Precalculus with Geometry and Trigonometry

by Avinash Sathaye, Professor of Mathematics [1]
Department of Mathematics, University of Kentucky

Āryabhaṭa

November 23, 2009

[1]Partially supported by NSF grant thru AMSP(Appalachian Math Science Partnership)

Introduction.

This book on **Precalculus with Geometry and Trigonometry** should be treated as simply an enhanced version of our book on **College Algebra**. Most of the topics that appear here have already been discussed in the Algebra book and often the text here is a verbatim copy of the text in the other book.
We expect the student to already have a strong Algebraic background and thus the algebraic techniques presented here are more a refresher course than a first introduction. We also expect the student to be heading for higher level mathematics courses and try to supply the necessary connections and motivations for future use.
Here is what is new in this book.

- In contrast with the Algebra book, we make a more extensive use of complex numbers. We use Euler's representation of complex numbers as well as the Argand diagrams extensively. Even though these are described and shown to be useful, we do not yet have tools to prove these techniques properly. They should be used as motivation and as an easy method to remember the trigonometric results.

- We have supplied a brief introduction to matrices and determinants. The idea is to supply motivation for further study and a feeling for the Linear Algebra.

- In the appendix, we give a more formal introduction to the structure of real numbers. While this is not necessary for calculations in this course, it is vital for understanding the finer concepts of Calculus which will be introduced in higher courses.

- We have also included an appendix discussing summation of series - both finite and infinite, as well as a discussion of power series. While details of convergence are left out, this should generate familiarity with future techniques and a better feeling for the otherwise mysterious trigonometric and exponential functions.

Contents

Chapter 1

Review of Basic Tools.

1.1 Underlying field of numbers.

Mathematics may be described as the science of manipulating numbers.
The process of using mathematics to analyze the physical universe often consists of representing events by a set of numbers, converting the laws of physical change into mathematical functions and equations and predicting or verifying the physical events by evaluating the functions or solving the equations.
We begin by describing various types of numbers in use. During thousands of years, mathematics has developed many systems of numbers. Even when some of these appear to be counterintuitive or artificial, they have proved to be increasingly useful in developing advanced solution techniques.
To be useful, our numbers must have a few fundamental properties.
We should be able to perform **the four basic operations of algebra:**

- addition
- subtraction
- multiplication and
- division (except by 0)

and produce well defined numbers as answers.
Any set of numbers having all these properties is said to be **a field** of numbers (or constants). Depending on our intended use, we work with different fields of numbers. Here is a description of fields of numbers that we typically use.

1. **Rational numbers $\mathbb{Q}$.** The most natural idea of numbers comes from simple counting $1, 2, 3, \cdots$ and these form the set of natural numbers often denoted by $\mathbb{N}$.

These are not yet good enough to make a field since subtraction like $2-5$ is undefined. To fix the subtraction property, we can add in the zero 0 and negative numbers. [1]

This produces the set of integers,

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}.$$

These are still deficient because the division does not work. You cannot divide 1 by 2 and get an integer back.

The natural next step is to introduce the fractions $\frac{m}{n}$ where m, n are integers and $n \neq 0$. You probably remember the explanation in terms of picking up parts; thus $\frac{3}{8}$-th of a pizza is three of the eight slices of one pizza.

We now have a natural field at hand, the so-called field of rational numbers

$$\mathbb{Q} = \{\frac{m}{n} | m, n \in \mathbb{Z}, n \neq 0\}$$

2. **Real numbers $\Re$.** One familiar way of thinking of numbers is as decimal numbers, say something like 2.34567 which is nothing but a rational number $\frac{234567}{100{,}000}$ whose denominator is a power of 10. [2]

[1] This simple sounding idea took several hundred years to develop and be accepted, because the idea of a negative count is hard to imagine. If we think of a number representing money owned, then a negative number can easily be thought of as money owed! The concept of negative numbers and zero was developed and expanded in India where a negative number is called ṛṇa which also means debt! The word used for a positive number, is similarly dhana which means wealth.

The point is that even though the idea of certain numbers sounds unrealistic or impossible, one should keep an open mind and accept and use them as needed. They can be useful and somebody may find a good interpretation for them some day.

[2] Indeed, the number 10 can and is often replaced by other convenient numbers. The computer scientists often prefer 2 in place of 10, leading to the binary numbers, or they also use 8 or 16 in other contexts, leading to octal or hexadecimal numbers.

In Number Theory, it is customary to replace 10 by some prime p and study the resulting p-adic numbers.

It is also possible and sometimes convenient to choose a product of suitable numbers, rather than power of a single one.

It is interesting to consider numbers of the form

$$a_1 + \frac{a_2}{2!} + \frac{a_3}{3!} + \cdots$$

where a_1 is an arbitrary integer, while $0 \leq a_2 < 2$, $0 \leq a_3 < 3$ and so on. Thus, you can verify that:

$$\frac{17}{9} = 1 + \frac{1}{2!} + \frac{2}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{4}{6!}$$

Thus, rational numbers whose denominators are powers of 10 are called decimal numbers.

A calculator, especially a primitive one, deals exclusively with such numbers.

One quickly realizes that even something as simple as $\frac{1}{3}$ runs into problems if we insist on using only decimal numbers. It has successive approximations $0.3, 0.33, 0.333, \cdots$ but no finite decimal will ever give the exact value $\frac{1}{3}$.

While this is somewhat disturbing, we do have the choice of writing $0.\overline{3}$ with the understanding that it is the decimal number obtained by repeating the digit 3 under the bar indefinitely.

It can be shown that any rational number $\frac{m}{n}$ can be written as a repeating decimal which consists of an infinite decimal number with a certain group of digits repeating indefinitely from some point on. For example,

$$\frac{1}{7} = 0.142857142857 \cdots = 0.\overline{142857}.$$

For a detailed explanation of why all rationals can be written as repeating decimals, please see the appendix.

