component, we find that the max horizontal distance is given by

$$d = \frac{|v_0|^2 \sin 2\alpha}{g}.$$

Tangential and Normal Components of Acceleration Notice that

$$\vec{T}(t) = \frac{\vec{r}'(t)}{|\vec{r}'(t)|} = \frac{\vec{v}(t)}{|\vec{v}(t)|} = \frac{\vec{v}}{v}$$

Solving for the velocity vector using the far left and far right of the above, we find that the velocity function equals the unit tangent function times the speed

$$\vec{v} = v\vec{T}.$$

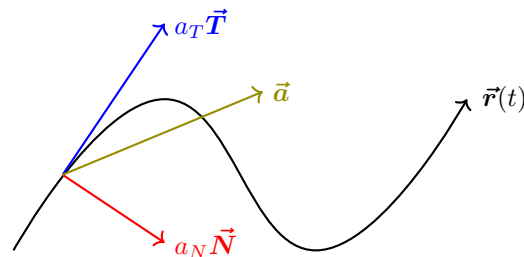
Note that the speed v is not static—it changes with respect to where you are. Differentiating both sides of the last equation, we get

$$\vec{a} = v'\vec{T} + v\vec{T}',$$

which can be rewritten as

$$\vec{a} = a_T \vec{T} + a_N \vec{T}',$$

where a_T is the tangential component and a_N is the normal component.



Notice that

$$\kappa = \frac{|\vec{T}'|}{|\vec{r}'|} = \frac{|\vec{T}'|}{v}$$

and recall that

$$\vec{N} = \frac{\vec{T}'}{|\vec{T}'|},$$

which allows us to write

$$\vec{T}' = \kappa v \vec{N}'.$$

so we can rewrite our equation for acceleration as

$$\vec{a} = v' \vec{T} + \kappa v^2 \vec{N}.$$

This means that $a_T = v'$ is the tangential component of acceleration and $a_N = \kappa v^2$ is the normal component of acceleration.

Taking the dot product of the two results for velocity and acceleration that we found earlier, we have that

$$\begin{aligned} \vec{v} \cdot \vec{a} &= v \vec{T} \cdot (v' \vec{T} + \kappa v^2 \vec{N}) \\ &= v v' \vec{T} \cdot \vec{T} + \kappa v^3 \vec{T} \cdot \vec{N} \\ &= v v' \end{aligned}$$

Therefore,

$$a_T = v' = \frac{\vec{v} \cdot \vec{a}}{v} = \frac{\vec{r}'(t) \cdot \vec{r}''(t)}{|\vec{r}'(t)|}$$

Using an earlier formula for curvature allows us to rewrite the normal component as

$$a_N = \frac{|\vec{r}'(t) \times \vec{r}''(t)|}{|\vec{r}'(t)|}$$

Recall from the definition of the magnitude of the cross product that

$$|\vec{r}'(t) \times \vec{r}''(t)| = |\vec{r}'(t)| |\vec{r}''(t)| \sin \theta,$$

where θ is the angle between the vectors. This formula can make it a lot easier to calculate the normal component of acceleration when the angle between $\vec{r}'(t)$ and $\vec{r}''(t)$ is known.

1.9 Partial Derivatives

1.9.1 Multivariable Functions

In the first part of our calculus education we considered functions of the single variables $y = f(x)$. This notation indicates that x is the independent variable and y is the dependent variable which depends on x . Such a function represents a curve in \mathbb{R}^2 .

Now, we also consider functions of two and more variables. A function of two variables can be expressed as $z = f(x, y)$. The notation indicates that x and y are independent variables and z is the dependent variable that depends on the two independent variables. Such a function represents a surface in \mathbb{R}^3 .

All of our previous work with limits, derivatives, and integrals, applied to functions of a single variable. We must relearn all of these for functions of more variables.

We can have functions of any number of variables $z = f(x_1, x_2, x_3, \dots, x_n)$. However, we cannot graph functions of more than two variables because we cannot see in more than 3-dimensions.

The **domain** of a single variable function is a length, or region, of the x -axis. Then the range of such a function is along the y -axis. The domain of a function of two variables is a region of the xy -plane. Then the range of such a function is along the z -axis.

When finding the domain of a function of two variables, we are essentially asking, what portion of the xy -plane will we be using? When finding the range of such a function, we are essentially asking, what part of the z -axis will we use? In general, finding the range is more difficult than finding the domain.

To find the domain of $y = f(x)$, we find all the locations where x is defined. To find the domain of $z = f(x, y)$, we find the region in which both x and y are defined. For polynomials like $f(x, y) = x^2 + y^2$, the domain is all real numbers. For a rational function like

$$f(x, y) = \frac{x + y}{x - y},$$

the denominator cannot be zero. In this case, the domain is all real numbers except $x = y$. What is $x = y$ in \mathbb{R}^3 ? It is an entire plane.

We can plot this function in MatLab using the code

```
fmesh(@(x,y) (x+y)./(x-y))
```

A function with one independent variable, $y = f(x)$ is a curve in \mathbb{R}^2 . A function with two independent variables $z = f(x, y)$ forms a surface in \mathbb{R}^3 . A function with three independent variables $w = f(x, y, z)$ forms a volume in \mathbb{R}^4 . Unfortunately, we can't visualize this because we cannot see in four dimensions.

What is the domain of a more complicated function like

$$f(x, y) = \sin^{-1}(x^2 + y^2 - 5)?$$

To simplify this, we start by asking, what is the domain of $\sin^{-1} x$? Since the domain is $[-1, 1]$, the domain of $f(x, y) = \sin^{-1}(x^2 + y^2 - 5)$ must be

$$-1 \leq x^2 + y^2 - 5 \leq 1,$$

which simplifies to

$$4 \leq x^2 + y^2 \leq 6.$$

This implies that $x^2 + y^2 \geq 4$ and $x^2 + y^2 \leq 6$, which means the domain of the function is an annulus, or ring, in the xy -plane with inner radius of $\sqrt{4}$ and outer radius of $\sqrt{6}$.

We can plot this function in MatLab using the code

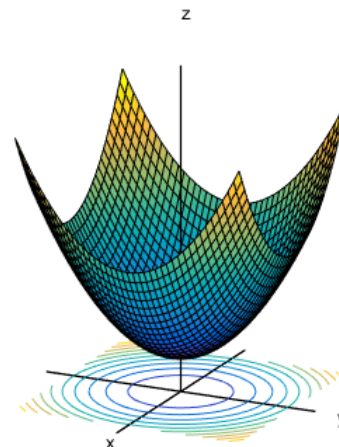
```
fmesh(@(x,y) asin(x.^2 + y.^2 - 5))
```

In addition to a **surface plot** one way to visualize functions of two variables is with **level curves**. A level curve is the set of all points where the function has the same value. To plot a level curve, first set the function equal to some constant,

$$f(x, y) = C,$$

then simplify it and plot the resulting set of points. To plot additional level curves, just repeat the process with different values for the constant C . Note that all the level curves will lie on the xy -plane.

Below is an example of a surface plot and some level curves of the function $f(x, y) = x^2 + y^2$. Notice that the level curves are all circles because $x^2 + y^2 = C$ is a circle for every value of C .



1.9.2 Limits and Continuity

For $y = f(x)$, which represents a curve in \mathbb{R}^2 , we can find the limit of the function at a certain point by approach-

ing that point on the curve from the left and right. For $z = f(x, y)$, it's more complicated. Since we now have a surface in \mathbb{R}^3 , we would have to approach the point from an infinite number of directions to guarantee the existence of a limit that way. This means, we cannot guarantee the existence of a limit by approaching a point on a surface in such a way. However, we *can* show that a limit does not exist if two limits from two different directions do not agree with each other.

Recall that differentiability implies continuity, but continuity does not imply differentiability.

When asked to define where a function is continuous, recall that polynomial and rational functions are continuous wherever they are defined. For those types of functions, it's simply a matter of finding their domain.

The limit of a function of two variables can be written as

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y),$$

which means

$$\lim_{x \rightarrow a} \left(\lim_{y \rightarrow b} f(x, y) \right) \quad \text{or} \quad \lim_{y \rightarrow b} \left(\lim_{x \rightarrow a} f(x, y) \right).$$

For a function of two variables to be continuous at a point (a, b) , it must satisfy the following conditions:

- $f(a, b)$ must be defined
- $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$ must “exist”
- $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b)$ for all (x, y) in the domain D

This definition of continuity is the same as for single variable functions, but for two variable functions, it is much more difficult to prove the second condition—that the limit exists.

There's not a nice L'Hopital's rule for multi-variate functions.

Direct Substitution

If a rational function is discontinuous at a single point, and the function has a limit at that point, then the function can be made continuous by defining that undefined point as the limit of the function. This only works for “holes,” it doesn't work for gap or jump discontinuities in which the limit does not exist.

For functions of two variables, just like with single variable functions, if you can plug the numbers in and evaluate, then do it.

Example:

Evaluate

$$\lim_{(x,y) \rightarrow (1,0)} \ln \left(\frac{1+x}{x^2+y^2} \right) = L.$$

Since the function is defined at $(1, 0)$, we can just plug this point in to get the limit

$$\begin{aligned} L &= \lim_{(x,y) \rightarrow (1,0)} \ln \left(\frac{1+(1)}{(1)^2 + (0)^2} \right) \\ &= \ln 2. \end{aligned}$$

Radial Limits

A **radial limit** is a limit approaching the point $(0, 0)$ along a straight line.

To show that a limit does not exist, approach the point from two different directions. If you get two different answers for the limit, then the limit does not exist.

Example:

Evaluate

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x-y}{x+y}.$$

Notice that the function is not defined at $(0, 0)$ so we cannot just plug in numbers. If we approach the point $(0, 0)$ along the x -axis, then y , will be zero, and we have that

$$\lim_{(x,0) \rightarrow (0,0)} \frac{x - (0)}{x + (0)} = 1.$$

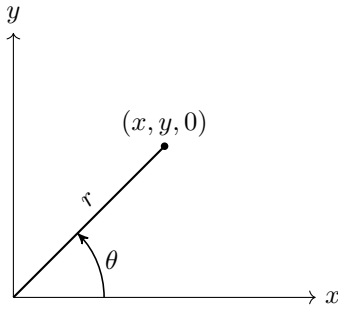
If we approach the origin along the y -axis, then x will be zero, and we get

$$\lim_{(0,y) \rightarrow (0,0)} \frac{(0) - y}{(0) + y} = -1.$$

Since we got two different results, the limit does not exist at that point. There are lots of different ways to approach a point in \mathbb{R}^3 . Notice that if we approached the origin along the plane $y = x$, the limit would be 0 (because of the numerator becoming zero).

Using Polar Coordinates

Sometimes, we can replace x and y by polar coordinates and take the limit of that to prove the existence of limits. Since θ is fixed, it can be any angle, that is, it can approach the point $(0, 0)$ in a straight line along an arbitrary angle.



Think of it this way—if we are in the xy -plane of \mathbb{R}^3 , we keep the angle θ fixed, and then we take the point (x, y) toward the origin. If θ is fixed, then $x = r \cos \theta$ and $y = r \sin \theta$ are also fixed. So as $(x, y) \rightarrow (0, 0)$, we have that $r \rightarrow 0^+$. Of course, this technique only works when the limit is at $(0, 0)$.

Example:

Switch to polar coordinates to evaluate the limit

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^4 + y^4}{x^2 + y^2}.$$

Substituting $x = r \cos \theta$ and $y = r \sin \theta$ and simplifying gives us

$$\lim_{r \rightarrow 0^+} r^2(\cos^4 \theta + \sin^4 \theta).$$

We can say that this limit is zero, but only because we know that the θ part ($\cos^4 \theta + \sin^4 \theta$) does not go to positive or negative infinity. We know that both sine and cosine vary between -1 and 1 and so does any power of them. So $0 \leq (\cos^4 \theta + \sin^4 \theta) \leq 2$.

So

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^4 + y^4}{x^2 + y^2} = 0.$$

The nice thing about this technique is that if all the θ terms cancel out and you're left with only r terms, then you're just dealing with a single variable limit anymore, and you can apply L'Hopital's rule and such.

Beware, switching to polar coordinates can only prove the existence of a limit if the function does not depend on θ (i.e. the θ parts cancel out) or if it can be shown that the θ part is bounded. For example, if the θ part can not be shown to be bounded, then you could have the case that the θ part goes to ∞ as $r \rightarrow 0^+$. Then you might have an indeterminate limit of the form $0 \cdot \infty$.

Squeeze Theorem

We can also use the squeeze theorem. Consider the limit

$$\lim_{(x,y) \rightarrow (0,0)} \frac{y^2 \sin^2 x}{x^2 + y^2}.$$

We know that x^2 and y^2 are both greater than or equal to zero since they are squared. This implies that $y^2 \leq x^2 + y^2$, which implies that

$$\frac{y^2}{x^2 + y^2} \leq 1.$$

Since we're effectively multiplying $\sin^2 x$ by a number between zero and one, we have that

$$0 \leq \frac{y^2}{x^2 + y^2} \sin^2 x \leq \sin^2 x,$$

and because the limit of the left and right are both zero, this implies that the limit of the middle is also zero.

1.9.3 Partial Derivatives

If a function f is a function of two variables $f(x, y)$ and we let only x vary while keeping y fixed, then we call it the **partial derivative** of f with respect to x . This can be denoted in several different ways:

$$\lim_{h \rightarrow 0} \frac{f(x+h, y) - f(x, y)}{h} = f_x(x, y) = f_x = \frac{\partial f}{\partial x} = D_x f$$

Similarly, the partial derivative of f with respect to y can be denoted

$$\lim_{h \rightarrow 0} \frac{f(x, y+h) - f(x, y)}{h} = f_y(x, y) = f_y = \frac{\partial f}{\partial y} = D_y f.$$

The ∂ symbols are called the **Jacobi d's**.

To find a partial derivative we simply differentiate with respect to the variable being varied and treat the fixed variable as a constant.

Example:

Find the partial derivative of $f(x, y) = x^2y$.

Since f is a function of two variables, it has two partial derivatives—one with respect to each variable. We find the partial derivative with respect to x by taking the regular derivative with respect to x while treating y as a constant,

$$\frac{\partial f}{\partial x} = f_x = 2xy.$$

Similarly, the partial with respect to y is

$$\frac{\partial f}{\partial y} = f_y = x^2.$$

To understand $f_x(a, b)$ geometrically, think of it this way. Remember that $f(x, y)$ is a surface in \mathbb{R}^3 . Slice this surface with a vertical plane that goes through (a, b) , and is parallel to the xy -axis since y is constant. Now, consider only the *curve* formed by the intersection of the surface and this plane. The value of $f_x(a, b)$ is the slope of this curve above the point (a, b) . Remember that (a, b)

is a point on the xy -plane—not on the surface, however, the curve we are considering, lies atop the surface.

