Applied Mathematics One

Dejen Ketema Mamo
Arba Minch University

Applied Mathematics IB **(Math-1051)**

Lecture Note

Version 1.0

By:- Dejen Ketema *

Arba Minch University

Department of Mathematics

December 31, 2016

"If I am anything, I have made myself so by hard work." Sir Isaac Newton

*dejen.ketema@amu.edu.et **Note**: it is the first version, if you have any comment and suggestion please contact me.

Contents

Introduction

These notes are intended to be a summary of the main ideas in course Math 1051: Applied Mathematics One . I may keep working on this document as the course goes on, so these notes will not be completely finished it is on progress. The textbook for this course is Robert Ellis But James Stewart, Linear algebra books and Wiki-books are main sources of this document.

If you find any typos or errors,or you have any suggestions, please, do not hesitate to let me know. You may email me,or use the web form for feedback on the web pages for the course in the AMU E-learning cite.

CHAPTER ONE

1 Vectors and Vector Spaces

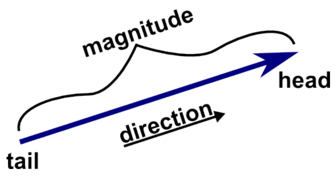
1.1 Scalars and Vectors

The study of Mitchel Joens is intimately tied in with the study of mathematics. Sometimes, the direction of a number or quantity is as important as the number itself. Mathematicians in the 19th century developed a convenient way of describing and interacting with quantities with and without direction by dividing them into two types: scalar quantities and vector quantities.

A physical quantity is expressed as the product of a numerical value and a physical unit, not merely a number. The quantity does not depend on the unit (e.g. for distance, 1 km is the same as 1000 m), although the number depends on the unit. Thus, following the example of distance, the quantity does not depend on the length of the base vectors of the coordinate system. Also, other changes of the coordinate system may affect the formula for computing the scalar (for example, the Euclidean formula for distance in terms of coordinates relies on the basis being orthonormal), but not the scalar itself. In this sense, physical distance deviates from the definition of metric in not being just a real number; however it satisfies all other properties. The same applies for other physical quantities which are not dimensionless. Direction does not apply to scalars; they are specified by magnitude or quantity alone.

Definition 1.1. *Quantities which can be described by magnitude only are called scalars. Scalars are represented by a single letter, such as a. Some examples of scalar quantities are mass (five kilograms), temperature (twenty-two degrees Celsius), and numbers without units (such as three).*

Definition 1.2. *A vector is an object that has both a magnitude and a direction. Geometrically, we can picture a vector as a directed line segment, whose length is the magnitude of the vector and with an arrow indicating the direction. The direction of the vector is from its tail to its head.*



Two vectors are the same if they have the same magnitude and direction. This means that if we take a vector and translate it to a new position (without rotating it), then the vector we obtain at the end of this process is the same vector we had in the beginning.

Two examples of vectors are those that represent force and velocity. Both force and velocity are in a particular direction. The magnitude of the vector would indicate the strength of the force or the speed associated with the velocity.

We denote vectors using boldface as in **a** or **b**. Especially when writing by hand where one cannot easily write in boldface, people will sometimes denote vectors using arrows as in $\vec{a}$ or $\vec{b}$, or they use other markings. We won't need to use arrows here. We denote the magnitude of the vector **a** by $\|a\|$. When we want to refer to a number and stress that it is not a vector, we can call the number a scalar. We will denote scalars with italics, as in **a** or **b**.

There is one important exception to vectors having a direction. The zero vector, denoted by a boldface **0**, is the vector of zero length. Since it has no length, it is not pointing in any particular direction. There is only one vector of zero length, so we can speak of the zero vector.

1.2 Operations on vectors

We can define a number of operations on vectors geometrically without reference to any coordinate system. Here we define addition, subtraction, and multiplication by a scalar. On separate pages, we discuss two different ways to multiply two vectors together: the dot product and the cross product.

1.2.1 Addition of vectors

Given two vectors **a** and **b** , we form their sum $\mathbf{a} + \mathbf{b}$, as follows. We translate the vector **b** until its tail coincides with the head of **a** . (Recall such translation does not change a vector.) Then, the directed line segment from the tail of $\mathbf{a}$ to the head of $\mathbf{b}$ is the vector $\mathbf{a} + \mathbf{b}$. The vector addition is the way forces and velocities combine.

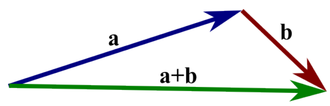


Figure 1.1: The sum of two vectors

Example 1.1. *If a car is traveling due north at 20 miles per hour and a child in the back seat behind the driver throws an object at 20 miles per hour toward his sibling who is sitting due east of him, then the velocity of the object (relative to the ground!) will be in a north-easterly direction. The velocity vectors form a right triangle, where the total velocity is the hypotenuse. Therefore, the total speed of the object (i.e., the magnitude of the velocity vector) is* $\sqrt{20^2 + 20^2} = 20\sqrt{2}$ *miles per hour relative to the ground.*

Addition of vectors satisfies two important properties.

1. The commutative law, which states the order of addition doesn't matter:

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}.$$

   This law is also called the parallelogram law, as illustrated in the below image. Two of the edges of the parallelogram define $\mathbf{a} + \mathbf{b}$, and the other pair of edges define $\mathbf{b} + \mathbf{a}$. But, both sums are equal to the same diagonal of the parallelogram.

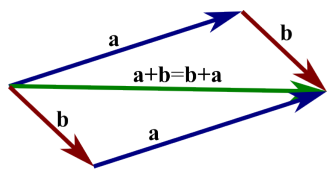


Figure 1.2: The parallelogram law, or commutative law, of vector addition

2. The associative law, which states that the sum of three vectors does not depend on which pair of vectors is added first:

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}).$$

1.2.2 Vector subtraction

Before we define subtraction, we define the vector $-\mathbf{a}$, which is the opposite of $\mathbf{a}$. The vector $\mathbf{a}$ is the vector with the same magnitude as $\mathbf{a}$ but that is pointed in the opposite direction.

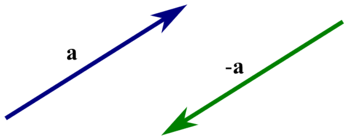


Figure 1.3: The opposite vector

We define subtraction as addition with the opposite of a vector: $\mathbf{b} - \mathbf{a} = \mathbf{b} + (-\mathbf{a})$. This is equivalent to turning vector a around in the applying the above rules for addition. Can you see how the vector $\mathbf{x}$ in the below figure is equal to $\mathbf{b} - \mathbf{a}$? Notice how this is the same as stating that $\mathbf{a} + \mathbf{x} = \mathbf{b}$, just like with subtraction of scalar numbers.

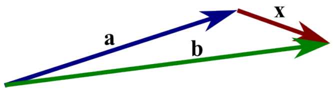


Figure 1.4: The difference of two vectors

1.2.3 Scalar multiplication

Given a vector $\mathbf{a}$ and a real number (scalar) λ, we can form the vector $\lambda\mathbf{a}$ as follows. If λ is positive, then $\lambda\mathbf{a}$ is the vector whose direction is the same as the direction of a and whose length is λ times the length of $\mathbf{a}$. In this case, multiplication by λ simply stretches (if $\lambda > \mathbf{1}$) or compresses (if $\mathbf{0} < \lambda < \mathbf{1}$) the vector $\mathbf{a}$.

If, on the other hand, λ is negative, then we have to take the opposite of $\mathbf{a}$ before stretching or compressing it. In other words, the vector $\lambda\mathbf{a}$ points in the opposite direction of $\mathbf{a}$, and the length of $\lambda\mathbf{a}$ is $|\lambda|$ times the length of $\mathbf{a}$. No matter the sign of λ, we observe that the magnitude of $\lambda\mathbf{a}$ is $|\lambda|$ times the magnitude of $\mathbf{a}$: $\|\lambda a\| = |\lambda|\|a\|$.

Scalar multiplications satisfies many of the same properties as the usual multiplication.

1. $\mathbf{s}(\mathbf{a} + \mathbf{b}) = \mathbf{sa} + \mathbf{sb}$ (distributive law, form 1)
2. $(\mathbf{s} + \mathbf{t})\mathbf{a} = \mathbf{sa} + \mathbf{ta}$ (distributive law, form 2)
3. $\mathbf{1a} = \mathbf{a}$
4. $(-\mathbf{1})\mathbf{a} = -\mathbf{a}$
5. $\mathbf{0a} = \mathbf{0}$

In the last formula, the zero on the left is the number **0**, while the zero on the right is the vector **0** , which is the unique vector whose length is zero.

If **a**= λ**b** for some scalar λ, then we say that the vectors a and b are parallel. If λ is negative, some people say that **a** and **b** are anti-parallel, but we will not use that language.

We were able to describe vectors, vector addition, vector subtraction, and scalar multiplication without reference to any coordinate system. The advantage of such purely geometric reasoning is that our results hold generally, independent of any coordinate system in which the vectors live. However, sometimes it is useful to express vectors in terms of coordinates, as discussed in a page about vectors in the standard Cartesian coordinate systems in the plane and in three-dimensional space.

1.3 Vectors in two- and three-dimensional Cartesian coordinates

In the introduction to vectors, we discussed vectors without reference to any coordinate system. By working with just the geometric definition of the magnitude and direction of vectors, we were able to define operations such as addition, subtraction, and multiplication by scalars. We also discussed the properties of these operation.

Often a coordinate system is helpful because it can be easier to manipulate the coordinates of a vector rather than manipulating its magnitude and direction directly. When we express a vector in a coordinate system, we identify a vector with a list of numbers, called coordinates or components, that specify the geometry of the vector in terms of the coordinate system. Here we will discuss the standard Cartesian coordinate systems in the plane and in three-dimensional space.

1.3.1 Vectors in the plane

We assume that you are familiar with the standard $(\mathbf{x}, \mathbf{y})$ Cartesian coordinate system in the plane. Each point **p** in the plane is identified with its **x** and **y** components: $\mathbf{p} = (\mathbf{p_1}, \mathbf{p_2})$.

To determine the coordinates of a vector **a** in the plane, the first step is to translate the vector so that its tail is at the origin of the coordinate system. Then, the head of the vector will be at some point $(\mathbf{a_1}, \mathbf{a_2})$ in the plane. We call $(\mathbf{a_1}, \mathbf{a_2})$ the coordinates or the components of the vector **a**. We often write $\mathbf{a} \in \mathbb{R}^2$ to denote that it can be described by two real coordinates. Using the Pythagorean Theorem, we can obtain an expression for the magnitude of a vector in terms of its components. Given a vector $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2})$, the vector is

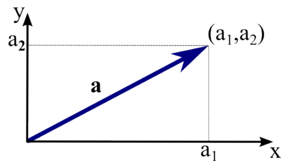


Figure 1.5: The coordinates of a vector in two dimensions

the hypotenuse of a right triangle whose legs are length $\mathbf{a_1}$ and $\mathbf{a_2}$. Hence, the length of the vector $\mathbf{a}$ is

$$\|\mathbf{a}\| = \sqrt{\mathbf{a_1^2} + \mathbf{a_2^2}}.$$

Example 1.2. *Consider the vector* $\mathbf{a}$ *represented by the line segment which goes from the point* $(1, 2)$ *to the point* $(4, 6)$*. Can you calculate the coordinates and the length of this vector?*

To find the coordinates, translate the line segment one unit left and two units down. The line segment begins at the origin and ends at $(4 - 1, 6 - 2) = (3, 4)$*. Therefore,* $\mathbf{a} = (\mathbf{3}, \mathbf{4})$*. The length of* $\mathbf{a}$ *is* $\|a\| = \sqrt{3^2 + 4^2} = 5$*.*

The vector operations we defined in the vector introduction are easy to express in terms of these coordinates. If $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2})$ and $\mathbf{b} = (\mathbf{b_1}, \mathbf{b_2})$, their sum is simply $\mathbf{a} + \mathbf{b} = (\mathbf{a_1} + \mathbf{b_1}, \mathbf{a_2} + \mathbf{b_2})$, as illustrated in the below figure. It is also easy to see that $\mathbf{b} - \mathbf{a} = (\mathbf{b_1} - \mathbf{a_1}, \mathbf{b_2} - \mathbf{a_2})$ and $\lambda\mathbf{a} = (\lambda\mathbf{a_1}, \lambda\mathbf{a_2})$ for any scalar λ. You may have noticed that we use the same notation to

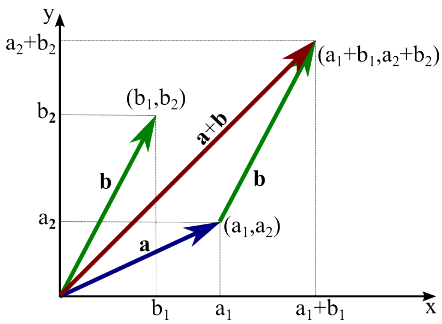


Figure 1.6: Adding two-dimensional vectors

denote a point and to denote a vector. We don't tend to emphasize any distinction between a point and a vector. You can think of a point as being represented by a vector whose tail is fixed at the origin. You'll have to figure out by context whether or not we are thinking of a vector as having its tail fixed at the origin.

Another way to denote vectors is in terms of the standard unit vectors denoted $\mathbf{i}$ and $\mathbf{j}$. A unit vector is a vector whose length is one. The vector $\mathbf{i}$ is the unit vector in the direction of the positive x-axis. In coordinates, we can write $\mathbf{i} = (\mathbf{1}, \mathbf{0})$. Similarly, the vector $\mathbf{j}$ is the unit vector in the direction of the positive y-axis: $\mathbf{j} = (\mathbf{0}, \mathbf{1})$. We can write any two-dimensional vector in terms of these unit vectors as $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2}) = \mathbf{a_1 i} + \mathbf{a_2 j}$.

1.3.2 Vectors in three-dimensional space

In three-dimensional space, there is a standard Cartesian coordinate system $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. Starting with a point which we call the origin, construct three mutually perpendicular axes, which we call the x-axis, the y-axis, and the z-axis. Here is one way to picture these axes. Stand near the corner of a room and look down at the point where the walls meet the floor. Then, the floor and the wall to your left intersect in a line which is the positive x-axis. The floor and the wall to your right intersect in a line which is the positive y-axis. The walls intersect in a vertical line which is the positive z-axis. These positive axes are depicted in the below applet labeled by **x**, **y** & **z**. The negative part of each axis is on the opposite side of the origin, where the axes intersect.

We have set the relative locations of the positive x, y, and z-axis to make the coordinate system a right-handed coordinate system. Note that if you curl the fingers of your right hand from the positive x-axis to the positive y-axis, the thumb of your right hand points in the direction of the positive z-axis.

If you switched the locations of the positive x-axis and positive y-axis, then you would end up having a left-handed coordinate system. If you do that, you will be living in a mathematical universe in which some formulas will differ by a minus sign from the formula in the universe we are using here. Your universe will be just as valid as ours, but there will be lots of confusion. We suggest you live in our universe while studying from these pages.

With these axes any point **p** in space can be assigned three coordinates $\mathbf{p} = (\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3})$. For example, given the above corner-of-room analogy, suppose you start at the corner of the room and move four meters along the x-axis, then turn left and walk three meters into the room. If you are two meters tall, then the top of your head is at the point $(4, 3, 2)$.

Just as in two-dimensions, we assign coordinates of a vector **a** by translating its tail to the origin and finding the coordinates of the point at its head. In this way, we can write the vector as $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3})$. We often write $\mathbf{a} \in \mathbb{R}^3$ to denote that it can be described by three real coordinates. Sums, differences, and scalar multiples of three-dimensional vectors are all performed on each component. If $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3})$ and $\mathbf{b} = (\mathbf{b_1}, \mathbf{b_2}, \mathbf{b_3})$, then $\mathbf{a} + \mathbf{b} = (\mathbf{a_1} + \mathbf{b_1}, \mathbf{a_2} + \mathbf{b_2}, \mathbf{a_3} + \mathbf{b_3}), \mathbf{b} - \mathbf{a} = (\mathbf{b_1} - \mathbf{a_1}, \mathbf{b_2} - \mathbf{a_2}, \mathbf{b_3} - \mathbf{a_3}), \& \lambda\mathbf{a} = (\lambda\mathbf{a_1}, \lambda\mathbf{a_2}, \lambda\mathbf{a_3})$.

Just as in two dimensions, we can also denote three-dimensional vectors is in terms

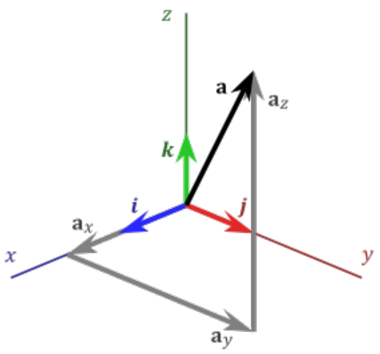


Figure 1.7: Three-dimensional vectors

of the standard unit vectors, $\mathbf{i}, \mathbf{j}, \&\mathbf{k}$. These vectors are the unit vectors in the positive x, y, and z direction, respectively. In terms of coordinates, we can write them as $\mathbf{i} = (\mathbf{1}, \mathbf{0}, \mathbf{0})$, $\mathbf{j} = (\mathbf{0}, \mathbf{1}, \mathbf{0})$, & $\mathbf{k} = (\mathbf{0}, \mathbf{0}, \mathbf{1})$. We can express any three-dimensional vector as a sum of scalar multiples of these unit vectors in the form $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3}) = \mathbf{a_1 i} + \mathbf{a_2 j} + \mathbf{a_3 k}$.

What is the length of the vector $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3})$? We can decompose the vector into $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3}) = (\mathbf{a_1}, \mathbf{a_2}, \mathbf{0}) + (\mathbf{0}, \mathbf{0}, \mathbf{a_3})$, where the two vectors on the right hand side correspond to the two green line segments in the above applet. These two line segments form a right triangle whose hypotenuse is the vector a (the blue line segment). The first vector can be thought of as a two dimensional vector, so its length is $\|(\mathbf{a_1}, \mathbf{a_2}, \mathbf{0})\| = \|(\mathbf{a_1}, \mathbf{a_2})\| = \sqrt{\mathbf{a_1^2} + \mathbf{a_2^2}}$. The second vector's length is $\|(\mathbf{0}, \mathbf{0}, \mathbf{a_3})\| = |\mathbf{a3}|$. Therefore, by the Pythagorean Theorem, the length of $\mathbf{a}$ is $\|\mathbf{a}\| = \sqrt{\|(\mathbf{a_1}, \mathbf{a_2}, \mathbf{0})\|^2 + \|(\mathbf{0}, \mathbf{0}, \mathbf{a_3})\|^2} = \sqrt{\mathbf{a_1^2} + \mathbf{a_2^2} + \mathbf{a_3^2}}$.

Note:-

1. Two vectors are said to be equal iff they have equal magnitude and the same direction. Equivalently they will be equal if their coordinates are equal. So two vectors $\vec{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ and $\vec{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3$ are equal iff $a_1 = b_1, \quad a_2 = b_2, \quad a_3 = b_3$.

2. Two vectors are opposite iff they have opposite direction but not necessarily the same magnitude.
$\vec{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ and $\vec{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3$ are opposite iff $a_1 = -b_1, \quad a_2 = -b_2, \quad a_3 = -b_3.$,

3. Two vectors are parallel if one is the scalar multiple of the other.

Example 1.3. *1. Add the following two vectors*

(a) $\vec{u} = (-1, 2, 1)$ *and* $\vec{v} = (2, -5, 0)$

(b) $\vec{u} = i + 4j - 2k$ *and* $\vec{v} = 2i - 5j$

2. *Let* $\vec{u} = 3i + 4j - k$, *then find*

   (a) $\frac{2}{3}\vec{u}$

   (b) $-1\vec{u}$

solution:

1. (a) $\vec{u} + \vec{v} = (-1, 2, 1) + (2, -5, 0)$
   $= (-1 + 2, 2 - 5, 1 + 0)$
   $= (1, -3, 1)$

   (b) $\vec{u} + \vec{v} = (i + 4j - k) + (2i - 5j)$
   $= (1 + 2)i + (4 - 5)j + (-1 + 0)k$

   $= 3i - j - k$

2. (a) $\frac{2}{3}\vec{u} = \frac{2}{3}(3i + 4j - k)$
   $= 2i + \frac{8}{3}j - \frac{2}{3}k$

   (b) $-1\vec{u} = -(3i + 4j - k)$
   $= -3i - 4j + k$

Theorem 1.1. *:- Let $\vec{u}$, $\vec{v}$, and $\vec{w}$ are any vectors and α and β are any scalars then the following holds true:*

1. $\vec{v} + \vec{w} = \vec{w} + \vec{v}$.

2. $(\vec{v} + \vec{w}) + \vec{u} = \vec{v} + (\vec{w} + \vec{u})$.

3. $\vec{v} + \vec{0} = \vec{v}$.

4. $\vec{v} + \vec{-v} = \vec{0}$.

5. $(\alpha\beta)\vec{v} = \alpha(\beta\vec{v})$

6. $(\alpha + \beta) \cdot \vec{v} = \alpha \cdot \vec{v} + \beta \cdot \vec{v}$.

7. $\alpha \cdot (\vec{v} + \vec{w}) = \alpha \cdot \vec{v} + \alpha \cdot \vec{w}$.

8. $1 \cdot \vec{v} = \vec{v}$.

Proof Exercise

1.4 Dot product

The dot product between two vectors is based on the projection of one vector onto another. Let's imagine we have two vectors **a** and **b**, and we want to calculate how much of **a** is pointing in the same direction as the vector **b**. We want a quantity that would be positive if the two vectors are pointing in similar directions, zero if they are perpendicular, and negative if the two vectors are pointing in nearly opposite directions. We will define the dot product between the vectors to capture these quantities.

But first, notice that the question "how much of **a** is pointing in the same direction as the vector **b**" does not have anything to do with the magnitude (or length) of **b**; it is based only on its direction. (Recall that a vector has a magnitude and a direction.) The answer to this question should not depend on the magnitude of **b**, only its direction. To sidestep any confusion caused by the magnitude of **b**, let's scale the vector so that it has length one. In other words, let's replace **b** with the unit vector that points in the same direction as **b**. We'll call this vector **u**, which is defined by

$$\mathbf{u} = \frac{\mathbf{b}}{\|\mathbf{b}\|}.$$

The dot product of **a** with unit vector **u**, denoted $\mathbf{a}.\mathbf{u}$, is defined to be the projection of **a** in the direction of **u**, or the amount that **a** is pointing in the same direction as unit vector **u**. Let's assume for a moment that **a** and **u** are pointing in similar directions. Then, you can imagine $\mathbf{a}.\mathbf{u}$ as the length of the shadow of **a** onto **u** if their tails were together and the sun was shining from a direction perpendicular to **u**. By forming a right triangle with **a** and this shadow, you can use geometry to calculate that

$$\mathbf{a}.\mathbf{u} = \|\mathbf{a}\| \cos\theta, \tag{1.1}$$

where θ is the angle between **a** and **u**.

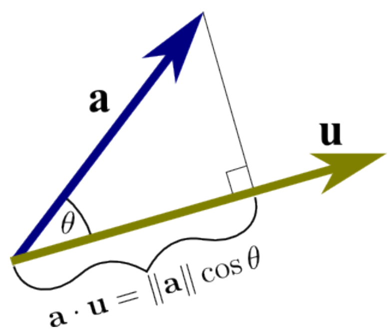


Figure 1.8: Dot product as projection onto a unit vector

If **a** and **u** were perpendicular, there would be no shadow. That corresponds to the case when $\cos\theta = \cos\pi/2 = 0$ and $\mathbf{a}.\mathbf{u} = \mathbf{0}$. If the angle θ between **a** and **u** were larger than

$\pi/2$, then the shadow wouldn't hit **u**. Since in this case $\cos\theta < 0$, the dot product**a**.**u** is also negative. You could think of $-\mathbf{a}.\mathbf{u}$ (which is positive in this case) as being the length of the shadow of **a** on the vector $-\mathbf{u}$, which points in the opposite direction of **u**.

But we need to get back to the dot product **a**.**u**, where **b** may have a magnitude different than one. This dot product **a**.**b** should depend on the magnitude of both vectors, $\|\mathbf{a}\|$ and $\|\mathbf{b}\|$, and be symmetric in those vectors. Hence, we don't want to define **a**.**b** to be exactly the projection of **a** on **b**; we want it to reduce to this projection for the case when **b** is a unit vector. We can accomplish this very easily: just plug the definition $\mathbf{u} = \dfrac{\mathbf{b}}{\|\mathbf{b}\|}$ into our dot product definition of equation 1.1. This leads to the definition that the dot product **a**.**b**, divided by the magnitude $\|\mathbf{b}\|$ of **b**, is the projection of **a** onto **b**.

$$\frac{\mathbf{a}.\mathbf{b}}{\|\mathbf{b}\|} = \|\mathbf{a}\| \cos\theta.$$

Then, if we multiply by through by $\|\mathbf{b}\|$, we get a nice symmetric definition for the dot product **a**.**b**.

$$\mathbf{a}.\mathbf{b} = \|\mathbf{a}\|.\|\mathbf{b}\| \cos\theta. \tag{1.2}$$

The picture of the geometric interpretation of **a**.**b** is almost identical to the above picture for **a**.**u** . We just have to remember that we have to divide through by $\|\mathbf{b}\|$ to get the projection of **a** onto **b** .

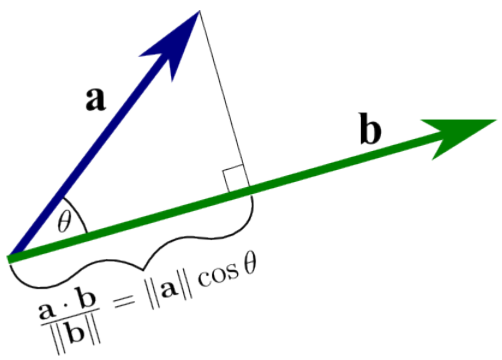


Figure 1.9: Dot product as projection of vectors

The geometric definition of equation 1.2 makes the properties of the dot product clear. One can see immediately from the formula that the dot product **a**.**b** is positive for acute angles and negative for obtuse angles. The formula demonstrates that the dot product grows linearly with the length of both vectors and is commutative, i.e., $\mathbf{a}.\mathbf{b} = \mathbf{b}.\mathbf{a}$.

However, the geometric formula 1.2 is not convenient for calculating the dot product when we are given the vectors **a** and **b** in terms of their components. To facilitate such calculations, we derive a formula for the dot product in terms of vector components. With such formula in hand, we can run through examples of calculating the dot product.

The formula for the dot product in terms of vector components

The geometric definition of the dot product says that the dot product between two vectors **a** and **b** is

$$\mathbf{a}.\mathbf{b} = \|\mathbf{a}\|.\|\mathbf{b}\| \cos\theta,$$

where θ is the angle between vectors **a** and **b**. Although this formula is nice for understanding the properties of the dot product, a formula for the dot product in terms of vector components would make it easier to calculate the dot product between two given vectors.

Since the standard unit vectors are orthogonal, we immediately conclude that the dot product between a pair of distinct standard unit vectors is zero:

$$i.j = i.k = j.k = 0.$$

The dot product between a unit vector and itself is also simple to compute. In this case, the angle is zero and $\cos\theta = 1$. Given that the vectors are all of length one, the dot products are

$$i.i = j.j = k.k = 1.$$

The second step is to calculate the dot product between two three-dimensional vectors

$$a = (a_1, a_2, a_3) = a_1 i + a_2 j + a_3 k$$

$$b = (b_1, b_2, b_3) = b_1 i + b_2 j + b_3 k.$$

To do this, we simply assert that for any three vectors **a**, **b**, and **c**, and any scalar λ,

$$(\lambda a).b = \lambda(a.b)$$

$$(a + b).c = a.c + b.c.$$

(These properties mean that the dot product is linear.)

Given these properties and the fact that the dot product is commutative, we can expand the dot product **a**.**b** in terms of components,

$$\mathbf{a}.\mathbf{b} = (a_1 i + a_2 j + a_3 k).(b_1 i + b_2 j + b_3 k)$$

$$= a_1 b_1 i.i + a_2 b_2 j.j + a_3 b_3 k.k$$

$$+(a_1b_2 + a_2b_1)i.j + (a_1b_3 + a_3b_1)i.k + (a_2b_3 + a_3b_2)j.k.$$

Since we know the dot product of unit vectors, we can simplify the dot product formula to

$$a.b = a_1b_1 + a_2b_2 + a_3b_3. \tag{1.3}$$

Equation 1.3 makes it simple to calculate the dot product of two three-dimensional vectors, $\mathbf{a},\ \mathbf{b} \in \mathbb{R}^3$. The corresponding equation for vectors in the plane, $\mathbf{a},\ \mathbf{b} \in \mathbb{R}^2$, is even simpler. Given

$$a = (a_1, a_2) = a_1i + a_2j$$

$$b = (b_1, b_2) = b_1i + b_2j,$$

we can use the same formula, but with $a_3 = b_3 = 0$,

$$a.b = a_1b_1 + a_2b_2 \tag{1.4}$$

Note:-

1. The cosine of the angle between two non-zero vectors $\vec{a}$ and $\vec{b}$ is

$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\|\vec{b}\|}$$
$$\implies \theta = \cos^{-1}\left(\frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\|\vec{b}\|}\right)$$

2. If an angle between two vectors $\vec{a}$ and $\vec{a}$ is $\frac{\pi}{2}$ then $\vec{a}\cdot\vec{a} = 0$ and the vectors $\vec{a}$ and $\vec{b}$ are known as orthogonal vectors. In general the dot product of two non-zero vectors is zero only if they are orthogonal.

Remark:- Suppose $\vec{a}$ and $\vec{b}$ are two non-zero vectors and θ is the angle between $\vec{a}$ and $\vec{b}$:

i) if θ is acute, then $\vec{a}\cdot\vec{b} > 0$.

ii) if θ is obtuse, then $\vec{a}\cdot\vec{b} < 0$.

iii) if θ is right angle, then $\vec{a}\cdot\vec{b} = 0$.

Theorem 1.2. *properties of dot product and magnitude*
If $\vec{v}$, $\vec{u}$ and $\vec{w}$ vectors in $\vec{V}$ in plane or space and α is scalar,then the following property holds true.

1. $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$ *commutative property*

2. $\vec{u}\cdot(\vec{v}+\vec{w}) = (\vec{u}\cdot\vec{v}) + (\vec{u}\cdot\vec{w})$ *distributive property of dot product over vector addition.*

3. *$\alpha(\vec{u}\cdot\vec{v}) = (\alpha\vec{u})\cdot\vec{v} = \vec{u}\cdot(\alpha\vec{v})$ associative property of scalar multiplication and dot product.*

4. *$\mathbf{0}\cdot\vec{u} = \mathbf{0}$ zero vector dot with other vector.*

5. *$\vec{u}\cdot\vec{u} = \|\vec{u}\|^2$ the relation between the dot product and magnitude.*

6. *$\vec{u} = \mathbf{0}$ if and only if $\|\vec{u}\| = 0$ property of zero vector.*

7. $\|\alpha\vec{u}\| = |\alpha|\,\|\vec{u}\|$
   Proof exercise

Armed with equations 1.3 and 1.4, you can make short work of calculating dot products, as shown in these examples.

Example 1.4. *Calculate the dot product of $a = (1, 2, 3)$ and $b = (4, -5, 6)$. Do the vectors form an acute angle, right angle, or obtuse angle?*
***Solution**: Using the component formula for the dot product of three-dimensional vectors, $a.b = a_1b_1 + a_2b_2 + a_3b_3$, we calculate the dot product to be $a.b = 1(4) + 2(-5) + 3(6) = 4 - 10 + 18 = 12$.*

Since a.b is positive, we can infer from the geometric definition, that the vectors form an acute angle.

Example 1.5. *Calculate the dot product of $c = (-4, -9)$ and $d = (-1, 2)$. Do the vectors form an acute angle, right angle, or obtuse angle?*

***Solution**: Using the component formula for the dot product of two-dimensional vectors, $a.b = a_1b_1 + a_2b_2$, we calculate the dot product to be $c.d = -4(-1) - 9(2) = 4 - 18 = -14$.*

Since c.d is negative, we can infer from the geometric definition, that the vectors form an obtuse angle.

Example 1.6. *If $a = (6, -1, 3)$, for what value of c is the vector $b = (4, c, -2)$ perpendicular to a?*

***Solution**: For a and b to be perpendicular, we need their dot product to be zero. Since $a.b = 6(4) - 1(c) + 3(-2) = 24 - c - 6 = 18 - c$, the number c must satisfy $18 - c = 0$, or $c = 18$. You can double-check that the vector $b = (4, 18, -2)$ is indeed perpendicular to a by verifying that $a.b = (6, -1, 3).(4, 18, -2) = 0$.*

Exercise

1. *Find the value of* $\boldsymbol{k}$ *such that the vectors* $(2, 3, -2)$ *and* $(5, k, 2)$ *are perpendicular.*

2. *Find the values of* β *such that the vectors* $(\beta, -3, 1)$ *and* $(\beta, \beta, 2)$ *are perpendicular.*

Example 1.7. 1. *Find the component form and the length of vector with initial point and terminal point* $p = (3, -7, 1)$ *and* $Q = (-2, 5, 3)$

2. *Find the unit vector in the direction of* $\vec{v} = (-2, 5, 3)$

3. *Let* $\vec{u} = (2, 5, -1)$ *and* $\vec{v} = (-1, 2, 3)$*,the find the scalar product of* $\vec{u}$ *and* $\vec{u}$

Solution

1. $\vec{v} = \vec{PQ} = (-2-3, 5+7, 3-1) = (-5, 12, 2)$
   and $\|\vec{v}\| = \sqrt{(-5)^2 + (12)^2 + 2^2} = \sqrt{173}$

2. $\|\vec{v}\| = \sqrt{-2^2 + 5^2 + 3^2} = \sqrt{38}$
   $\hat{u} = \frac{\vec{v}}{\|\vec{v}\|} = \dfrac{1}{\sqrt{38}}(-2, 5, 3)$
   $= (\dfrac{-2}{\sqrt{38}}, \dfrac{5}{\sqrt{38}}, \dfrac{3}{\sqrt{38}})$

3. $\vec{u} \cdot \vec{v} = (2, 5, -1) \cdot (-1, 2, 3)$
   $= 2. - 1 + 5.2 + -1.3$
   $= -2 + 10 - 3$
   $= 5$

Example 1.8. 1. *Find the cosine of the angle and the angle between the following vectors*

   (a) $\vec{u} = 2i - 5j + k$ *and* $\vec{v} = i + j + 2k$

   (b) $\vec{v} = (-\sqrt{3}, 1)$ *and* $\vec{w} = (0, 5)$

   (c) $\vec{v} = (0, -1, 0)$ *and* $\vec{w} = (0, 0, 3)$

Solution:- exercise

1.4.1 Orthogonal projection

Suppose that two non-zero vectors **A** and **U** are positioned as in the figure below and that the light casts a shadow on the line containing the vector **U**. Informally we think of the shadow as determining a vector parallel to vector **U**, we call this vector the projection of vector **A** on to vector **U** which is denoted by $proj_{\mathbf{U}}\ {}^{\mathbf{A}}$ and defined as $proj_U\ {}^A = (\frac{U \cdot A}{\|U\|^2})U$.

Note:- The scalar projection of vector A along the vector **U** is $\left\|proj_{\mathbf{U}}\ {}^{\mathbf{A}}\right\| = |\frac{\vec{U} \cdot \vec{A}}{\|\vec{U}\|}|$

Figure 1.10: Projections with acute and obtuse angles

Example 1.9. *1. Let vector $\vec{u} = 2i - j + 3k$, $\quad \vec{v} = i + 2j + k$, then find:*

(a) $proj_{\vec{v}}\ \vec{u}$

(b) $proj_{\vec{u}}\ \vec{v}$

(c) $proj_{\frac{1}{2}\vec{v}}\ 3\vec{u}$

(d) *The scalar projection of* $proj_{\vec{v}}\ \vec{u}$ *and* $proj_{\vec{u}}^{\vec{v}}$

Solution:-

1. (a)

$$\begin{aligned} proj_{\vec{v}}^{\vec{u}} &= \left(\frac{\vec{v}\cdot\vec{u}}{\|\vec{v}\|^2}\right)\vec{v} \\ &= \left(\frac{(i+2j+k)\cdot(2i-j+3k)}{\|i+2j+k\|^2}\right)(i+2j+k) \\ &= (\frac{2-2+3}{1^2+2^2+1^2})(i+2j+k) \\ &= \frac{3}{6}(i+2j+k) \\ &= \frac{1}{2}i + j + \frac{1}{2}k \end{aligned}$$

(b)

$$\begin{aligned} proj_{\vec{u}}^{\vec{v}} &= \left(\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|^2}\right)\vec{u} \\ &= (\frac{(2i-j+3k)\cdot(i+2j+k)}{\|2i-j+3k\|^2})(2i-j+3k) \\ &= (\frac{2-2+3}{1^2+(-2)^2+3^2})(2i-j+3k) \\ &= \frac{3}{14}(2i-j+3k) \\ &= \frac{6}{14}i - \frac{3}{14}j + \frac{9}{14}k \end{aligned}$$

(c)

$$\begin{aligned}
proj_{\frac{1}{2}\vec{v}}^{3\vec{u}} &= \left(\frac{\frac{1}{2}\vec{v}\cdot 3\vec{u}}{\|\frac{1}{2}\vec{v}\|^2}\right)\frac{1}{2}\vec{v} \\
&= (\frac{\frac{1}{2}(i+2j+k)\cdot 3(2i-j+3k)}{\frac{1}{4}\left\|i+2j+k\right\|^2})\frac{1}{2}(i+2j+k) \\
&= (\frac{\frac{3}{2}(2-2+3)}{\frac{1}{4}(1^2+2^2+1^2)})\frac{1}{2}(i+2j+k) \\
&= 3(i+2j+k) \\
&= 3i+6j+3k
\end{aligned}$$

(d)

$$\begin{aligned}
\|proj_{\vec{v}}\ ^{\vec{u}}\| &= |(\frac{\vec{v}\cdot\vec{u}}{\|\vec{v}\|})| \\
&= |(\frac{(i+2j+k)\cdot(2i-j+3k)}{\|i+2j+k\|})| \\
&= |(\frac{2-2+3}{\sqrt{1^2+2^2+1^2}})| \\
&= \frac{3}{\sqrt{6}} \\
\&\|proj_{\vec{u}}\ ^{\vec{v}}\| &= |(\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|})| \\
&= |(\frac{(2i-j+3k)\cdot(i+2j+k)}{\|2i-j+3k\|})| \\
&= |\frac{2-2+3}{\sqrt{1^2+(-2)^2+3^2}}| \\
&\frac{3}{\sqrt{14}}
\end{aligned}$$

1.4.2 Direction angles and direction cosine

Definition 1.3. *Let* **p** *be a vector and* α, β *and* γ *are angles that a vector* **p** *makes with the positive x, y and z axis respectively. These angles are called directional angles of the vector* **p** *and* $\cos\alpha$, $\cos\beta$ *and* $\cos\gamma$ *are directional cosines of vector* **p**.

Note:-

1. $\cos\alpha = \frac{x}{r}$, $\cos\beta = \frac{y}{r}$ and $\cos\gamma = \frac{z}{r}$ where $r = \sqrt{x^2+y^2+z^2}$.
2. $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$

Example 1.10. *1. consider the following vectors*

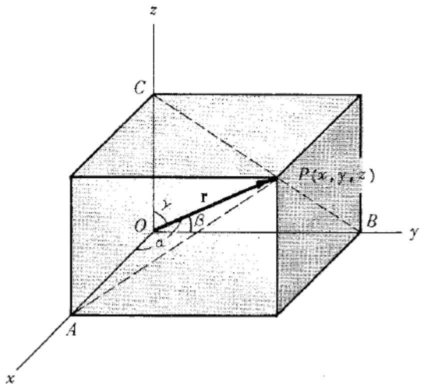


Figure 1.11: direction angeles

(a) $\vec{a} = (2, 3)$

(b) $\vec{b} = (1, 1, 1)$

(c) $\vec{c} = (3, 5, 2)$

Then find the direction angles and direction cosines

2. *If the direction angle of a vector $\vec{a}$ makes an angle of $\frac{\pi}{3}$ with the positive x-axis and an angle of $\frac{\pi}{4}$ with the positive y-axis then what angle will make with the positive z-axis.*

solution:- *Exercise*

1.5 Cross product

The cross product of two vectors a and b is defined only in three-dimensional space and is denoted by $a \times b$. In physics, sometimes the notation $a \bigwedge b$ is used, though this is avoided in mathematics to avoid confusion with the exterior product.

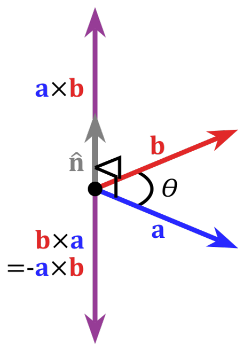


Figure 1.12: The cross-product in respect to a right-handed coordinate system

Definition 1.4. *The cross product $a \times b$ is defined as a vector c that is perpendicular to both a and b, with a direction given by the right-hand rule and a magnitude equal to the area of the parallelogram that the vectors span.*
The cross product is defined by the formula

$$\mathbf{a} \times \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \sin(\theta) \ \hat{\mathbf{n}} \tag{1.5}$$

where θ is the angle between a and b in the plane containing them (hence, it is between 0^o and 180^o), $\|a\|$ and $\|b\|$ are the magnitudes of vectors a and b, and n is a unit vector perpendicular to the plane containing a and b in the direction given by the right-hand rule (illustrated). If the vectors a and b are parallel (i.e., the angle $\hat{I}_{\text{¸}}$ between them is either 0^o or 180^o), by the above formula, the cross product of a and b is the zero vector 0.

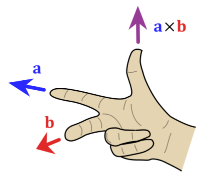


Figure 1.13: Finding the direction of the cross product by the right-hand rule

1.5.1 Computing the cross product

Coordinate notation

The standard basis vectors $i, j,$ and k satisfy the following equalities in a right hand coordinate system:

$$\mathbf{i} = \mathbf{j} \times \mathbf{k}$$
$$\mathbf{j} = \mathbf{k} \times \mathbf{i}$$
$$\mathbf{k} = \mathbf{i} \times \mathbf{j}$$

which imply, by the anticommutativity of the cross product, that

$$\mathbf{k} \times \mathbf{j} = -\mathbf{i}$$
$$\mathbf{i} \times \mathbf{k} = -\mathbf{j}$$
$$\mathbf{j} \times \mathbf{i} = -\mathbf{k}$$

The definition of the cross product also implies that

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0} (\text{the zero vector}).$$

These equalities, together with the distributivity and linearity of the cross product (but both do not follow easily from the definition given above), are sufficient to determine the cross product of any two vectors u and v. Each vector can be defined as the sum of three orthogonal components parallel to the standard basis vectors:

$$\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$$
$$\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$$

Their cross product $u \times v$ can be expanded using distributivity:

$$\begin{aligned}\mathbf{u}\times\mathbf{v} &= (u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}) \times (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k})\\ &= u_1v_1(\mathbf{i}\times\mathbf{i}) + u_1v_2(\mathbf{i}\times\mathbf{j}) + u_1v_3(\mathbf{i}\times\mathbf{k}) +\\ &\quad u_2v_1(\mathbf{j}\times\mathbf{i}) + u_2v_2(\mathbf{j}\times\mathbf{j}) + u_2v_3(\mathbf{j}\times\mathbf{k}) +\\ &\quad u_3v_1(\mathbf{k}\times\mathbf{i}) + u_3v_2(\mathbf{k}\times\mathbf{j}) + u_3v_3(\mathbf{k}\times\mathbf{k})\end{aligned}$$

This can be interpreted as the decomposition of $u \times v$ into the sum of nine simpler cross products involving vectors aligned with i, j, or k. Each one of these nine cross products operates on two vectors that are easy to handle as they are either parallel or orthogonal to each other. From this decomposition, by using the above-mentioned equalities and collecting similar terms, we obtain:

$$\begin{aligned}\mathbf{u}\times\mathbf{v} = &- u_1v_1\mathbf{0} + u_1v_2\mathbf{k} - u_1v_3\mathbf{j}\\ &- u_2v_1\mathbf{k} - u_2v_2\mathbf{0} + u_2v_3\mathbf{i}\\ &+ u_3v_1\mathbf{j} - u_3v_2\mathbf{i} - u_3v_3\mathbf{0}\\ = &\,(u_2v_3 - u_3v_2)\mathbf{i} + (u_3v_1 - u_1v_3)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}\end{aligned}$$

meaning that the three scalar components of the resulting vector $s = s_1i + s_2j + s_3k = u \times v$ are

$$s_1 = u_2v_3 - u_3v_2$$
$$s_2 = u_3v_1 - u_1v_3$$
$$s_3 = u_1v_2 - u_2v_1$$

Using column vectors, we can represent the same result as follows:

$$\begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}$$

Matrix notation

The cross product can also be expressed as the formal determinant:

$$\mathbf{u}\times\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

This determinant can be computed using Sarrus' rule or cofactor expansion. Using Sarrus' rule, it expands to

$$\begin{aligned}\mathbf{u}\times\mathbf{v} &= (u_2v_3\mathbf{i}+u_3v_1\mathbf{j}+u_1v_2\mathbf{k})-(u_3v_2\mathbf{i}+u_1v_3\mathbf{j}+u_2v_1\mathbf{k})\\ &= (u_2v_3-u_3v_2)\mathbf{i}+(u_3v_1-u_1v_3)\mathbf{j}+(u_1v_2-u_2v_1)\mathbf{k}.\end{aligned}$$

Using cofactor expansion along the first row instead, it expands to

$$\mathbf{u}\times\mathbf{v}=\begin{vmatrix}u_2 & u_3\\ v_2 & v_3\end{vmatrix}\mathbf{i}+\begin{vmatrix}u_3 & u_1\\ v_3 & v_1\end{vmatrix}\mathbf{j}+\begin{vmatrix}u_1 & u_2\\ v_1 & v_2\end{vmatrix}\mathbf{k}$$

which gives the components of the resulting vector directly.

Properties of the cross product

The following results are consequences of the definition of the cross product.

1. Cross product is not commutative

$$a\times b=-b\times a \tag{1.6}$$

2. The cross product is distributive over addition

$$a\times(b+c)=(a\times b)+(a\times c). \tag{1.7}$$

3. Parallel vectors ($\theta=0$): If vectors a and b are parallel, then

$$a\times b=0. \tag{1.8}$$

4. Orthogonal vectors ($\theta=\pi/2$): If vectors a and b are orthogonal, then

$$a\times b=\|a\|.\|b\|\hat{n}. \tag{1.9}$$

5. Product of unit vectors: If a and b are unit vectors, then

$$a\times b=\sin\theta\hat{n}. \tag{1.10}$$

The scalar triple product (also called the box product or mixed triple product) is not really a new operator, but a way of applying the other two multiplication operators to three vectors. The scalar triple product is sometimes denoted by (a b c) and defined as:

$$\mathbf{a}\ \cdot(\mathbf{b}\ \times\mathbf{c})=(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}. \tag{1.11}$$

Example 1.11. *Given that* $\mathbf{a} = 3i - 2j - k$ *and* $\mathbf{b} = i + 4j + 2k$*, find* $\mathbf{a} \times \mathbf{b}$ *and a unit vector* $\hat{n}$ *normal to the plane containing* $\mathbf{a}$ *and* $\mathbf{b}$ *such that* $\mathbf{a}$, $\mathbf{b}$*, and* $\hat{n}$*, in this order, obey the right-hand rule.*

Solution:-

$$\begin{aligned} a \times b &= \begin{vmatrix} i & j & k \\ 3 & 2 & -1 \\ 1 & 4 & 2 \end{vmatrix} \\ &= [(-2) \cdot 2 - 4 \cdot (-1)]i - [3 \cdot 2 - 1 \cdot (-1)]j + [3 \cdot 4 - 1 \cdot (-2)]k \\ &= -7j + 14k. \end{aligned}$$

The required unit vector $\hat{n}$ *is simply the unit vector in the direction of* $a \times b$*, so*

$$\begin{aligned} \hat{n} &= \frac{a \times b}{\|a \times b\|} \\ &= \frac{(-7j + 14k)}{7\sqrt{5}} \\ &= (\frac{-1}{\sqrt{5}})j + (\frac{2}{\sqrt{5}})k. \end{aligned}$$

1.6 Application of cross and dot product

The area of a parallelogram defined by the vectors $\vec{a}$ and $\vec{b}$ is determined by the formula:

$$\begin{aligned} A &= \|a\|\|b\| \sin \theta \\ &= \|a \times b\| \end{aligned}$$

where θ is an angle between $\vec{a}$ and $\vec{b}$

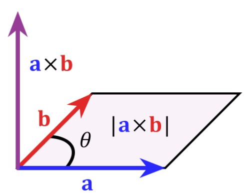


Figure 1.14: The area of a parallelogram as the magnitude of a cross product

Example 1.12. *Find the area of the parallelogram defined by the vectors*

$\vec{a} = (1, -1, 0)$ *and* $\vec{b} = (0, 1, 2)$

$$\begin{aligned}
\vec{a} \times \vec{b} &= \begin{vmatrix} i & j & k \\ 1 & -1 & 0 \\ 0 & 1 & 2 \end{vmatrix} \\
&= ((-1)(2) - (1)(0))i + ((0)(0) - 1(2))j + ((1)(1) - (-1)(0))k \\
&= -2i - 2j + k \\
A &= \| \vec{a} \times \vec{b} \| \\
&= \sqrt{(-2)^2 + (-2)^2 + 1^2} \\
&= \sqrt{4 + 4 + 1} \\
&= \sqrt{9} \\
&= 3
\end{aligned}$$

1.6.1 The area of triangle

The area of a triangle defined by the vectors $\vec{a}$ and $\vec{b}$ is given by:

$$\begin{aligned}
A &= 1/2 \|\vec{a}\| \|\vec{b}\| \sin\theta \\
&= 1/2 \|\vec{a} \times \vec{b}\|
\end{aligned}$$

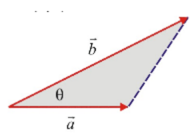


Figure 1.15: the area of triangle with sides $\vec{a}$ and $\vec{b}$

Example 1.13. *Find the area of the triangle* $\triangle ABC$ *where* $A(0, 1, 2)$, $B(-1, 0, 2)$, *and* $C(1, -2, 0)$

Solution:-

$$
\begin{aligned}
AB &= (-1,0,2) - (0,1,2) = (-1,-1,0)\\
\overrightarrow{AC} &= (1,-2,0) - (0,1,2) = (1,-3,-2)\\
\overrightarrow{AB} \times \overrightarrow{Ac} &= \begin{vmatrix} i & j & k \\ -1 & -1 & 0 \\ 1 & -3 & -2 \end{vmatrix}\\
&= (2-0)i + (0-2)j + (3+1)k\\
&= 2i + -2j + 4k.
\end{aligned}
$$

Then

$$
\begin{aligned}
A &= 1/2\|\overrightarrow{AB} \times \overrightarrow{Ac}\|\\
&= 1/2\sqrt{2^2 + (-2)^2 + 4^2}\\
&= 1/2\sqrt{4+4+16}\\
&= 1/2\sqrt{24}\\
&= \sqrt{6}
\end{aligned}
$$

1.6.2 Volume of Parallelepiped

The volume of a parallelepiped defined by the vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ is given by:

$$
\begin{aligned}
V &= | \, \vec{c} \cdot (\vec{a} \times \vec{b}) \, |\\
&= \vec{a} \cdot (\vec{b} \times \vec{c})\\
&= \vec{b} \cdot (\vec{c} \times \vec{a})
\end{aligned}
$$

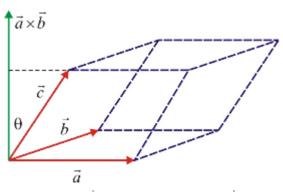


Figure 1.16: the volume of parallelopiped

$$\begin{aligned} V &= A_{base} \times h \\ &= \| (\vec{a} \times \vec{b}) \| | Sproj^{c}_{\vec{a}\times\vec{b}} | \\ &= \| (\vec{a} \times \vec{b}) \| | \frac{(\vec{a} \times \vec{b}) \cdot \vec{c}}{\| (\vec{a} \times \vec{b}) \|} | \\ &= | \vec{a} \times \vec{b}) \cdot \vec{c} | \end{aligned}$$

Example 1.14. *Find the volume of the parallelepiped defined by the vectors* $\vec{a} = (0, 1, -3)$, $\vec{b} = (1, 2, 3)$ *and* $\vec{c} = (-1, 0, 1)$.

Solution:-

$$\begin{aligned} \vec{a} \times \vec{b} &= \begin{vmatrix} i & j & k \\ 0 & 1 & -3 \\ 1 & 2 & 3 \end{vmatrix} \\ &= (3 + 6)i + (3 - 0)j + (0 - 1)k = 9i + 3j - k \\ (\vec{a} \times \vec{b}) \cdot \vec{c} &= (9i + 3j - k) \cdot (-i + k) \\ &= 9(-1) + 3(0) + (-1)(1) \\ &= -9 - 1 = -10 \\ V &= | (\vec{a} \times \vec{b}) \cdot \vec{c} | \\ &= | -10 | = 10 \end{aligned}$$

1.7 Lines and planes

Consider a line through the point $p_0(x_0, y_0, z_0)$ in the direction defined by the vector (a, b, c). See the fig below. Let $P(x, y, z)$ be any other point on the line. We get

$$\overrightarrow{p_0p} = (x - x_0, y - y_0, z - z_0)$$

The vectors $\overrightarrow{p_0p}$ and $(a,\ b,\ c)$ are parallel. Thus there exists a scalar t such that $\overrightarrow{p_0p} = t(a,\ b,\ c)$

$$(x - x_0,\ y - y_0,\ z - z_0) = t(a,\ b,\ c) \tag{1.12}$$

This is called the vector equation of the line. Comparing the components of the vectors on the left and right of this equation gives $x - x_0 = ta, y - y_0 = tb,\ z - z_0 = tc$

Rearranging these equations as follows gives the parametric equations of a line in R^3.

$$x = x_0 + ta, \quad y = y_0 + bt, \quad z = z_0 + ct \qquad -\infty < t < \infty \tag{1.13}$$

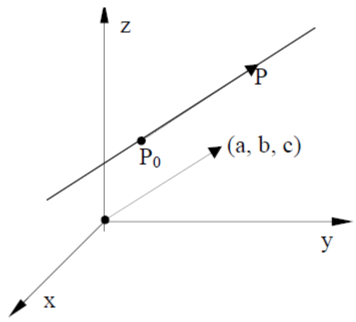


Figure 1.17: A parallel vector of a line

Example 1.15. *Find the vector and parametric equation of the line that pass through the point* $(1, 2, 5)$ *in the direction of* $(4, 3, 1)$*, and determine any two points on the line.*
***Solution**: Let* $(a, b, c) = (4, 3, 1)$ *and* $(x_0, y_0, z_0) = (1, 2, 5)$ *then from equation (1.12) we can write the vector equation of the line as* $(x - 1, y - 2, z - 5) = t(4, 3, 1)$ *.*
And from equation (1.13) we give the parametric equation of the line is
$x = 1 + 4t,\ y = 2 + 3t\ z = 5 + t \qquad -\infty < t < \infty.$
To find to points on the line we give **t** *two arbitrary values, for instance* $t = 1$ *leads to the point* $(5, 5, 6)$*, and* $t = -1$ *leads to the point* $(-3, -1, 4)$*.*

Example 1.16. *Find the parametric equation of the line pass through the points* $(-1, 2, 6)$ *and* $(1, 5, 4)$*.*

Let $(x_0, y_0, z_0) = (-1, 2, 6)$*. The direction of the line is given by the vector* $(a, b, c) = (1, 5, 4) - (-1, 2, 6) = (2, 3, -2)$*.*
Consequently the parametric equations of the line are given by
$x = -1 + 2t,\ y = 2 + 3t,\ z = 6 - 2t \qquad -\infty < t < \infty.$

1.7.1 Symmetric Equations of a Line

From equation (1.13) we can clear the parameter t by writing it as
$t = \dfrac{x - x_0}{a} = \dfrac{y - y_0}{b} = \dfrac{z - z_0}{c}$
provided that the three numbers **a**, **b**, and **c** are non-zero. The resulting equations
$\dfrac{x - x_0}{a} = \dfrac{y - y_0}{b} = \dfrac{z - z_0}{c}$
are said to be symmetric equations for the line pass through the points P_0 and P.

Example 1.17. *Find the symmetric equations for the line pass through the points* $(4, 10, -6)$ *and* $(7, 9, 2)$*.*
***Solution**: First let us find the reference vector as below* $(a, b, c) = (4, 10, -6) - (7, 9, 2) = (-3, 1, -8)$*.*

Then if we let $(x_0, y_0, z_0) = (7, 9, 2)$ *the symmetric equation of the line is given by.*

$$\frac{x-7}{-3} = \frac{y-9}{1} = \frac{z-2}{-8}$$

Note: If one of the numbers a, b, or c is zero in (1.13), we use the remaining two equations to eliminate the parameter $\mathbf{t}$. For example if $a = 0, \quad b \neq 0, \quad c \neq 0$, then (1.13) yields the symmetric equations for the line to be $x = x_0, \quad \dfrac{y-y_0}{b} = \dfrac{z-z_0}{c}$.

Theorem 1.3. *Let L be a line parallel to a vector $\vec{v}$ and Let p_1 be a point not on a line L then the distance "D" between p_1 and L is given by*

$$D = \frac{\| \vec{v} \times \overrightarrow{p_0p_1} \|}{\| \vec{v} \|}$$

where p_0 is any point on L and $(\|\vec{v}\| \neq 0)$
Proof exercise

Example 1.18. *Find the distance from the point* $p = (1, 1, 5)$ *to the line* $L: \ x = 1 + t, \ y = 3 - t, \quad z = 2t$

***Solution**: we see from the equation for L that L passes through $P_0 = (1, 3, 0)$ parallel to $\vec{v} = i - j + 2k$ with $\overrightarrow{p_0p} = -2j + 5k$ and $\overrightarrow{p_0p} \times \vec{v} = i + 5j + 2k$. Then from the above theorem,*

$$D = \frac{\| \overrightarrow{p_0p} \times \vec{v} \|}{\| \vec{v} \|}$$

$$= \frac{\sqrt{30}}{\sqrt{6}}$$

$$= \sqrt{5}$$

Example 1.19. *Find the distance between the following two parallel lines with the given parametric equations line $L_1: \ x = 2 - t, y = 2t, z = 1 + t$ and line $L_2: x = 1 + 2t, y = 3 - 4t, z = 5 - 2t$.*

***Solution**: Here the two lines are parallel, (i.e they do not intersect and they are on the some plane) (show!). Now to find the distance, take $\vec{v} = (-1, 2, 1)$ the vector parallel to a line L_1, the point $p_0(2, 0, 1)$ in line L_1 and $p_1(1, 3, 5)$ in line L_2 such that $\overrightarrow{p_0p_1} = (-1, 3, 4)$.*

But

$$\vec{v}\times\overrightarrow{p_0p_1} = \begin{vmatrix} i & j & k \\ -1 & 2 & 1 \\ -1 & 3 & 4 \end{vmatrix}$$
$$= (8-9)i - (-4+3)j + (-3+2)k$$
$$= -i + j - k, \quad \| \vec{v}\times\overrightarrow{p_0p_1} \|$$
$$= \sqrt{3}$$
$$\& \; \| \vec{v} \| = \sqrt{6}$$

Thus

$$D = \frac{\| \vec{v}\times\overrightarrow{p_0p_1} \|}{\| \vec{v} \|}$$
$$= \frac{\sqrt{3}}{\sqrt{6}}$$
$$= 1/\sqrt{2}.$$

.

1.7.2 Equations of Planes in $\mathbb{R}^3$

Let $p_0(x_0,\ y_0,\ z_0)$ be a point in a plane. Let $(a,\ b,\ c)$ be a vector perpendicular to the plane, called a normal to the plane. These two quantities, namely a point in a plane and a normal vector to the plane characterize the plane. There is only one plane through a given point and having a given normal. We will now drive the equation of a plane passing through the point $p_0(x_0, y_0, z_0)$ and having normal (a, b, c). Let $P(x, y, z)$ be any arbitrary point in the plane. We get

$$\overrightarrow{p_0p} = (x, y, z) - (x_0, y_0, z_)$$
$$= (x - x_0, y - y_0, z - z_0)$$

The vector $\overrightarrow{p_0p}$ lies in the plane. Thus the vector (a, b, c) and $\overrightarrow{p_0p}$ are orthogonal. Their dot product is zero. This observation leads to a vector equation of the plane

$$(a, b, c)\cdot\overrightarrow{p_0p} = 0$$
$$(a, b, c)\cdot(x - x_0, y - y_0, z - z_0) = 0$$

Specifically the last equation yields the point-normal form of the equation of the plane

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \tag{1.14}$$

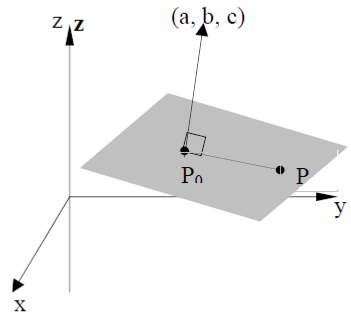


Figure 1.18: Normal vector and plane

and expanding the last equation and putting $d = ax_0 + by_0 + cz_0$ we obtain the general form of the equation of the plane

$$ax + by + cy = d \tag{1.15}$$

Observe that the components of (a, b, c) appear as coefficients in (1.14) and (1.15), ane the coordinates of the points $p_0(x_0, y_0, z_0)$ in the plane appear inside the parenthesis in (1.14)

Example 1.20. *Find the point-normal and general forms of the equation of the plane passing through the point* $(1, 2, 3)$ *and having normal* $(-1, 4, 6)$.
Solution: *Let* $(x_0, y_0, z_0) = (1, 2, 3)$ *and* $(a, b, c) = (-1, 4, 6)$. *Then the point normal form equation of the plane is given by*

$$-(x - 1) + 4(y - 2) + 6(z - 3) = 0$$

multiplying and simplifying the last equation we get the general form

$$x + 4y + 6z = 25$$

Example 1.21. *Determine the equation of the plane through the three points* $P(2, -1, 1)$, $Q(-1, 1, 3)$ *and* $R(2, 0, -3)$.
Solution: *The vectors* $\overrightarrow{PQ}$ *and* $\overrightarrow{PR}$ *lie in the plane. Thus* $\overrightarrow{PQ} \times \overrightarrow{PR}$ *will be normal to the plane. So since*

$$\overrightarrow{PQ} = (-1, 1, 3) - (2, -1, 1)$$
$$\overrightarrow{PQ} = (2, 0, -3) - (2, -1, 1)$$
$$\overrightarrow{PQ} \times \overrightarrow{PR} = -10i - 12j - 3k$$

Finally putting $(x_0, y_0, z_0) = (2, -1, 1)$ *and* $(a, b, c) = (-10, -12, -3)$ *we give the point normal equation by*
$-10(x - 2) - 12(y + 1) - 3(z - 1) = 0$
or the general equation by $-10x - 12y - 3z = -11$.

Note: Two planes are parallel whenever their normal vectors are parallel and the planes are orthogonal whenever their normal vectors are orthogonal.