Thus, we realize that if we are willing to handle infinite decimals, we have all the rational numbers and as a bonus we get a whole lot of new numbers whose decimal expansions don't repeat. What do we get?

We get the so-called field of real numbers. Because we are so familiar with the decimals in everyday life, these feel natural and easy, except for the fact that almost always we are dealing with approximations and issues of limiting values. Thus for example, there cannot be any difference between the decimal 0.1 and the decimal $0.09999 \cdots = 0.0\overline{9}$ even though they appear different!

The ideas that the sequence of longer and longer decimals

$$0.09, 0.099, 0.0999, \cdots$$

gets closer and closer to the decimal 0.1 and we can precisely formulate a meaning for the infinite decimal as a limit equal to 0.1 are natural, but rather sophisticated. While these were informally being used all the time, they were formalized only in the last four hundred years. The subject of Calculus pushes them to their natural logical limit.

You might recall that the set of real numbers can be represented as points of the number line by the following procedure:

---

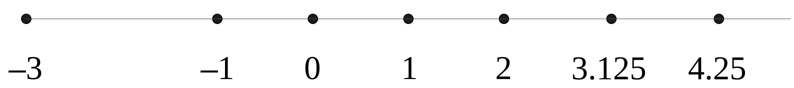
and it can be thought of as represented by $1 \cdot 12114$. We may call this the **factorial system**. This idea has the advantage of keeping all rationals as finite expressions, but is clearly not as easy to use as the decimal system!

- Mark some convenient point as **the origin** associated with **zero.**
- Mark some **other** convenient point as the **unit point**, identified with **1**. We define its distance from the origin to be the **unit distance** or simply **the unit.**
- Then every decimal number gets an appropriate position on the line with its distance from the origin set equal to its absolute value and is placed on the correct side of the zero depending on its sign. Thus the number 2 gets marked in the same direction as the unit point but at twice the unit distance.

  The number -3 gets marked on the opposite side of the unit point and at three times the unit distance.

  In general, a positive number x is on the same side of the unit point at distance equal x times the unit distance and the point $-x$ is on the other side at the same distance.

  See some illustrated points below.



It seems apparent that the real numbers must fill up the whole number line and we should not need any further numbers.

Actually, this "feeling" is misleading, since you can get the same feeling by plotting lots of rational points on the number line and see it visibly fill up!

It takes an algebraic manipulation and some clever argument to show that the positive square root of 2, usually denoted as $\sqrt{2}$ cannot be a rational number and yet deserves its rightful position on the number line.

If you have not thought about it, it is a very good challenge to prove that there are no positive integers m, n which will satisfy: [3]

$$\sqrt{2} = \frac{m}{n} \text{ in other words } m^2 = 2n^2.$$

It can, however, be "proved" that in some sense the real numbers (or the set of all decimal numbers, allowing infinite decimals) do fill up the number line. The precise proof of this fact will be presented in a good first course in Calculus.

It would seem natural to be satisfied that we have found all the numbers that need to be found and can stop at $\Re$. However, we now show that there is something more interesting to find!

[3] In the appendix, you will find a precise argument to show how $\sqrt{2}$ can be given a proper place on the real number line.

3. **Complex numbers. $\mathbb{C}$.** Rather than the geometric idea of filling up the number line, we could think of numbers as needed solutions of natural equations. Thus, the negative integer -2 is needed to be able to solve $x+5=3$. The fraction $\frac{1}{2}$ is needed to be able to solve $2x=1$.

   The number $\sqrt{2}$ is needed to solve $x^2=2$.

   In fact, in a beginning Calculus course you will meet a proof that every polynomial equation:

   $$x^n + a_1 x^{n-1} + \cdots a_n = 0$$

   of **odd degree n** has a real solution!

   Thus, real numbers provide solutions to many polynomial equations, but miss one important equation:

   $$x^2 + 1 = 0.$$

   Why does it not have a real solution?

   Suppose we **name a solution of the equation as i** and try to think where to put it on the real number line.

   Note that $i^2 = -1$.

   If i were positive, then $i \cdot i = i^2$ would be positive, but it is -1 and hence negative.

   Similarly, if i were negative, then again i^2 would be positive since a product of two negative numbers is positive. But it is -1 and hence negative.

   Thus, either assumption leads to a contradiction! Thus, i cannot be put on the left or the right side of 0.

   This number i is certainly not zero, for otherwise its square would be zero!

   Thus, i **has to be somewhere outside the real number line** and we have found a new number that we must have, if we want to solve all polynomial equations.

   As a result, we have a whole new set of numbers of the form $a+bi$ where a, b are any real numbers.