Alternatively, to understand $f_x(a, b)$ geometrically, draw a line in the xy -plane parallel to the x -axis and passing through (a, b) . Now, project this line straight up or straight down onto the surface. The “shadow” of this line projected in this manner, is a curve on the surface. Then $f_x(a, b)$ is the slope of that curve at the point on the surface directly above or below (a, b) .

We can also take higher order partial derivatives. For example, the second partial derivative of f with respect to x can be denoted

$$f_{xx}(x, y) = f_{xx} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial x^2} = D_x^2 f.$$

We can even take **mixed partial derivatives**. If we take the derivative of f with respect to x and then take the derivative of that with respect to y , we could denote that

$$f_{xy}(x, y) = f_{xy} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial y \partial x}.$$

Notice the order of the partials in the denominator of the last notation! In general, the order of the x and y matters, no matter which notation is being used.

Clairaut’s theorem states that if f_{xy} and f_{yx} are continuous, then

$$f_{xy}(a, b) = f_{yx}(a, b).$$

If f_x and f_y are both continuous, then

$$f_{xy} = f_{yx},$$

if they are not continuous, then

$$f_{xy} \neq f_{yx}.$$

For example, consider the function $f(x, y) = e^x \sin y$. Since both f_x and f_y are continuous functions, by Clairaut’s theorem, the mixed partials are the same

$$f_{xy} = f_{yx} = e^x \cos y.$$

So if both first partial derivatives are continuous, then it is okay to switch the order when dealing with mixed partials.

If a function is symmetric in x and y , which means if you switch x and y , the function stays the same, then the partial derivatives show that same symmetry. Consider the function $f(x, y) = x^2 + xy + y^2$, which satisfies this symmetry. The partial derivative with respect to x is $f_x(x, y) = 2x + y$, and the partial derivative with respect to y is $f_y(x, y) = x + 2y$. Notice the symmetry between the two partial derivatives. You can get one from the other simply by switching x and y .

A **Laplace equation** is any partial differential equation that satisfies

$$z_{xx} + z_{yy} = \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0.$$

Any function z satisfying Laplace’s equation is called a **harmonic function**. There’s an area of physics called “potential theory” based almost entirely on this equation. There’s also an area in math called “harmonic analysis” about these functions. The equation can also be extended to higher dimensions as for the function $w(x, y, z)$

$$w_{xx} + w_{yy} + w_{zz} = 0.$$

There’s a similarity between conic sections of algebra and partial differential equations. Recall that the general equation for a conic section is

$$Ax^2 + Bxy + Cy^2 + Ex + Fy + G = 0.$$

By looking at the discriminant $B^2 - 4AC$, we can tell what the conic section will be. If the discriminant is less than zero, the result is an ellipse (or circle), if it’s equal to zero, the result is a parabola, and if it’s greater than zero, the result is a hyperbola. There’s an analogous partial differential equation, which has the form

$$Az_{xx} + Bz_{xy} + Cz_{yy} + Ez_x + Fz_y + G = 0.$$

If the discriminant $B^2 - 4AC < 0$, it’s called an elliptic PDE (partial differential equation), if $B^2 - 4AC = 0$, it’s called a parabolic PDE, and if $B^2 - 4AC > 0$, it’s called a hyperbolic PDE. Each different type has a slightly different solution.

The **heat conduction equation** is given as

$$\alpha^2 u_{xx} - u_t = 0.$$

Notice that since $A = \alpha^2$, $B = 0$, and $C = 0$, then $B^2 - 4AC = 0$, this is a parabolic PDE.

The **wave equation** is given as

$$a^2 u_{xx} - u_{tt} = 0.$$

Notice that since $A = a^2$, $B = 0$, and $C = -1$, then $B^2 - 4AC > 0$, this is a hyperbolic PDE. The wave equation describes the motion of a waveform.

To show that an equation satisfies the heat or wave equations, we simply take the relevant partial derivatives and plug them in. For example, to show that an equation satisfies the heat equation, we would take the second partial derivative with respect to x , the first partial derivative with respect to t , and then plug them into the heat equation to see if it evaluates to zero.

Tip:

When taking the partial derivatives of composed functions, don’t forget to apply the chain rule!

Recall that if the integrand of an integral is continuous, then the function set equal to that integral is differentiable. For example, if

$$f(x) = \int_4^x (t + 1)^{10} dt,$$

then $f(x)$ is differentiable. Recall from the first fundamental theorem of calculus, that

$$\frac{df}{dt} = f'(x) = (x+1)^{10}.$$

As another example:

$$\frac{d}{dt} \int_a^x \sqrt{t^3+1} dt = \sqrt{x^3+1}.$$

We can extend this same thinking to functions in two variables containing an integral. For example, for

$$F(x, y) = \int_y^x \cos(e^t) dt,$$

from the FFTC (first fundamental theorem of calculus), we have that

$$\frac{\partial F}{\partial x} = \cos(e^x).$$

To find the other partial derivative, we simply flip the integral

$$F(x, y) = \int_y^x \cos(e^t) dt = - \int_x^y \cos(e^t) dt,$$

so we can apply the FFTC to get

$$\frac{\partial F}{\partial y} = -\cos(e^y).$$

To find all the partial derivatives of a function in n variables for example $u = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ use i to indicate the variable under consideration. The function can be rewritten as

$$u = \sqrt{x_1^2 + x_2^2 + \dots + x_i^2 + \dots + x_n^2},$$

to indicate this. Then take the partial derivative with respect to x_i to get

$$u_{x_i} = \frac{x_i}{\sqrt{x_1^2 + x_2^2 + \dots + x_i^2 + \dots + x_n^2}}.$$

Notice that you have to apply the chain rule in this case.

Implicit Differentiation

We can also implicitly differentiate a multivariable function. For example

$$x^2 + 2y^2 + 3z^2 = 1,$$

can be implicitly differentiated. To find the partial derivative of z with respect to x , we get

$$2x + 6z \frac{\partial z}{\partial x} = 0.$$

Notice that the partial derivative of y^2 is zero since y is being treated as a constant. Solving for $\frac{\partial z}{\partial x}$ gives us

$$\frac{\partial z}{\partial x} = -\frac{x}{3z}.$$

1.9.4 Tangent Planes and Linear Approximation

Recall that, a line that is tangent to a function $f(x)$ at the point (x_0, y_0) can be represented as

$$y - y_0 = f'(x_0)(x - x_0),$$

in point-slope form. Notice that $f'(x_0)$ is the slope of the line and y_0 can be written as $f(x_0)$. Hence, this function can be rewritten as

$$y = f(x_0) + f'(x_0)(x - x_0).$$

Around the point (x_0, y_0) , this tangent line/function approximates the original function $f(x)$. We call this the **linear approximation** of $f(x)$.

In \mathbb{R}^3 an analogous idea can be developed. The problem is, our functions are now two variable functions, so instead of a tangent line, we have a tangent plane. Recall that the **equation of a plane** is

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

Rearranging, and letting $-\frac{A}{C} = a$ and $-\frac{B}{C} = b$, we get

$$z - z_0 = a(x - x_0) + b(y - y_0),$$

where $z_0 = f(x_0, y_0)$. Now, since a and b represent slopes in the x and y directions, respectively, it must be the case that $a = f_x(x_0, y_0)$ and $b = f_y(x_0, y_0)$. Notice that these are partial derivatives. Now we can write the equation of the tangent plane (and the linear approximation) of the function $z = f(x, y)$ at the point (x_0, y_0, z_0) as

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Here's an important theorem that can be used to show that a function is differentiable at a point: $f(x, y)$ is differentiable at (x_0, y_0) if both partial derivatives f_x and f_y are continuous at (x_0, y_0) .

Recall that if $y = f(x)$ is differentiable, that is, $\frac{dy}{dx} = y'$ exists, then its differential is given by $dy = f'(x) dx$. How do we find the differential for multivariate functions?

If $z = f(x, y)$ then the **total differential** is

$$dz = f_x(x, y)dx + f_y(x, y)dy.$$

This equation can be shortened by writing

$$dz = z_x dx + z_y dy.$$

1.9.5 The Chain Rule

Recall that if $y = f(x)$ and $x = x(t)$, then y is indirectly a function of t , and we can talk about $\frac{dy}{dt}$. For such a composed function, the chain rule states that

$$\frac{dy}{dx} = \frac{dy}{dt} \cdot \frac{dt}{dx}.$$

For multivariate functions, there are several different cases we have to consider.

Case 1

In the first case, we have $z = f(x, y)$ and $x = g(t)$ and $y = h(t)$. Then z is indirectly a single variable function that depends only on t . In other words, z is some unknown function of t . There are two different ways of finding $\frac{dz}{dt}$. The first way is to use the chain rule specific to this case, which is

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial z}{\partial y} \cdot \frac{dy}{dt}.$$

Notice that this chain rule contains both partial derivatives (because z is two variable function of x and y) and regular derivatives (because x and y are single variable functions of t).

The first way of finding $\frac{dz}{dt}$ is to calculate the two partial derivatives and the two regular derivatives of $z = f(x, y)$ and plug them into the formula above. The second way is to replace x and y in $z = f(x, y)$ by their functions of t . Now, z is only a function of t and its derivative $\frac{dz}{dt}$ can be computed directly.

Example:

Given

$$z = x^2 + xy,$$

where $x = e^t$ and $y = e^{-t}$, compute the derivative $\frac{dz}{dt}$.

Using the first method, we compute the two partial derivatives and the two regular derivatives

$$\begin{aligned}\frac{\partial z}{\partial x} &= 2x + y \\ \frac{\partial z}{\partial y} &= x \\ \frac{dx}{dt} &= e^t \\ \frac{dy}{dt} &= -e^{-t}.\end{aligned}$$

Plugging them into the formula given above gives us

$$\frac{dz}{dt} = (2x + y)e^t - xe^{-t}.$$

Using the second method, we replace x and y in z by their respective functions of t and then take the derivative with respect to t

$$\begin{aligned}z &= x^2 + xy \\ z &= (e^t)^2 + e^t e^{-t} \\ z &= e^{2t} + 1 \\ \frac{dz}{dt} &= 2e^{2t}.\end{aligned}$$

These two results appear different, but if we replace x and y in the first with their corresponding functions of t , we see that it simplifies to the second result.

Case 2

The second case of the multivariate chain rule occurs when x and y no longer depend on a single variable. For example, when $z = f(x, y)$ and $x = g(s, t)$ and $y = h(s, t)$, this means that z is not indirectly a function of just t , but rather, it is indirectly a function of s and t . The chain rule, then, becomes two separate partial derivatives

$$\begin{aligned}\frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial s}, \\ \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t}.\end{aligned}$$

Example:

For example, if we have $z = x^2y$ and $x = s^2e^t$ and $y = e^{st}$, then the relevant partial derivatives are

$$\begin{aligned}\frac{\partial z}{\partial x} &= 2xy \\ \frac{\partial z}{\partial y} &= x^2 \\ \frac{\partial x}{\partial s} &= 2se^t \\ \frac{\partial x}{\partial t} &= s^2e^t \\ \frac{\partial y}{\partial s} &= te^{st} \\ \frac{\partial y}{\partial t} &= se^{st}.\end{aligned}$$

Putting it all together, we get

$$\begin{aligned}\frac{\partial z}{\partial s} &= 2xy \cdot 2se^t + x^2 \cdot te^{st}, \\ \frac{\partial z}{\partial t} &= 2xy \cdot s^2e^t + x^2 \cdot se^{st}.\end{aligned}$$

Just as with the other case, you could replace x and y in the original function with the functions containing s and t , and partially differentiate it. This is a good way to check your work.

This concept can also be expanded to functions of more than two variables, whether it's z that contains more variables or whether it's the variables in z that contains more variables. For example, if you have $z = f(x, y)$ where $x = g(s, t, u)$ and $y = h(s, t, u)$, then your relevant partial derivatives become

$$\begin{aligned}\frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial s}, \\ \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t}, \\ \frac{\partial z}{\partial u} &= \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial u} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial u}.\end{aligned}$$

On the other hand, if you have $w = f(x, y, z)$, where $x = g(s, t)$, $y = h(s, t)$, and $z = l(s, t)$, then your relevant

partial derivatives become

$$\begin{aligned}\frac{\partial w}{\partial s} &= \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial s} + \frac{\partial w}{\partial z} \cdot \frac{\partial z}{\partial s}, \\ \frac{\partial w}{\partial t} &= \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial t} + \frac{\partial w}{\partial z} \cdot \frac{\partial z}{\partial t}.\end{aligned}$$

2nd Partial Derivatives by Product and Chain Rules

If you have a function $z = f(x, y)$ where $x = g(s, t)$ is a function of s and t and $y = h(s, t)$ is also a function of x

and t , how do you calculate the second partial derivatives with respect to s and t ?