Example 1.22. *Determine whether the planes* $x - 3y + 6z = 4$ *and* $5x + y - z = 4$ *are orthogonal, parallel, or neither. Find the angle of intersection and the set of parametric equations for the line of intersection of the plane.*
Solution: *For the plane* $x - 3y + 6z = 4$ *, the normal vector is* $n_1 = (1, -3, 6)$ *and for the plane* $5x + y - z = 4$*, the normal vector is* $n_2 = (5, 1, -1)$ *. The two planes will be orthogonal only if their corresponding normal vectors are orthogonal, that is, if* $n_1 \cdot n_2 = 0$*. However, we see that*

$$n_1 \cdot n_2 = (1, -3, 6) \cdot (5, 1, -1) = (1)(5) + (-3)(1) + (6)(-1) = 5 - 3 - 6 = -4 \neq 0$$

Hence, the planes are not orthogonal.
If the planes are parallel, then their corresponding normal vectors must be parallel. For that to occur, there must exist a scalar k *where*

$$n_2 = kn_1$$

Rearranging this equation as $kn_1 = n_2$ *and substituting for* n_1 *and* n_2 *gives*
$k(1, -3, 6) = (5, 1, -1)$ *or* $(k, -3k, 6k) = (5, 1, -1)$.
Equating components gives the equations
$k = 5, \quad -3k = 1, \quad 6k = -1$
This gives
$k = 5, \quad k = -1/3 \quad k = -1/6$.
Since the values of k *are not the same for each component to make the vector* n_2 *a scalar multiple of the vector* n_1 *, the planes are not parallel. Thus, the planes must intersect in a straight line at a given angle. To find this angle, we use the equation*

$$\cos\theta = \frac{n_1 \cdot n_2}{\| n_1 \| \| n_2 \|}$$

For this formula, we have the following:

$$\begin{aligned} n_1 \cdot n_2 &= (1, -3, 6) \cdot (5, 1, -1) \\ &= (1)(5) + (-3)(1) + (6)(-1) \\ &= 5 - 3 - 6 = -4 \\ \parallel n_1 \parallel &= \sqrt{1^2 + (-3)^2 + 6^2} \\ &= \sqrt{1 + 9 + 36} \\ &= \sqrt{46} \\ \parallel n_2 \parallel &= \sqrt{5^2 + 1^2 + (-1)^2} \\ &= \sqrt{25 + 1 + 1} \\ &= \sqrt{27} \end{aligned}$$

Thus,

$$\cos\theta = \frac{-4}{\sqrt{46}\sqrt{27}}.$$

Solving for θ gives

$$\begin{aligned} \theta &= \cos^{-1}(\frac{-4}{\sqrt{46}\sqrt{27}}) \\ &= 1.68 \end{aligned}$$

radians $= 96.5^0$.

To find the equation of the line of intersection between the two planes, we need a point on the line and a parallel vector. To find a point on the line, we can consider the case where the line touches the x y plane, that is, z = 0. If we take the two equations of the plane

$$\begin{aligned} x - 3y + 6z &= 4 \\ 5x + y - z &= 4 \end{aligned}$$

and substitute z = 0, we obtain the system of equations

$$\begin{aligned} x - 3y &= 4 \qquad (1) \\ 5x + y &= 4 \qquad (2) \end{aligned}$$

Taking the first equation and multiplying by -5 *gives*

$$\begin{aligned} -5x + 15y &= -20 \\ 5x + y &= 4 \end{aligned}$$

Adding the two equations gives $16y = -16$ *or* $y = \dfrac{-16}{16} = -1$. *Substituting* $y = -1$ *back into equation (1) gives*

$x - 3(-1) = 4$ *or* $x + 3 = 4$. *Solving for* x *gives* $x = 4 - 3 = 1$. *Thus, the point* $(1, -1, 0)$ *is on the plane. To find a parallel vector for the line, we use the fact that since the line is on both planes, it must be orthogonal to both normal vectors* n_1 *and* n_2 *. Since the cross product* $n_1 \times n_2$ *gives a vector orthogonal to both* n_1 *and* n_2*,* $n_1 \times n_2$ *will be a parallel vector for the line. Thus, we say that*

$$\begin{aligned} l' &= n_1 \times n_2 \\ &= \begin{vmatrix} i & j & k \\ 1 & -3 & 6 \\ 5 & 1 & -1 \end{vmatrix} \\ &= (3-6)i - (-1-30)j + (1+15)k \\ &= -3i + 31j + 16k. \end{aligned}$$

Hence, using the point $(1, -1, 0)$ *and the parallel vector* $l' = -3i + 31j + 16k$ *, we find the parametric equations of the line are*

$$\begin{aligned} x &= 1 - 3t \\ y &= -1 + 31t \\ z &= 16t. \end{aligned}$$

1.7.3 Distance between Points and a Plane

Suppose that we wanted to find the distance from the plane $ax + by + cz + d = 0$ to a point $P_0(x_0, y_0, z_0)$ not on the plane. Notice that the distance is measured along a line segment connecting the point to the plane that is orthogonal to the plane (see the following Figure). To compute this distance, pick any point $P_1(x_1, y_1, z_1)$ lying in the plane and let $\vec{a} = (a, b, c)$ denote a vector normal to the plane.

Next, notice from the figure that the distance from P_0 to the plane is simply the norm of $proj_{\vec{a}}^{\overrightarrow{p_0p_1}}$ Also notice that $\overrightarrow{p_0p_1} = (x_1 - x_0, y_1 - y_0, z_1 - z_0)$.

Now we can write the distance as

$$\begin{aligned} \mathbf{D} &= \| proj_{\vec{a}}^{\overrightarrow{p_0p_1}} \| \\ &= | \overrightarrow{p_0p_1} \cdot \frac{\vec{a}}{\| \vec{a} \|} | \\ &= \frac{| ax_0 + by_0 + cz_0 - d |}{\sqrt{a^2 + b^2 + c^2}} \end{aligned}$$

Note: To find the distance between two parallel planes in space, take any point of one of the planes and find the distance between the point and the other plane.

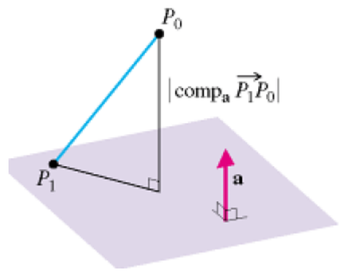


Figure 1.19: The distance of the plane and the point

Example 1.23. *Find the distance between the point* $(1, 2, 3)$ *and plane* $2x - y + z = 4$.
Solution: *Since we are given the point* $Q = (1, 2, 3)$, *we need to find a point on the plane* $2x - y + z = 4$ *in order to find the vector* $\overrightarrow{PQ}$. *We can simply do this by setting* $y = 0$ *and* $z = 0$ *in the plane equation* $2x - y + z = 4$ *and solving for* x. *Thus we have* $2x - 0 + 0 = 4 \quad \Rightarrow \quad 2x = 4 \quad \Rightarrow \quad x = 4/2 \quad \Rightarrow \quad x = 2$.
Thus the point $P(2, 0, 0)$ *and the vector* $\overrightarrow{PQ}$ *is* $(1 - 2, 2 - 0, 3 - 0) = (-1, 2, 3)$.
Hence, using the fact that the normal vector for the plane is $n = (2, -1, 1)$, *we have distance* ***D*** *Between* Q *and the plane*

$$\begin{aligned}\mathbf{D} &= \left| \overrightarrow{PQ} \cdot \frac{\vec{n}}{\| \vec{n} \|} \right| \\ &= \frac{|(-1, 2, 3) \cdot (2, -1, 1)|}{\sqrt{2^2 + (-1)^2 + 1^2}} \\ &= \frac{|-2 - 2 + 3|}{\sqrt{4 + 1 + 1}} \\ &= 1/\sqrt{6}.\end{aligned}$$

Example 1.24. *Find the distance between the following two parallel planes given by* $3x - y + 2z = 6$ *and* $6x - 2y + 4z - 4$.
Solution*: To find the distance between the two planes take a point, say* $(x_0, y_0, z_0) = (2, 0, 0)$ *from the first plane. Then from the second plane, you can determine that* $a = 6, \quad b = -2$ *and* $c = 4$ *,then conclude that the distance* **D** *is*

$$\begin{aligned}\mathbf{D} &= \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}} \\ &= \frac{|(6)(2) + (-2)(0) + (4)(0) - (-4)|}{\sqrt{6^2 + (-2)^2 + 4^2}} \\ &= \frac{12 + 0 + 0 + 4}{\sqrt{36 + 4 + 16}} \\ &= 16/\sqrt{56}\end{aligned}$$

1.8 Vector space and Subspace

Definition 1.5. *A vector space (over $\mathbb{R}$) consists of a set V along with two operations $+$ and $\cdot$ subject to these conditions.*

1. *For any $\vec{v}, \vec{w} \in V$, their vector sum $(\vec{v}+\vec{w})$ is an element of V .*
2. *If $\vec{v}, \vec{w} \in V$, then $\vec{v}+\vec{w}=\vec{w}+\vec{v}$.*
3. *For any $\vec{u}, \vec{v}, \vec{w} \in V$, $\quad (\vec{v}+\vec{w})+\vec{u}=\vec{v}+(\vec{w}+\vec{u})$.*
4. *There is a zero vector $\boldsymbol{0} \in V$ such that $\vec{v}+\boldsymbol{0}=\vec{v}$ for all $\vec{v} \in V$.*
5. *Each $\vec{v} \in V$ has an additive inverse $\vec{w} \in V$ such that $\vec{w}+\vec{v}=\boldsymbol{0}$.*
6. *If r is a scalar, that is, a member of $\mathbb{R}$ and $\vec{v} \in V$ then the scalar multiple $r \cdot \vec{v}$ is in V .*
7. *If $r, s \in \mathbb{R}$ and $\vec{v} \in V$ then $(r+s) \cdot \vec{v}=r \cdot \vec{v}+s \cdot \vec{v}$.*
8. *If $r \in \mathbb{R}$ and $\vec{v}, \vec{w} \in V$, then $r \cdot(\vec{v}+\vec{w})=r \cdot \vec{v}+r \cdot \vec{w}$.*
9. *If $r, s \in \mathbb{R}$ and $\vec{v} \in V$, then $(r s) \cdot \vec{v}=r \cdot(s \cdot \vec{v})$.*
10. *For any $\vec{v} \in V$, $\quad 1 \cdot \vec{v}=\vec{v}$.*

Example 1.25. *Show that the set $\mathbb{R}^2$ is a vector space if the operations "$+$" and "$\cdot$" have their usual meaning.*

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}+\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \end{pmatrix}$$
$$\& r \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} r x_1 \\ r x_2 \end{pmatrix}$$

Solution:*To prove $\mathbb{R}^2$ is a vector space we shall check all of the conditions.*
There are five conditions in item 1.
For 1, closure of addition, note that for any $v_1, v_2, w_1, w_2 \in \mathbb{R}$ the result of the sum

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}+\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}=\begin{pmatrix} v_1+w_1 \\ v_2+w_2 \end{pmatrix}$$

is a column array with two real entries, and so is in $\mathbb{R}^2$.

For 2, that addition of vectors commutes, take all entries to be real numbers and compute

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix}$$
$$= \begin{pmatrix} w_1 + v_1 \\ w_2 + v_2 \end{pmatrix}$$
$$= \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

(the second equality follows from the fact that the components of the vectors are real numbers, and the addition of real numbers is commutative).

Condition 3, associativity of vector addition, is similar.

$$\{\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}\} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} (v_1 + w_1) + u_1 \\ (v_2 + w_2) + u_2 \end{pmatrix}$$
$$= \begin{pmatrix} v_1 + (w_1 + u_1) \\ v_2 + (w_2 + u_2) \end{pmatrix}$$
$$= \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \{\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\}$$

For the fourth condition we must produce a zero element the vector of zeroes is it.

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

For 5, to produce an additive inverse, note that for any $v_1, v_2 \in \mathbb{R}$ *we have*

$$\begin{pmatrix} -v_1 \\ -v_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

so the first vector is the desired additive inverse of the second.

To check the next five conditions having to do with scalar multiplication are just as routine.

For 6, closure under scalar multiplication, where

$$r, v_1, v_2 \in \mathbb{R},$$
$$r \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} rv_1 \\ rv_2 \end{pmatrix}$$

is a column array with two real entries, and so is in $\mathbb{R}^2$ *.*
Next, this checks 7.

$$\begin{aligned}(r+s)\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix} &= \begin{pmatrix}(r+s)v_1\\(r+s)v_2\end{pmatrix}\\ &= \begin{pmatrix}rv_1+sv_1\\rv_2+sv_2\end{pmatrix}\\ &= r\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix} + s\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix}\end{aligned}$$

For 8, that scalar multiplication distributes from the left over vector addition, we have this.

$$\begin{aligned}r\cdot(\begin{pmatrix}v_1\\v_2\end{pmatrix}+\begin{pmatrix}w_1\\w_2\end{pmatrix}) &= \begin{pmatrix}r(v_1+w_1)\\r(v_2+w_2)\end{pmatrix}\\ &= \begin{pmatrix}rv_1+rw_1\\rv_2+rw_2\end{pmatrix}\\ &= r\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix} + r\cdot\begin{pmatrix}w_1\\w_2\end{pmatrix}\end{aligned}$$

The ninth

$$\begin{aligned}(rs)\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix} &= \begin{pmatrix}(rs)v_1\\(rs)v_2\end{pmatrix}\\ &= \begin{pmatrix}r(sv_1)\\r(sv_2)\end{pmatrix}\\ &= r\cdot(s\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix})\end{aligned}$$

and tenth conditions are also straightforward.

$$1\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix} = \begin{pmatrix}1v_1\\1v_2\end{pmatrix} = \begin{pmatrix}v_1\\v_2\end{pmatrix}$$

In a similar way, each $\mathbb{R}^n$ *where* $\boldsymbol{n}\in\mathbb{N}$ *is a vector space with the usual operations of vector addition and scalar multiplication.*

Example 1.26. *The singleton set* $\{\begin{pmatrix}0\\0\\0\\0\end{pmatrix}\}$ *is a vector space under the operations*

$$\begin{pmatrix}0\\0\\0\\0\end{pmatrix}+\begin{pmatrix}0\\0\\0\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

$$\&r\cdot\begin{pmatrix}0\\0\\0\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

Exercise

1. Show that $\mathcal{P}_3 = \{a_0 + a_1x + a_2x^2 + a_3x^3 \,\big|\, a_0, \ldots, a_3 \in \mathbb{R}\}$, the set of polynomials of degree three or less is a vector space under the operations

   $$(a_0+a_1x+a_2x^2+a_3x^3)+(b_0+b_1x+b_2x^2+b_3x^3) = (a_0+b_0)+(a_1+b_1)x+(a_2+b_2)x^2+(a_3+b_3)x^3$$

   and
   $r \cdot (a_0 + a_1x + a_2x^2 + a_3x^3) = (ra_0) + (ra_1)x + (ra_2)x^2 + (ra_3)x^3$

2. The set $\mathcal{M}_{2\times2}$ of 2×2 matrices with real number entries is a vector space under the natural entry-by-entry operations.
   $\begin{pmatrix}a & b\\c & d\end{pmatrix}+\begin{pmatrix}w & x\\y & z\end{pmatrix}=\begin{pmatrix}a+w & b+x\\c+y & d+z\end{pmatrix}$ and $r\cdot\begin{pmatrix}a & b\\c & d\end{pmatrix}=\begin{pmatrix}ra & rb\\rc & rd\end{pmatrix}$

3. The set of polynomials with real coefficients $\{a_0+a_1x+\cdots+a_nx^n \,\big|\, n \in \mathbb{N}$ and $a_0, \ldots, a_n \in \mathbb{R}\}$
   makes a vector space when given the natural " + "

   $$(a_0+a_1x+\cdots+a_nx^n)+(b_0+b_1x+\cdots+b_nx^n) = (a_0+b_0)+(a_1+b_1)x+\cdots+(a_n+b_n)x^n$$

   and

   $$r \cdot (a_0 + a_1x + \ldots a_nx^n) = (ra_0) + (ra_1)x + \ldots (ra_n)x^n$$

1.8.1 Subspace

A subspace is a vector space that is contained within another vector space. So every subspace is a vector space in its own right, but it is also defined relative to some other (larger) vector space. We will discover shortly that we are already familiar with a wide variety of subspaces from previous sections.

Definition 1.6. *Suppose that V and W are two vector spaces that have identical definitions of vector addition and scalar multiplication, and that W is a subset of V, $W \sqsubseteq V$. Then W is a subspace of V.*

Proposition:- A subset **S** of a vector space **V** is a subspace of **V** if and only if **S** is nonempty and closed under linear operations, i.e., $x, y \in \mathbf{S} \quad \Rightarrow \quad x + y \in \mathbf{S}$, and $x \in \mathbf{S} \quad \Rightarrow \quad rx \in \mathbf{S}$ for all $r \in \mathbf{R}$.

Example 1.27. *The set* $P = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Big| x + y + z = 0\}$ *is a subspace of* $\mathbb{R}^3$ *if "+" and "·" are interpreted as* $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}$ *and* $r \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} rx \\ ry \\ rz \end{pmatrix}$

Solution:- *Since* $\mathbb{R}^3$ *is a vector space and also* ***P*** *is a subset of* $\mathbb{R}^3$ *then by the above preposition we prove that P is closed under addition and scalar multiplication , take two elements of P*
$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$ *and* $\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$ *(membership in P means that* $x_1 + y_1 + z_1 = 0$ *and* $x_2 + y_2 + z_2 = 0$*), and observe that their sum* $\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}$ *is also in P since its entries add* $(x_1 + x_2) + (y_1 + y_2) + (z_1 + z_2) = (x_1 + y_1 + z_1) + (x_2 + y_2 + z_2) = 0 + 0 = 0$. *To show that P is closed under scalar multiplication, start with a vector from P*
$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ *(so that* $x + y + z = 0$ *) and then for* $r \in \mathbb{R}$ *observe that the scalar multiple*
$r \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} rx \\ ry \\ rz \end{pmatrix}$ *satisfies that* $rx + ry + rz = r(x + y + z) = r.0 = 0$. *Thus the two closure conditions are satisfied. There fore* ***P*** *is a subspace of* $\mathbb{R}^3$.

Exercise: Show that a nonempty subset S of a real vector space is a subspace if and only if it is closed under linear combinations of pairs of vectors: whenever $c_1, c_2 \in \mathbb{R}$ and $\vec{v}_1, \vec{v}_2 \in S$ then the combination $c_1\vec{v}_1 + c_2\vec{v}_2$ is in S.

1.9 Linear Dependence and independence; Basis of a vector space

A set of vectors is linearly independent if no vector in the set **V** is a scalar multiple of another vector or a linear combination of other vectors in the set;
conversely, a set of vectors is linearly dependent if any vector in the set **V** is a scalar multiple of another vector in the set or a linear combination of other vectors in the set.

Definition 1.7. *Suppose that* $\mathbf{V}$ *is a vector space. Given* n *vectors* $u_1, u_2, u_3, \cdots u_n$ *and* $\boldsymbol{n}$ *scalars* $\alpha_1, \alpha_2, \alpha_3 \cdots, \alpha_n$*, their linear combination is the vector* $\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \cdots + \alpha_n u_n$*.*

Definition 1.8. *If* $S = \{v_1,\ v_2,\ v_3, \cdots,\ v_n\}$ *is a non-empty set of vectors, then the vector equation* $k_1 v_1 + k_2 v_2 + k_3 v_3 + \cdots + k_n v_n = 0$ *has at least one trivial solution, namely* $k_1 = k_2 = k_3 = \cdots = k_n = 0$*. If this solution is the only solution of the set* $\mathbf{S}$ *then the set* $\mathbf{S}$ *is called linearly independence set.*
if there are other solutions then the set $\mathbf{S}$ *is called linearly dependent.*

Example 1.28. *Let* $v_1 = (1, 0, 1), \quad v_2 = (-1, 1, 0)$ *and* $v_3 = (1, 2, 3)$*. Express* v_3 *as a linear combination of* v_1 *and* v_2*.*
Solution*: We must find scalars* c_1 *and* c_2 *so that* $v_3 = c_1 v_1 + c_2 v_2$ *Using our knowledge of scalar multiplication and addition of vectors, we set*

$$(1, 2, 3) = c_1(1, 0, 1) + c_2(-1, 1, 0)$$
$$(1, 2, 3) = (c_1, 0, c_1) + (-c_2, c_2, 0)$$
$$(1, 2, 3) = (c_1 - c_2, c_2, c_1)$$

Equating corresponding components we have :
$1 = c_1 - c_2,\ 2 = c_2\ \&\ 3 = c_1$
Since $c_1 = 3$ *and* $c_2 = 2$ *does satisfy* $c_1 - c_2 = 1$*, we see that this is the solution. Hence* $(1, 2, 3) = 3(1, 0, 1) + 2(-1, 1, 0)$*. so we have found the required linear combination to be* $v_3 = 3v_1 + 2v_2$*.*

Example 1.29. *Determine whether* $\{v_1, v_2, v_3\}$ *form a linearly independent set or not, where* $v_1 = (1, 1, 2, 1), \quad v_2 = (0, 2, 1, 1),$ *and* $v_3 = (3, 1, 2, 0)$
Solution*: Using the second definition, we wish to know whether* $c_1 v_1 + c_2 v_2 + c_3 v_3 = 0$ *has only the trivial solution.*
Suppose that $c_1 v_1 + c_2 v_2 + c_3 v_3 = \boldsymbol{0}$*.*

Then

$$c_1(1,1,2,1) + c_2(0,2,1,1) + c_3(3,1,2,0) = \boldsymbol{0}$$
$$(c_1, c_1, 2c_1, c_1) + (0, 2c_2, c_2, c_2) + (3c_3, c_3, 2c_3, 0) = \boldsymbol{0}$$
$$(c_1 + 3c_3, c_1 + 2c_2 + c_3, 2c_1 + c_2 + 2c + 3, c_1 + c_2) = \boldsymbol{0}$$
$$c_1 + 3c_3 = 0$$
$$c_1 + 2c_2 + c_3 = 0$$
$$2c_1 + c_2 + 2c + 3 = 0$$
$$c_1 + c_2 = 0$$

has the unique solution $c_1 = c_2 = c_3 = 0$.
There fore we see that the set $\{v_1, v_2, v_3\}$ *is linearly independent.*

Theorem 1.4. *Suppose that* $S = \{u_1, u_2, u_3, \cdots, u_n\}$ *is a set of vectors. Then* $\mathbf{S}$ *is a linearly dependent set if and only if there is an index* $\mathbf{t}$ $1 \leqslant t \leqslant n$ *such that* u_t *is a linear combination of the vectors* $u_1, u_2, u_3, \cdots u_{t-1}, u_{t+1}, \cdots, u_n$.

Definition 1.9. *Let* $\mathbf{V}$ *denote a vector space and* $\boldsymbol{S} = \{u_1, u_2, u_3, \cdots, u_n\}$ *a subset of* $\mathbf{V}$. $\boldsymbol{S}$ *is called a basis for* V *if the following is true:*

1. $\mathbf{S}$ *spans* $\mathbf{V}$.
2. $\mathbf{S}$ *is linearly independent.*

Definition 1.10. *Let* $\mathbf{V}$ *be a vector space and let* $\boldsymbol{S} = \{v_1, v_2, \cdots, v_n\}$ *be a subset of* $\mathbf{V}$. *We say that* $\mathbf{S}$ *spans* $\mathbf{V}$ *if every vector* v *in* $\mathbf{V}$ *can be written as a linear combination of vectors in* $\mathbf{S}$. $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$

Example 1.30. *Show that the set* $S = \{(0,1,1), (1,0,1), (1,1,0)\}$ *spans* R^3 *and write the vector* $(2,4,8)$ *as a linear combination of vectors in S.*
Solution:- *A vector in* R^3 *has the form* $v = (x, y, z)$ *Hence we need to show that every such* v *can be written as*

$$(x, y, z) = c_1(0,1,1) + c_2(1,0,1) + c_3(1,1,0)$$
$$= (c_2 + c_3, c_1 + c_3, c_1 + c_2)$$

$$x = c_2 + c_3$$
$$y = c_1 + c_3$$
$$z = c_1 + c_2$$

$$c_3 = x - c_2,$$
$$\&c_1 = z - c_2$$
$$y = z - c_2 + x - c_2$$
$$= z + x - 2c_2$$
$$c_2 = (x - y + z)/2,$$
$$c_1 = (z - x + y)/2$$
$$c_3 = (x + y - z)/2$$

$$\therefore \textit{ for every } (x, y, z) = \frac{(z - x + y)}{2}(0,1,1) + \frac{(x - y + z)}{2}(1,0,1) + \frac{(x + y - z)}{2}(1,1,0).$$

Hence $S = \{(0,1,1),(1,0,1),(1,1,0)\}$ *spans* R^3
And We have $(2,4,8) = 5(0,1,1) + 3(1,0,1) + (-1)(1,1,0)$

Example 1.31. *Prove that* $\boldsymbol{S} = \{1,\ x,\ x^2\}$ *is a basis for the set of polynomials of degree less than or equal to 2.*
Solution: *We need to prove that* **S** *spans* P_2 *and is linearly independent.*
S *spans* P_2. *We already did this in the section on spanning sets. A typical polynomial of degree less than or equal to* ***2*** *is* $ax^2 + bx + c$

S is linearly independent. Here, we need to show that the only solution to $a + bx + cx^2 = 0$ *(where* ***0*** *is the zero polynomial) is* $a = b = c = 0$.
From algebra, we remember that two polynomials are equal if and only if their corresponding coefficients are equal. The zero polynomial has all its coefficients equal to zero. So, $a(1) + bx + cx^2 = \boldsymbol{0}$ *if and only if* $a = 0,\quad b = 0,\quad c = 0$. *Which proves that* **S** *is linearly independent.*

Theorem 1.5. *Let* **V** *denote a vector space and* $S = \{u_1,\ u_2,\ u_3,\ \cdots, u_n\}$ *a basis of* ***V****. Every vector in* **V** *can be written in a unique way as a linear combination of vectors in* **S**

1.10 Exercises and solutions

1. A kite is pulled with a force $\vec{F} = \langle 2, 1, 4\rangle$. It has velocity $\vec{v} = \langle 1, -1, 1\rangle$. The dot product of $\vec{F}$ with $\vec{v}$ is called power.

   (a) Find the angle between the force and the velocity.

   (b) Find the vector projection of the force onto the velocity vector.

2. Light shines long the vector $\vec{a} = \langle a_1, a_2, a_3\rangle$ and reflects at the three coordinate planes where the angle of incidence equals the angle of reflection. Verify that the reflected ray is $\vec{a}$.
   Hint. Reflect first at the xy-plane and find the components of the ray after reflection. You can assume that in that case, the reflected vector is in the plane spanned by $\vec{k} = \langle 0, 0, 1\rangle$ and $\vec{a}$.

3. Given the vectors $\vec{a} = \langle 1, 2, 1\rangle$, $\vec{b} = \langle 1, -1, 1\rangle$, $\vec{c} = \langle 0, 1, 1\rangle$, $d = \langle 2, 3, 4\rangle$, compute all possible dot products and determine which pairs are perpendicular.

4. (a) Find the angle between a diagonal of a cube and the diagonal in one of its faces.

   (b) The hypercube or tesseract has vertices $(\pm1, \pm1, \pm1, \pm1)$. Find the angle between the hyper diagonal connecting $(1, 1, 1, 1)$ with $(-1, -1, -1, -1)$ and the space diagonal connecting $(1, 1, 1, 1)$ with $(-1, -1, -1, -1)$

5. (a) Verify that if $\vec{a}$. $\vec{b}$ are nonzero, then $\vec{c} =\mid \vec{a} \mid \vec{b}+ \mid \vec{b} \mid \vec{a}$ bisects the angle between $\vec{a}$. $\vec{b}$ if $\vec{c}$ is not zero.

   (b) Verify the parallelogram law

$$\mid \vec{a} + \vec{a} \mid^2 + \mid \vec{a} - \vec{a} \mid^2 = 2 \mid \vec{a} \mid^2 + 2 \mid \vec{b} \mid^2 .$$

6. (a) Find a nonzero vector orthogonal to the plane through the points $P = (-2, 3, 1)$, $Q = (1, 5, 2)$, $R = (4, 3, -1)$ and containing P.

   (b) Find the equation of this plane.

   (c) Find the area of the triangle PQR.

7. (a) Parametrize a line perpendicular to the plane containing $A = (1, 1, .1)$, $B = (2, 3, 4)$ and $C = (4, 5, 6)$ and passing through A.

   (b) Find the equation $ax + by + cz = d$ of the plane through A, B, C.

8. (a) Use volume to determine whether $A = (1, 1, 2)$, $B = (3, 1, 6)$, $C = (5, 2, 0)$ and $D = (1, 4, 12)$ are in the same plane.

(b) Find the distance between the line L through A, B and the line M through C, D.

(c) How come that whenever A, B, C, D are not in the same plane, then the distance between L and M is positive?

9. (a) Find an equation of the plane containing the line of intersection of the planes $x - z = 1$ and $y + z = 3$ which is perpendicular to the plane $x + y - 2z = 1$.

(b) Find the distance of the plane found in a) to the origin $(0, 0, 0)$.

10. (a) Parametrize the line L through $P = (2, 1, 2)$ that intersects the line $x = 1+t,\ y = 1 - t,\ z = 2t$ perpendicularly.

(b) What is the distance from this line L to the origin $(0, 0, 0)$?

CHAPTER TWO

2 Matrices and Determinants

Objective:

- By the end of this chapter, students are expected to:
  - Define and identify different types of matrices
  - Understand the arithmetic operations on matrices
  - Reduce the given matrix to row reduced echelon form using elementary row operations
  - Find the inverse of some matrices using elementary row operations
  - Define system of linear equations in terms of matrices
  - Apply Gaussian elimination method, Gaussian Jordan method, and matrix inversion method to solve the given system of linear equations
  - Define and compute eigenvalue and eigenvectors

2.1 Definition of matrix and basic operations

Definition 2.1. - *Matrix is a rectangular array or arrangement of numbers of the form*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

is called a matrix of size m by n (written as $m \times n$).

Each number $\mathbf{a_{ij}}$ is called an element or entry of the matrix and it is an element appearing in the i^{th} row and j^{th} column of a matrix. Elements in the horizontal line are said to form rows, and elements in the vertical lines are said to form columns. Here we say A has m rows and n columns. The i^{th} row of matrix A is $R_i = \begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \end{bmatrix}$ $1 \leq i \leq m$

The j^{th} column of the matrix A is $c_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ a_{3j} \\ \vdots \\ a_{mj} \end{bmatrix}$ $1 \leq j \leq n.$

The order of a matrix denotes the number of rows and columns in the matrix. Thus, a matrix of order $m \times n$ has **m** rows and **n** columns. We often write the matrix A as $A = [a_{ij}]_{m\times n},\ 1 \le i \le m,\ 1 \le j \le n$, i-denotes the row and j- denotes the column.
For example, in the matrix

$$A = \begin{bmatrix} 5 & 9 & 6 & 8 \\ 3 & 2 & 3 & 1 \\ 1 & 0 & 4 & 7 \end{bmatrix}$$

, there are 3 rows and 4 columns. Therefore, matrix A can be called a matrix of order or size 3×4.

The rows are $\begin{bmatrix} 5 & 9 & 6 & 8 \end{bmatrix}$, $\begin{bmatrix} 3 & 2 & 3 & 1 \end{bmatrix}$ & $\begin{bmatrix} 1 & 0 & 4 & 7 \end{bmatrix}$, and the columns $\begin{bmatrix} 5 \\ 3 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 9 \\ 2 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 6 \\ 3 \\ 4 \end{bmatrix}$ & $\begin{bmatrix} 8 \\ 1 \\ 7 \end{bmatrix}$.
Here $a_{11} = 5,\ a_{12} = 9, a_{13} = 6,\ a_{14} = 8,\ a_{21} = 3,\ a_{22} = 2,\ a_{23} = 3,\ a_{24} = 1$, etc. and A has 12 elements.

Definition 2.2. *Two matrices A and B are said to be equal, written $A = B$, if they are of the same order and if all corresponding entries are equal i.e. $a_{ij} = b_{ij}$.*

For example, $\begin{bmatrix} 5 & 1 & 0 \\ 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 2+3 & 1 & 0 \\ 2 & 3 & 2\times 2 \end{bmatrix}$
but $\begin{bmatrix} 2 \\ 9 \end{bmatrix} \neq \begin{bmatrix} 2 & 9 \end{bmatrix}$. Why?

Example 2.1. *Given the matrix equation*

$$\begin{bmatrix} x+y & 8 \\ x-y & 6 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 1 & 6 \end{bmatrix}.$$

Find x and y.
***Solution**: By the definition of equality of matrices,*
$x + y = 3$ and $x - y = 1$ solving this system of equations gives $x = 2$ and $y = 1$.

Exercise:

1. Find the values of x, y, z and w which satisfy the matrix equation

   (a) $\begin{bmatrix} x-y & 2x+z \\ 2x-y & 3z+w \end{bmatrix} = \begin{bmatrix} -1 & 5 \\ 0 & 13 \end{bmatrix}$

   (b) $\begin{bmatrix} x+3 & 2y+x \\ z-1 & 4w+6 \end{bmatrix} = \begin{bmatrix} 0 & -7 \\ 3 & 2w \end{bmatrix}$

2.1.1 Types of Matrices

1. A matrix having exactly one row is called a row matrix.

   Example 2.2. $\begin{pmatrix} 1 & 0 & 4 & 7 \end{pmatrix}$, $\begin{pmatrix} 1 & 5 \end{pmatrix}$, *are row matrices. A row matrix is often referred to as a row vector.*

2. A matrix having exactly one column is called a column matrix

   Example 2.3. $\begin{bmatrix} 4 \\ 3 \\ -6 \end{bmatrix}$ *is a column matrix.*

3. A zero matrix or null matrix is a matrix in which all of its elements are zero.

   Example 2.4. *The matrices,* $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ & $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ *are the zero matrices*

4. A matrix, in which the numbers of rows and the number of columns are equal, that is, an $n \times n$ matrix, is called a square matrix of order n.

   Example 2.5. $\begin{bmatrix} 3 & 1 \\ 4 & 0 \end{bmatrix}$ *and* $\begin{bmatrix} 1 & 2 & 8 \\ 4 & 6 & 0 \\ 1 & 3 & 5 \end{bmatrix}$ *are square matrices.*

   If $A = (a_{ij})_{n\times n}$ is a square matrix the elements a_{ii}'s are called the diagonal elements. The main diagonal or simply diagonal of **A** consists of the elements $a_{11},\ a_{22},\ a_{33},\ \cdots,\ a_{nn}$.

5. A square matrix in which all the non-diagonal elements are zero is called a diagonal matrix.

   Example 2.6. $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ *and* $\begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$ *are diagonal matrices.*

   Note that the diagonal elements in a diagonal matrix may also be zero.

Example 2.7. $\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ *and* $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ *are also diagonal matrix.*

6. A diagonal matrix whose diagonal elements are equal is called a scalar matrix

Example 2.8. $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$, $\begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ & $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ *are scalar matrices*

7. Diagonal matrix of order n in which every diagonal element is unity (one) is called the identity matrix or unit matrix of order n . The identity matrix of order n is denoted by I_n.

Example 2.9. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ *is an identity matrix of order two.*

8. A square matrix having only zeros below its diagonal is called upper triangular matrix. A square matrix having only zeros above its diagonal is called lower triangular matrix. A matrix that is either upper triangular or lower triangular is called triangular matrix.

Example 2.10. $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 4 & 3 \end{bmatrix}$, & $\begin{bmatrix} 3 & 0 \\ 6 & 4 \end{bmatrix}$ *are lower triangular matrices and* $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, *and* $\begin{bmatrix} 3 & 7 \\ 0 & 4 \end{bmatrix}$ *are upper triangular matrices.*

Definition 2.3. *A matrix obtained by deleting one or more rows and/or columns of A is called sub matrix of A.*

Example 2.11. *If* $A = \begin{bmatrix} 4 & 6 & 1 \\ 3 & 8 & 2 \\ 2 & 0 & 3 \end{bmatrix}$, *then* $\begin{bmatrix} 8 & 2 \\ 0 & 3 \end{bmatrix}$, $\begin{bmatrix} 4 & 6 \\ 3 & 8 \end{bmatrix}$ *and* $\begin{bmatrix} 4 & 6 \\ 3 & 8 \\ 2 & 0 \end{bmatrix}$ *are some of the sub matrices of A.*

2.1.2 Operations on Matrices

Addition and Subtraction of matrices

Definition 2.4. *If A and B are matrices of the same order, then sum of A and B , denoted by* $A+B$*, is the new matrix of the same order obtained by adding the corresponding elements of A and B. Similarly, the difference of A and B , denoted by* $A-B$*, is the matrix obtained by subtracting the corresponding elements ofA and B.*

Example 2.12. *1. If* $A = \begin{bmatrix} 1 & 2 & 8 \\ 4 & 6 & 0 \\ 1 & 3 & 5 \end{bmatrix}$ *and* $B = \begin{bmatrix} 2 & 0 & 3 \\ 3 & 8 & 2 \\ 4 & 6 & 1 \end{bmatrix}$*, then we have that*

$$(a)\ A + B = \begin{bmatrix} 1 & 2 & 8 \\ 4 & 6 & 0 \\ 1 & 3 & 5 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 3 \\ 3 & 8 & 2 \\ 4 & 6 & 1 \end{bmatrix} = \begin{bmatrix} 1+2 & 2+0 & 8+3 \\ 4+3 & 6+8 & 0+2 \\ 1+4 & 3+6 & 5+1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 11 \\ 7 & 14 & 2 \\ 5 & 9 & 6 \end{bmatrix}$$

$$(b)\ A - B = \begin{bmatrix} 1 & 2 & 8 \\ 4 & 6 & 0 \\ 1 & 3 & 5 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 3 \\ 3 & 8 & 2 \\ 4 & 6 & 1 \end{bmatrix} = \begin{bmatrix} 1-2 & 2-0 & 8-3 \\ 4-3 & 6-8 & 0-2 \\ 1-4 & 3-6 & 5-1 \end{bmatrix} = \begin{bmatrix} -1 & 2 & 5 \\ 1 & -2 & -2 \\ -3 & -3 & 4 \end{bmatrix}.$$

2.1.3 Multiplication of Matrices by Scalar

Let A be any matrix and α be a scalar (real number), then αA is the matrix obtained from A multiplying each element of A by α. This operation is called scalar multiplication.
In particular, $-A$ is the matrix obtained from A by multiplying each element of A by -1 and is called the additive inverse of A.

Example 2.13. *If* $A = \begin{bmatrix} 1 & 2 & 6 \\ 5 & 0 & 4 \\ 3 & 1 & 2 \end{bmatrix}$ *, then*

$$3A = 3\begin{bmatrix} 1 & 2 & 6 \\ 5 & 0 & 4 \\ 3 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 3(1) & 3(2) & 3(6) \\ 3(5) & 3(0) & 3(4) \\ 3(3) & 3(1) & 3(2) \end{bmatrix} = \begin{bmatrix} 3 & 6 & 18 \\ 15 & 0 & 12 \\ 9 & 3 & 6 \end{bmatrix}$$

Properties on Matrix addition and Scalar Multiplication

Let A,B and C be $m \times n$ matrices and $\mathbf{0}$ be a zero matrix of size $m \times n$ and $\alpha,\ \beta$ be scalars. Then

1. $A + B = B + A$ (Commutative law for addition)
2. $(A + B) + C = A + (B + C)$ (Associative law for addition)
3. $A + \mathbf{0} = \mathbf{0} + A = A$ ($\mathbf{0}$ is called Additive identity)
4. For each matrix A , there exists a unique $m \times n$ matrix $-A$ such that $A + (-A) = \mathbf{0} = -A + A$
5. $\alpha(A + B) = \alpha A + \alpha B$
6. $(\alpha\beta)A = \alpha(\beta A) = \beta(\alpha A)$

7. $(\alpha+\beta)A = \alpha A + \beta A$

From the above properties, the set of all matrices having the same order forms a vector space with the operations addition and scalar multiplication.

2.2 Product of matrices and some algebraic properties, Transpose of a matrix

Definition 2.5. *(Matrix Product) Let A be an $m \times r$ matrix and B be an $r \times n$. The $(ij)^{th}$ entry of $C = AB$ is the dot product of the i^{th} row vector of A and the j^{th} column vector of B:*

$$c_{ij} = \begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & \cdots & a_{ir} \end{bmatrix} \cdot \begin{bmatrix} b_{1j} \\ b_{2j} \\ b_{3j} \\ \vdots \\ b_{rj} \end{bmatrix}$$

$$= a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{ir}b_{rj}$$

$$= \Sigma_{k=1}^{r} a_{ik}b_{kj}$$

The product C has order $m \times n$.

Example 2.14. *1. If* $A = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 0 & 5 \\ 3 & 2 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 5 \\ 2 & 3 \\ 4 & 1 \end{bmatrix}$ *then*

$$AB = \begin{bmatrix} 1(2)+1(2)+3(4) & 1(5)+1(3)+3(1) \\ 1(2)+0(2)+5(4) & 1(5)+0(3)+5(1) \\ 3(2)+2(2)+1(4) & 3(5)+2(3)+1(1) \end{bmatrix} = \begin{bmatrix} 2+2+12 & 5+3+3 \\ 2+0+20 & 5+0+5 \\ 6+4+4 & 15+6+1 \end{bmatrix} = \begin{bmatrix} 14 & 11 \\ 22 & 10 \\ 14 & 22 \end{bmatrix}$$

2. $A = \begin{bmatrix} 2 & 3 \\ 7 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 3 \\ 2 & 1 \end{bmatrix}$ *, then* $AB = \begin{bmatrix} 6 & 9 \\ 0 & 21 \end{bmatrix}$ *and* $BA = \begin{bmatrix} 21 & 0 \\ 11 & 6 \end{bmatrix}$*. So,* $AB \neq BA$

Note: In general matrix multiplication is not commutative.

Properties of Matrix Multiplication

Let A,B and C be three matrices of the appropriate sizes. Let α be a scalar. Then

1. $A(BC) = (AB)C$.
2. $A(B+C) = AB + AC$ and $(A+B)C = AC + BC$.
3. $\alpha(AB) = (\alpha A)B = A(\alpha B)$

2.2.1 The Transpose of a Matrix

Definition 2.6. *A matrix obtained from a given matrix* $\boldsymbol{A}$ *by interchanging the rows and columns is called the transpose of A and it is denoted by* A^t *.That is, If* $A = (a_{ij})_{m\times n}$ *then* $A^t = (a_{ji})_{n\times m}$

Example 2.15. *1. If* $A = \begin{bmatrix} 2 & 5 \\ -2 & 3 \\ 4 & 1 \end{bmatrix}$, *then* $A^t = \begin{bmatrix} 2 & -2 & 4 \\ 5 & 3 & 1 \end{bmatrix}$

2. If $B = \begin{bmatrix} 4 \\ 3 \\ -6 \end{bmatrix}$, *then* $B^t = \begin{bmatrix} 4 & 3 & -6 \end{bmatrix}$

Let $A = (a_{ij})_{n\times n}$ be a square matrix. Then A is said to be
i) Symmetric matrix if $A^t = A$
ii) Skew symmetric if $A^t = -A$

Example 2.16. *1.* $A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 0 & -2 \\ 4 & -2 & 1 \end{bmatrix}$, $A^t = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 0 & -2 \\ 4 & -2 & 1 \end{bmatrix}$,

$\implies$ $A^t = A$, *therefore, A is symmetric.*

2. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $A^t = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, *therefore, A is skew-symmetric.*

Properties of Transpose of Matrix

Let A and B be matrices such that addition and multiplication is defined. Then

1. $(A^t)^t = A$
2. $(A + B)^t = A^t + B^t$ And $(AB)^t = B^tA^t$
3. $(\alpha A)^t = \alpha A^t$, $\alpha -$ is a scalar
4. $A = \underbrace{\frac{1}{2}(A + A^t)}_{symmetric} + \underbrace{\frac{1}{2}(A - A^t)}_{skew}$

2.2.2 Trace of a Matrix

Definition 2.7. *let* $A = (a_{ij})_{n\times n}$, *be a square matrix of order n .Then trace of A is defined to be the sum of the diagonal elements of A. That is* $trace(A) = \sum_{i=1}^{n} a_{ii}$.

Notation: The trace of a matrix A is also commonly denoted as $trace(A)$ or $tr(A)$.

Properties of trace of a matrix

If A and B are square matrices, then

* $trace(A + B) = trace(A) + trace(B)$
* $trace(A) + trace(A^t)$
* $trace(cA) = c(trace(A))$
* $trace(AB) = trace(BA)$

Example 2.17. *Find the trace of* $A = \begin{bmatrix} 15 & 6 & 7 \\ 2 & -4 & 2 \\ 3 & 2 & 6 \end{bmatrix}$

Solution: $tr(A) = \sum_{i=1}^{3} a_{ii} = 15 + -4 + 6 = 17$

2.3 Elementary Row Operations and its properties

Definition 2.8. *(Elementary row operation) given any matrix A of order $m \times n$. Any one of the following operations on the matrix is called elementary row operation.*

1. Interchanging any two rows of A $R_i \Leftrightarrow R_j$ (Interchange the i^{th} and j^{th} row)
2. Multiplying a row of **A** by a nonzero constant **k** $R_i \Longrightarrow kR_i$ (Multiply the i^{th} row by scalar **k**)
3. Adding a multiple of one row of A to another row of A . $R_j \Longrightarrow R_j + kR_i$ (add **k** times i^{th} row to j^{th} row).

Example 2.18. *1. Give a matrix* $A = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & -2 \\ 3 & 3 & 6 & -9 \end{bmatrix}$

(a) Interchange rows 1 and 3 of A

$$\Longrightarrow \begin{bmatrix} 3 & 3 & 6 & -9 \\ 2 & 3 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

(b) Multiply the third row of A by 1/3

$$\Longrightarrow \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & -2 \\ 1 & 1 & 2 & -3 \end{bmatrix}$$

(c) *Multiply the second row of A by -2, then add to the third row of A*

$$\Longrightarrow \quad \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & -2 \\ -1 & -3 & 6 & -5 \end{bmatrix}$$

Definition 2.9. *Two matrices A and B are called row equivalent or simply (equivalent matrices) if one matrix can be obtained from the other matrix by applying finite number of elementary operations. In this case we write $A \sim B$.*

Example 2.19. *As we observe from the above example,*

$$\begin{bmatrix} 3 & 3 & 6 & -9 \\ 2 & 3 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 3 & 3 & 6 & -9 \\ 2/3 & 1 & 0 & -2/3 \\ 0 & 0 & 1 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 3 & 3 & 6 & -9 \\ 2 & 3 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & -2 \\ -1 & -3 & 6 & -5 \end{bmatrix}$$

Definition 2.10. *(Matrix in reduced row echelon form):*
A matrix in reduced row echelon form has the following properties:

1. *All rows consisting entirely of 0 are at the bottom of the matrix.*
2. *For each nonzero row, the first entry is 1. The first entry is called a leading 1.*
3. *For two successive non zero rows, the leading 1 in the higher row appears farther to the left than the leading 1 in the lower row.*
4. *If a column contains a leading 1, then all other entries in that column are 0.*

Note: A matrix is in row echelon form as the matrix has the first 3 properties.

Example 2.20. $A = \begin{bmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
are the matrices in reduced row echelon form. Where as the matrix.

$$\begin{bmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & -2 & 5 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is not in reduced row echelon form but it is row echelon form since the matrix has the first 3 properties and all the other entries above the leading 1 in the third column are not **0**. The matrix

$$\begin{bmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & -2 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is not in row echelon form (also not in reduced row echelon form) since the leading 1 in the second row is not in the left of the leading 1 in the third row and all the other entries above the leading 1 in the third column are not **0**.

Definition 2.11. *(Rank of a matrix) The rank of a matrix A , denoted by rank (A), is the number of nonzero rows remaining after it has been changed into row echelon or reduced row echelon form.*

Remark: If A is zero matrix then **rank(A)** is **0**.

Example 2.21. *Determine the rank of the following matrices.*

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 3 & 1 & 2 & 6 \\ -1 & 2 & 5 & -4 \\ 2 & 3 & 7 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 3 & 6 & 9 & -3 \\ 2 & 4 & 6 & -2 \end{bmatrix}$$

Solution:

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 3 & 1 & 2 & 6 \\ -1 & 2 & 5 & -4 \\ 2 & 3 & 7 & 2 \end{bmatrix} \begin{matrix} R_2 \to R_2 - 3R_1 \\ R_3 \to R_3 + R_1 \\ R_4 \to R_4 - 2R_1 \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 5 & -3 \\ 0 & 3 & 7 & 0 \end{bmatrix} \begin{matrix} R_3 \to R_3 - 2R_2 \\ R_4 \to R_4 - 3R_2 \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & -9 \\ 0 & 0 & 1 & -9 \end{bmatrix}$$

$$R_4 \to R_4 - R_3 \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & -9 \\ 0 & 0 & 0 & 0 \end{bmatrix} R_2 \to R_2 - 2R_3 \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 21 \\ 0 & 0 & 1 & -9 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence, ***rank(A) = 3*** *by using elementary row operations.*

$$B = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 3 & 6 & 9 & -3 \\ 2 & 4 & 6 & -2 \end{bmatrix} \begin{matrix} R_2 \to R_2 - 3R_1 \\ R_3 \to R_3 - 2R_1 \end{matrix} \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Therefore, ***rank(B)****=1*

Exercise:

1. Transform the following matrix in to reduced row echelon form and determine the rank of the following matrices.

(a) $A = \begin{bmatrix} 10 & -8 & 0 \\ 1 & 3 & -5 \\ 7 & 0 & 9 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 1 & -1 & 3 \\ 1 & -1 & 2 & 1 \\ -4 & 6 & -7 & 1 \\ 2 & 0 & 1 & 3 \end{bmatrix}$

(b) $C = \begin{bmatrix} 2 & 6 & 7 & 9 \\ 3 & 4 & 5 & -1 \\ 1 & 2 & 3 & 1 \\ 2 & 5 & 8 & 4 \\ -1 & 2 & 2 & 10 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 1 & -1 & 3 \\ 1 & -1 & 2 & 1 \\ -4 & 6 & -7 & 1 \\ 2 & 8 & 0 & 3 \end{bmatrix}$

2.4 Inverse of a matrix and its properties

Suppose A and B are square matrices of size n such that $AB = I_n$ and $BA = I_n$. Then A is invertible or non-singular and B is the inverse of A. In this situation, we write $B = A^{-1}$. Notice that if B is the inverse of A, then we can just as easily say A is the inverse of B, or A and B are inverses of each other.

Example 2.22. *Show that* $B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ *is an inverse for the matrix* $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$*:*

Solution:-*By the definition there are two multiplications to confirm. (We will show later that this isn't necessary, but right now we are working strictly from the definition.) We have*

$$\begin{aligned} AB &= \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2(2) + (-1)1 & 2(1) + (-1)2 \\ (-1)1 + 1(1) & -1(1) + 1(2) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= I_2 \end{aligned}$$

and similarly

$$\begin{aligned} BA &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1(2)+1(-1) & 1(-1)+1(1) \\ 1(2)+2(1) & 1(-1)+2(2) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= I_2 \end{aligned}$$

Therefore the definition for inverse is satisfied, so that A and B work as inverses to each other.

Example 2.23. *Matrix* $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ *cannot have an inverse.*

Theorem 2.1. *Suppose that* $\boldsymbol{A}$ *is invertible and that both* $\boldsymbol{B}$ *and* $\boldsymbol{C}$ *are inverses of* $\boldsymbol{A}$*. Then* $\boldsymbol{B} = \boldsymbol{C}$ *and we will denote the inverse as* A^{-1} *.*

Computing the Inverse of a Non Singular Matrix

Suppose A is a non singular square matrix of size **n**. Create the $n \times n$ matrix **M** by placing the $n \times n$ identity matrix in to the right of the matrix A. Let **N** be a matrix that is row-equivalent to **M** and in reduced row-echelon form then the first $\boldsymbol{n}$ columns of **N** is I_n and the last $\boldsymbol{n}$ columns of **N** is A^{-1}.

Example 2.24. *Computing a Matrix Inverse of* $B = \begin{bmatrix} -7 & 6 & 12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix}$

Solution:*The augmented matrix is*

$$[B \mid I] = \begin{bmatrix} -7 & -6 & -12 & 1 & 0 & 0 \\ 5 & 5 & 7 & 0 & 1 & 0 \\ 1 & 0 & 4 & 0 & 0 & 1 \end{bmatrix}$$

by applying elementary row operation the equivalent reduced row echelon form is

$$\begin{bmatrix} 1 & 0 & 0 & -10 & -12 & 9 \\ 0 & 1 & 0 & 13/2 & 8 & 11/2 \\ 0 & 0 & 1 & 5/2 & 3 & 5/2 \end{bmatrix}$$

So

$$B^{-1} = \begin{bmatrix} -10 & -12 & 9 \\ 13/2 & 8 & 11/2 \\ 5/2 & 3 & 5/2 \end{bmatrix}$$

Properties of inverse matrix

Let **A, B, C** be matrices of the appropriate sizes so that the following multiplications make sense, **I** a suitably sized identity matrix, and α a nonzero scalar. Then

1. (Uniqueness) The matrix A has at most one inverse, henceforth denoted as A^{-1}, provided A is invertible.

2. (Double Inverse) If A is invertible, then $(A^{-1})^{-1} = A$:

3. (2=3 Rule) If any two of the three matrices **A, B** and **AB** are invertible, then so is the third, and moreover $(AB)^{-1} = B^{-1}A^{-1}$:

4. A^n is invertible and $(A^n)^{-1} = (A^{-1})^n$.

5. If A is invertible, then $(\alpha A)^{-1} = (\frac{1}{\alpha})A^{-1}$:

6. (Inverse/Transpose) If A is invertible, then $(A^t)^{-1} = (A^{-1})^t$.

7. (Cancellation) Suppose A is invertible. If $AB = AC$ or $BA = CA$, then $B = C$:

2.5 Determinant of a matrix and its properties

Definition 2.12. ***(Determinant of a matrix)*** *Let A be an $n \times n$ matrix. Then the determinant of A denoted as $det(A)$ or $\mid A \mid$ is defined recursively by:*
If $A = [a]$ is a 1×1 matrix, then $det(A) = a$. If A is a matrix of size ***n*** *within ,$n \geqslant 2$ then* $det(A) = A_{11}det(A_{11}) - A_{12}det(A_{12}) + A_{13}det(A_{13}) - \cdots + (-1)^{n+1}A_{1n}det(A_{1n})$ *where A_{1j} a sub matrix of A obtaining by deleting the fist row and the j^{th} column.*

So to compute the determinant of a 5×5 matrix we must build 5 sub matrices, each of size 4. To compute the determinants of each the 4×4 matrices we need to create 4 sub matrices each, these now of size 3 and so on. To compute the determinant of a 10×10 matrix would require computing the determinant of $10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 = 3,628,800$ 1×1 matrices. Fortunately there are better ways. However this does suggest an excellent computer programming exercise to write a recursive procedure to compute a determinant.

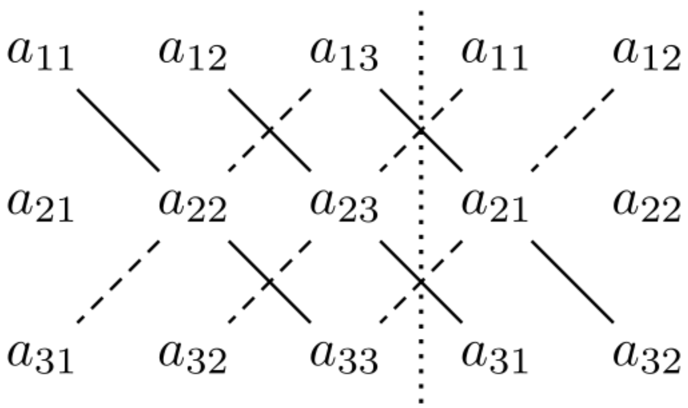
Lets compute the determinant of a reasonable sized matrix by hand.

Suppose that we have the 3×3 matrix $A = \begin{bmatrix} 3 & 2 & -1 \\ 4 & 1 & 6 \\ -3 & -1 & 2 \end{bmatrix}$ then

$$\begin{aligned} det(A) &= \mid A \mid \\ &= \begin{vmatrix} 3 & 2 & -1 \\ 4 & 1 & 6 \\ -3 & -1 & 2 \end{vmatrix} \\ &= 3 \begin{vmatrix} 1 & 6 \\ -1 & 2 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ -3 & 2 \end{vmatrix} - 1 \begin{vmatrix} 4 & 1 \\ -3 & -1 \end{vmatrix} \\ &= 3(1\left|2\right| - 6\left|-1\right|) - 2(4\left|2\right| - 6\left|-3\right|) - (4\left|-1\right| - \left|-3\right|) \\ &= 3(1(2) - 6(-1)) - 2(4(2) - 6(-3)) - (4(-1) - (-3)) \\ &= 24 - 52 + 1 \\ &= -27 \end{aligned}$$

Theorem 2.2. *(Exchanging Columns Changes the Sign of a Determinant). If the matrix A' is obtained from A by interchanging any two columns, and their determinants exist, then $|A'| = -|A|$.*

The rule of Sarrus is a mnemonic for the 3×3 matrix determinant: the sum of the products of three diagonal north-west to south-east lines of matrix elements, minus the sum of the products of three diagonal south-west to north-east lines of elements, when the copies of the first two columns of the matrix are written beside it as in the illustration. This scheme for calculating the determinant of a 3×3 matrix does not carry over into higher dimensions.



Properties of determinants of matrix

1. If A is a triangular matrix, then the determinant of A is the product of all the diagonal elements of A.

2. If B is obtained from A by multiplying one row of A by the scalar α, then $det(B) = \alpha(det(A))$.

3. If B is obtained from A by adding a multiple of one row of A to another row of A, then $det(B) = det(A)$.

4. The matrix A is invertible if and only if $det(A) \neq 0$ and $det(A^{-1}) = \dfrac{1}{det(A)}$.

5. The determinant of a product of two matrices is the product of their determinants. That is,
   $det(AB) = det(A)det(B) \quad \Longrightarrow \quad det(A^n) = (det(A))^n$

6. If B is the transpose of a matrix A, then $det(B) = det(A)$

Minor In a Matrix

Suppose A is an $n \times n$ matrix and A_{ij} is the $(n-1) \times (n-1)$ sub matrix formed by removing row i and column j. Then the **minor** for A at location i j is the determinant of the sub matrix, $M_{ij}(A) = det(A_{ij})$.

Co factor In a Matrix

Suppose A is an $n \times n$ matrix and A_{ij} is the $(n-1) \times (n-1)$ sub matrix formed by removing row i and column j. Then the **Co factor** for A at location i j is the determinant of the sub matrix, $C_{ij}(A) = (-1)^{i+j}det(A_{ij})$.

Definition 2.13. *(Adjoint) If $A = (a_{ij})$ is an $n \times n$ matrix, the adjoint of A, denoted by* ***adjA****, is the transpose of the matrix of cofactors.*

Hence

$$adj A = \begin{bmatrix} c_{11} & c_{21} & c_{31} & \cdots & c_{n1} \\ c_{12} & c_{22} & c_{32} & \cdots & c_{n2} \\ c_{13} & c_{23} & c_{33} & \cdots & c_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & c_{3n} & \cdots & c_{nn} \end{bmatrix}$$

Theorem 2.3. *Let A be an $n \times n$ matrix. Then*

$$A(adj\ A) = (det\ A)I_n = (adj\ A)A.$$

Note If the $det(A) \neq 0$, then $A^{-1} = \frac{1}{det(A)} adj(A)$

Example 2.25. *Let* $\begin{bmatrix} 2 & 3 & -1 \\ 0 & -4 & 2 \\ 1 & -1 & 5 \end{bmatrix}$.

1. *Determine*
   (a) *The minors of all elements A.*
   (b) *The co factors of all elements of A.*
   (c) *The adj(A).*
   (d) *The inverse of A.*

Solution: Exercise

2.6 Solving system of linear equations

Definition 2.14. *A linear equation in the variables $x_1,\ x_2,\ \cdots,\ x_n$ is an equation of the form $a_1x_1 + a_2x_2 + \ \cdots \ + a_nx_n = b$ where the coefficients $a_1,\ a_2,\ \cdots,\ a_n$ and right hand side constant term b are given constants.*

Definition 2.15. *A general system of m linear equations with n unknowns can be written as*

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n & = & b_m. \end{array} \tag{2.1}$$

Here $x_1, x_2, \ldots, x_n$, are the unknowns, $a_{11}, a_{12}, \ldots, a_{mn}$ are the coefficients of the system, and $b_1, b_2, \ldots, b_m$ are the constant terms.

A solution of a linear equation is any n-tuple of values $(s_1, s_2, ..., s_n)$ which satisfies the linear equation. For example, $(-1, -1)$ is a solution of the linear equation $x + 3y = -4$ since$-1 + (3 \times -1) = -1 + (-3) = -4$, but $(1, 5)$ is not.

Similarly, a solution to a linear system is any n-tuple of values $(s_1, s_2, ..., s_n)$ which simultaneously satisfies all the linear equations given in the system.

For example,

$$\begin{aligned} 3x + 2y - z &= 1 \\ 2x - 2y + 4z &= -2 \\ -x + \tfrac{1}{2}y - z &= 0 \end{aligned}$$

has as its solution $(1, -2, -2)$. This can also be written as:

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

We also refer to the collection of all possible solutions as the solution set.

In general, for any linear system of equations there are three possibilities regarding solutions:

1. **A unique solution** In this case only one specific solution set exists. Geometrically this implies the n-planes specified by each equation of the linear system all intersect at a unique point in the space that is specified by the variables of the system.
2. **No solution:** The equations are termed inconsistent and specify n-planes in space which do not intersect or overlap. It is not possible to specify a solution set that satisfies all equations of the system.
3. **An infinite range of solutions:** The equations specify n-planes whose intersection is an m-plane where $m \leq n$. This being the case, it is possible to show that an infinite set of solutions within a specific range exists that satisfy the set of linear equations.

Example 2.26. *Given the system of linear equations,*

$$\begin{aligned} x_1 + 2x_2 + x_4 &= 7 \\ x_1 + x_2 + x_3 - x_4 &= 3 \\ 3x_1 + x_2 + 5x_3 - 7x_4 &= 1 \end{aligned}$$

we have $n = 4$ variables and $m = 3$ equations. Also,

$$\begin{array}{lllll} a_{11}=1 & a_{12}=2 & a_{13}=0 & a_{14}=1 & b_1=7 \\ a_{21}=1 & a_{22}=1 & a_{23}=1 & a_{24}=-1 & b_2=3 \\ a_{31}=3 & a_{32}=1 & a_{33}=5 & a_{34}=-7 & b_3=1 \end{array}$$

Additionally, convince yourself that $x_1 = -2, \quad x_2 = 4, \quad x_3 = 2, \quad x_4 = 1$ *is one solution (but it is not the only one!).*

Note that the above system can be written concisely as

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

we may write the above simultaneous equations as
$AX = b$ where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \text{ and } b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

A matrix **A** is called the coefficient matrix of the system, while the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

obtained by adjoining **b** to **A** is called the augmented matrix of the system.
Remark: If $b_i = 0, \quad \forall i = 1, 2, \cdots, m$ then the linear system is called Homogenous otherwise, non Homogenous.

Theorem 2.4. *If* $[A|b]$ *and* $[C|d]$ *are row equivalent, then the systems* $AX = b$ *and* $CX = d$ *have exactly the same solutions.*

Remark: If **A** is an **m** by **n** matrix then the linear system $AX = \mathbf{0}$ has trivial solution $X = \mathbf{0}$.

Theorem 2.5. *If* ***A*** *is an* ***m*** *by* ***n*** *matrix then the equation* $AX = \mathbf{0}$ *has non trivial solution only, when* $Rank(A) < n$ *otherwise if* $Rank(A) = n$ *then the trivial solution is unique.*

The system of equation $AX = b$ can be directly solved in the following cases.

1. If $A = D$, the equation (2.1) become

$$\begin{array}{llll} a_{11}x_1 & & & = b_1 \\ & a_{22}x_2 & & = b_2 \\ & & \ddots & \vdots \\ & & a_{nn}x_n & = b_n \end{array}$$

   The solution is given by $x_i = \dfrac{b_i}{a_{ii}}$, $a_{ii} \neq 0$

2. If $A = L$, the equation (2.1) become

$$\begin{array}{lllll} a_{11}x_1 & & & & = b_1 \\ a_{21}x_1 & +a_{22}x_2 & & & = b_2 \\ & & & \ddots & \vdots \\ a_{n1}x_1 & +a_{n2} & +\cdots+ & a_{nn}x_n & = b_n \end{array}$$

   Solving the first equation and then successively solving the 2nd, 3rd and so on.
   We obtain $x_1 = \dfrac{b_1}{a_{11}}$, $x_2 = \dfrac{b_2 - a_{21}x_1}{a_{22}}, \cdots, x_n = \dfrac{b_2 - (a_{n1}x_1 + a_{n2} + \cdots + a_{n(n-1)}x_n)}{a_{nn}}$,
   where $a_{ii} \neq 0, \quad i = 1, 2, \cdots, n$
   This method of solving equation is called forward substitution method.