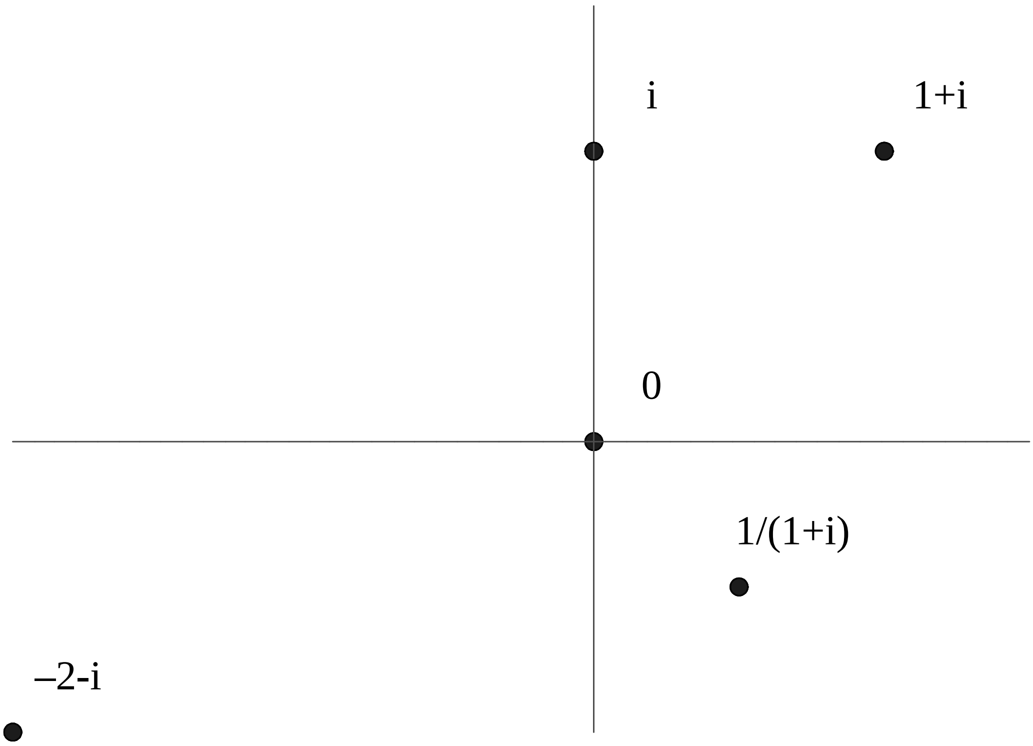
   Do these already form a field? Surprisingly, they do!

   You only need to verify the following formulas by comparing both sides of the equations, by cross multiplication, if necessary.

   $$(a+ib)(c+id) = (ac-bd) + i(ad+bc) \text{ and } \frac{1}{(a+ib)} = \frac{a}{(a^2+b^2)} - \frac{b}{(a^2+b^2)}i.$$

   Thus we have a new field at hand called the field of complex numbers and denoted by $\mathbb{C}$.

Its points can be conveniently represented as points of a plane by simply plotting the complex number $a + bi$ at the point $P(a, b)$ the point with real coordinates a, b. Here is a "complex number plane" with some points plotted. Such pictures are called **Argand diagrams**. The complex number plane is also called just the complex line by some people.



To verify your understanding of complex numbers, assign yourself the task of verifying the following: [4]

$$\frac{1}{1+i} = \frac{1-i}{2}\ ,\ (-1 \pm \sqrt{3}i)^3 = 8\ ,\ (1+i)^4 = -4.$$

[4]We illustrate how you would check the first of these. By cross multiplying, we see that we want to show:

$$2(1) = (1+i)(1-i).$$

By simple expansion the right hand side (usually shortened to RHS) is

$$(1)(1-i) + (i)(1-i)\ =\ 1 - i + i - i^2\ =\ 1 - i^2.$$

Using the basic fact $i^2 = -1$, we see that this becomes $1 + 1 = 2$ or the left hand side (LHS).
Done!

One of the most satisfying properties of complex numbers is known as the **Fundamental Theorem of Algebra** It states that **every polynomial equation** with complex coefficients:

$$x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

has a complex solution!
Of course, the real numbers are included in complex numbers, so the coefficients can be all real!
Many clever proofs of this theorem are known, but the problem of "actually finding" the solution to a given polynomial equation is hard and often has to be left as an approximate procedure. Only equations of degree 1 and 2 have well known formulas to find roots, equations of degrees 3 and 4 have known formulas, but they are rather messy and for higher degrees, it has been established that there cannot be any simple minded formulas. Only suitable procedures for finding approximate solutions to a desired accuracy can be established.
We may, occasionally, denote our set of underlying numbers by the letter F (to remind us of its technical mathematical term - a field). Usually, we may just call them numbers and we will usually be talking about the field of real numbers i.e. $F = \Re$, but other fields are certainly possible.
There are some fields which have only finitely many numbers, but have rich algebraic properties. These are known as finite fields and have many applications. From time to time, we may invite you to consider these number fields and imagine what will happen to our results when you switch over to them. However, they will not form an integral part of this course.

1.1.1 Working with Complex Numbers.

We list a few useful definitions about complex numbers which help in calculations.

1. **Real and Imaginary Parts.**

   Let $z = x + iy$ be a complex number so that x, y are real. Then x is called its real part and is denoted as $Re(z)$.

   Similarly, y is called its imaginary part and is denoted as $Im(z)$.

   It is important to note that the imaginary part itself is very real and the "i" is not included in it.

   Thus, any complex number z is always equal to $Re(z) + iIm(z)$.

2. **Complex conjugate.**

   Let $z = x + iy$ be a complex number so that x, y are real. Then we define its complex conjugate (or just called the conjugate) by

   $$\overline{z} = x - iy = Re(z) - iIm(z).$$

This works without simplifying z. Thus,

$$\text{if } z = \frac{1+i}{1-i} \text{ then } \overline{z} = \frac{1-i}{1+i}.$$

Thus, we do not have to explicitly write it as $x + iy$ to find its conjugate.

Computing conjugates. It is easy to check that given complex numbers z, w we have:

$$\overline{z+w} = \overline{z} + \overline{w},\ \overline{zw} = \overline{z}\,\overline{w} \text{ and if } w \neq 0 \text{ then } \overline{\left(\frac{z}{w}\right)} = \frac{\overline{z}}{\overline{w}}.$$

In turn, the conjugate can help us figure out the real and imaginary parts thus:

$$Re(z) = \frac{z+\overline{z}}{2} \text{ and } Im(z) = \frac{z-\overline{z}}{2i}.$$

Thus, for $z = \frac{1+i}{1-i}$ as above, we get:

$$\begin{array}{rcl} Re(z) & = & \frac{1}{2}\left(\frac{1+i}{1-i} + \frac{1-i}{1+i}\right) \\ & = & \frac{1}{2}\left(\frac{(1-2i+i^2)+(1+2i+i^2)}{1-i^2}\right) \\ & = & \frac{1}{2}\left(\frac{2+2i^2}{1-i^2}\right) \\ & = & 0 \end{array}.$$

Also,

$$\begin{array}{rcl} Im(z) & = & \frac{1}{2i}\left(\frac{1+i}{1-i} - \frac{1-i}{1+i}\right) \\ & = & \frac{1}{2i}\left(\frac{(1-2i+i^2)-(1+2i+i^2)}{1-i^2}\right) \\ & = & \frac{1}{2i}\left(\frac{-4i}{1-i^2}\right) \\ & = & -2 \end{array}.$$

Thus, our z was simply $Re(z) + iIm(z) = 0 - 2i = -2i$.