Let's calculate the second partial derivative of z with respect to s . The first partial derivative, by the chain rule, is

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial s}.$$

The second partial derivative, then, is

$$\frac{\partial^2 z}{\partial s^2} = \frac{\partial}{\partial s} \left(\frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial s} \right).$$

Applying the product rule to each of the two products, we get

$$\frac{\partial^2 z}{\partial s^2} = \frac{\partial z}{\partial x} \cdot \frac{\partial}{\partial s} \left(\frac{\partial x}{\partial s} \right) + \frac{\partial x}{\partial s} \cdot \frac{\partial}{\partial s} \left(\frac{\partial z}{\partial x} \right) + \frac{\partial z}{\partial y} \cdot \frac{\partial}{\partial s} \left(\frac{\partial y}{\partial s} \right) + \frac{\partial y}{\partial s} \cdot \frac{\partial}{\partial s} \left(\frac{\partial z}{\partial y} \right).$$

The first and third terms simplify because

$$\begin{aligned}\frac{\partial}{\partial s} \left(\frac{\partial x}{\partial s} \right) &= \frac{\partial^2 x}{\partial s^2}, \\ \frac{\partial}{\partial s} \left(\frac{\partial y}{\partial s} \right) &= \frac{\partial^2 y}{\partial s^2},\end{aligned}$$

so we get

$$\frac{\partial^2 z}{\partial s^2} = \frac{\partial z}{\partial x} \cdot \frac{\partial^2 x}{\partial s^2} + \frac{\partial x}{\partial s} \cdot \frac{\partial}{\partial s} \left(\frac{\partial z}{\partial x} \right) + \frac{\partial z}{\partial y} \cdot \frac{\partial^2 y}{\partial s^2} + \frac{\partial y}{\partial s} \cdot \frac{\partial}{\partial s} \left(\frac{\partial z}{\partial y} \right).$$

The second and last terms require an additional application of the chain rule. To keep it straight, just remember that

$$\frac{\partial \square}{\partial s} = \frac{\partial \square}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial \square}{\partial y} \cdot \frac{\partial y}{\partial s}$$

where \square is some function of x and y . Note that the derivative of a function of x and y is still a function of x and y . Now our second derivative simplifies to

$$\frac{\partial^2 z}{\partial s^2} = \frac{\partial z}{\partial x} \cdot \frac{\partial^2 x}{\partial s^2} + \frac{\partial x}{\partial s} \cdot \left[\frac{\partial}{\partial x} \left(\frac{\partial z}{\partial x} \right) \cdot \frac{\partial x}{\partial s} + \frac{\partial}{\partial y} \left(\frac{\partial z}{\partial x} \right) \cdot \frac{\partial y}{\partial s} \right] + \frac{\partial z}{\partial y} \cdot \frac{\partial^2 y}{\partial s^2} + \frac{\partial y}{\partial s} \cdot \left[\frac{\partial}{\partial x} \left(\frac{\partial z}{\partial y} \right) \cdot \frac{\partial x}{\partial s} + \frac{\partial}{\partial y} \left(\frac{\partial z}{\partial y} \right) \cdot \frac{\partial y}{\partial s} \right],$$

which simplifies further to

$$\frac{\partial^2 z}{\partial s^2} = \frac{\partial z}{\partial x} \cdot \frac{\partial^2 x}{\partial s^2} + \frac{\partial x}{\partial s} \cdot \left[\frac{\partial^2 z}{\partial x^2} \cdot \frac{\partial x}{\partial s} + \frac{\partial^2 z}{\partial y \partial x} \cdot \frac{\partial y}{\partial s} \right] + \frac{\partial z}{\partial y} \cdot \frac{\partial^2 y}{\partial s^2} + \frac{\partial y}{\partial s} \cdot \left[\frac{\partial^2 z}{\partial x \partial y} \cdot \frac{\partial x}{\partial s} + \frac{\partial^2 z}{\partial y^2} \cdot \frac{\partial y}{\partial s} \right].$$

The second derivative of z with respect to r is obtained in the same manner.

Implicit Differentiation

Recall that if y is a function of x , that is, if $y = f(x)$, then $x^2 + y^2 = 9$ is an example of an implicit equation since it is not solved for the dependent variable. Implicit differentiation is used when it's difficult to solve an equation explicitly for the dependent variable to make an explicit function. To differentiate such an equation explicitly we differentiate both sides with respect to the independent

variable to get

$$\begin{aligned}\frac{d}{dx} (x^2 + y^2) &= \frac{d}{dx} (9) \\ \frac{d}{dx} (x^2) + \frac{d}{dx} (y^2) &= 0 \\ 2x + 2y \frac{dy}{dx} &= 0\end{aligned}$$

The only tricky part of it is $\frac{d}{dx}(y^2)$, which is differentiated using the chain rule

$$\frac{d}{dx}(y^2) = \frac{d(y^2)}{dy} \cdot \frac{dy}{dx} = 2y \frac{dy}{dx}.$$

Finally, we can solve for $\frac{dy}{dx}$ to get

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

With our knowledge of partial derivatives and the chain rule, we can now give a better method for differentiating such an implicit function. Supposing that we have an equation $F(x, y) = 0$ that defines y implicitly as a differentiable function of x , that is, $y = f(x)$, where $F(x, f(x)) = 0$, then applying the chain rule, to differentiate both sides of $F(x, y) = 0$, we get

$$\frac{\partial F}{\partial x} \cdot \frac{\partial x}{\partial x} + \frac{\partial F}{\partial y} \cdot \frac{\partial y}{\partial x} = 0.$$

Notice that our implicit equation $x^2 + y^2 = 9$ fits these conditions as it can be rewritten as $F(x, y) = x^2 + y^2 - 9 = 0$. Solving the above equation for $\frac{dy}{dx}$, we get

$$\frac{dy}{dx} = -\frac{F_x}{F_y}.$$

Considering our implicit equation again, we see that we get the same answer using this method

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

For functions of two or more variables, implicit differentiation is much the same. For example, with an implicitly defined function of z that depends on x and y , that is, $z = f(x, y)$, we don't know exactly how z is a function of x and y , so we use implicit differentiation to find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$. Using a similar process as the one shown above for $F(x, y) = 0$, we can find similar formulas for $F(x, y, z) = 0$ where $z = f(x, y)$:

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z} \quad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}$$

1.9.6 Directional Derivatives

In \mathbb{R}^3 , $z = f(x, y)$ forms a surface and the partial derivative gives the rate of change of the surface along the x or y coordinate. What if we want the rate of change along some other direction given as a unit vector? In other words, what is the derivative along the unit vector \hat{u} ? It is given by

$$D_{\hat{u}}f(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h},$$

where $\hat{u} = \langle a, b \rangle$ is a unit vector. If it is not a unit vector then it must be normalized first. In the end $D_{\hat{u}}f(x_0, y_0)$ yields a plain number. Using the limit method shown here is not a practical method for calculating the directional derivative.

Note that the direction vector does not just point in any direction—it is in (or parallel to) the xy -plane.

If $z = f(x, y)$ is a function with partial derivatives f_x and f_y and $\vec{u} = \langle a, b \rangle$ is a unit vector, then we can calculate the directional derivative as

$$D_{\vec{u}}f(x_0, y_0) = f_x(x, y)a + f_y(x, y)b.$$

Using this method for calculating the directional derivative is much more practical. Since \hat{u} is a unit vector, we can write it as $\hat{u} = \langle \cos \theta, \sin \theta \rangle$ where θ is the angle with respect to the positive x -axis of the direction of \hat{u} . So to calculate the directional derivative at the point (x, y) in the direction θ , we can use

$$D_{\hat{u}}f(x_0, y_0) = f_x(x, y) \cos \theta + f_y(x, y) \sin \theta.$$

Example:

Calculate the directional derivative at the point $(2, 3)$ in the direction $\theta = \frac{\pi}{4}$ counterclockwise from the positive x -axis of the function $f(x, y) = x^2 + 2xy + 3y^2$.

Taking the relevant partial derivatives, we get

$$\begin{aligned} f_x &= 2x + 2y \\ f_y &= 2x + 6y. \end{aligned}$$

Plugging in the point $(2, 3)$, we get $f_x(2, 3) = 10$ and $f_y(2, 3) = 22$, so our directional derivative is

$$\begin{aligned} D_{\hat{u}}f(2, 3) &= 10 \cos \frac{\pi}{4} + 22 \sin \frac{\pi}{4} \\ &= 16\sqrt{2}. \end{aligned}$$

This is the rate of change in the direction of the function at $f(x) = x^2 + 2xy + 3y^2$ at the point $(2, 3)$ along the direction $\theta = \frac{\pi}{2}$.

Gradient

If $z = f(x, y)$ is a function with derivatives f_x and f_y , then its **gradient** is given by

$$\text{grad}f(x, y) = \nabla f(x, y) = \langle f_x(x, y), f_y(x, y) \rangle.$$

The quantity $\nabla f(x, y)$ is read as “the gradient of f ”. The upside down capital delta is called “nabla” or “del”. It is a *vector operator*, and it can be represented

$$\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle.$$

Notice that

$$\begin{aligned} D_{\hat{u}}f(x_0, y_0) &= f_x(x, y)a + f_y(x, y)b \\ &= \nabla f(x, y) \bullet \hat{u}. \end{aligned}$$

This gives us a more concise definition of the **directional derivative**: The directional derivative of f in the direction \hat{u} is the dot product of the gradient of f with \hat{u} .

Tip:

Remember that the gradient of a scalar function f is a vector field. The gradient evaluated at a specific point is a vector.

Another useful bit of notation is

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Notice that this is just the dot product of the gradient operator with itself. This notation allows us to write the short form of the **Laplace equation** as

$$\nabla^2 \bullet \vec{u} = 0.$$

Expanding the del operator to include t is called the **D'Alembert operator** and is defined as

$$\square^2 = \nabla^2 + \frac{\partial^2}{\partial t^2}.$$

When will the directional derivative be at a maximum? Using

$$D_u f(x, y) = \nabla f(x, y) \bullet \hat{u},$$

and applying one definition of the dot product, we have that

$$\nabla f \bullet \hat{u} = |\nabla f| |\hat{u}| \cos \theta.$$

We know that $|\hat{u}| = 1$ because we've defined it as a unit vector. We also know that $\cos \theta$ is greatest at 1, which occurs when $\theta = 0$ and least at -1 when $\theta = \pi$. This means $D_u f(x, y)$ is at a maximum when the gradient is the same direction as \vec{u} and at a minimum when they are in opposite directions.

Example:

Find the maximum rate of change of $f(x, y) = 3xy^2$ at $(x, y) = (1, 2)$ and the direction in which it occurs.

We know the maximum of the directional derivative occurs when the direction equals the direction of the gradient vector, and its value is the magnitude of the gradient vector. Taking the gradient, we get

$$\begin{aligned} \nabla f &= \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \\ &= \langle 3y^2, 6xy \rangle. \end{aligned}$$

Evaluating the gradient at $(1, 2)$ gives us $\nabla f(1, 2) = \langle 12, 12 \rangle$. Its magnitude is $\sqrt{12^2 + 12^2} = 12\sqrt{2}$.

So the maximum directional derivative at $(1, 2)$ is $12\sqrt{2}$, and it is in the direction $\langle 12, 12 \rangle$.

A **tangent plane** is to a surface as a tangent line is to a curve. A function of the form $f(x, y, z) = c$ describes

a level *surface* of a function of three variables. To find the tangent plane to such a surface at the point (x_0, y_0, z_0) we use

$$\begin{aligned} f_x(x_0, y_0, z_0)(x - x_0) &+ f_y(x_0, y_0, z_0)(y - y_0) \\ &+ f_z(x_0, y_0, z_0)(z - z_0) = 0. \end{aligned}$$

In the case of a function of two variables of the form $z = f(x, y)$, this equation simplifies to

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) - (z - z_0) = 0.$$

The **normal line** is the line perpendicular to a surface (and therefore its tangent plane) at some point. The normal line will go through the given point, and it will be perpendicular to the tangent plane. The gradient vector is perpendicular to the level curves of a function, so for a level surface of a function of three variables, the gradient ∇f gives us the direction of the normal vector. The symmetric equations for our normal line are given by

$$\frac{x - x_0}{f_x(x_0, y_0, z_0)} = \frac{y - y_0}{f_y(x_0, y_0, z_0)} = \frac{z - z_0}{f_z(x_0, y_0, z_0)}.$$

Example:

Find the tangent plane for the level surface $f(x, y, z) = x^2yz = 6$ at the point $(1, 2, 3)$.

We start by finding the three partial derivatives and evaluating them at the point.

$$\begin{aligned} f_x(x, y, z) &= 2xyz & f_x(3, 2, 1) &= 12 \\ f_y(x, y, z) &= x^2z & f_y(3, 2, 1) &= 3 \\ f_z(x, y, z) &= x^2y & f_z(3, 2, 1) &= 2. \end{aligned}$$

Plugging these numbers and the point into the equation given above and simplifying, gives us the equation for the tangent plane $12x + 3y + 2z = 24$.

When finding the directional derivative at a specific point, it usually saves time to plug the point in after finding the gradient and before taking its dot product with the direction vector.

For two functions of two variables $u = u(x, y)$, $v = v(x, y)$, and $a, b \in \mathbb{R}$, prove that

$$\nabla(au + bv) = a\nabla u + b\nabla v.$$

We have that

$$\begin{aligned} \nabla(au + bv) &= \left\langle \frac{\partial}{\partial x}(au + bv), \frac{\partial}{\partial y}(au + bv) \right\rangle \\ &= \langle au_x + bv_x, au_y + bv_y \rangle \\ &= \langle au_x, au_y \rangle + \langle bv_x, bv_y \rangle \\ &= a\langle u_x, u_y \rangle + b\langle v_x, v_y \rangle \\ &= a\nabla u + b\nabla v \end{aligned}$$

This shows that the gradient is a **linear operator** meaning that the operation done on a sum is the sum of the operations done on the terms. The derivative is also a linear operator.

For two functions of two variables $u = u(x, y)$, $v = v(x, y)$, and $a, b \in \mathbb{R}$, prove that

$$\nabla(uv) = v\nabla u + u\nabla v.$$

We have that

$$\begin{aligned} \nabla(uv) &= \left\langle \frac{\partial}{\partial x}(uv), \frac{\partial}{\partial y}(uv) \right\rangle \\ &= \left\langle u_x v + uv_x, u_y v + uv_y \right\rangle \\ &= \langle u_x v, u_y v \rangle + \langle v_x u + v_y u \rangle \\ &= v \langle u_x, u_y \rangle + u \langle v_x, v_y \rangle \\ &= v\nabla u + u\nabla v \end{aligned}$$

1.9.7 Maximum and Minimum Values

Recall that with $y = f(x)$, the maximum and minimum can be determined by looking at where the derivative is zero. Anywhere that the derivative is zero or undefined is a critical point, and these can be maximums, minimums, or inflection points. There are analogous ideas for functions of more variables, although it's a little more complicated and involves gradients.

We say that the point $f(a, b)$ on a curve is a local maximum if $f(x, y) \leq f(a, b)$ in the neighborhood of (a, b) . Similarly, $f(a, b)$ is a local minimum if $f(x, y) \geq f(a, b)$ in the neighborhood of (a, b) . If a function has a minimum or a maximum at a point (a, b) , then it is the case that

$$f_x(a, b) = 0 \quad \text{and} \quad f_y(a, b) = 0$$

This is a necessary, but not sufficient condition, because this does not tell us whether such a *critical point* is a maximum or a minimum, and not all points satisfying this condition will be a maximum or a minimum. What we need is something analogous to the second derivative test for functions of a single variable.

The test for the minimum or maximum that we use here is only valid for functions of two variables. There are similar tests for functions of more variables, but they

are a little more complicated and aren't taught in basic calculus.