3. If $A = U$, the equation (2.1) become

$$\begin{array}{llllll} a_{11}x_1 & +a_{12}x_2 & +a_{13}x_3 & +\cdots & +a_{1n}x_n & = b_1 \\ & a_{22}x_2 & +a_{23}x_3 & +\cdots & +a_{2n}x_n & = b_2 \\ & \ddots & & & & \vdots \\ & & & & a_{nn}x_n & = b_n \end{array}$$

   Solving for the unknowns in the in the order $x_n, \quad x_{n-1}, \cdots, x_1$, we get

$$x_n = \frac{b_n}{a_{nn}}, \quad x_{n-1} = \frac{b_{n-1} - a_{(n-1)nx_n}}{a_{(n-1)(n-1)}}, \quad \cdots, \quad x_1 = \frac{b_1 - (a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n)}{a_{11}}$$

   This method of solving equation is called backward substitution method. Therefore, matrix **A** is solvable if it can be transformed in to any one of the forms **D**, **U**, **L**.

Theorem 2.6. *Consider an* $\boldsymbol{m}$ *equations with* $\boldsymbol{n}$ *variables* $AX = b$ *then*
a) If $rank(A|b) = rank(A) = n$ *then the system has unique solution*
b) If $rank(A|b) = rank(A) < n$ *then the system has infinitely many solutions.*
c) If $rank(A|b) > n$ *then the system has no solution.*
To solve a linear system , we have the following Methods;

2.6.1 Cramer's rule

If $AX = b$ is a linear system of **n** equations in **n** unknowns, and if $detA \neq 0$, then the system has unique solution which can be determined by. $x_i = \dfrac{|A_i|}{|A|}, \quad i = 1,\ 2,\ \cdots,\ n$
Where A_i is the matrix obtained from **A** when i^{th} column of A is replaced by **b**.

Consider the linear system

$$\begin{cases} a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2 \end{cases}$$

which in matrix format is

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

Assume $a_1b_2 - b_1a_2$ nonzero. Then, with help of determinants x and y can be found with Cramer's rule as

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} = \frac{c_1b_2 - b_1c_2}{a_1b_2 - b_1a_2}, \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} = \frac{a_1c_2 - c_1a_2}{a_1b_2 - b_1a_2} \quad .$$

The rules for 3×3 matrices are similar. Given

$$\begin{cases} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \\ a_3x + b_3y + c_3z &= d_3 \end{cases}$$

which in matrix format is

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}.$$

Then the values of x, y and z can be found as follows:

$$x = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}, \quad y = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}, \text{ and } z = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}.$$

Example 2.27. *Solve the following system by cramer's rule.*

$$2x_1 + 3x_2 + 4x_3 = 19$$
$$x_1 + 2x_2 + x_3 = 4$$
$$3x_1 - x_2 + x_3 = 9$$

***Solution**: The coefficient matrix is*

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 2 & 1 \\ 3 & -1 & 1 \end{bmatrix}$$

and column matrix

$$b = \begin{bmatrix} 19 \\ 4 \\ 9 \end{bmatrix}$$

, then

$$det(A) = \begin{vmatrix} 2 & 3 & 4 \\ 1 & 2 & 1 \\ 3 & -1 & 1 \end{vmatrix} = 4 + 9 - 4 - 24 - 3 + 2 = -16 \neq 0$$

then the system has unique solution.

$$A_1 = \begin{bmatrix} 19 & 3 & 4 \\ 4 & 2 & 1 \\ 9 & -1 & 1 \end{bmatrix} \& det(A_1) = \begin{vmatrix} 19 & 3 & 4 \\ 4 & 2 & 1 \\ 9 & -1 & 1 \end{vmatrix} = 38 + 27 - 16 - 72 - 12 + 19 = -16$$

$$A_2 = \begin{bmatrix} 2 & 19 & 4 \\ 1 & 4 & 1 \\ 3 & 9 & 1 \end{bmatrix} \& det(A_2 = \begin{vmatrix} 2 & 19 & 4 \\ 1 & 4 & 1 \\ 3 & 9 & 1 \end{vmatrix} = 8 + 57 + 36 - 48 - 19 - 18 = 16$$

$$A_3 = \begin{bmatrix} 2 & 3 & 19 \\ 1 & 2 & 4 \\ 3 & -1 & 9 \end{bmatrix} \& det(A_3) = \begin{vmatrix} 2 & 3 & 19 \\ 1 & 2 & 4 \\ 3 & -1 & 9 \end{vmatrix} = 36 + 36 - 19 - 114 - 27 + 8 = -80$$

$\therefore \ x_1 = \dfrac{det(A_1)}{det(A)} = \dfrac{-16}{-16} = 1$
$x_2 = \dfrac{det(A_2)}{det(A)} = \dfrac{16}{-16} = -1$
$x_3 = \dfrac{det(A_3)}{det(A)} = \dfrac{-80}{-16} = 5.$
This is the solution of the system.

1. **Problem 1** Use Cramer's Rule to solve each for each of the variables.

$$\begin{array}{rcrcr} x & - & y & = & 4 \\ -x & + & 2y & = & -7 \end{array}$$

$$\begin{array}{rcrcr} -2x & + & y & = & -2 \\ x & - & 2y & = & -2 \end{array}$$

2. **Problem 2** Use Cramer's Rule to solve this system for z.

$$\begin{array}{rcrcrcr} 2x & + & y & + & z & = & 1 \\ 3x & & & + & z & = & 4 \\ x & - & y & - & z & = & 2 \end{array}$$

2.6.2 Gaussian elimination method

Gauss elimination method is used to solve system of linear equations. In this method the linear system of equation is reduced to an upper triangular system by using successive elementary row operations. Finally we solve the value variables by using back ward substitution method. This method will be fail if any of the pivot element a_{ii}, $i = 1, 2, \cdots, n$ becomes zero. In such case we re-write equation in such manner so that pivots are non zero. This procedure is called pivoting.
Consider system $AX = b$
Step 1: Form the augmented matrix $[A|b]$
Step 2: Transform $[A|b]$ to row echelon form $[U|d]$ using row operations.
Step 3: Solve the system $UX = d$ by back substitution.

Example 2.28. *Solve the following system using Gauss elimination method.*

$$\begin{gathered} 2x_1 - 3x_2 + x_3 = 5 \\ 4x_1 + 14x_2 + 12x_3 = 10 \\ 6x_1 + x_2 + 5x_3 = 9 \end{gathered}$$

Solution*: The augmented matrix of the system is*

$$\begin{bmatrix} 2 & -3 & 1 & 5 \\ 4 & 14 & 12 & 10 \\ 6 & 1 & 5 & 9 \end{bmatrix}$$

Applying, elementary row operations on this matrix to change into its echelon form.

$$\begin{bmatrix} 2 & -3 & 1 & 5 \\ 4 & 14 & 12 & 10 \\ 6 & 1 & 5 & 9 \end{bmatrix} \begin{array}{l} R_2 \longrightarrow R_2 - 2R_1 \\ R_3 \longrightarrow R_3 - 3R_1 \end{array} \begin{bmatrix} 2 & -3 & 1 & 5 \\ 0 & 20 & 10 & 0 \\ 0 & 10 & 2 & -6 \end{bmatrix} R_3 \longrightarrow R_3 - 1/2R_2 \begin{bmatrix} 2 & -3 & 1 & 5 \\ 0 & 20 & 10 & 0 \\ 0 & 0 & -3 & -6 \end{bmatrix}$$

Since rank(A)=rank($\underline{A}$)=3=n the solution exists and is unique.

$$\begin{array}{llll} 2x_1 & -3x_2 & +x_3 & = 5 \\ & 20x_2 & +10x_3 & = 0 \\ & & -3x_3 & = -6 \end{array}$$

From this we get $x_3 = 2$. *And using back substitution we have* $x_2 = -1$ *and* $x_1 = 0$
Hence $(0, -1, 2)$ *is the solution of the system.*

2.6.3 Inverse matrix method

Let $AX = b$ is a system of **n** linear equations with **n** unknowns and A is invertible, then the system has unique solution given by inversion method $X = A^{-1}b$.
Note:- When A is not square or is singular, the system may not have a solution or may have more than one solution.

Example 2.29. *Use the inverse of the coefficient matrix to solve the following system*

$$3x_1 + x_2 = 6$$
$$-x_1 + 2x_2 + 2x_3 = -7$$
$$5x_1 - x_3 = 10$$

Solution: *Okay, let's first write down the matrix form of this system.*

$$\begin{bmatrix} 3 & 9 & 0 \\ -1 & 2 & 2 \\ 5 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -7 \\ 10 \end{bmatrix}$$

Now, we found the inverse of the coefficient matrix by using methods of finding Inverses and is the following;

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -1 & 2 & 2 \\ 5 & 0 & -1 \end{bmatrix}$$

$$\implies C_A = \begin{bmatrix} -2 & 9 & -10 \\ 1 & 3 & 5 \\ 2 & -6 & 7 \end{bmatrix}$$

$$\implies adj(A) = \begin{bmatrix} 2 & -1 & 2 \\ 9 & -3 & -6 \\ -10 & 5 & 7 \end{bmatrix}$$

and $det(A) = 3(-2) + 1(9) + 0(-10) = -6 + 9 = 3$, then

$$A^{-1} = 1/3 \begin{bmatrix} 2 & -1 & 2 \\ 9 & -3 & -6 \\ -10 & 5 & 7 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 & 2/3 \\ 3 & -1 & -2 \\ -10/3 & 5/3 & 7/3 \end{bmatrix}$$

$$\therefore \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 & 2/3 \\ 3 & -1 & -2 \\ -10/3 & 5/3 & 7/3 \end{bmatrix} \begin{bmatrix} 6 \\ -7 \\ 10 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 5 \\ -25/3 \end{bmatrix}$$

Now each of the entries of X are $x_1 = 1/3,\ x_2 = 5$ and $x_3 = -25/3$

Exercise

1. Solve the following linear system of equation by using Cramer's rule, Gaussian elimination method, and inverse method.

   a) $\begin{aligned} 2x_1 + 5x_2 + 3x_3 &= 9 \\ 3x_1 + x_2 + 2x_3 &= 3 \\ x_1 + 2x_2 - x_3 &= 6 \end{aligned}$ b) $\begin{aligned} x + z &= 1 \\ 2x + y + z &= 0 \\ x + y + 2z &= 1 \end{aligned}$ c) $\begin{aligned} x + 2y + z &= 3 \\ 2x + 5y - z &= -4 \\ 3x - 2y - z &= 5 \end{aligned}$

2. Use rank of matrix to determine the values of a, b and c so that the following system has:

   a) no solution b) more than one solution c) a unique solution and solve it.

   i) $\begin{aligned} 1x + y - bz &= 1 \\ 2x + 3y + az &= 3 \\ x + ay + 3z &= 2 \end{aligned}$ ii) $\begin{aligned} x + 2y - 3z &= a \\ 2x + 6y - 11z &= b \\ x - 2y + 7z &= c \end{aligned}$ iii) $\begin{aligned} x - 2y + bz &= 3 \\ ax + 2z &= 2 \\ 5x + 2y &= 2 \end{aligned}$

2.7 Eigenvalues and Eigenvectors

Definition 2.16. *: (Eigenvalue, eigenvector):- Let A be a square matrix. Then if λA, is a real number and $\boldsymbol{X}$ a non zero column vector satisfying $AX = \lambda X$, we call $\boldsymbol{X}$ an eigenvector of A, while λ is called an eigenvalue of A. We also say that X is an eigenvector corresponding to the eigenvalue λ.*

Example 2.30. *Let* $A = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$*, then show that* $X = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ *is eigenvector of* $\boldsymbol{A}$ *with* $\lambda = 1$.

If λ is an eigenvalue of an $n \times n$ matrix $\mathbf{A}$, with corresponding eigenvector $\mathbf{X}$, then $(A - \lambda I_n)X = 0$, with $X \neq 0$, so $det(A - \lambda I_n) = 0$ and there are at most $\mathbf{n}$ distinct eigenvalues of A.

Conversely if $det(A - \lambda I_n) = 0$, then $(A - \lambda I_n)X = 0$ has a nontrivial solution **X**.

The equation $det(A - \lambda I_n) = 0$ is called the **characteristic equation** of A, while the polynomial $det(A - \lambda I_n)$ is called the **characteristic polynomial** of **A**. The characteristic polynomial of A is often denoted by $ch_A(\lambda)$.

Hence the eigenvalues of A are the roots of the characteristic polynomial of A.

Example 2.31. *Find the eigenvalues for the matrix*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

Theorem 2.7. *if A is an $n \times n$ triangular matrix (upper triangular, lower triangular, or diagonal), then the eigenvalues of A are the entries on the main diagonal of A.*

Example 2.32. *Find the eigenvalues of $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and find all eigenvectors.*
Solution: *The characteristic equation of A is $\lambda^2 - 4\lambda + 3 = 0$, or $(\lambda - 1)(\lambda - 3) = 0$ Hence $\lambda = 1$ or $\lambda = 3$. The eigenvector equation $(A - I_n)X = 0$ reduces to*

$$\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or

$$(2-\lambda)x + y = 0$$
$$x + (2-\lambda)y = 0$$

Taking $\lambda = 1$ gives

$$x + y = 0$$
$$x + y = 0.$$

which has solution $x = -y$, and let $y = t$ is arbitrary non zero. Consequently the eigenvectors corresponding to $\lambda = 1$ are the vectors

$$\begin{bmatrix} -t \\ t \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

with $t \neq 0$ *which is the scalar multiple of* $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Taking $\lambda = 3$ *gives*

$$\begin{aligned} -x + y &= 0 \\ x - y &= 0. \end{aligned}$$

which has solution $x = y$, *and let* $y = t$ *is arbitrary non zero. Consequently the eigenvectors corresponding to* $\lambda = 3$ *are the vectors*

$$\begin{bmatrix} t \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

with $t \neq 0$ *hence the scalar multiple of* $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Therefore $\lambda_1 = 1$ *and* $\lambda_2 = 3$ *are the eigenvalues of* $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ *and the corresponding eigenvector are* $X_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *and* $X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *respectively.*

Theorem 2.8. *If* A *is an* $n \times n$ *matrix, the following statements are equivalent.*

(a) λ *is an eigenvalue of* A.

(b) The system of equations $(A - \lambda I)\mathbf{x} = 0$ *has nontrivial solutions.*

(c) There is a nonzero vector $\mathbf{x}$ *such that* $A\mathbf{x} = \lambda\mathbf{x}$.

(d) λ *is a solution of the characteristic equation* $det(A - \lambda I) = 0$.

2.7.1 Diagonalization

Problem 1 Given an $n \times n$ matrix A, does there exist an invertible matrix P such that

$$P^{-1}AP$$

is diagonal?

Problem 2 Given an $n \times n$ matrix A, does A have n linearly independent eigenvectors?

Theorem 2.9. *Let* $\boldsymbol{A}$ *be an* $n \times n$ *matrix having distinct eigenvalues* $\lambda_1,\ \lambda_2, \cdots, \lambda_n$ *and corresponding eigenvectors* $X_1,\ X_2,\ \cdots\ X_n$ *respectively. Let* P *be the matrix whose columns are respectively* $X_1,\ X_2,\ \cdots,\ X_n$*. Then* P *is non singular and*

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Definition 2.17. *If* A *and* B *are square matrices, then we say that* B *is* **similar** *to* A *if there is an invertible matrix* P *such that*

$$B = P^{-1}AP.$$

Definition 2.18. *A square matrix* A *is said to be* **diagonalizable** *if it is similar to some diagonal matrix. In other words,* A *is diagonalizable if there exists an invertible matrix* P *such that* $P^{-1}AP$ *is diagonal. In this case the matrix* P *is said to* **diagonalize** A*.*

Theorem 2.10. *If* A *is an* $n \times n$ *matrix, the following statements are equivalent.*

(a) A *is diagonalizable.*

(b) A *has* n *linearly independent eigenvectors.*

Procedure for Diagonalizing a Matrix

1. Confirm that the matrix is actually diagonalizable by finding n linearly independent eigenvectors. One way to do this is by finding a basis for each eigenspace and merging these basis vectors into a single set S. If this set has fewer than n vectors, then the matrix is not diagonalizable.

2. Form the matrix

$$P = [\mathbf{p}_1\ \mathbf{p}_2\ \cdots\ \mathbf{p}_n]$$

that has the vectors in S as its column vectors.

3. The matrix $P^{-1}AP$ will be diagonal and have the eigenvalues

$$\lambda_1,\ \lambda_2 \cdots, \lambda_n$$

corresponding to the eigenvectors

$$\mathbf{p}_1,\ \mathbf{p}_2,\ \cdots, \mathbf{p}_n$$

as its successive diagonal entries.

Example 2.33. *In each of the following, determine if the the matrix is diagonalizable*

$$A = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} \quad A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix} \quad A = \begin{bmatrix} -1 & 4 & -2 \\ -3 & 4 & 0 \\ -3 & 1 & 3 \end{bmatrix}$$

CHAPTER THREE

3 Limit and Continuity

3.1 Definition of Limit

Informal Definition of a Limit

As the precise definition of a limit is a bit technical, it is easier to start with an informal definition; we'll explain the formal definition later.

We suppose that a function f is defined for x near c (but we do not require that it be defined when $x = c$).

Definition 3.1. *(Informal definition of a limit)*
We call L the limit of $f(x)$ as x approaches c if $f(x)$ becomes close to L when x is close (but not equal) to c , and if there is no other value L' with the same property.

When this holds we write

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

or

$$f(x) \to L \quad as \quad x \to c$$

Notice that the definition of a limit is not concerned with the value of $f(x)$ when $x = c$ (which may exist or may not). All we care about are the values of $f(x)$ when x is close to c , on either the left or the right (i.e. less or greater).

Example 3.1. *Let $f(x) = x^2 + 5$ determine the limit of $f(x)$ as x approaches to 2.*
***Solution**: The value of $f(x)$ as x approaches to 2 listed in the following table.*

x	*1.9*	*1.99*	*1.999*	*2.001*	*2.01*	*2.1*
f(x)	*8.61*	*8.9601*	*8.99601*	*9.004*	*9.04*	*9.4*

On the basis of the values in the tables, we can conclude that

$$\lim_{x\to 2}(x^2+5)=9$$

Remark: The limit of the function does not depend on the value of the function at the point for which the limit is computed.

Example 3.2. *Guess the value of* $\lim_{x\to 1}\dfrac{x-1}{x^2-1}$.

Solution:
Notice that the function $f(x)=\dfrac{x-1}{x^2-1}$ *is not defined when* $x=1$, *but that doesn't matter because the definition of* $\lim_{x\to 1} f(x)$ *says that we consider values of* x *that are close to* 1 *but not equal to* 1 . *The tables below give values of* $f(x)$ *(correct to six decimal places) for values of* x *that approach* 1 *(but are not equal to* 1*).*

$x<1$	$f(x)$
0.5	0.666667
0.9	0.526316
0.99	0.502513
0.999	0.500250
0.9999	0.500025

&

$x>1$	$f(x)$
1.5	0.400000
1.1	0.476190
1.01	0.497512
1.001	0.499750
1.0001	0.499975

On the basis of the values in the tables, we make the guess that

$$\lim_{x\to 1}\frac{x-1}{x^2-1}=0.5$$

Example is illustrated by the graph of $y=\dfrac{x-1}{x^2-1}$ in Figure below.

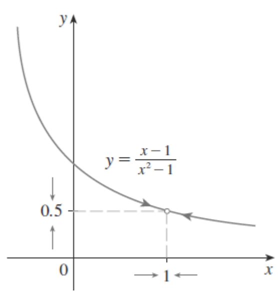


Figure 3.1: limit f as x approuches to 1

Exercise:

1. Estimate the value of

(a) $\lim_{x\to 0} \frac{\sqrt{x^2+9}-3}{x^2}$

(b) $\lim_{x\to 0} \frac{x-1}{x^2-1}$

(c) $\lim_{x\to 0} \sin\frac{\pi}{x}$

Definition 3.2. *Let f be a function defined on some open interval that contains the number c, except possibly at c itself. Then we say that the limit of f as x approaches c is L, and we write*

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

if for every number $\epsilon > 0$ there is a number $\delta > 0$ such that

$$\textit{if } 0 <\mid x-c \mid< \delta \textit{ then } \mid f(x) - L \mid< \epsilon$$

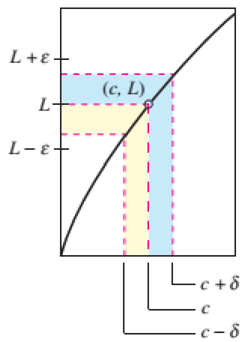


Figure 3.2: The $\epsilon - \delta$ definition of the limit of $f(x)$ as x approaches c

Since $\mid x - c \mid$ is the distance from x to c and $\mid f(x) - L \mid$ is the distance from $f(x)$ to L,and ϵ can be arbitrarily small, the definition of a limit can be expressed in words as follows:

$\lim_{x\to c} f(x) = L$ means that the distance between $f(x)$ and L can be made arbitrarily small by taking the distance from x to a sufficiently small (but not 0).

Alternatively, $\lim_{x\to c} f(x) = L$ means that the values of $f(x)$ can be made as close as we please to L by taking x close enough to c (but not equal to c). We can also reformulate definition in terms of intervals by observing that the inequality $\mid x - c \mid< \delta$ is equivalent to $-\delta < x - c < \delta$, which in turn can be written as $c - \delta < x < c + \delta$. Also $0 <\mid x - c \mid$ is true if and only if $x - c \neq 0$, that is, $x \neq a$. Similarly, the inequality $\mid f(x) - L \mid< \epsilon$ is equivalent to the pair of inequalities $L - \epsilon < f(x) < L + \epsilon$. Therefore, in terms of intervals, definition can be stated as follows:

$\lim_{x\to c} f(x)$ means that for every $\epsilon > 0$ (no matter how ϵ is small) we can find $\delta > 0$ such that if x lies in the open interval $(c - \delta, c + \delta)$ and $x \neq c$, then $f(x)$ lies in the open interval $(L - \epsilon, L + \epsilon)$

Example 3.3. *Prove that* $\lim_{x\to 3}(4x-5)=7$.

Solution:-

Let ϵ be a given positive number. We want to find a number δ such that $|(4x-5)-7|<\epsilon$ whenever $0<|x-3|<\delta$. But $|(4x-5)-7|=|4x-12|=4|x-3|$. Therefore, we want $4|x-3|<\epsilon$ whenever $|x-3|<\delta$ this implies $|x-3|<\dfrac{\epsilon}{4}$ whenever $|x-3|<\delta$ This suggests that we should choose $\delta=\dfrac{\epsilon}{4}$.

Example 3.4. *Using formal definition of limit prove that* $\lim_{x\to 2}x^2=4$

Solution:

We must show that for any $\epsilon>0$ there exists $\delta>0$ such that $|x^2-4|<\epsilon$ when ever $0<|x-2|<\delta$ Factoring, we get $|x^2-4|=|x-2||x+2|$ we want to show that $|x^2-4|$ is small when x is close to 2. To do this, we first find an upper bound for the factor $|x+2|$. If x is close to 2, we know that the factor $|x-2|$ is small, and that the factor $|x+2|$ is close to 4. Because we are considering values of x close to 2, we can concern ourselves with only those values of x for which $|x-2|<1$; that is, we are requiring the δ, for which we are looking, to be less than or equal to 1. The inequality $|x-2|<1$ is equivalent to $-1<x-2<1$ which is equivalent to $1<x<3$ or, equivalently, $3<x+2<5$. This means that if $|x-2|<1$, then $3<|x+2|<5$ therefore, we have $|x^2-4|=|x-2||x+2|<5|x-2|$. Now we want $5|x-2|<\epsilon$ or, equivalently $|x-2|<\dfrac{1}{5}\epsilon$. Thus, if we choose δ to be the smaller of 1 and $\dfrac{\epsilon}{5}$, then when ever $|x-2|<\delta$, it follows that $|x-2|<\dfrac{1}{5}\epsilon$ and $|x+2|<5$ because this is true when $|x-2|<1$ and so $|x^2-4|<\dfrac{\epsilon}{5}$.

Therefore, we conclude that $|x^2-4|<\epsilon$ whenever $0<|x-2|<\delta$ if δ is the smaller of the two numbers 1 and $\dfrac{\epsilon}{5}$, which we write as $\delta=min(1,\dfrac{\epsilon}{5})$.

Exercise Prove that

$$a)\lim_{x\to 3}(-2x+7)=1 \quad b)\lim_{x\to 0}\frac{x^2}{x^2+1}=0, \quad c)\lim_{x\to 3}(x^2)=9$$
$$d)\lim_{x\to -3}(x^2-9)=0, \quad e)\lim_{x\to 0}(x^2)=0 \quad f)\lim_{x\to a}(-bx+c)-ab+c$$

3.2 Basic limit theorems

Theorem 3.1. *Suppose that c is a constant and the limits* $\lim_{x\to a}f(x)=L$ *and* $\lim_{x\to a}g(x)=M$ *exist. Then*

1. *The limit of a sum is the sum of the limits.*

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

2. *The limit of a difference is the difference of the limits.*

$$\lim_{x\to a}[f(x) - g(x)] = \lim_{x\to a} f(x) - \lim_{x\to a} g(x) = L - M.$$

3. *The limit of a constant times a function is the constant times the limit of the function.*

$$\lim_{x\to a}[cf(x)] = c\lim_{x\to a} f(x) = cL.$$

4. *The limit of a product is the product of the limits*

$$\lim_{x\to a}[f(x)g(x)] = \lim_{x\to a} f(x) \lim_{x\to a} g(x) = LM.$$

5. *The limit of a quotient is the quotient of the limits (provided that the limit of the denominator is not 0)*

$$\lim_{x\to a}[\frac{f(x)}{g(x)}] = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{L}{M} \quad if \lim_{x\to a} g(x) \neq 0.$$

6. $\lim_{x\to a}[f(x)]^n = [\lim_{x\to a} f(x)]^n$, *where n is a positive integer.*

7. $\lim_{x\to a} c = c$, *where c is any real number.*

8. $\lim_{x\to a} x = a$.

9. $\lim_{x\to a} x^n = a^n$, *where n is a positive integer.*

10. $\lim_{x\to a} \sqrt[n]{x} = \sqrt[n]{a}$, *where n is a positive integer (If n is even, we assume that $a > 0$). More generally, we have the following law, which is proved as a consequence of Law 10*

11. $\lim_{x\to a}(\sqrt[n]{f(x)}) = \sqrt[n]{\lim_{x\to a} f(x)}$, *where n is a positive integer (If n is even, we assume that $\lim_{x\to a} f(x) > 0$).*

Direct Substitution Property

If $f(x)$ is a polynomial or a rational function and a is in the domain of $f(x)$, then

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

Example 3.5. 1. *Evaluate the following limits and justify each step.*

(a) $\lim_{x\to 5} 2x^2 - 3x + 4$

(b) $\lim_{x\to 4} \sqrt[3]{\frac{x}{-7x+1}}$

Solution:

1. (a)

$$\begin{aligned}\lim_{x\to 5}(2x^2-3x+4) &= \lim_{x\to 5}(2x^2)-\lim_{x\to 5}(-3x)+4 \quad \text{(by rule 1 and 2)}\\ &= 2\lim_{x\to 5}x^2-3\lim_{x\to 5}x+4 \ \ \text{(by rule 3 and 7)}\\ &= 2(5^2)-3(5)+4 \ \ \text{(by rule 9)}\\ &= 50-15+4\\ &= 39\end{aligned}$$

(b)

$$\begin{aligned}\lim_{x\to 4}\sqrt[3]{\frac{x}{-7x+1}} &= \sqrt[3]{\lim_{x\to 4}[\frac{x}{-7x+1}]} \ \ \text{(by low 11)}\\ &= \sqrt[3]{\frac{\lim_{x\to 4}x}{\lim_{x\to 4}(-7x+1)}} \ \ \text{(by low 5)}\\ &= \sqrt[3]{\frac{4}{-7(4)+1}} \ \ \text{(by low 3,9 and 1)}\\ &= \sqrt[3]{\frac{4}{-27}}\\ &= -\frac{\sqrt[3]{4}}{3}\end{aligned}$$

Theorem 3.2. *If $f(x) \leqslant g(x)$ when x is near a (except possibly at a) and the limits of f and g both exist as x approaches a, then*

$$\lim_{x\to a} f(x) \leqslant \lim_{x\to a} g(x)$$

Theorem 3.3. *Squeeze Theorem: If $f(x) \leqslant g(x) \leqslant h(x)$ when x is near a (except possibly at a) and*

$$\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = L \quad \textit{then} \ \lim_{x\to a} g(x) = L$$

which is sometimes called the Sandwich Theorem.

Example 3.6. *Show that*

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

Solution for 1:

First note that we cannot use

$$\lim_{x\to 0} x^2 \sin\frac{1}{x} = \lim_{x\to 0} x^2 \cdot \lim_{x\to 0} \sin\frac{1}{x}$$

because $\lim_{x\to 0} \sin\frac{1}{x}$ does not exist. However, since

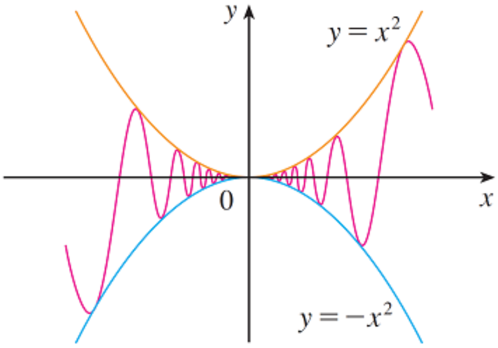
$$-1 \le \sin\frac{1}{x} \le 1$$

Multiply both side by x^2 we get

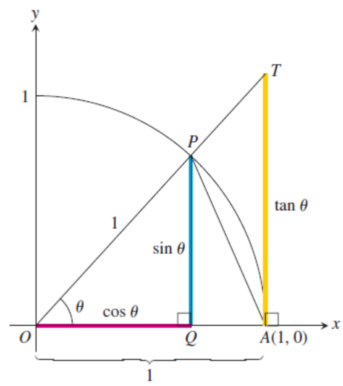
$$-x^2 \le x^2 \sin\frac{1}{x} \le x^2$$

We know that $\lim_{x\to 0} -x^2 = 0$ and $\lim_{x\to 0} x^2 = 0$ taking $f(x) = -x^2$, $h(x) = x^2$ and $g(x) = x^2 \sin\frac{1}{x}$ by the Squeeze theorem, we obtain

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



Proof for 2: To show that the limit is 1, we begin with positive values of θ less than π. Notice that from the following figure



We have $Area\ \triangle OAP \leqslant Area\ sector\ OAP \leqslant Area\ \triangle OAT$.
We can express these areas in terms of θ as follows:
$Area\ \triangle OAP = \frac{1}{2} base \times hieght = \dfrac{1}{2}(1)(\sin\theta) = \dfrac{\sin\theta}{2}$
$area\ sector\ OAP = \frac{1}{2} r^2\theta = \dfrac{1}{2}(1^2)(\theta) = \dfrac{\theta}{2}$
$area\ \triangle OAT = \dfrac{1}{2} base \times hieght = \dfrac{1}{2}(1)(\tan\theta) = \dfrac{\tan\theta}{2}$
Thus $\dfrac{\sin\theta}{2} \leqslant \dfrac{\theta}{2} \leqslant \dfrac{\tan\theta}{2}$.
This last inequality goes the same way if we divide all three terms by the number $\frac{2}{\sin\theta}$ which is positive since $0 < \theta < \pi/2$

$$1 \leqslant \frac{\theta}{\sin\theta} \leqslant \frac{1}{\cos\theta}$$

Taking reciprocals reverses the inequalities:

$$1 \geqslant \frac{\sin\theta}{\theta} \geqslant \frac{\cos\theta}{1}$$

Since $\lim\limits_{\theta\to 0} \dfrac{\cos\theta}{1} = 1$ and $\lim\limits_{\theta\to 0} 1 = 1$ then by the Sandwich theorem gives
$\lim\limits_{\theta\to 0} \dfrac{\sin\theta}{\theta} = 1$

Recall that $\sin\theta$ and θ are both odd functions . Therefore, $f(\theta) = \dfrac{\sin\theta}{\theta}$ is an even function, with a graph symmetric about the y-axis (see the above figure). This symmetry implies that the left-hand limit at 0 exists and has the same value as the right-hand limit:

$$\lim_{\theta\to 0^+} \frac{\sin\theta}{\theta} = 1 = \lim_{\theta\to 0^-} \frac{\sin\theta}{\theta}$$

So $\lim\limits_{\theta\to 0} \dfrac{\sin\theta}{\theta} = 1$

3.3 One sided limits

Definition 3.3. *Let f be a function which is defined at every number in some open interval (a, c). Then the limit of $f(x)$, as x approaches from the right a, is L, written*

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

if for any $\epsilon > 0$, however small, there exists a $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $0 < x - a < \delta$

Example 3.7. *using the definition to prove that* $\lim\limits_{x \to 0^+} \sqrt{x} = 0$

Solution:

1. Guessing a value for δ. Let ϵ be a given positive number. Here $a = 0$ and $L = 0$ so we want to find a number δ such that $|\sqrt{x} - 0| < \epsilon$ whenever $0 < x - 0 < \delta$ that is $\sqrt{x} < \epsilon$ whenever $0 < x < \delta$ or, squaring both sides of the inequality $\sqrt{x} < \epsilon$, we get $x < \epsilon^2$ whenever $0 < x < \delta$ This suggests that we should choose $\delta = \epsilon$

2. Showing that this δ works. Given $\epsilon > 0$, let $\delta = \epsilon^2$. If $0 < x < \delta$, then $\sqrt{x} < \sqrt{\delta} = \sqrt{\epsilon^2} = \epsilon$ So $|\sqrt{x} - 0| < \epsilon$
   According to the definition, this shows that $\lim_{x \to 0^+} \sqrt{x} = 0$

Definition 3.4. *Let f be a function which is defined at every number in some open interval (d, a). Then the limit of $f(x)$, as x approaches from the left a, is L, written*

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

if for any $\epsilon > 0$, however small, there exists a $\delta > 0$ such that
$|f(x) - L| < \epsilon$ whenever $-\delta < x - a < 0$

Example 3.8. *Show that* $\lim\limits_{x \to 4^-} \sqrt{4 - x} = 0$

Solution: For every $\epsilon > 0$ we need to find a $\delta > 0$ such that
if $-\delta < x - 4 < 0$ then $|\sqrt{4 - x} - 0| < \epsilon$
Since $|\sqrt{4 - x} - 0| = |\sqrt{4 - x}| < \epsilon \Rightarrow |4 - x| < \epsilon^2 \Rightarrow -\epsilon^2 < (x - 4) < \epsilon^2 \Rightarrow -\epsilon^2 < (x - 4) < 0$
then if $-\delta < (x - 4) < 0$ then $-\epsilon^2 < (x - 4) < 0$
This suggests that we should choose $\delta = \epsilon^2$
Therefore, we conclude that

$$\lim_{x \to 4^-} \sqrt{4 - x} = 0$$

Theorem 3.4. $\lim_{x\to a} f(x) = L$ *if and only if*

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

Exercise: Show that a) $\lim_{x\to 0} |x| = 0$ and $\lim_{x\to 0} \frac{|x|}{x}$ does not exist.

3.4 Infinite limits, limit at infinity and asymptotes

Definition 3.5. *Let $f(x)$ be a function defined on some open interval that contains the number a, except possibly at a itself. Then*

i $\lim_{x\to a^+} f(x) = \infty$ *means that for every positive number M there is a positive number δ such that*

$f(x) > M$ *whenever* $0 < x - a < \delta$

ii $\lim_{x\to a^-} f(x) = \infty$ *means that for every positive number M there is a positive number δ such that*

$f(x) > M \quad$ *whenever* $\quad -\delta < x - a < 0$

iii $\lim_{x\to a} f(x) = \infty$ *means that for every positive number M there is a positive number δ such that*

$f(x) > M$ *whenever* $|x - a| < \delta$

Example 3.9. *By using the definition prove that* $\lim_{x\to 0} \frac{1}{x^2} = \infty$

Solution: Let M be given large number, we want to find a $\delta > 0$ such that
$f(x) > M \quad$ when ever $\quad 0 < |x - 0| < \delta$
or $\frac{1}{x^2} > M \quad$ when ever $\quad 0 < |x| < \delta$
that is, $x^2 < \frac{1}{M} \quad$ when ever $\quad 0 < |x| < \delta$
or $|x| < \frac{1}{\sqrt{M}} \quad$ when ever $\quad 0 < |x| < \delta$
This suggests that we should take $\delta = \frac{1}{\sqrt{M}}$
Therefore $\lim_{x\to 0} \frac{1}{x^2} = \infty$

Definition 3.6. *Let $f(x)$ be a function defined on some open interval that contains the number a, except possibly at a itself. Then*

i $\lim_{x\to a^+} f(x) = -\infty$ *means that for every negative number N there is a positive number δ such that*

$f(x) < N \quad$ *whenever* $\quad 0 < x - a < \delta$

ii $\lim_{x\to a^-} f(x) = -\infty$ *means that for every negative number N there is a positive number δ such that*

$f(x) < N$ *whenever* $-\delta < x - a < 0$

iii $\lim_{x\to a} f(x) = -\infty$ *means that for every negative number N there is a positive number δ such that*

$f(x) < N$ *whenever* $0 < |x - a| < \delta$

Example 3.10. *Show that* $\lim_{x\to 0^-} \frac{1}{x} = -\infty$

Solution: Let N be given negative small number, we want to find a $\delta > 0$ such that
$f(x) < N$ when ever $-\delta < x - 0 < 0$
or $\frac{1}{x} < N$ when ever $-\delta < x - 0 < 0$
that is, $x > \frac{1}{N}$ when ever $-\delta < x < 0$
This suggests that we should take $\delta = -\frac{1}{N}$
Therefore $\lim_{x\to 0^-} \frac{1}{x} = -\infty$

Definition 3.7. *We say that $f(x)$ has the limit L as x approaches infinity and write*

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

if, for every number $\epsilon > 0$ there exists a corresponding positive number M such that for all x

$$x > M \quad \Rightarrow \quad |f(x) - L| < \epsilon$$

Example 3.11. *Show that*

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

Solution: Let $\epsilon > 0$ be given. We must find a positive number M such that for all x
$x > M \quad \Rightarrow \quad |\frac{1}{x} - 0| = |\frac{1}{x}| < \epsilon$
$x > M \quad \Rightarrow \quad x > \frac{1}{\epsilon}$ because x is positive
The implication will hold if $M = \frac{1}{\epsilon}$ or any larger positive number.
This proves that

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

Definition 3.8. *We say that $f(x)$ has the limit L as x approaches negative infinity and write if, for every number $\epsilon > 0$ there exists a corresponding negative number N such that for all x*

$$x < N \quad \Rightarrow \quad |f(x) - L| < \epsilon$$

Example 3.12. *Show that*

$$\lim_{x\to-\infty}\frac{1}{x}=0$$

Solution: Let $\epsilon > 0$ be given. We must find a negative number N such that for all x

$x < N \quad \Rightarrow \quad |\frac{1}{x}-0| = |\frac{1}{x}| < \epsilon$

$x < N \quad \Rightarrow \quad x < -\frac{1}{\epsilon}$ because x is negative

The implication will hold if $N = -\frac{1}{\epsilon}$ or any small number.

This proves that

$$\lim_{x\to-\infty}\frac{1}{x}=0$$

Theorem 3.5. *Suppose that c is a constant and the limits $\lim_{x\to\pm\infty} f(x) = L$ and $\lim_{x\to\pm\infty} g(x) = M$ exist. Then*

1. *Sum rule* $\lim_{x\to\pm\infty}[f(x)+g(x)] = \lim_{x\to\pm\infty} f(x) + \lim_{x\to\pm\infty} g(x) = L + M.$

2. *Difference rule* $\lim_{x\to\pm\infty}[f(x)-g(x)] = \lim_{x\to\pm\infty} f(x) - \lim_{x\to\pm\infty} g(x) = L - M.$

3. *Constant Rule* $\lim_{x\to\pm\infty}[cf(x)] = c\lim_{x\to\pm\infty} f(x) = cL.$

4. *Product Rule* $\lim_{x\to\pm\infty}[f(x)g(x)] = \lim_{x\to\pm\infty} f(x)\lim_{x\to\pm\infty} g(x) = LM.$

5. *Quotient Rule* $\lim_{x\to\pm\infty}[\frac{f(x)}{g(x)}] = \frac{\lim_{x\to\pm\infty} f(x)}{\lim_{x\to\pm\infty} g(x)} = \frac{L}{M}$ *if* $\lim_{x\to\pm\infty} g(x) \neq 0.$

Example 3.13. *Using the above theorem evaluate the following limits at infinity*

a). $\lim_{x\to\infty}\frac{\sqrt{4x^2+2}}{3x+1}$ *b).* $\lim_{x\to\infty}\frac{x^2+1}{2x-3}$

Solution:

a).

$$\begin{aligned}
\lim_{x\to\infty} \frac{\sqrt{4x^2+2}}{3x+1} &= \lim_{x\to\infty} \frac{\sqrt{x^2(4+\frac{1}{x^2})}}{x(3+\frac{1}{x})} \\
&= \lim_{x\to\infty} \frac{\sqrt{(4+\frac{1}{x^2})}}{(3+\frac{1}{x})} \\
&= \frac{\sqrt{\lim_{x\to\infty}(4+\frac{1}{x^2})}}{\lim_{x\to\infty}(3+\frac{1}{x})} \\
&= \frac{\sqrt{(\lim_{x\to\infty} 4+\lim_{x\to\infty}\frac{1}{x^2})}}{(\lim_{x\to\infty} 3+\lim_{x\to\infty}\frac{1}{x})} \\
&= \frac{\sqrt{(4+0)}}{(3+0)} \\
&= \frac{2}{3}
\end{aligned}$$

b).

$$\begin{aligned}
\lim_{x\to\infty} \frac{x^2+1}{2x-3} &= \lim_{x\to\infty} \frac{x+\frac{1}{x}}{2-\frac{3}{x}} \\
&= \frac{\lim_{x\to\infty} x+\lim_{x\to\infty}\frac{1}{x}}{\lim_{x\to\infty} 2-\lim_{x\to\infty}\frac{3}{x}} \\
&= \frac{\infty+0}{2-0} \\
&= \frac{\infty}{2} \\
&= \infty
\end{aligned}$$

Definition 3.9. *The line $y = b$ is called a horizontal asymptote of the curve $y = f(x)$ if either*

$\lim_{x\to\infty} f(x) = b$ *or/and* $\lim_{x\to-\infty} f(x) = b$

Example 3.14. *Find $\lim_{x\to\infty} \frac{1}{x}$ and $\lim_{x\to-\infty} \frac{1}{x}$.*

Solution: Observe that when is large, is small. For instance, In fact, by taking large enough, we can make as close to 0 as we please. Therefore, according to Definition 1, we have Similar reasoning shows that when x is large negative, $1/x$.
is small negative, so we also have

Definition 3.10. *A line $x = a$ is a vertical asymptote of the graph of a function $y = f(x)$ if either*
$\lim_{x\to a^+} f(x) = \pm\infty$ *or/and* $\lim_{x\to a^-} f(x) = \pm\infty$

Example 3.15. *Find* $\lim_{x\to 3^-}(\frac{2x}{x-3})$ *and* $\lim_{x\to 3^+}(\frac{2x}{x-3})$.

Solution: If x is close to 3 but larger than 3, then the denominator $x - 3$ is a small positive number and $2x$ is close to 6. So the quotient $2x/x - 3$ is a large positive number.

$$\lim_{x\to 3^+}(\frac{2x}{x-3}) = \infty$$

Likewise If x is close to 3 but smaller than 3, then the denominator $x-3$ is a small negative number and $2x$ is close to 6. So the quotient $2x/x - 3$ is a large negative number.

$$\lim_{x\to 3^-}(\frac{2x}{x-3}) = -\infty$$

The graph of the curve $y = \frac{2x}{x-3}$ is given in the following figure. The line is a vertical asymptote.

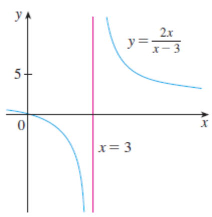


Figure 3.3: Asymptote and graph of $y = \frac{2x}{x-3}$

Example 3.16. *Find the vertical and horizontal asymptotes for the graph of*
a). $f(x) = \dfrac{\sqrt{x^2+2}}{x-1}$ *b).* $f(x) = 2x+1-\sqrt{4x^2+5}$ *c).* $f(x) = \sqrt[3]{\dfrac{x^2+3}{27x^2-1}}$

solution:

a)Solution We are interested in the behavior as $x \to \pm\infty$ and as $x \to 1$ where the denominator is zero.
and then $\lim\limits_{x\to 1^-} \dfrac{\sqrt{x^2+2}}{x-1} = -\infty$, $\lim\limits_{x\to 1^+} \dfrac{\sqrt{x^2+2}}{x-1} = \infty$, $\lim\limits_{x\to\infty} \dfrac{\sqrt{x^2+2}}{x-1} = 1$ and $\lim\limits_{x\to-\infty} \dfrac{\sqrt{x^2+2}}{x-1} = -1$
$\therefore$ $x = 1$ is the vertical asymptote of $f(x) = \dfrac{\sqrt{x^2+2}}{x-1}$ and also $y = 1$ and $y = -1$ are horizontal asymptote of $f(x) = \dfrac{\sqrt{x^2+2}}{x-1}$.

3.5 Continuity and One Sided Continuity

We are now ready to define the concept of a function being continuous. The idea is that we want to say that a function is continuous if you can draw its graph without taking your pencil off the page. But sometimes this will be true for some parts of a graph but not for others. Therefore, we want to start by defining what it means for a function to be continuous at one point. The definition is simple, now that we have the concept of limits:

Definition 3.11. *(continuity at a point) If $f(x)$ is defined on an open interval containing c, then $f(x)$ is said to be continuous at c if and only if*

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

Note that for f to be continuous at c, the definition in effect requires three conditions:

1. that f is defined at c , so $f(c)$ exists,
2. the limit as x approaches c exists, and
3. the limit and $f(c)$ are equal.

If any of these do not hold then f is not continuous at c.

The idea of the definition is that the point of the graph corresponding to c will be close to the points of the graph corresponding to nearby x-values. Now we can define what it means for a function to be continuous in general, not just at one point.

Definition 3.12. *A function is said to be continuous on (a, b) if it is continuous at every point of the interval (a, b) .*

We often use the phrase "the function is continuous" to mean that the function is continuous at every real number. This would be the same as saying the function was continuous on $(-\infty, \infty)$, but it is a bit more convenient to simply say "continuous".

Note that, by what we already know, the limit of a rational, exponential, trigonometric or logarithmic function at a point is just its value at that point, so long as it's defined there. So, all such functions are continuous wherever they're defined. (Of course, they can't be continuous where they're not defined!)

3.5.1 Discontinuities

A discontinuity is a point where a function is not continuous. There are lots of possible ways this could happen, of course. Here we'll just discuss two simple ways.

Removable discontinuities

The function $f(x) = \dfrac{x^2-9}{x-3}$ is not continuous at $x = 3$. It is discontinuous at that point because the fraction then becomes $\dfrac{0}{0}$, which is undefined. Therefore the function fails the first of our three conditions for continuity at the point 3; 3 is just not in its domain.

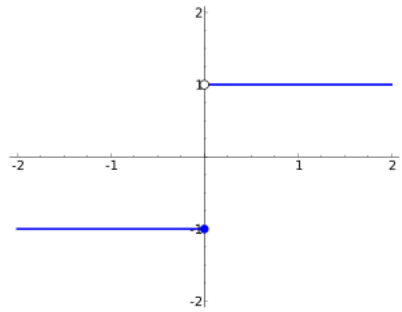
However, we say that this discontinuity is removable. This is because, if we modify the function at that point, we can eliminate the discontinuity and make the function continuous. To see how to make the function $f(x)$ continuous, we have to simplify $f(x)$, getting $f(x) = \dfrac{x^2-9}{x-3} = \dfrac{(x+3)(x-3)}{(x-3)} = \dfrac{x+3}{1} \cdot \dfrac{x-3}{x-3}$. We can define a new function $g(x))$ where $g(x) = x + 3$. Note that the function $g(x))$ is not the same as the original function $f(x)$, because $g(x)$ is defined at $x = 3$, while $f(x)$ is not. Thus, $g(x)$ is continuous at $x = 3$, since $\lim_{x\to 3}(x + 3) = 6 = g(3)$. However, whenever $x \neq 3$, $f(x) = g(x)$; all we did to f to get g was to make it defined at $x = 3$.

In fact, this kind of simplification is often possible with a discontinuity in a rational function. We can divide the numerator and the denominator by a common factor (in our example $x - 3$ to get a function which is the same except where that common factor was 0 (in our example at $x = 3$. This new function will be identical to the old except for being defined at new points where previously we had division by 0.

However, this is not possible in every case. For example, the function $f(x) = \dfrac{x-3}{x^2-6x+9}$ has a common factor of $x - 3$ in both the numerator and denominator, but when you simplify you are left with $g(x) = \dfrac{1}{x-3}$, which is still not defined at $x = 3$. In this case the domain of $f(x)$ and $g(x)$ are the same, and they are equal everywhere they are defined, so they are in fact the same function. The reason that $g(x)$ differed from $f(x)$ in the first example was because we could take it to have a larger domain and not simply that the formulas defining $f(x)$ and $g(x)$ were different.

Jump discontinuities

Illustration of a jump discontinuity



Not all discontinuities can be removed from a function. Consider this function: ot all discontinuities can be removed from a function. Consider this function: $k(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x \leq 0 \end{cases}$ Since $\lim_{x \to 0} k(x)$ does not exist, there is no way to redefine k kat one point so that it will be continuous at 0. These sorts of discontinuities are called nonremovable discontinuities.

Note, however, that both one-sided limits exist; $\lim_{x \to 0^-} k(x) = -1$ and $\lim_{x \to 0^+} k(x) = 1$. The problem is that they are not equal, so the graph "jumps" from one side of 0 to the other. In such a case, we say the function has a jump discontinuity. (Note that a jump discontinuity is a kind of nonremovable discontinuity.)

3.5.2 One-Sided Continuity

Just as a function can have a one-sided limit, a function can be continuous from a particular side. For a function to be continuous at a point from a given side, we need the following three conditions:

1. the function is defined at the point,
2. the function has a limit from that side at that point and
3. the one-sided limit equals the value of the function at the point.

A function $f(x)$ is

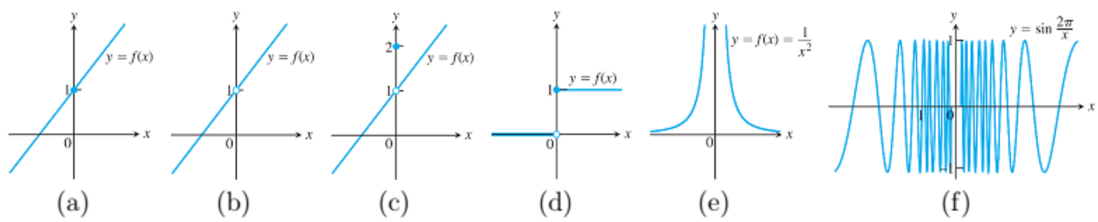
- *Left-continuous* at $x = c$ if $\lim_{x \to c^-} f(x) = f(c)$
- *Right-continuous* at $x = c$ if $\lim_{x \to c^+} f(x) = f(c)$.

A function will be continuous at a point if and only if it is continuous from both sides at that point. Now we can define what it means for a function to be continuous on a closed interval.

Definition 3.13. *(continuity on a closed interval) A function is said to be continuous on* $[a, b]$ *if and only if*

1. *it is continuous on* (a, b)*,*
2. *it is continuous from the right at* a *and*
3. *it is continuous from the left at* b*.*

Notice that, if a function is continuous, then it is continuous on every closed interval contained in its domain.



REMARK: The discontinuities in parts (b) and (c) are called removable discontinuities because we could remove them by redefining f at just the single number 0. The discontinuity in part (d) is called jump discontinuity because the function "jumps" from one value to another. The discontinuities in parts (e) and (f) are called infinite or essential discontinuities.

3.6 Intermediate value theorem

Theorem 3.6. *If a function* f *is continuous on a closed interval* $[a, b]$ *, then for every value* y *between* $f(a)$ *and* $f(b)$ *there is a value* $c \in (a, b)$ *such that* $f(c) = y$ *.*

If we think of a continuous function as a function whose graph has no hole or break, then it is easy to believe that the Intermediate Value Theorem is true. In geometric terms it says that if any horizontal line $y = u$ is given between $y = f(a)$ and $y = f(b)$ as in the above Figure, then the graph of can't jump over the line. It must intersect $y = u$ somewhere.
It is important that the function f in the above theorem be continuous. The Intermediate

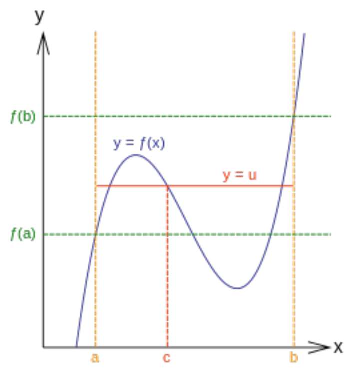


Figure 3.4: the graph of $y = f(x)$ and the line $y = u$.

Value theorem is not true in general for discontinuous functions.
One use of the Intermediate Value theorem is in locating roots of equations as in the following example.

Application: bisection method

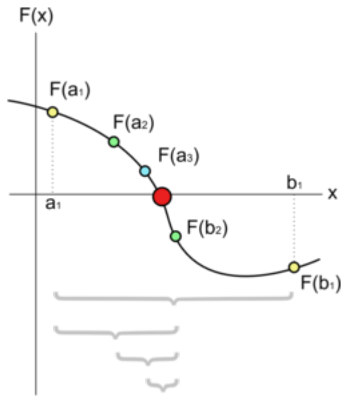
A few steps of the bisection method applied over the starting range $[a_1, b_1]$. The bigger red dot is the root of the function. The bisection method is the simplest and most reliable algorithm to find zeros of a continuous function.

Suppose we want to solve the equation $f(x) = 0$. Given two points a and b such that $f(a)$ and $f(b)$ have opposite signs, the intermediate value theorem tells us that f must have at least one root between a and b as long as f is continuous on the interval $[a, b]$. If we know f is continuous in general (say, because it's made out of rational, trigonometric, exponential and logarithmic functions), then this will work so long as f is defined at all points between a and b. So, let's divide the interval $[a, b]$ in two by computing $c = \dfrac{a+b}{2}$. There are now three possibilities:

1. $f(c) = 0$,
2. $f(a)$ and $f(c)$ have opposite signs, or
3. $f(c)$ and $f(b)$ have opposite signs.

In the first case, we're done. In the second and third cases, we can repeat the process on the sub-interval where the sign change occurs. In this way we hone in to a small sub-interval containing the 0. The midpoint of that small sub-interval is usually taken as a good approximation to the 0.

Note that, unlike the methods you may have learned in algebra, this works for any continuous function that you (or your calculator) know how to compute.



Example 3.17. *Show that there is a root of the equation*

$$4x^3 - 6x^2 + 3x - 2 = 0$$

between 1 and 2.

Solution

Let $f(x) = 4x^3 - 6x^2 + 3x - 2$ we are looking for a solution of the given equation, that is, a number c between 1 and 2 such that $f(c) = 0$. Therefore we take $a = 1,\ b = 2$ and $N = 0$ in the above theorem. We have
$f(1) = 4 - 6 + 3 - 2 = -1 < 0$
and $\quad f(2) = 32 - 24 + 6 - 2 = 12 > 0$
Thus $f(1) < 0 < f(2)$ that is, $N = 0$ is a number between $f(1)$ and $f(2)$. Now f is continuous since it is a polynomial, so the Intermediate Value Theorem says there is a number c between 1 and 2 such that $f(c)$. In other words, the equation $f(x) = 4x^3 - 6x^2 + 3x - 2$ has at least one root c in the interval $(1, 2)$.

3.7 Exercises and solutions

Basic Limit Exercises

1. Evaluate $\lim_{x \to 2}(4x^2 - 3x + 1)$
   Since this is a polynomial, two can simply be plugged in. This results in $4(4) - 2(3) + 1 = 16 - 6 +$

2. Evaluate $\lim_{x \to 5}(x^2)$ $5^2 = \mathbf{25}$

One-Sided Limits

Evaluate the following limits or state that the limit does not exist.

3. $\lim_{x \to 0^-} \frac{x^3 + x^2}{x^3 + 2x^2}$
Factor as $\frac{x^2}{x^2} \frac{x+1}{x+2}$. In this form we can see that there is a removable discontinuity at x=0 and that the limit is $\mathbf{\frac{1}{2}}$

4. $\lim_{x \to 7^-} |x^2 + x| - x$
$|7^2 + 7| - 7 = \mathbf{49}$

5. $\lim_{x \to -1^-} \sqrt{1 - x^2}$
$\sqrt{1 - x^2}$ is defined if $x^2 < 1$, so the limit is $\sqrt{1 - 1^2} = \mathbf{0}$

6. $\lim_{x \to -1^+} \sqrt{1 - x^2}$
$\sqrt{1 - x^2}$ is not defined if $x^2 > 1$, so the limit does not exist.

Two-Sided Limits

Evaluate the following limits or state that the limit does not exist.

7. $\lim_{x \to -1} \frac{1}{x - 1}$
$-\mathbf{\frac{1}{2}}$

8. $\lim_{x \to 4} \frac{1}{x - 4}$
$\lim_{x \to 4^-} \frac{1}{x - 4} = -\infty \lim_{x \to 4^+} \frac{1}{x - 4} = +\infty$ The limit does not exist.

9. $\lim_{x \to 2} \frac{1}{x - 2}$
$\lim_{x \to 2^-} \frac{1}{x - 2} = -\infty \lim_{x \to 2^+} \frac{1}{x - 2} = +\infty$.The limit does not exist.

10. $\lim_{x \to -3} \frac{x^2 - 9}{x + 3}$
$\lim_{x \to -3} \frac{(x + 3)(x - 3)}{x + 3} = \lim_{x \to -3} x - 3 = -3 - 3 = \mathbf{-6}$

11. $\lim_{x \to 3} \frac{x^2 - 9}{x - 3}$
$\lim_{x \to 3} \frac{(x - 3)(x + 3)}{x - 3} = \lim_{x \to 3} x + 3 = 3 + 3 = \mathbf{6}$

12. $\lim_{x \to -1} \frac{x^2 + 2x + 1}{x + 1}$
$\lim_{x \to -1} \frac{(x + 1)(x + 1)}{x + 1} = \lim_{x \to -1} x + 1 = -1 + 1 = \mathbf{0}$

13. $\lim_{x\to -1} \frac{x^3+1}{x+1}$
$\lim_{x\to -1} \frac{(x^2-x+1)(x+1)}{x+1} = \lim_{x\to -1} x^2-x+1 = (-1)^2-(-1)+1 = 1+1+1 = \mathbf{3}$

14. $\lim_{x\to 4} \frac{x^2+5x-36}{x^2-16}$
$\lim_{x\to 4} \frac{(x-4)(x+9)}{(x-4)(x+4)} = \lim_{x\to 4} \frac{x+9}{x+4} = \frac{4+9}{4+4} = \mathbf{\frac{13}{8}}$

15. $\lim_{x\to 25} \frac{x-25}{\sqrt{x}-5}$
$\lim_{x\to 25} \frac{(\sqrt{x}-5)(\sqrt{x}+5)}{\sqrt{x}-5} = \lim_{x\to 25} \sqrt{x}+5) = \sqrt{25}+5) = 5+5 = \mathbf{10}$

16. $\lim_{x\to 0} \frac{|x|}{x}$
$\lim_{x\to 0^-} \frac{|x|}{x} = \lim_{x\to 0^-} \frac{-x}{x} = \lim_{x\to 0^-} -1 = -1$ $\lim_{x\to 0^+} \frac{|x|}{x} = \lim_{x\to 0^+} \frac{x}{x} = \lim_{x\to 0^+} 1 = 1$. The limit does not exist.

17. $\lim_{x\to 2} \frac{1}{(x-2)^2}$
As x approaches 2, the denominator will be a very small positive number, so the whole fraction will be a very large positive number. Thus, the limit is ∞.

18. $\lim_{x\to 3} \frac{\sqrt{x^2+16}}{x-3}$
As x approaches 3, the numerator goes to 5 and the denominator goes to 0. Depending on whether you approach 3 from the left or the right, the denominator will be either a very small negative number, or a very small positive number. So the limit from the left is $-\infty$ and the limit from the right is $+\infty$. Thus, the limit does not exist.

19. $\lim_{x\to -2} \frac{3x^2-8x-3}{2x^2-18}$
$\frac{3(-2)^2-8(-2)-3}{2(-2)^2-18} = \frac{3(4)+16-3}{2(4)-18} = \frac{12+16-3}{8-18} = \frac{25}{-10} = -\mathbf{\frac{5}{2}}$

20. $\lim_{x\to 2} \frac{x^2+2x+1}{x^2-2x+1}$
$\frac{2^2+2(2)+1}{2^2-2(2)+1} = \frac{4+4+1}{4-4+1} = \frac{9}{1} = \mathbf{9}$

21. $\lim_{x\to 3} \frac{x+3}{x^2-9}$
$\lim_{x\to 3} \frac{x+3}{(x+3)(x-3)} = \lim_{x\to 3} \frac{1}{x-3}$ $\lim_{x\to 3^-} \frac{1}{x-3} = -\infty$ $\lim_{x\to 3^+} \frac{1}{x-3} = +\infty$. The limit does not exist.

22. $\lim\limits_{x\to -1} \dfrac{x+1}{x^2+x}$
    $\lim\limits_{x\to -1} \dfrac{x+1}{x(x+1)} = \lim\limits_{x\to -1} \dfrac{1}{x} = \dfrac{1}{-1} = \mathbf{-1}$

23. $\lim\limits_{x\to 1} \dfrac{1}{x^2+1}$
    $\dfrac{1}{1^2+1} = \dfrac{1}{1+1} = \dfrac{\mathbf{1}}{\mathbf{2}}$

24. $\lim\limits_{x\to 1} x^3 + 5x - \dfrac{1}{2-x}$
    $1^3 + 5(1) - \dfrac{1}{2-1} = 1+5-\dfrac{1}{1} = 6-1 = \mathbf{5}$

25. $\lim\limits_{x\to 1} \dfrac{x^2-1}{x^2+2x-3}$
    $\lim\limits_{x\to 1} \dfrac{(x-1)(x+1)}{(x-1)(x+3)} = \lim\limits_{x\to 1} \dfrac{x+1}{x+3} = \dfrac{1+1}{1+3} = \dfrac{2}{4} = \dfrac{\mathbf{1}}{\mathbf{2}}$

26. $\lim\limits_{x\to 1} \dfrac{5x}{x^2+2x-3}$
    Notice that as x approaches 1, the numerator approaches 5 while the denominator approaches 0. However, if you approach from below, the denominator is negative, and if you approach from above, the denominator is positive. So the limits from the left and right will be $-\infty$ and $+\infty$ respectively. Thus, the limit does not exist.

Limits to Infinity

Evaluate the following limits or state that the limit does not exist.

27. $\lim\limits_{x\to \infty} \dfrac{-x+\pi}{x^2+3x+2}$
    This rational function is bottom-heavy, so the limit is $\mathbf{0}$.

28. $\lim\limits_{x\to -\infty} \dfrac{x^2+2x+1}{3x^2+1}$
    This rational function has evenly matched powers of x in the numerator and denominator, so the limit will be the ratio of the coefficients, i.e. $\dfrac{\mathbf{1}}{\mathbf{3}}$.

29. $\lim\limits_{x\to -\infty} \dfrac{3x^2+x}{2x^2-15}$
    Balanced powers in the numerator and denominator, so the limit is the ratio of the coefficients, i.e. $\dfrac{\mathbf{3}}{\mathbf{2}}$.

30. $\lim\limits_{x\to -\infty} 3x^2 - 2x + 1$
    This is a top-heavy rational function, where the exponent of the ratio of the leading terms is 2. Since it is even, the limit will be ∞.

31. $\lim_{x\to\infty} \frac{2x^2 - 32}{x^3 - 64}$
Bottom-heavy rational function, so the limit is **0**.