3. **Absolute Value.** Recall that a real number x has an absolute value $|x|$ which is defined as:

   $$|x| = x \text{ if } x \geq 0 \text{ and } -x \text{ otherwise.}$$

   It is important to keep in mind that $|x|$ represents the **distance to the origin** from x.

   If $x \neq 0$, then $\frac{x}{|x|} = \pm 1$. We may simply **declare this to be the direction (from 0) of a non zero number x**. Note that this direction evaluates to 1 if the number x is positive and -1 if it is negative.

   Naturally, we have an absolute value for a complex number $z = x + iy$ defined as:

   $$|z| = \sqrt{x^2 + y^2} \text{ and this is clearly a non negative real number.}$$

What takes place of the sign? Since the complex numbers live in the plane, there are infinitely many directions available to them.

Any complex number z with $|z| = 1$ represents a direction from the origin to it and all such numbers lie on the unit circle defined by $x^2 + y^2 = 1$. [5]

As before, we can compute the absolute value more efficiently by noting:

$$z\overline{z} = x^2 - i^2y^2 = x^2 + y^2 = |z|^2 \text{ or, simply} |z| = \sqrt{z\overline{z}}.$$

We note that $|z|$ is in fact a real number and this is simply the usual square root of the non negative number $z\overline{z}$!

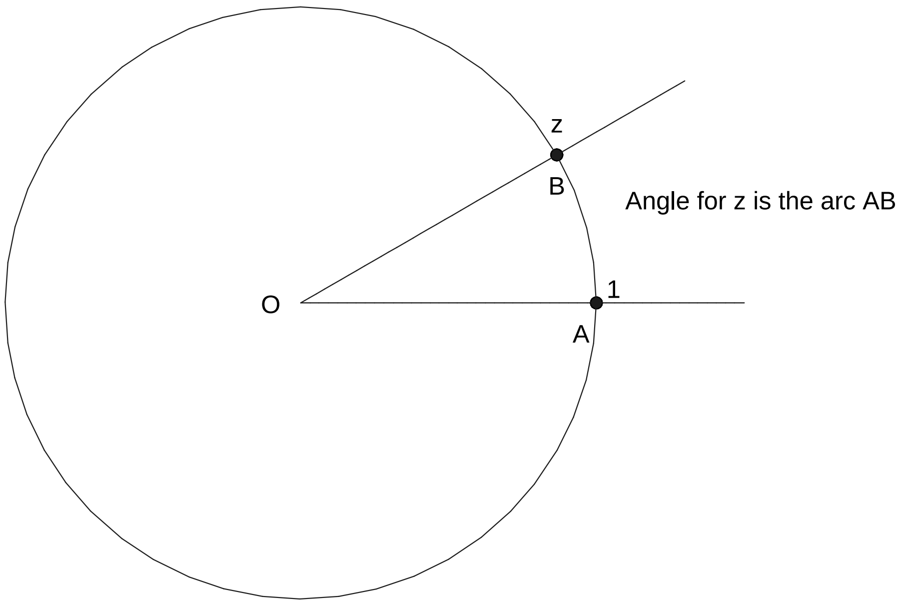
We may find it useful to declare $|z|$ as the length of the complex number z. As before, when $z \neq 0$ is a complex number, then we **declare $\frac{\mathbf{z}}{|\mathbf{z}|}$ to be its direction (from 0)**.

4. Euler Representation of a Complex Number.

   There is a very convenient way of thinking of complex numbers and calculating with them which often attributed to Euler. We don't have enough tools to give its complete proof, but it is very useful to learn to use it and learn the proof later.

   As noted above, every complex number $z = x + iy$ such that $|z| = 1$ lies on the unit circle $x^2 + y^2 = 1$. It would be convenient to describe all these points by convenient single real numbers. We simply take the angle for z to be the measure of the arclength s measured along the unit circle from the point corresponding to the number 1 to the point corresponding to the number z. See the picture below.

[5]This is consistent with the real definition. Since a real number can be thought of as a complex number with $y = 0$ the only real numbers with unit length are x with $x^2 = 1$. Thus we get $x = \pm 1$, giving the two directions from origin!



Note that the arclength of the whole circle is 2π and thus adding any multiple of 2π to s will land us in the same point as z.

For a complex number z with $|z| = 1$, we define its **argument** to be the arclength s from 1 to z measured counter clockwise. Formally, we define $Arg(z) = s$, the **principal value** of the argument which is a non negative real number less than 2π. The general values of the argument are said to be any members of the set $\{s + 2n\pi\}$ where n is an integer. This set is called $arg(z)$.

Finally, when z is any non zero complex number, then we define $Arg(z) = Arg(w)$ where w is complex number $\frac{z}{|z|}$ of length 1 in the direction of z.

The Euler representation of a complex number states that any non zero complex number z can be **uniquely** written as:

$$z = r\exp(i\theta) \text{ where } r = |z| \text{ and } \theta = Arg(z).$$

This requires some explanation.

The exponential function $\exp(t)$ can be formally defined as

$$\exp(t) = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \cdots + \frac{t^n}{n!} + \cdots .$$

The right hand side makes complete sense if we add up terms up to some large enough value n.

It takes much deeper analysis to define the infinite sum.

Here, we shall only worry about the formal algebraic properties, leaving the finer details for future study.