We're given a function $f(x, y)$ that has the derivatives $f_x, f_y, f_{xy} = f_{yx}, f_{xx}, f_{yy}$. First we solve the system

$$\begin{cases} f_x(x, y) = 0 \\ f_y(x, y) = 0 \end{cases}$$

If it turns out that the two equations are the same, then there are infinitely many critical points—an entire plane of solutions. Usually, the systems arrived at will not be nice—they won't be linear, and we won't be able to solve them using matrices or simple substitution. We have to use a little cleverness. However, we can always solve for x in one equation and plug it into the other, but it usually results in a nasty computation.

From this, we get the critical point (a, b) . We plug this critical point into the formula

$$D(a, b) = f_{xx}(a, b)f_{yy}(a, b) - f_{xy}^2(a, b),$$

which comes from the **Hessian determinant**

$$D(a, b) = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}.$$

The $D(a, b)$ test gives a plain number at the end. There are four possible cases:

- If $D > 0$ and $f_{xx}(a, b) > 0$, then we have a local minimum at $(a, b, f(a, b))$.
- If $D > 0$ and $f_{xx}(a, b) < 0$, then we have a local maximum at $(a, b, f(a, b))$.
- If $D < 0$ then $(a, b, f(a, b))$ is neither a maximum nor a minimum. It is rather, a saddle point, which is analogous to an inflection point in \mathbb{R}^2 . Along one direction, it acts as a local maximum, and along another direction, it acts as a local minimum.
- If $D = 0$, the test fails—it tells us nothing.

Example:

Find and characterize the critical values of

$$f(x, y) = x^2y - 10x^2 - 10y^2.$$

The two partial derivatives are

$$\begin{aligned} f_x &= 2xy - 20x \\ f_y &= x^2 - 20y. \end{aligned}$$

Next, we set the two equations to zero and solve for the intersection.

$$\begin{cases} 2xy - 20x = 0 \\ x^2 - 20y = 0 \end{cases}$$

If these functions were nice, we could solve them as a system of equations, but it's not a linear system, so

we'll have to solve it a different way. Notice that we can factor the first function to get $2x(y - 10) = 0$ which tells us that $x = 0$ or $y = 10$. Plugging $x = 0$ into the second equation gives us $y = 0$, so $(0, 0)$ is a critical point. Plugging $y = 10$ into the second equation gives us $x = \pm 10\sqrt{2}$. This gives us two more critical points— $(10\sqrt{2}, 10)$ and $(-10\sqrt{2}, 10)$.

When we test our three candidate critical points in $D(a, b)$, we find that

$$\begin{aligned} D(0, 0) &\implies \text{maximum} \\ D(10\sqrt{2}, 10) &\implies \text{saddle point} \\ D(-10\sqrt{2}, 10) &\implies \text{saddle point.} \end{aligned}$$

The tricky part is finding all of the critical points and excluding all the points that are not critical points. The safest way to do that is to follow this algorithm: 1) Find the values for x and y for which the first equation will be zero, 2) Then, for each potential value of x or y , what must the opposite variable be in the *second* equation in order for that equation to equal zero? Consider the following system on the region $-\pi < x, y < \pi$

$$\begin{cases} \cos x \sin y = 0 \\ \sin x \cos y = 0 \end{cases}$$

Looking at the first equation, we see that it is 0 when $x = \pm \frac{\pi}{2}$ and when $y = 0$. Now we turn to the *second* equation, asking

- If $x = \frac{\pi}{2}$, then what value of y makes the second equation zero? The answer is $y = \pm \frac{\pi}{2}$.
- If $x = -\frac{\pi}{2}$, then what value of y makes the second equation zero? The answer is $y = \pm \frac{\pi}{2}$.

- If $y = 0$, then what value of x makes the second equation zero? The answer is $y = 0$.

Putting all of these together, we have that our critical points are $(0, 0)$, $(\frac{\pi}{2}, \frac{\pi}{2})$, $(-\frac{\pi}{2}, \frac{\pi}{2})$, $(\frac{\pi}{2}, -\frac{\pi}{2})$, and $(-\frac{\pi}{2}, -\frac{\pi}{2})$. If you don't follow this algorithm, you tend to find points that aren't actually critical points, miss some of the critical points, or both.

Absolute Minimums and Maximums

If $f(x, y)$ is continuous over a closed and bounded region R , then it has an absolute maximum and an absolute minimum at some points in R . To find the absolute maximums and minimums

1. Find the values of f at the critical points of f in R .
2. Find the extreme values of f on the boundary of R .
3. The largest of these is the absolute maximum and the smallest is the absolute minimum.

Example:

Find the absolute minimum and maximum of

$$f(x, y) = x^2y - 10x^2 - 10y^2,$$

on the region $-10 \leq x \leq 10$ and $-10 \leq y \leq 10$.

In the previous example, we found the three critical points $(0, 0)$ and $(\pm 10\sqrt{2}, 10)$, but we didn't evaluate the function at those points. We also need to find the extreme values along the boundary. Notice that our region is a 20×20 square, and we can consider each side of the square individually.

Along the first line, we have that $y = 10$ and $-10 \leq x \leq 10$. Plugging this in, our function simplifies to

$$f(x, 10) = -1000,$$

so along this border of our region, the function has a constant value of -1000 .

Along the second line, we have that $y = -10$ and $-10 \leq x \leq 10$. Plugging this in, our function simplifies to

$$f(x, -10) = -20x^2 - 1000.$$

This is a parabola centered at $x = 0$, so its extreme values occur at the middle $(0, -10)$ and the two endpoints $(\pm 10, -10)$.

Along the third line, we have that $x = 10$ and $-10 \leq y \leq 10$. Plugging this in, our function simplifies to

$$f(10, y) = 100y - 1000 - 10y^2.$$

Completing the square, we can put this in the more convenient vertex form as

$$f(10, y) = -10(y - 5)^2 - 750.$$

So along the line $x = 10$, the function is a quadratic centered at $y = 5$. So its extreme values occur at the vertex $(10, 5)$ and its endpoints $(10, \pm 10)$.

Along the final line, we have that $x = -10$ and $-10 \leq y \leq 10$. Plugging this in, our function simplifies again to

$$f(-10, y) = 100y - 1000 - 10y^2.$$

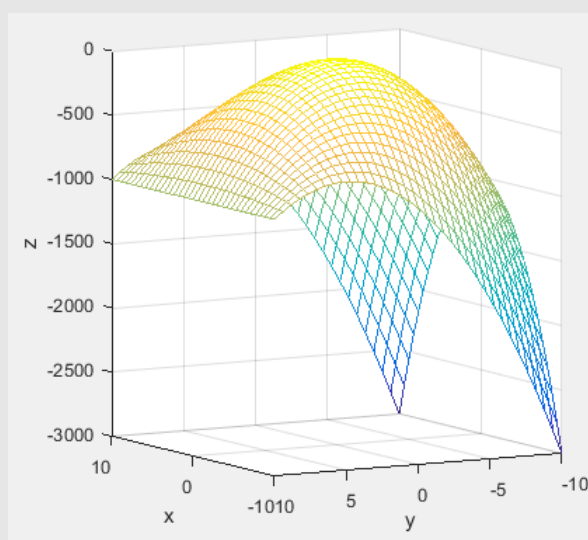
This is the same as for the previous line except we're now along the line $x = -10$. So the extreme values along this line occur at $(-10, 5)$ and $(-10, \pm 10)$.

Evaluating all of these points, we get

$$\begin{aligned} f(0, 0) &= 0 \\ f(x, 10) &= -1000 \\ f(0, -10) &= -1000 \\ f(10, -10) &= -3000 \\ f(-10, -10) &= -3000 \\ f(10, 5) &= -750 \\ f(-10, 5) &= -750. \end{aligned}$$

We can ignore some of the points, for example, $(\pm 10\sqrt{2}, 10)$ because we know that along the line $y = 10$, the function has a constant value of -1000 . The evaluations above tell us that the absolute maximum of the function in the given region is $f(0, 0) = 0$ and the absolute minimum is $f(10, -10) = f(-10, -10) = -3000$.

A surface plot of the function on the specified region is shown below. Notice that one border is a straight line, and the other three are quadratics.



Given a plane in \mathbb{R}^3 , such as $2x + 3y + z = 1$, and a point (x_0, y_0, z_0) , how do you calculate the shortest distance from the point to the plane? The distance function in 3 dimensions is

$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

But we know from the equation of the plane that $z = -2x - 3y + 1$, so we can rewrite our distance function as

$$d = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (-2x - 3y + 1 - z_0)^2}.$$

To find the shortest distance from the point to the plane, we minimize this quantity. That is, we find the partial

derivatives, set them equal to zero, and solve the system of equations to find the critical points. Minimizing this function containing a huge square root, would be messy, but squaring both sides of the function preserves everything that is relevant. If d is minimized, then d^2 will also be minimized. So, we minimize this function, instead

$$d^2 = (x - x_0)^2 + (y - y_0)^2 + (-2x - 3y + 1 - z_0)^2.$$

Finding the distance between points and a line is used when finding the **least squares approximation**. The best-fitting line for a collection of points minimizes the sum of the distances between the points and the lines.

This is a terrible calculation to do by hand, by computers can do it quickly.

1.9.8 Lagrange Multipliers

Let's say you're asked to build the largest rectangle in the upper half of the Cartesian plane such that two corners lie on the x -axis, and two corners lie on a semicircle of specified radius. This is what is called a maximization problem with a condition or constraint. In this case, the function being maximized is the area of the rectangle

Linear programming involves maximizing or minimizing a function $z = f(x, y)$ subject to certain conditions. For example, those conditions might be

$$\begin{aligned}x + y &\leq 1 \\x - y &\leq \frac{1}{2} \\x &\geq 0 \\y &\geq 0\end{aligned}$$

If we plot these equations on the xy -plane, we get what is called a **feasible region**. The maximum of z must occur at one of the corner points of the feasible region.

Lagrange multipliers act like catalysts in the sense that they help us solve these maximization problems, but in the process, they disappear. We don't even have to know what they are. Consider that we are given a function to maximize $z = f(x, y)$ subject to the condition $g(x, y) = c$. In order to maximize f subject to g , their gradients must be parallel. This means that the gradient of one is a scalar multiple of another, which allows us to write

$$\nabla f = \lambda \nabla g,$$

where λ , called the "Lagrange multiplier" is the scalar factor that equates the two gradients. In other words, to maximize f subject to g , we solve the system

$$\begin{cases} \nabla f = \lambda \nabla g \\ g(x, y) = c \end{cases}$$

Since $\langle f_x, f_y \rangle = \lambda \langle g_x, g_y \rangle$ subject to $g(x, y) = c$, we have three unknowns and three equations

$$\begin{aligned}f_x &= \lambda g_x \\f_y &= \lambda g_y \\g(x, y) &= c\end{aligned}$$

The method used in the previous section only works for two variable functions, but this method can be used with any number of variables.

Example:

Maximize the function $f(x, y) = x^2 + xy + y^2$ subject to the condition $g(x, y) = xy = 1$.

To maximize f subject to g means that we want to maximize $x^2 + xy + y^2$ given the relationship $xy = 1$ between x and y .

Since the maximum and/or minimum occurs at $\nabla f = \lambda \nabla g$, we first calculate the gradients of f and g to get

$$\begin{aligned}\nabla f &= \langle 2x + y, 2y + x \rangle \\ \nabla g &= \langle y, x \rangle\end{aligned}$$

Since $\nabla f = \lambda \nabla g$ and $g(x, y) = xy = 1$, we have the system of equations

$$\begin{cases} 2x + y = \lambda y \\ 2y + x = \lambda x \\ xy = 1 \end{cases}$$

Solving for λ and setting the results equal to each other, we get

$$\frac{2x}{y} + 1 = \frac{2y}{x} + 1.$$

Subtracting 1 from both sides then dividing both sides by 2 and simplifying gives us

$$y = \pm x.$$

From our constraint $xy = 1$, we know that $y = \frac{1}{x}$. Plugging this in and simplifying gives us $x = \pm 1$, which implies that $y = \pm 1$. Notice that the signs must be the same for x and y . So we get the points $(1, 1)$ and $(-1, -1)$. Plugging these into f gives us

$$\begin{aligned}f(1, 1) &= 3 \\ f(-1, -1) &= 3.\end{aligned}$$

So the maximum of minimum of f constrained by g is 3 and it occurs at $(1, 1)$ and $(-1, -1)$. To find out whether it's a maximum or minimum, we have to plug in some other value for f that *also* satisfies $xy = 1$. An example is the point $(2, \frac{1}{2})$. Plugging this into f gives us 5.25. Since this value is larger than 3, the points $(1, 1)$ and $(-1, -1)$ must give the minimum of f subject to g .

Example:

Suppose you have 10 square meters of cardboard from which you want to construct a rectangular box (with no top) whose volume is maximized.

If we consider x and y to be the sides of the box and z the top, then we want to maximize $V = f(x, y, z) = xyz$. We are constrained to a surface area of $g(x, y, z) = 2xz + xy + 2yz = 10$ square meters.

When V is at maximum or minimum, $\nabla f = \lambda \nabla g$. Find the gradients, we get

$$\begin{aligned}\nabla f &= \langle yz, xz, xy \rangle \\ \nabla g &= \langle 2z + y, 2z + x, 2y + 2x \rangle\end{aligned}$$

So, we have the system of equations

$$\begin{aligned}yz &= \lambda(2z + y) \\ xz &= \lambda(2z + x) \\ xy &= \lambda(2y + 2x) \\ 2yz + 2xz + xy &= 10.\end{aligned}$$

We have 4 unknowns and 4 equations. Solving the first three for λ and setting them equal to each other, we have

$$\frac{yz}{2z + y} = \frac{xz}{2z + x} = \frac{xy}{2x + 2y}.$$

Equating the first two and solving for x gives us $x = y$. Equating the first and third one and solving for x gives us $x = 2z$. Finally, equating the last two and solving for y gives us $y = 2z$. So we have that

$$x = y = 2z.$$

Using $g(x, y, z) = 2yz + 2xz + xy = 10$, we find that

$$\begin{aligned}x &= \frac{\sqrt{30}}{3} \\ y &= \frac{\sqrt{30}}{3} \\ z &= \frac{\sqrt{30}}{6}.\end{aligned}$$

So the volume of our box will be

$$\frac{\sqrt{30}}{3} \cdot \frac{\sqrt{30}}{3} \cdot \frac{\sqrt{30}}{6} = \frac{5}{9}\sqrt{30}.$$

In the end, we could figure out what the Lagrange multiplier λ is, but there is no need to. It acts as a catalyst that disappears after being used.