32. $\lim_{x\to\infty} 6$
This is a rational function, as can be seen by writing it in the form $\frac{6x^0}{1x^0}$. Since the powers of x in the numerator and denominator are evenly matched, the limit will be the ratio of the coefficients, i.e. **6** .

33. $\lim_{x\to\infty} \frac{3x^2 + 4x}{x^4 + 2}$
Bottom-heavy, so the limit is **0**.

34. $\lim_{x\to-\infty} \frac{2x + 3x^2 + 1}{2x^2 + 3}$
Evenly matched highest powers of x in the numerator and denominator, so the limit will be the ratio of the corresponding coefficients, i.e. $\mathbf{\frac{3}{2}}$.

35. $\lim_{x\to-\infty} \frac{x^3 - 3x^2 + 1}{3x^2 + x + 5}$
Top-heavy rational function, where the exponent of the ratio of the leading terms is 1, so the limit is $-\infty$.

36. $\lim_{x\to\infty} \frac{x^2 + 2}{x^3 - 2}$
Bottom-heavy, so the limit is **0**.

Limits of Piecewise Functions

Evaluate the following limits or state that the limit does not exist.

37. Consider the function $f(x) = \begin{cases} (x-2)^2 & \text{if } x < 2 \\ x - 3 & \text{if } x \geq 2. \end{cases}$

 (a) $\lim_{x\to2^-} f(x)$
 $(2-2)^2 = \mathbf{0}$

 (b) $\lim_{x\to2^+} f(x)$
 $2 - 3 = -\mathbf{1}$

 (c) $\lim_{x\to2} f(x)$

Since the limits from the left and right don't match, the limit does not exist.

38. Consider the function

$$g(x) = \begin{cases} -2x+1 & \text{if } x \leq 0 \\ x+1 & \text{if } 0 < x < 4. \\ x^2+2 & \text{if } x \geq 4. \end{cases}$$

(a) $\lim_{x\to 4^+} g(x)$
$4^2 + 2 = 16 + 2 = \mathbf{18}$

(b) $\lim_{x\to 4^-} g(x)$ $4 + 1 = \mathbf{5}$

(c) $\lim_{x\to 0^+} g(x)$ $0 + 1 = \mathbf{1}$

(d) $\lim_{x\to 0^-} g(x)$
$-2(0) + 1 = \mathbf{1}$

(e) $\lim_{x\to 0} g(x)$ Since the left and right limits match, the overall limit is also **1** .

(f) $\lim_{x\to 1} g(x)$

39. Consider the function $h(x) = \begin{cases} 2x-3 & \text{if } x < 2 \\ 8 & \text{if } x = 2 \\ -x+3 & \text{if } x > 2. \end{cases}$

(a) $\lim_{x\to 0} h(x)$
$2(0) - 3 = -\mathbf{3}$

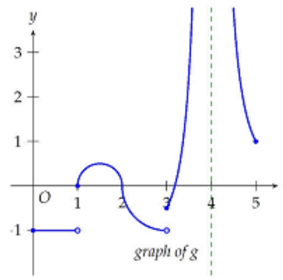
(b) $\lim_{x\to 2^-} h(x)$
$2(2) - 3 = 4 - 3 = \mathbf{1}$

(c) $\lim_{x\to 2^+} h(x)$
$-(2) + 3 = \mathbf{1}$

(d) $\lim_{x\to 2} h(x)$

Since the limits from the right and left match, the overall limit is **1** . Note that in this case, the limit at 2 does not match the function value at 2, so the function is discontinuous at this point, hence the function is nondifferentiable at this point as well.

40. The graph of a function g is shown.



(a) At which points a in $(0, 1, 2, 3, 4, 5)$ is g continuous?

(b) At which points a in $(0, 1, 2, 3, 4, 5)$ is g continuous from the right?

(c) At which points a in $(0, 1, 2, 3, 4, 5)$ is g continuous from the left?

CHAPTER FOUR

4 Derivatives

4.1 Introduction

What is Differentiation?

Differentiation is an operation that allows us to find a function that outputs the rate of change of one variable with respect to another variable.

Informally, we may suppose that we're tracking the position of a car on a two-lane road with no passing lanes. Assuming the car never pulls off the road, we can abstractly study the car's position by assigning it a variable, x . Since the car's position changes as the time changes, we say that x is dependent on time, or $x = f(t)$. This tells where the car is at each specific time. Differentiation gives us a function $\frac{dx}{dt}$ which represents the car's speed, that is the rate of change of its position with respect to time.

Equivalently, differentiation gives us the slope at any point of the graph of a non-linear function. For a linear function, of form $f(x) = ax + b$, a is the slope. For non-linear functions, such as $f(x) = 3x^2$, the slope can depend on x ; differentiation gives us a function which represents this slope.

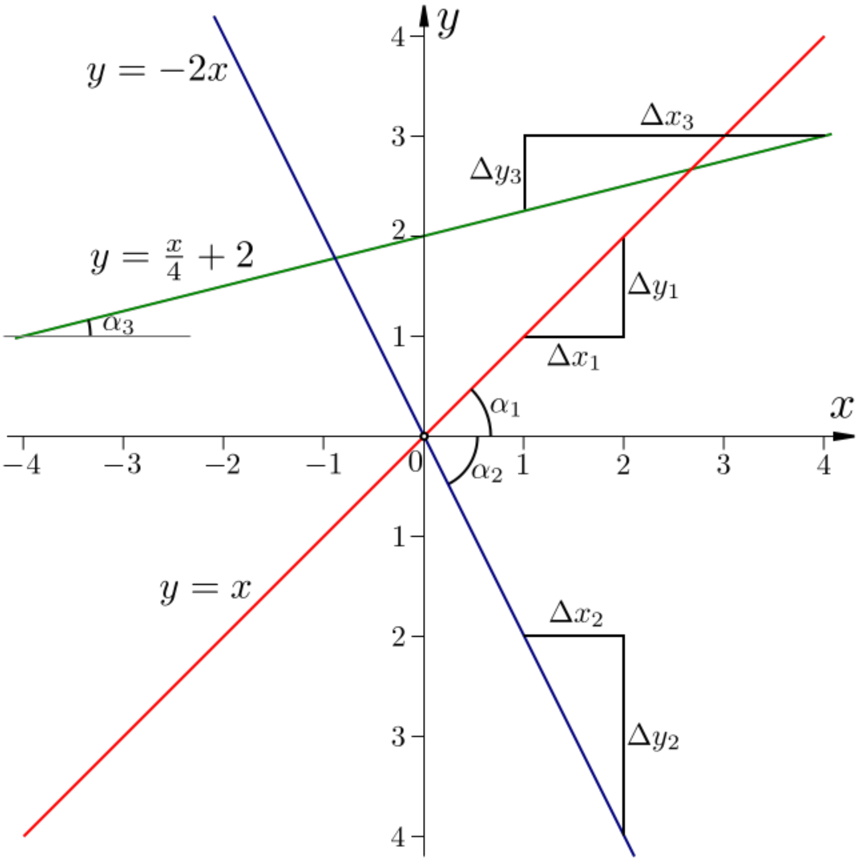
The Definition of Slope

Historically, the primary motivation for the study of differentiation was the tangent line problem: for a given curve, find the slope of the straight line that is tangent to the curve at

a given point. The word tangent comes from the Latin word tangens, which means touching. Thus, to solve the tangent line problem, we need to find the slope of a line that is "touching" a given curve at a given point, or, in modern language, that has the same slope. But what exactly do we mean by "slope" for a curve?
The solution is obvious in some cases: for example, a line $y = mx + c$ is its own tangent; the slope at any point is m. For the parabola $y = x^2$, the slope at the point $(0, 0)$ is 0 ; the tangent line is horizontal.

But how can you find the slope of, say, $y = \sin(x) + x^2$ at $x = 1.5$? This is in general a nontrivial question, but first we will deal carefully with the slope of lines.

Of a line



The slope of a line, also called the gradient of the line, is a measure of its inclination. A line that is horizontal has slope 0, a line from the bottom left to the top right has a positive slope and a line from the top left to the bottom right has a negative slope.

The slope can be defined in two (equivalent) ways. The first way is to express it as how much the line climbs for a given motion horizontally. We denote a change in a quantity using the symbol Δ (pronounced "delta"). Thus, a change in x is written as Δx . We can therefore write this definition of slope as:

$$\text{Slope} = \frac{\Delta y}{\Delta x}.$$

An example may make this definition clearer. If we have two points on a line, $P(x_1, y_1)$ and $Q(x_2, y_2)$, the change in x from P to Q is given by:

$$\Delta x = x_2 - x_1$$

Likewise, the change in y from P to Q is given by:

$$\Delta y = y_2 - y_1$$

This leads to the very important result below.

The slope of the line between the points (x_1, y_1) and (x_2, y_2) is

$$\frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

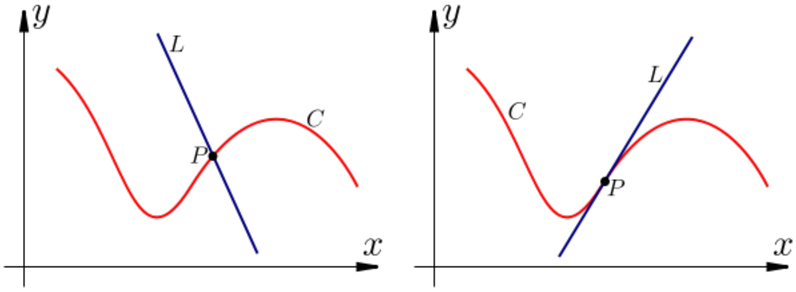
Alternatively, we can define slope trigonometrically , using the tangent function:

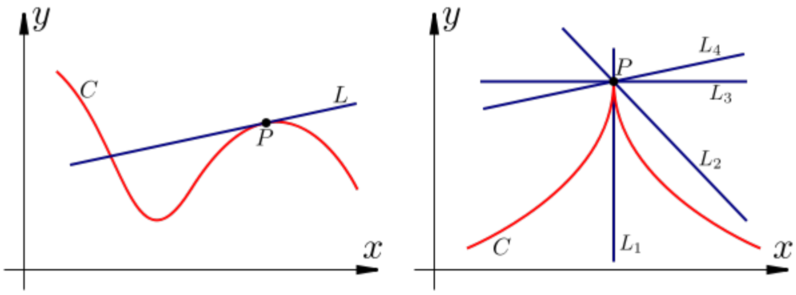
$$\text{Slope} = \tan(\alpha)$$

where α is the angle from the rightward-pointing horizontal to the line, measured counter-clockwise. If you recall that the tangent of an angle is the ratio of the y-coordinate to the x-coordinate on the unit circle, you should be able to spot the equivalence here.

Of a graph of a function.

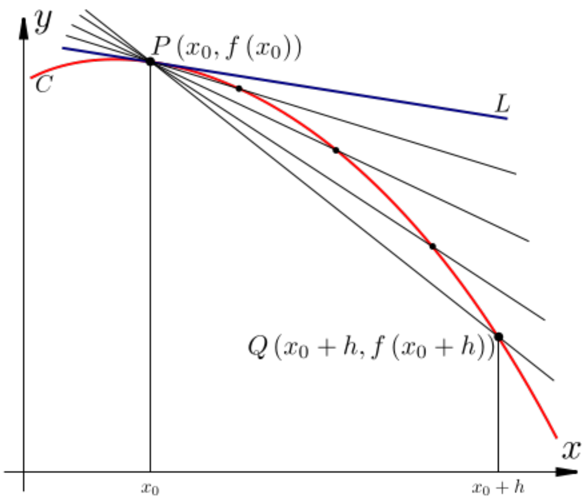
The graphs of most functions we are interested in are not straight lines (although they can be), but rather curves. We cannot define the slope of a curve in the same way as we can for a line. In order for us to understand how to find the slope of a curve at a point, we will first have to cover the idea of tangency. Intuitively, a tangent is a line which just touches a curve at a point, such that the angle between them at that point is 0. Consider the following four curves and lines:





The line L crosses, but is not tangent to C at P . The line L crosses, and is tangent to C at P. The line L crosses C at two points, but is tangent to C only at P. There are many lines that cross C at P, but none are tangent. In fact, this curve has no tangent at P.

A secant is a line drawn through two points on a curve. We can construct a definition of a tangent as the limit of a secant of the curve taken as the separation between the points tends to zero. Consider the diagram below.



As the distance h tends to 0, the secant line becomes the tangent at the point x_0. The two points we draw our line through are:

$$P(x_0, f(x_0))$$

and

$$Q(x_0 + h, f(x_0 + h))$$

As a secant line is simply a line and we know two points on it, we can find its slope, m_h, using the formula from before:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

(We will refer to the slope as m_h because it may, and generally will, depend on h .) Substituting in the points on the line,

$$m_h = \frac{f(x_0 + h) - f(x_0)}{(x_0 + h) - x_0}$$

This simplifies to

$$m_h = \frac{f(x_0 + h) - f(x_0)}{h}$$

This expression is called the difference quotient. Note that h can be positive or negative it is perfectly valid to take a secant through any two points on the curve but cannot be 0.

The definition of the tangent line we gave was not rigorous, since we've only defined limits of numbers or, more precisely, of functions that output numbers not of lines. But we can define the slope of the tangent line at a point rigorously, by taking the limit of the slopes of the secant lines from the last paragraph. Having done so, we can then define the tangent line as well. Note that we cannot simply set h to 0 as this would imply division of 0 by 0 which would yield an undefined result. Instead we must find the limit of the above expression as h tends to 0:

Definition 4.1. *(Slope of the graph of a function) The slope of the graph of $f(x)$ at the point $(x_0, f(x_0))$ is*

$$\lim_{h \to 0} \left[\frac{f(x_0 + h) - f(x_0)}{h} \right]$$

If this limit does not exist, then we say the slope is undefined. If the slope is defined, say m, then the tangent line to the graph of $f(x)$ at the point $(x_0, f(x_0))$ is the line with equation

$$y - f(x_0) = m \cdot (x - x_0)$$

This last equation is just the point-slope form for the line through $(x_0, f(x_0))$ with slope m.

Exercises

1. Find the slope of the tangent to the curve $y = x^2$ $(1, 1)$.

The Rate of Change of a Function at a Point

Consider the formula for average velocity in the x direction, $\frac{\Delta x}{\Delta t}$, where Δx is the change in x over the time interval Δt. This formula gives the average velocity over a period of time, but suppose we want to define the instantaneous velocity. To this end we look at the change in position as the change in time approaches 0. Mathematically this is written as: $\lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$, which we abbreviate by the symbol $\frac{dx}{dt}$. (The idea of this notation is that the letter d

denotes change.) Compare the symbol d with Δ. The idea is that both indicate a difference between two numbers, but Δ denotes a finite difference while d denotes an infinitesimal difference. Please note that the symbols dx and dt have no rigorous meaning on their own, since $\lim_{\Delta t \to 0} \Delta t = 0$, and we can't divide by 0.

(Note that the letter s is often used to denote distance, which would yield $\frac{ds}{dt}$. The letter d is often avoided in denoting distance due to the potential confusion resulting from the expression $\frac{dd}{dt}$.)

4.2 The Definition of the Derivative

You may have noticed that the two operations we've discussed -computing the slope of the tangent to the graph of a function and computing the instantaneous rate of change of the function -involved exactly the same limit. That is, the slope of the tangent to the graph of $y = f(x)$ is $\frac{dy}{dx}$. Of course, $\frac{dy}{dx}$ can, and generally will, depend on x , so we should really think of it as a function of x. We call this process (of computing $\frac{dy}{dx}$) is differentiation. Differentiation results in another function whose value for any value x is the slope of the original function at x. This function is known as the derivative of the original function.

Since lots of different sorts of people use derivatives, there are lots of different mathematical notations for them. Here are some:

$f'(x)$ (read "f prime of x") for the derivative of $f(x)$, $D_x[f(x)]$, $Df(x)$, $\frac{dy}{dx}$ for the derivative of y as a function of x or $\frac{d}{dx}\left[y\right]$, which is more useful in some cases.

Most of the time the brackets are not needed, but are useful for clarity if we are dealing with something like $D(f \cdot g)$, where we want to differentiate the product of two functions, f and g .

The first notation has the advantage that it makes clear that the derivative is a function. That is, if we want to talk about the derivative of $f(x)$ at $x = 2$, we can just write $f'(2)$.

In any event, here is the formal definition:

Definition 4.2. *(derivative) Let $f(x)$ be a function. Then* $f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$ *wherever this limit exists. In this case we say that f is differentiable at x and its derivative at x is $f'(x)$.*

Example 4.1. *The derivative of* $f(x) = \frac{x}{2}$ *is*

$$f'(x) = \lim_{\Delta x \to 0} \left(\frac{\frac{x+\Delta x}{2} - \frac{x}{2}}{\Delta x} \right) = \lim_{\Delta x \to 0} \left(\frac{\frac{x}{2} + \frac{\Delta x}{2} - \frac{x}{2}}{\Delta x} \right) = \lim_{\Delta x \to 0} \left(\frac{\frac{\Delta x}{2}}{\Delta x} \right) = \lim_{\Delta x \to 0} \left(\frac{\Delta x}{2\Delta x} \right) = \lim_{\Delta x \to 0} \left(\frac{1}{2} \right) = \frac{1}{2}$$

no matter what x is. This is consistent with the definition of the derivative as the slope of a function.

Example 4.2. *What is the slope of the graph of* $y = 3x^2$ *at* $(4, 48)$*? We can do it "the hard (and imprecise) way", without using differentiation, as follows, using a calculator and using small differences below and above the given point:*

When $x = 3.999$, $y = 47.976003$.
When $x = 4.001$, $y = 48.024003$.
Then the difference between the two values of x is $\Delta x = 0.002$.
Then the difference between the two values of y is $\Delta y = 0.048$.
Thus, the slope $\frac{\Delta y}{\Delta x} = 24$ at the point of the graph at which $x = 4$.
But, to solve the problem precisely, we compute

$$\begin{aligned}\lim_{\Delta x \to 0} \frac{3(4+\Delta x)^2 - 48}{\Delta x} &= 3 \lim_{\Delta x \to 0} \frac{(4+\Delta x)^2 - 16}{\Delta x} \\ &= 3 \lim_{\Delta x \to 0} \frac{16 + 8\Delta x + (\Delta x)^2 - 16}{\Delta x} \\ &= 3 \lim_{\Delta x \to 0} \frac{8\Delta x + (\Delta x)^2}{\Delta x} \\ &= 3 \lim_{\Delta x \to 0} (8 + \Delta x) \\ &= 3(8) = 24\end{aligned}$$

We were lucky this time; the approximation we got above turned out to be exactly right. But this won't always be so, and, anyway, this way we didn't need a calculator.

In general, the derivative of $f(x) = 3x^2$ is

$$\begin{aligned}f'(x) &= \lim_{\Delta x \to 0} \frac{3(x+\Delta x)^2 - 3x^2}{\Delta x} \\ &= 3 \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x} \\ &= 3 \lim_{\Delta x \to 0} \frac{x^2 + 2x\Delta x + (\Delta x)^2 - x^2}{\Delta x} \\ &= 3 \lim_{\Delta x \to 0} \frac{2x\Delta x + (\Delta x)^2}{\Delta x} \\ &= 3 \lim_{\Delta x \to 0} (2x + \Delta x) \\ &= 3(2x) \\ &= 6x\end{aligned}$$

Example 4.3. $f(x) = |x|$ *(the absolute value function) then* $f'(x) = \frac{x}{|x|}$*, which can also be stated as*

$$f'(x) = \begin{cases} -1 & x < 0 \\ \text{undefined} & x = 0 \\ 1 & x > 0 \end{cases}$$

Finding this derivative is a bit complicated, so we won't prove it at this point. Here, $f(x)$ is not smooth (though it is continuous) at $x = 0$ and so the limits $\lim_{x\to 0^+} f'(x)$ and $\lim_{x\to 0^-} f'(x)$ (the limits as 0 is approached from the right and left respectively) are not equal. From the definition,$f'(0) = \lim_{\Delta x\to 0} \frac{|\Delta x|}{\Delta x}$, which does not exist. Thus, $f'(0)$ is undefined, and so $f'(x)$ has a discontinuity at 0. This sort of point of non-differentiability is called a cusp. Functions may also not be differentiable because they go to infinity at a point, or oscillate infinitely frequently.

The derivative notation is special and unique in mathematics. The most common notation for derivatives you'll run into when first starting out with differentiating is the Leibniz notation, expressed as $\frac{dy}{dx}$. You may think of this as "rate of change in y with respect to x. You may also think of it as "infinitesimal value of y divided by infinitesimal value of x". Either way is a good way of thinking, although you should remember that the precise definition is the one we gave above. Often, in an equation, you will see just $\frac{d}{dx}$, which literally means "derivative with respect to x". This means we should take the derivative of whatever is written to the right; that is, $\frac{d}{dx}(x+2)$ means $\frac{dy}{dx}$ where $y = x + 2$.

As you advance through your studies, you will see that we sometimes pretend that dy and dx are separate entities that can be multiplied and divided, by writing things like $dy = x^4 dx$. Eventually you will see derivatives such as $\frac{dx}{dy}$, which just means that the input variable of our function is called y and our output variable is called x; sometimes, we will write $\frac{d}{dy}$, to mean the derivative with respect to y of whatever is written on the right. In general, the variables could be anything, say $\frac{d\theta}{dr}$.

All of the following are equivalent for expressing the derivative of $y = x^2$

$$\frac{dy}{dx} = 2x\frac{d}{dx}x^2 = 2x$$

$$dy = 2x\,dx$$

$$f'(x) = 2x$$

$$D(f(x)) = 2x$$

Exercises

1. Using the definition of the derivative find the derivative of the function $f(x) = 2x + 3$.
2. Using the definition of the derivative find the derivative of the function $f(x) = x^3$. Now try $f(x) = x^4$. Can you see a pattern? In the next section we will find the derivative of $f(x) = x^n$ for all n.
3. The text states that the derivative of $|x|$ is not defined at $x = 0$. Use the definition of the derivative to show this.

4. Graph the derivative to $y = 4x^2$ on a piece of graph paper without solving for $\frac{dy}{dx}$. Then, solve for $\frac{dy}{dx}$ and graph that; compare the two graphs.

5. Use the definition of the derivative to show that the derivative of $\sin(x)$ is $\cos(x)$. Hint: Use a suitable sum to product formula and the fact that $\lim_{t\to 0} \frac{\sin(t)}{t} = 1$ and $\lim_{t\to 0} \frac{\cos(t) - 1}{t} = 0$.

Theorem 4.1. *If f is differentiable at a, then f is continuous at a.*

4.3 Differentiation Rules

The process of differentiation is tedious for complicated functions. Therefore, rules for differentiating general functions have been developed, and can be proved with a little effort. Once sufficient rules have been proved, it will be fairly easy to differentiate a wide variety of functions. Some of the simplest rules involve the derivative of linear functions.

4.3.1 Derivative of a constant function

For any fixed real number c,

$$\frac{d}{dx}[c] = 0$$

Intuition

The graph of the function $f(x) = c$ is a horizontal line, which has a constant slope of 0. Therefore, it should be expected that the derivative of this function is zero, regardless of the values of x and c.

Proof

The definition of a derivative is

$$\lim_{\Delta x\to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Let $f(x) = c$ for all x. (That is, f is a constant function.) Then $f(x + \Delta x) = c$. Therefore

$$\frac{d}{dx}[c] = \lim_{\Delta x\to 0} \frac{c - c}{\Delta x} = \lim_{\Delta x\to 0} \frac{0}{\Delta x}$$

Let $g(\Delta x) = \frac{0}{\Delta x}$. To prove that $\lim_{\Delta x\to 0} g(\Delta x) = 0$, we need to find a positive δ such that, for any given positive ε, $\left|g(\Delta x) - 0\right| < \varepsilon$ whenever $0 < |\Delta x - 0| < \delta$. But $\left|g(\Delta x) - 0\right| = 0$, so $\left|g(\Delta x) - 0\right| < \varepsilon$ for any choice of δ.

$$\frac{d}{dx}[3] = 0$$
$$\frac{d}{dx}[z] = 0$$

Note that, in the second example, z is just a constant.

4.3.2 Derivative of a linear function

For any fixed real numbers m and c, $\frac{d}{dx}[mx + c] = m$

The special case $\frac{dx}{dx} = 1$ shows the advantage of the $\frac{d}{dx}$ notation-rules are intuitive by basic algebra, though this does not constitute a proof, and can lead to misconceptions to what exactly dx and dy actually are.

Intuition

The graph of $y = mx + c$ is a line with constant slope m.

Proof

If $f(x) = mx + c$, then $f(x + \Delta x) = m(x + \Delta x) + c$. So,

$$\begin{aligned} f'(x) &= \lim_{\Delta x \to 0} \frac{m(x + \Delta x) + c - mx - c}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{m(x + \Delta x) - mx}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{mx + m\Delta x - mx}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{m\Delta x}{\Delta x} \\ &= m \end{aligned}$$

4.3.3 Constant multiple and addition rules

Since we already know the rules for some very basic functions, we would like to be able to take the derivative of more complex functions by breaking them up into simpler functions. Two tools that let us do this are the constant multiple rule and the addition rule.

The Constant Rule

For any fixed real number c , $\frac{d}{dx}\Big[c \cdot f(x)\Big] = c \cdot \frac{d}{dx}\Big[f(x)\Big]$. The reason, of course, is that one can factor c out of the numerator, and then of the entire limit, in the definition. The details are left as an exercise.

Example 4.4. *We already know that* $\frac{d}{dx}\left[x^2\right] = 2x$

Suppose we want to find the derivative of $3x^2$

$$\begin{aligned}\frac{d}{dx}\left[3x^2\right] &= 3\frac{d}{dx}\left[x^2\right]\\ &= 3 \cdot 2x\\ &= 6x\end{aligned}$$

Another simple rule for breaking up functions is the addition rule.

The Addition and Subtraction Rules

$$\frac{d}{dx}\left[f(x) \pm g(x)\right] = \frac{d}{dx}\left[f(x)\right] \pm \frac{d}{dx}\left[g(x)\right]$$

Proof

From the definition:

$$\begin{aligned}\lim_{\Delta x \to 0}\left[\frac{\left[f(x+\Delta x) \pm g(x+\Delta x)\right] - \left[f(x) \pm g(x)\right]}{\Delta x}\right] &= \lim_{\Delta x \to 0}\left[\frac{\left[f(x+\Delta x) - f(x)\right] \pm \left[g(x+\Delta x) - g(x)\right]}{\Delta x}\right.\\ &= \lim_{\Delta x \to 0}\left[\frac{f(x+\Delta x) - f(x)}{\Delta x}\right] \pm \lim_{\Delta x \to 0}\left[\frac{g(x+\Delta x) - g(}{\Delta x}\right.\end{aligned}$$

By definition then, this last term is $\frac{d}{dx}[f(x)] \pm \frac{d}{dx}[g(x)]$

Example 4.5. *What is the derivative of* $3x^2 + 5x$*?*

$$\begin{aligned}\frac{d}{dx}\left[3x^2 + 5x\right] &= \frac{d}{dx}\left[3x^2 + 5x\right]\\ &= \frac{d}{dx}\left[3x^2\right] + \frac{d}{dx}\left[5x\right]\\ &= 6x + \frac{d}{dx}\left[5x\right]\\ &= 6x + 5\end{aligned}$$

The fact that both of these rules work is extremely significant mathematically because it means that differentiation is linear. You can take an equation, break it up into terms, figure out the derivative individually and build the answer back up, and nothing odd will happen.

We now need only one more piece of information before we can take the derivatives of any polynomial.

4.3.4 The Power Rule

$$\frac{d}{dx}\Big[x^n\Big] = nx^{n-1}$$

This has been proved in an example in Derivatives of Exponential and Logarithm Functions where it can be best understood. For example, in the case of x^2 the derivative is $2x^1 = 2x$ as was established earlier. A special case of this rule is that $\frac{dx}{dx} = \frac{dx^1}{dx} = 1x^0 = 1$.

Since polynomials are sums of monomials, using this rule and the addition rule lets you differentiate any polynomial. A relatively simple proof for this can be derived from the binomial expansion theorem.

This rule also applies to fractional and negative powers. Therefore

$$\begin{aligned}\frac{d}{dx}\Big[\sqrt{x}\Big] &= \frac{d}{dx}\Big[x^{\frac{1}{2}}\Big] \\ &= \frac{1}{2}x^{-\frac{1}{2}} \\ &= \frac{1}{2\sqrt{x}}\end{aligned}$$

4.3.5 Derivatives of polynomials

With these rules in hand, you can now find the derivative of any polynomial you come across. Rather than write the general formula, let's go step by step through the process.

$$\frac{d}{dx}\Big[6x^5 + 3x^2 + 3x + 1\Big]$$

The first thing we can do is to use the addition rule to split the equation up into terms:

$$\frac{d}{dx}\Big[6x^5\Big] + \frac{d}{dx}\Big[3x^2\Big] + \frac{d}{dx}\Big[3x\Big] + \frac{d}{dx}[1]$$

We can immediately use the linear and constant rules to get rid of some terms:

$$\frac{d}{dx}\Big[6x^5\Big] + \frac{d}{dx}\Big[3x^2\Big] + 3 + 0$$

Now you may use the constant multiplier rule to move the constants outside the derivatives:

$$6\frac{d}{dx}\Big[x^5\Big] + 3\frac{d}{dx}\Big[x^2\Big] + 3$$

Then use the power rule to work with the individual monomials:

$$6(5x^4) + 3(2x) + 3$$

And then do some algebra to get the final answer:

$$30x^4 + 6x + 3$$

These are not the only differentiation rules. There are other, more advanced, differentiation rules, which will be described in a later sections.

Exercises Find the derivatives of the following equations:

1. $f(x) = 42$
2. $f(x) = 6x + 10$
3. $f(x) = 2x^2 + 12x + 3$

4.3.6 Product Rule

When we wish to differentiate a more complicated expression such as

$$h(x) = (x^2 + 5x + 7)(x^3 + 2x - 4)$$

our only way (up to this point) to differentiate the expression is to expand it and get a polynomial, and then differentiate that polynomial. This method becomes very complicated and is particularly error prone when doing calculations by hand. A beginner might guess that the derivative of a product is the product of the derivatives, similar to the sum and difference rules, but this is not true. To take the derivative of a product, we use the product rule. Derivatives of products (Product Rule)

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

It may also be stated as

$$(f \cdot g)' = f' \cdot g + f \cdot g'$$

or in the Leibniz notation as

$$\frac{d}{dx}(u \cdot v) = u \cdot \frac{dv}{dx} + v \cdot \frac{du}{dx}$$

The derivative of the product of three functions is:

$$\frac{d}{dx}(u \cdot v \cdot w) = \frac{du}{dx} \cdot v \cdot w + u \cdot \frac{dv}{dx} \cdot w + u \cdot v \cdot \frac{dw}{dx}.$$

Since the product of two or more functions occurs in many mathematical models of physical phenomena, the product rule has broad application in physics, chemistry, and engineering.

Example 4.6. *Suppose one wants to differentiate* $f(x) = x^2 \sin(x)$.

By using the product rule, one gets the derivative

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

One special case of the product rule is the constant multiple rule, which states: if c is a real number and $f(x)$ is a differentiable function, then $c \cdot f(x)$ is also differentiable, and its derivative is $(c \cdot f)'(x) = c \cdot f'(x)$. This follows from the product rule since the derivative of any constant is 0. This, combined with the sum rule for derivatives, shows that differentiation is linear.

Physics Example I: rocket acceleration The acceleration of model rockets can be computed with the product rule.

Consider the vertical acceleration of a model rocket relative to its initial position at a fixed point on the ground. Newton's second law says that the force is equal to the time rate change of momentum. If $\vec{F}$ is the net force (sum of forces), $\vec{p}$ is the momentum, and t is the time,

$$\vec{F} = \frac{d\vec{p}}{dt}$$

Since the momentum is equal to the product of mass and velocity, this yields

$$\vec{F} = \frac{d}{dt}(m\vec{v})$$

where m is the mass and v is the velocity. Application of the product rule gives

$$\vec{F} = \vec{v}\frac{dm}{dt} + m\frac{d\vec{v}}{dt}$$

Since the acceleration, $\vec{a}$, is defined as the time rate change of velocity,

$$\vec{a} = \frac{d\vec{v}}{dt}$$

$$\vec{F} = \vec{v}\frac{dm}{dt} + m\vec{a}$$

Solving for the acceleration,

$$\vec{a} = \frac{\vec{F} - \vec{v}\frac{dm}{dt}}{m}$$

Since the rocket is losing mass, $\frac{dm}{dt}$ is negative, and the changing mass term results in increased acceleration.

Note: Here $\vec{F}$ is considered to be the net force.

Physics Example II: electromagnetic induction

Faraday's law of electromagnetic induction states that the induced electromotive force is the negative time rate of change of magnetic flux through a conducting loop.

$$\mathcal{E} = -\frac{d\Phi_B}{dt}$$

where $\mathcal{E}$ is the electromotive force (emf) in volts and $\phi B$is the magnetic flux in webers. For a loop of area, A, in a magnetic field, B, the magnetic flux is given by

$$\Phi_B = B \cdot A \cdot \cos(\theta)$$

where θ is the angle between the normal to the current loop and the magnetic field direction.

Taking the negative derivative of the flux with respect to time yields the electromotive force gives

$$\begin{aligned} \mathcal{E} &= -\frac{d}{dt}\Big(B \cdot A \cdot \cos(\theta)\Big) \\ &= -\frac{dB}{dt} \cdot A\cos(\theta) - B \cdot \frac{dA}{dt}\cos(\theta) - B \cdot A\frac{d}{dt}\cos(\theta) \end{aligned}$$

In many cases of practical interest only one variable (A, B, or θ) is changing, so two of the three above terms are often 0.

Physics Example III: Kinematics

The position of a particle on a number line relative to a fixed point O is $4t^3\sin(t)\sec^2(t)$, where t represents the time. What is its instantaneous velocity at $t = 7$ relative to O? Distances are in meters and time in seconds.

Answer:

Note: To solve this problem, we need some 'tools' from the next section.

We can simplify the function to

$$4t^3\tan(t)\sec(t)$$
$$\sin(t)\sec(t) = \tan(t)$$

$$v(t) = \frac{d}{dt}\Big[4t^3\tan(t)\sec(t)\Big] = \tan(t)\sec(t)\cdot\frac{d}{dt}[4t^3] + 4t^3\sec(t)\cdot\frac{d}{dt}[\tan(t)] + 4t^3\tan(t)\cdot\frac{d}{dt}[\sec(t)]$$
$$= 12t^2\tan(t)\sec(t) + 4t^3\sec^3(t) + 4t^3\tan^2(t)\sec(t)$$

Substituting $t = 7$ into our velocity function:

$v(7) = 12(7)^2\tan(7)\sec(7) + 4(7)^3\sec^3(7) + 4(7)^3\tan^2(7)\sec(7) = 1496.72\,\frac{m}{s}$ (to 2 decimal places).

Proof of the Product Rule

Proving this rule is relatively straightforward, first let us state the equation for the derivative:

$$\frac{d}{dx}\Big[f(x)\cdot g(x)\Big] = \lim_{h\to 0}\frac{f(x+h)\cdot g(x+h) - f(x)\cdot g(x)}{h}$$

We will then apply one of the oldest tricks in the book-adding a term that cancels itself out to the middle:

$$\frac{d}{dx}\Big[f(x)\cdot g(x)\Big] = \lim_{h\to 0}\frac{f(x+h)\cdot g(x+h) - f(x+h)\cdot g(x) + f(x+h)\cdot g(x) - f(x)\cdot g(x)}{h}$$

Notice that those terms sum to 0, and so all we have done is add 0 to the equation. Now we can split the equation up into forms that we already know how to solve:

$$\frac{d}{dx}\Big[f(x)\cdot g(x)\Big] = \lim_{h\to 0}\left[\frac{f(x+h)\cdot g(x+h) - f(x+h)\cdot g(x)}{h} + \frac{f(x+h)\cdot g(x) - f(x)\cdot g(x)}{h}\right]$$

Looking at this, we see that we can factor the common terms out of the numerators to get:

$$\frac{d}{dx}\Big[f(x)\cdot g(x)\Big] = \lim_{h\to 0}\left[f(x+h)\cdot\frac{g(x+h)-g(x)}{h} + g(x)\cdot\frac{f(x+h)-f(x)}{h}\right]$$
$$= \lim_{h\to 0}\left[f(x+h)\cdot\frac{g(x+h)-g(x)}{h}\right] + \lim_{h\to 0}\left[g(x)\cdot\frac{f(x+h)-f(x)}{h}\right]$$
$$= \lim_{h\to 0} f(x+h)\cdot\lim_{h\to 0}\frac{g(x+h)-g(x)}{h} + \lim_{h\to 0} g(x)\cdot\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$$

Which, when we take the limit, becomes:

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

, or the mnemonic "one D-two plus two D-one"

This can be extended to 3 functions:

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

For any number of functions, the derivative of their product is the sum, for each function, of its derivative times each other function.

Back to our original example of a product, $h(x) = (x^2 + 5x + 7)(x^3 + 2x - 4)$, we find the derivative by the product rule is

$$h'(x) = (x^2 + 5x + 7)(3x^2 + 2) + (2x + 5)(x^3 + 2x - 4) = 5x^4 + 20x^3 + 27x^2 + 12x - 6$$

Note, its derivative would not be

$$(2x + 5)(3x^2 + 2) = 6x^3 + 15x^2 + 4x + 10$$

which is what you would get if you assumed the derivative of a product is the product of the derivatives.

To apply the product rule we multiply the first function by the derivative of the second and add to that the derivative of first function multiply by the second function. Sometimes it helps to remember the phrase "First times the derivative of the second plus the second times the derivative of the first."

4.3.7 Quotient Rule

There is a similar rule for quotients. To prove it, we go to the definition of the derivative:

$$\begin{aligned}
\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] &= \lim_{h\to 0}\frac{\dfrac{f(x+h)}{g(x+h)}-\dfrac{f(x)}{g(x)}}{h} \\
&= \lim_{h\to 0}\frac{f(x+h)\cdot g(x)-f(x)\cdot g(x+h)}{h\cdot g(x)\cdot g(x+h)} \\
&= \lim_{h\to 0}\frac{f(x+h)\cdot g(x)-f(x)\cdot g(x)+f(x)\cdot g(x)-f(x)\cdot g(x+h)}{h\cdot g(x)\cdot g(x+h)} \\
&= \lim_{h\to 0}\frac{g(x)\cdot\dfrac{f(x+h)-f(x)}{h}-f(x)\cdot\dfrac{g(x+h)-g(x)}{h}}{g(x)\cdot g(x+h)} \\
&= \frac{\lim\limits_{h\to 0}\left[g(x)\cdot\dfrac{f(x+h)-f(x)}{h}-f(x)\cdot\dfrac{g(x+h)-g(x)}{h}\right]}{\lim\limits_{h\to 0}\left[g(x)\cdot g(x+h)\right]} \\
&= \frac{\lim\limits_{h\to 0}\left[g(x)\cdot\dfrac{f(x+h)-f(x)}{h}\right]-\lim\limits_{h\to 0}\left[f(x)\cdot\dfrac{g(x+h)-g(x)}{h}\right]}{\lim\limits_{h\to 0}\left[g(x)\cdot g(x+h)\right]} \\
&= \frac{\lim\limits_{h\to 0}g(x)\cdot\lim\limits_{h\to 0}\dfrac{f(x+h)-f(x)}{h}-\lim\limits_{h\to 0}f(x)\cdot\lim\limits_{h\to 0}\dfrac{g(x+h)-g(x)}{h}}{\lim\limits_{h\to 0}g(x)\cdot\lim\limits_{h\to 0}g(x+h)} \\
&= \frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}
\end{aligned}$$

This leads us to the so-called "quotient rule":

Derivatives of quotients (Quotient Rule)

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

Some people remember this rule with the mnemonic "low D-high minus high D-low, square the bottom and away we go!"

Example 4.7. *The derivative of* $\dfrac{4x-2}{x^2+1}$ *is:*

$$\begin{aligned}
\frac{d}{dx}\left[\frac{4x-2}{x^2+1}\right] &= \frac{(4)(x^2+1)-(2x)(4x-2)}{(x^2+1)^2} \\
&= \frac{(4x^2+4)-(8x^2-4x)}{(x^2+1)^2} \\
&= \frac{-4x^2+4x+4}{(x^2+1)^2}
\end{aligned}$$

Remember: the derivative of a product/quotient is not the product/quotient of the derivatives. (That is, differentiation does not distribute over multiplication or division.) However one can distribute before taking the derivative. That is $\frac{d}{dx}\Big((a+b)\times(c+d)\Big) = \frac{d}{dx}(ac+ad+b$

4.4 Derivatives of Trigonometric Functions

Sine, cosine, tangent, cosecant, secant, cotangent. These are functions that crop up continuously in mathematics and engineering and have a lot of practical applications. They also appear in more advanced mathematics, particularly when dealing with things such as line integrals with complex numbers and alternate representations of space like spherical and cylindrical coordinate systems.

We use the definition of the derivative, i.e.,

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

to work these first two out.

Let us find the derivative of sin(x), using the above definition.

$$f(x) = \sin(x)$$

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{\sin(x+h)-\sin(x)}{h} \text{ Definition of derivative} \\
&= \lim_{h\to 0} \frac{\cos(x)\sin(h)+\cos(h)\sin(x)-\sin(x)}{h} \quad \text{trigonometric identity} \\
&= \lim_{h\to 0} \frac{\cos(x)\sin(h)+(\cos(h)-1)\sin(x)}{h} \text{ factoring} \\
&= \lim_{h\to 0} \frac{\cos(x)\sin(h)}{h} + \lim_{h\to 0} \frac{(\cos(h)-1)\sin(x)}{h} \text{ separation of terms} \\
&= \cos(x)\times 1 + \sin(x)\times 0 \text{ application of limit} \\
&= \cos(x) \text{ solution}
\end{aligned}$$

Now for the case of $\cos(x)$.

$$f(x) = \cos(x)$$

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{\cos(x+h)-\cos(x)}{h} \text{ Definition of derivative} \\
&= \lim_{h\to 0} \frac{\cos(x)\cos(h)-\sin(h)\sin(x)-\cos(x)}{h} \quad \text{trigonometric identity} \\
&= \lim_{h\to 0} \frac{\cos(x)(\cos(h)-1)-\sin(x)\sin(h)}{h} \text{ factoring} \\
&= \lim_{h\to 0} \frac{\cos(x)(\cos(h)-1)}{h} - \lim_{h\to 0} \frac{\sin(x)\sin(h)}{h} \text{ separation of terms} \\
&= \cos(x)\times 0 - \sin(x)\times 1 \text{ application of limit} \\
&= -\sin(x) \text{ solution}
\end{aligned}$$

Therefore we have established

Derivative of Sine and Cosine

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

To find the derivative of the tangent, we just remember that:

$$\tan(x) = \frac{\sin(x)}{\cos(x)}$$

which is a quotient. Applying the quotient rule, we get:

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

Then, remembering that $\cos^2(x) + \sin^2(x) = 1$, we simplify:

$$\begin{aligned}\frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)}\\ &= \frac{1}{\cos^2(x)}\\ &= \sec^2(x)\end{aligned}$$

Derivative of the Tangent

$$\frac{d}{dx}\tan(x) = \sec^2(x)$$

For secants, we again apply the quotient rule.

$$\begin{aligned}\sec(x) &= \frac{1}{\cos(x)}\\ \frac{d}{dx}\sec(x) &= \frac{d}{dx}\frac{1}{\cos(x)}\\ &= \frac{\cos(x)\frac{d1}{dx} - 1\frac{d\cos(x)}{dx}}{\cos(x)^2}\\ &= \frac{\cos(x)0 - 1(-\sin(x))}{\cos(x)^2}\end{aligned}$$

Leaving us with:

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

Simplifying, we get:

Derivative of the Secant

$$\frac{d}{dx}\sec(x) = \sec(x)\tan(x)$$

Using the same procedure on cosecants:

$$\csc(x) = \frac{1}{\sin(x)}$$

We get:
Derivative of the Cosecant

$$\frac{d}{dx}\csc(x) = -\csc(x)\cot(x)$$

Using the same procedure for the cotangent that we used for the tangent, we get: Derivative of the Cotangent

$$\frac{d}{dx}\cot(x) = -\csc^2(x)$$

4.5 Chain Rule

The chain rule is a method to compute the derivative of the functional composition of two or more functions.

If a function, f , depends on a variable, u, which in turn depends on another variable, x, that is $f = y\big(u(x)\big)$, then the rate of change of f with respect to x can be computed as the rate of change of y with respect to u multiplied by the rate of change of u with respect to x.

Chain Rule

If a function f is composed to two differentiable functions $y(x)$ and $u(x)$, so that $f(x) = y\big(u(x)\big)$, $f(x)$ is differentiable and,

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

The method is called the "chain rule" because it can be applied sequentially to as many functions as are nested inside one another. For example, if f is a function of g which is in turn a function of h, which is in turn a function of x, that is

$$f\big(g(h(x))\big)$$

the derivative of f with respect to x is given by

$$\frac{df}{dx} = \frac{df}{dg}\cdot\frac{dg}{dh}\cdot\frac{dh}{dx}$$

and so on. A useful mnemonic is to think of the differentials as individual entities that can be canceled algebraically, such as

$$\frac{df}{dx} = \frac{df}{\cancel{dg}}\cdot\frac{\cancel{dg}}{\cancel{dh}}\cdot\frac{\cancel{dh}}{dx}$$

However, keep in mind that this trick comes about through a clever choice of notation rather than through actual algebraic cancellation.

The chain rule has broad applications in physics, chemistry, and engineering, as well as being used to study related rates in many disciplines. The chain rule can also be generalized to multiple variables in cases where the nested functions depend on more than one variable.

Example 4.8. *Suppose that a mountain climber ascends at a rate of 0.5 kilometer per hour. The temperature is lower at higher elevations; suppose the rate by which it decreases is 6^oC per kilometer. To calculate the decrease in air temperature per hour that the climber experiences, one multiplies 6^oC per kilometer by 0.5 kilometer per hour, to obtain 3^oC per hour. This calculation is a typical chain rule application.*

Example 4.9. *Consider the function $f(x) = (x^2 + 1)^3$.*

It follows from the chain rule that

$$
\begin{aligned}
&f(x) = (x^2+1)^3 \ \text{ Function to differentiate}\\
&u(x) = x^2+1 \text{ Define } u(x) \ \text{ as inside function}\\
&f(x) = u(x)^3 \text{ Express } \ f(x) \text{ in terms of } u(x)\\
&\frac{df}{dx} = \frac{df}{du}\cdot\frac{du}{dx} \ \text{ Express chain rule applicable here}\\
&\frac{df}{dx} = \frac{d}{du}u^3\cdot\frac{d}{dx}(x^2+1) \ \text{ Substitute in } f(u) \text{ and } u(x)\\
&\frac{df}{dx} = 3u^2\cdot 2x \text{ Compute derivatives with power rule}\\
&\frac{df}{dx} = 3(x^2+1)^2\cdot 2x \text{ Substitute } u(x) \ \text{ back in terms of } x\\
&\frac{df}{dx} = 6x(x^2+1)^2 \text{ Simplify.}
\end{aligned}
$$

Example 4.10. *In order to differentiate the trigonometric function $f(x) = \sin(x^2)$*

one can write:

$f(x) = \sin(x^2)$ Function to differentiate

$u(x) = x^2$ Define $u(x)$ as inside function

$f(x) = \sin(u)$ Express$f(x)$ in terms of $u(x)$

$\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx}$ Express chain rule applicable here

$\frac{df}{dx} = \frac{d}{du}\sin(u) \cdot \frac{d}{dx}(x^2)$ Substitute in $f(u)$ and $u(x)$

$\frac{df}{dx} = \cos(u) \cdot 2x$ Evaluate derivatives

$\frac{df}{dx} = \cos(x^2) \cdot 2x$ Substitute uu in terms of x.

Example 4.11. *The chain rule can be used to differentiate $|x|$, the absolute value function:*

$f(x) = |x|$ Function to differentiate

$f(x) = \sqrt{x^2}$ Equivalent function

$u(x) = x^2$ Define $u(x)$ as inside function

$f(x) = u(x)^{\frac{1}{2}}$ Express $f(x)$in terms of $u(x)$

$\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx}$ Express chain rule applicable here

$\frac{df}{dx} = \frac{d}{du}u^{\frac{1}{2}} \cdot \frac{d}{dx}(x^2)$ Substitute in$f(u)$ and $u(x)$

$\frac{df}{dx} = \frac{u^{-\frac{1}{2}}}{2} \cdot 2x$ Compute derivatives with power rule

$\frac{df}{dx} = \frac{(x^2)^{-\frac{1}{2}}}{2} \cdot 2x$ Substitute $u(x)$ back in terms of$x$

$\frac{df}{dx} = \frac{x}{\sqrt{x^2}}$ Simplify

$\frac{df}{dx} = \frac{x}{|x|}$ Express $\sqrt{x^2}$ as absolute value.

Example 4.12. *three nested functions*

The method is called the "chain rule" because it can be applied sequentially to as many functions as are nested inside one another. For example, if $f\Big(g(h(x))\Big) = e^{\sin(x^2)}$, sequential application of the chain rule yields the derivative as follows (we make use of the fact that

$\frac{d}{dx}e^x = e^x$, which will be proved in a later section):

$$f(x) = e^{\sin(x^2)} = e^g \text{ Original (outermost) function}$$
$$h(x) = x^2 \text{ Define } h(x) \text{ as innermost function}$$
$$g(x) = \sin(h) = \sin(x^2)$$
$$g(h) = sin(h) \text{ as middle function}$$
$$\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dh} \cdot \frac{dh}{dx} \text{ Express chain rule applicable here}$$
$$\frac{df}{dg} = e^g = e^{\sin(x^2)} \text{ Differentiate } f(g)$$
$$\frac{dg}{dh} = \cos(h) = \cos(x^2) \text{ Differentiate } g(h)$$
$$\frac{dh}{dx} = 2x \text{ Differentiate } h(x)$$
$$\frac{d}{dx}e^{\sin(x^2)} = e^{\sin(x^2)} \cdot \cos(x^2) \cdot 2x \text{ Substitute into chain rule.}$$

Chain Rule in Physics

Chain Rule in Chemistry

4.6 Higher Order Derivatives

The second derivative, or second order derivative, is the derivative of the derivative of a function. The derivative of the function $f(x)$ may be denoted by $f'(x)$, and its double (or "second") derivative is denoted by $f''(x)$. This is read as "f double prime of x," or "The second derivative of f(x)". Because the derivative of function f is defined as a function representing the slope of function f, the double derivative is the function representing the slope of the first derivative function.

Furthermore, the third derivative is the derivative of the derivative of the derivative of a function, which can be represented by $f'''(x)$. This is read as "f triple prime of x", or "The third derivative of f(x)" . This can continue as long as the resulting derivative is itself differentiable, with the fourth derivative, the fifth derivative, and so on. Any derivative beyond the first derivative can be referred to as a higher order derivative.

Notation

Let $f(x)$ be a function in terms of x. The following are notations for higher order derivatives.

2^{nd} Derivative	3^{rd} Derivative	4^{th} Derivative	n^{th}	Derivative Notes
$f''(x)$	$f'''(x)$	$f^{(4)}(x)$	$f^{(n)}(x)$	Probably the most common notation
$\frac{d^2 f}{dx^2}$	$\frac{d^3 f}{dx^3}$	$\frac{d^4 f}{dx^4}$	$\frac{d^n f}{dx^n}$	Leibniz notation.
$\frac{d^2}{dx^2}[f(x)]$	$\frac{d^3}{dx^3}[f(x)]$	$\frac{d^4}{dx^4}[f(x)]$	$\frac{d^n}{dx^n}[f(x)]$	Another form of Leibniz notation.
$D^2 f$	$D^3 f$	$D^4 f$	$D^n f$	Euler's notation.

Warning: You should not write $f^n(x)$ to indicate the nth derivative, as this is easily confused with the quantity $f(x)$ all raised to the nth power.

The Leibniz notation, which is useful because of its precision, follows from

$$\frac{d}{dx}\left(\frac{df}{dx}\right) = \frac{d^2 f}{dx^2}.$$

Newton's dot notation extends to the second derivative, $\ddot{y}$, but typically no further in the applications where this notation is common.

Example 4.13. *Find the third derivative of*

$$f(x) = 4x^5 + 6x^3 + 2x + 1$$

with respect to x.

Repeatedly apply the Power Rule to find the derivatives.

$$f'(x) = 20x^4 + 18x^2 + 2$$
$$f''(x) = 80x^3 + 36x$$
$$f'''(x) = 240x^2 + 36$$

Example 4.14. *Find the third derivative of*

$$f(x) = 12\sin(x) + \frac{1}{x+2} + 2x$$

with respect to x.

$$f'(x) = 12\cos(x) - \frac{1}{(x+2)^2} + 2$$
$$f''(x) = -12\sin(x) + \frac{2}{(x+2)^3}$$
$$f'''(x) = -12\cos(x) - \frac{6}{(x+2)^4}$$

Example 4.15. *If a)* $f(x) = x^4 - 8x^3 + 5x + 4,$ *b)* $f(x) = \frac{1}{x}$ *then find* $f^{(n)}(x)$.

Solution

a) $f(x) = x^4 - 8x^3 + 5x + 4$, then

$$\begin{aligned} f'(x) &= 4x^3 - 24x^2 + 5 \\ f''(x) &= 12x^2 - 48x \\ f'''(x) &= 24x - 48 \\ f^{(4)}(x) &= 24 \\ f^{(5)}(x) &= 0 \end{aligned}$$

and in fact $f^{(n)}(x) = 0$ for all $n \geq 5$.

b)

$$\begin{aligned} f(x) &= \frac{1}{x} = x^{-1} \\ f'(x) &= (-1)x^{-1-1} = -x^{-2} = -\frac{1}{x^2} \\ f''(x) &= (-2)(-1)x^{-2-1} = 2x^{-3} = 2\frac{1}{x^3} \\ f'''(x) &= (-3)(-2)(-1)x^{-3-1} = -6x^{-3-1} = -6\frac{1}{x^4} \\ f^{(4)}(x) &= (-4)(-3)(-2)(-1)x^{-4-1} = 24x^{-5} = 2\frac{1}{x^5} \\ f^{(5)}(x) &= (-5)(4)(3)(2)(1)x^{-5-1} = (x^{-6} = -(5!)\frac{1}{x^6} \\ &\vdots \\ f^{(n)}(x) &= (-1)^n(n)(n-1)(n-2)\cdots(4)(3)(2)(1)x^{-n-1} = (-1)^n(n!)\frac{1}{x^{n+1}} \end{aligned}$$

4.7 Implicit Differentiation

Generally, you will encounter functions expressed in explicit form, that is, in the form $y = f(x)$. To find the derivative of y with respect to x, you take the derivative with respect to x of both sides of the equation to get

$$\frac{dy}{dx} = \frac{d}{dx}[f(x)] = f'(x)$$

But suppose you have a relation of the form $f(x, y(x)) = g(x, y(x))$. In this case, it may be inconvenient or even impossible to solve for y as a function of x. A good example is the relation $y^2 + 2yx + 3 = 5x$. In this case you can utilize implicit differentiation to find the derivative. To do so, one takes the derivative of both sides of the equation with respect to x and solves for y'. That is, form

$$\frac{d}{dx}[f(x, y(x))] = \frac{d}{dx}[g(x, y(x))]$$

and solve for dy/dx. You need to employ the chain rule whenever you take the derivative of a variable with respect to a different variable. For example,

$$\frac{d}{dx}(y^3) = \frac{d}{dy}[y^3] \cdot \frac{dy}{dx} = 3y^2 \cdot y'$$

Implicit Differentiation and the Chain Rule

To understand how implicit differentiation works and use it effectively it is important to recognize that the key idea is simply the chain rule. First let's recall the chain rule. Suppose we are given two differentiable functions f(x) and g(x) and that we are interested in computing the derivative of the function f(g(x)), the chain rule states that:

$$\frac{d}{dx}\Big(f(g(x))\Big) = f'(g(x))g'(x)$$

That is, we take the derivative of f as normal and then plug in g, finally multiply the result by the derivative of g.

Now suppose we want to differentiate a term like y^2 with respect to x where we are thinking of y as a function of x, so for the remainder of this calculation let's write it as y(x) instead of just y. The term y^2 is just the composition of $f(x) = x^2$ and y(x). That is, $f(y(x)) = y^2(x)$. Recalling that $f'(x) = 2x$ then the chain rule states that:

$$\frac{d}{dx}\Big(f(y(x))\Big) = f'(y(x))y'(x) = 2y(x)y'(x)$$

Of course it is customary to think of y as being a function of x without always writing y(x), so this calculation usually is just written as

$$\frac{d}{dx}y^2 = 2yy'$$

Don't be confused by the fact that we don't yet know what y' is, it is some function and often if we are differentiating two quantities that are equal it becomes possible to explicitly solve for y' (as we will see in the examples below.) This makes it a very powerful technique for taking derivatives.

Explicit Differentiation

For example, suppose we are interested in the derivative of y with respect to x where x and y are related by the equation

$$x^2 + y^2 = 1$$

This equation represents a circle of radius 1 centered on the origin. Note that y is not a function of x since it fails the vertical line test $y = \pm 1$ when $x = 0$, for example).

To find y', first we can separate variables to get

$$y^2 = 1 - x^2$$

Taking the square root of both sides we get two separate functions for y:

$$y = \pm\sqrt{1 - x^2}$$

We can rewrite this as a fractional power:

$$y = \pm(1 - x^2)^{\frac{1}{2}}$$

Using the chain rule we get,

$$y' = \pm\frac{(1 - x^2)^{-\frac{1}{2}} \cdot (-2x)}{2} = \pm\frac{x}{\sqrt{1 - x^2}}$$

And simplifying by substituting y back into this equation gives

$$y' = -\frac{x}{y}$$

Implicit Differentiation

Using the same equation

$$x^2 + y^2 = 1$$

First, differentiate with respect to x on both sides of the equation:

$$\frac{d}{dx}[x^2 + y^2] = \frac{d}{dx}[1]$$
$$\frac{d}{dx}[x^2] + \frac{d}{dx}[y^2] = 0$$

To differentiate the second term on the left hand side of the equation (call it $f(y(x)) = y^2$), use the chain rule:

$$\frac{df}{dx} = \frac{df}{dy} \cdot \frac{dy}{dx} = 2y \cdot y'$$

So the equation becomes

$$2x + 2yy' = 0$$

Separate the variables:

$$2yy' = -2x$$

Divide both sides by $2y$, and simplify to get the same result as above:

$$y' = -\frac{2x}{2y}$$
$$y' = -\frac{x}{y}$$

Implicit differentiation is useful when differentiating an equation that cannot be explicitly differentiated because it is impossible to isolate variables.

Example 4.16. *consider the equation,* $x^2 + xy + y^2 = 16$

Differentiate both sides of the equation (remember to use the product rule on the term xy):

$$2x + y + xy' + 2yy' = 0$$

Isolate terms with y':

$$xy' + 2yy' = -2x - y$$

Factor out a y' and divide both sides by the other term:

$$y' = \frac{-2x - y}{x + 2y}$$

Example 4.17. $xy = 1$

can be solved as:

$$y = \frac{1}{x}$$

then differentiated:

$$\frac{dy}{dx} = -\frac{1}{x^2}$$

However, using implicit differentiation it can also be differentiated like this:

$$\frac{d}{dx}[xy] = \frac{d}{dx}[1]$$

use the product rule:

$$x\frac{dy}{dx} + y = 0$$

solve for $\frac{dy}{dx}$:

$$\frac{dy}{dx} = -\frac{y}{x}$$

Note that, if we substitute $y = \frac{1}{x}\frac{dy}{dx} = -\frac{y}{x}$, we end up with $\frac{dy}{dx} = -\frac{1}{x^2}$ again.

4.8 Derivatives of Inverse Functions

The general relationship between derivatives of f and f^{-1}. For this purpose, suppose that both functions are differentiable, and let $y = f^{-1}(x)$. Rewriting this equation as $x = f(y)$ and differentiating both side with respect to x yields $\frac{d}{dx}[x] = \frac{d}{dx}[f(y)] \Rightarrow 1 = f'(y)\frac{dy}{dx}$ then

$$\frac{dy}{dx} = \frac{1}{f'(y)}$$

Since $y = f^{-1}(x)$ we obtain the following formula that relates the derivative of f^{-1} to the derivative of f.

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

Example 4.18. *$y = f^{-1}(x) = \sqrt{x}$ then $f(x) = x^2$ and $f'(x) = 2x$ therefore*

$$\begin{aligned} \frac{dy}{dx} &= \frac{1}{f'(y)} \\ &= \frac{1}{2y} \\ &= \frac{1}{2\sqrt{x}} \end{aligned}$$

Example 4.19. *Let $y = f^{-1}(x) = \ln x$ then $f(x) = e^x$ and $f'(x) = e^x$ hence $f'(y) = e^y = e^{\ln x} = x$ therefore by derivative of inverse function we have*

$$\frac{d}{dx}(\ln x) = \frac{1}{f'(y)} = \frac{1}{x}$$

4.8.1 Derivatives of inverse trigonometric functions

Arcsine, arccosine, arctangent. These are the functions that allow you to determine the angle given the sine, cosine, or tangent of that angle.

First, let us start with the arcsine such that:

$$y = \arcsin(x)$$

To find dy/dx we first need to break this down into a form we can work with:

$$x = \sin(y)$$

Then we can take the derivative of that:

$$1 = \cos(y) \cdot \frac{dy}{dx}$$

and solve for dy/dx : $y = \arcsin(x)$ gives us this unit triangle.

$$\frac{dy}{dx} = \frac{1}{\cos(y)}$$

At this point we need to go back to the unit triangle. Since y is the angle and the opposite side is sin(y) (which is equal to x), the adjacent side is cos(y) (which is equal to the square root of $1 - x^2$, based on the Pythagorean theorem), and the hypotenuse is 1. Since we have determined the value of cos(y) based on the unit triangle, we can substitute it back in to the above equation and get:

Derivative of the Arcsine

$$\frac{d}{dx}\arcsin(x) = \frac{1}{\sqrt{1-x^2}}$$

We can use an identical procedure for the arccosine and arctangent:

Derivative of the Arccosine

$$\frac{d}{dx}\arccos(x) = -\frac{1}{\sqrt{1-x^2}}$$

Derivative of the Arctangent

$$\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$$

4.9 Derivatives of Exponential and Logarithm Functions

4.9.1 Logarithm Function

We shall first look at the value of e :

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

Now we find the derivative of $\ln x$ using the formal definition of the derivative:

$$\frac{d}{dx}\ln(x) = \lim_{\Delta x\to 0}\frac{\ln(x+\Delta x)-\ln(x)}{\Delta x} =$$
$$\lim_{\Delta x\to 0}\frac{1}{\Delta x}\ln\left(\frac{x+\Delta x}{x}\right) = \lim_{\Delta x\to 0}\ln\left(\frac{x+\Delta x}{x}\right)^{\frac{1}{\Delta x}}$$

Let $n = \frac{x}{\Delta x}$. Note that as n approaches ∞, Δx approaches 0. So we can redefine our limit as:

$$\lim_{n\to\infty}\ln\left(1+\frac{1}{n}\right)^{\frac{n}{x}} = \frac{1}{x}\ln\left(\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n\right) = \frac{1}{x}\ln(e) = \frac{1}{x}$$

Here we could take the natural logarithm outside the limit because it doesn't have anything to do with the limit (we could have chosen not to). We then substituted the value of e.

Derivative of the Natural Logarithm

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

If we wanted, we could go through that same process again for a generalized base, but it is easier just to use properties of logs and realize that:

$$\log_b(x) = \frac{\ln(x)}{\ln(b)}$$

Since $1/\ln(b)$ is a constant, we can just take it outside of the derivative:

$$\frac{d}{dx}\log_b(x) = \frac{1}{\ln(b)} \cdot \frac{d}{dx}\ln(x)$$

Which leaves us with the generalized form of:

4.9.2 Derivative of the Logarithm

$$\frac{d}{dx}\log_b(x) = \frac{1}{x\ln(b)}$$

4.9.3 Exponential Function

We shall take two different approaches to finding the derivative of $\ln(e^x)$. The first approach:

$$\frac{d}{dx}\ln(e^x) = \frac{d}{dx}x = 1$$

The second approach:

$$\frac{d}{dx}\ln(e^x) = \frac{1}{e^x}\left(\frac{d}{dx}e^x\right)$$

Note that in the second approach we made some use of the chain rule.Thus:

$$\frac{1}{e^x}\left(\frac{d}{dx}e^x\right) = 1$$
$$\frac{d}{dx}e^x = e^x$$

so that we have proved the following rule:

4.9.4 Derivative of the exponential function

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

Now that we have derived a specific case, let us extend things to the general case. Assuming that a is a positive real constant, we wish to calculate:

$$\frac{d}{dx}a^x$$

One of the oldest tricks in mathematics is to break a problem down into a form that we already know we can handle. Since we have already determined the derivative of e^x, we will attempt to rewrite ax in that form.