5. We will now show how the series for $\exp(t)$ is related to the following series for the important sine and cosine functions:

$$\cos(t) = 1 - \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots + (-1)^n \frac{t^{2n}}{(2n)!} + \cdots$$

and

$$\sin(t) = \frac{t}{1!} - \frac{t^3}{3!} + \cdots + (-1)^{n-1} \frac{t^{2n-1}}{(2n-1)!} + \cdots .$$

6. Note that:

$$\begin{aligned} \exp(it) &= 1 + \tfrac{it}{1!} + \tfrac{(it)^2}{2!} + \tfrac{(it)^3}{3!} + \cdots \\ &= \left(1 + \tfrac{i^2t^2}{2!} + \tfrac{i^4t^4}{4!} + \cdots + \tfrac{i^{2n}t^{2n}}{(2n)!} + \cdots\right) \\ &\quad + \left(\tfrac{it}{1!} + \tfrac{i^3t^3}{3!} + \cdots + \tfrac{i^{2n-1}t^{2n-1}}{(2n-1)!} + \cdots\right) \\ &= \left(1 - \tfrac{t^2}{2!} + \tfrac{t^4}{4!} + \cdots + (-1)^n \tfrac{t^{2n}}{(2n)!} + \cdots\right) \\ &\quad + i\left(\tfrac{t}{1!} - \tfrac{t^3}{3!} + \cdots + (-1)^{n-1} \tfrac{t^{2n-1}}{(2n-1)!} + \cdots\right). \end{aligned}$$

To understand the above work, note that

$$i^{2n} = (i^2)^n = (-1)^n$$

and

$$i^{2n-1} = (i)(i^{2n-2}) = i(i^2)^{n-1} = i(-1)^{n-1}.$$

In higher mathematics, the trigonometric functions $\sin(t), \cos(t)$ (that you may have seen in high school) are defined by:

$$\cos(t) = Re(exp(it)) \textbf{ and } \sin(t) = Im(exp(it)).$$

This means: [6]

$$\mathbf{exp(it) = cos(t) + i\,sin(t)}$$

7. It is instructive to use your calculator to compare the two sides of this formula.

   Evaluate the value of say $\sin(1)$ and $\cos(1)$ directly by using the calculator key. (Be sure to put the calculator in the "radian" mode.) Then calculate the

[6] This leads to the well known "formula"

$$e^{i\pi} = -1$$

which connects three famous numbers $e = \exp(1)$, i and π with the usual 1.

sum of several terms of this formula and compare how the answers change. Of course, when a calculator gives an answer, it is subject to approximation error in evaluation as well as addition, so don't expect precise agreement!

An infinite precision computer program (like Maple) might be more suited for this exercise.

The main point of this discussion is this: Euler's formulas give a completely precise definition of complicated functions $\sin(t), \cos(t)$ with built in polynomial approximations.

The final aim of Calculus is to push this technique to more and more functions, ideally to all well behaved functions.

1.2 Indeterminates, variables, parameters

It is useful to set up some precise meaning of usual algebraic terms to avoid future misunderstanding.
If you see an expression like $ax^2 + bx + c$; you usually think of this as an expression in a variable x with constants a, b, c.
At this stage, x as well as a, b, c are unspecified. However, we are assuming that a, b, c are fixed while x is a variable which may be pinned down after some extra information - for example, if the expression is set equal to zero.
Thus, in algebraic expressions, the idea of who is constant and who is a variable is a matter of declaration or convention.

Variables

A variable is a symbol, typically a Greek or Standard English letter, used to represent an unknown quantity. However, not all symbols are variables and typically a simple declaration tells us which is a variable and which is a constant.
Take this familiar expression for example:

$$mx + c.$$

On the face of it, all the letters "m, c, x" are potential variables.
You may already remember seeing this in connection with the equation of a line.
Recall that m in $mx + c$ used to be a constant and not a variable. The letter m was the slope of a certain line and was a fixed number when we knew the line.
The letter c in $mx + c$ also used to be a constant. It used to be the y-intercept (i.e. the y-coordinate of the intersection of the line and the y-axis).
The x in $mx + c$ is the variable in this expression and represents a changing value of the x-coordinate. The expression $mx + c$ then gives the y-coordinate of the corresponding point on the line.
Thus, when dealing with a specific line, only one of the three letters is a true "variable".

Parameters

Certain variables are often called parameters; here the idea is that a **parameter** is a variable which is not intended to be pinned down to a specific value, but is expected to move through a preassigned set of values generating interesting quantities. In other words, a parameter is to be thought of as unspecified constant in a certain range of values.

Often, a parameter is described by the oxymoron **"a variable constant"**, to remind us of this feature.

In the above example of $mx + c$ we may think of m, c as parameters. They can change, but once a line is pinned down, they are fixed. Then only the x stays a variable.

As another example, let $1 + 2t$ be the distance of a particle from a starting position at time t, then t might be thought of as a parameter describing the motion of the particle. It is pinned down as soon as we locate the moving point.

The same t can also be simply called a variable if we start asking a question like

"At what time is the distance equal to 5?".

Then we will write an equation $1 + 2t = 5$ and by algebraic manipulation, declare that $t = 2$.

Indeterminates

An **indeterminate** usually is a symbol which is not intended to be substituted with values; thus it is like a variable in appearance, but we don't care to assign or change values for it. Thus the statement

$$X^2 - a^2 = (X - a)(X + a)$$

is a formal statement about expansion and X, a can be thought of as indeterminates. When we use it to factor $y^2 - 9$ as $(y - 3)(y + 3)$ we are turning X, a into variables and substituting values $y, 3$ to them! [7]

1.3 Basics of Polynomials

We now show how polynomials are constructed and handled. This material is very important for future success in Algebra.

Monomials

Suppose that we are given a variable x, a constant c and a non negative integer n. Then the expression cx^n is said to be **a monomial**.