In the example above, our system yielded a single result. Sometimes, they yield multiple solutions, in which case the largest one is the maximum and the smallest one is the minimum.

What if your function is subject to more than one condition? What if you're asked to maximize $f(x, y, z)$

subject to $g(x, y, z) = c_1$ and $h(x, y, z) = c_2$? In such a case, we would need to solve the system

$$\begin{aligned}\nabla f &= \lambda_1 \nabla g + \lambda_2 \nabla h \\ g(x, y, z) &= c_1 \\ h(x, y, z) &= c_2\end{aligned}$$

1.10 Multiple Integrals

We can also do multiple integrals of the form

$$\int \int f(x, y) dx dy$$

$$\int \int \int f(x, y, z) dx dy dz$$

These “composed” integrals we just solve from the inside out. When the limits of the integrals are all numbers such as

$$\int_6^7 \int_4^5 \int_0^1 f(x, y, z) dx dy dz,$$

they are really easy to integrate. It becomes a bit more difficult when the limits are functions, such as

$$\int_6^7 \int_{z^2}^{\cos z} \int_{y+z}^{y-z} f(x, y, z) dx dy dz.$$

The outermost integral will always be from a number to a number because it can't refer back to a variable that has already been taken care of.

1.10.1 Double Integrals over Rectangles

Recall that definite integration of a single variable function is defined by the Riemann sum

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*) \Delta x.$$

Geometrically, single variable definite integration can be thought of as area under the curve, and it can be thought of as being the sum of the areas of an infinite number of rectangles.

Note that single-variable integration integrates over a one-dimensional region of the x -axis, $[a, b]$, and it yields *area* under a *curve*.

Integration of two-variable functions is defined in much the same way as a double Riemann sum

$$\iint_A f(x, y) dA = \lim_{m, n \rightarrow \infty} \sum_{i=1}^m \sum_{j=1}^n f(x_{ij}^*, y_{ij}^*) \Delta A.$$

Recall that a function of two variables describes a surface as opposed to the curve described by a function of one variable. Geometrically, definite integration of two-variable functions can be thought of as volume under the surface, and it can be thought of as being the sum of the volumes of an infinite number of rectangular prisms. So instead of dividing an interval of the x axis into an infinite number of subintervals of width dx , we now divide a region of the xy -plane into an infinite number subregions (squares) of area $dA = dx dy$. Now instead of having a rectangle of width dx reaching up to a curve, we have a rectangular prism of base area $dx dy$ reaching up to a surface.

Note that two-variable integration integrates over a two-dimensional region of the xy -plane, and it yields *volume* under a *surface*.

The way we define definite integration is called **Riemann integration** and it uses infinite rectangles and rectangular prisms to partition the area under a curve. **Lebesgue integration**, on the other hand, does it with simple functions. Riemann integration works for our purposes, but it cannot integrate a simple function like the **Dirichlet function**

$$f(x) = \begin{cases} 1 & : x \text{ is irrational} \\ 0 & : x \text{ is rational} \end{cases}$$

but Lebesgue integration can.

The order of the differentials in a double integral is important. For example in the integral

$$\int_a^b \int_c^d f(x, y) dy dx$$

the order $dy dx$, tells us that $f(x, y)$ is integrated first with respect to y from c to d and then the result is integrated with respect to x from a to b .

This integral can also be written as

$$\iint_R f(x, y) dA, \quad R = [a, b] \times [c, d]$$

When it's written like this, R means the region and in $R = [a, b] \times [c, d]$, the first pair of numbers are the x bounds, and the second pair are the y bounds. In this case, dA is called the area element of the region since it is the length dy times the width dx of an infinitesimal rectangle on R . Since $dA = dx dy = dy, dx$, when written in this form, we can either integrate first with respect to either variable or else the mechanics of the problem dictate the one we must integrate with respect to first.

Note that the multiplication symbol used in $R = [a, b] \times [c, d]$ is called the **Cartesian product**.

Recall that definite integration will always result in just a number.

Recall that the average value of a single-variable function over $[a, b]$ is given by

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

Extending this concept to higher dimensions is easy. For a function of two variables over a region, R , the average value is given by

$$f_{ave} = \frac{1}{A(R)} \iint_R f(x, y) dA,$$

where $A(R)$ is the area of the region R . Calculating the area of R is easy when we're dealing with a rectangular region, but it's not so easy when R is defined by nonlinear functions.

Recall that for single variable functions, integration is a linear operation meaning that the integral of a sum is the sum of integrals. Furthermore, the integral of a function times a constant is the constant times the integral of the function. The same properties hold for double integrals.

$$\iint_R (\alpha f(x) + \beta g(x)) dA = \alpha \iint_R f(x) dA + \beta \iint_R g(x) dA.$$

Also, recall that for single integration, we can split the region into multiple pieces

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx.$$

Similarly, for double integrals, the region R can be split into multiple pieces

$$\iint_R f(x, y) dA = \iint_{R_1} f(x, y) dA + \iint_{R_2} f(x, y) dA,$$

provided that R_1 and R_2 do not overlap and that they cover the entire region R .

Recall that for single integration, the area between curves can be found. Similarly, with double integrals, the volume between two surfaces can be found. If $g(x, y) \geq f(x, y)$ over R , then

$$\iint_R f(x, y) dA \leq \iint_R g(x, y) dA.$$

Note that if $f(x, y) = k$ is a constant function over a rectangular region, then the double integral of it gives the volume of a rectangular box

$$\int_a^b \int_c^d f(x, y) dy dx = k(b-a)(d-c).$$

Fubini's theorem tells us that if f is continuous on $R = [a, b] \times [c, d]$, then

$$\int_c^d \int_a^b f(x, y) dx dy = \int_a^b \int_c^d f(x, y) dy dx.$$

In other words, when the function is continuous on our region, then we can switch the order of integration without a problem.

A double integral can be thought of as an **iterated** integral, and we can evaluate it by evaluating the inside integral and then evaluating the outside integral.

Example:

Evaluate

$$I = \int_1^2 \int_0^\pi \sin(x+y) dy dx.$$

Evaluating, we get

$$\begin{aligned} I &= \int_1^2 \left(\int_0^\pi \sin(x+y) dy \right) dx \\ &= \int_1^2 \left(-\cos(x+y) \Big|_0^\pi \right) dx \\ &= \int_1^2 (-\cos(x+\pi) + \cos x) dx \\ &= \left(-\sin(x+\pi) + \sin x \right) \Big|_1^2 \\ &= -\sin(2+\pi) + \sin 2 + \sin(1+\pi) - \sin 1. \end{aligned}$$

The double integral of two separable functions can be separated. For example

$$\int_a^b \int_c^d f(x)g(y) dy dx = \left(\int_a^b f(x) dx \right) \left(\int_c^d g(y) dy \right)$$

When doing double integration, you usually have to do partial integration with the inner integral—treating the other variable as constant.

When given a choice with respect to which variables to integrate with, as in the example

$$\iint_R x \sin(xy) dA,$$

choose the variable to integrate with first, such that the double integral is easiest. In this case, you should integrate with respect to y first, because doing it with respect to x requires integration by parts.

1.10.2 Double Integrals over General Regions

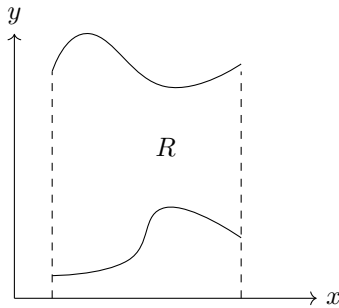
Previously, we saw double integrals over a rectangular region. Now we look at double integrals over more general regions. Keep in mind that the region R , for double integrals, is always in the xy -plane.

The hardest part about doing double integrals over general regions is figuring out the region, and thereby, the structure of the double integral.

The region R is a region of the first kind if it lies between the graphs of two continuous functions of x . This kind of region can be defined as $R = \{(x, y) | a \leq x \leq b, g_1(x) \leq y \leq g_2(x)\}$. For such a region,

$$\iint_R f(x, y) dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y) dy dx.$$

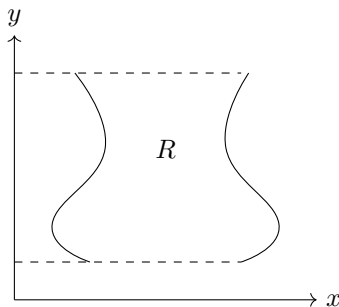
For this case, $g_2(x)$ must always be above $g_1(x)$. If it isn't then you have to break it into multiple regions and use multiple integrals to solve the problem. An example of a region of this kind is shown below.



The region R is a region of the second kind if it lies between the graphs of two continuous functions of y . This kind of region can be defined as $R = \{(x, y) | c \leq y \leq d, h_1(y) \leq x \leq h_2(y)\}$. For such a region,

$$\iint_R f(x, y) dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y) dx dy.$$

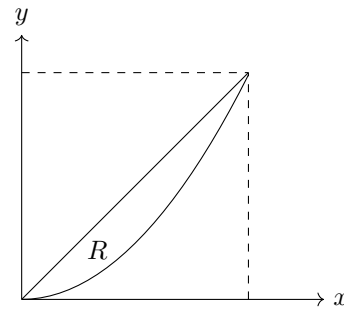
For this case, $h_2(y)$ must always be to the right of $h_1(y)$. If it isn't then you have to break it into multiple regions and use multiple integrals to solve the problem. An example of this kind of region is shown below.



To keep it straight, remember that the outside integral must always be from a constant to a constant. Only the inner integrals can have limits that are functions. Whichever pair of sides of the region that are constants, will tell you what your outer integrals are. If your region is bounded on two sides by the constants $x = a$ and $x = b$ and on the other two sides by curves, then you know that the outer integral will have those bounds, and it will be with respect to x . On the other hand, if two sides of your region is bounded by the constants $y = c$ and $y = d$, and the other two sides are curves, then you know that your outer integral will be with respect to y .

The first, and most important step is to draw your region. Then the curvy lines will form the limits of your inside integral and the straight lines will form the limits of your outside integral.

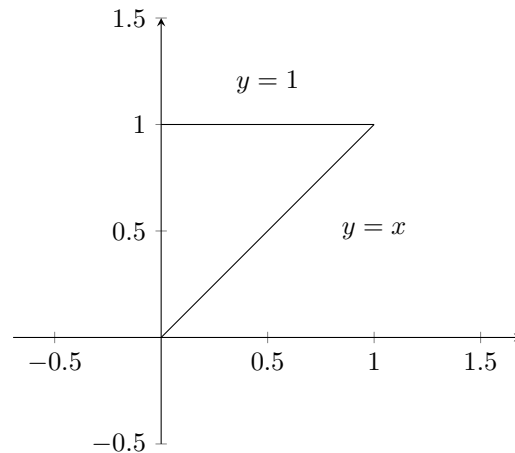
As shown in the picture below, some regions can be interpreted as either the first kind or the second kind. In such a case, just pick whichever way makes the integration easier.



Some functions can only be integrated one way. Consider

$$\int_0^1 \int_x^1 \sin y^2 dy dx.$$

We cannot do this integration, as is, because we don't have the tools to integrate $\sin y^2$. Our only choice is to change the order of integration. Drawing our region, we get



Changing the order of integration, we get a do-able integral

$$\int_0^1 \int_0^y \sin y^2 dx dy.$$

We can estimate a double integral as follows. If $m \leq f(x, y) \leq M$ for all (x, y) in a region R , then

$$mA(R) \leq \iint_R f(x, y) dA \leq MA(R),$$

where $A(R)$ is the area of the region R .

Here's an interesting way to find the area of a triangle when given the three vertex coordinates (a, b) , (c, d) , and (e, f) .

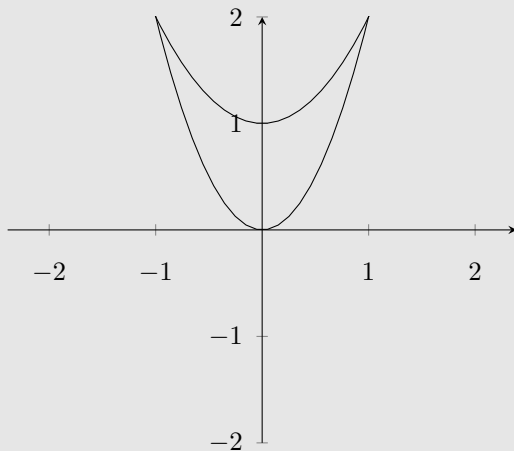
$$A = \frac{1}{2} \begin{vmatrix} 1 & 1 & 1 \\ a & c & e \\ b & d & f \end{vmatrix}.$$

Example:

Given $R = \{(x, y) | y = 2x^2 \text{ and } y = 1 + x^2\}$, find

$$I = \iint_R (x + 1) dA.$$

The first step is to find the region. Since no constants are given in the definition of the problem's region—only two functions of x , then it must be the case that the two functions enclose an area. We start by graphing $y = x^2$ and $y = 1 + x^2$ on the xy -plane.



By looking at the graph, we see that one curve is always above the other curve, which means it is a region of the first kind. One curve is not always to the right of the other, so it is not a region of the second kind. The

other way we could have figured out the type of region is simply by realizing that the bounding functions are functions of x rather than functions of y .

This means, we first integrate with respect to y , using the bounds $2x^2 \leq y \leq 1 + x^2$. Next we integrate that result with respect to x . We can find the constant bounds for the outer integral by setting the two equal to each other to find their intersections. We find that the limits for the outer integral are $-1 \leq x \leq 1$. Putting it all together, our double integral is

$$\begin{aligned} I &= \int_{-1}^1 \int_{2x^2}^{1+x^2} (x + 1) dy dx \\ &= \int_{-1}^1 (xy + y) \Big|_{y=2x^2}^{x^2+1} dx \\ &= \int_{-1}^1 [-x^3 - x^2 + x + 1] dx \\ &= \left[-\frac{1}{4}x^4 - \frac{1}{3}x^3 + \frac{1}{2}x^2 + x \right]_{-1}^1 \\ &= \frac{4}{3}. \end{aligned}$$

Because our function is continuous over the region, by Fubini's theorem, we could switch the order of integration, that is, we could convert it to region of the second kind. However, we would have to change the limits of integration, and we would end up having to solve multiple integrals. It would be a lot messier that way.