Using that $e^{\ln(c)} = c$ and that $\ln(a^b) = b\ln(a)$, we find that:

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

Thus, we simply apply the chain rule:

$$\frac{d}{dx}e^{x\cdot\ln(a)} = \frac{d}{dx}[x\cdot\ln(a)]e^{x\cdot\ln(a)} = \ln(a)a^x$$

Derivative of the exponential function

$$\frac{d}{dx}a^x = \ln(a)a^x$$

4.9.5 Logarithmic Differentiation

We can use the properties of the logarithm, particularly the natural log, to differentiate more difficult functions, such a products with many terms, quotients of composed functions, or functions with variable or function exponents. We do this by taking the natural logarithm of both sides, re-arranging terms using the logarithm laws below, and then differentiating both sides implicitly, before multiplying through by y.

$$\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$$

$$\log(a^n) = n\log(a)$$

$$\log(a) + \log(b) = \log(ab)$$

Example 4.20. *We shall now prove the validity of the power rule using logarithmic differentiation.*

$$\frac{d}{dx}\ln(x^n) = n\frac{d}{dx}\ln(x) = nx^{-1}$$

$$\frac{d}{dx}\ln(x^n) = \frac{1}{x^n}\cdot\frac{d}{dx}x^n$$

Thus:

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

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

Example 4.21. *Suppose we wished to differentiate*

$$y = \frac{(6x^2+9)^2}{\sqrt{3x^3-2}}$$

We take the natural logarithm of both sides

$$\begin{aligned}\ln(y) &= \ln\left(\frac{(6x^2+9)^2}{\sqrt{3x^3-2}}\right)\\ &= \ln(6x^2+9)^2 - \ln(3x^3-2)^{\frac{1}{2}}\\ &= 2\ln(6x^2+9) - \frac{1}{2}\ln(3x^3-2)\end{aligned}$$

Differentiating implicitly, recalling the chain rule

$$\begin{aligned}\frac{1}{y}\cdot\frac{dy}{dx} &= 2\times\frac{12x}{6x^2+9} - \frac{1}{2}\times\frac{9x^2}{3x^3-2}\\ &= \frac{24x}{6x^2+9} - \frac{\frac{9}{2}x^2}{3x^3-2}\\ &= \frac{24x(3x^3-2) - \frac{9}{2}x^2(6x^2+9)}{(6x^2+9)(3x^3-2)}\end{aligned}$$

Multiplying by y, the original function

$$\frac{dy}{dx} = \frac{(6x^2+9)^2}{\sqrt{3x^3-2}}\times\frac{24x(3x^3-2) - \frac{9}{2}x^2(6x^2+9)}{(6x^2+9)(3x^3-2)}$$

Example 4.22. *Let us differentiate a function*

$$y = x^x$$

Taking the natural logarithm of left and right

$$\begin{aligned}\ln(y) &= \ln(x^x)\\ &= x\ln(x)\end{aligned}$$

We then differentiate both sides, recalling the product and chain rules

$$\begin{aligned}\frac{1}{y}\cdot\frac{dy}{dx} &= \ln(x) + x\frac{1}{x}\\ &= \ln(x) + 1\end{aligned}$$

Multiplying by the original function y

$$\frac{dy}{dx} = x^x(\ln(x)+1)$$

Example 4.23. *Take a function*

$$y = x^{6\cos(x)}$$

Then

$$\begin{aligned}\ln(y) &= \ln(x^{6\cos(x)}) \\ &= 6\cos(x)\ln(x)\end{aligned}$$

We then differentiate

$$\frac{1}{y}\cdot\frac{dy}{dx} = -6\sin(x)\ln(x) + \frac{6\cos(x)}{x}$$

And finally multiply by y

$$\begin{aligned}\frac{dy}{dx} &= y\left(-6\sin(x)\ln(x) + \frac{6\cos(x)}{x}\right) \\ &= x^{6\cos(x)}\left(-6\sin(x)\ln(x) + \frac{6\cos(x)}{x}\right)\end{aligned}$$

4.10 Derivative of Hyperbolic and Inverse Hyperbolic Function

4.10.1 Definition of the Hyperbolic Functions

Certain even and odd combinations of the exponential functions $e^x$and e^{-x} arise so frequently in mathematics and its applications that they deserve to be given special names. In many ways they are analogous to the trigonometric functions, and they have the same relationship to the hyperbola that the trigonometric functions have to the circle. For this reason they are collectively called **hyperbolic functions** and individually called **hyperbolic sine, hyperbolic cosine**, and so on.

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

The hyperbolic functions satisfy a number of identities that are similar to well-known trigonometric identities. We list some of them here

4.10.2 Hyperbolic Identities

$$\sinh(-x) = -\sinh x \qquad \cosh(-x) = \cosh x$$
$$\cosh^2 x - \sinh^2 x = 1 \qquad 1 - \tanh^2 x = sech^2 x$$
$$\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y$$
$$\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y$$

4.10.3 Derivatives of Hyperbolic Functions

The derivatives of the hyperbolic functions are easily computed.
For example,
$$\frac{d}{dx}(\cosh x) = \frac{d}{dx}(\frac{e^x + e^{-x}}{2}) = \frac{e^x - e^{-x}}{2} = \sinh x$$
Note the analogy with the differentiation formulas for trigonometric functions, but beware that the signs are different in some cases.
We list the differentiation formulas for the hyperbolic functions as
$$\frac{d}{dx}(\sinh x) = \cosh x \qquad \frac{d}{dx}(\cosh x) = \sinh x$$
$$\frac{d}{dx}(cschx) = -cschx \tanh x \qquad \frac{d}{dx}(sechx) = -sechx \coth x$$
$$\frac{d}{dx}(\tanh x) = sech^2 x \qquad \frac{d}{dx}(\coth x) = -csch^2 x$$

4.10.4 Inverse of Hyperbolic Functions

Both sinh and tanh are one-to-one functions and so they have inverse functions denoted by $\sinh^{-1}$ and $\tanh^{-1}$ but cosh is not one to one, when restricted to the domain $[0, \infty)$ it becomes one-to-one. The inverse hyperbolic cosine function is defined as the inverse of this restricted function.

$$y = \sinh^{-1} x \quad \Leftrightarrow \quad \sinh y = x$$
$$y = \cosh^{-1} x \quad \Leftrightarrow \quad \cosh y = x$$
$$y = \tanh^{-1} x \quad \Leftrightarrow \quad \tanh y = x$$

The remaining inverse hyperbolic functions are defined similarly way
Since the hyperbolic functions are defined in terms of exponential functions, it's not surprising to learn that the inverse hyperbolic functions can be expressed in terms of logarithms.

In particular, we have:

$$\sinh^{-1} x = \ln{(x + \sqrt{x^2 + 1})} \quad x \in \mathbb{R}$$
$$\cosh^{-1} x = \ln{(x + \sqrt{x^2 - 1})} \quad x \geq 1$$
$$\tanh^{-1} x = \frac{1}{2}\ln{(\frac{1+x}{1-x})} \quad -1 \leq x \leq 1$$

Example 4.24. *Show that* $\cosh^{-1} x = \ln(x + \sqrt{x^2 - 1})$

Solution

Let $y = \cosh^{-1} x$. Then

$$x = \cosh y = \frac{e^y + e^{-y}}{2}$$
$$2x = e^y + e^{-y}$$
$$e^y - 2x + e^{-y} = 0$$
$$\Longrightarrow e^{2y} - 2xe^y + 1 = 0 \quad \text{multiply by } e^y$$
$$\Longrightarrow (e^y)^2 - 2xe^y + 1 = 0$$

This is really a quadratic equation in e^y then Solving by the quadratic formula, we get

$$e^y = \frac{2x \pm \sqrt{4x^2 - 4}}{2} = x \pm \sqrt{x^2 - 1}$$

Note that $e^y > 0$, and $x \pm \sqrt{x^2 - 1} > 0$ for all $x \geq 1$ but the inverse hyperbolic cosine function is defined in the domain $[1, \infty)$ it becomes one-to-one *i.e* $y \geq 0$. Thus, the minus sign is inadmissible and we have

$$y = \ln{(x + \sqrt{x^2 - 1})}$$

Therefore

$$\cosh^{-1} x = \ln{(x + \sqrt{x^2 - 1})}$$

In similar way

4.10.5 Derivatives of inverse hyperbolic function

$$\frac{d}{dx}(\sinh^{-1} x) = \frac{1}{\sqrt{x^2 + 1}} \qquad \frac{d}{dx}(\cosh^{-1} x) = \frac{1}{\sqrt{x^2 - 1}}$$
$$\frac{d}{dx}(\tanh^{-1} x) = \frac{1}{1 - x^2} \qquad \frac{d}{dx}(\coth^{-1} x) = \frac{1}{1 - x^2}$$
$$\frac{d}{dx}(csch^{-1} x) = -\frac{1}{|x|\sqrt{x^2 - 1}} \qquad \frac{d}{dx}(sech^{-1} x) = \frac{1}{|x|\sqrt{x^2 - 1}}$$

Example 4.25. *Prove that* $\dfrac{d}{dx}(\sinh^{-1} x) = \dfrac{1}{\sqrt{x^2 + 1}}$

Solution

$$\begin{aligned}
\frac{d}{dx}(\sinh^{-1} x) &= \frac{d}{dx}(\ln{(x+\sqrt{x^2+1})}) \\
&= \frac{1}{x+\sqrt{x^2+1}}\frac{d}{dx}(x+\sqrt{x^2+1}) \\
&= \frac{1}{x+\sqrt{x^2+1}}[1+\frac{x}{\sqrt{x^2+1}}] \\
&= \frac{1}{x+\sqrt{x^2+1}}[\frac{x+\sqrt{x^2+1}}{\sqrt{x^2+1}}] \\
&= \frac{1}{\sqrt{x^2+1}}
\end{aligned}$$

or Let $y = sinh^{-1}x$. Then $sinhy = x$. If we differentiate this equation implicitly with respect to x, we get

$$\cosh y \frac{dy}{dx} = 1$$

since $\cosh^2 x - \sinh^2 x = 1$ and $\cosh y \geq 0$, we have $\cosh y = \sqrt{1+\sinh^2 x}$, so

$$\frac{dy}{dx} = \frac{1}{\cosh y} = \frac{1}{\sqrt{1+\sinh^2 x}} = \frac{1}{\sqrt{1+x^2}}$$

5 Application of Derivative

One of the most important applications of the derivative is its use as a tool for finding the optimal (best) solutions to problems. Optimization problems abound in mathematics, physical science and engineering, business and economics, and biology and medicine.

In this section we use derivatives to find extreme values of functions, to determine and analyze the shapes of graphs, and evaluate indeterminate form of limits.

5.1 Extreme values of Functions

This subsection shows how to locate and identify extreme (maximum or minimum) values of a function from its derivative. Once we can do this, we can solve a variety of optimization problems .

Definition 5.1. ***Global Maximum*** *A global maximum (also called an absolute maximum) of a function f on a closed interval I is a value $f(c)$ such that $f(c) \geq f(x)$ for all x in I.*

Definition 5.2. ***Global Minimum:*** *A global minimum (also called an absolute minimum) of a function f on a closed interval I is a value $f(c)$ such that $f(c) \leq f(x)$ for all x in I.*

Maxima and minima are collectively known as extrema.

Theorem 5.1. ***Extreme Value Theorem*** *If f is a function that is continuous on the closed interval $[a, b]$, then f has both a global minimum and a global maximum on $[a, b]$. It is assumed that a and b are both finite.*

Definition 5.3. ***Relative Extreme Values***
A function f has a local maximum value at an interior point c of its domain if $f(x) \leq f(c)$ for all x in some open interval containing c.

In Figure 5.1, the function f has local maxima at c and d and local minima at a, e, and b. Local extrema are also called relative extrema. An absolute maximum is also a local maximum. Being the largest value overall, it is also the largest value in its immediate neighborhood. Hence, a list of all local maxima will automatically include the absolute maximum if there is one. Similarly, a list of all local minima will include the absolute minimum if there is one

Theorem 5.2. ***Fermat's Theorem*** *If f has a local maximum or minimum at c, and if $f'(c)$ exists, then $f'(c) = 0$.*

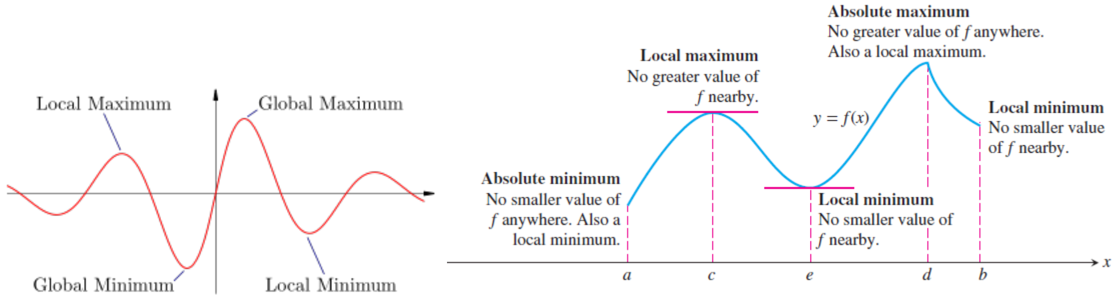


Figure 5.1: How to identify types of maxima and minima for a function with domain $a \leq x \leq b$.

Definition 5.4. *A critical number of a function f is a number c in the domain of f such that either $f'(c) = 0$ or $f'(c)$ does not exist.*

Note: If f has a local maximum or minimum at c, then c is a critical number of f .

The Closed Interval Method

To find the absolute maximum and minimum values of a continuous function f on a closed interval [a, b]:

1. Find the values of f at the critical numbers of f in (a, b).
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 value; the smallest of these values is the absolute minimum value.

Theorem 5.3. ***Rolle's Theorem:*** *If a function, $f(x)$, is continuous on the closed interval $[a, b]$, is differentiable on the open interval (a, b), and $f(a) = f(b)$, then there exists at least one number c, in the interval (a, b) such that $f'(c) = 0$.*

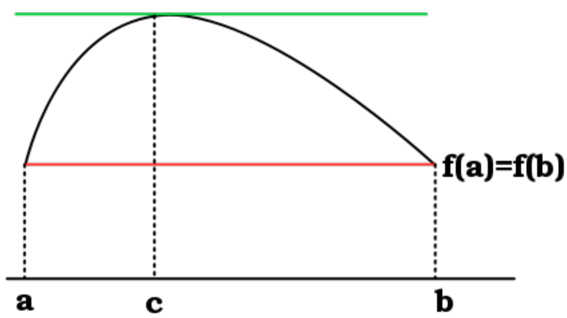


Figure 5.2: Rolle's theorem.

Theorem 5.4. ***Mean Value Theorem***

If $f(x)$ is continuous on the closed interval $[a, b]$, where $a < b$,and differentiable on the open interval (a, b), there exists at least one c in the open interval (a, b) such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The Mean Value Theorem is an important theorem of differential calculus. It basically says that for a differentiable function defined on an interval, there is some point on the interval whose instantaneous slope is equal to the average slope of the interval. Note that Rolle's Theorem is the special case of the Mean Value Theorem when $f(a) = f(b)$.

5.2 Monotonic Functions and the First Derivative test

A function that is increasing or decreasing on an interval is said to be **monotonic** on the interval

5.2.1 Increasing, Decreasing Function

Let f be a function defined on an interval I and let x_1 and x_2 be any two points in I.

1. $f(x_1) < f(x_2)$ whenever $x_1 < x_2$ then f is said to be increasing on I.
2. $f(x_1) > f(x_2)$ whenever $x_1 < x_2$ then f is said to be decreasing on I.

It is important to realize that the definitions of increasing and decreasing functions must be satisfied for every pair of points x_1 and x_2 in I with $x_1 < x_2$. Because of the inequality $<$ comparing the function values, and not $\leq$ some books say that f is strictly increasing or decreasing on I. The interval I may be finite or infinite.

The function $f(x) = x^2$ decreases on $(-\infty, 0]$ and increases on $[0, \infty)$ as can be seen from its graph below. The function f is monotonic on $(-\infty, 0]$ and $[0, \infty)$ but it is not monotonic on $(-\infty, \infty)$.

Notice that on the interval $(-\infty, 0]$ the tangents have negative slopes, so the first derivative is always negative there; for $[0, \infty)$ the tangents have positive slopes and the first derivative is positive. The following result confirms these observations.

First Derivative Test for Monotonic Functions

Suppose that f is continuous on $[a, b]$ and differentiable on (a, b).

1. If $f'(x) > 0$ at each point $x \in (a, b)$, then f is increasing on $[a, b]$
2. If $f'(x) < 0$ at each point $x \in (a, b)$ then f is decreasing on $[a, b]$.

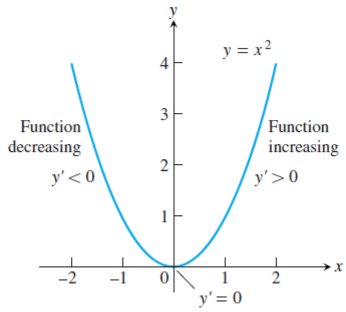


Figure 5.3: The function $f(x) = x^2$ is monotonic on the intervals $(-\infty, 0]$ and $[0, \infty)$ but it is not monotonic on $(-\infty, \infty)$.

First Derivative Test for local Extreme Values

In Figure 5.4, at the points where f has a minimum value, $f' < 0$ immediately to the left and $f' > 0$ immediately to the right. (If the point is an endpoint, there is only one side to consider.) Thus, the function is decreasing on the left of the minimum value and it is increasing on its right. Similarly, at the points where f has a maximum value, $f' > 0$ immediately to the left and $f' < 0$ immediately to the right. Thus, the function is increasing on the left of the maximum value and decreasing on its right. In summary, at a local extreme point, the sign of $f'(x)$changes.

Suppose that c is a critical number of a continuous function f, and that f is differentiable at every point in some interval containing c except possibly at c itself. Moving across this interval from left to right,

1. if f' changes from negative to positive at c, then f has a local minimum at c;
2. if f' changes from positive to negative at c, then f has a local maximum at c;
3. if f' does not change sign at c(that is, f' is positive on both sides of c or negative on both sides), then f has no local extremum at c.

5.3 Concavity and Curve sketching

Definition 5.5. *If the graph of f lies above all of its tangents on an interval I, then it is called concave upward on I. If the graph of f lies below all of its tangents on I, it is called concave downward on I*
Or
The graph of a differentiable function $y = f(x)$ is

1. *concave up on an open interval I if f' is increasing on I;*

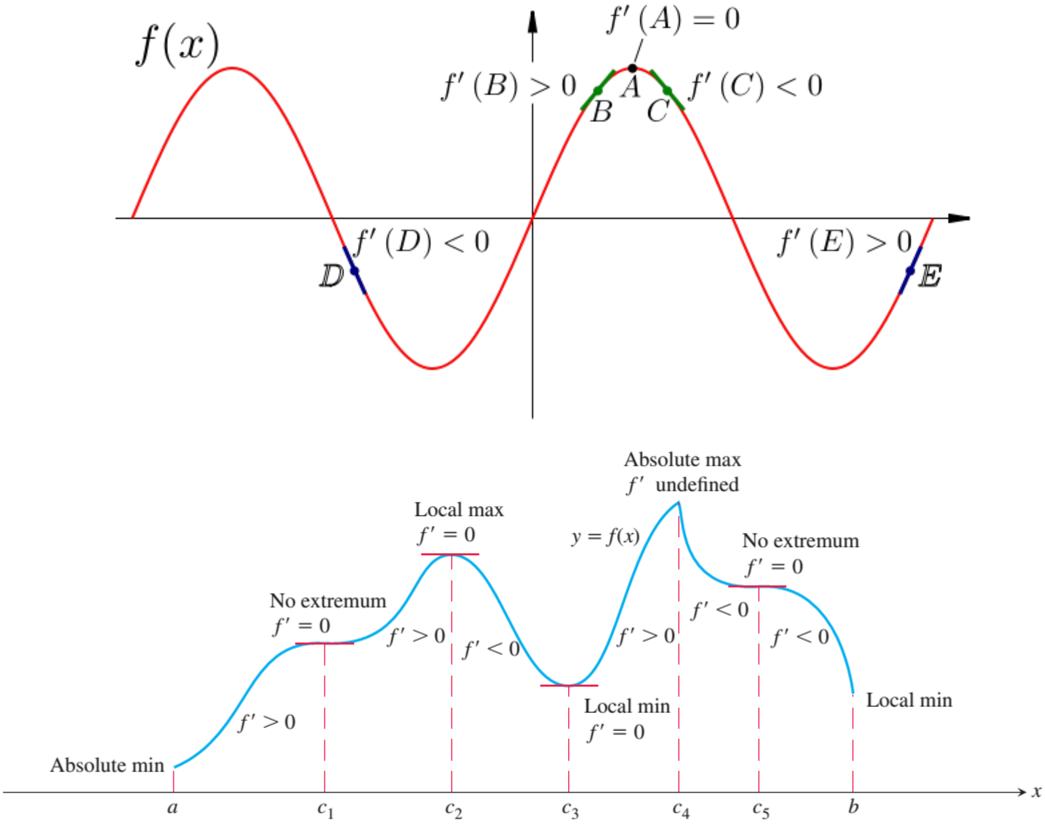


Figure 5.4: The critical points of a function locate where it is increasing and where it is decreasing. The first derivative changes sign at a critical point where a local extremum occurs.

2. concave down on an open interval I if f' is decreasing on I.

The Second Derivative Test for Concavity

Let $y = f(x)$ be twice differentiable on an interval I.

1. 1. If $f'' > 0$ on I, the graph of f over I is concave up.
2. If $f'' < 0$ on I, the graph of f over I is concave down.

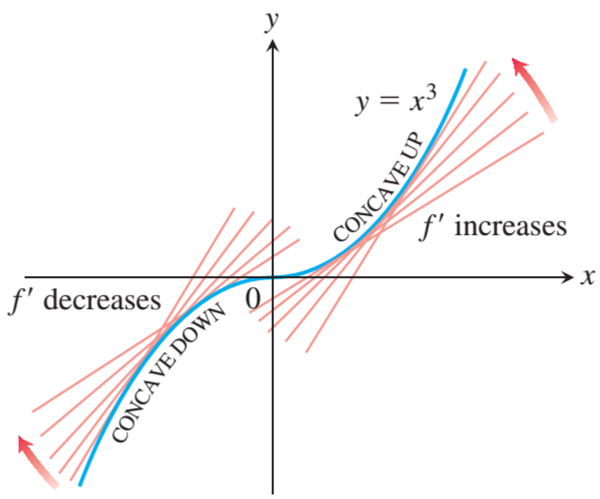


Figure 5.5: The graph of $f(x) = x^3$ is concave down on $(-\infty, 0)$ and concave up on $(0, \infty)$

Definition 5.6. *A point $(c, f(c))$ where the graph of a function has a tangent line and where the concavity changes is a* ***point of inflection****.*
Or
A point P on a curve $y = f(x)$ is called an ***inflection point*** *if f is continuous there and the curve changes from concave upward to concave downward or from concave downward to concave upward at P.*

At a point of inflection $(c, f(c))$, either $f''(c) = 0$ or $f''(c)$ fails to exist.

5.3.1 Second Derivative test for Local extrema

Instead of looking for sign changes in f' at critical points, we can sometimes use the following test to determine the presence and nature of local extrema.

Theorem 5.5. *Suppose f'' is continuous on an open interval that contains $x = c$.*

1. If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at $x = c$.

2. *If $f'(c) = 0$ and $f''(c) > 0$,then f has a local minimum at $x = c$.*

3. *If $f'(c) = 0$ and $f''(c) = 0$,then the test fails. The function f may have a local maximum, a local minimum, or neither.*

Strategy for Graphing $y = f(x)$

1. Identify the domain of f and any symmetries the curve may have.

2. Find y' and y''.

3. Find the critical points of f, and identify the functions behavior at each one.

4. Find where the curve is increasing and where it is decreasing.

5. Find the points of inflection, if any occur, and determine the concavity of the curve.

6. Identify any asymptotes.

7. Plot key points, such as the intercepts and the points found in Steps $3-5$, and sketch the curve.

Example 5.1. *Sketch a graph of the function*

$$f(x) = x^4 - 4x^3 + 10$$

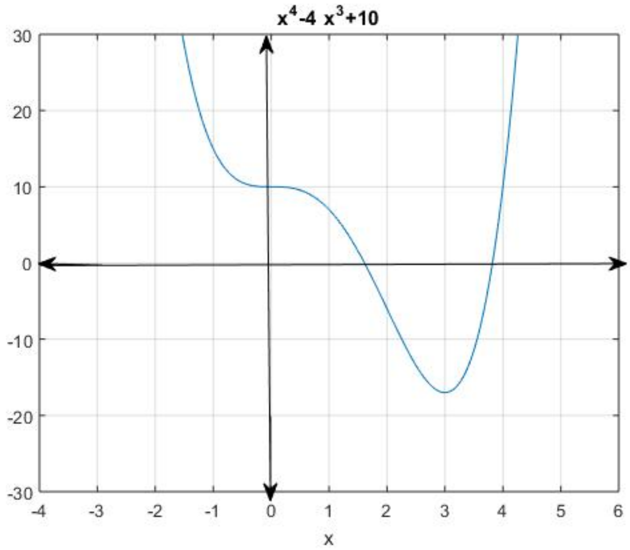


Figure 5.6: The graph of $f(x) = x^4 - 4x^3 + 10$

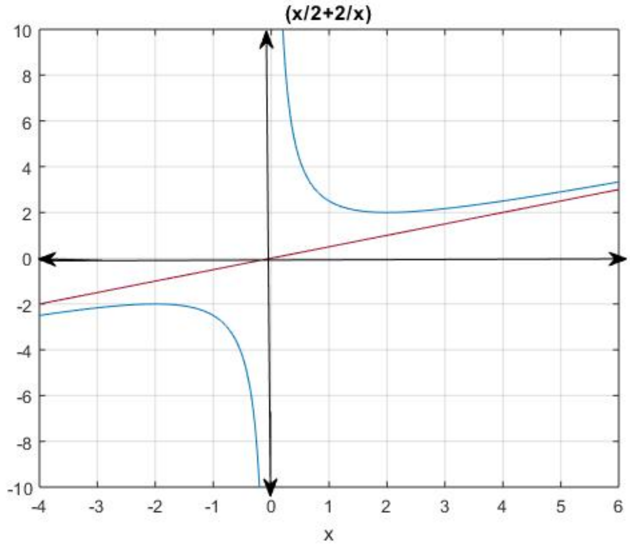


Figure 5.7: Graph of the function $f(x) = \dfrac{x^2+4}{2x}$

Example 5.2. *Sketch a graph of the function*

$$f(x) = \frac{x^2+4}{2x}$$

Example 5.3. *Sketch the graph of* $f(x) = \dfrac{2x^2}{x^2-1}$

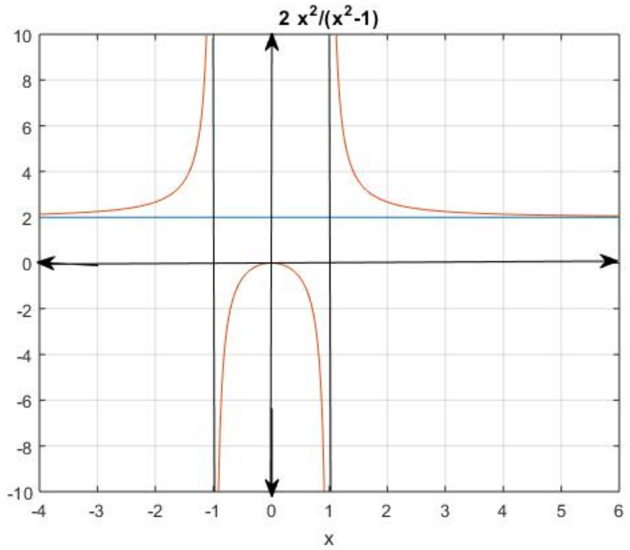


Figure 5.8: Graph of $f(x) = \dfrac{2x^2}{x^2-1}$

5.4 Indeterminate Forms L'Hospital's Rule

Suppose we are trying to analyze the behavior of the function

$$\frac{\ln x}{x-1}$$

Although f is not defined when $x = 1$, we need to know how f behaves near 1. In particular, we would like to know the value of the limit

$$\lim_{x\to 1} \frac{\ln x}{x-1} \tag{5.1}$$

In computing this limit we can't apply the limit of a quotient is the quotient of the limits, because the limit of the denominator is 0. In fact, although the limit in (5.1) exists, its value is not obvious because both numerator and denominator approach 0 and $\frac{0}{0}$ is not defined.

In general, if we have a limit of the form

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

where both $f(x) \to 0$ and $g(x) \to 0$ as $x \to a$, then this limit may or may not exist and is called an **indeterminate form of type** $\dfrac{0}{0}$.

Another situation in which a limit is not obvious occurs when we look for a horizontal asymptote of f and need to evaluate the limit

$$\lim_{x\to \infty} \frac{\ln x}{x-1}$$

It isn't obvious how to evaluate this limit because both numerator and denominator become large as $x \to \infty$. There is a struggle between numerator and denominator. If the numerator wins, the limit will be ∞ ; if the denominator wins, the answer will be 0. Or there may be some compromise, in which case the answer may be some finite positive number.

In general, if we have a limit of the form

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

where both $f \to \infty$ (or $f \to -\infty$) and $g \to \infty$(or $g \to -\infty$), then the limit may or may not exist and is called an **indeterminate form of type** $\dfrac{\infty}{\infty}$

So in this section we introduce a systematic method, known as **l'Hospital's Rule**, for the evaluation of indeterminate forms.

Theorem 5.6. ***L'Hospital's Rule*** *Suppose f and g are differentiable and $g'(x) \neq 0$ near a(except possibly at a). Suppose that $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$ or that $\lim_{x\to a} f(x) = \pm\infty$ and $\lim_{x\to a} g(x) = \pm\infty$*
(In other words, we have an indeterminate form of type $\frac{0}{0}$ or $\frac{\infty}{\infty}$.) Then

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

if the limit on the right side exists (or ∞ is or $-\infty$).

Example 5.4. *Find* $\lim_{x\to 1} \dfrac{\ln x}{x-1}$

Solution

Since $\lim_{x\to 1} \ln x = 0$ and $\lim_{x\to 1}(x-1) = 0$
we can apply l'Hospital's Rule:

$$\lim_{x\to 1} \frac{\ln x}{x-1} = \lim_{x\to 1} \frac{\frac{d}{dx}\ln x}{\frac{d}{dx}(x-1)} = \lim_{x\to 1} \frac{\frac{1}{x}}{1}$$
$$= \lim_{x\to 1} \frac{1}{x} = 1$$

Example 5.5. *Find* $\lim_{x\to 0} \dfrac{\tan x - x}{x^3}$

Solution

: Since $\lim_{x\to 0}(\tan x - x) = 0$ and $\lim_{x\to 0}(x^3) = 0$
we can apply l'Hospital's Rule:

$$\lim_{x\to 0} \frac{\tan x - x}{x^3} = \lim_{x\to 0} \frac{\sec^2 x - 1}{3x^2}$$

Since the limit on the right side is still indeterminate of type $\dfrac{0}{0}$, we apply l'Hospital's Rule again:

$$\lim_{x\to 0} \frac{\sec^2 x - 1}{3x^2} = \lim_{x\to 0} \frac{2\sec^2 x \tan x}{6x}$$

Because $\lim_{x\to 0} \sec^x = 1$, we simplify the calculation by writing

$$\lim_{x\to 0} \frac{2\sec^2 x \tan x}{6x} = \frac{1}{3} \lim_{x\to 0} \sec^2 x \lim_{x\to 0} \frac{\tan x}{x} = \frac{1}{3} \lim_{x\to 0} \frac{\tan x}{x}$$

We can evaluate this last limit either by using l'Hospital's Rule a third time or by writing $\tan x$ as and $\dfrac{\sin}{\cos x}$ making use of our knowledge of trigonometric limits. Putting together all the steps, we get

$$\lim_{x\to 0} \frac{\tan x - x}{x^3} = \lim_{x\to 0} \frac{\sec^2 x - 1}{3x^2} = \lim_{x\to 0} \frac{2\sec^2 x \tan x}{6x} = \frac{1}{3} \lim_{x\to 0} \frac{\tan x}{x} = \lim_{x\to 0} \frac{\sec^2 x}{1} = \frac{1}{3}$$

Example 5.6. *Find* $\lim\limits_{x\to\infty} \dfrac{e^x}{x^2}$

Solution

Since $\lim\limits_{x\to\infty} \ln x = \infty$ and $\lim\limits_{x\to\infty}(x^2) = \infty$
we can apply l'Hospital's Rule:

$$\lim_{x\to\infty} \frac{e^x}{x^2} = \lim_{x\to\infty} \frac{\frac{d}{dx}e^x}{\frac{d}{dx}(x^2)} = \lim_{x\to\infty} \frac{e^x}{2x}$$

Since $e^x \to \infty$ and $2x \to \infty$ as $x \to \infty$, the limit on the right side is also indeterminate, but a second application of l'Hospital's Rule gives

$$\lim_{x\to\infty} \frac{e^x}{x^2} = \lim_{x\to\infty} \frac{e^x}{2x} = \lim_{x\to\infty} \frac{e^x}{2} = \infty$$

Indeterminate Products

If $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = \infty$(or $\lim_{x\to a} g(x) = -\infty$), then it isn't clear what is the value of $\lim_{x\to a} g(x)f(x)$, if any, will be. There is a struggle between f and g. If f wins, the answer will be 0; if g wins, the answer will be ∞ (or $-\infty$). Or there may be a compromise where the answer is a finite nonzero number. This kind of limit is called an **indeterminate form of type** $0 \cdot \infty$. We can deal with it by writing the product fg as a quotient:

$$fg = \frac{f}{1/g} \text{ or } fg = \frac{g}{1/f}$$

This converts the given limit into an indeterminate form of type $\frac{0}{0}$ or $\frac{\infty}{\infty}$ so that we can use L'Hospital's Rule.

Example 5.7. *Evaluate* $\lim_{x\to 0+}(x \ln x)$

Solution

The given limit is indeterminate because, as $x \to 0$, the first factor x approaches 0 while the second factor $\ln x$ approaches $-\infty$. Writing $x = \frac{1}{x}$, we have $\frac{1}{x} \to \infty$ as $x \to 0^+$, so l'Hospital's Rule gives

$$\lim_{x\to 0^+}(x \ln x) = \lim_{x\to 0^+} \frac{\ln x}{1/x} = \lim_{x\to 0^+} \frac{1/x}{-1/x^2} = \lim_{x\to 0^+}(-x) = 0$$

Indeterminate Differences

If $\lim_{x\to a} f(x) = \infty$ and $\lim_{x\to a} g(x) = \infty$, then the limit

$$\lim_{x\to a}[f(x) - g(x)]$$

is called an indeterminate form of type $\infty - \infty$. Again there is a struggle between f and g. Will the answer be ∞ (if f wins) or will it be $-\infty$ (if g wins) or will they compromise on a finite number. To find out, we try to convert the difference into a quotient (for instance, by using a common denominator, or rationalization, or factoring out a common factor) so that we have an indeterminate form of type $\frac{0}{0}$ or $\frac{\infty}{\infty}$.

Example 5.8. *Compute* $\lim_{x\to \frac{\pi}{2}^-}[\sec x - \tan x]$.

Solution

First notice that $\sec x \to \infty$ and $\tan x \to \infty$ as $x \to \frac{\pi}{2}^-$, so the limit is indeterminate type $\infty - \infty$. Here we use a common denominator:

$$\lim_{x\to \frac{\pi}{2}^-}[\sec x - \tan x] = \lim_{x\to \frac{\pi}{2}^-}[\frac{1}{\cos x} - \frac{\sin x}{\cos x}] = \lim_{x\to \frac{\pi}{2}^-}\frac{1-\sin x}{\cos x} = \lim_{x\to \frac{\pi}{2}^-}\frac{-\cos x}{\sin x} = 0$$

Note that the use of l'Hospital's Rule is justified because $1 - \sin x \to 0$ and $\cos x \to 0$ as $x \to \frac{\pi}{2}^-$.

Indeterminate Powers

Several indeterminate forms arise from the limit

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

1. $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$ type 0^0
2. $\lim_{x\to a} f(x) = \infty$ and $\lim_{x\to a} g(x) = 0$ type ∞^0
3. $\lim_{x\to a} f(x) = 1$ and $\lim_{x\to a} g(x) = \infty$ type 1^∞

Each of these three cases can be treated either by taking the natural logarithm:
Let $y = [f(x)]^{g(x)}$, then $\ln y = g(x)\ln f(x)$
or by writing the function as an exponential:

$$y = [f(x)]^{g(x)} = e^{g(x)\ln f(x)}$$

(Recall that both of these methods were used in differentiating such functions.) In either method we are led to the indeterminate product $g(x)\ln f(x)$, which is of type $0\cdot\infty$.

Example 5.9. *Calculate* $\lim_{x\to 0+}(1+\sin 4x)^{\cot x}$

Solution

First notice that as $x \to 0+$, we have $1+\sin 4x \to 1$ and $\cot x \to \infty$, so the given limit is indeterminate 1^{∞}.
Let $y = (1+\sin 4x)^{\cot x}$ the $\ln y = \cot x \ln(1+\sin 4x)$
So L'Hospital's Rule gives

$$\lim_{x\to 0^+} \ln y = \lim_{x\to 0^+} \cot x \ln(1+\sin 4x) = \lim_{x\to 0^+} \frac{\ln 1+\sin 4x}{\tan x} = \frac{\frac{4\cos x}{1+\sin x}}{\sec^2 x} = 4$$

So far we have computed the limit of $\ln y$, but what we want is the limit of y. To find this we use the fact that $y = e^{\ln y}$

$$\lim_{x\to 0+}(1+\sin 4x)^{\cot x} = \lim_{x\to 0+} y = \lim_{x\to 0+} e^{\ln y} = e^4$$

Example 5.10. *Find* $\lim_{x\to 0+} x^x$.

Solution

Notice that this limit is indeterminate since $0^x = 0$ for any $x > 0$ but $x^0 = 1$ for any $x > 0$. We could proceed by writing the function as an exponential:

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

we used l'Hospital's Rule to show that $\lim_{x\to 0^+}(x \ln x) = 0$
Therefore

$$\lim_{x\to 0^+} x^x = \lim_{x\to 0^+} e^{x\ln x} = e^0 = 1$$

5.5 Relative Rate

Main concepts of this section are

- Identifying word problems as related rate problems
- translate word problem into mathematical expressions.
- calculate derivative of expression with multiple variables implicity
- Understand the process of solving related rate problems

Suppose we have two variables x and y (in most problems the letters will be different, but for now let's use x and y) which are both changing with time. A "related rates" problem is a problem in which we know one of the rates of change at a given instant say, $x' == dx/dt$ and we want to find the other rate $y = dy/dt$ at that instant. (The use of x' to mean dx/dt goes back to Newton and is still used for this purpose, especially by physicists.)

If y is written in terms of x, i.e., $y = f(x)$, then this is easy to do using the chain rule:

$$y = \frac{dy}{dt} = \frac{dy}{dx} * \frac{dx}{dt}$$

That is, find the derivative of f(x), plug in the value of x at the instant in question, and multiply by the given value of $x' = dx/dt$ to get $y' = dy/dt$.

Example 5.11. *Suppose an object is moving along a path described by $y = x^2$, that is, it is moving on a parabolic path. At a particular time, say $t = 5$, the x coordinate is 6 and we measure the speed at which the x coordinate of the object is changing and find that $dx/dt = 3$. At the same time, how fast is the y coordinate changing?*

solution

Using the chain rule, $dy/dt = 2x \cdot dx/dt$. At $t = 5$ we know that $x = 6$ and $dx/dt = 3$, so $dy/dt = 2 \cdot 6 \cdot 3 = 36$.

In many cases, particularly interesting ones, x and y will be related in some other way, for example $x = f(y)$, or $F(x, y) = k$, or perhaps $F(x, y) = G(x, y)$, where $F(x, y)$ and $G(x, y)$ are expressions involving both variables. In all cases, you can solve the related rates problem by taking the derivative of both sides, plugging in all the known values (namely, x, y, and x'), and then solving for y'.

To summarize, here are the steps in doing a related rates problem:

1. Decide what the two variables are.
2. Find an equation relating them.
3. Take d/dt of both sides.
4. Plug in all known values at the instant in question.
5. Solve for the unknown rate.

Example 5.12. *A plane is flying directly away from you at 500 mph at an altitude of 3 miles. How fast is the plane's distance from you increasing at the moment when the plane is flying over a point on the ground 4 miles from you?*

solution

To see what's going on, we first draw a schematic representation of the situation, as in figure5.9. Because the plane is in level flight directly away from you, the rate at which x

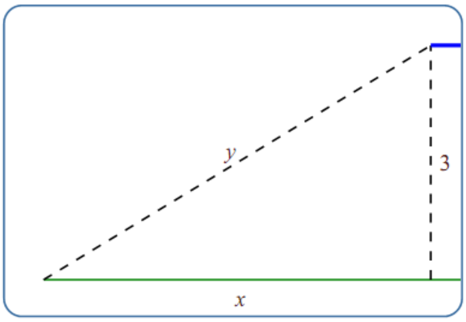


Figure 5.9: Receding airplane.

changes is the speed of the plane, $dx/dt = 500$. The distance between you and the plane is y; it is dy/dt that we wish to know. By the Pythagorean Theorem we know that x2+9=y2. Taking the derivative:

$$2x * x' = 2y * y'$$

. We are interested in the time at which $x = 4$, at this time we know that $4^2 + 9 = y^2$, so $y = 5$. Putting together all the information we get $2(4)(500) = 2(5)'y$. Thus, $y' = 400mph$

Example 5.13. *Suppose that we have two resistors connected in parallel with resistances R_1 and R_2 measured in ohms (Ω). The total resistance, R, is then given by,*

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

Suppose that R_1 is increasing at a rate of $0.4\Omega/min$ and is decreasing at a rate of $0.7\Omega/min$. At what rate is R changing when $R_1 = 80\Omega$ and $R_2 = 105\Omega$?

solution

First we look for R' and that we know $R_1' = 0.4$ and $R_2' = -0.7$.(The sign is due to decreasing behavior of R_2).
Now we evaluate R at the requiring time. Thus

$$\frac{1}{R} = \frac{1}{80} + \frac{1}{105} = \frac{37}{1680} \quad \Rightarrow R = \frac{1680}{37} = 45.4054\Omega$$

Next differentiate the given statement we gate $-\frac{1}{R^2}R' = -\frac{1}{R_1^2}R_1' - \frac{1}{R_2^2}R_2'$
$R' = R^2\left(\frac{1}{R_1^2}R_1' + \frac{1}{R_2^2}R_2'\right)$ Final we plug into this equation and to do same computation

then we obtain

$$R' = (45.4054)^2(\frac{1}{80^2}(0.4) + \frac{1}{105^2}(-0.7)) = -0.0024045$$

So r is decreasing at a rate of $0.0024045\Omega/min$

Example 5.14. *A tank of water in the shape of a cone is leaking water at a constant rate of $12ft^3/hour$. The base radius of the tank is 5 ft and the height of the tank is $14ft$.*
(a) At what rate is the depth of the water in the tank changing when the depth of the water is 6 ft?
(b) At what rate is the radius of the top of the water in the tank changing when the depth of the water is 6 ft?

5.6 Optimization

The methods we have learned in this chapter for finding extreme values have practical applications in many areas of life. A businessperson wants to minimize costs and maximize profits. A traveler wants to minimize transportation time. Fermat's Principle in optics states that light follows the path that takes the least time. In this section we solve such problems as maximizing areas, volumes, and profits and minimizing distances, times, and costs.

Steps in Solving Optimization Problems

1. **Understand the problem:** The first step is to read the problem carefully until it is clearly understood. Ask yourself: What is the unknown? What are the given quantities? What are the given conditions?

2. **Draw a Diagram:** In most problems it is useful to draw a diagram and identify the given and required quantities on the diagram.

3. **Introduce Notation:** Assign a symbol to the quantity that is to be maximized or minimized (say Q). Also select symbols $(a, b, c, \cdots, x, y)$ for other unknown quantities and label the diagram with these symbols.

4. Express Q in terms of some of the other symbols from Step 3.

5. Find the relationship between Q and the unknown quantities.

6. Find the absolute maximum or minimum value off .

Example 5.15. *Find two nonnegative numbers whose sum is 9 and so that the product of one number and the square of the other number is a maximum.*

solution

Let variables x and y represent two nonnegative numbers. The sum of the two numbers is given to be

$$9 = x + y,$$

so that

$$y = 9 - x.$$

We wish to MAXIMIZE the PRODUCT

$$P = xy^2.$$

However, before we differentiate the right-hand side, we will write it as a function of x only. Substitute for y getting

$$\begin{aligned} P &= xy^2 \\ &= x(9 - x)^2. \end{aligned}$$

Now differentiate this equation using the product rule and chain rule, getting

$$\begin{aligned} P' &= x(2)(9 - x)(-1) + (1)(9 - x)^2 \\ &= (9 - x)[-2x + (9 - x)] \\ &= (9 - x)[9 - 3x] \\ &= (9 - x)(3)[3 - x] \\ &= 0 \end{aligned}$$

for $x = 9 or x = 3$. Note that since both x and y are nonnegative numbers and their sum is 9, it follows that $0 \leq x \leq 9$. See the adjoining sign chart for P' . at p=0 and 9 the value is zero . Therefore $x = 3$ and $y = 6$ with the value P= 108 is the largest possible product.

Example 5.16. *Build a rectangular pen with three parallel partitions using 500 feet of fencing. What dimensions will maximize the total area of the pen ?*

solution

Let variable x be the width of the pen and variable y the length of the pen. The total amount of fencing is given to be

$$500 = 5(width) + 2(length) = 5x + 2y,$$

so that $42y = 500 - 5x4$ or

$$y = 250 - (5/2)x.$$

We wish to MAXIMIZE the total AREA of the pen

$$A = (width)(length) = xy.$$

However, before we differentiate the right-hand side, we will write it as a function of x only. Substitute for y getting

$$\begin{aligned} A &= xy \\ &= x(250 - (5/2)x) \\ &= 250x - (5/2)x^2. \end{aligned}$$

Now differentiate this equation, getting

$$\begin{aligned} A' &= 250 - (5/2)2x \\ &= 250 - 5x \\ &= 5(50 - x) \\ &= 0 \end{aligned}$$

for

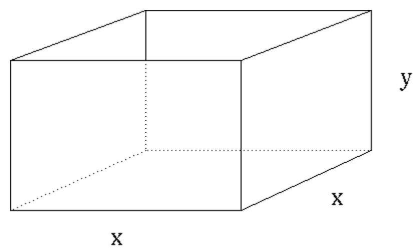
$$x = 50.$$

Note that since there are 5 lengths of x in this construction and 500 feet of fencing, it follows that $0 \leq x \leq 100$. See the adjoining sign chart for A' . If $x = 50ft.$ and $y = 125ft.$, then $A = 6250ft.2$ is the largest possible area of the pen.

Example 5.17. *An open rectangular box with square base is to be made from* $48ft.^2$ *of material. What dimensions will result in a box with the largest possible volume ?*

solution

Let variable x be the length of one edge of the square base and variable y the height of the box.



The total surface area of the box is given to be

$$48 = (areaofbase) + 4(areaofoneside) = x2 + 4(xy),$$

so that

$$4xy = 48 - x^2$$

or

$$\begin{aligned} y &= \frac{48 - x^2}{4x} \\ &= \frac{48}{4x} - \frac{x^2}{4x} \\ &= \frac{12}{x} - (1/4)x. \end{aligned}$$

We wish to MAXIMIZE the total VOLUME of the box

$$V = (length)(width)(height) = (x)(x)(y) = x2y.$$

However, before we differentiate the right-hand side, we will write it as a function of x only. Substitute for y getting

$$\begin{aligned} V &= x^2 y \\ &= x^2\Big(\frac{12}{x} - (1/4)x\Big) \\ &= 12x - (1/4)x^3. \end{aligned}$$

Now differentiate this equation, getting

$$\begin{aligned} V' = 12 - (1/4)3x^2 & \\ &= 12 - (3/4)x^2 \\ &= (3/4)(16 - x^2) \\ &= (3/4)(4 - x)(4 + x) \\ &= 0 \end{aligned}$$

for $x = 4$ or $x = -4$. But $x \neq -4$ since variable x measures a distance and $x > 0$. Since the base of the box is square and there are 48 ft.2 of material, it follows that $0 < x \leq \sqrt{48}$. See the adjoining sign chart for V' .
If $x = 4ft.$ and $y = 2ft.$, then

$$V = 32ft.^3$$

is the largest possible volume of the box.

Example 5.18. *A container in the shape of a right circular cylinder with no top has surface area $3\pi ft.^2$ What height h and base radius r will maximize the volume of the cylinder ?*

Example 5.19. *A sheet of cardboard 3 ft. by 4 ft. will be made into a box by cutting equal-sized squares from each corner and folding up the four edges. What will be the dimensions of the box with largest volume ?*

Example 5.20. *Find the point (x, y) on the graph of* $y = \sqrt{x}$ *nearest the point (4, 0).*

Example 5.21. *An open-top box is to be made by cutting small congruent squares from the corners of a 12cm. by 12cm. sheet of tin and bending up the sides. How large should the squares cut from the corners be to make the box hold as much as possible?*

Solution

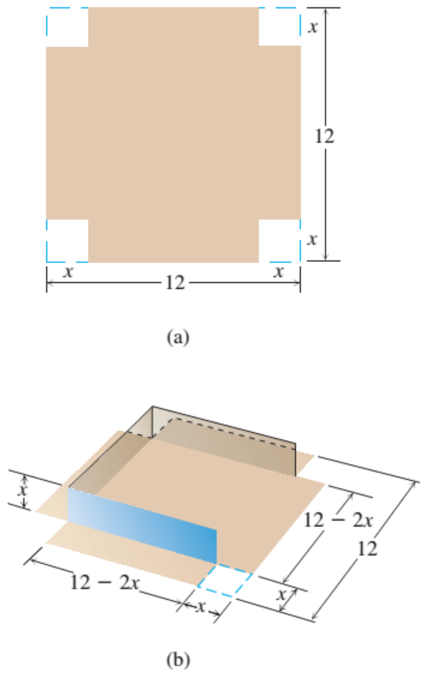


Figure 5.10: An open box made by cutting the corners from a square sheet of tin.

We start with a picture (Figure5.10). In the figure, the corner squares are x in. on a side. The volume of the box is a function of this variable:

$$V(x) = x(12 - x)^2 = 144x - 48x^2 + 4x^3 \qquad (5.2)$$

Since the sides of the sheet of tin are only 12 in. long, $x \leq 6$ and the domain of V is the interval $0 \leq x \leq 6$.

Form the equation 5.2 we have two critical numbers $x = 2$ and $x = 6$ but $x = 2$ is lies between the interval.

Now we evaluate the functional values at the critical number and the two end points and

obtain the results. $V(2) = 28$ at the critical point $V(0) = 0$ and $V(6) = 0$ at the ends of the interval.
Hence the maximum value (volume) occurred at the critical point, that is $V(2) = 28in^3$

Example 5.22. *A farmer has 2400 ft of fencing and wants to fence off a rectangular field that borders a straight river. He needs no fence along the river. What are the dimensions of the field that has the largest area?*

Example 5.23. *The manager of a department store wants to build a 600-square-foot rectangular enclosure on the store's parking lot to display some equipment. Three sides of the enclosure will be built of redwood fencing at a cost of* \$14 *per running foot. The fourth side will be built of cement blocks, at a cost of* \$28 *per running foot. Find the dimensions of the enclosure that will minimize the total cost of the building materials.*

Example 5.24. *You have been asked to design a one-liter can shaped like a right circular cylinder. What dimensions will use the least material?*

Example 5.25. *A rectangle is to be inscribed in a semicircle of radius r. What is the largest area the rectangle can have, and what are its dimensions?*

CHAPTER Five

6 Integration

In this chapter we define the integral of a function on an interval [a, b], introduce the Fundamental theorem of Calculus relating integration and differentiation, and develop basic techniques for computing integrals.

The most common interpretation of the integral is in terms of the area under the graph of the given function, so that is where we begin.

6.1 Definite integral

The rough idea of defining the area under the graph of f is to approximate this area with a finite number of rectangles. Since we can easily work out the area of the rectangles, we get an estimate of the area under the graph. If we use a larger number of smaller-sized rectangles we expect greater accuracy with respect to the area under the curve and hence a better approximation. Somehow, it seems that we could use our old friend from differentiation, the limit, and "approach" an infinite number of rectangles to get the exact area. Let's look at such an idea more closely.

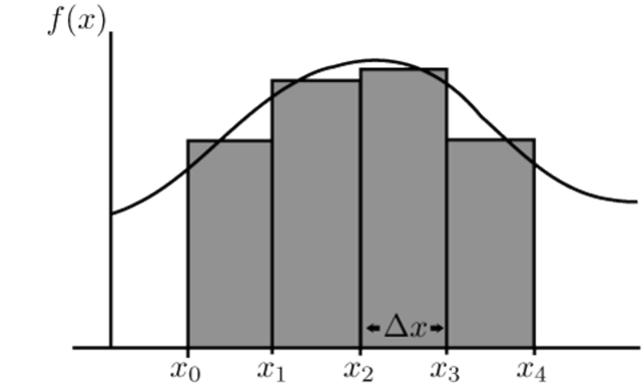


Figure 6.1: Approximation of the area under the curve $f(x)$ from $x = x_0$ to $x = x_4$.

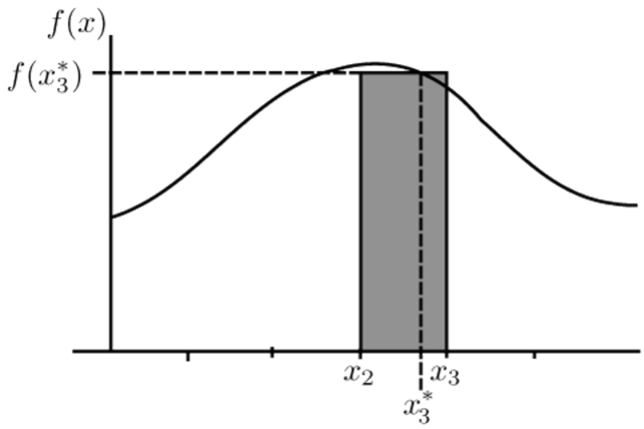


Figure 6.2: Rectangle approximating the area under the curve from x_2 to x_3 with sample point x_3^*.

Suppose we have a function f that is positive on the interval $[a, b]$ and we want to find the area S under f between a and b. Let's pick an integer n and divide the interval into n subintervals of equal width (see Figure 6.1). As the interval $[a, b]$ has width $b - a$, each subinterval has width $\Delta x = \dfrac{b-a}{n}$. We denote the endpoints of the subintervals by $x_0, x_1, \ldots, x_n$. This gives us $x_i = a + i\Delta x$ for $i = 0, 1, \ldots, n$

Now for each $i = 1, \ldots, n$ pick a sample point x_i^* in the interval $[x_{i-1}, x_i]$ and consider the rectangle of height $f(x_i^*)$ and width Δx(see Figure 6.2). The area of this rectangle is $f(x_i^*)\Delta x$. By adding up the area of all the rectangles for $i = 1, \ldots, n$ we get that the area S is approximated by

$$A_n = f(x_1^*)\Delta x + \cdots + f(x_n^*)\Delta x$$

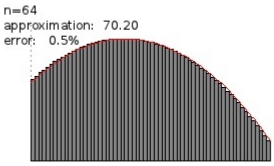


Figure 6.3: Riemann sums with an increasing number of subdivisions yielding better approximations.

A more convenient way to write this is with summation notation:

$$A_n = \sum_{i=1}^{n} f(x_i^*)\Delta x$$

For each number n we get a different approximation. As n gets larger the width of the rectangles gets smaller which yields a better approximation (see Figure 6.3). In the limit of A_n as n tends to infinity we get the area S.

Definition of the Definite Integral

Definition 6.1. *Suppose f is a continuous function on $[a, b]$ and $\Delta x = \dfrac{b-a}{n}$. Then the definite integral of f between a and b is*

$$\int_a^b f(x)dx = \lim_{n\to\infty} A_n = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

where x_i^* are any sample points in the interval $[x_{i-1}, x_i]$ and $x_k = a + k \cdot \Delta x$.

It is a fact that if f is continuous on $[a, b]$ then this limit always exists and does not depend on the choice of the points $x_i^* \in [x_{i-1}, x_i]$. For instance they may be evenly spaced, or distributed ambiguously throughout the interval. The proof of this is technical and is beyond the scope of this section.

Notation

When considering the expression, $\int_a^b f(x)dx$ (read "the integral from a a to b of the f of x dx”). The symbol ” $\int$ ” is an integral sign. The function f is called the integrand and and x is the variable of integration. The interval $[a, b]$ is the interval of integration and also a is called the lower limit and b the upper limit of integration. One important feature of this definition is that we also allow functions which take negative values. If $f(x) < 0$ for all x then $f(x_i^*) < 0$ so $f(x_i^*)\Delta x < 0$. So the definite integral of f will be strictly negative. More generally if f takes on both positive and negative values then $\int_a^b f(x)dx$ will be the area under the positive part of the graph of f minus the area above the graph of the negative

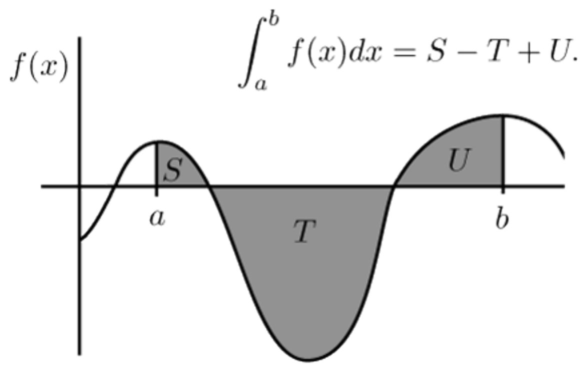


Figure 6.4: The integral gives the signed area under the graph.

part of the graph (see Figure 6.4). For this reason we say that $\int\limits_a^b f(x)dx$ is the signed area under the graph.

Independence of Variable

It is important to notice that the variable x did not play an important role in the definition of the integral. In fact we can replace it with any other letter, so the following are all equal:

$$\int\limits_a^b f(x)dx = \int\limits_a^b f(t)dt = \int\limits_a^b f(u)du = \int\limits_a^b f(w)dw$$

Each of these is the signed area under the graph of f between a and b. Such a variable is often referred to as a dummy variable or a bound variable.

Left and Right Handed Riemann Sums

The following methods are sometimes referred to as L-RAM and R-RAM, RAM standing for "Rectangular Approximation Method."

We could have decided to choose all our sample points x_i^* to be on the right hand side of the interval $[x_{i-1}, x_i]$ (see Figure 6.5). Then $x_i^* = x_i$ for all i and the approximation that we called A_n for the area becomes

$$A_n = \sum_{i=1}^{n} f(x_i)\Delta x$$

This is called the right-handed Riemann sum, and the integral is the limit

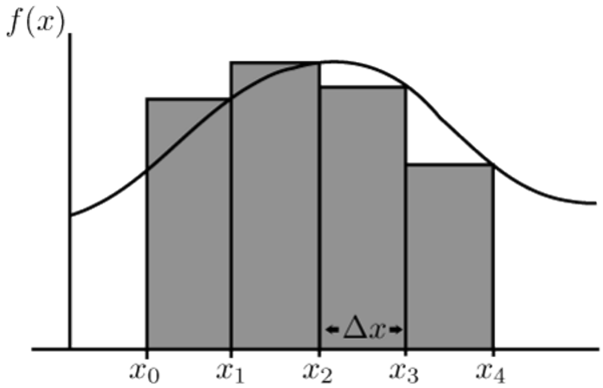


Figure 6.5: Right-handed Riemann sum

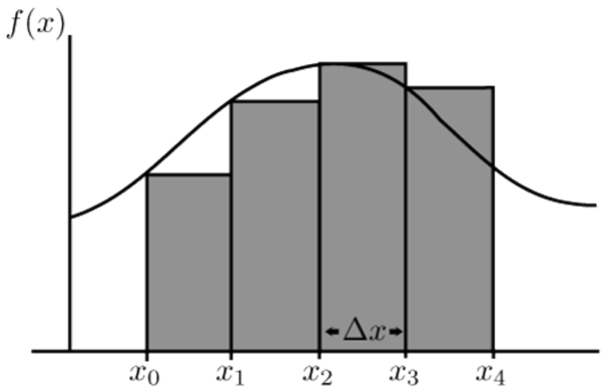


Figure 6.6: Left-handed Riemann sum

$$\int_a^b f(x)dx = \lim_{n\to\infty} A_n = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x$$

Alternatively we could have taken each sample point on the left hand side of the interval. In this case $x_i^* = x_{i-1}$ (see Figure 6.6) and the approximation becomes

$$A_n = \sum_{i=1}^{n} f(x_{i-1})\Delta x$$

Then the integral of f is

$$\int_a^b f(x)dx = \lim_{n\to\infty} A_n = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_{i-1})\Delta x$$

The key point is that, as long as f is continuous, these two definitions give the same answer for the integral.

Example 6.1. *In this example we will calculate the area under the curve given by the graph of $f(x) = x$ for x between 0 and 1.*

First we fix an integer n and divide the interval $[0, 1]$ into n subintervals of equal width. So each subinterval has width

$$\Delta x = \frac{1}{n}$$

To calculate the integral we will use the right-handed Riemann sum. (We could have used the left-handed sum instead, and this would give the same answer in the end). For the right-handed sum the sample points are

$$x_i^* = 0 + i\Delta x = \frac{i}{n} \quad i = 1, \ldots, n$$

Notice that $f(x_i^*) = x_i^* = \frac{i}{n}$. Putting this into the formula for the approximation,

$$A_n = \sum_{i=1}^{n} f(x_i^*)\Delta x = \sum_{i=1}^{n} f\left(\frac{i}{n}\right)\Delta x = \sum_{i=1}^{n} \frac{i}{n} \cdot \frac{1}{n} = \frac{1}{n^2}\sum_{i=1}^{n} i$$

Now we use the formula

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

to get

$$A_n = \frac{1}{n^2} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2n}$$

To calculate the integral of f between 0 and 1 we take the limit as n tends to infinity,

$$\int_0^1 f(x)dx = \lim_{n\to\infty} \frac{n+1}{2n} = \frac{1}{2}$$

Example 6.2. *Next we show how to find the integral of the function* $f(x) = x^2$ *between* $x = a$ *and* $x = b$.