- The constant c is said to be its **coefficient** .(Sometimes we call it the coefficient of x^n).

[7] In short, the distinction between an indeterminate, variable and parameter, is, like beauty, in the eye of the beholder!

- If the coefficient c is non zero, then the **degree** of the monomial is defined to be n.

  If $c = 0$, then we get the **zero monomial** and its degree is undefined.

- It is worth noting that n is allowed to be zero, so 2 is also a monomial with coefficient 2 and degree 0 in any variable (or variables) of your choice!

Sometimes, we need **monomials with more general exponents**, meaning we allow n to be a negative integer as well. The rest of the definition is the same. Sometimes, we may even allow n to be a positive or negative fraction or even a real number. Formally, this is easy, but it takes some effort and great care to give a meaning to the resulting quantity.
We will always make it clear if we are using such general exponents.
For example, consider the monomial $4x^3$. Its variable is x, its coefficient is 4 and its degree is 3.
Consider another example $\frac{2}{3}x^5$. Its variable is x, its coefficient is $\frac{2}{3}$ and its degree is 5.
Consider a familiar expression for the area of a circle of radius r, namely πr^2. This is a monomial of degree 2 in r with coefficient π. Thus, even though it is a Greek letter, the symbol π is a well defined fixed real number. [8]
This basic definition can be made fancier as needed.
Usually, n is an integer, but in higher mathematics, n can be a more general object. As already stated, we can accept $-\frac{5}{2x^3} = (-5/2)x^{-3}$ as an acceptable **monomial with general exponents**. We declare that its variable is x and it has degree -3 with coefficient $-5/2$.
Notice that we collected all parts (including the minus sign) of the expression except the power of x to build the coefficient!

Monomials in several variables.

It is permissible to use several variables and write a monomial of the form cx^py^q. This is a monomial in x, y with **coefficient** c. Its **degree** is said to be $p + q$. Its **exponents** are said to be (p, q) respectively (with respect to variables x, y). As before, more general exponents may be allowed, if necessary. [9]

---

[8]Note that this famous number has a long history associated with it and people have spent enormous energy in determining its decimal expansion to higher and higher accuracy. There is no hope of ever listing all the infinitely many digits, since they cannot have any repeating pattern because the number is not rational. Indeed, it has been proved to be transcendental, meaning it is not the solution to any polynomial equation with integer coefficients!
It is certainly very interesting to know more about this famous number!

[9]For the reader who likes formalism, here is a general definition. A monomial in variables $x_1, \cdots, x_r$ is an expression of the form $cx^{n_1} \cdots x^{n_r}$ where $n_1, \cdots, n_r$ are non negative integers and c is any expression which is free of the variables $x_1, \cdots, x_r$. The coefficient of the monomial is c, its degree is $n_1 + \cdots + n_r$ and the exponents are said to be $n_1, \cdots, n_r$.
If convenient, $n_1, \cdots, n_r$ may be allowed to be more general.

Example.
Consider the following expression which becomes a monomial after simplification.
expression in $\mathbf{x}, \mathbf{y}$.

$$\frac{21x^3y^4}{-3xy}.$$

First, we **must simplify** the expression thus:

$$\frac{21x^3y^4}{-3xy} = \frac{21}{-3} \cdot \frac{x^3}{x} \cdot \frac{y^4}{y} = -7x^2y^3.$$

Now it is a monomial in x, y.
What are the exponents, degree etc.? Its coefficient is -7, its degree is $2 + 3 = 5$ and the exponents are $(2, 3)$.
What happens if we change our idea of who the variables are?
Consider the same simplified monomial $-7x^2y^3$.
Let us think of it as a monomial in y. Then we rewrite it as:

$$(-7x^2)y^3.$$

Thus as a monomial in y alone, its **coefficient** is $-7x^2$ and its degree is 3.
What happens if we take the same monomial $\mathbf{-7x^2y^3}$ and make x as our variable?
Then **we rewrite it as:**

$$(-7y^3)x^2.$$

Be sure to note the rearrangement!
Thus as a monomial in x alone, its coefficient is $-7y^3$ and its degree is 2.
Here is yet another example. Consider the monomial

$$\frac{2x}{y} = 2x^1y^{(-1)}.$$

Think of this as a monomial in x, y with **more general exponents** allowed.
As a monomial in x, y its degree is $1 - 1 = 0$ and its coefficient is 2. Its exponents are $(1, -1)$ respectively.

Binomials, Trinomials, $\cdots$ and Polynomials

We now use the monomials as building blocks to build other algebraic structures. When we add monomials together, we create binomials, trinomials, and more generally polynomials.
It is easy to understand these terms by noting that the prefix "mono" means one. The prefix "bi" means two, so it is a sum of two monomials. The prefix "tri" means three, so it is a sum of three monomials. [10]

[10]Can you find or make up names for a polynomial with four or five or six terms?

In general "poly" means many so we define:

Definition: Polynomial
A polynomial is defined as a sum of finitely many monomials whose exponents are all non negative. Thus $f(x) = x^5 + 2x^3 - x + 5$ is a polynomial.
The four monomials $x^5, 2x^3, -x, 5$ are said to be **the terms of the polynomial f(x).**
Usually, a polynomial is defined to be **simplified** if all terms with the same exponent are combined into a single term.
For example, our polynomial is the simplified form of $x^5 + x^3 + x^3 + 3x - 4x + 5$ and many others.
An important convention to note: Before you apply any definitions like the degree and the coefficients etc., you should make sure that you have collected all like terms together and identified the terms which are non zero after this collection.
For example, $f(x)$ above can also be written as