1.10.3 Change of Variables

Note that “change of variables” is simply a fancy way of saying “substitution”. For single integrals, substitution is straightforward, but it's not so straightforward for double integrals. The hard part is figuring out what the new area element (i.e. dA) is or for triple integrals, what the volume element (i.e. dV) is.

To evaluate double integrals, it is sometimes easier to switch coordinate systems, which requires making appropriate substitutions. We may convert between the Cartesian, polar, cylindrical, and spherical coordinate systems.

The **Jacobian** tells us how to convert between two coordinate systems, and it tells us what the area or volume element is. The general Jacobian to transform from (x, y) to (u, v) is given by

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}.$$

To find the Jacobian of a transformation in 3 vari-

ables, this idea can easily be extended:

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}.$$

The absolute value of the Jacobian is used to figure out the new area element when we're making a substitution. In other words, when going from an (x, y) to a (u, v) coordinate system,

$$dA = \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv.$$

The whole process of transforming a double integral from (x, y) to (u, v) is shown here

$$\iint_R f(x, y) dA = \iint_S f(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv.$$

Note: If we reverse the order of the transformation, that is, we go from (u, v) to (x, y) , then we get the inverse for the Jacobian. For example, to go from (x, y) to (r, θ) , the Jacobian is r . To go the other way, the Jacobian is $\frac{1}{r}$.

Example:

Solve

$$\iint_R \frac{x-2y}{3x-y} dA,$$

where R is the parallelogram enclosed by the lines $x-2y=0$, $x-2y=4$, $3x-y=1$, and $3x-y=8$. ^a

Notice that the double integral is not easy to solve as it is, however, we might be able to transform the integral into something much easier to solve. The structure of the integrand suggests that we make the transformation

$$u = x - 2y \quad \text{and} \quad v = 3x - y$$

This should help because our integral will then be of the form

$$\iint_S \frac{u}{v} \left| \frac{\partial(x,y)}{\partial(u,v)} \right| du dv.$$

Our next step is to find the Jacobian. To do that, we first need to solve our transformation equations for x and y . We can do this by solving both equations for x and then substituting one into the other. When we do so, we get

$$x = \frac{2}{5}v - \frac{1}{5}u \quad \text{and} \quad y = \frac{1}{5}v - \frac{3}{5}u$$

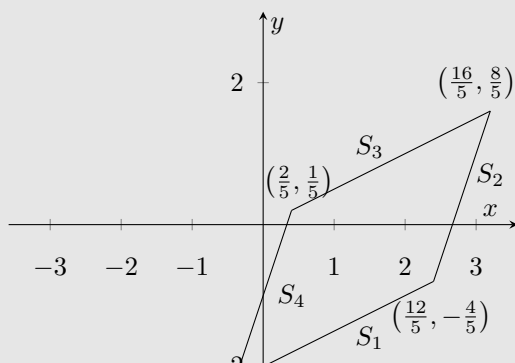
The Jacobian, then, is

$$\left| \frac{\partial(x,y)}{\partial(u,v)} \right| = \left| \begin{vmatrix} -\frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{1}{5} \end{vmatrix} \right| = \left| -\frac{1}{25} + \frac{6}{25} \right| = \frac{1}{5}.$$

Now, our transformed integral is

$$\frac{1}{5} \iint_S \frac{u}{v} du dv.$$

The hardest part, though, is figuring out what the limits of integration are, and to do that, we have to figure out what the new region, S , looks like. We start by sketching the original region, R . We know that it is a parallelogram enclosed by the lines $x-2y=0$, $x-2y=4$, $3x-y=1$, and $3x-y=8$. We sketch R on the left below. If we solve all these equations for y , we get ones that are easier to graph.



$$y = \frac{1}{2}x$$

$$y = \frac{1}{2}x - 2$$

$$y = 3x - 1$$

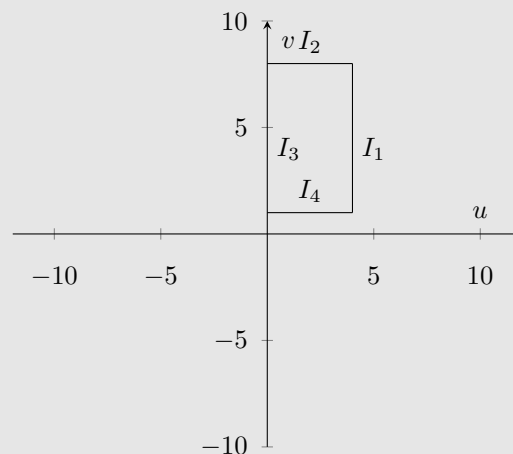
$$y = 3x - 8$$

To help with finding the new region of integration, we denote the sides of our first region with S_1 , S_2 , and so on making sure to do it either in clockwise or counterclockwise order. We also mark all the points at the corners of our old region. Now for each side of the old region, we have to figure out the corresponding line or curve in the new coordinate system. Let's start with S_1 .

S_1 is defined by $y = \frac{1}{2}x - 2$ with x running from $-\frac{2}{5}$ to $\frac{12}{5}$. In other words, $-\frac{2}{5} \leq x \leq \frac{12}{5}$. Plugging $y = \frac{1}{2}x - 2$ into our transformation equations $u = x - 2y$, and $v = 3x - y$ gives us $u = 4$ and $v = \frac{5}{2}x + 2$. Plugging $-\frac{2}{5} \leq x \leq \frac{12}{5}$ into $v = \frac{5}{2}x + 2$ shows us the v will run from 1 to 8. This gives us our first line I_1 (i.e. the image of S_1 in our new coordinate system).

S_2 is defined by $y = 3x - 8$ with x running from $\frac{12}{5}$ to $\frac{16}{5}$. In other words, $\frac{12}{5} \leq x \leq \frac{16}{5}$. Plugging $y = 3x - 8$ into our transformation equations $u = x - 2y$, and $v = 3x - y$ gives us $u = -5x + 8$ and $v = 8$.

We use the same process to find the other two boundary lines for our new region, S .



Now that we know our new region of integration, S , we can set up our final integral and solve it:

$$\begin{aligned} \frac{1}{5} \int_1^8 \int_4^8 \frac{u}{v} du dv &= \frac{1}{5} \left(\int_1^8 \frac{1}{v} dv \right) \left(\int_4^8 u du \right) \\ &= \frac{1}{5} \left(\ln(v) \Big|_1^8 \right) \left(\frac{1}{2} u^2 \Big|_4^8 \right) \\ &= \frac{1}{5} (\ln(8) - \ln(1)) (8) \end{aligned}$$

1.10.4 Double Integrals in Polar Coordinates

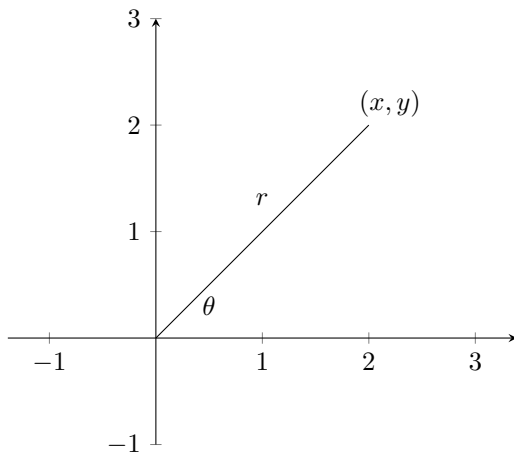
Recall that a point in the xy -plane can also be represented by a distance r from the origin and an angle, θ , from the positive x -axis.

Recall that the transformation equations to go from the (x, y) coordinate system to polar coordinates are

$$\begin{aligned}x &= r \cos \theta \\y &= r \sin \theta\end{aligned}$$

By squaring the two equations and by taking an arc tangent, we get the transformation equations to go from the polar coordinate system to the Cartesian coordinate system:

$$\begin{aligned}r &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1} \left(\frac{y}{x} \right)\end{aligned}$$



The Jacobian to convert from Cartesian to polar coordinates is

$$\begin{aligned}\frac{\partial(x, y)}{\partial(r, \theta)} &= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \\ &= \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} \\ &= r \cos^2 \theta + r \sin^2 \theta = r.\end{aligned}$$

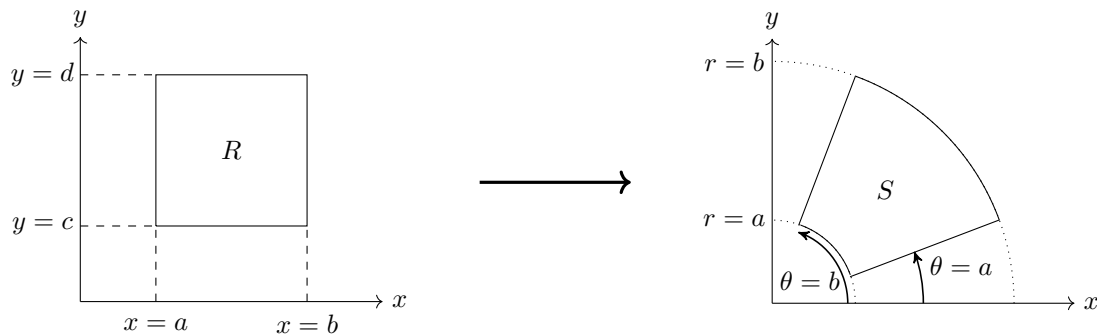
This means that

$$\iint_R f(x, y) dA = \iint_S f(r, \theta) r dr d\theta.$$

If we wanted to go the other way—from polar to Cartesian coordinates, we could work out the new Jacobian, or we could just realize that one is the inverse of the other. In this case, the Jacobian would be $\frac{1}{r}$ and

$$\iint_S f(r, \theta) dr d\theta = \iint_R f(x, y) \frac{1}{r} dA.$$

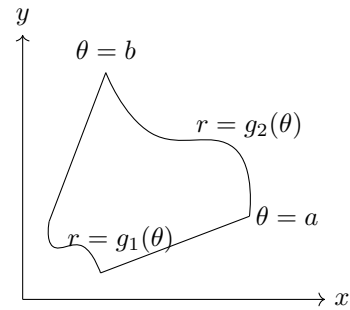
Converting a double integral from rectangular to polar coordinates is a **transformation**.



Recall that the Jacobian for the transformation from Cartesian to polar is simply r . The full transformation is given by

$$\iint_R f(x, y) dA = \int_{\alpha}^{\beta} \int_a^b f(r \cos \theta, r \sin \theta) r dr d\theta.$$

What we don't want in polar coordinates is a region of integration in which the radius varies. We want everything to be constant. We don't want a region where the radius is a function of θ , $r = h(\theta)$.



Example:

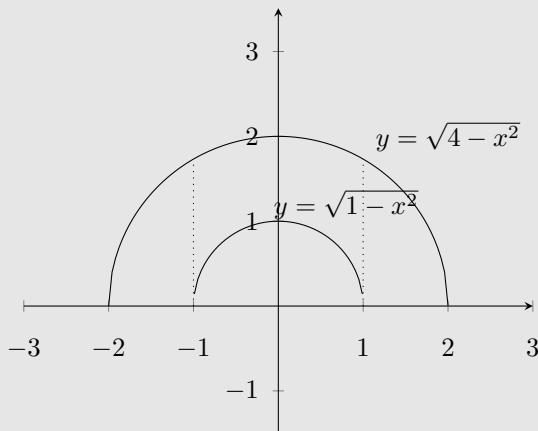
Solve

$$\iint_R (3x + 4y^2) dA,$$

where $R = \{(x, y) | y \geq 0 \text{ and } 1 \leq x^2 + y^2 \leq 4\}$. Sketching this region, we see that it's half of the annulus between two circles—one with radius 2, and the other with radius 1.

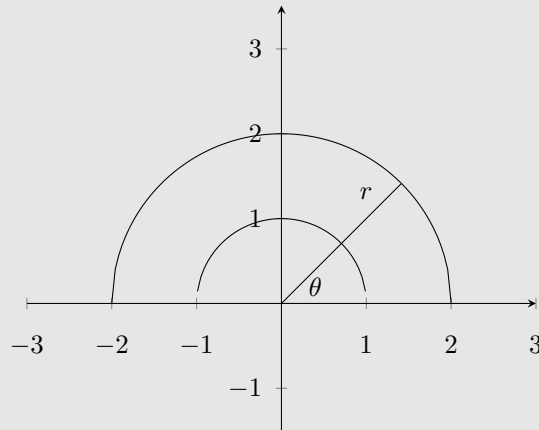
Doing the integration with respect to y first would require 3 integrals as demonstrated by the graph below.

$$\begin{aligned} & \int_{-2}^{-1} \int_0^{\sqrt{4-x^2}} (3x + 4y^2) dy dx \\ & + \int_{-1}^1 \int_{\sqrt{1-x^2}}^{\sqrt{4-x^2}} (3x + 4y^2) dy dx \\ & + \int_1^2 \int_0^{\sqrt{4-x^2}} (3x + 4y^2) dy dx. \end{aligned}$$



2

Doing the integration with respect to x first, wouldn't be any better. That would also require multiple integrals. Notice however, that if we used polar coordinates it would be a lot easier. The radius r would go from 1 to 2, and the angle would sweep from 0 to π .



Our new, and much easier double integral becomes

$$\int_0^\pi \int_1^2 (3r \cos \theta + 4r^2 \sin^2 \theta) r dr d\theta.$$

Example:

Since we don't have the tools to integrate $\sin(x^2)$, we cannot evaluate the double integral

$$\iint_R \sin(x^2 + y^2) dA,$$

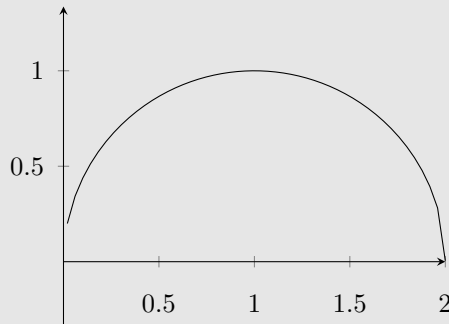
as is. However, if we transform into polar coordinates, the Jacobian r that we gain makes it possible for us to do the integration.

$$\iint_R \sin(x^2 + y^2) dA = \iint_S \sin(r^2) r dr d\theta.$$

Example:Solve ^a

$$\int_0^2 \int_0^{\sqrt{2x-x^2}} \sqrt{x^2+y^2} dy dx.$$

Graphing the region of integration, we see that we have an off-center circle. How do we find the limits for the radius when we convert to polar coordinates.