This time the interval $[a, b]$ has width $b - a$ so

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

Once again we will use the right-handed Riemann sum. So the sample points we choose are

$$x_i^* = a + i\Delta x = a + \frac{i(b-a)}{n}$$

Thus

$$\begin{aligned} A_n &= \sum_{i=1}^{n} f(x_i^*)\Delta x \\ &= \sum_{i=1}^{n} f\left(a + \frac{(b-a)i}{n}\right)\Delta x \\ &= \frac{b-a}{n}\sum_{i=1}^{n}\left(a + \frac{(b-a)i}{n}\right)^2 \\ &= \frac{b-a}{n}\sum_{i=1}^{n}\left(a^2 + \frac{2a(b-a)i}{n} + \frac{(b-a)^2 i^2}{n^2}\right) \end{aligned}$$

We have to calculate each piece on the right hand side of this equation. For the first two,

$$\sum_{i=1}^{n} a^2 = a^2\sum_{i=1}^{n} 1 = na^2$$
$$\sum_{i=1}^{n}\frac{2a(b-a)i}{n} = \frac{2a(b-a)}{n}\sum_{i=1}^{n} i = \frac{2a(b-a)}{n}\cdot\frac{n(n+1)}{2}$$

For the third sum we have to use a formula

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

to get

$$\sum_{i=1}^{n}\frac{(b-a)^2 i^2}{n^2} = \frac{(b-a)^2}{n^2}\frac{n(n+1)(2n+1)}{6}$$

Putting this together

$$A_n = \frac{b-a}{n}\left(na^2 + \frac{2a(b-a)}{n}\cdot\frac{n(n+1)}{2} + \frac{(b-a)^2}{n^2}\frac{n(n+1)(2n+1)}{6}\right)$$

Taking the limit as n tend to infinity gives

$$\begin{aligned}\int_a^b x^2dx &= (b-a)\left(a^2+a(b-a)+\frac{1}{3}(b-a)^2\right)\\ &= (b-a)\left(a^2+ab-a^2+\frac{1}{3}(b^2-2ab+a^2)\right)\\ &= \frac{1}{3}(b-a)(b^2+ab+a^2)\\ &= \frac{1}{3}(b^3-a^3)\end{aligned}$$

Exercises

1. Use left- and right-handed Riemann sums with 5 subdivisions to get lower and upper bounds on the area under the function $f(x)=x^6$ from $x=0$ to $x=1$.

   Solution

i	x_i	x_i^6	$0.2\times x_{i-1}^6$	$\sum_{k=1}^{i} 0.2\times x_{i-1}^6$	$0.2\times x_i^6$	$\sum_{k=1}^{i} 0.2\times x_i^6$
0	0.0	0			0	
1	0.2	0.000064	0	0	0.0000128	0.0000128
2	0.4	0.004096	0.0000128	0.0000128	0.0008192	0.000832
3	0.6	0.046656	0.0008192	0.000832	0.0093312	0.0101632
4	0.8	0.262144	0.0093312	0.0101632	0.0524288	0.062592
5	1.0	1	0.0524288	0.062592	.2	0.262592

   Lower bound: 0.062592

   Upper bound: 0.262592

2. Use left- and right-handed Riemann sums with 5 subdivisions to get lower and upper bounds on the area under the function $f(x)=x^6$ from $x=1$ to $x=2$

   Solutions

i	x_i	x_i^6	$0.2\times x_{i-1}^6$	$\sum_{k=1}^{i} 0.2\times x_{i-1}^6$	$0.2\times x_i^6$	$\sum_{k=1}^{i} 0.2\times x_i^6$
0	1.0	1			.2	
1	1.2	2.985984	.2	.2	0.5971968	0.5971968
2	1.4	7.529536	0.5971968	0.7971968	1.5059072	2.103104
3	1.6	16.777216	1.5059072	2.303104	3.3554432	5.4585472
4	1.8	34.012224	3.3554432	5.6585472	6.8024448	12.260992
5	2.0	64	6.8024448	12.460992	12.8	25.060992

Lower bound: 12.460992

Upper bound: 25.060992

6.1.1 Basic Properties of the Integral

From the definition of the integral we can deduce some basic properties. For all the following rules, suppose that f and g are continuous on $[a, b]$.

The Constant Rule

$$\int_a^b c \cdot f(x)dx = c \int_a^b f(x)dx$$

When f is positive, the height of the function $c \cdot f$ at a point x is c times the height of the function f. So the area under $c \cdot f$ between a and b is c times the area under f. We can also give a proof using the definition of the integral, using the constant rule for limits,

$$\int_a^b c \cdot f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} c \cdot f(x_i^*) = c \cdot \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*) = c \int_a^b f(x)dx$$

Example 6.3. *We saw in the previous section that*

$$\int_0^1 xdx = \frac{1}{2}$$

Using the constant rule we can use this to calculate that

$$\int_0^1 3xdx = 3\int_0^1 xdx = 3 \cdot \frac{1}{2} = 1.5,$$

$$\int_0^1 -7xdx = -7\int_0^1 xdx = (-7) \cdot \frac{1}{2} = -3.5.$$

Example 6.4. *We saw in the previous section that*

$$\int_a^b x^2 dx = \frac{b^3 - a^3}{3}$$

We can use this and the constant rule to calculate that

$$\int_1^3 2x^2 dx = 2\int_1^3 x^2 dx = 2 \cdot \frac{1}{3} \cdot (3^3 - 1^3) = \frac{2}{3}(27 - 1) = \frac{52}{3}$$

There is a special case of this rule used for integrating constants:

Integrating Constants

If c is constant then $\int_a^b c\,dx = c(b-a)$ When$c > 0$ and $a < b$ this integral is the area of a rectangle of height c and width $b-a$ which equals $c(b-a)$.

Example 6.5.

$$\int_1^3 9dx = 9(3-1) = 9 \cdot 2 = 18$$
$$\int_{-2}^6 11dx = 11(6-(-2)) = 11 \cdot 8 = 88$$
$$\int_2^{17} 0dx = 0 \cdot (17-2) = 0$$

The addition and subtraction rule

$$\int_a^b \Big(f(x)+g(x)\Big)dx = \int_a^b f(x)dx + \int_a^b g(x)dx$$
$$\int_a^b \Big(f(x)-g(x)\Big)dx = \int_a^b f(x)dx - \int_a^b g(x)dx$$

As with the constant rule, the addition rule follows from the addition rule for limits:

$$\begin{aligned}\int_a^b \Big(f(x)+g(x)\Big)dx &= \lim_{n\to\infty}\sum_{i=1}^n \Big(f(x_i^*)+g(x_i^*)\Big)\\ &= \lim_{n\to\infty}\sum_{i=1}^n f(x_i^*) + \lim_{n\to\infty}\sum_{i=1}^n g(x_i^*)\\ &= \int_a^b f(x)dx + \int_a^b g(x)dx\end{aligned}$$

The subtraction rule can be proved in a similar way.

Example 6.6. *From above* $\int_1^3 9dx = 18$ *and* $\int_1^3 2x^2dx = \frac{52}{3}$ *so*

$$\int_1^3 (2x^2+9)dx = \int_1^3 2x^2dx + \int_1^3 9dx = \frac{52}{3} + 18 = \frac{106}{3}$$
$$\int_1^3 (2x^2-9)dx = \int_1^3 2x^2dx - \int_1^3 9dx = \frac{52}{3} - 18 = -\frac{2}{3}$$

Example 6.7.

$$\int_0^2 (4x^2+14)dx = 4\int_0^2 x^2dx + \int_0^2 14dx = 4\cdot\frac{1}{3}(2^3-0^3)+2\cdot 14 = \frac{32}{3}+28 = \frac{116}{3}$$

Exercise

1. Use the subtraction rule to find the area between the graphs of $f(x) = x$ and $g(x) = x^2$ between $x = 0$ and $x = 1$

The Comparison Rule

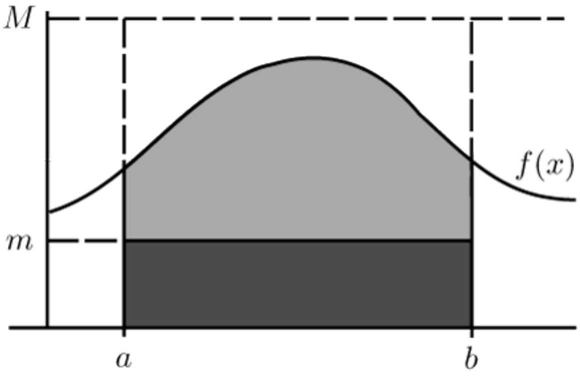


Figure 6.7: Bounding the area under $f(x)$ on $[a, b]$ Comparison Rule

Suppose $f(x) \geq 0$ for all $x \in [a, b]$. Then $\int_a^b f(x)dx \geq 0$ Suppose $f(x) \geq g(x)$ for all $x \in [a, b]$. Then $\int_a^b f(x)dx \geq \int_a^b g(x)dx$ Suppose $M \geq f(x) \geq m$. Then $M(b-a) \geq \int_a^b f(x)dx \geq m(b-a)$

If $f(x) \geq 0$ then each of the rectangles in the Riemann sum to calculate the integral of f will be above the y axis, so the area will be non-negative. If $f(x) \geq g(x)$ then $f(x) - g(x) \geq 0$ and by the first property we get the second property. Finally if $M \geq f(x) \geq m$ then the area under the graph of f will be greater than the area of rectangle with height m and less than the area of the rectangle with height M (see Figure 6.7). So

$$M(b-a) = \int_a^b M \geq \int_a^b f(x)dx \geq \int_a^b m = m(b-a)$$

Linearity with respect to endpoints

Additivity with respect to endpoints Suppose $a < c < b$. Then

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

Again suppose that f is positive. Then this property should be interpreted as saying that the area under the graph of f between a and b is the area between a and c plus the area between c and b (see Figure 6.8).

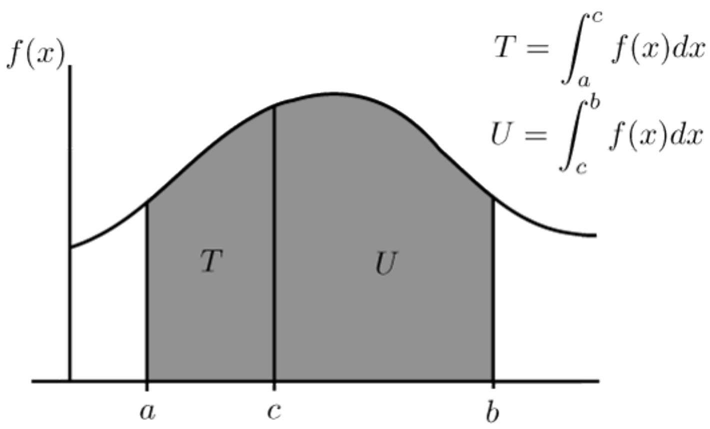


Figure 6.8: Illustration of the property of additivity with respect to endpoints

Extension of Additivity with respect to limits of integration When $a = b$ we have that $\Delta x = \dfrac{b-a}{n} = 0$ so

$$\int_a^a f(x)dx = 0$$

Also in defining the integral we assumed that $a < b$. But the definition makes sense even when $b < a$, in which case $\Delta x = \dfrac{b-a}{n}$ has changed sign. This gives

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

With these definitions,

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

whatever the order of a, b, c.

Exercise

1. Use the results of exercises 1 and 2 and the property of linearity with respect to endpoints to determine upper and lower bounds on $\int_0^2 x^6 dx$

Even and odd functions

Recall that a function f is called odd if it satisfies $f(-x) = -f(x)$ and is called even if $f(-x) = f(x)$.

Suppose f is a continuous odd function then for any a ,

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

If f is a continuous even function then for any a,

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

Suppose f is an odd function and consider first just the integral from $-a$ to 0. We make the substitution $u = -x$ so $du = -dx$. Notice that if $x = -a$ then $u = a$ and if $x = 0$ then $u = 0$. Hence

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

Now as f is odd, $f(-u) = -f(u)$ so the integral becomes

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

. Now we can replace the dummy variable u with any other variable. So we can replace it with the letter x to give

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

. Now we split the integral into two pieces

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

. The proof of the formula for even functions is similar.

1. Prove that if f is a continuous even function then for any a,

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

.

6.1.2 Fundamental Theorem of Calculus

Antiderivatives

Definition 6.2. *A function F is an antiderivative of f on an interval I if $F'(x) = f(x)$ for all in I.*

Theorem 6.1. *If F is an antiderivative of f on an interval I then G is an antiderivative of f on the interval I if and only if G is of the form $G(x(= F(x) + C$, for all x in I where C is a constant.*

Mean Value Theorem for Integration

We will need the following theorem in the discussion of the Fundamental Theorem of Calculus.

Suppose $f(x)$ is continuous on $[a, b]$. Then $\frac{1}{b-a}\int_a^b f(x)dx = f(c)$ for some c in $[a, b]$.

Proof of the Mean Value Theorem for Integration

$f(x)$ satisfies the requirements of the Extreme Value Theorem, so it has a minimum m and a maximum M in $[a, b]$. Since

$$\int_a^b f(x)dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\frac{b-a}{n} = \lim_{n\to\infty}\frac{b-a}{n}\sum_{i=1}^{n} f(x_i^*)$$

and since $m \leq f(x_i^*) \leq M$ for all x_i^* in $[a, b]$, we have

$$\lim_{n\to\infty}\frac{b-a}{n}\sum_{i=1}^{n} m \leq \lim_{n\to\infty}\frac{b-a}{n}\sum_{i=1}^{n} f(x_i^*) \leq \lim_{n\to\infty}\frac{b-a}{n}\sum_{i=1}^{n} M$$

$$\lim_{n\to\infty}\frac{b-a}{n}nm \leq \int_a^b f(x)dx \leq \lim_{n\to\infty}\frac{b-a}{n}nM$$

$$\lim_{n\to\infty}(b-a)m \leq \int_a^b f(x)dx \leq \lim_{n\to\infty}(b-a)M$$

$$(b-a)m \leq \int_a^b f(x)dx \leq (b-a)M$$

$$m \leq \frac{1}{b-a}\int_a^b f(x)dx \leq M$$

Since f is continuous, by the Intermediate Value Theorem there is some $f(c)$ with c in $[a, b]$ such that

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

Fundamental Theorem of Calculus

Theorem 6.2. ***Fundamental Theorem of Calculus***
Suppose that f is continuous on $[a, b]$. We can define a function F by

$$F(x) = \int_a^x f(t)dt \text{ for } x \text{ in } [a, b]$$

Fundamental Theorem of Calculus Part I
Suppose f is continuous on $[a, b]$ and F is defined by

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

Then F is differentiable on (a, b) and for all $x \in (a, b)$,

$$F'(x) = f(x)$$

When we have such functions F and f where $F'(x) = f(x)$ for every x in some interval I we say that F is the antiderivative of f on I.

Theorem 6.3. ***Fundamental Theorem of Calculus Part II***
Suppose that f is continuous on $[a, b]$ and that F is any antiderivative of f. Then

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

Note: a minority of mathematicians refer to part one as two and part two as one. All mathematicians refer to what is stated here as part 2 as The Fundamental Theorem of Calculus.

Proof. **Proof of Fundamental Theorem of Calculus Part I**
Suppose x is in (a, b). Pick Δx so that $x + \Delta x$ is also in (a, b). Then

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

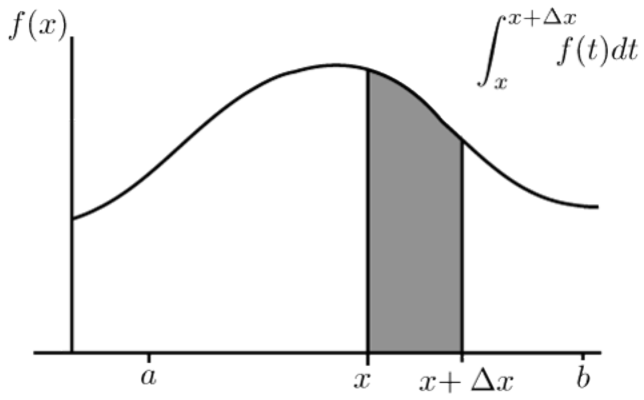


Figure 6.9

and

$$F(x+\Delta x) = \int_a^{x+\Delta x} f(t)dt.$$

Subtracting the two equations gives

$$F(x+\Delta x) - F(x) = \int_a^{x+\Delta x} f(t)dt - \int_a^{x} f(t)dt.$$

Now

$$\int_a^{x+\Delta x} f(t)dt = \int_a^{x} f(t)dt + \int_x^{x+\Delta x} f(t)dt$$

so rearranging this we have

$$F(x+\Delta x) - F(x) = \int_x^{x+\Delta x} f(t)dt$$

According to the Mean Value Theorem for Integration, there exists a $c \in [x, x+\Delta x]$ such that

$$\int_x^{x+\Delta x} f(t)dt = f(c)\Delta x.$$

Notice that c depends on Δx. Anyway what we have shown is that

$$F(x+\Delta x) - F(x) = f(c)\Delta x$$

, and dividing both sides by Δx gives

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

Take the limit as $\Delta x \to 0$ we get the definition of the derivative of F at x so we have

$$F'(x) = \lim_{\Delta x \to 0} f(c).$$

To find the other limit, we will use the squeeze theorem. The number c is in the interval $[x, x + \Delta x]$, so $x \leq c \leq x + \Delta x$. Also, $\lim_{\Delta x \to 0} x = x$ and $\lim_{\Delta x \to 0} \left[x + \Delta x \right] = x$. Therefore, according to the squeeze theorem,

$$\lim_{\Delta x \to 0} c = x.$$

As f is continuous we have

$$F'(x) = \lim_{\Delta x \to 0} f(c) = f\left(\lim_{\Delta x \to 0} c \right) = f(x)$$

which completes the proof.

Proof of Fundamental Theorem of Calculus Part II

Define $P(x) = \int\limits_a^x f(t)dt$. Then by the Fundamental Theorem of Calculus part I we know that P is differentiable on (a, b) and for all $x \in (a, b)$

$$P'(x) = f(x)$$

So P is an antiderivative of f. Since we were assuming that F was also an antiderivative for all $x \in (a, b)$,

$$\begin{aligned} P'(x) &= F'(x) \\ P'(x) - F'(x) &= 0 \\ \Big(P(x) - F(x) \Big)' &= 0 \end{aligned}$$

Let $g(x) = P(x) - F(x)$. The Mean Value Theorem applied to $g(x)$ on $[a, \xi]$ with $a < \xi < b$ says that

$$\frac{g(\xi) - g(a)}{\xi - a} = g'(c)$$

for some c in (a, ξ). But since $g'(x) = 0$ for all $x$in $[a, b]$, $g(\xi)$ must equal $g(a)$ for all ξ in (a, b), i.e. $g(x)$ is constant on (a, b).

This implies there is a constant $C = g(a) = P(a) - F(a) = -F(a)$ such that for all $x \in (a, b)$,

$$P(x) = F(x) + C,$$

and as g is continuous we see this holds when $x = a$ and $x = b$ as well. And putting $x = b$ gives

$$\int\limits_a^b f(t)dx = P(b) = F(b) + C = F(b) - F(a).$$

$\square$

Notation for Evaluating Definite Integrals

The second part of the Fundamental Theorem of Calculus gives us a way to calculate definite integrals. Just find an antiderivative of the integrand, and subtract the value of the antiderivative at the lower bound from the value of the antiderivative at the upper bound. That is

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

where $F'(x) = f(x)$. As a convenience, we use the notation

$$F(x)\Big|_a^b$$

to represent $F(b) - F(a)$.

Integration of Polynomials

Using the power rule for differentiation we can find a formula for the integral of a power using the Fundamental Theorem of Calculus. Let $f(x) = x^n$. We want to find an antiderivative for f. Since the differentiation rule for powers lowers the power by 1 we have that

$$\frac{d}{dx}x^{n+1} = (n+1)x^n$$

As long as $n + 1 \neq 0$ we can divide by $n + 1$ to get

$$\frac{d}{dx}\left(\frac{x^{n+1}}{n+1}\right) = x^n = f(x)$$

So the function $F(x) = \dfrac{x^{n+1}}{n+1}$ is an antiderivative of f. If 0 is not in $[a, b]$ then F is continuous on $[a, b]$ and, by applying the Fundamental Theorem of Calculus, we can calculate the integral of f to get the following rule.

Power Rule of Integration I

$\displaystyle\int_a^b x^n dx = \frac{x^{n+1}}{n+1}\Big|_a^b = \frac{b^{n+1} - a^{n+1}}{n+1}$ as long as $n \neq -1$ and 0 is not in $[a, b]$. Notice that we allow all values of n, even negative or fractional. If $n > 0$ then this works even if $[a, b]$ includes 0.

Power Rule of Integration II

$\displaystyle\int_a^b x^n dx = \frac{x^{n+1}}{n+1}\Big|_a^b = \frac{b^{n+1} - a^{n+1}}{n+1}$ as long as $n > 0$.

Example 6.8. *1. To find* $\int\limits_1^2 x^3dx$ *we raise the power by 1 and have to divide by 4. So*

$$\int\limits_1^2 x^3dx = \left.\frac{x^4}{4}\right|_1^2 = \frac{2^4}{4} - \frac{1^4}{4} = \frac{15}{4}.$$

2. The power rule also works for negative powers. For instance

$$\int\limits_1^3 \frac{dx}{x^3} = \int\limits_1^3 x^{-3}dx = \left.\frac{x^{-2}}{-2}\right|_1^3 = \frac{1}{-2}\left(3^{-2} - 1^{-2}\right) = -\frac{1}{2}\left(\frac{1}{3^2} - 1\right) = -\frac{1}{2}\left(\frac{1}{9} - 1\right) = \frac{1}{2}\cdot\frac{8}{9} = \frac{4}{9}.$$

3. We can also use the power rule for fractional powers. For instance

$$\int\limits_0^5 \sqrt{x}dx = \int\limits_0^5 x^{\frac{1}{2}}dx = \left.\frac{x\sqrt{x}}{\frac{3}{2}}\right|_0^5 = \frac{2}{3}\left(5^{\frac{3}{2}} - 0^{\frac{3}{2}}\right) = \frac{10\sqrt{5}}{3}$$

4. Using linearity the power rule can also be thought of as applying to constants. For example,

$$= \int\limits_3^{11} 7dx = \int\limits_3^{11} 7x^0dx = 7\int\limits_3^{11} x^0dx = 7x\Big|_3^{11} = 7(11-3) = 56.$$

5. Using the linearity rule we can now integrate any polynomial. For example

$$\int\limits_0^3 (3x^2 + 4x + 2)dx = (x^3 + 2x^2 + 2x)\Big|_0^3 = 3^3 + 2\cdot 3^2 + 2\cdot 3 - 0 = 27 + 18 + 6 = 51$$

Exercises

1. Evaluate $\int\limits_0^1 x^6dx.$

$$\int_0^1 x^6dx = \left.\frac{x^7}{7}\right|_0^1 = \frac{1^7}{7} - \frac{0^7}{7} = \mathbf{\frac{1}{7}} = \mathbf{0.\overline{142857}}$$

2. Evaluate $\int\limits_1^2 x^6dx$

$$\int_1^2 x^6dx = \left.\frac{x^7}{7}\right|_1^2 = \frac{2^7}{7} - \frac{1^7}{7} = \frac{128}{7} - \frac{1}{7} = \mathbf{\frac{127}{7}} = \mathbf{18.\overline{142857}}$$

3. Evaluate $\int\limits_0^2 x^6dx$

$$\int_0^2 x^6dx = \left.\frac{x^7}{7}\right|_0^2 = \frac{2^7}{7} - \frac{0^7}{7} = \mathbf{\frac{128}{7}} = \mathbf{18.\overline{285714}}$$

6.2 Indefinite integral

Now recall that F is said to be an antiderivative of f if $F'(x) = f(x)$. However, F is not the only antiderivative. We can add any constant to F without changing the derivative. With this, we define the indefinite integral as follows: $\int f(x)dx = F(x) + C$ where F satisfies $F'(x) = f(x)$ and C is any constant. The function $f(x)$, the function being integrated, is known as the integrand. Note that the indefinite integral yields a family of functions or the sets of all antiderivatives.

Example 6.9. *Since the derivative of x^4 is $4x^3$, the general antiderivative of $4x^3$ is x^4 plus a constant. Thus,*

$$\int 4x^3 dx = x^4 + C$$

Example 6.10. *Let's take a look at $6x^2$. How would we go about finding the integral of this function? Recall the rule from differentiation that*

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

In our circumstance, we have:

$$\frac{d}{dx}x^3 = 3x^2$$

This is a start! We now know that the function we seek will have a power of 3 in it. How would we get the constant of 6? Well,

$$2\frac{d}{dx}x^3 = 2 \times 3x^2 = 6x^2$$

Thus, we say that $2x^3$ is an antiderivative of $6x^2$.

Exercises

1. Evaluate $\int \frac{3x}{2} dx$

$$\frac{3}{4}x^2 + C$$

2. Find the general antiderivative of the function $f(x) = 2x^4$

   $\frac{2x^5}{5} + C$

Indefinite integral identities

Constant Rule for indefinite integrals

If c is a constant then $\int c \cdot f(x)dx = c \int f(x)dx$

Sum/Difference Rule for indefinite integrals

$$\int \Big(f(x) + g(x)\Big)dx = \int f(x)dx + \int g(x)dx$$
$$\int \Big(f(x) - g(x)\Big)dx = \int f(x)dx - \int g(x)dx$$

Indefinite integrals of Polynomials

Say we are given a function of the form, $f(x) = x^n$, and would like to determine the antiderivative of f . Considering that

$$\frac{d}{dx}\frac{1}{n+1}x^{n+1} = x^n$$

we have the following rule for indefinite integrals:

Power rule for indefinite integrals
$\int x^n dx = \dfrac{x^{n+1}}{n+1} + C$ for all $n \neq -1$

Integral of the Inverse function

To integrate $f(x) = \dfrac{1}{x}$, we should first remember

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

Therefore, since $\dfrac{1}{x}$ is the derivative of $\ln(x)$ we can conclude that

$$\int \frac{dx}{x} = \ln|x| + C$$

Note that the polynomial integration rule does not apply when the exponent is -1. This technique of integration must be used instead. Since the argument of the natural logarithm function must be positive (on the real line), the absolute value signs are added around its argument to ensure that the argument is positive.

Integral of the Exponential function
Since

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

we see that e^x is its own antiderivative. This allows us to find the integral of an exponential function:

$$\int e^x dx = e^x + C$$

Integral of Sine and Cosine

Recall that

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

So $\sin(x)$ is an antiderivative of $\cos(x)$ and $-\cos(x)$ is an antiderivative of $\sin(x)$. Hence we get the following rules for integrating $\sin(x)$ and $\cos(x)$

$$\int \cos(x)dx = \sin(x) + C$$
$$\int \sin(x)dx = -\cos(x) + C$$

We will find how to integrate more complicated trigonometric functions in the chapter on integration techniques.

Example 6.11. *Suppose we want to integrate the function* $f(x) = x^4 + 1 + 2\sin(x)$. *An application of the sum rule from above allows us to use the power rule and our rule for integrating* $\sin(x)$ *as follows,*

$$\begin{aligned}\int f(x)dx &= \int \Big(x^4 + 1 + 2\sin(x)\Big)dx \\ &= \int x^4 dx + \int 1\,dx + \int 2\sin(x)dx \\ &= \frac{x^5}{5} + x - 2\cos(x) + C.\end{aligned}$$

Exercises

1. Evaluate $\int (7x^2 + 3\cos(x) - e^x)dx$

$$\begin{aligned}\int (7x^2 + 3\cos(x) - e^x)dx &= 7\int x^2 dx + 3\int \cos(x)dx - \int e^x dx \\ &= 7(\frac{x^3}{3}) + 3\sin(x) - e^x + C \\ &= \mathbf{\frac{7}{3}x^3 + 3\sin(x) - e^x + C}\end{aligned}$$

2. Evaluate $\int \left(\frac{2}{5x} + \sin(x)\right)dx$

$$\begin{aligned}\int (\frac{2}{5x} + \sin(x))dx &= \frac{2}{5}\int \frac{dx}{x} + \int \sin(x)dx \\ &= \mathbf{\frac{2}{5}\ln|x| - \cos(x) + C}\end{aligned}$$

7 Integration techniques

7.1 Infinite Sums

The most basic, and arguably the most difficult, type of evaluation is to use the formal definition of a Riemann integral.

Exact Integrals as Limits of Sums

Using the definition of an integral, we can evaluate the limit as n goes to infinity. This technique requires a fairly high degree of familiarity with summation identities. This technique is often referred to as evaluation "by definition," and can be used to find definite integrals, as long as the integrands are fairly simple. We start with definition of the integral:

$$\int_a^b f(x)dx = \lim_{n\to\infty}\left[\frac{b-a}{n}\sum_{k=1}^{n} f(x_k^*)\right].$$

Then picking x_k^* to be $x_k = a + k\dfrac{b-a}{n}$ we get,

$$= \lim_{n\to\infty}\left[\frac{b-a}{n}\sum_{k=1}^{n} f\Big(a + k\tfrac{b-a}{n}\Big)\right]$$

In some simple cases, this expression can be reduced to a real number, which can be interpreted as the area under the curve if $f(x)$ is positive on $[a, b]$.

Example 7.1. *Find* $\int\limits_0^2 x^2dx$ *by writing the integral as a limit of Riemann sums.*

$$\begin{aligned}\int\limits_0^2 x^2dx &= \lim_{n\to\infty}\left[\frac{b-a}{n}\sum_{k=1}^{n} f(x_k^*)\right]\\ &= \lim_{n\to\infty}\left[\frac{2}{n}\sum_{k=1}^{n} f\left(\tfrac{2k}{n}\right)\right]\\ &= \lim_{n\to\infty}\left[\frac{2}{n}\sum_{k=1}^{n}\left(\frac{2k}{n}\right)^2\right]\\ &= \lim_{n\to\infty}\left[\frac{2}{n}\sum_{k=1}^{n}\frac{4k^2}{n^2}\right]\\ &= \lim_{n\to\infty}\left[\frac{8}{n^3}\sum_{k=1}^{n} k^2\right]\\ &= \lim_{n\to\infty}\left[\frac{8}{n^3}\cdot\frac{n(n+1)(2n+1)}{6}\right]\\ &= \lim_{n\to\infty}\left[\frac{4}{3}\cdot\frac{2n^2+3n+1}{n^2}\right]\\ &= \lim_{n\to\infty}\left[\frac{8}{3}+\frac{4}{n}+\frac{4}{3n^2}\right]\\ &= \frac{8}{3}\end{aligned}$$

In other cases, it is even possible to evaluate indefinite integrals using the formal definition. We can define the indefinite integral as follows:

$$\begin{aligned}\int f(x)dx &= \int\limits_0^x f(t)dt = \lim_{n\to\infty}\left[\frac{x-0}{n}\sum_{k=1}^{n} f(t_k^*)\right]\\ &= \lim_{n\to\infty}\left[\frac{x}{n}\sum_{k=1}^{n} f\left(0+\tfrac{k(x-0)}{n}\right)\right]\\ &= \lim_{n\to\infty}\left[\frac{x}{n}\sum_{k=1}^{n} f\left(\tfrac{kx}{n}\right)\right]\end{aligned}$$

Example 7.2. *Suppose $f(x) = x^2$, then we can evaluate the indefinite integral as follows.*

$$\begin{aligned}\int_0^x f(t)dt &= \lim_{n\to\infty}\left[\frac{x}{n}\sum_{k=1}^{n} f\left(\frac{kx}{n}\right)\right] \\ &= \lim_{n\to\infty}\left[\frac{x}{n}\sum_{k=1}^{n}\left(\frac{kx}{n}\right)^2\right] \\ &= \lim_{n\to\infty}\left[\frac{x}{n}\sum_{k=1}^{n}\frac{k^2\cdot x^2}{n^2}\right] \\ &= \lim_{n\to\infty}\left[\frac{x^3}{n^3}\sum_{k=1}^{n}k^2\right] \\ &= \lim_{n\to\infty}\left[\frac{x^3}{n^3}\cdot\frac{n(n+1)(2n+1)}{6}\right] \\ &= \lim_{n\to\infty}\left[\frac{x^3}{n^3}\cdot\frac{2n^3+3n^2+n}{6}\right] \\ &= x^3\cdot\lim_{n\to\infty}\left[\frac{2n^3}{6n^3}+\frac{3n^2}{6n^3}+\frac{n}{6n^3}\right] \\ &= x^3\cdot\lim_{n\to\infty}\left[\frac{1}{3}+\frac{1}{2n}+\frac{1}{6n^2}\right] \\ &= x^3\cdot\left(\frac{1}{3}\right) \\ &= \frac{x^3}{3}\end{aligned}$$

7.2 Recognizing Derivatives

If we recognize a function $g(x)$ as being the derivative of a function $f(x)$, then we can easily express the antiderivative of $g(x)$:

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

For example, since

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

we can conclude that

$$\int \cos(x)dx = \sin(x) + C$$

Similarly, since we know e^x is its own derivative,

$$\int e^x dx = e^x + C$$

The power rule for derivatives can be reversed to give us a way to handle integrals of powers of x. Since

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

we can conclude that

$$\int nx^{n-1}dx = x^n + C$$

or, a little more usefully,

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

7.3 Integration by Substitution

The Substitution Rule

The substitution rule is a valuable asset in the toolbox of any integration grease monkey. It is essentially the chain rule (a differentiation technique you should be familiar with) in reverse. First, let's take a look at an example:

Preliminary Example

Suppose we want to find $\int x\cos(x^2)dx$. That is, we want to find a function such that its derivative equals $x\cos(x^2)$. Stated yet another way, we want to find an antiderivative of $f(x) = x\cos(x^2)$. Since $\sin(x)$ differentiates to $\cos(x)$, as a first guess we might try the function $\sin(x^2)$. But by the Chain Rule,

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

Which is almost what we want apart from the fact that there is an extra factor of 2 in front. But this is easily dealt with because we can divide by a constant (in this case 2). So,

$$\frac{d}{dx}\frac{\sin(x^2)}{2} = \frac{1}{2}\cdot\frac{d}{dx}\sin(x^2) = \frac{1}{2}\cdot 2\cos(x^2)x = x\cos(x^2) = f(x)$$

Thus, we have discovered a function, $F(x) = \dfrac{\sin(x^2)}{2}$, whose derivative is $x\cos(x^2)$. That is, F is an antiderivative of $f(x) = x\cos(x^2)$. This gives us

$$\int x\cos(x^2)dx = \frac{\sin(x^2)}{2} + C$$

Evaluate $\int nx^{n-1}$ or $\int \cos(x)$ is. We usually get

Evaluate $\int \dfrac{\sin(\ln(x))}{x}$ instead. These look hard, but there is a way to do them. Mathematicians call it Integration by Substitution, and for many integrals, this can be used to re-express the integrand in a way that makes finding of an antiderivative possible and easy. Sure, depending on the form of the integrand, the substitution to make may be different, but there is no doubt that the overall method is useful.

The objective of Integration by substitution is to substitute the integrand from an expression with variable x and the right side of the integral dx to an expression with variable u where $u = g(x)$ and the right side of the integral du, where $du = g'(x)dx$. How? By identifying a function and its derivative that makes up a part of the overall equation.

Goal

The general gist of Integration by Substitution is to transform the integral so that instead of referencing x , it references the function u. We can show how this method works by abstracting each step using math. In math, we can write down what we want to do (write the steps of Integration by Substitution in math) by writing

Given $u = g(x)$,

$$\int_{x=a}^{x=b} f(x)dx \rightarrow \int_{u=c}^{u=d} h(u)du$$

Steps

$$\int_{x=a}^{x=b} f(x)dx = \int_{x=a}^{x=b} f(x)\,\frac{du}{du}\,dx \quad i.e.\frac{du}{du} = 1 \tag{7.1}$$

$$= \int_{x=a}^{x=b} \left(f(x)\,\frac{dx}{du}\right)\left(\frac{du}{dx}\right)\,dx \quad i.e.\frac{dx}{du}\cdot\frac{du}{dx} = 1 \tag{7.2}$$

$$= \int_{x=a}^{x=b} \left(f(x)\,\frac{dx}{du}\right) g'(x)\,dx \quad i.e.\frac{du}{dx} = g'(x) \tag{7.3}$$

$$= \int_{x=a}^{x=b} h(g(x))g'(x)dx \quad \text{i.e. Now equate } \left(f(x)\,\frac{dx}{du}\right) \text{ with } h(g(x)) \tag{7.4}$$

$$= \int_{x=a}^{x=b} h(u)g'(x)dx \quad i.e.g(x) = u \tag{7.5}$$

$$= \int_{u=g(a)}^{u=g(b)} h(u)du \quad i.e.du = \frac{du}{dx}dx = g'(x)dx \tag{7.6}$$

$$= \int_{u=c}^{u=d} h(u)du \quad \text{i.e. We have achieved our desired result} \tag{7.7}$$

Procedure

If the previous mathematical steps are difficult to grasp all at once or difficult to put into practice, don't worry! Here are the steps written in plain English. It even includes the Goal too.

Find a function $u = g(x)$ that has a $g'(x)$ also in the expression somewhere . This may involve experimenting or staring at the expression in the integrand long enough If the question is hard, finding the $g'(x)$ may involve synthesizing numbers (constants) from nowhere so that it can be used to cancel out portions of $g'(x)$. However, if the entirety of $g'(x)$ needs to cancel artificially, then this may be a sign that you are making a question harder.

- Calculate $g'(x) = \dfrac{du}{dx}$
- Calculate $h(u)$ which is $f(x)\dfrac{dx}{du} = \dfrac{f(x)}{g'(x)}$ and make sure the final expression $h(u)$ does not have x in it
- Calculate $c = g(a)$
- Calculate $d = g(b)$

In summary, Integration by Substitution tells us the following

Substitution rule for definite integrals
Assume u is differentiable with continuous derivative and that f is continuous on the range of u. Suppose $c = u(a), d = u(b)$. Then

$$\int_a^b f(u(x))\frac{du}{dx}dx = \int_c^d f(u)du.$$

Substitution rule for indefinite integrals
Assume u is differentiable with continuous derivative and that f is continuous on the range of u. Then

$$\int f\Big(u(x)\Big)\frac{du}{dx}dx = \int f(u)du$$

Notice that it looks like you can "cancel" in the expression $\dfrac{du}{dx}dx$ to leave just a du. This does not really make any sense because $\dfrac{du}{dx}$ is not a fraction. But it's a good way to remember the substitution rule.

Examples

Integrating with the derivative present
Under ideal circumstances for Integration by Substitution, a component of the integrand can be viewed as the derivative of another component of the integrand. This makes it so that the substitution can be easily applied to simplify the integrand.

For example, in the integral

$$\int 3x^2(x^3+1)^5 dx$$

we see that $3x^2$ is the derivative of x^3+1. Letting

$$u = x^3 + 1$$

we have

$$\frac{du}{dx} = 3x^2$$

or, in order to apply it to the integral,

$$du = 3x^2 dx$$

With this we may write

$$\int 3x^2(x^3+1)^5 dx = \int u^5 du = \frac{u^6}{6} + C = \frac{(x^3+1)^6}{6} + C$$

Note that it was not necessary that we had exactly the derivative of u in our integrand. It would have been sufficient to have any constant multiple of the derivative.

For instance, to treat the integral

$$\int x^4 \sin(x^5) dx$$

we may let $u = x^5$. Then

$$du = 5x^4 dx$$

and so

$$\frac{du}{5} = x^4 dx$$

the right-hand side of which is a factor of our integrand. Thus,

$$\int x^4 \sin(x^5) dx = \int \frac{\sin(u)}{5} du = -\frac{\cos(u)}{5} + C = -\frac{\cos(x^5)}{5} + C$$

In general, the integral of a power of a function times that function's derivative may be integrated in this way. Since $\frac{d[g(x)]}{dx} = g'(x)$,

we have $dx = \frac{d[g(x)]}{g'(x)}$.

Therefore,

$$\int g'(x)g(x)^n dx = \int g'(x)g(x)^n \frac{d[g(x)]}{g'(x)}$$

$$= \int g(x)^n d[g(x)]$$

$$= \frac{g(x)^{n+1}}{n+1}$$

There is a similar rule for definite integrals, but we have to change the endpoints.

Synthesizing Terms

What if the derivative does not show up one-for-one in the expression? This is okay! For some integrals, it may be necessary to synthesize constants in order to solve the integral. Usually, this looks like a multiplication between the expression and $\frac{n}{n} = 1$, for some number n. Note that this usually works for variables as well, but synthesizing variables should not be a common thing and should only be an absolute last resort.

As an example of this practice put into the Integration by Substitution method, consider the integral

$$\int_0^2 x\cos(1+x^2)dx$$

By using the substitution $u = 1 + x^2$, we obtain $du = 2x\,dx$. However, notice that the constant 2 does not show up in the expression in the integrand. This is where this extra step applies. Notice that

$$\begin{aligned}\int_0^2 x\cos(1+x^2)dx &= \frac{1}{2}\int_0^2 2x\cos(1+x^2)dx\\ &= \frac{1}{2}\int_1^5 \cos(u)du\\ &= \frac{\sin(5)-\sin(1)}{2}\end{aligned}$$

and remember to calculate the new bounds for this integral. The lower limit for this integral was $x = 0$ but is now $u = 1 + 0^2 = 1$ and the upper limit was $x = 2$ but is now $u = 1 + 2^2 = 5$.

The following example shows how powerful a technique substitution can be. At first glance the following integral seems intractable, but after a little simplification, it's possible to tackle using substitution.

Example 7.3. *We will show that*

$$\int \frac{dx}{(x^2+a^2)\sqrt{x^2+a^2}} = \frac{x}{a^2\sqrt{x^2+a^2}} + C$$

First, we re-write the integral:

$$\begin{aligned}\int \frac{dx}{(x^2+a^2)\sqrt{x^2+a^2}} &= \int (x^2+a^2)^{-\frac{3}{2}} dx \\ &= \int \left(x^2\left(1+\tfrac{a^2}{x^2}\right)\right)^{-\frac{3}{2}} dx \\ &= \int x^{-3}\left(1+\tfrac{a^2}{x^2}\right)^{-\frac{3}{2}} dx \\ &= \int \left(1+\tfrac{a^2}{x^2}\right)^{-\frac{3}{2}} \left(x^{-3}dx\right)\end{aligned}$$

Now we perform the following substitution:

$$\begin{aligned} u &= 1+\frac{a^2}{x^2} \\ \frac{du}{dx} &= -2a^2x^{-3} \quad \Longrightarrow \quad x^{-3}dx = -\frac{du}{2a^2}\end{aligned}$$

Which yields:

$$\begin{aligned}\int \left(1+\tfrac{a^2}{x^2}\right)^{-\frac{3}{2}} \left(x^{-3}dx\right) &= \int u^{-\frac{3}{2}}\left(-\frac{du}{2a^2}\right) \\ &= -\frac{1}{2a^2}\int u^{-\frac{3}{2}} du \\ &= -\frac{1}{2a^2}\left(-\frac{2}{\sqrt{u}}\right) + C \\ &= \frac{1}{a^2\sqrt{1+\frac{a^2}{x^2}}} + C \\ &= \left(\frac{x}{x}\right)\frac{1}{a^2\sqrt{1+\frac{a^2}{x^2}}} + C \\ &= \frac{x}{a^2\sqrt{x^2+a^2}} + C\end{aligned}$$

Exercises

1. Evaluate the following using a suitable substitution.

   (a) $\int x\sin(2x^2)dx$

   Solution

by making the substitution $u = 2x^2$ Since $u = 2x^2, du = 4xdxanddx = \frac{du}{4x}$

$$\begin{aligned}\int x\sin(2x^2)dx &= \int x\sin(u)\frac{du}{4x}\\ &= \frac{1}{4}\int \sin(u)du\\ &= -\frac{\cos(u)}{4} + C\\ &= -\frac{\cos(\mathbf{2x^2})}{\mathbf{4}} + \mathbf{C}\end{aligned}$$

(b) $\int -3\cos(x)e^{\sin(x)}dx$
Solution
Let $u = \sin(x), du = \cos(x)dxsothatdx = \frac{du}{\cos(x)}$

$$\begin{aligned}\int -3\cos(x)e^{\sin(x)}dx &= -3\int \cos(x)e^u\frac{du}{\cos(x)}\\ &= -3\int e^u du\\ &= -3e^u + C\\ &= \mathbf{-3e^{\sin(x)} + C}\end{aligned}$$

(c) $\int \frac{19}{\sqrt{9x-38}}dx$
Solution

Let

$$u = 9x - 38 \qquad du = 9dx \qquad dx = \frac{du}{9}$$

Then

$$\begin{aligned}\int \frac{19}{\sqrt{9x-38}}dx &= \int \frac{19}{9\sqrt{u}}du\\ &= \frac{19}{9}(2\sqrt{u}) + C\\ &= \frac{\mathbf{38\sqrt{9x-38}}}{\mathbf{9}} + \mathbf{C}\end{aligned}$$

(d) $\int -15\sqrt{9x+43}dx$
Solution

Let

$$u = \cos(x) \qquad du = -\sin(x)dx \qquad dx = -\frac{du}{\sin(x)}$$

Then

$$\begin{aligned}\int \frac{17\sin(x)}{\cos(x)}dx &= 17\int -\frac{du}{u} \\ &= -17\ln|u| + C \\ &= \mathbf{-17\ln|\cos(x)| + C}\end{aligned}$$

(e) $\displaystyle\int \frac{17\sin(x)}{\cos(x)}dx$

Solution

Let

$$u = 9x + 43 \qquad du = 9dx \qquad dx = \frac{du}{9}$$

Then

$$\begin{aligned}\int -15\sqrt{9x+43}dx &= -15\int \frac{\sqrt{u}}{9}du \\ &= -\frac{15}{9}\frac{2}{3}u^{3/2} + C \\ &= -\mathbf{\frac{10(9x+43)^{3/2}}{9} + C}\end{aligned}$$

(f) $\displaystyle\int 5\cos(x)\sin(x)dx$

Solution

Let

$$u = \sin(x) \qquad du = \cos(x)dx \qquad dx = \frac{du}{\cos(x)}$$

Then

$$\begin{aligned}\int 5\cos(x)\sin(x)dx &= 5\int udu \\ &= 5\frac{u^2}{2} + C \\ &= \mathbf{\frac{5\sin^2(x)}{2} + C}\end{aligned}$$

(g) $\displaystyle\int_0^1 -\frac{10}{(-5x-32)^4}dx$

Solution

Let

$$u = -5x - 32 \qquad du = -5dx \qquad dx = -\frac{du}{5}$$

Then

$$\begin{aligned}
\int_0^1 -\frac{10}{(-5x-32)^4}dx &= -10\int_{u(0)}^{u(1)} \frac{-du}{5u^4} \\
&= -2\frac{1}{3u^3}\Bigg|_{u(0)}^{u(1)} \\
&= -2\frac{1}{3(-5x-32)^3}\Bigg|_0^1 \\
&= -\frac{2}{3}\left(\frac{1}{(-5-32)^3} - \frac{1}{(-32)^3}\right) \\
&= -\frac{2}{3}\left(\frac{1}{(-37)^3} + \frac{1}{(32)^3}\right) \\
&= \frac{2}{3}\left(\frac{1}{37^3} - \frac{1}{32^3}\right) \\
&= \frac{2}{3}\cdot\frac{32^3 - 37^3}{32^3 \cdot 37^3} \\
&= \frac{2}{3}\cdot\frac{32^3 - 37^3}{2^{15} \cdot 37^3} \\
&= \frac{32^3 - 37^3}{2^{14}\cdot 3 \cdot 37^3} \\
&= -\mathbf{\frac{17885}{2489696256}}
\end{aligned}$$

(h) $\int -3e^{3x+12}dx$

Solution

Let

$$u = 3x + 12 \qquad du = 3dx \qquad dx = \frac{du}{3}$$

Then

$$\begin{aligned}
\int -3e^{3x+12}dx &= -3\int \frac{e^u}{3}du \\
&= -e^u + C \\
&= -\mathbf{e^{3x+12}} + \mathbf{C}
\end{aligned}$$

7.4 Integration by Parts

Integration by parts is another powerful tool for integration. It was mentioned above that one could consider integration by substitution as an application of the chain rule in reverse.

In a similar manner, one may consider integration by parts as the product rule in reverse.
Integration by parts for indefinite integrals

If $y = uv$ where u and v are functions of x, then

$$y' = (uv)' = u'v + uv'$$

Rearranging,

$$uv' = (uv)' - u'v$$

Therefore,

$$\int uv'dx = \int (uv)'\ dx - \int u'v\ dx$$

Therefore,

$$\int uv'\ dx = uv - \int vu'\ dx$$

, or

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

This is the integration by parts formula. It is very useful in many integrals involving products of functions, as well as others.
Or
Suppose f and g are differentiable and their derivatives are continuous. Let $u = f(x)$ and $v = \int g(x)dx$. Then

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

it is also very important to notice that

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

is not equal to

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

to set the $f(x)$ and $g(x)$ we need to follow the rule called **I.L.A.T.E.**
ILATE defines the order in which we must set the $f(x)$

1. I for inverse trigonometric function

2. L for log functions

3. A for algebraic functions

4. T for trigonometric functions

5. E for exponential function

$f(x)$ and $g(x)$ must be in the order of **ILATE** or else your final answers will not match with the main key

For instance, to treat

$$\int x\sin(x)dx$$

we choose $u = x$ and $dv = \sin x dx$. With these choices, we have $du = dx$ and $v = -\cos x$, and we have

$$\int x\sin(x)dx = -x\cos(x) - \int -\cos(x)dx = -x\cos(x) + \int \cos(x)dx = -x\cos(x) + \sin(x) + C$$

Note that the choice of u and dv was critical. Had we chosen the reverse, so that $u = \sin(x)$ and $dv = x\ dx$, the result would have been

$$\frac{x^2\sin(x)}{2} - \int \frac{x^2\cos(x)}{2}dx$$

The resulting integral is no easier to work with than the original; we might say that this application of integration by parts took us in the wrong direction.

So the choice is important. One general guideline to help us make that choice is, if possible, to choose u to be the factor of the integrand which becomes simpler when we differentiate it. In the last example, we see that $\sin(x)$ does not become simpler when we differentiate it: $\cos(x)$ is no simpler than $\sin(x)$.

With definite integral

For definite integrals the rule is essentially the same, as long as we keep the endpoints.

Integration by parts for definite integrals Suppose f and g are differentiable and their derivatives are continuous. Then

$$\int_a^b f(x)g'(x)dx = \Big(f(x)g(x)\Big)\Big|_a^b - \int_a^b f'(x)g(x)dx$$

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

This can also be expressed in Leibniz notation.

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

Example 7.4. *Find* $\int x\cos(x)dx$

Here we let: $u = x$*, so that* $du = dx$*,* $dv = \cos(x)dx$*, so that* $v = \sin(x)$*. Then:*

$$\begin{aligned}\int x\cos(x)dx &= \int u\,dv \\ &= uv - \int v\,du \\ &= x\sin(x) - \int \sin(x)dx \\ &= x\sin(x) + \cos(x) + C\end{aligned}$$

Example 7.5. *Find* $\int x^2e^xdx$ *In this example we will have to use integration by parts twice.*

Here we let

$u = x^2$*, so that* $du = 2xdx$*,* $dv = e^xdx$*, so that* $v = e^x$*. Then:*

$$\begin{aligned}\int x^2e^xdx &= \int u\,dv \\ &= uv - \int v\,du \\ &= x^2e^x - \int 2xe^xdx \\ &= x^2e^x - 2\int xe^xdx\end{aligned}$$

Now to calculate the last integral we use integration by parts again. Let $u = x$*, so that* $du = dx$*,* $dv = e^xdx$*, so that* $v = e^x$ *and integrating by parts gives* $\int xe^xdx = xe^x - \int e^xdx = e^x(x-1)$
So, finally we obtain $\int x^2e^xdx = x^2e^x - 2e^x(x-1) + C = e^x(x^2-2x+2) + C$

Example 7.6. *Find* $\int \ln(x)dx$

The trick here is to write this integral as

$$\int \ln(x)\cdot 1\,dx$$

Now let $u = \ln(x)$ *so* $du = \dfrac{dx}{x}$*,* $v = x$ *so* $dv = 1\,dx$*. Then using integration by parts,*

$$\begin{aligned}\int \ln(x)dx &= x\ln(x) - \int \frac{x}{x}dx \\ &= x\ln(x) - \int 1\,dx \\ &= x\ln(x) - x + C \\ &= x\Big(\ln(x) - 1\Big) + C\end{aligned}$$

Example 7.7. *Find* $\int \arctan(x)dx$

Again the trick here is to write the integrand as $\arctan(x) = \arctan(x) \cdot 1$. *Then let* $u = \arctan(x)$ *so* $du = \dfrac{dx}{1+x^2}$ $v = x$ *so* $dv = 1\,dx$ *so using integration by parts,*

$$\begin{aligned}\int \arctan(x)dx &= x\arctan(x) - \int \frac{x}{1+x^2}dx \\ &= x\arctan(x) - \tfrac{1}{2}\ln(1+x^2) + C\end{aligned}$$

Example 7.8. *Find* $\int e^x \cos(x)dx$

This example uses integration by parts twice. First let,

$u = e^x$ *so* $du = e^x dx$ $dv = \cos(x)dx$ *so* $v = \sin(x)$ *so*

$$\int e^x \cos(x)dx = e^x \sin(x) - \int e^x \sin(x)dx$$

Now, to evaluate the remaining integral, we use integration by parts again, with $u = e^x$ *so* $du = e^x dx$ $v = -\cos(x)$ *so* $dv = \sin(x)dx$ *Then*

$$\begin{aligned}\int e^x \sin(x)dx &= -e^x\cos(x) - \int -e^x\cos(x)dx \\ &= -e^x\cos(x) + \int e^x \cos(x)dx\end{aligned}$$

Putting these together, we have $\int e^x\cos(x)dx = e^x\sin(x) + e^x\cos(x) - \int e^x\cos(x)dx$ *Notice that the same integral shows up on both sides of this equation, but with opposite signs. The integral does not cancel; it doubles when we add the integral to both sides to get*

$$\begin{aligned}2\int e^x\cos(x)dx &= e^x\Big(\sin(x)+\cos(x)\Big) \\ \int e^x\cos(x)dx &= \frac{e^x\Big(\sin(x)+\cos(x)\Big)}{2}\end{aligned}$$

Exercises

1. Evaluate the following using integration by parts.

   (a) $\int \dfrac{2x-5}{x^3}dx$

   Solution

   using integration by parts with $u = 2x - 5$ and $dv = \frac{dx}{x^3}$

$du = 2dx; v = \int \frac{dx}{x^3} = -\frac{1}{2x^2}$

$$\begin{aligned}
\int \frac{2x-5}{x^3}dx &= \int udv \\
&= uv - \int vdu \\
&= (2x-5)(-\frac{1}{2x^2}) - \int (-\frac{1}{2x^2})2dx \\
&= \frac{5-2x}{2x^2} + \int \frac{dx}{x^2} \\
&= \frac{5-2x}{2x^2} - \frac{1}{x} \\
&= \frac{5-2x}{2x^2} - \frac{2x}{2x^2} \\
&= \mathbf{\frac{5-4x}{2x^2}}
\end{aligned}$$

(b) $\int (2x-1)e^{-3x+1}dx$

Solution

Let $u = 2x - 1; dv = e^{-3x+1}dx$ Then $du = 2dx and dv = \int e^{-3x+1}dx$ To evaluate v, make the substitution $w = -3x + 1; dw = -3dx; dx = \frac{-dw}{3}$. Then $v = \int e^{-3x+1}dx = \int e^w(\frac{-1}{3})dw = \frac{-e^w}{3} = \frac{-e^{-3x+1}}{3}$. So

$$\begin{aligned}
\int (2x-1)e^{-3x+1}dx &= \int udv \\
&= uv - \int vdu \\
&= (2x-1)\frac{-e^{-3x+1}}{3} - \int \frac{-e^{-3x+1}}{3}(2)dx \\
&= \frac{(1-2x)e^{-3x+1}}{3} + \frac{2}{3}\int e^{-3x+1}dx \\
&= \frac{(1-2x)e^{-3x+1}}{3} + \frac{2}{3}\int \frac{-e^w}{3}dw \\
&= \frac{3(1-2x)e^{-3x+1}}{9} - \frac{2}{9}e^w \\
&= \frac{(3-6x)e^{-3x+1}}{9} - \frac{2}{9}e^{-3x+1} \\
&= \mathbf{\frac{(1-6x)e^{-3x+1}}{9}}
\end{aligned}$$

(c) $\int -4\ln(x)dx$

Solution

Let $u = \ln(x) \qquad du = \frac{dx}{x}$

$v = -4x \qquad dv = -4dx$

Then

$$\int -4\ln(x)dx = -4x\ln(x) - \int -4dx$$
$$= \mathbf{-4x\ln(x) + 4x + C}$$

(d) $\int (38 - 7x)\cos(x)dx$

Solution

Let $u = -7x + 38 \qquad du = -7dx$

$v = \sin(x) \qquad dv = \cos(x)dx$

Then

$$\int (-7x + 38)\cos(x)dx = (-7x + 38)\sin(x) + 7\int \sin(x)dx$$
$$= \mathbf{(-7x + 38)\sin(x) - 7\cos(x) + C}$$

(e) $\int_0^{\frac{\pi}{2}} (-6x + 45)\cos(x)dx$

Solution

Let $u = -6x + 45 \qquad du = -6dx$

$v = \sin(x) \qquad dv = \cos(x)dx$

Then

$$\int_0^{\frac{\pi}{2}} (-6x + 45)\cos(x)dx = ((-6x + 45)\sin(x))\Big|_0^{\frac{\pi}{2}} + 6\int_0^{\frac{\pi}{2}} \sin(x)dx$$
$$= (-3\pi + 45) - 6\cos(x)\Big|_0^{\frac{\pi}{2}}$$
$$= (-3\pi + 45) + 6$$
$$= \mathbf{-3\pi + 51}$$

(f) $\int (5x + 1)(x - 6)^4 dx$

Solution

Let $u = 5x + 1 \qquad du = 5dx$

$v = \frac{(x-6)^5}{5} \qquad dv = (x - 6)^4 dx$

Then

$$\int (5x + 1)(x - 6)^4 dx = \frac{(5x + 1)(x - 6)^5}{5} - \int (x - 6)^5 dx$$
$$= \mathbf{\frac{(5x + 1)(x - 6)^5}{5} - \frac{(x - 6)^6}{6} + C}$$

(g) $\int\limits_{-1}^{1} (2x+8)^3(2-x)dx$

Solution

Let $\begin{array}{ll} u = -x+2 & du = -dx \\ v = \frac{(2x+8)^4}{8} & dv = (2x+8)^3 dx \end{array}$

Then

$$\begin{aligned}
\int_{-1}^{1} (2x+8)^3(-x+2)dx &= \left.\frac{(-x+2)(2x+8)^4}{8}\right|_{-1}^{1} + \int_{-1}^{1} \frac{(2x+8)^4}{8}dx \\
&= \frac{10^4-(3\cdot 6^4)}{8} + \left.\frac{1}{8}\frac{(2x+8)^5}{10}\right|_{-1}^{1} \\
&= \frac{10^4-3\cdot 6^4}{8} + \frac{1}{80}(10^5-6^5) \\
&= \frac{10^5-30\cdot 6^4+10^5-6^5}{80} \\
&= \frac{2\cdot 10^5-6^4(30+6)}{80} \\
&= \frac{2\cdot 10^5-6^6}{80} \\
&= \mathbf{\frac{9584}{5}}
\end{aligned}$$

(h) $\int \sin(x)e^x dx$

7.5 Trigonometric Substitution

The idea behind the trigonometric substitution is quite simple: to replace expressions involving square roots with expressions that involve standard trigonometric functions, but no square roots. Integrals involving trigonometric functions are often easier to solve than integrals involving square roots.

Let us demonstrate this idea in practice. Consider the expression $\sqrt{1-x^2}$. Probably the most basic trigonometric identity is $\sin^2(\theta)+\cos^2(\theta) = 1$ for an arbitrary angle θ. If we replace x in this expression by $\sin(\theta)$, with the help of this trigonometric identity we see

$$\sqrt{1-x^2} = \sqrt{1-\sin^2(\theta)} = \sqrt{\cos^2(\theta)} = \cos(\theta)$$

Note that we could write $\theta = \arcsin(x)$ since we replaced x^2 with $\sin^2(\theta)$.

We would like to mention that technically one should write the absolute value of $\cos(\theta)$, in other words $\left|\cos(\theta)\right|$ as our final answer since $\sqrt{a^2} = |a|$ for all possible a. But as long as we are careful about the domain of all possible x and how $\cos(\theta)$ is used in the final computation, omitting the absolute value signs does not constitute a problem. However, we cannot directly interchange the simple expression $\cos(\theta)$ with the complicated $\sqrt{1-x^2}$ wherever it may appear, we must remember when integrating by substitution we need to take the derivative into account. That is we need to remember that $dx = \cos(\theta)d\theta$, and to get a integral that only involves θ we need to also replace dx by something in terms of $d\theta$. Thus, if we see an integral of the form

$$\int \sqrt{1-x^2}dx$$

we can rewrite it as

$$\int \cos(\theta)\cos(\theta)d\theta = \int \cos^2(\theta)d\theta.$$

Notice in the expression on the left that the first $\cos(\theta)$ comes from replacing the $\sqrt{1-x^2}$ and the $\cos(\theta)d\theta$ comes from substituting for the dx. Since $\cos^2(\theta) = \dfrac{1+\cos(2\theta)}{2}$ our original integral reduces to:

$$\tfrac{1}{2}\int d\theta + \tfrac{1}{2}\int \cos(2\theta)d\theta.$$

These last two integrals are easily handled. For the first integral we get

$$\tfrac{1}{2}\int d\theta = \tfrac{1}{2}\theta$$

For the second integral we do a substitution, namely $u = 2\theta$ (and $du = 2d\theta$) to get:

$$\tfrac{1}{2}\int \cos(2\theta)d\theta = \tfrac{1}{2}\int \cos(u)\tfrac{1}{2}du = \tfrac{1}{4}\sin(u) = \frac{\sin(2\theta)}{4}$$

Finally we see that:

$$\int \cos^2(\theta)d\theta = \frac{\theta}{2} + \frac{\sin(2\theta)}{4} = \frac{\theta + \sin(\theta)\cos(\theta)}{2}$$

However, this is in terms of θ and not in terms of x , so we must substitute back in order to rewrite the answer in terms of x.

That is we worked out that:

$$\sin(\theta) = x \qquad \cos(\theta) = \sqrt{1-x^2} \qquad \theta = \arcsin(x)$$

So we arrive at our final answer

$$\int \sqrt{1-x^2}dx = \frac{\arcsin(x) + x\sqrt{1-x^2}}{2}$$

As you can see, even for a fairly harmless looking integral this technique can involve quite a lot of calculation. Often it is helpful to see if a simpler method will suffice before turning to trigonometric substitution. On the other hand, frequently in the case of integrands involving square roots, this is the most tractable way to solve the problem. We begin with giving some rules of thumb to help you decide which trigonometric substitutions might be helpful.

If the integrand contains a single factor of one of the forms $\sqrt{a^2 - x^2}$ or $\sqrt{a^2 + x^2}$ or $\sqrt{x^2 - a^2}$ we can try a trigonometric substitution.

If the integrand contains $\sqrt{a^2 - x^2}$ let $x = a\sin(\theta)$ and use the identity $1 - \sin^2(\theta) = \cos^2(\theta)$.
If the integrand contains $\sqrt{a^2 + x^2}$ let $x = a\tan(\theta)$ and use the identity $1 + \tan^2(\theta) = \sec^2(\theta)$.
If the integrand contains $\sqrt{x^2 - a^2}$ let $x = a\sec(\theta)$ and use the identity $\sec^2(\theta - 1) = \tan^2(\theta)$.

Sine substitution

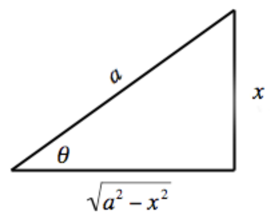


Figure 7.1: This substitution is easily derived from a triangle, using the Pythagorean Theorem.

If the integrand contains a piece of the form $\sqrt{a^2 - x^2}$ we use the substitution

$$x = a\sin(\theta) \quad dx = a\cos(\theta)d\theta$$

This will transform the integrand to a trigonometric function. If the new integrand can't be integrated on sight then the tan-half-angle substitution described below will generally transform it into a more tractable algebraic integrand.

E.g., if the integrand is $\sqrt{1 - x^2}$,

$$\begin{aligned}\int_0^1 \sqrt{1-x^2}dx &= \int_0^{\frac{\pi}{2}} \sqrt{1-\sin^2(\theta)}\cos(\theta)d\theta \\ &= \int_0^{\frac{\pi}{2}} \cos^2(\theta)d\theta \\ &= \frac{1}{2}\int_0^{\frac{\pi}{2}} \Big(1+\cos(2\theta)\Big)d\theta \\ &= \frac{\pi}{4}\end{aligned}$$

If the integrand is $\sqrt{\dfrac{1+x}{1-x}}$, we can rewrite it as

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

Then we can make the substitution

$$\begin{aligned}\int_0^a \frac{1+x}{\sqrt{1-x^2}}dx &= \int_0^\alpha \frac{1+\sin(\theta)}{\cos(\theta)}\cos(\theta)d\theta \quad 0<a<1\\ &= \int_0^\alpha \Big(1+\sin(\theta)\Big)d\theta\\ \alpha &= \arcsin(a)\\ &= \alpha + \Big[-\cos(\theta)\Big]_0^\alpha\\ &= \alpha+1-\cos(\alpha)\\ &= 1+\arcsin(a)-\sqrt{1-a^2}\end{aligned}$$

Tangent substitution

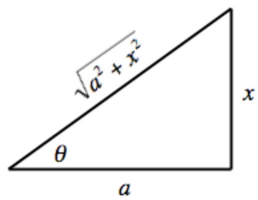


Figure 7.2: This substitution is easily derived from a triangle, using the Pythagorean Theorem.