$$f(x) = x^5 + 2x^3 + x^2 - x + 5 - x^2$$

but we don't count x^2 or $-x^2$ among its terms!
A polynomial with no terms has to admitted for algebraic reasons and it is denoted by just 0 and called the **zero polynomial** .
About the notation for a polynomial: When we wish to identify the main variable of a polynomial, we include it in the notation. If several variables are involved, several variables may be mentioned. Thus we may write:
$p(x) = x^3 + ax + 5$, $h(x, y) = x^2 + y^2 - r^2$ and so on. With this notation, all variables other than the ones mentioned in the left hand side are treated as constants for that polynomial.
The coefficient of a specific power (or exponent) of the variable in a polynomial is the coefficient of the corresponding monomial, **provided we have collected all the monomials having the same exponent.**
Thus the coefficient of x^3 in $f(x) = x^5 + 2x^3 - x + 5$ is 2. Sometimes, we find it convenient to say that the x^3 term of $f(x)$ is $2x^3$.
The coefficient of the missing monomial x^4 in $f(x)$ is declared to be 0. The coefficient of x^{100} in $f(x)$ is also 0 by the same reasoning.
Indeed, we can think of infinitely many monomials which have zero coefficients in a given polynomial. They are not to be counted among the terms of the polynomial.
What is the coefficient of x^3 in $x^4 + x^3 - 1 - 2x^3$. Remember that we must collect like terms first to rewrite

$$x^4 + (x^3 - 2x^3) - 1 = x^4 - x^3 - 1$$

and then we see that the coefficient is -1.

1.3.1 Rational functions.

Just as we create rational numbers from ratios of two integers, we create rational functions from ratios of two polynomials.

Definition: Rational function A rational function is a ratio of two polynomials $\frac{p(x)}{q(x)}$ where $q(x)$ is assumed to be a non zero polynomial.
Examples.

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

Suppose we have rational functions $h_1(x) = \frac{p_1(x)}{q_1(x)}$ and $h_2(x) = \frac{p_2(x)}{q_2(x)}$.
We define algebraic operations of rational functions, just as in rational numbers.

$$h_1(x) \pm h_2(x) = \frac{p_1(x)q_2(x) \pm p_2(x)q_1(x)}{q_1(x)q_2(x)}$$

and

$$h_1(x)h_2(x) = \frac{p_1(x)p_2(x)}{q_1(x)q_2(x)}.$$

It is easy to deduce that $h_1(x) = h_2(x)$ or $h_1(x) - h_2(x) = 0$ if and only if the the numerator of $h_1(x) - h_2(x)$ is zero, i.e.

$$p_1(x)q_2(x) - p_2(x)q_1(x) = 0.$$

For a rational number, we know that multiplying the numerator and the denominator by the same non zero integer does not change its value. (For example $\frac{4}{5} = \frac{8}{10}$.)
Similarly, we have:

$$\frac{p(x)}{q(x)} = \frac{d(x)p(x)}{d(x)q(x)}$$

for any non zero polynomial $d(x)$. (For example $\frac{x}{x+1} = \frac{x^2-x}{x^2-1}$.)

1.4 Working with polynomials

Given several polynomials, we can perform the usual operations of addition, subtraction and multiplication on them. We can also do division, but if we expect the answer to be a polynomial again, then we have a problem. We will discuss these matters later. [11]

[11]The situation is similar to integers. If we wish to divide 7 by 2 then we get a rational number $\frac{7}{2}$ which is not an integer any more. It is fine if you are willing to work with the rational numbers.

As a simple example, let

$$f(x) = 3x^5 + x \ , \ g(x) = 2x^5 - 2x^2 \ , \ h(x) = 3 \ , \ \text{ and } w(x) = 2.$$

Then

$f(x)+g(x) = 3x^5+x+2x^5-2x^2$ and after collecting terms $f(x)+g(x) = 5x^5-2x^2+x$.

Notice that we have also arranged the monomials in decreasing degrees, this is a recommended practice for polynomials.
What is $w(x)f(x) - h(x)g(x)$?
We see that

$$(2)(3x^5 + x) - (3)(2x^5 - 2x^2) = 6x^5 + 2x - 6x^5 + 6x^2 = 6x^2 + 2x.$$

Definition: Degree of a polynomial with respect to a variable x is defined to be the *highest* degree of any monomial present in the simplified form of the polynomial (i.e. any term cx^m with $c \neq 0$ in the simplified form of the polynomial). We shall write $\deg_x(p)$ for the degree of p with respect to x.
The degree of a zero polynomial is defined differently by different people. Some declare it not to have a degree, others take it as -1 and yet others take it as $-\infty$. We shall **declare it undefined** and hence we always have to be careful to determine if our polynomial reduces to 0.
We shall use:

Definition: Leading coefficient of a polynomial
For a polynomial $p(x)$ with degree n in x, by its leading coefficient, we mean the coefficient of x^n in the polynomial.
Thus given a polynomial $x^3 + 2x^5 - x - 1$ we first rewrite it as $2x^5 + x^3 - x - 1$ and then we can say that its degree is 5, and its leading coefficient is 2.
The following are some of the evident facts about polynomials and their degrees.
Assume that

$$u = ax^n + \cdots \ , \ v = bx^m + \cdots$$

are **non zero polynomials** in x of **degrees** $\mathbf{n}, \mathbf{m}$ respectively.
We shall need the following **important observations**.

1. If we say that "u has degree n", then we mean that the coefficient a of x^n is non zero and that none of the other (monomial) terms has degree as high as n.

   Similarly, if we say that "v has degree m", then we mean that $b \neq 0$ and no other terms in v are of degree as high as m.

You might have also seen the idea of division with remainder which says divide 7 by 2 and you get a quotient of 3 and a remainder of 1, i.e. $7 = (2)(3) + 1$.
As promised, we shall take this up later.

2. If c is a non zero number, then

$$cu = cax^n + \cdots$$

has the same degree n and its leading coefficient is ca.

For example, if c is a non zero number, then the degree of $c(2x^5 + x - 2)$ is always 5 and the leading coefficient is $(c)(2) = 2c$.