To find the function for the radius, we recall that $x^2 + y^2 = r^2$. In our case, the circle in rectangular coordinates is given by $(x-1)^2 + y^2 = 1$. Substituting $x = r \cos \theta$ and $y = r \sin \theta$, we get $r = 2 \cos \theta$, and so our transformed integral becomes

$$\int_0^{\frac{\pi}{2}} \int_0^{2 \cos \theta} r^2 dr d\theta = \frac{16}{9}.$$

^aThis is problem 32 in chapter 15.4 in Stewart's Calculus with Early Transcendentals.

1.10.5 Surface Area

The formula to calculate the surface area of $z = f(x, y)$ over the region R is

$$A(S) = \iint_R \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} dA.$$

Calculating the surface area in \mathbb{R}^3 is analogous to calculating the arc length in \mathbb{R}^2 .

1.10.6 Triple Integrals

Triple integrals are defined in much the same way as double integrals. If the region of integration is $B = [a, b] \times [c, d] \times [m, n]$, then

$$\iiint_B f(x, y, z) dV = \int_a^b \int_c^d \int_r^s f(x, y, z) dz dy dx,$$

where dV is called the volume element and it has 6 possibilities—the 6 different ways to order dx , dy , and dz .

If the limits are all constants, then Fubini's theorem applies, and we are able to switch to any of the six different orders of integration.

If the function $w = f(x, y, z)$ is separable, that is, $z = F(x)G(y)H(z)$, then we can separate the triple integral as

$$\int_a^b \int_c^d \int_m^n f(x, y, z) dz dy dx = \left(\int_a^b F(x) dx\right) \left(\int_c^d G(y) dy\right) \left(\int_m^n H(z) dz\right).$$

The limits do not all have to be constant. The inner limits could be functions of two variables, the middle limits could be functions of single variable, but the outer limits still have to be constant. For example, we could have

$$\int_a^b \int_{g_1(x)}^{g_2(x)} \int_{h_1(x,y)}^{h_2(x,y)} f(x, y, z) dz dy dx.$$

Triple integrals are harder to evaluate because we cannot visualize the three variable function or the “hypervolume” that we're computing. In fact, just our *region* of integration is now a three-dimensional shape.

Some triple integrals are much easier if we switch to cylindrical or spherical coordinates. Keep in mind that you can even switch coordinates part way through the

process. For example, you might evaluate the inside integral without a problem, but then decide to switch to polar coordinates to evaluate the middle integral.

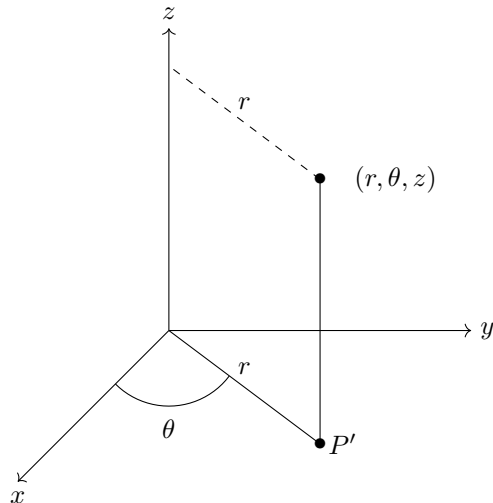
1.10.7 Triple Integrals in Cylindrical Coordinates

The cylindrical coordinate system is a three dimensional extension of polar coordinates. It's essentially the same as polar coordinates except that it adds the z axis of a three-dimensional system.

The transformation equations are

$$\begin{aligned}x &= r \cos \theta \\y &= r \sin \theta \\z &= z\end{aligned}$$

In the cylindrical coordinate system, a point is represented by the coordinates (r, θ, z) , where r is the distance from the origin to the point, θ is the angle between the point's projection in the xy -plane and the positive x axis, and z is the point's height above the xy -plane. The point, then, sits on the surface of a cylinder.



To convert from Cartesian coordinates to cylindrical coordinates, we find that the Jacobian is

$$\frac{\partial(x, y, z)}{\partial(r, \theta, z)} = r,$$

which is the same as from Cartesian coordinates to polar coordinates.

The total volume element is when converting from rectangular to cylindrical coordinates is

$$dV = r \, dz \, dr \, d\theta.$$

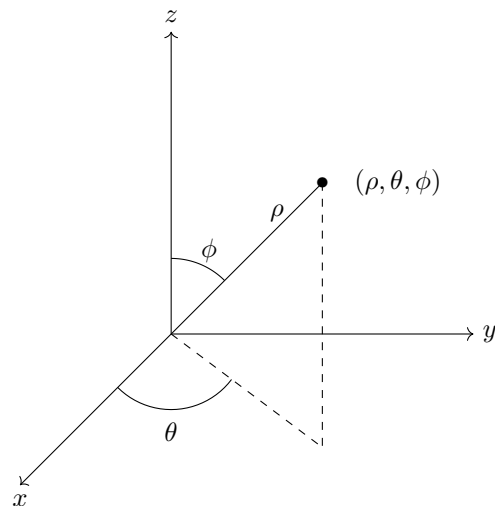
Note: The spherical coordinate system is defined a little differently in physics.

1.10.8 Triple Integrals in Spherical Coordinates

The spherical coordinate system is a three dimensional coordinate system. In it, a point sits on the surface of a sphere.

The transformation equations are

$$\begin{aligned}x &= \rho \sin \phi \cos \theta \\y &= \rho \sin \phi \sin \theta \\z &= \rho \cos \phi\end{aligned}$$



A point in spherical coordinates is given as (ρ, θ, ϕ) , where ρ is the distance from the origin to the point, ϕ is the angle between ρ and the positive z -axis, and θ is the angle between the point's projection onto the xy -plane and the positive x -axis.

Notice that

$$x^2 + y^2 + z^2 = \rho^2.$$

To convert from Cartesian coordinates to spherical coordinates, we find that the Jacobian is

$$\frac{\partial(x, y, z)}{\partial(\rho, \theta, \phi)} = \rho^2 \sin \phi.$$

This implies that to convert from spherical coordinates to Cartesian coordinates, our Jacobian will be

$$\frac{\partial(\rho, \theta, \phi)}{\partial(x, y, z)} = \frac{1}{\rho^2 \sin \phi}.$$

The total volume element is when converting from rectangular to spherical coordinates is

$$dV = \rho^2 \sin \phi \, d\rho \, d\theta \, d\phi.$$

Note: The spherical coordinate system is defined a little differently in physics.

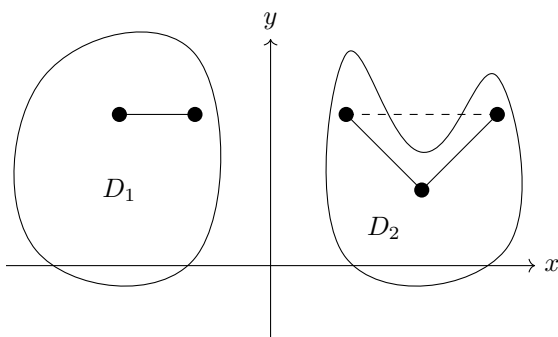
1.11 Vector Calculus

1.11.1 Vector Fields

A **vector field** is a function that assigns vectors to points in space. Given a region R in \mathbb{R}^2 , a vector field is a function \vec{F} that assigns to every point (x, y) in R a vector $\vec{F}(x, y)$. This can be expressed as

$$\vec{F}(x, y) = \langle P(x, y), Q(x, y) \rangle = P(x, y)\vec{i} + Q(x, y)\vec{j}.$$

A vector field can also be defined in R^3 in which case a vector function assigns a vector to every point in some region of R^3 .



If the straight line between any two points in D is completely within D , then D is called a *convex region*. This is the nicest kind of region and is illustrated in region D_1 above. On the other hand, for a region like D_2 , if it's not a convex region, it might still be a *path-connected region*.

A vector field is basically like the **direction field** from differential equations.

Gradient Field

Recall that the gradient of a scalar function $f(x, y)$ is defined as

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y) \rangle.$$

Notice that $\nabla f(x, y)$ is actually a vector field, and it is called the **gradient vector field**.

A vector field is called **conservative** if it is the gradient of a scalar function. That is, when

$$\vec{F}(x, y) = \nabla f(x, y).$$

1.11.2 Curl and Divergence

Curl and divergence are two operations that can be performed on vector fields. Curl yields another vector field and divergence yields a scalar field.

Recall that the vector differential operator ∇ , called “del” is defined as

$$\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle,$$

and when applied to a scalar function, ∇f , it is called the gradient of f .

The **curl** of a vector field \vec{F} is defined as the cross product of the del operator with the vector field

$$\text{curl } \vec{F} = \nabla \times \vec{F}.$$

Recall that the cross product of two vectors is defined as

$$\vec{u} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix},$$

so if

$$\vec{F}(x, y, z) = \langle P(x, y, z), Q(x, y, z), R(x, y, z) \rangle,$$

then curl is

$$\begin{aligned} \nabla \times \vec{F} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} \\ &= \langle R_y - Q_z, P_z - R_x, Q_x - P_y \rangle \end{aligned}$$

Note: When expanding this determinant, you have to be careful about the order that you multiply the elements. For example, $\frac{\partial}{\partial x}P$ is meaningful, but $P\frac{\partial}{\partial x}$ is not.

The curl vector is associated with rotations. If \vec{F} represents the velocity field in fluid flow, then a particle near (x, y, z) in the fluid will rotate about the axis that points in the direction of $\text{curl } \vec{F}$. The length of this curl vector is a measure of how quickly, the particle moves around that axis.

The **divergence** of a vector field \vec{F} is defined as the dot product of the del operator with the vector field

$$\text{div } \vec{F} = \nabla \cdot \vec{F} = P_x + Q_y + R_z.$$

Notice that this yields a scalar function.

The $\text{div } \vec{F}$ measures the tendency for a fluid to diverge from the point (x, y, z) . If $\text{div } \vec{F} = 0$, then \vec{F} is **incompressible**.

If we compute the divergence of a gradient field, we get the **Laplace equation**.

$$\text{div } (\nabla f) = \nabla \cdot \nabla f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \nabla^2 f.$$

If you take the curl of the gradient of a function, you get the zero vector $\vec{\mathbf{0}}$.

$$\text{curl}(\nabla \vec{F}) = \nabla \times \nabla \vec{F} = \vec{\mathbf{0}}.$$

Here's the proof

$$\begin{aligned} \nabla \times \nabla \vec{F} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_x & f_y & f_z \end{vmatrix} \\ &= \langle f_{zy} - f_{yz}, -f_{zx} + f_{xz}, f_{yx} - f_{xy} \rangle \\ &= \langle 0, 0, 0 \rangle \\ &= \vec{\mathbf{0}} \end{aligned}$$

To reiterate, if you have the gradient of a function, taking the curl of that yields the zero vector.

Recall that a vector field is called **conservative** if it is the gradient of a scalar function. That is, when

$$\vec{F}(x, y) = \nabla f(x, y).$$

Putting these two concepts together, we can say that if $\vec{F}(x, y)$ is conservative, then $\text{curl } F(x, y) = \vec{\mathbf{0}}$. This gives us a way of verifying that a vector field is *not* conservative by showing that its curl is not equal to the zero vector. In other words, showing that $\text{curl } F(x, y) = \vec{\mathbf{0}}$ is a necessary, but not sufficient condition to show that $\vec{F}(x, y)$ is conservative.

If \vec{F} is a vector field defined for the whole of \mathbb{R}^3 , and all its partial derivatives are continuous and $\text{curl } F(x, y) = \vec{\mathbf{0}}$, then \vec{F} is conservative.

If \vec{F} is a conservative vector field then the scalar function f , from $\vec{F}(x, y) = \nabla f(x, y)$ is called the **potential** of the vector field.

1.11.3 Line Integrals

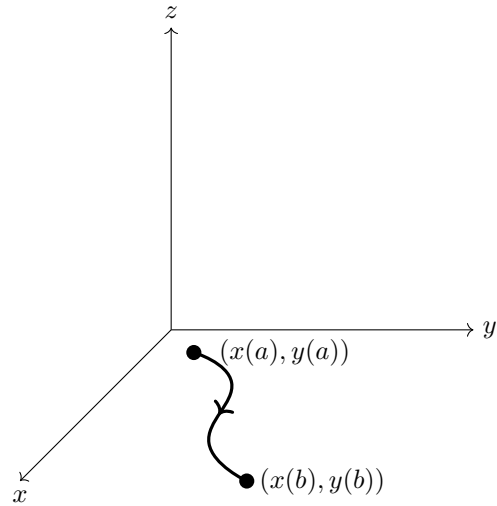
Recall that in two dimensions, the definite integral $\int_a^b f(x) dx$ can be interpreted as the area under the curve $f(x)$ from a to b along the x -axis.

In three dimensions, function $f(x, y)$ defines a surface. Again, the definite integral $\int_a^b f(x, y) dx$ can be interpreted as the area under the surface $f(x, y)$ from a to b along the x -axis. What if you want to determine the area under the surface $f(x, y)$ from a point a to a point b that is not along the x -axis? If we have a curve C in the xy -plane, then

$$\int_C f(x, y) dS,$$

called a **line integral**, can be interpreted as the area under the surface $f(x, y)$ along the curve C . You can imagine the curve C forming the base of a wall that reaches up

to the surface $f(x, y)$, then this line integral is the area of one side of that wall.



The orientation of the curve is determined by its initial point and its terminal point.

If the curve, C is merely a portion of the x -axis, then the line integral reduces to the regular Riemann integral.

If f is a continuous function then the formula

$$\int_C f(x, y) ds = \int_a^b f(x(t), y(t)) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt.$$

can be used to compute the line integral.

Note: The line integral does not depend on how exactly the curve is parametrized, provided that the whole curve is covered and no segment of it is covered more than once.

This only works with smooth (i.e. no cusps and no self-intersections) curves.

If a curve is not smooth but has multiple connected sections that are individually smooth, then we call it **piecewise continuous**. Such a line integral poses no problem—we just have to evaluate it as the sum of multiple line integrals—one for each segment.

When doing these problems, we have to be able to parametrize functions—meaning we have to be able to express functions of multiple variables in terms of a single variable.

Here is a “silly” example where we simply substitute $x = t$ then solve for y . This simple parametrization is not always the most advantageous one:

$$y = x^2 \quad \Longrightarrow \quad \begin{cases} x = t \\ y = t^2 \end{cases}$$

Here's a parametrization that utilizes trigonometry functions:

$$y = \sqrt{1-x^2} \quad \Rightarrow \quad \begin{cases} x = \cos t \\ y = \sin t \end{cases}$$

where $0 \leq t \leq \pi$.