When the integrand contains a piece of the form $\sqrt{a^2+x^2}$ we use the substitution

$$x=a\tan(\theta) \qquad \sqrt{x^2+a^2}=a\sec(\theta) \qquad dx=a\sec^2(\theta)d\theta$$

E.g., if the integrand is $(x^2+a^2)^{-\frac{3}{2}}$ then on making this substitution we find

$$\begin{aligned}\int_0^z (x^2+a^2)^{-\frac{3}{2}}dx &= a^{-2}\int_0^\alpha \cos(\theta)d\theta \qquad z>0\\ &= a^{-2}\Big[\sin(\theta)\Big]_0^\alpha \qquad \alpha=\arctan\left(\tfrac{z}{a}\right)\\ &= a^{-2}\sin(\alpha)\\ &= a^{-2}\frac{\frac{z}{a}}{\sqrt{1+\frac{z^2}{a^2}}}\\ &= \frac{z}{a^2\sqrt{a^2+z^2}}\end{aligned}$$

If the integral is

$$I = \int_0^z \sqrt{x^2 + a^2} \qquad z > 0$$

then on making this substitution we find

$$\begin{aligned} I &= a^2 \int_0^\alpha \sec^3\theta \, d\theta && \alpha = \tan^{-1}(z/a) \\ &= a^2 \int_0^\alpha \sec\theta \, d\tan\theta \\ &= a^2[\sec\theta\tan\theta]_0^\alpha - a^2 \int_0^\alpha \sec\theta\tan^2\theta \, d\theta \\ &= a^2\sec\alpha\tan\alpha - a^2\int_0^\alpha \sec^3\theta\, d\theta + a^2\int_0^\alpha \sec\theta\, d\theta \\ &= a^2\sec\alpha\tan\alpha - I + a^2\int_0^\alpha \sec\theta\, d\theta \end{aligned}$$

After integrating by parts, and using trigonometric identities, we've ended up with an expression involving the original integral. In cases like this we must now rearrange the equation so that the original integral is on one side only

$$\begin{aligned} I &= \tfrac{a^2}{2}\left(\sec(\alpha)\tan(\alpha) + \int_0^\alpha \sec(\theta) d\theta\right) \\ &= \tfrac{a^2}{2}\left(\sec(\alpha)\tan(\alpha) + \left[\ln\left(\sec(\theta)+\tan(\theta)\right)\right]_0^\alpha\right) \\ &= \tfrac{a^2}{2}\left(\sec(\alpha)\tan(\alpha) + a^2\ln\left(\sec(\alpha)+\tan(\alpha)\right)\right) \\ &= \tfrac{a^2}{2}\left(\tfrac{z}{a^2}\sqrt{a^2+z^2} + \ln\left(\tfrac{z+\sqrt{a^2+z^2}}{a}\right)\right) \\ &= \tfrac{1}{2}\left(z\sqrt{z^2+a^2} + a^2\ln\left(\tfrac{z+\sqrt{a^2+z^2}}{a}\right)\right) \end{aligned}$$

As we would expect from the integrand, this is approximately $\dfrac{z^2}{2}$ for large z.

In some cases it is possible to do trigonometric substitution in cases when there is no $\sqrt{\ }$ appearing in the integral.

Example 7.9. $\displaystyle\int \frac{dx}{x^2+1}$

The denominator of this function is equal to $\left(\sqrt{1+x^2}\right)^2$. This suggests that we try to substitute $x = \tan(u)$ and use the identity $1 + \tan^2(u) = \sec^2(u)$. With this substitution, we obtain that $dx = \sec^2(u)du$ and thus

$$\begin{aligned} \int \frac{dx}{x^2+1} = \int \frac{\sec^2(u)}{\tan^2(u)+1} du &= \int \frac{\sec^2(u)}{\sec^2(u)} du \\ &= \int du \\ &= u + c \end{aligned}$$

Using the initial substitution $u = \arctan(x)$ gives

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

Secant substitution

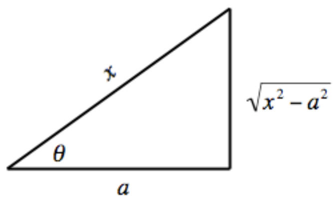


Figure 7.3: This substitution is easily derived from a triangle, using the Pythagorean Theorem.

If the integrand contains a factor of the form $\sqrt{x^2 - a^2}$ we use the substitution

$$x = a\sec(\theta) \quad dx = a\sec(\theta)\tan(\theta)d\theta \quad \sqrt{x^2-a^2} = a\tan(\theta)$$

Example 7.10. *Find* $\int\limits_0^z \frac{\sqrt{x^2-1}}{x}dx$.

$$\begin{aligned}
\int\limits_0^z \frac{\sqrt{x^2-1}}{x}dx &= \int\limits_0^\alpha \frac{\tan(\theta)}{\sec(\theta)}\sec(\theta)\tan(\theta)d\theta && z > 1 \\
&= \int\limits_0^\alpha \tan^2(\theta)d\theta && \alpha = \operatorname{arcsec}(z) \\
&= \Big[\tan(\theta) - \theta\Big]_0^\alpha && \tan(\alpha) = \sqrt{\sec^2(\alpha) - 1} \\
&= \tan(\alpha) - \alpha && \tan(\alpha) = \sqrt{z^2-1} \\
&= \sqrt{z^2-1} - \operatorname{arcsec}(z)
\end{aligned}$$

Example 7.11. *Find* $\int\limits_1^z \frac{\sqrt{x^2-1}}{x^2}dx$.

$$\begin{aligned}
\int\limits_1^z \frac{\sqrt{x^2-1}}{x^2}dx &= \int\limits_1^\alpha \frac{\tan(\theta)}{\sec^2(\theta)}\sec(\theta)\tan(\theta)d\theta && z > 1 \\
&= \int\limits_0^\alpha \frac{\sin^2(\theta)}{\cos(\theta)}d\theta && \alpha = \operatorname{arcsec}(z)
\end{aligned}$$

We can now integrate by parts

$$\begin{aligned}\int_1^z \frac{\sqrt{x^2-1}}{x^2}dx &= -\Big[\tan(\theta)\cos(\theta)\Big]_0^\alpha + \int_0^\alpha \sec(\theta)d\theta \\ &= -\sin(\alpha) + \Big[\ln\Big(\sec(\theta)+\tan(\theta)\Big)\Big]_0^\alpha \\ &= \ln\Big(\sec(\alpha)+\tan(\alpha)\Big) - \sin(\alpha) \\ &= \ln\Big(z+\sqrt{z^2-1}\Big) - \frac{\sqrt{z^2-1}}{z}\end{aligned}$$

Exercise

Evaluate the following using an appropriate trigonometric substitution. $\int \frac{10}{25x^2+a}dx$

7.6 Trigonometric Integrals

Powers of Sine and Cosine

We will give a general method to solve generally integrands of the form $\cos^m(x)\cdot\sin^n(x)$. First let us work through an example.

$$\int \cos^3(x)\sin^2(x)dx$$

Notice that the integrand contains an odd power of cos. So rewrite it as

$$\int \cos^2(x)\sin^2(x)\cos(x)dx$$

We can solve this by making the substitution $u = \sin(x)$ so $du = \cos(x)dx$. Then we can write the whole integrand in terms of u by using the identity

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

So

$$\begin{aligned}\int \cos^3(x)\sin^2(x)dx &= \int \cos^2(x)\sin^2(x)\cos(x)dx \\ &= \int (1-u^2)u^2du \\ &= \int u^2du - \int u^4du \\ &= \frac{u^3}{3}+\frac{u^5}{5}+C \\ &= \frac{\sin^3(x)}{3}-\frac{\sin^5(x)}{5}+C\end{aligned}$$

This method works whenever there is an odd power of sine or cosine.

To evaluate $\int \cos^m(x)\sin^n(x)dx$ when either m or n is odd.

If m is odd substitute $u = \sin(x)$ and use the identity $\cos^2(x) = 1 - \sin^2(x) = 1 - u^2$. If n is odd substitute $u = \cos(x)$ and use the identity

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

Example 7.12. *Find* $\int_0^{\frac{\pi}{2}} \cos^{40}(x)\sin^3(x)dx$.

As there is an odd power of sin we let $u = \cos(x)$ so $du = -\sin(x)dx$. Notice that when $x = 0$ we have $u = \cos(0) = 1$ and when $x = \dfrac{\pi}{2}$ we have $u = \cos(\frac{\pi}{2}) = 0$

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} \cos^{40}(x)\sin^3(x)dx &= \int_0^{\frac{\pi}{2}} \cos^{40}(x)\sin^2(x)\sin(x)dx \\
&= -\int_1^0 u^{40}(1-u^2)du \\
&= \int_0^1 u^{40}(1-u^2)du \\
&= \int_0^1 (u^{40} - u^{42})du \\
&= \left(\frac{u^{41}}{41} - \frac{u^{43}}{43}\right)\Bigg|_0^1 \\
&= \frac{1}{41} - \frac{1}{43}
\end{aligned}$$

When both m and n are even things get a little more complicated.

To evaluate $\int \cos^m(x)\sin^n(x)dx$ when both m and n are even.

Use the identities $\sin^2(x) = \dfrac{1-\cos(2x)}{2}$ and $\cos^2(x) = \dfrac{1+\cos(2x)}{2}$.

Example 7.13. *Find* $\int \sin^2(x)\cos^4(x)dx$.

As $\sin^2(x) = \dfrac{1-\cos(2x)}{2}$ and $\cos^2(x) = \dfrac{1+\cos(2x)}{2}$ we have

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

and expanding, the integrand becomes

$$\frac{1}{8}\int \left(1 - \cos^2(2x) + \cos(2x) - \cos^3(2x)\right) dx$$

Using the multiple angle identities

$$
\begin{aligned}
I &= \frac{1}{8}\left(\int 1dx - \int \cos^2(2x)dx + \int \cos(2x)dx - \int \cos^3(2x)dx\right)\\
&= \frac{1}{8}\left(x - \frac{1}{2}\int \Big(1+\cos(4x)\Big)dx + \frac{\sin(2x)}{2} - \int \cos^2(2x)\cos(2x)dx\right)\\
&= \frac{1}{16}\left(x + \sin(2x) + \int \cos(4x)dx - 2\int \Big(1-\sin^2(2x)\Big)\cos(2x)dx\right)
\end{aligned}
$$

then we obtain on evaluating $I = \dfrac{x}{16} - \dfrac{\sin(4x)}{64} + \dfrac{\sin^3(2x)}{48} + C$

Powers of Tan and Secant

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

If n is even and $n \geq 2$ then substitute $u = tan(x)$ and use the identity $\sec^2(x) = 1 + \tan^2(x)$. If n and m are both odd then substitute $u = \sec(x)$ and use the identity $\tan^2(x) = \sec^2(x) - 1$. If n is odd and m is even then use the identity $\tan^2(x) = \sec^2(x) - 1$ and apply a reduction formula to integrate $\sec^j(x)dx$, using the examples below to integrate when $j = 1, 2$.

Example 7.14. *Find* $\int \sec^2(x)dx$.

There is an even power of $\sec(x)$. Substituting $u = \tan(x)$ gives $du = \sec^2(x)dx$ so

$$\int \sec^2(x)dx = \int du = u + C = \tan(x) + C.$$

Example 7.15. *Find* $\int \tan(x)dx$.

Let $u = \cos(x)$ so $du = -\sin(x)dx$. Then

$$
\begin{aligned}
\int \tan(x)dx &= \int \frac{\sin(x)}{\cos(x)}dx\\
&= \int -\frac{du}{u}\\
&= -\ln|u| + C\\
&= -\ln\Big|\cos(x)\Big| + C\\
&= \ln\Big|\sec(x)\Big| + C
\end{aligned}
$$

Example 7.16. *Find* $\int \sec(x)dx$.

The trick to do this is to multiply and divide by the same thing like this:

$$
\begin{aligned}
\int \sec(x)dx &= \int \sec(x)\frac{\sec(x)+\tan(x)}{\sec(x)+\tan(x)}dx\\
&= \int \frac{\sec^2(x)+\sec(x)\tan(x)}{\sec(x)+\tan(x)}dx
\end{aligned}
$$

Making the substitution $u = \sec(x) + \tan(x)$ so $du = \sec(x)\tan(x) + \sec^2(x)dx$,

$$\begin{aligned}\int \sec(x)dx &= \int \frac{du}{u} \\ &= \ln|u| + C \\ &\ln\Big|\sec(x) + \tan(x)\Big| + C\end{aligned}$$

More trigonometric combinations

For the integrals $\int \sin(nx)\cos(mx)dx$ or $\int \sin(nx)\sin(mx)dx$ or $\int \cos(nx)\cos(mx)dx$ use the identities

$$\begin{aligned}\sin(a)\cos(b) &= \frac{\sin(a+b) + \sin(a-b)}{2} \\ \sin(a)\sin(b) &= \frac{\cos(a-b) - \cos(a+b)}{2} \\ \cos(a)\cos(b) &= \frac{\cos(a-b) + \cos(a+b)}{2}\end{aligned}$$

Example 7.17. *Find* $\int \sin(3x)\cos(5x)dx$.

We can use the fact that $\sin(a)\cos(b) = \dfrac{\sin(a+b) + \sin(a-b)}{2}$, so

$$\sin(3x)\cos(5x) = \frac{\sin(8x) + \sin(-2x)}{2}$$

Now use the oddness property of $\sin(x)$ to simplify

$$\sin(3x)\cos(5x) = \frac{\sin(8x) - \sin(2x)}{2}$$

And now we can integrate

$$\begin{aligned}\int \sin(3x)\cos(5x)dx &= \int \Big(\sin(8x) - \sin(2x)\Big) \\ &= \frac{\cos(2x)}{4} - \frac{\cos(8x)}{16} + C\end{aligned}$$

Example 7.18. *Find:* $\int \sin(x)\sin(2x)dx$.

Using the identities

$$\sin(x)\sin(2x) = \frac{\cos(-x) - \cos(3x)}{2} = \frac{\cos(x) - \cos(3x)}{2}$$

Then

$$\begin{aligned}\int \sin(x)\sin(2x)dx &= \frac{1}{2}\int \Big(\cos(x) - \cos(3x)\Big)dx \\ &= \frac{\sin(x)}{2} - \frac{\sin(3x)}{6} + C\end{aligned}$$

7.7 Reduction Formula

A reduction formula is one that enables us to solve an integral problem by reducing it to a problem of solving an easier integral problem, and then reducing that to the problem of solving an easier problem, and so on.

For example, if we let

$$I_n = \int x^n e^x dx$$

Integration by parts allows us to simplify this to

$$I_n = x^n e^x - n \int x^{n-1} e^x dx =$$

$$I_n = x^n e^x - n I_{n-1}$$

which is our desired reduction formula. Note that we stop at

$$I_0 = e^x.$$

Similarly, if we let

$$I_n = \int \sec^n(\theta) d\theta$$

then integration by parts lets us simplify this to

$$I_n = \sec^{n-2}(\theta)\tan(\theta) - (n-2) \int \sec^{n-2}(\theta)\tan^2(\theta) d\theta$$

Using the trigonometric identity, $\tan^2(\theta) = \sec^2(\theta) - 1$, we can now write

$$\begin{aligned} I_n &= \sec^{n-2}(\theta)\tan(\theta) + (n-2)\left(\int \sec^{n-2}(\theta)d\theta - \int \sec^n(\theta)d\theta\right) \\ &= \sec^{n-2}(\theta)\tan(\theta) + (n-2)\left(I_{n-2} - I_n\right) \end{aligned}$$

Rearranging, we get

$$I_n = \frac{\sec^{n-2}(\theta)\tan(\theta)}{n-1} + \frac{n-2}{n-1} I_{n-2}$$

Note that we stop at $n = 1$ or 2 if n is odd or even respectively.

As in these two examples, integrating by parts when the integrand contains a power often results in a reduction formula.

7.8 Partial Fraction Decomposition

Suppose we want to find $\int \frac{3x+1}{x^2+x} dx$. One way to do this is to simplify the integrand by finding constants A and B so that

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

This can be done by cross multiplying the fraction which gives

$$\frac{3x+1}{x(x+1)} = \frac{A(x+1)+Bx}{x(x+1)}$$

As both sides have the same denominator we must have

$$3x+1 = A(x+1)+Bx$$

This is an equation for x so it must hold whatever value x is. If we put in $x=0$ we get $A=1$ and putting $x=-1$ gives $=-B=-2$ so $B=2$. So we see that

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

Returning to the original integral

$$\begin{aligned}\int \frac{3x+1}{x^2+x}dx &= \int \frac{dx}{x} + \int \frac{2}{x+1}dx \\ &= \int \frac{dx}{x} + 2\int \frac{dx}{x+1} \\ &= \ln|x| + 2\ln\left|x+1\right| + C\end{aligned}$$

Rewriting the integrand as a sum of simpler fractions has allowed us to reduce the initial integral to a sum of simpler integrals. In fact this method works to integrate any rational function.

Method of Partial Fractions

To decompose the rational function $\dfrac{P(x)}{Q(x)}$:

- **Step 1** Use long division (if necessary) to ensure that the degree of $P(x)$ is less than the degree of $Q(x)$.
- **Step 2** Factor $Q(x)$ as far as possible.
- **Step 3** Write down the correct form for the partial fraction decomposition (see below) and solve for the constants.

To factor $Q(x)$ we have to write it as a product of linear factors (of the form $ax+b$ and irreducible quadratic factors (of the form ax^2+bx+c with $b^2-4ac<0$.

Some of the factors could be repeated. For instance if $Q(x)=x^3-6x^2+9x$ we factor $Q(x)$ as

$$Q(x) = x(x^2-6x+9) = x(x-3)(x-3) = x(x-3)^2$$

It is important that in each quadratic factor we have $b^2 - 4ac < 0$, otherwise it is possible to factor that quadratic piece further. For example if $Q(x) = x^3 - 3x^2 + 2x$ then we can write

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

We will now show how to write $\frac{P(x)}{Q(x)}$ as a sum of terms of the form $\frac{A}{(ax+b)^k}$ and $\frac{Ax+B}{(ax^2+bx+c)^k}$ Exactly how to do this depends on the factorization of $Q(x)$ and we now give four cases that can occur.

$Q(x)$ **is a product of linear factors with no repeats**

This means that $Q(x) = (a_1x + b_1)(a_2x + b_2)\cdots(a_nx + b_n)$ where no factor is repeated and no factor is a multiple of another.

For each linear term we write down something of the form $\frac{A}{(ax+b)}$, so in total we write

$$\frac{P(x)}{Q(x)} = \frac{A_1}{a_1x + b_1} + \frac{A_2}{a_2x + b_2} + \cdots + \frac{A_n}{a_nx + b_n}$$

Example 7.19. *Find* $\int \frac{1+x^2}{(x+3)(x+5)(x+7)}dx$

Here we have $P(x) = 1 + x^2$, $Q(x) = (x+3)(x+5)(x+7)$ and $Q(x)$ is a product of linear factors. So we write

$$\frac{1+x^2}{(x+3)(x+5)(x+7)} = \frac{A}{x+3} + \frac{B}{x+5} + \frac{C}{x+7}$$

Multiply both sides by the denominator

$$1 + x^2 = A(x+5)(x+7) + B(x+3)(x+7) + C(x+3)(x+5)$$

Substitute in three values of x to get three equations for the unknown constants,

$$\begin{aligned} x = -3 \quad & 1 + 3^2 = 2 \cdot 4A \\ x = -5 \quad & 1 + 5^2 = -2 \cdot 2B \\ x = -7 \quad & 1 + 7^2 = (-4) \cdot (-2)C \end{aligned}$$

so $A = \frac{5}{4}$, $B = -\frac{13}{2}$, $C = \frac{25}{4}$, and

$$\frac{1+x^2}{(x+3)(x+5)(x+7)} = \frac{5}{4x+12} - \frac{13}{2x+10} + \frac{25}{4x+28}$$

We can now integrate the left hand side.

$$\int \frac{1+x^2}{(x+3)(x+5)(x+7)}dx = \tfrac{5}{4}\ln\left|x+3\right| - \tfrac{13}{2}\ln\left|x+5\right| + \tfrac{25}{4}\ln\left|x+7\right| + C$$

Exercises

1. Evaluate the following by the method partial fraction decomposition.

   (a) $\int \frac{2x+11}{(x+6)(x+5)}dx$

   Solution

   Decompose the fraction:

   $$\frac{2x+11}{(x+6)(x+5)} = \frac{A}{x+6} + \frac{B}{x+5} = \frac{Ax+5A+Bx+6B}{(x+6)(x+5}$$

   Equate coefficients of x:

   $A + B = 2$

   $5A + 6B = 11$

   Solve the system of equations:

   $A = 1, \ B = 1$

   Rewrite the integral and solve:

   $$\int \frac{2x+11}{(x+6)(x+5)}dx = \int \frac{dx}{x+6} + \int \frac{dx}{x+5}$$
   $$= \ln|\mathbf{x+6}| + \ln|\mathbf{x+5}| + \mathbf{C}$$

   (b) $\int \frac{7x^2-5x+6}{(x-1)(x-3)(x-7)}dx$

   Solution

   Decompose the fraction:

   $$\begin{aligned}\frac{7x^2-5x+6}{(x-1)(x-3)(x-7)} &= \frac{A}{x-1} + \frac{B}{x-3} + \frac{C}{x-7} \\ &= \frac{A(x-3)(x-7)+B(x-1)(x-7)+C(x-1)(x-3)}{(x-1)(x-3)(x-7)} \\ &= \frac{A(x^2-10x+21)+B(x^2-8x+7)+C(x^2-4x+3)}{(x-1)(x-3)(x-7)} \\ &= \frac{Ax^2-10Ax+21A+Bx^2-8Bx+7B+Cx^2-4Cx+3C}{(x-1)(x-3)(x-7)}\end{aligned}$$

   Equate coefficients:

   $$A + B + C = 7$$
   $$-10A - 8B - 4C = -5$$
   $$21A + 7B + 3C = 6$$

   Solve the system of equations:

   $$D = \begin{vmatrix} 1 & 1 & 1 \\ -10 & -8 & -4 \\ 21 & 7 & 3 \end{vmatrix} = -24 - 84 - 70 - (-30 - 28 - 168) = 48$$

$$A = \frac{\begin{vmatrix} 7 & 1 & 1 \\ -5 & -8 & -4 \\ 6 & 7 & 3 \end{vmatrix}}{D} = \frac{(7)(-8)(3)-24-35-(-15+(7)(-4)(7)-48)}{48} = \frac{32}{48} = \frac{2}{3}$$

$$B = \frac{\begin{vmatrix} 1 & 7 & 1 \\ -10 & -5 & -4 \\ 21 & 6 & 3 \end{vmatrix}}{D} = \frac{-15+(7)(-4)(21)-60-(-210-24+(-5)(21))}{48} = \frac{-324}{48} = -\frac{27}{4}$$

$$C = \frac{\begin{vmatrix} 1 & 1 & 7 \\ -10 & -8 & -5 \\ 21 & 7 & 6 \end{vmatrix}}{D} = \frac{(-8)(6)+(-5)(21)+(7)(-10)(7)-((-10)(6)+(-5)(7)+(7)(-8)(21))}{48} = \frac{628}{48} = \frac{157}{12}$$

Rewrite the integral and solve:

$$\begin{aligned}\int \frac{7x^2-5x+6}{(x-1)(x-3)(x-7)}dx &= \frac{2}{3}\int \frac{dx}{x-1} - \frac{27}{4}\int \frac{dx}{x-3} + \frac{157}{12}\int \frac{dx}{x-7} \\ &= \mathbf{\frac{2}{3}\ln|x-1| - \frac{27}{4}\ln|x-3| + \frac{157}{12}\ln|x-7| + C}\end{aligned}$$

$Q(x)$ is a product of linear factors some of which are repeated

If $(ax+b)$ appears in the factorisation of $Q(x)$ k-times then instead of writing the piece $\dfrac{A}{ax+b}$ we use the more complicated expression

$$\frac{A_1}{ax+b} + \frac{A_2}{(ax+b)^2} + \frac{A_3}{(ax+b)^3} + \cdots + \frac{A_k}{(ax+b)^k}$$

Example 7.20. *Find* $\displaystyle\int \frac{dx}{(x+1)(x+2)^2}$

Here $P(x) = 1$ and $Q(x) = (x+1)(x+2)^2$ We write

$$\frac{1}{(x+1)(x+2)^2} = \frac{A}{x+1} + \frac{B}{x+2} + \frac{C}{(x+2)^2}$$

Multiply both sides by the denominator $1 = A(x+2)^2 + B(x+1)(x+2) + C(x+1)$

Substitute in three values of x to get 3 equations for the unknown constants,

$$\begin{aligned} x = 0 &\quad 1 = 2^2A + 2B + C \\ x = -1 &\quad 1 = A \\ x = -2 &\quad 1 = -C \end{aligned}$$

so $A = 1$, $B = -1$, $C = -1$ and

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

We can now integrate the left hand side.

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

We now simplify the function with the property of Logarithms.

$$\ln\left|\frac{1}{x+1}\right| - \ln\left|\frac{1}{x+2}\right| + \frac{1}{x+2} + C = \ln\left|\frac{x+2}{x+1}\right| + \frac{1}{x+2} + C$$

Exercise

1. Evaluate $\int \frac{x^2-x+2}{x(x+2)^2}dx$ using the method of partial fractions.

   Solution

   using the method of partial fractions. Decompose the fraction:

$$\begin{aligned}\frac{x^2-x+2}{x(x+2)^2} &= \frac{A}{x} + \frac{B}{x+2} + \frac{C}{(x+2)^2}\\ &= \frac{A(x+2)^2 + Bx(x+2) + Cx}{x(x+2)^2}\\ &= \frac{A(x^2+4x+4) + Bx^2 + 2Bx + Cx}{x(x+2)^2}\\ &= \frac{Ax^2 + 4Ax + 4A + Bx^2 + 2Bx + Cx}{x(x+2)^2}\end{aligned}$$

   Equate the coefficients:

$$\begin{array}{ccccccc} A & + & B & & & = & 1\\ 4A & + & 2B & + & C & = & -1\\ 4A & & & & & = & 2\end{array}$$

   Solve the system of equations:

   $4A = 2 \implies A = \frac{1}{2}$
   $A + B = 1 \implies B = \frac{1}{2}$
   $4A + 2B + C = -1 \implies C = -4$

   Rewrite the integral and solve:

$$\begin{aligned}\int \frac{x^2-x+2}{x(x+2)^2}dx &= \frac{1}{2}\int\frac{dx}{x} + \frac{1}{2}\int\frac{dx}{x+2} - 4\int\frac{dx}{(x+2)^2}\\ &= \mathbf{\frac{1}{2}\ln|x| + \frac{1}{2}\ln|x+2| + \frac{4}{x+2} + C}\end{aligned}$$

$Q(x)$ contains some quadratic pieces which are not repeated

If $ax^2 + bx + c$ appears we use $\dfrac{Ax + B}{ax^2 + bx + c}$.

1. Evaluate the following using the method of partial fractions.

   (a) $\displaystyle\int \frac{2}{(x+2)(x^2+3)}dx$

   Solution

   $\frac{2}{(x+2)(x^2+3)} = \frac{A}{x+2} + \frac{Bx+C}{x^2+3} = \frac{Ax^2+3A+Bx^2+Cx+2Bx+2C}{(x+2)(x^2+3)}$

   Equate the coefficients for each power of x. For x^2:

   $A + B = 0$

   , for x:

   $C + 2B = 0$

   , and for the constant terms:

   $3A + 2C = 2$

   Solve the system of equations however you see fit (Gaussian elimination with back-substitution used here):

   $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 3 & 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & -3 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{7}{2} & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{4}{7} \end{bmatrix}$

   $C = \frac{4}{7}$, $B = -\frac{2}{7}$, $A = \frac{2}{7}$

   So

   $\int \frac{2}{(x+2)(x^2+3)}dx = \frac{2}{7}\int \frac{1}{x+2}dx - \frac{2}{7}\int \frac{x}{x^2+3} + \frac{4}{7}$
   $\int \frac{1}{x^2+3}dx$

   To evaluate the first integral use substitution, letting $u = x + 2, du = dx$.

   To evaluate the second integral use substitution, letting $u = x^2+3, du = 2xdx, dx = \frac{du}{2x}$.

   To evaluate the third integral, use the trigonometric substitution $x = \sqrt{3}\tan(\theta), dx = \sqrt{3}\sec^2(\theta)d\theta$.

$$\begin{aligned}\int \frac{2}{(x+2)(x^2+3)}dx &= \frac{2}{7}\int \frac{1}{x+2}dx - \frac{2}{7}\int \frac{x}{x^2+3} + \frac{4}{7}\int \frac{1}{x^2+3}dx \\ &= \frac{2}{7}\ln|x+2| - \frac{2}{14}\ln|x^2+3| + \frac{4}{7}\int \frac{\sqrt{3}\sec^2(x)d\theta}{3\tan^2(x)+3} \\ &= \mathbf{\frac{2}{7}\ln|x+2| - \frac{1}{7}\ln|x^2+3| + \frac{4}{7\sqrt{3}}\arctan(\frac{x}{\sqrt{3}})}\end{aligned}$$

(b) $\int \frac{dx}{(x+2)(x^2+2)}$
Solution

Decompose the fraction:

$$\begin{aligned}\frac{1}{(x+2)(x^2+2)} &= \frac{A}{x+2} + \frac{Bx+C}{x^2+2} \\ &= \frac{A(x^2+2)+(Bx+C)(x+2)}{(x+2)(x^2+2)} \\ &= \frac{Ax^2+2A+Bx^2+2Bx+Cx+2C}{(x+2)(x^2+2)}\end{aligned}$$

Equate the coefficients:

$$\begin{array}{ccccccc} A & + & B & & & = & 0 \\ & & 2B & + & C & = & 0 \\ 2A & & & + & 2C & = & 1 \end{array}$$

Solve the system of equations:

$D = \begin{vmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 2 & 0 & 2 \end{vmatrix} = 6$

$A = \frac{\begin{vmatrix} 0 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{vmatrix}}{D} = \frac{1}{6}$

$B = \frac{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 2 & 1 & 2 \end{vmatrix}}{D} = -\frac{1}{6}$

$C = \frac{\begin{vmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 2 & 0 & 1 \end{vmatrix}}{D} = \frac{2}{6} = \frac{1}{3}$

Rewrite the integral and solve:

$\int \frac{dx}{(x+2)(x^2+2)} = \frac{1}{6}\int \frac{dx}{x+2} - \frac{1}{6}\int \frac{xdx}{x^2+2} + \frac{1}{3}\int \frac{dx}{x^2+2}$

Making the substitution

$u = x^2+2 \qquad du = 2xdx \qquad dx = \frac{du}{2x}$

in the second integral and

$\frac{x}{\sqrt{2}} = \tan(\theta) \qquad \frac{dx}{\sqrt{2}} = \sec^2(\theta)d\theta \qquad dx = \sqrt{2}\sec^2(\theta)d\theta$

in the third integral, we have

$$\begin{aligned}\int \frac{dx}{(x+2)(x^2+2)} &= \frac{1}{6}\ln|x+2| - \frac{1}{6}\int \frac{du}{2u} + \frac{1}{3}\int \frac{\sqrt{2}\sec^2(\theta)d\theta}{2(\tan^2(\theta)+1)} \\ &= \frac{1}{6}\ln|x+2| - \frac{1}{12}\ln|u| + \frac{\sqrt{2}}{6}\int \frac{\sec^2(\theta)d\theta}{\sec^2(\theta)} \\ &= \frac{1}{6}\ln|x+2| - \frac{1}{12}\ln|x^2+2| + \frac{\sqrt{2}}{6}\int d\theta \\ &= \frac{1}{6}\ln|x+2| - \frac{1}{12}\ln|x^2+2| + \frac{\sqrt{2}}{6}\theta + C \\ &= \mathbf{\frac{1}{6}\ln|x+2| - \frac{1}{12}\ln|x^2+2| + \frac{\sqrt{2}}{6}\arctan(\frac{x}{\sqrt{2}}) + C}\end{aligned}$$

$Q(x)$ **contains some repeated quadratic factors**

If ax^2+bx+c appears k-times then use

$$\frac{A_1x+B_1}{ax^2+bx+c} + \frac{A_2x+B_2}{(ax^2+bx+c)^2} + \frac{A_3x+B_3}{(ax^2+bx+c)^3} + \cdots + \frac{A_kx+B_k}{(ax^2+bx+c)^k}$$

Exercise

1. Evaluate the following using the method of partial fractions.

$$\int \frac{dx}{(x-1)(x^2+1)^2}$$

solution

Decompose the fraction:

$$\begin{aligned}\frac{1}{(x^2+1)^2(x-1)} &= \frac{Ax+B}{x^2+1} + \frac{Cx+D}{(x^2+1)^2} + \frac{E}{x-1} \\ &= \frac{(Ax+B)(x^2+1)(x-1) + (Cx+D)(x-1) + E(x^2+1)^2}{(x^2+1)^2(x-1)} \\ &= \frac{(Ax+B)(x^3-x^2+x-1) + Cx^2 + (D-C)x - D + E(x^4+2x^2+1)}{(x^2+1)^2(x-1)} \\ &= \frac{Ax^4 + (B-A)x^3 + (A-B)x^2 + (B-A)x - B + Cx^2 + (D-C)x - D + Ex^4 + 2Ex^2 + E}{(x^2+1)^2(x-1)}\end{aligned}$$

Equate coefficients:

$$\begin{array}{rcrcrcrcrcl} A & & & & & & & + & E & = & 0 \\ -A & + & B & & & & & & & = & 0 \\ A & - & B & + & C & & & + & 2E & = & 0 \\ -A & + & B & - & C & + & D & & & = & 0 \\ & - & B & & & - & D & + & E & = & 0 \end{array}$$

Solve the system of equations any way you see fit. Here, we'll solve for A by Cramer's rule, then plug in to solve for the other variables. The denominator in Cramer's rule will be

$$d = \begin{vmatrix} 1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 2 \\ -1 & 1 & -1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 1 \end{vmatrix}$$

Expanding across the top row gives

$$d = \begin{vmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 2 \\ 1 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 \end{vmatrix} + \begin{vmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 1 \\ 0 & -1 & 0 & -1 \end{vmatrix}$$

Expanding across the top rows in both matrices gives

$$d = \begin{vmatrix} 1 & 0 & 2 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{vmatrix} - \begin{vmatrix} -1 & 1 & 0 \\ 1 & -1 & 1 \\ -1 & 0 & -1 \end{vmatrix} - \begin{vmatrix} 1 & 1 & 0 \\ -1 & -1 & 1 \\ 0 & 0 & -1 \end{vmatrix}$$

Solve the individual determinants

$$\begin{vmatrix} 1 & 0 & 2 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{vmatrix} = (1)(1)(1) + (0)(0)(1) + (2)(-1)(-1) - (0)(-1)(2) - (1)(0)(-1) - (2)(1)(0)$$

$$= 1 + 2 = 3$$

$$\begin{vmatrix} -1 & 1 & 0 \\ 1 & -1 & 1 \\ -1 & 0 & -1 \end{vmatrix} = (-1)(-1)(-1) + (1)(1)(-1) + (0)(1)(0) - (1)(1)(-1) - (-1)(1)(0) - (0)(-1)(-1)$$

$$= -1 - 1 - (-1) = -1$$

$$\begin{vmatrix} 1 & 1 & 0 \\ -1 & -1 & 1 \\ 0 & 0 & -1 \end{vmatrix} \begin{matrix} 1 & 1 \\ -1 & -1 \\ 0 & 0 \end{matrix} = (1)(-1)(-1) + (1)(1)(0) + (0)(-1)(0) - (1)(-1)(-1) - (1)(1)(0) - (0)(-1)(0)$$

$$= 1 - 1 = 0$$

So

$d = 3 - (-1) - 0 = 4$

Now use Cramer's rule to solve for A :

$$A = \frac{1}{4}\begin{vmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 2 \\ 0 & 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & -1 & 1 \end{vmatrix}$$

Expanding down the first column gives

$$A = \frac{1}{4}\begin{vmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 2 \\ 1 & -1 & 1 & 0 \end{vmatrix}$$

Expanding across the first row gives

$$A = -\frac{1}{4}\begin{vmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{vmatrix}$$

Expanding down the last column gives

$$A = -\frac{1}{4}\begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = -\frac{1}{4}$$

Now that we know A, we can solve for E using the first equation

$$-\frac{1}{4} + E = 0 \implies E = \frac{1}{4}$$

We can solve for B using the second equation and the value of A

$$-(-\frac{1}{4}) + B = 0 \implies B = -\frac{1}{4}$$

We can solve for C using the third equation and the values we've found so far

$$-\frac{1}{4} - (-\frac{1}{4}) + C + 2\frac{1}{4} = 0 \implies C = -\frac{1}{2}$$

We can solve for D using the last equation and the values of B and E

$$-(-\frac{1}{4}) - D + \frac{1}{4} = 1 \implies D = -\frac{1}{2}$$

Finally, we can check our solution using the 4th equation and the values we've found

$$-(-\frac{1}{4}) + (-\frac{1}{4}) - (-\frac{1}{2}) + (-\frac{1}{2}) = 0\checkmark$$

Rewrite the integral and solve

$$\int \frac{dx}{(x^2+1)^2(x-1)} = -\frac{1}{4}\int \frac{xdx}{x^2+1} - \frac{1}{4}\int \frac{dx}{x^2+1} - \frac{1}{2}\int \frac{xdx}{(x^2+1)^2} - \frac{1}{2}\int \frac{dx}{(x^2+1)^2} + \frac{1}{4}\int \frac{dx}{x-1}$$

Let's solve each integral separately. To solve the first, use the substitution

$$u = x^2+1; \qquad du = 2xdx; \qquad dx = \frac{du}{2x}$$

$$\begin{aligned}\int \frac{xdx}{x^2+1} &= \int \frac{du}{2u}\\ &= \frac{1}{2}\ln|u| + C_1\\ &= \frac{1}{2}\ln|x^2+1| + C_1\end{aligned}$$

To solve the second integral, use the substitution

$$x = \tan(\theta); \qquad dx = \sec^2(\theta)d\theta$$

$$\begin{aligned}\int \frac{dx}{x^2+1} &= \int \frac{\sec^2(\theta)d\theta}{\tan^2(\theta)+1}\\ &= \int \frac{\sec^2(\theta)d\theta}{\sec^2(\theta)}\\ &= \int d\theta\\ &= \theta + C_2\\ &= \arctan(x) + C_2\end{aligned}$$

To solve the third integral, use the substitution

$$u = x^2+1; \qquad du = 2xdx; \qquad dx = \frac{du}{2x}$$

$$\int \frac{xdx}{(x^2+1)^2} = \int \frac{du}{2u^2} = -\frac{1}{2u} + C_3 = -\frac{1}{2(x^2+1)} + C_3$$

To solve the fourth integral, use the substitution

$x = \tan(\theta); \qquad dx = \sec^2(\theta)d\theta$

$$\begin{aligned}\int \frac{dx}{(x^2+1)^2} &= \int \frac{\sec^2(\theta)d\theta}{(\tan^2(\theta)+1)^2}\\ &= \int \frac{\sec^2(\theta)d\theta}{(\sec^2(\theta))^2}\\ &= \int \frac{d\theta}{\sec^2(\theta)}\\ &= \int \cos^2(\theta)d\theta\\ &= \int \frac{1+\cos(2\theta)}{2}d\theta\\ &= \int \frac{d\theta}{2} + \frac{1}{2}\int \cos(2\theta)d\theta\\ &= \frac{\theta}{2} + \frac{1}{4}\sin(2\theta) + C_4\\ &= \frac{\theta}{2} + \frac{1}{2}\cos(\theta)\sin(\theta) + C_4\\ &= \frac{1}{2}\arctan(x) + \frac{1}{2}\cos(\theta)\sin(\theta) + C_4\\ &= \frac{1}{2}\arctan(x) + \frac{1}{2}\frac{1}{\sqrt{1+x^2}}\frac{x}{\sqrt{1+x^2}} + C_4\\ &= \frac{1}{2}\arctan(x) + \frac{x}{2(1+x^2)} + C_4\end{aligned}$$

To solve the last integral, use the substitution

$$u = x-1; \qquad du = dx$$

$$\int \frac{dx}{x-1} = \int \frac{du}{u} = \ln|u| + C_5 = \ln|x-1| + C_5$$

Putting it all together, we have

$$\begin{aligned}\int \frac{dx}{(x^2+1)^2(x-1)} &= -\frac{1}{4}(\frac{1}{2}\ln|x^2+1|) - \frac{1}{4}\arctan(x) - \frac{1}{2}(-\frac{1}{2(x^2+1)}) - \frac{1}{2}(\frac{1}{2}\arctan(x) + \frac{x}{2(1+x^2)}) + \frac{1}{4}\ln|x-1| + C\\ &= -\frac{1}{8}\ln|x^2+1| - \frac{1}{4}\arctan(x) + \frac{1}{4(x^2+1)} - \frac{1}{4}\arctan(x) - \frac{x}{4(1+x^2)} + \frac{1}{4}\ln|x-1| + C\\ &= \mathbf{-\frac{1}{2}\arctan(x) + \frac{1-x}{4(x^2+1)} + \frac{1}{8}\ln\left(\frac{(x-1)^2}{x^2+1}\right) + C}\end{aligned}$$

7.9 Tangent Half Angle

Another useful change of variables is the Weierstrass substitution, named after Karl Weierstrass:

$$t = \tan\left(\frac{x}{2}\right)$$

With this transformation, using the double-angle trigonometric identities,

$$\sin(x) = \frac{2t}{1+t^2} \quad \cos(x) = \frac{1-t^2}{1+t^2} \quad \tan(x) = \frac{2t}{1-t^2} \quad dx = \frac{2\ dt}{1+t^2}$$

This transforms a trigonometric integral into an algebraic integral, which may be easier to integrate.

For example, if the integrand is $\dfrac{1}{1+\sin(x)}$ then

$$\int_0^{\frac{\pi}{2}} \frac{dx}{1+\sin(x)} = \int_0^1 \frac{\left(\frac{2\ dt}{1+t^2}\right)}{1+\left(\frac{2t}{1+t^2}\right)} = \int_0^1 \frac{2\ dt}{(t+1)^2}$$

This method can be used to further simplify trigonometric integrals produced by the changes of variables described earlier.
For example, if we are considering the integral

$$I = \int_{-1}^{1} \frac{\sqrt{1-x^2}}{1+x^2} dx$$

we can first use the substitution $x = \sin(\theta)$, which gives

$$I = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\cos^2(\theta)}{1+\sin^2(\theta)} d\theta$$

then use the tan-half-angle substition to obtain

$$I = \int_{-1}^{1} \frac{(1-t^2)^2}{1+6t^2+t^4} \cdot \frac{2\ dt}{1+t^2}$$

In effect, we've removed the square root from the original integrand. We could do this with a single change of variables, but doing it in two steps gives us the opportunity of doing the trigonometric integral another way.

Having done this, we can split the new integrand into partial fractions, and integrate.

$$I = \int_{-1}^{1} \frac{2-\sqrt{2}}{t^2+3-2\sqrt{2}} dt + \int_{-1}^{1} \frac{2+\sqrt{2}}{t^2+3+2\sqrt{2}} dt - \int_{-1}^{1} \frac{2}{1+t^2} dt$$

$$= \frac{4-2\sqrt{2}}{\sqrt{3-2\sqrt{2}}} \cdot \arctan\left(\sqrt{3+2\sqrt{2}}\right) + \frac{4+2\sqrt{2}}{\sqrt{3+2\sqrt{2}}} \cdot \arctan\left(\sqrt{3-2\sqrt{2}}\right) - \pi$$

This result can be further simplified by use of the identities

$$3 \pm 2\sqrt{2} = \left(\sqrt{2} \pm 1\right)^2 \quad \arctan\left(\sqrt{2} \pm 1\right) = \left(\frac{1}{4} \pm \frac{1}{8}\right)\pi$$

ultimately leading to

$$I = \left(\sqrt{2} - 1\right)\pi$$

In principle, this approach will work with any integrand which is the square root of a quadratic multiplied by the ratio of two polynomials. However, it should not be applied automatically.

E.g., in this last example, once we deduced

$$I = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\cos^2(\theta)}{1 + \sin^2(\theta)} d\theta$$

we could have used the double angle formula, since this contains only even powers of cos and sin. Doing that gives

$$I = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1 + \cos(2\theta)}{3 - \cos(2\theta)} d\theta = \frac{1}{2} \int_{-\pi}^{\pi} \frac{1 + \cos(\phi)}{3 - \cos(\phi)} d\phi$$

Using tan-half-angle on this new, simpler, integrand gives

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \frac{1}{1 + 2t^2} \cdot \frac{dt}{1 + t^2} \\ &= \int_{-\infty}^{\infty} \frac{2\ dt}{1 + 2t^2} - \int_{-\infty}^{\infty} \frac{dt}{1 + t^2} \end{aligned}$$

This can be integrated on sight to give

$$I = \frac{4}{\sqrt{2}} \cdot \frac{\pi}{2} - 2\frac{\pi}{2} = \left(\sqrt{2} - 1\right)\pi$$

This is the same result as before, but obtained with less algebra, which shows why it is best to look for the most straightforward methods at every stage.

A more direct way of evaluating the integral I is to substitute $t = \tan(\theta)$ right from the start, which will directly bring us to the line

$$I = \int_{-\infty}^{\infty} \frac{1}{1 + 2t^2} \cdot \frac{dt}{1 + t^2}$$

above. More generally, the substitution $t = \tan(x)$ gives us

$$\sin(x) = \frac{t}{\sqrt{1 + t^2}} \quad \cos(x) = \frac{1}{\sqrt{1 + t^2}} \quad dx = \frac{dt}{1 + t^2}$$

so this substitution is the preferable one to use if the integrand is such that all the square roots would disappear after substitution, as is the case in the above integral.

Example 7.21. *Using the trigonometric substitution* $t = a\tan(x)$, *then* $dt = a\sec^2(x)dx$ *and* $\sqrt{t^2+a^2} = a\sec(x)$ *when* $-\frac{\pi}{2} < x < \frac{\pi}{2}$. *So,*

$$\begin{aligned}\int \frac{dt}{(t^2+a^2)\sqrt{t^2+a^2}} &= \int \frac{a\sec^2(x)}{a^3\sec^3(x)}dx \\ &= \frac{1}{a^2}\int \cos(x)dx \\ &= \frac{\sin(x)}{a^2} + C \\ &= \frac{1}{a^2}\cdot\frac{a\tan(x)}{a\sec(x)} + C \\ &= \frac{t}{a^2\sqrt{t^2+a^2}} + C\end{aligned}$$

Alternate Method

In general, to evaluate integrals of the form

$$\int \frac{A + B\cos(x) + C\sin(x)}{a + b\cos(x) + c\sin(x)}\, dx,$$

it is extremely tedious to use the aforementioned "tan half angle" substitution directly, as one easily ends up with a rational function with a 4th degree denominator. Instead, we may first write the numerator as

$$A + B\cos(x) + C\sin(x) \equiv p\Big(a + b\cos(x) + c\sin(x)\Big) + q\frac{d}{dx}\Big(a + b\cos(x) + c\sin(x)\Big) + r.$$

Then the integral can be written as

$$\int \left(p + q\cdot\frac{\frac{d}{dx}\big(a + b\cos(x) + c\sin(x)\big)}{a + b\cos(x) + c\sin(x)} + \frac{r}{a + b\cos(x) + c\sin(x)}\right)dx$$

which can be evaluated much more easily.

Example 7.22. *Evaluate* $\int \frac{\cos(x)+2}{\cos(x)+\sin(x)}dx$,

Let

$$\cos(x) + 2 \equiv p\Big(\cos(x) + \sin(x)\Big) + q\frac{d}{dx}\Big(\cos(x) + \sin(x)\Big) + r.$$

Then

$$\begin{aligned}\cos(x) + 2 &\equiv p\Big(\cos(x) + \sin(x)\Big) + q\Big(\cos(x) - \sin(x)\Big) + r \\ \cos(x) + 2 &\equiv (p+q)\cos(x) + (p-q)\sin(x) + r\end{aligned}$$

Comparing coefficients of cos(x), sin(x) and the constants on both sides, we obtain

$$\begin{cases} p+q &= 1 \\ p-q &= 0 \\ r &= 2 \end{cases}$$

yielding $p = q = 1/2, r = 2$. Substituting back into the integrand,

$$\int \frac{\cos(x)+2}{\cos(x)+\sin(x)}dx = \int \frac{dx}{2} + \frac{1}{2}\int \frac{d\big(\cos(x)+\sin(x)\big)}{\cos(x)+\sin(x)} + \int \frac{2}{\cos(x)+\sin(x)}dx$$

The last integral can now be evaluated using the "tan half angle" substitution described above, and we obtain

$$\int \frac{2}{\cos(x)+\sin(x)}dx = \sqrt{2}\ln\left|\frac{\tan\left(\frac{x}{2}\right)-1+\sqrt{2}}{\tan\left(\frac{x}{2}\right)-1-\sqrt{2}}\right| + C.$$

The original integral is thus

$$\int \frac{\cos(x)+2}{\cos(x)+\sin(x)}dx = \frac{x+\ln\left|\cos(x)+\sin(x)\right|}{2} + \sqrt{2}\ln\left|\frac{\tan\left(\frac{x}{2}\right)-1+\sqrt{2}}{\tan\left(\frac{x}{2}\right)-1-\sqrt{2}}\right| + C.$$

7.10 Irrational Functions

Integration of irrational functions is more difficult than rational functions, and many cannot be done. However, there are some particular types that can be reduced to rational forms by suitable substitutions.

1. **Type 1:** Integrand contains $\sqrt[n]{\dfrac{ax+b}{cx+d}}$

   Use the substitution $u = \sqrt[n]{\dfrac{ax+b}{cx+d}}$.

   Example 7.23. *Find* $\displaystyle\int \frac{1}{x}\sqrt{\frac{1-x}{x}}\,dx$.

   Find $\displaystyle\int \frac{x}{\sqrt[3]{ax+b}}\,dx$.

2. **Type 2:** Integral is of the form $\displaystyle\int \frac{Px+Q}{\sqrt{ax^2+bx+c}}\,dx$ Write $Px+Q$ as $Px+Q = p\cdot\dfrac{d[ax^2+bx+c]}{dx}$

   Example 7.24. *Find* $\displaystyle\int \frac{4x-1}{\sqrt{5-4x-x^2}}\,dx$.

3. **Type 3** Integrand contains $\sqrt{a^2-x^2}$, $\sqrt{a^2+x^2}$ or $\sqrt{x^2-a^2}$

   This was discussed in "trigonometric substitutions above". Here is a summary:

   For $\sqrt{a^2-x^2}$, use $x = a\sin(\theta)$.
   For $\sqrt{a^2+x^2}$, use $x = a\tan(\theta)$.
   For $\sqrt{x^2-a^2}$, use $x = a\sec(\theta)$.

4. **Type 4** Integral is of the form $\displaystyle\int \frac{dx}{(px+q)\sqrt{ax^2+bx+c}}$

   Use the substitution $u = \dfrac{1}{px+q}$.

   Example 7.25. *Find* $\displaystyle\int \frac{dx}{(1+x)\sqrt{3+6x+x^2}}$.

5. **Type 5** Other rational expressions with the irrational function $\sqrt{ax^2+bx+c}$

   If $a > 0$, we can use $u = \sqrt{ax^2+bx+c} \pm \sqrt{a}x$.
   If $c > 0$, we can use $u = \dfrac{\sqrt{ax^2+bx+c} \pm \sqrt{c}}{x}$.
   If ax^2+bx+c can be factored as $a(x-\alpha)(x-\beta)$, we can use $u = \sqrt{\dfrac{a(x-\alpha)}{x-\beta}}$.
   If $a < 0$ and ax^2+bx+c can be factored as $-a(\alpha-x)(x-\beta)$, we can use $x = \alpha\cos^2(\theta) + \beta\sin^2(\theta)$

8 Improper Integrals

The definition of a definite integral:

$\int\limits_a^b f(x)dx$ requires the interval $[a, b]$ be finite. The Fundamental Theorem of Calculus requires that f be continuous on $[a, b]$. In this section, you will be studying a method of evaluating integrals that fail these requirements - either because their limits of integration are infinite, or because a finite number of discontinuities exist on the interval $[a, b]$. Integrals that fail either of these requirements are improper integrals. (If you are not familiar with L'Hopital's rule, it is a good idea to review it before reading this section.)

8.1 Improper Integrals with Infinite Limits of Integration

Consider the integral

$$\int\limits_1^\infty \frac{dx}{x^2}$$

Assigning a finite upper bound b in place of infinity gives

$$\lim_{b\to\infty}\int\limits_1^b \frac{dx}{x^2} = \lim_{b\to\infty}\left(\frac{1}{1}-\frac{1}{b}\right) = \lim_{b\to\infty}\left(1-\frac{1}{b}\right) = 1$$

This improper integral can be interpreted as the area of the unbounded region between $f(x) = \frac{1}{x^2}$, $y = 0$ (the x-axis), and $x = 1$.

Definition

1. Suppose $\int_a^b f(x)dx$ exists for all $b \geq a$. Then we define

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

as long as this limit exists and is finite. If it does exist we say the integral is convergent and otherwise we say it is divergent.

2. Similarly if $\int_a^b f(x)dx$ exists for all $a \leq b$ we define

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

3. Finally suppose c is a fixed real number and that $\int_{-\infty}^c f(x)dx$ and $\int_c^\infty f(x)dx$ are both convergent. Then we define

$$\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^c f(x)dx + \int_c^\infty f(x)dx$$

Example 8.1. *Convergent Improper Integral*
We claim that

$$\int_0^\infty e^{-x}dx = 1$$

To do this we calculate

$$\begin{aligned}\int_0^\infty e^{-x}dx &= \lim_{b\to\infty} \int_0^b e^{-x}dx \\ &= \lim_{b\to\infty} (-e^{-x})\Big|_0^b \\ &= \lim_{b\to\infty} \left(-e^{-b} + 1\right) \\ &= 1\end{aligned}$$

Example 8.2. *Divergent Improper Integral*
We claim that the integral $\int_1^\infty \frac{dx}{x}$ *diverges.*

This follows as

$$\begin{aligned}\int_1^\infty \frac{dx}{x} &= \lim_{b\to\infty}\int_1^b \frac{dx}{x}\\ &= \lim_{b\to\infty} \ln(x)\Big|_1^b\\ &= \lim_{b\to\infty} (\ln(b)-0)\\ &= \infty\end{aligned}$$

Therefore $\int_1^\infty \frac{dx}{x}$ diverges.

Example 8.3. *Improper Integral*
Find $\int_0^\infty x^2e^{-x}dx$.

To calculate the integral use integration by parts twice to get

$$\begin{aligned}\int_0^b x^2e^{-x}dx &= (-x^2e^{-x})\Big|_0^b + 2\int_0^b xe^{-x}dx\\ &= -b^2e^{-b} + 2\left((-xe^{-x})\Big|_0^b + \int_0^b e^{-x}dx\right)\\ &= -b^2e^{-b} + 2\left(-be^{-b} - (e^{-x})\Big|_0^b\right)\\ &= -b^2e^{-b} + 2(-be^{-b} - e^{-b} + 1)\end{aligned}$$

Now $\lim_{b\to\infty} e^{-b} = 0$ and because exponentials overpower polynomials, we see that $\lim_{b\to\infty} b^2e^{-b} = 0$ and $\lim_{b\to\infty} be^{-b} = 0$ as well. Hence,

$$\int_0^\infty x^2e^{-x}dx = \lim_{b\to\infty}\int_0^b x^2e^{-x}dx = 0 + 2(0-0+1) = 2$$

Example 8.4. *Powers Show* $\int_1^\infty \frac{dx}{x^p} = \begin{cases}\frac{1}{p-1}, & \text{if } p>1\\ \text{diverges}, & \text{if } p\le 1\end{cases}$

If $p \neq 1$ then

$$\int_1^\infty \frac{dx}{x^p} = \lim_{b\to\infty}\int_1^b x^{-p}dx$$
$$= \lim_{b\to\infty}\left(\frac{x^{-p+1}}{-p+1}\right)\Bigg|_1^b$$
$$= -\frac{1}{1-p}\lim_{b\to\infty}\left(b^{-p+1}-1\right)$$
$$= \begin{cases} \frac{1}{p-1}, & \text{if } p>1 \\ \text{diverges}, & \text{if } p<1 \end{cases}$$

Example 8.5. *Determine weather the integrals are convergent or divergent* $\int_1^\infty \frac{1}{x}dx$

Solution:

$$\int_1^\infty \frac{1}{x}dx = \lim_{t\to\infty}\int_1^t \frac{1}{x}dx$$
$$= \lim_{t\to\infty} \ln x\Big|_1^t$$
$$= \lim_{t\to\infty}[\ln t]$$
$$= \infty$$

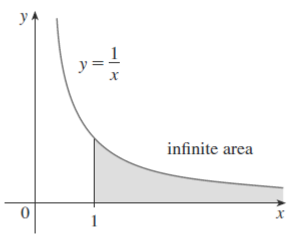


Figure 8.1: This is appropriate because $\int_1^\infty \frac{1}{x}dx$ is the limit as $\lim_{t\to\infty}$ of the area under the graph of f from 1 to t.

$\therefore \int_1^\infty \frac{1}{x}dx$ is divergent.

Example 8.6. *Evaluate* $\int_{-\infty}^0 e^{2x}dx$

$$\begin{aligned}\int_{-\infty}^{0} e^{2x}dx &= \lim_{t\to-\infty}\int_{t}^{0} e^{2x}dx\\ &= \lim_{t\to-\infty}\frac{1}{2}e^{2x}\Big|_{t}^{0}\\ &= \lim_{t\to-\infty}\frac{1}{2}[1-e^{2t}]\\ &= \frac{1}{2}\end{aligned}$$

$\therefore \int_{-\infty}^{0} e^{2x}dx$ is convergent.

Example 8.7. *Evaluate* $\int_{-\infty}^{\infty}\frac{1}{1+x^2}dx$

Solution:

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{1}{1+x^2}dx &= \int_{-\infty}^{0}\frac{1}{1+x^2}dx + \int_{0}^{\infty}\frac{1}{1+x^2}dx\\ &= \lim_{t\to-\infty}\int_{t}^{0}\frac{1}{1+x^2}dx + \lim_{t\to\infty}\int_{0}^{t}\frac{1}{1+x^2}dx\\ &= \lim_{t\to-\infty}[\tan^{-1}x]\Big|_{t}^{0} + \lim_{t\to\infty}[\tan^{-1}x]\Big|_{0}^{t}\\ &= \lim_{t\to-\infty}[0-\tan^{-1}t] + \lim_{t\to\infty}[\tan^{-1}t]\\ &= \frac{\pi}{2}+\frac{\pi}{2}\\ &= \pi\end{aligned}$$

$\therefore \int_{-\infty}^{\infty}\frac{1}{1+x^2}dx$ is convergent.

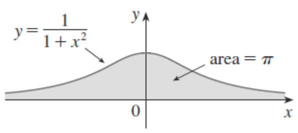


Figure 8.2: This is appropriate $\int_{-\infty}^{\infty}\frac{1}{1+x^2}dx$ because is the limit as $\lim_{t\to-\infty}$ and $\lim_{t\to-\infty}$ of the area under the graph of from to (∞,∞).

Notice that we had to assume that $p\neq 1$ to avoid dividing by 0. However the $p=1$ case was done in a previous example.

8.2 Improper Integrals with a Finite Number Discontinuities

First we give a definition for the integral of functions which have a discontinuity at one point.

Definition 8.1. *If f is continuous on the interval $[a, b)$ and is discontinuous at b, we define*

$$\int_a^b f(x)dx = \lim_{c \to b^-} \int_a^c f(x)dx$$

If the limit in question exists we say the integral converges and otherwise we say it diverges.

Similarly if f is continuous on the interval $(a, b]$ and is discontinuous at a, we define

$$\int_a^b f(x)dx = \lim_{c \to a^+} \int_c^b f(x)dx$$

Finally suppose f has an discontinuity at a point $c \in (a, b)$ and is continuous at all other points in $[a, b]$. If $\int_a^c f(x)dx$ and $\int_c^b f(x)dx$ converge we define

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

Example 8.8. *Show* $\int_0^1 \frac{dx}{x^p} = \begin{cases} \frac{1}{1-p}, & \text{if } p < 1 \\ \text{diverges}, & \text{if } p \geq 1 \end{cases}$

If $p \neq 1$ then

$$\begin{aligned} \int_0^1 \frac{dx}{x^p} &= \lim_{a \to 0^+} \int_a^1 x^{-p} dx \\ &= \lim_{a \to 0^+} \left(\frac{x^{-p+1}}{-p+1} \right) \Bigg|_a^1 \\ &= -\frac{1}{1-p} \lim_{a \to 0^+} \left(1 - a^{-p+1} \right) \\ &= \begin{cases} \frac{1}{1-p}, & \text{if } p < 1 \\ \text{diverges}, & \text{if } p > 1 \end{cases} \end{aligned}$$

Notice that we had to assume that $p \neq 1$ do avoid dividing by 0. So instead we do the $p = 1$ case separately,

$$\int_0^1 \frac{dx}{x} = \lim_{a \to 0^+} \left[\ln \left(|x| \right) \Big|_a^1 \right] = \lim_{a \to 0^+} \left[-\ln(a) \right]$$

which diverges.

Example 8.9. *The integral* $\int\limits_{-1}^{3} \frac{dx}{x-2}$ *is improper because the integrand is not continuous at* $x = 2$ *However had we not noticed that we might have been tempted to apply the fundamental theorem of calculus and conclude that it equals*

$$\ln\Big(|x-2|\Big)\Big|_{-1}^{3} = \ln(5) - \ln(3) = \ln\left(\tfrac{5}{3}\right)$$

which is not correct. In fact the integral diverges since

$$\begin{aligned}
\int\limits_{-1}^{3} \frac{dx}{x-2} &= \lim_{b\to 2^-} \int\limits_{-1}^{b} \frac{dx}{x-2} + \lim_{a\to 2^+} \int\limits_{a}^{3} \frac{dx}{x-2} \\
&= \lim_{b\to 2^-} \ln\Big(|x-2|\Big)\Big|_{-1}^{b} + \lim_{a\to 2^+} \ln\Big(|x-2|\Big)\Big|_{a}^{3} \\
&= \lim_{b\to 2^-} \Big[\ln(2-b) - \ln(3)\Big] + \lim_{a\to 2^+} \Big[\ln(1) - \ln(a-2)\Big] \\
&= \lim_{b\to 2^-} \Big[\ln(2-b)\Big] - \ln(3) + \lim_{a\to 2^+} \Big[-\ln(a-2)\Big]
\end{aligned}$$

and $\lim\limits_{b\to 2^-} \Big[\ln(2-b)\Big]$ *and* $\lim\limits_{a\to 2^+} \Big[-\ln(a-2)\Big]$ *both diverge.*

We can also give a definition of the integral of a function with a finite number of discontinuities.

Definition 8.2. *Suppose* f *is continuous on* $[a, b]$ *except at points* $c_1 < c_2 < \ldots < c_n$ *in* $[a, b]$. *We define* $\int\limits_{a}^{b} f(x)dx = \int\limits_{a}^{c_1} f(x)dx + \int\limits_{c_1}^{c_2} f(x)dx + \int\limits_{c_2}^{c_3} f(x)dx + \cdots + \int\limits_{c_{n-1}}^{c_n} f(x)dx + \int\limits_{c_n}^{b} f(x)dx$ *as long as each integral on the right converges.*

Notice that by combining this definition with the definition for improper integrals with infinite endpoints, we can define the integral of a function with a finite number of discontinuities with one or more infinite endpoints.

General Integration Table

Integration is the basic operation in integral calculus. While differentiation has easy rules by which the derivative of a complicated function can be found by differentiating its simpler component functions, integration does not, so tables of known integrals are often useful. This page lists some of the most common anti derivatives.

Integrals with a singularity

When there is a singularity in the function being integrated such that the antiderivative becomes undefined or at some point (the singularity), then C does not need to be the same on both sides of the singularity. The forms below normally assume the Cauchy principal value around a singularity in the value of C but this is not in general necessary. For instance in

$$\int \frac{1}{x}\, dx = \ln |x| + C$$

there is a singularity at 0 and the antiderivative becomes infinite there. If the integral above would be used to compute a definite integral between -1 and 1, one would get the wrong answer 0. This however is the Cauchy principal value of the integral around the singularity. If the integration is done in the complex plane the result depends on the path around the origin, in this case the singularity contributes $-i\pi$ when using a path above the origin and $i\pi$ for a path below the origin. A function on the real line could use a completely different value of C on either side of the origin as in:

$$\int \frac{1}{x}\, dx = \ln |x| + \begin{cases} A & \text{if } x > 0; \\ B & \text{if } x < 0. \end{cases}$$

Rational functions

More integrals: List of integrals of rational functions These rational functions have a non-integrable singularity at 0 for $a \leq -1$.

$$\int k\, dx = kx + C$$

$$\int x^a\, dx = \frac{x^{a+1}}{a+1} + C \qquad (\text{for } a \neq -1)$$

(Cavalieri's quadrature formula)

$$\int (ax+b)^n\, dx = \frac{(ax+b)^{n+1}}{a(n+1)} + C \qquad (\text{for } n \neq -1)$$

$$\int \frac{1}{x}\, dx = \ln |x| + C$$

More generally,

$$\int \frac{1}{x}\, dx = \begin{cases} \ln |x| + C^- & x < 0 \\ \ln |x| + C^+ & x > 0 \end{cases}$$

$$\int \frac{c}{ax+b}\, dx = \frac{c}{a} \ln |ax+b| + C$$

Exponential functions

More integrals: List of integrals of exponential functions

$$\int e^{ax}\,dx = \frac{1}{a}e^{ax} + C$$
$$\int f'(x)e^{f(x)}\,dx = e^{f(x)} + C$$
$$\int a^x\,dx = \frac{a^x}{\ln a} + C$$

Logarithms

More integrals: List of integrals of logarithmic functions

$$\int \ln x\,dx = x\ln x - x + C$$
$$\int \log_a x\,dx = x\log_a x - \frac{x}{\ln a} + C$$

Trigonometric functions

More integrals: List of integrals of trigonometric functions

$$\int \sin x\,dx = -\cos x + C$$
$$\int \cos x\,dx = \sin x + C$$
$$\int \tan x\,dx = -\ln|\cos x| + C = \ln|\sec x| + C$$
$$\int \cot x\,dx = \ln|\sin x| + C$$
$$\int \sec x\,dx = \ln|\sec x + \tan x| + C$$

(See Integral of the secant function. This result was a well-known conjecture in the 17th century.)

$$\int \csc x\,dx = -\ln|\csc x + \cot x| + C$$
$$\int \sec^2 x\,dx = \tan x + C$$
$$\int \csc^2 x\,dx = -\cot x + C$$
$$\int \sec x\ \tan x\,dx = \sec x + C$$
$$\int \csc x\ \cot x\,dx = -\csc x + C$$
$$\int \sin^2 x\,dx = \frac{1}{2}\left(x - \frac{\sin 2x}{2}\right) + C = \frac{1}{2}(x - \sin x\cos x) + C$$
$$\int \cos^2 x\,dx = \frac{1}{2}\left(x + \frac{\sin 2x}{2}\right) + C = \frac{1}{2}(x + \sin x\cos x) + C$$
$$\int \sec^3 x\,dx = \frac{1}{2}\sec x\tan x + \frac{1}{2}\ln|\sec x + \tan x| + C$$

(see integral of secant cubed)

$$\int \sin^n x\,dx = -\frac{\sin^{n-1} x\cos x}{n} + \frac{n-1}{n}\int \sin^{n-2} x\,dx$$
$$\int \cos^n x\,dx = \frac{\cos^{n-1} x\sin x}{n} + \frac{n-1}{n}\int \cos^{n-2} x\,dx$$

Inverse trigonometric functions

More integrals: List of integrals of inverse trigonometric functions

$$\int \arcsin x\,dx = x\arcsin x + \sqrt{1-x^2} + C,\ \text{for } |x| \le +1$$
$$\int \arccos x\,dx = x\arccos x - \sqrt{1-x^2} + C,\ \text{for } |x| \le +1$$
$$\int \arctan x\,dx = x\arctan x - \frac{1}{2}\ln|1+x^2| + C,\ \text{for all real } x$$
$$\int \operatorname{arccot} x\,dx = x\operatorname{arccot} x + \frac{1}{2}\ln|1+x^2| + C,\ \text{for all real } x$$
$$\int \operatorname{arcsec} x\,dx = x\operatorname{arcsec} x - \ln\left|x\left(1+\sqrt{1-x^{-2}}\right)\right| + C,\ \text{for } |x| \ge 1$$
$$\int \operatorname{arccsc} x\,dx = x\operatorname{arccsc} x + \ln\left|x\left(1+\sqrt{1-x^{-2}}\right)\right| + C,\ \text{for } |x| \ge 1$$

Hyperbolic functions

More integrals: List of integrals of hyperbolic functions

$$\int \sinh x\,dx = \cosh x + C$$
$$\int \cosh x\,dx = \sinh x + C$$
$$\int \tanh x\,dx = \ln\cosh x + C$$
$$\int \coth x\,dx = \ln|\sinh x| + C, \text{ for } x \neq 0$$
$$\int \text{sech } x\,dx = \arctan(\sinh x) + C$$
$$\int \text{csch } x\,dx = \ln\left|\tanh\frac{x}{2}\right| + C, \text{ for } x \neq 0$$

subsectionInverse hyperbolic functions More integrals: List of integrals of inverse hyperbolic functions

$$\int \text{arsinh } x\,dx = x \text{ arsinh } x - \sqrt{x^2+1} + C, \text{ for all real } x$$
$$\int \text{arcosh } x\,dx = x \text{ arcosh } x - \sqrt{x^2-1} + C, \text{ for } x \geq 1$$
$$\int \text{artanh } x\,dx = x \text{ artanh } x + \frac{\ln(1-x^2)}{2} + C, \text{ for } |x| < 1$$
$$\int \text{arcoth } x\,dx = x \text{ arcoth } x + \frac{\ln(x^2-1)}{2} + C, \text{ for } |x| > 1$$
$$\int \text{arsech } x\,dx = x \text{ arsech } x + \arcsin x + C, \text{ for } 0 < x \leq 1$$
$$\int \text{arcsch } x\,dx = x \text{ arcsch } x + |\text{ arsinh } x| + C, \text{ for } x \neq 0$$

Products of functions proportional to their second derivatives

$$\int \cos ax\, e^{bx}\,dx = \frac{e^{bx}}{a^2+b^2}(a\sin ax + b\cos ax) + C$$
$$\int \sin ax\, e^{bx}\,dx = \frac{e^{bx}}{a^2+b^2}(b\sin ax - a\cos ax) + C$$
$$\int \cos ax\ \cosh bx\,dx = \frac{1}{a^2+b^2}(a\sin ax\ \cosh bx + b\cos ax\ \sinh bx) + C$$
$$\int \sin ax\ \cosh bx\,dx = \frac{1}{a^2+b^2}(b\sin ax\ \sinh bx - a\cos ax\ \cosh bx) + C$$

CHAPTER SIX

9 Application of Integration

10 Application of Integration

10.1 Area

We met areas under curves earlier in the Integration section, but here we develop the concept further.

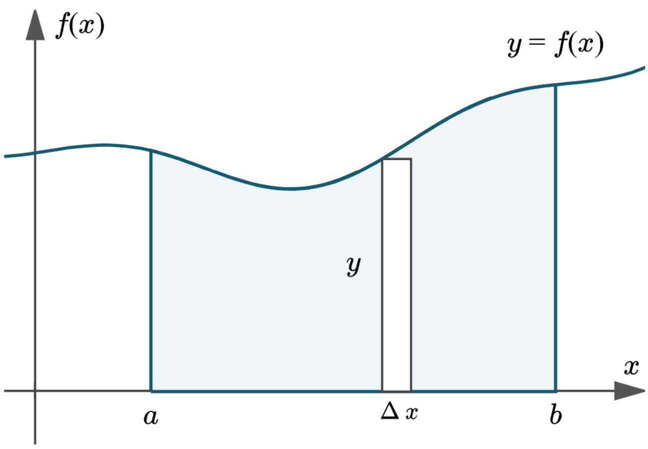


Figure 10.1: The curve $y = f(x)$, completely above x-axis. Shows a "typical" rectangle, Δx wide and y high.

Finding the area between two curves, usually given by two explicit functions, is often useful in calculus.

In general the rule for finding the area between two curves is

$$A = A_{\text{top}} - A_{\text{bottom}}$$

or

If f(x) is the upper function and g(x) is the lower function

$$A = \int_a^b \Big(f(x) - g(x)\Big)dx$$

This is true whether the functions are in the first quadrant or not.

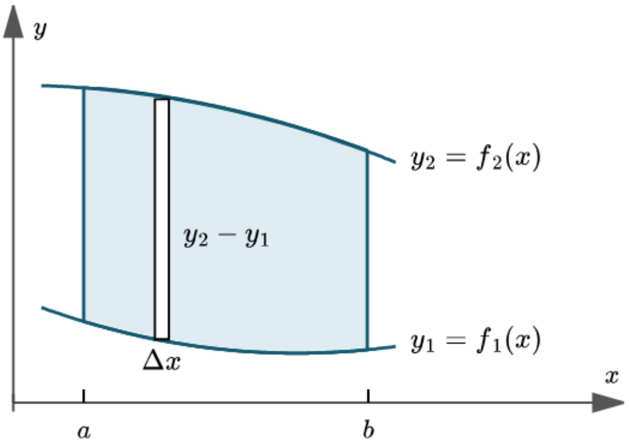


Figure 10.2: Area bounded by the curves y_1 and y_2, & the lines $x = a$ and $x = b$, including a typical rectangle.

10.1.1 Area between two curves

Suppose we are given two functions $y_1 = f(x)$ and $y_2 = g(x)$ and we want to find the area between them on the interval $[a, b]$. Also assume that $f(x) \geq g(x)$ for all x on the interval $[a, b]$. Begin by partitioning the interval $[a, b]$ into n equal subintervals each having a length of $\Delta x = \dfrac{b-a}{n}$. Next choose any point in each subinterval, x_i^*. Now we can 'create' rectangles on each interval. At the point x_i*, the height of each rectangle is $f(x_i^*) - g(x_i^*)$ and the width is Δx. Thus the area of each rectangle is $\Big[f(x_i^*) - g(x_i^*)\Big]\Delta x$. An approximation of the area, A, between the two curves is

$$A := \sum_{i=1}^{n} \Big[f(x_i^*) - g(x_i^*)\Big]\Delta x.$$

Now we take the limit as n approaches infinity and get

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

which gives the exact area. Recalling the definition of the definite integral we notice that

$$A = \int_a^b \Big(f(x) - g(x)\Big)dx.$$

This formula of finding the area between two curves is sometimes known as applying integration with respect to the x-axis since the rectangles used to approximate the area have their bases lying parallel to the x-axis. It will be most useful when the two functions are of the form $y_1 = f(x)$ and $y_2 = g(x)$. Sometimes however, one may find it simpler to integrate with respect to the y-axis. This occurs when integrating with respect to the x-axis would result in more than one integral to be evaluated. These functions take the form $x_1 = f(y)$

and $x_2 = g(y)$ on the interval $[c, d]$. Note that $[c, d]$ are values of y. The derivation of this case is completely identical. Similar to before, we will assume that $f(y) \geq g(y)$ for all y on $[c, d]$. Now, as before we can divide the interval into n subintervals and create rectangles to approximate the area between $f(y)$ and $g(y)$. It may be useful to picture each rectangle having their 'width', Δy, parallel to the y-axis and 'height', $f(y_i^*) - g(y_i^*)$ at the point y_i^*, parallel to the x-axis. Following from the work above we may reason that an approximation of the area, A, between the two curves is

$$A := \sum_{i=1}^{n} \Big[f(y_i^*) - g(y_i^*) \Big] \Delta y.$$

As before, we take the limit as n approaches infinity to arrive at

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} \Big[f(y_i^*) - g(y_i^*) \Big] \Delta y,$$

which is nothing more than a definite integral, so

$$A = \int_{c}^{d} \Big(f(y) - g(y) \Big) dy.$$

Regardless of the form of the functions, we basically use the same formula.

Example 10.1. *Find the area of the region that is enclosed between the curves $y = x^2$ and $y = x + 6$.*

Solution.

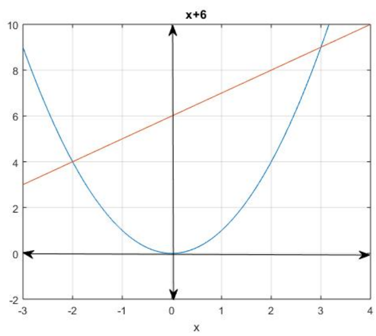


Figure 10.3

A sketch of the region (Figure 10.3) shows that the lower boundary is $y = x^2$ and the upper boundary is $y = x + 6$. At the endpoints of the region, the upper a boundaries have the samey-coordinates; thus, to find the endpoints we equate

$y = x^2$ and $y = x + 6$
This yields

$$x^2 = x + 6 \text{ or } x^2 - x - 6 = 0 \text{ or } (x+3)(x-2) = 0$$

from which we obtain $x = -3$ and $x = 2$

$$A = \int_{-2}^{3} [(x+6) - x^2]dx = \frac{125}{6}$$

Example 10.2. *Find the area of the region enclosed by $x = y^2$ and $y = x - 2$.*

Solution.

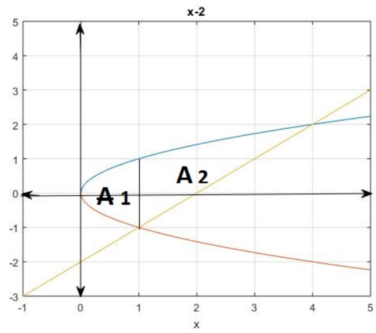


Figure 10.4

To determine the appropriate boundaries of the region, we need to know where the curves $x = y^2$ and $y = x - 2$ intersect. The figure 10.4 show as the region enclosed by this two curves. Ans. $A_1 = \frac{4}{3}$ $A_2 = \frac{19}{6}$

$$A = \frac{9}{2}$$

Example 10.3. *Find the area between the curves $y = x^2 + 5x$ and $y = 3 - x^2$ between $x = -2$ and $x = 0$.*

Solution

From the graph, we see that $y = 3 - x^2$ is above $y = x^2 + 5x$ in the region of interest, so we'll use:
$y_2 = 3 - x^2$, and

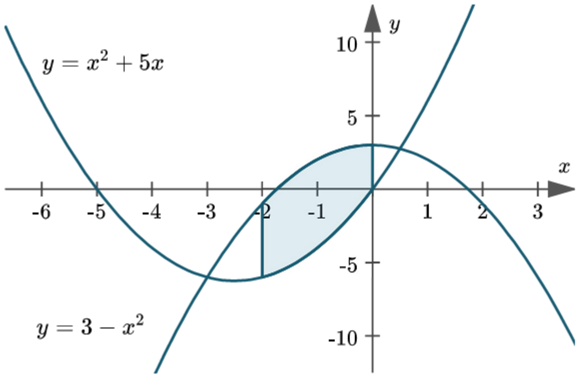


Figure 10.5: Graphs of $y = x^2 + 5x$ and $y = 3 - x^2$, showing the portion between $-2 < x < 0$.

$y_1 = x^2 + 5x$. So we need to find:

$$\begin{aligned} \text{Area} &= \int_a^b (y_2 - y_1)\,dx \\ &= \int_{-2}^{0} \left[\left(3 - x^2\right) - \left(x^2 + 5x\right)\right] dx \\ &= \int_{-2}^{0} \left[\left(-2x^2 - 5x + 3\right)\right] dx \\ &= \left[-\frac{2}{3}x^3 - \frac{5}{2}x^2 + 3x\right]_{-2}^{0} \\ &= 0 - \left[\frac{16}{3} - 10 - 6\right] \\ &= 10\frac{2}{3} \text{ sq units} \end{aligned}$$

Some of the shaded area is above the x-axis and some of it is below. Don't worry about taking absolute value - the formula takes care of that automatically.

Example 10.4. *Find the area bounded by the curves* $y = x^2$, $y = 2 - x$ *and* $y = 1$.

Solution

We take horizontal elements in this case. So we need to solve $y = x^2$ for x:

$$x = \pm\sqrt{y}$$

We need the left hand portion, so $x = -\sqrt{y}$.
Notice that $x = 2 - y$ is to the right of $x = -\sqrt{y}$
so we choose $x_2 = 2 - y$ and $x_1 = -\sqrt{y}$.
The intersection of the graphs occurs at $(-2, 4)$ and $(1, 1)$.

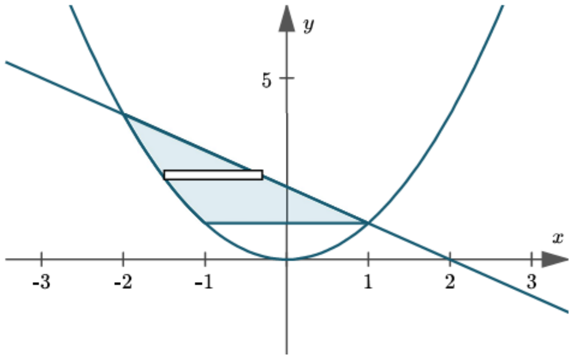


Figure 10.6: Area bounded by $y = x^2$, $y = 2 - x$ and $y = 1$, including a typical rectangle.

So we have: $c = 1$ and $d = 4$.

$$\begin{aligned}
\text{Area} &= \int_c^d (x_2 - x_1)\,dy \\
&= \int_1^4 ([2 - y] - [-\sqrt{y}])\,dy \\
&= \int_1^4 (2 - y + \sqrt{y})\,dy \\
&= \left[2y - \frac{y^2}{2} + \frac{2}{3}y^{3/2}\right]_1^4 \\
&= \left(\frac{16}{3}\right) - \left(\frac{13}{6}\right) \\
&= \frac{19}{6} \text{ sq units}
\end{aligned}$$

11 Volume

When we think about volume from an intuitive point of view, we typically think of it as the amount of "space" an item occupies. Unfortunately assigning a number that measures this amount of space can prove difficult for all but the simplest geometric shapes. Calculus provides a new tool that can greatly extend our ability to calculate volume. In order to understand the ideas involved it helps to think about the volume of a cylinder. The volume of a cylinder is calculated using the formula $V = \pi r^2 h$. The base of the cylinder is a circle whose area is given by $A = \pi r^2$. Notice that the volume of a cylinder is derived by taking the area of its base and multiplying by the height h. For more complicated shapes, we could think of approximating the volume by taking the area of some cross section at some height x and multiplying by some small change in height Δx then adding up the heights of all of these approximations from the bottom to the top of the object. This would appear to be a Riemann sum. Keeping this in mind, we can develop a more general formula for the volume of solids in $\mathbb{R}^3$ (3 dimensional space).

11.1 Formal Definition

Formally the ideas above suggest that we can calculate the volume of a solid by calculating the integral of the cross-sectional area along some dimension. In the above example of a cylinder, the every cross section was given by the same circle, so the cross-sectional area is therefore a constant function, and the dimension of integration was vertical (although it could have been any one we desired). Generally, if S is a solid that lies in $\mathbb{R}^3$ between $x = a$ and $x = b$, let $A(x)$ denote the area of a cross section taken in the plane perpendicular to the x direction, and passing through the point x.

If the function $A(x)$ is continuous on $[a, b]$, then the volume V_S of the solid S is given by:

$$V_S = \int_a^b A(x)dx.$$

Example 11.1. ***A right cylinder:***

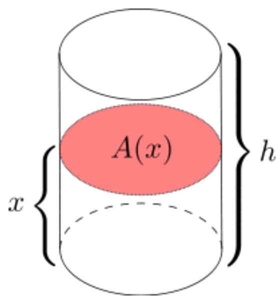


Figure 11.1

Now we will calculate the volume of a right cylinder using our new ideas about how to calculate volume. Since we already know the formula for the volume of a cylinder this will give us a "sanity check" that our formulas make sense. First, we choose a dimension along which to integrate. In this case, it will greatly simplify the calculations to integrate along the height of the cylinder, so this is the direction we will choose. Thus we will call the vertical direction x (see Figure 1). Now we find the function, $A(x)$, which will describe the cross-sectional area of our cylinder at a height of x. The cross-sectional area of a cylinder is simply a circle. Now simply recall that the area of a circle is πr^2, and so $A(x) = \pi r^2$. Before performing the computation, we must choose our bounds of integration. In this case, we simply define $x = 0$ to be the base of the cylinder, and so we will integrate from $x = 0$

to $x = h$, where h is the height of the cylinder. Finally, we integrate:

$$\begin{aligned} V_{\text{cylinder}} &= \int_a^b A(x)dx \\ &= \int_0^h \pi r^2 dx \\ &= \pi r^2 \int_0^h dx \\ &= \pi r^2 x \Big|_{x=0}^{h} \\ &= \pi r^2 (h - 0) \\ &= \pi r^2 h. \end{aligned}$$

This is exactly the familiar formula for the volume of a cylinder.

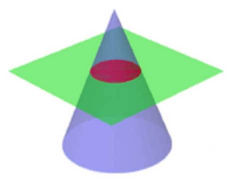


Figure 11.2: The cross-section of a right circular cone by a plane perpendicular to the axis of the cone is a circle.

Example 11.2.

For our next example we will look at an example where the cross sectional area is not constant. Consider a right circular cone. Once again the cross sections are simply circles. But now the radius varies from the base of the cone to the tip. Once again we choose x to be the vertical direction, with the base at $x = 0$ and the tip at $x = h$, and we will let R denote the radius of the base. While we know the cross sections are just circles we cannot calculate the area of the cross sections unless we find some way to determine the radius of the circle at height x. Luckily in this case it is possible to use some of what we know from geometry. We can imagine cutting the cone perpendicular to the base through some diameter of the circle all the way to the tip of the cone. If we then look at the flat side we just created, we will see simply a triangle, whose geometry we understand well. The right triangle from the tip to the base at height x is similar to the right triangle from the tip to the base at height h. This tells us that $\dfrac{r}{h-x} = \dfrac{R}{h}$. So that we see that the radius of the circle at height

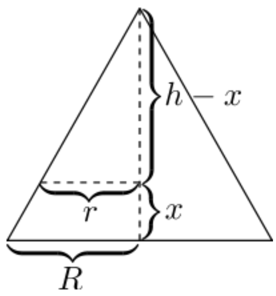


Figure 11.3: Cross-section of the right circular cone by a plane perpendicular to the base and passing through the tip.

x is $r(x) = \frac{R}{h}(h-x)$. Now using the familiar formula for the area of a circle we see that $A(x) = \pi\frac{R^2}{h^2}(h-x)^2$.

Now we are ready to integrate.

$$\begin{aligned} V_{\text{cone}} &= \int_a^b A(x)dx \\ &= \int_0^h \pi\frac{R^2}{h^2}(h-x)^2 dx \\ &= \pi\frac{R^2}{h^2}\int_0^h (h-x)^2 dx \end{aligned}$$

By u-substitution we may let $u = h - x$, then $du = -dx$ and our integral becomes

$$\begin{aligned} &= \pi\frac{R^2}{h^2}\left(-\int_h^0 u^2 du\right) \\ &= \pi\frac{R^2}{h^2}\left(-\frac{u^3}{3}\Big|_h^0\right) \\ &= \pi\frac{R^2}{h^2}\left(-0+\frac{h^3}{3}\right) \\ &= \frac{\pi}{3}R^2h \end{aligned}$$

Example 11.3. ***A sphere***

In a similar fashion, we can use our definition to prove the well known formula for the volume of a sphere. First, we must find our cross-sectional area function, $A(x)$. Consider a sphere of radius R which is centered at the origin in $\mathbb{R}^3$. If we again integrate vertically then x will vary from $-R$ to R. In order to find the area of a particular cross section it helps to draw a right triangle whose points lie at the center of the sphere, the center of the circular cross section, and at a point along the circumference of the cross section. As shown in the

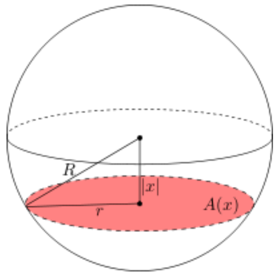


Figure 11.4: Determining the radius of the cross-section of the sphere at a distance $|x|$from the sphere's center.

diagram the side lengths of this triangle will be R , $|x|$, and r. Where r is the radius of the circular cross section. Then by the Pythagorean theorem $r = \sqrt{R^2 - |x|^2}$ and find that $A(x) = \pi(R^2 - |x|^2)$. It is slightly helpful to notice that $|x|^2 = x^2$ so we do not need to keep the absolute value.

So we have that

$$\begin{aligned} V_{\text{sphere}} &= \int_a^b A(x)dx \\ &= \int_{-R}^{R} \pi(R^2 - x^2)dx \\ &= \pi \int_{-R}^{R} R^2 dx - \pi \int_{-R}^{R} x^2 dx \\ &= \pi R^2 x \Big|_{-R}^{R} - \pi \frac{x^3}{3}\Big|_{-R}^{R} \\ &= \pi R^2(R - (-R)) - \pi \left(\frac{R^3}{3} - \frac{(-R)^3}{3}\right) \\ &= 2\pi R^3 - \frac{2\pi}{3}R^3 = \frac{4\pi}{3}R^3 \end{aligned}$$

11.2 Volume of solids of revolution

In this section we cover solids of revolution and how to calculate their volume. A solid of revolution is a solid formed by revolving a 2-dimensional region around an axis. For example, revolving the semi-circular region bounded by the curve $y = \sqrt{1 - x^2}$ and the line $y = 0$ around the x-axis produces a sphere. There are two main methods of calculating the volume of a solid of revolution using calculus: the disk method and the shell method.

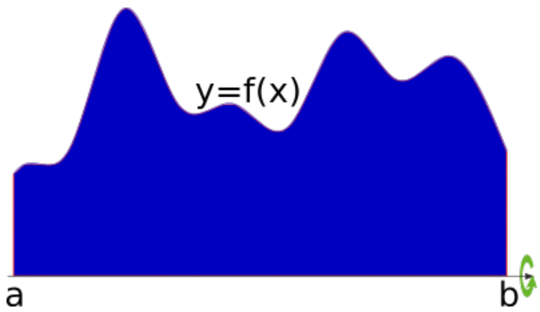


Figure 11.5: A solid of revolution is generated by revolving this region around the x-axis.

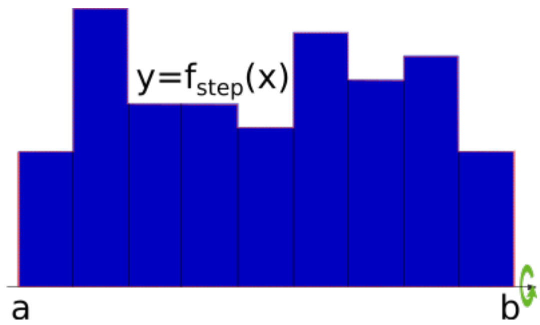


Figure 11.6: Approximation to the generating region in Figure 11.5.

11.2.1 Disk Method

Consider the solid formed by revolving the region bounded by the curve $y = f(x)$, which is continuous on $[a, b]$, and the lines $x = a$, $x = b$ and $y = 0$ around the x-axis. We could imagine approximating the volume by approximating $f(x)$ with the stepwise function $g(x)$ shown in figure 11.6, which uses a right-handed approximation to the function. Now when the region is revolved, the region under each step sweeps out a cylinder, whose volume we know how to calculate, i.e.

$$V_{\text{cylinder}} = \pi r^2 h$$

where r is the radius of the cylinder and h is the cylinder's height. This process is reminiscent of the Riemann process we used to calculate areas earlier. Let's try to write the volume as a Riemann sum and from that equate the volume to an integral by taking the limit as the subdivisions get infinitely small.

Consider the volume of one of the cylinders in the approximation, say the k-th one from the left. The cylinder's radius is the height of the step function, and the thickness is the length of the subdivision. With n subdivisions and a length of $b - a$ for the total length of the region, each subdivision has width

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

Since we are using a right-handed approximation, the k-th sample point will be

$$x_k = k\Delta x$$

So the volume of the k-th cylinder is

$$V_k = \pi f(x_k)^2 \Delta x$$

Summing all of the cylinders in the region from a to b , we have

$$V_{\text{approx}} = \sum_{k=1}^{n} \pi f(x_k)^2 \Delta x$$

Taking the limit as n approaches infinity gives us the the exact volume

$$V = \lim_{n\to\infty} \sum_{k=1}^{n} \pi f(x_k)^2 \Delta x$$

which is equivalent to the integral

$$V = \int_a^b \pi f(x)^2 dx$$

Example: Volume of a Sphere

Let's calculate the volume of a sphere using the disk method. Our generating region will be the region bounded by the curve $f(x) = \sqrt{r^2 - x^2}$ and the line $y = 0$. Our limits of integration will be the x-values where the curve intersects the line $y = 0$, namely, $x = \pm r$. We have

$$\begin{aligned}
V_{\text{sphere}} &= \int_{-r}^{r} \pi(r^2 - x^2)dx \\
&= \pi \left(\int_{-r}^{r} r^2 dx - \int_{-r}^{r} x^2 dx \right) \\
&= \pi \left(r^2 x \Big|_{-r}^{r} - \frac{x^3}{3} \Big|_{-r}^{r} \right) \\
&= \pi \Big(r^2 \big(r - (-r) \big) - \tfrac{1}{3} \big(r^3 - (-r)^3 \big) \Big) \\
&= \pi \left(2r^3 - \frac{2r^3}{3} \right) \\
&= \pi \frac{6r^3 - 2r^3}{3} \\
&= \frac{4\pi}{3} r^3
\end{aligned}$$

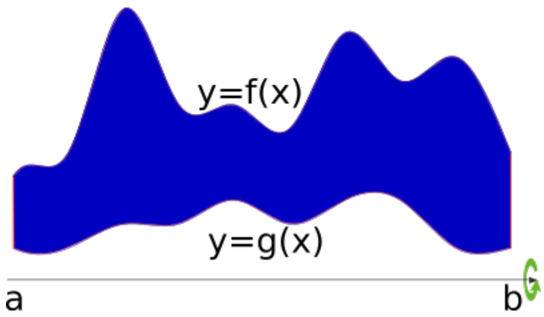


Figure 11.7: A solid of revolution containing an irregularly shaped hole through its center is generated by revolving this region around the x-axis.

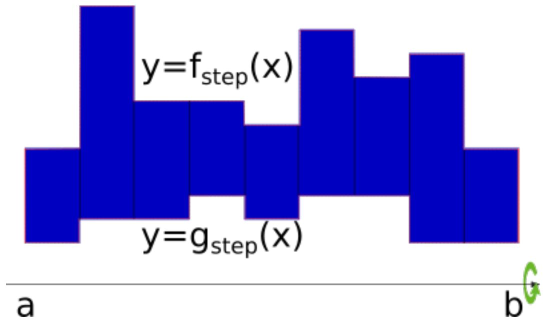


Figure 11.8: Approximation to the generating region in Figure 11.7.

11.2.2 Washer Method

The washer method is an extension of the disk method to solids of revolution formed by revolving an area bounded between two curves around the x-axis. Consider the solid of revolution formed by revolving the region in figure 11.7 around the x-axis. The curve $f(x)$ is the same as that in figure 11.5, but now our solid has an irregularly shaped hole through its center whose volume is that of the solid formed by revolving the curve $g(x)$ around the x-axis. Our approximating region has the same upper boundary, $f_{\text{step}}(x)$ as in figure 11.6, but now we extend only down to $g_{\text{step}}(x)$ rather than all the way down to the x-axis. Revolving each block around the x-axis forms a washer-shaped solid with outer radius $f_{\text{step}}(x)$ and inner radius $g_{\text{step}}(x)$. The volume of the k-th hollow cylinder is

$$\begin{aligned} V_k &= \pi \cdot f(x_k)^2 \Delta x - \pi \cdot g(x_k)^2 \Delta x \\ &= \pi \Big(f(x_k)^2 - g(x_k)^2 \Big) \Delta x \end{aligned}$$

where $\Delta x = \dfrac{b-a}{n}$ and $x_k = k\Delta x$. The volume of the entire approximating solid is

$$V_{\text{approx}} = \sum_{k=1}^{n} \pi \Big(f(x_k)^2 - g(x_k)^2 \Big) \Delta x$$

Taking the limit as n approaches infinity gives the volume

$$V = \lim_{n\to\infty} \sum_{k=1}^{n} \pi\Big(f(x_k)^2 - g(x_k)^2\Big)\Delta x$$
$$= \int_a^b \pi\Big(f(x)^2 - g(x)^2\Big)dx$$

11.2.3 Shell Method

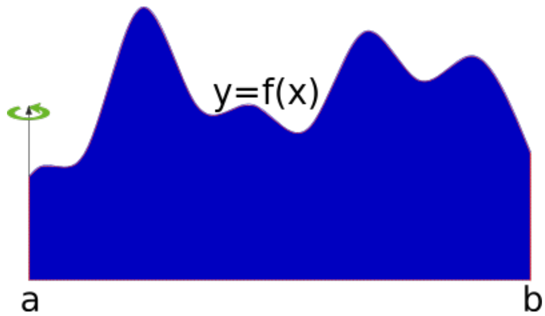


Figure 11.9: A solid of revolution is generated by revolving this region around the y-axis.

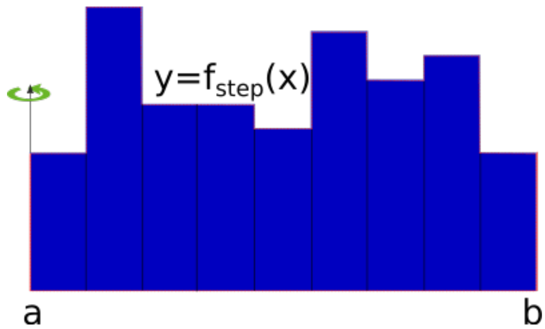


Figure 11.10: Approximation to the generating region in Figure 11.9

The shell method is another technique for finding the volume of a solid of revolution. Using this method sometimes makes it easier to set up and evaluate the integral. Consider the solid of revolution formed by revolving the region in figure 11.9 around the y-axis. While the generating region is the same as in figure 11.5, the axis of revolution has changed, making the disk method impractical for this problem. However, dividing the region up as we did previously suggests a similar method of finding the volume, only this time instead of adding up the volume of many approximating disks, we will add up the volume of many cylindrical shells. Consider the solid formed by revolving the region in figure 6 around the y-axis. The k-th rectangle sweeps out a hollow cylinder with height $\Big|f(x_k)\Big|$ and with inner radius x_k and outer radius $x_k + \Delta x$, where $\Delta x = \dfrac{b-a}{n}$ and $x_k = k\Delta x$, the volume of which is

$$\begin{aligned} V_k &= \pi\Big((x_k + \Delta x)^2 - x_k^2\Big)\Big|f(x_k)\Big| \\ &= \pi\Big((x_k^2 + 2x_k\Delta x + \Delta x^2) - x_k^2\Big)\Big|f(x_k)\Big| \\ &= \pi(2x_k\Delta x + \Delta x^2)\Big|f(x_k)\Big| \end{aligned}$$

The volume of the entire approximating solid is

$$V_{\text{approx}} = \sum_{k=1}^{n} \pi(2x_k\Delta x + \Delta x^2)\Big|f(x_k)\Big|$$

Taking the limit as n approaches infinity gives us the exact volume

$$\begin{aligned} V &= \lim_{n\to\infty} \sum_{k=1}^{n} \pi(2x_k\Delta x + \Delta x^2)\Big|f(x_k)\Big| \\ &= \pi \cdot \lim_{n\to\infty} \left(\sum_{k=1}^{n} 2x_k\Delta x\Big|f(x_k)\Big| + \sum_{k=1}^{n} \Delta x^2\Big|f(x_k)\Big|\right) \end{aligned}$$

Since $|f|$ is continuous on $[a, b]$, the Extreme Value Theorem implies that $|f|$ has some maximum, M , on $[a, b]$. Using this and the fact that $\Delta x^2\Big|f(x_k)\Big| > 0$, we have

$$\sum_{k=1}^{n} 2x_k\Delta x\Big|f(x_k)\Big| \le \sum_{k=1}^{n} 2x_k\Delta x\Big|f(x_k)\Big| + \sum_{k=1}^{n} \Delta x^2\Big|f(x_k)\Big| \le \sum_{k=1}^{n} 2x_k\Delta x\Big|f(x_k)\Big| + \sum_{k=1}^{n} \Delta x^2 M$$

But

$$\begin{aligned} \lim_{n\to\infty} \sum_{k=1}^{n} \Delta x^2 M &= \lim_{n\to\infty} \sum_{k=1}^{n} \left(\frac{b-a}{n}\right)^2 M \\ &= \lim_{n\to\infty} \frac{(b-a)^2}{n} M \\ &= 0 \end{aligned}$$

So by the Squeeze Theorem

$$\pi \cdot \lim_{n\to\infty} \left(\sum_{k=1}^{n} 2x_k\Delta x\Big|f(x_k)\Big| + \sum_{k=1}^{n} \Delta x^2\Big|f(x_k)\Big|\right) = \pi \cdot \lim_{n\to\infty} \sum_{k=1}^{n} 2x_k\Delta x\Big|f(x_k)\Big|$$

which is just the integral

$$\int_a^b 2\pi x\Big|f(x)\Big|dx$$

Exercise of Volume of solids of revolution

1. . Calculate the volume of the cone with radius r and height h which is generated by the revolution of the region bounded by $y = r - \frac{r}{h}x$ and the lines $y = 0$ and $x = 0$ around the x-axis.

   The region extends in the x-direction from $x = 0$ to $x = h$. The volume of the solid of revolution is given by

$$\begin{aligned}\int_0^h \pi \left(r - \frac{r}{h}x\right)^2 dx &= \pi \int_0^h \left(r^2 - \frac{2r^2}{h}x + \frac{r^2}{h^2}x^2\right) dx \\ &= \pi r^2 \left(\int_0^h dx - \frac{2}{h}\int_0^h x dx + \frac{1}{h^2}\int_0^h x^2 dx\right) \\ &= \pi r^2 \left(x\Big|_0^h - \frac{x^2}{h}\Big|_0^h + \frac{x^3}{3h^2}\Big|_0^h\right) \\ &= \pi r^2 \left(h - h + \frac{h}{3}\right) \\ &= \frac{\boldsymbol{\pi r^2 h}}{\mathbf{3}}\end{aligned}$$

2. Calculate the volume of the solid of revolution generated by revolving the region bounded by the curve $y = x^2$ and the lines $x = 1$ and $y = 0$ around the x-axis.

   The region extends in the x-direction from $x = 0$ to $x = 1$. The volume of the solid of revolution is given by

$$\begin{aligned}\int_0^1 \pi \left(x^2\right)^2 dx &= \pi \int_0^1 x^4 dx \\ &= \pi \frac{x^5}{5}\Big|_0^1 \\ &= \frac{\pi}{\mathbf{5}}\end{aligned}$$

3. Use the washer method to find the volume of a cone containing a central hole formed by revolving the region bounded by $y = R - \frac{R}{h}x$ and the lines $y = r$ and $x = 0$ around the x-axis.

The x values of the region extend from $x = 0$ to $x = h$. The volume is

$$\begin{aligned}
V &= \int_0^h \pi \left(\left(R - \frac{R}{h}x \right)^2 - r^2 \right) dx \\
&= \int_0^h \pi \left(R^2 - 2\frac{R^2}{h}x + \frac{R^2}{h^2}x^2 - r^2 \right) dx \\
&= \pi \left(R^2 x - \frac{R^2}{h}x^2 + \frac{R^2}{3h^2}x^3 - r^2 x \right) \Big|_0^h \\
&= \pi \left(R^2 h - R^2 h + \frac{R^2 h}{3} - r^2 h \right) \\
&= \pi \mathbf{h} \left(\frac{\mathbf{R^2}}{\mathbf{3}} - \mathbf{r^2} \right)
\end{aligned}$$

4. Calculate the volume of the solid of revolution generated by revolving the region bounded by the curves $y = x^2$ and $y = x^3$ and the lines $x = 1$ and $y = 0$ around the x-axis.

$$\begin{aligned}
V &= \int_0^1 \pi \left(\left(x^2\right)^2 - \left(x^3\right)^2 \right) dx \\
&= \int_0^1 \pi \left(x^4 - x^6 \right) dx \\
&= \pi \left(\frac{x^5}{5} - \frac{x^7}{7} \right) \Big|_0^1 \\
&= \pi \left(\frac{1}{5} - \frac{1}{7} \right) \\
&= \pi \frac{7-5}{35} \\
&= \frac{\mathbf{2}\pi}{\mathbf{35}}
\end{aligned}$$

5. Find the volume of a cone with radius r and height h by using the shell method on the appropriate region which, when rotated around the y-axis, produces a cone with the given characteristics.

   You could set up the appropriate region in any of the four quadrants. Here we set it up in the first quadrant. Since we are revolving around the y-axis, the y direction will be the height and the radius will be along the x direction. So we need a line that passes through the points $(0, h)$ and $(R, 0)$. The slope of this line is

$$m = \frac{0-h}{R-0} = -\frac{h}{R}$$

   and the y-intercept is h. Thus, the equation of the line is $y = -\frac{h}{R}x + h$

   The x-values of the region run from $x = 0$ to $x = R$. Since the function is positive throughout the region we can drop the absolute value sign. The volume will be

$$\begin{aligned} V &= \int_0^R 2\pi x(-\frac{h}{R}x + h)dx \\ &= \int_0^R 2\pi(-\frac{h}{R}x^2 + hx)dx \\ &= 2\pi(-\frac{h}{3R}x^3 + \frac{h}{2}x^2)\Big|_0^R \\ &= 2\pi(-\frac{hR^2}{3} + \frac{hR^2}{2}) \\ &= 2\pi\frac{-2hR^2 + 3hR^2}{6} \\ &= \frac{\pi \mathbf{hR^2}}{\mathbf{3}} \end{aligned}$$

6. Calculate the volume of the solid of revolution generated by revolving the region bounded by the curve $y = x^2$ and the lines $x = 1$ and $y = 0$ around the y-axis.

$$\begin{aligned} V &= \int_0^1 2\pi x x^2 dx \\ &= \int_0^1 2\pi x^3 dx \\ &= 2\pi\frac{x^4}{4}\Big|_0^1 \\ &= \frac{\pi}{\mathbf{2}} \end{aligned}$$

Example 11.4. *Find the volume if the area bounded by the curve $y = x^3 + 1$, the x-axis and the limits of $x = 0$ and $x = 3$ is rotated around the x-axis.*

This is the region as described, under a cubic curve. When the shaded area is rotated 360^o

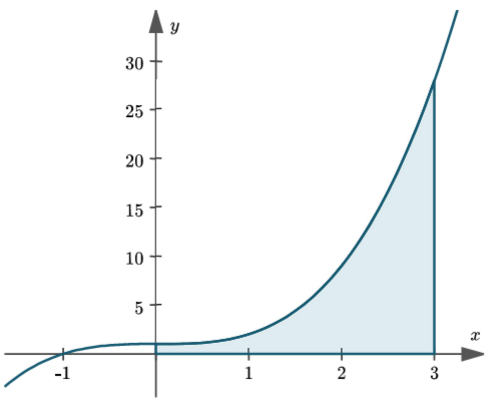


Figure 11.11: Area bounded by $y = x^3 + 1$, $x = 0$ and $x = 3$.

about the x-axis, we observe that a volume is generated: Applying the formula for the solid

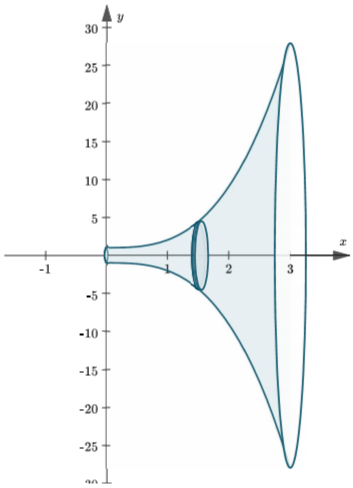


Figure 11.12: Area under the curve $y = x^3 + 1$ from $x = 0$ to $x = 3$ rotated around the x-axis, showing a typical disk.

of revolution, we get

$$\begin{aligned} V &= \pi \int_a^b y^2\,dx \\ &= \pi \int_0^3 \left(x^3+1\right)^2 dx \\ &= \pi \int_0^3 \left(x^6 + 2x^3 + 1\right) dx \\ &= \pi \left[\frac{x^7}{7} + \frac{x^4}{2} + x\right]_0^3 \\ &= \pi(|355.93| - |0|) \\ &= 1118.2 \text{ units}^3 \end{aligned}$$

Example 11.5. *A cup is made by rotating the area between $y = 2x^2$ and $y = x + 1$ with $x \geq 0$ around the x-axis. Find the volume of the material needed to make the cup. Units are cm.*

We sketch the upper and lower bounding curves:
The lower limit of integration is $x = 0$ (since the question says $x \geq 0$). Next, we need to find where the curves intersect so we know the upper limit of integration.

Equating the 2 expressions and solving:

$$\begin{aligned} 2x^2 &= x + 1 \\ 2x^2 - x - 1 &= 0 \\ (2x+1)(x-1) &= 0 \end{aligned}$$

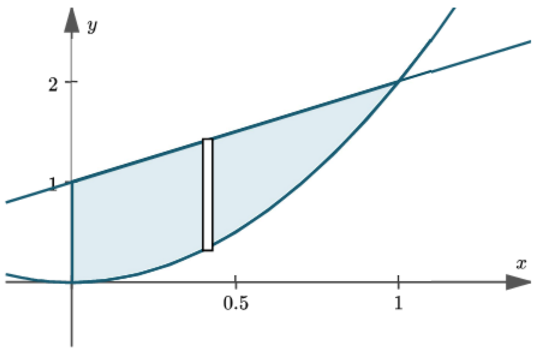


Figure 11.13: Area bounded by $y = 2x^2$ (the bottom curve), $y = x + 1$ (the line above), and $x = 0$, showing a typical rectangle.

$x = 1$ (since we only need to consider $x \leq 0$. This is consistent with what we see in the graph above.)

So with $y_2 = x + 1$ and$y_1 = 2x^2$, the volume required is given by:

$$\begin{aligned}\text{Volume} &= \pi\int_0^1 \left[(x+1)^2 - \left(2x^2\right)^2\right] dx\\ &= \pi\int_0^1 \left[\left(x^2+2x+1\right) - \left(4x^4\right)\right] dx\\ &= \pi\left[\frac{x^3}{3} + x^2 + x - \frac{4x^5}{5}\right]_0^1\\ &= \pi\left[\left(\frac{1}{3} + 1 + 1 - \frac{4}{5}\right) - (0)\right]\\ &= \pi\left[\frac{5+30-12}{15}\right]\\ &= \frac{23\pi}{15}\\ &= 4.817 \text{ cm}^3\end{aligned}$$

Here's an illustration of the volume we have found. A typical "washer" with outer radius $y_2 = x + 1$ and inner radius $y_1 = 2x^2$ is shown.

Example 11.6. *Find the volume of the solid of revolution generated by rotating the curve $y = x^3$ between $y = 0$ and $y = 4$ about the y-axis.*

Solution

Here is the region we need to rotate: And here is the volume generated when we rotate the region around the y-axis:

We first must express x in terms of y, so that we can apply the volume of solid of revolution formula.

If $y = x^3$ then $x = y^{1/3}$

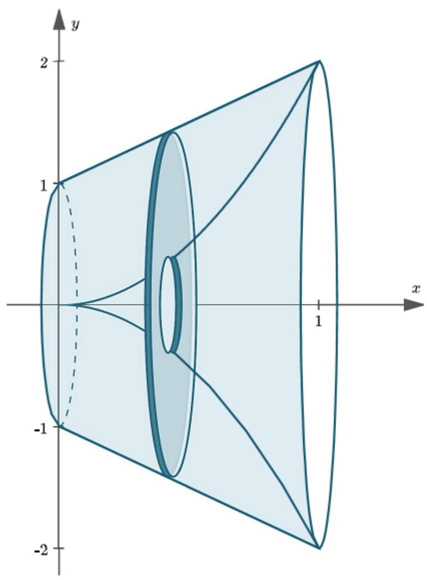


Figure 11.14: The cup resulting from rotating the area bounded by $y = 2x^2$, $y = x + 1$, and $x = 0$ about the x-axis.

The formula requires x^2, and on squaring we have $x^2 = y^{2/3}$

$$\begin{aligned}
\text{Vol} &= \pi \int_c^d x^2 \, dy \\
&= \pi \int_0^4 y^{2/3} \, dy \\
&= \pi \left[\frac{3y^{5/3}}{5} \right]_0^4 \\
&= \frac{3\pi}{5} \left[y^{5/3} \right]_0^4 \\
&= \frac{3\pi}{5} [10.079 - 0] \\
&= 19.0 \text{ units}^3
\end{aligned}$$

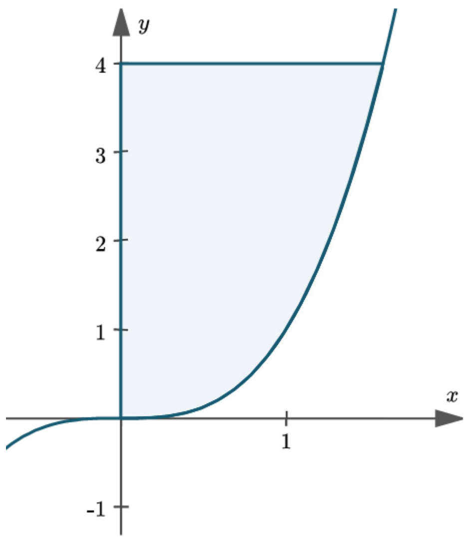


Figure 11.15: The graph of the area bounded by $y = x^3$, $x = 0$ and $y = 4$.

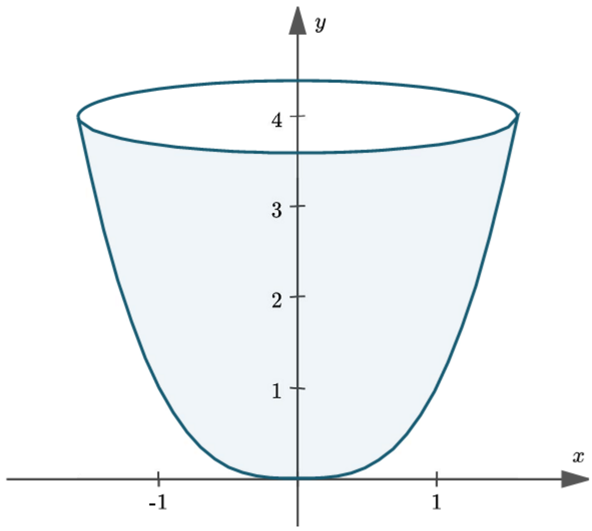


Figure 11.16: The volume generated when revolving the curve bounded by $y = x^3$, $x = 0$ and $y = 4$ around the y-axis.

12 Arc length

Suppose that we are given a function f that is continuous on an interval $[a, b]$ and we want to calculate the length of the curve drawn out by the graph of $f(x)$ from $x = a$ to $x = b$. If the graph were a straight line this would be easy the formula for the length of the line is given by Pythagoras' theorem. And if the graph were a piecewise linear function we can calculate the length by adding up the length of each piece.

The problem is that most graphs are not linear. Nevertheless we can estimate the length of the curve by approximating it with straight lines. Suppose the curve C is given by the formula $y = f(x)$ for $a \leq x \leq b$. We divide the interval $[a, b]$ into n subintervals with equal width Δx and endpoints $x_0, x_1, \ldots, x_n$. Now let $y_i = f(x_i)$ so $P_i = (x_i, y_i)$ is the point on the curve above x_i. The length of the straight line between P_i and P_{i+1} is

$$\left|P_i P_{i+1}\right| = \sqrt{(y_{i+1} - y_i)^2 + (x_{i+1} - x_i)^2}$$

So an estimate of the length of the curve C is the sum

$$\sum_{i=0}^{n-1} \left|P_i P_{i+1}\right|$$

As we divide the interval$[a, b]$ into more pieces this gives a better estimate for the length of C. In fact we make that a definition.

Length of a Curve

The length of the curve $y = f(x)$ for $a \leq x \leq b$ is defined to be

$$L = \lim_{n \to \infty} \sum_{i=0}^{n-1} \left|P_{i+1} P_i\right|$$

12.1 The Arclength Formula

Suppose that f' is continuous on $[a, b]$. Then the length of the curve given by $y = f(x)$ between a and b is given by

$$L = \int_a^b \sqrt{1 + f'(x)^2} dx$$

And in Leibniz notation

$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx$$

Proof: Consider $y_{i+1} - y_i = f(x_{i+1}) - f(x_i)$. By the Mean Value Theorem there is a point z_i in (x_{i+1}, x_i) such that

$$y_{i+1} - y_i = f(x_{i+1}) - f(x_i) = f'(z_i)(x_{i+1} - x_i)$$

So

$$\begin{aligned}\left|P_iP_{i+1}\right| &= \sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2}\\ &= \sqrt{(x_{i+1}-x_i)^2+f'(z_i)^2(x_{i+1}-x_i)^2}\\ &= \sqrt{\left(1+f'(z_i)^2\right)(x_{i+1}-x_i)^2}\\ &= \sqrt{1+f'(z_i)^2}\Delta x\end{aligned}$$

Putting this into the definition of the length of C gives

$$L = \lim_{n\to\infty}\sum_{i=0}^{n-1}\sqrt{1+f'(z_i)^2}\Delta x$$

Now this is the definition of the integral of the function $g(x)=\sqrt{1+f'(x)^2}$ between a and b (notice that g is continuous because we are assuming that f' is continuous). Hence

$$L = \int_a^b \sqrt{1+f'(x)^2}dx$$

as claimed.

Example 12.1. *Length of the curve $y = 2x$ from $x = 0$ to $x = 1$*

As a sanity check of our formula, let's calculate the length of the "curve" $y = 2x$ from $x = 0$ to $x = 1$. First let's find the answer using the Pythagorean Theorem.

$$P_0 = (0, 0)$$

and

$$P_1 = (1, 2)$$

so the length of the curve, s, is

$$s = \sqrt{2^2+1^2} = \sqrt{5}$$

Now let's use the formula

$$s = \int_0^1 \sqrt{1+\left(\tfrac{d(2x)}{dx}\right)^2}\,dx = \int_0^1 \sqrt{1+2^2}\,dx = \sqrt{5}x\Big|_0^1 = \sqrt{5}$$

12.2 Arclength of a parametric curve

For a parametric curve, that is, a curve defined by $x = f(t)$ and $y = g(t)$, the formula is slightly different:

$$L = \int_a^b \sqrt{f'(t)^2+g'(t)^2}\,dt$$

Proof: The proof is analogous to the previous one: Consider $y_{i+1} - y_i = g(t_{i+1}) - g(t_i)$ and $x_{i+1} - x_i = f(t_{i+1}) - f(t_i)$.

By the Mean Value Theorem there are points c_i and d_i in (t_{i+1}, t_i) such that

$$y_{i+1} - y_i = g(t_{i+1}) - g(t_i) = g'(c_i)(t_{i+1} - t_i)$$

and

$$x_{i+1} - x_i = f(t_{i+1}) - f(t_i) = f'(d_i)(t_{i+1} - t_i)$$

So

$$\begin{aligned}\left|P_iP_{i+1}\right| &= \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \\ &= \sqrt{f'(d_i)^2(t_{i+1} - t_i)^2 + g'(c_i)^2(t_{i+1} - t_i)^2} \\ &= \sqrt{\left(f'(d_i)^2 + g'(c_i)^2\right)(t_{i+1} - t_i)^2}\end{aligned}$$

$$= \sqrt{f'(d_i)^2 + g'(c_i)^2}\Delta t$$

Putting this into the definition of the length of the curve gives

$$L = \lim_{n\to\infty} \sum_{i=0}^{n-1} \sqrt{f'(d_i)^2 + g'(c_i)^2}\Delta t$$

This is equivalent to:

$$L = \int_a^b \sqrt{f'(t)^2 + g'(t)^2}\, dt$$

Exercises of Arc length

1. Find the length of the curve $y = x^{3/2}$ from $x = 0$ to $x = 1$.

$$\begin{aligned}s &= \int_0^1 \sqrt{1 + \left(\frac{d\left(x^{3/2}\right)}{dx}\right)^2} dx \\ &= \int_0^1 \sqrt{1 + \left(\frac{3}{2}\sqrt{x}\right)^2} dx \\ &= \int_0^1 \sqrt{1 + \frac{9}{4}x dx}\end{aligned}$$

Let

$$u = 1 + \frac{9}{4}x; \qquad du = \frac{9}{4}dx; \qquad dx = \frac{4}{9}du$$

Then

$$\begin{aligned}
s &= \int_{u(0)}^{u(1)} \sqrt{u}\frac{4}{9}du \\
&= \frac{4}{9}\frac{2}{3}u^{3/2}\Big|_{u(0)}^{u(1)} \\
&= \frac{8}{27}\left(1+\frac{9}{4}x\right)^{3/2}\Big|_0^1 \\
&= \frac{8}{27}\left(\left(1+\frac{9}{4}\right)^{3/2}-1\right) \\
&= \frac{8}{27}\left(\left(\frac{13}{4}\right)^{3/2}-1\right) \\
&= \frac{8}{27}\left(\left(\frac{13}{2^2}\right)^{3/2}-1\right) \\
&= \frac{8}{27}\left(\frac{13^{3/2}}{2^3}-1\right) \\
&= \frac{8}{27}\left(\frac{13^{3/2}}{8}-1\right) \\
&= \mathbf{\frac{13^{3/2}-8}{27}}
\end{aligned}$$

2. Find the length of the curve $y = \dfrac{e^x + e^{-x}}{2}$ from $x = 0$ to $x = 1$.

$$\begin{aligned}
s &= \int_0^1 \sqrt{1+\left(\frac{d\left(\frac{e^x+e^{-x}}{2}\right)}{dx}\right)^2}dx \\
&= \int_0^1 \sqrt{1+\left(\frac{e^x-e^{-x}}{2}\right)^2}dx \\
&= \int_0^1 \sqrt{1+\frac{e^{2x}-2+e^{-2x}}{4}}dx \\
&= \int_0^1 \sqrt{\frac{e^{2x}+2+e^{-2x}}{4}}dx \\
&= \int_0^1 \sqrt{\left(\frac{e^x+e^{-x}}{2}\right)^2}dx \\
&= \int_0^1 \frac{e^x+e^{-x}}{2}dx \\
&= \frac{e^x-e^{-x}}{2}\Big|_0^1 \\
&= \mathbf{\frac{e-\frac{1}{e}}{2}}
\end{aligned}$$

3. Find the circumference of the circle given by the parametric equations $x(t) = R\cos(t)$,

$y(t) = R\sin(t)$, with t running from 0 to 2π.

$$\begin{aligned} s &= \int_0^{2\pi} \sqrt{\left(\frac{d(R\cos(t))}{dt}\right)^2 + \left(\frac{d(R\sin(t))}{dt}\right)^2} dt \\ &= \int_0^{2\pi} \sqrt{(-R\sin(t))^2 + (R\cos(t))^2} dt \\ &= \int_0^{2\pi} \sqrt{R^2(\sin^2(t) + \cos^2(t))} dt \\ &= \int_0^{2\pi} R dt \\ &= Rt\Big|_0^{2\pi} \\ &= \mathbf{2\pi R} \end{aligned}$$

4. Find the length of one arch of the cycloid given by the parametric equations $x(t) = R(t - \sin(t))$, $y(t) = R(1 - \cos(t))$, with t running from 0 to 2π.

$$\begin{aligned} s &= \int_0^{2\pi} \sqrt{\left(\frac{d(R(t - \sin(t)))}{dt}\right)^2 + \left(\frac{d(R(1 - \cos(t)))}{dt}\right)^2} dt \\ &= \int_0^{2\pi} \sqrt{(R(1 - \cos(t)))^2 + (R\sin(t))^2} dt \\ &= \int_0^{2\pi} R\sqrt{(1 - \cos(t))^2 + \sin^2(t)} dt \\ &= \int_0^{2\pi} R\sqrt{1 - 2\cos(t) + \cos^2(t) + \sin^2(t)} dt \\ &= \int_0^{2\pi} R\sqrt{2 - 2\cos(t)} dt \end{aligned}$$

Using the trigonometric identity

$$\sin^2\left(\frac{t}{2}\right) = \frac{1 - \cos(t)}{2}$$

, we have

$$\begin{aligned} s &= \int_0^{2\pi} R\sqrt{4\sin^2\left(\frac{t}{2}\right)} dt \\ &= \int_0^{2\pi} 2R\sin\left(\frac{t}{2}\right) dt \\ &= -4R\cos\left(\frac{t}{2}\right)\Big|_0^{2\pi} \\ &= -4R(-1 - 1) \\ &= \mathbf{8R} \end{aligned}$$

13 Surface area

Suppose we are given a function f and we want to calculate the surface area of the function f rotated around a given line. The calculation of surface area of revolution is related to the arc length calculation.

If the function f is a straight line, other methods such as surface area formula for cylinders and conical frustra can be used. However, if f is not linear, an integration technique must be used.

Recall the formula for the lateral surface area of a conical frustum:

$$A = 2\pi rl$$

where r is the average radius and l is the slant height of the frustum.

For $y = f(x)$ and $a \leq x \leq b$, we divide $[a, b]$ into subintervals with equal width δx and endpoints $x_0, x_1, \ldots, x_n$. We map each point $y_i = f(x_i)$ to a conical frustum of width Δx and lateral surface area A_i.

We can estimate the surface area of revolution with the sum

$$A = \sum_{i=0}^{n} A_i$$

As we divide $[a, b]$ into smaller and smaller pieces, the estimate gives a better value for the surface area.

13.1 Definition (Surface of Revolution)

The surface area of revolution of the curve $y = f(x)$ about a line for $a \leq x \leq b$ is defined to be

$$A = \lim_{n\to\infty} \sum_{i=0}^{n} A_i$$

13.2 The Surface Area Formula

Suppose f is a continuous function on the interval $[a, b]$ and $r(x)$ represents the distance from $f(x)$to the axis of rotation. Then the lateral surface area of revolution about a line is given by

$$A = 2\pi \int_a^b r(x)\sqrt{1 + f'(x)^2}\, dx$$

And in Leibniz notation

$$A = 2\pi \int_a^b r(x)\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

Proof:

$$\begin{aligned} A &= \lim_{n\to\infty}\sum_{i=1}^{n} A_i \\ &= \lim_{n\to\infty}\sum_{i=1}^{n} 2\pi r_i l_i \\ &= 2\pi\cdot\lim_{n\to\infty}\sum_{i=1}^{n} r_i l_i \end{aligned}$$

As $n\to\infty$ and $\Delta x\to 0$, we know two things:

the average radius of each conical frustum r_i approaches a single value the slant height of each conical frustum l_i equals an infitesmal segment of arc length From the arc length formula discussed in the previous section, we know that

$$l_i = \sqrt{1+f'(x_i)^2}$$

Therefore

$$\begin{aligned} A &= 2\pi\cdot\lim_{n\to\infty}\sum_{i=1}^{n} r_i l_i \\ &= 2\pi\cdot\lim_{n\to\infty}\sum_{i=1}^{n} r_i\sqrt{1+f'(x_i)^2}\Delta x \end{aligned}$$

Because of the definition of an integral $\int_a^b f(x)dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(c_i)\Delta x_i$, we can simplify the sigma operation to an integral.

$$A = 2\pi\int_a^b r(x)\sqrt{1+f'(x)^2}dx$$

Or if f is in terms of y on the interval $[c,d]$

$$A = 2\pi\int_c^d r(y)\sqrt{1+f'(y)^2}dy$$