Note, that by convention, we may write

$$\int_C P(x) dx + \int_C Q(x) dy = \int_C P(x) dx + Q(x) dy.$$

Line Integrals over Vector Fields

If \vec{F} is a continuous vector field defined on a smooth curve C given by $\vec{r}(t) = \langle x(t), y(t), z(t) \rangle$ for $a \leq t \leq b$, then its line integral over C is given by

$$\int_C \vec{F} \cdot d\vec{r}.$$

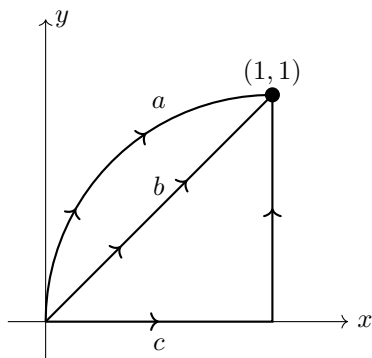
The differential $d\vec{r}$ is the infinitesimal displacement vector along the curve. To find $d\vec{r}$, we simply differentiate \vec{r} with respect to t , to get $d\vec{r} = \vec{r}'(t) dt$. The line integral becomes

$$\int_C \vec{F} \cdot d\vec{r} = \int_C \vec{F} \cdot \vec{r}'(t) dt.$$

Consider the vector field $\vec{F} = \langle x, y \rangle$. How do we calculate

$$\int_C \vec{F} \cdot d\vec{r},$$

over the three curves shown below?



Path a can be parametrized as

$$\vec{r} = \langle x, y \rangle = \langle 1 - \cos t, \sin t \rangle,$$

from $t = 0$ to $t = \frac{\pi}{2}$, which defines a quarter circle centered at $(1, 0)$. Parametrizing \vec{F} by plugging in $x = 1 - \cos t$ and $y = \sin t$ gives us

$$\vec{F} = \langle x, y \rangle = \langle 1 - \cos t, \sin t \rangle.$$

Taking the derivative of \vec{r} , we have that

$$d\vec{r} = \langle \sin t, \cos t \rangle dt,$$

and taking the relevant dot product, we get

$$\begin{aligned} \vec{F} \cdot d\vec{r} &= (1 - \cos t)(\sin t) dt + \cos t \sin t dt \\ &= \sin t dt. \end{aligned}$$

So our integral becomes

$$\int_C \vec{F} \cdot d\vec{r} = \int_0^{\frac{\pi}{2}} \sin t dt = 1.$$

Path b we can parametrize as

$$\vec{r} = \langle x, y \rangle = \langle t, t \rangle,$$

for $0 \leq t \leq 1$. Parametrizing the vector field, we get

$$\vec{F} = \langle x, y \rangle = \langle t, t \rangle.$$

Taking the derivative of \vec{r} , we have that

$$d\vec{r} = \langle 1, 1 \rangle dt,$$

so $\vec{F} \cdot d\vec{r} = 2t dt$ and our integral becomes

$$\int_C \vec{F} \cdot d\vec{r} = 2 \int_0^1 t dt = 1.$$

For path c , we have to break the line integral into two pieces. We can parametrize the first portion as $\vec{r}_1 = \langle t, 0 \rangle$ for $0 \leq t \leq 1$ then the differential is $d\vec{r}_1 = \langle 1, 0 \rangle dt$, the parametrized vector field is $\vec{F}_1 = \langle t, 0 \rangle$, and the dot product is $\vec{F}_1 \cdot d\vec{r}_1 = t dt$. We can parametrize the second portion as $\vec{r}_2 = \langle 1, t \rangle$ for $0 \leq t \leq 1$ then the differential is $d\vec{r}_2 = \langle 0, 1 \rangle dt$, the parametrized vector field is $\vec{F}_2 = \langle 1, t \rangle$, and the dot product is $\vec{F}_2 \cdot d\vec{r}_2 = t dt$. Putting it all together, we have

$$\int_C \vec{F} \cdot d\vec{r} = \int_0^1 t dt + \int_0^1 t dt = 1.$$

Notice that our result was the same regardless of which path we integrated over. This is because our vector field \vec{F} is conservative. In general, integrating over a different path will yield a different result.

In general, to find the line integral $\int_C \vec{F} \cdot d\vec{r}$, we follow the steps

1. Parametrize the curve C as $\vec{r}(t)$ —a function of t .
2. Since \vec{r} gives you the values for x and y , use these to parametrize the vector field $\vec{F}(x, y)$.
3. Differentiate \vec{r} with respect to t to find $d\vec{r}$.
4. Take the dot product $\vec{F} \cdot d\vec{r}$.
5. Perform the integration.

We can also think of these line integrals in a different way—by breaking them into x and y components from the beginning. In general,

$$\int_C \vec{F} \cdot d\vec{r} = \int F_x dx + \int F_y dy.$$

When we do this, we still have to parametrize the path z by finding some vector \vec{r} function of a single parameter.

Using the same examples given above, for path a we have the parametrization $\vec{r} = \langle 1 - \cos t, \sin t \rangle$, so $x = 1 - \cos t$, which gives us $dx = \sin t dt$, and $y = \sin t$, which gives us $dy = \cos t dt$. The parameter t runs from $t = 0$ to $t = \frac{\pi}{2}$. The x component of the force is just $F_x = x = 1 - \cos t$ and the y -component is $F_y = \sin t$, so

$$\begin{aligned} \int_C \vec{F} \cdot d\vec{r} &= \int F_x dx + \int F_y dy \\ &= \int_0^{\frac{\pi}{2}} (1 - \cos t) \sin t dt + \int_0^{\frac{\pi}{2}} \sin t \cos t dt \\ &= 1. \end{aligned}$$

Path b we have the parametrization $\vec{r} = \langle x, y \rangle = \langle t, t \rangle$ for $0 \leq t \leq 1$. Now we have $x = t$, and $y = t$ with $dx = dy = dt$, so

$$\begin{aligned} \int_C \vec{F} \cdot d\vec{r} &= \int F_x dx + \int F_y dy \\ &= \int_0^1 t dt + \int_0^1 t dt \\ &= 1. \end{aligned}$$

For path c , our problem is simple. Since the x dependence of the path occurs only on the first half and the y dependence of the path depends only on the second half, we don't even have to parametrize the path.

$$\begin{aligned} \int_C \vec{F} \cdot d\vec{r} &= \int F_x dx + \int F_y dy \\ &= \int_0^1 x dx + \int_0^1 y dy \\ &= 1. \end{aligned}$$

Even if F_x didn't depend only on x and F_y didn't depend only y , we could still do this type of problem, where the path is formed of straight lines in the x or y direction, very easily since for whichever variable you're integrating, the other will be constant, and you can just substitute that constant number in for the variable. For example, if $\vec{F} = \langle 3y, 2x^2y \rangle$ then on the first portion of the path, you would integrate the $F_x = 3y$ with respect to x , but since $y = 0$ over this part of the path, the integral would be 0. Similarly for the second portion of the path, you would integrate $F_y = 2x^2y$ with respect to y , which is easy given

that $x = 1$, so you really just integrate $2y$ with respect to y .

1.11.4 Gradient Theorem

Recall that the fundamental theorem of calculus states that

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

There is an analogous theorem for line integrals, which states that

$$\int_C \nabla f \cdot dr = f(\vec{r}(b)) - f(\vec{r}(a)).$$

In our case, ∇f acts like the derivative of the function. This is known as the **gradient theorem** or as the **fundamental theorem for line integrals**.

What would it take for two different paths (i.e. two different C 's) with the same initial and terminal points to yield the same line integral? In other words, when is the integral

$$\int_C \vec{F} \cdot dr,$$

independent of the path C ? The answer is, it is independent of the path C if \vec{F} is a conservative field, that is, if $\vec{F} = \nabla f$.

The field $\vec{F} = P\hat{i} + Q\hat{j} = \langle P(x, y), Q(x, y) \rangle$ is conservative is

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

This gives us a quick and easy way of showing that a field is conservative.

1.11.5 Green's Theorem

If C is a positively-oriented (i.e. counterclockwise), piecewise smooth, simple closed curve in the plane and R is the region bounded by C , then if P and Q have continuous partial derivatives on an open region that contains R , then

$$\int_C P dx + Q dy = \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA.$$

Index

3D Coordinate Systems, 48

Absolute convergence, 46

Absolute maximum, 78

Absolute minimum, 78

Acceleration, 66

Additive identity, 49

Additive inverse, 49

Alternating series, 45

Alternating series test, 45

Antiderivative, 27, 30

Approximation, 20

Arc length, 42, 63

Associativity, 49

Asymptote, 24

Average value of a function, 41

Average velocity, 18

Binormal vector, 65

Bounded sequence, 43

Cartesian product, 82

- Cauchy-Schwarz inequality, 53
- Centripetal force, 66
- Chain rule, 14, 72
- Change of variables, 85
- Clairaut's theorem, 71
- Closure, 49
- Commutativity, 49
- Comparison theorem, 38
- Component form, 51
- Composition law, 23
- Concavity, 25
- Concavity test, 24
- Conditional convergence, 46
- Conic section, 48
- Conservative, 91
- Continuity, 6, 68
- Critical value, 21
- Cross Product, 54
- Curl, 91
- Curvature, 63, 64
- Curve sketching, 23
- Cylindrical coordinates, 89

- D'Alembert operator, 76
- Definite integral, 27, 28
- Derivative test, 24
- Difference quotient, 3, 10
- Differential, 21
- Direct substitution, 69
- Direction cosines, 53
- Direction numbers, 58
- Directional derivative, 75
- Dirichlet function, 82
- Discontinuity, 7
- Distance formula, 48
- Distance function, 79
- Distributivity, 49
- Divergence, 91
- Dot product, 52
- Double integrals, 82

- Epsilon-delta definition, 4
- Equation of a plane, 72
- Extreme value, 21
- Extreme value theorem, 21
- Extreme values, 24, 77

- Feasible region, 80
- Fermat's theorem, 21
- First derivative test, 24
- Fubini's theorem, 83
- Fundamental theorem of calculus, 29, 30

- Geometric series, 44
- Gradient, 75, 91
- Gradient field, 91
- Gradient theorem, 94
- Green's theorem, 94

- Harmonic function, 71
- Heat equation, 71
- Hessian determinant, 77
- Hooke's law, 41
- Horizontal asymptote, 24
- Horizontal asymptote, 8
- Hyperbolic functions, 21

- Implicit differentiation, 15, 72, 74
- Improper integrals, 37
- Increasing/decreasing test, 23
- Indeterminate difference, 23
- Indeterminate forms, 22
- Indeterminate power, 23
- Indeterminate product, 23
- Infinite discontinuity, 7
- Inflection point, 25
- Initial point, 49
- Inner product, 52
- Instantaneous acceleration, 18
- Instantaneous jerk, 18
- Instantaneous velocity, 18
- Integration by parts, 33
- Intermediate value theorem, 8
- Interval of convergence, 47
- Inverse trig functions, 17
- Iterated integrals, 83

- Jacobian, 85
- Jump discontinuity, 7

- L'Hopital's rule, 22
- Lagrange multipliers, 80
- Laplace equation, 76
- Laplace's equation, 71
- Law of cosines, 52
- Least squares approximation, 80
- Limit, 13, 68
- Limit laws, 5
- Line integrals, 94
- Line segment, 58
- Linear approximation, 20, 72
- Linear programming, 80
- Linearization, 20
- Logarithm, 34
- Logarithms, 17

- Maclaurin series, 47
- Magnitude, 49, 52
- Mathematical induction, 43
- Maximum, 21, 24, 77
- Mean value theorem, 41
- Midpoint, 48
- Minimum, 21, 24, 77
- Mixed partial derivatives, 71
- Modulus, 52
- Monotonic sequence, 43
- Monotonic sequence theorem, 43
- Multiplicative identity, 49
- Multiplicative inverse, 49
- Multivariable functions, 68

- Natural logarithm, 34
- Net change, 31
- Net change theorem, 31
- Newton's method, 26
- Norm, 52
- Normal component, 66
- Normal line, 11, 76
- Normal vector, 59, 65

- Oblique asymptote, 25
- Optimization, 25
- Orthogonal projection, 53
- Osculating circle, 65

- P-integral, 38
- P-series, 45
- Parallelepiped, 55
- Parallelogram, 25, 56
- Parametric form, 57
- Partial derivative, 70
- Partial fraction decomposition, 37
- Partial sum, 44
- Plane, 59
- Position vector, 66
- Power rule, 11
- Power series, 46
- Product rule, 12
- Projectile motion, 66
- Pythagorean theorem, 52

- Quadric surfaces, 60
- Quotient rule, 12

- Radial limit, 69
- Radius of convergence, 47
- Rate of change, 18
- Ratio test, 46
- Rationalizing, 48
- Rationalizing substitutions, 37
- Ray, 49
- Real numbers, 49
- Rectifiability, 63
- Related rates, 20
- Removable discontinuity, 7
- Resultant vector, 50
- Riemann sum, 27, 28
- Right-hand rule, 54
- Rolle's theorem, 22
- Root test, 46

- Scalar product, 52
- Scalar projection, 53
- Second derivative test, 24
- Sequence, 42
- Series, 44
- Skew lines, 59
- Slant asymptote, 25
- Solid of revolution, 40
- Speed, 66
- Sphere, 48
- Spherical coordinates, 90
- Squeeze theorem, 6, 70
- Standard basis, 51

- Surface area, 89
- Surface of revolution, 42
- Symmetry, 24
- Tangent, 3
- Tangent plane, 72, 76
- Tangent vector, 65
- Tangential component, 66
- Taylor series, 47
- Terminal point, 49
- Test for divergence, 44
- Total change, 31
- Total differential, 72
- Trace, 60
- Trapezoid, 25
- Trigonometric integrals, 35
- Trigonometric substitution, 36
- Triple integrals, 89
- Triple product, 55
- Turning point, 12
- Unit tangent, 64
- Unit vectors, 51
- Vector, 49
- Vector addition, 49, 50
- Vector components, 50
- Vector equation, 58
- Vector field, 91
- Vector functions, 61
- Vector operator, 75
- Vector projection, 53
- Vectors and Space, 48
- Velocity, 66
- Vertical asymptote, 8, 25
- Wave equation, 71
- Work, 41