For $c = 0$ the degree becomes undefined!

3. Suppose that the degrees n, m of u, v are unequal.

Given constants c, d, what is the degree of $cu + dv$?

If c, d are non zero, then the degree of $cu + dv$ is the maximum of n, m. If one of c, d is zero, then we have to look closely.

For example, if $u = 2x^5 + x - 2$ and $v = -x^3 + 3x - 1$, then the degree of $cu + dv$ is determined thus:

$$cu+dv = c(2x^5+x-2)+d(-x^3+3x-1) = (2c)x^5+(-d)x^3+(c+3d)x+(-2c-d)$$

Thus, if $c \neq 0$, then degree will be 5. The leading coefficient is $2c$.

If $c = 0$ but $d \neq 0$, then

$$cu + dv = (-d)x^3 + (3d)x + (-d)$$

and clearly has degree 3. The leading coefficient is $-d$.

If c, d are both zero then $cu + dv = 0$ and the degree becomes undefined.

4. Now suppose that the degrees of u, v are the same, i.e. $n = m$.

Let c, d be constants and consider $h = cu + dv$.

Then the degree of h needs a careful analysis.

Since $m = n$, we see that

$$h = (ca + db)x^n + \cdots$$

and hence we have:

- either $h = 0$ and hence $\deg_x(h)$ is undefined, or
- $0 \leq \deg_x(h) \leq n = m$.

For example, let $u = -x^3 + 3x^2 + x - 1$ and $v = x^3 - 3x^2 + 2x - 2$. Calculate $cu + dv$ with c, d constants.

$$\begin{aligned} h = cu + dv &= c(-x^3 + 3x^2 + x - 1) + d(x^3 - 3x^2 + 2x - 2) \\ &= (-c + d)x^3 + (3c - 3d)x^2 + (c + 2d)x + (-c - 2d) \end{aligned}$$

We determine the degree of the expression thus:

- If $-c + d \neq 0$, i.e. $c \neq d$ then the degree is 3.
- If $c = d$, then the x^2 term also vanishes and we get

  $$cu + dv = (c + 2d)x + (-c - 2d) = (3d)x + (-3d).$$

  Thus if $d = 0$ then the degree is undefined and if $d \neq 0$ then the degree is 1.

You are encouraged to figure out the formula for the leading coefficients in each case.

You should also experiment with other polynomials.

5. **The product rule for degrees.** The degree of the product of two *non zero* polynomials is always the sum of their degrees. This means:

   $$\deg_x(uv) = \deg_x(u) + \deg_x(v) = m + n.$$

   Indeed, we can see that

   $$\begin{aligned} uv &= (ax^n + \text{ terms of degree less than } n\ )(bx^m + \text{ terms of degree less than } m\ ) \\ &= (ab)x^{(m+n)} + \text{ terms of degree less than } m + n\ . \end{aligned}$$

   The leading coefficient of the product is ab, i.e. the product of their leading coefficients. Thus we have an **interesting principle.**

   The degree and the leading coefficient of a product of two non zero polynomials can be calculated by ignoring all but the leading terms in each!

   By repeated application of this principle, we can see the

   The power rule for degrees. If $u = ax^n + \cdots$ is a polynomial of degree n in x, then for any positive integer d, the polynomial u^d has degree dn and leading coefficient a^d.

   You should think and convince yourself that it is natural to define $u^0 = 1$ for all non zero polynomials u. Thus, the same formula can be assumed to hold for $d = 0$.

1.5 Examples of polynomial operations.

- **Example 1:** Consider $p(x) = x^3 + x + 1$ and $q(x) = -x^3 + 3x^2 + 5x + 5$. What are their degrees?

  Also, calculate the following expressions:

  $$p(x) + q(x) \quad , \quad p(x) + 2q(x) \text{ and } p(x)^2$$

  and their degrees.

  Determine the leading coefficients for each of the answers.

  Answers: The degrees of $p(x), q(x)$ are both 3 and their leading coefficients are respectively $1, -1$.

  The remaining answers are:

  $$\begin{array}{lcl} p(x) + q(x) & = & x^3(1-1) + x^2(0+3) + x(1+5) + (1+5) \\ & = & 3x^2 + 6x + 6 \\ \hline p(x) + 2q(x) & = & x^3(1-2) + x^2(0+6) + x(1+10) + (1+10) \\ & = & -x^3 + 6x^2 + 11x + 11 \\ \hline (x^3+x+1)(x^3+x+1) & = & (x^6 + x^4 + x^3) + (x^4 + x^2 + x) + (x^3 + x + 1) \\ & = & x^6 + 2x^4 + 2x^3 + x^2 + 2x + 1 \\ \hline \end{array}$$

  The respective degrees are, $2, 3, 6$. The leading coefficients are, respectively, $3, -1, 1$.

  Here is **an alternate technique** for finding the last answer. It will be useful for future work.

  First, note that the degree of the answer is $3 + 3 = 6$ from the product rule above. Thus we need to calculate coefficients of all powers x^i for $i = 0$ to $i = 6$.

  Each x^i term comes from multiplying an x^j term and an x^{i-j} term from $p(x)$. Naturally, both j and $i - j$ need to be between 0 and 3.

  Thus the only way to get x^0 in the answer is to take $j = 0$ and $i - j = 0 - 0 = 0$. Thus the x^0 term of the answer is the square of the x^0 term of $p(x)$, hence is $1^2 = 1$.

  Similarly, the x^6 term can only come from x^3 term multiplied by the x^3 term and is x^6.

  Now for the x^5 term, the choices are $j = 3, i - j = 2$ or $j = 2, i - j = 3$ and in $p(x)$ the coefficient of x^2 is 0, so both these terms are 0. Hence the x^5 term is missing in the answer.

Now for x^4 terms, we have three choices for $(j, i-j)$, namely $(3,1),(2,2),(1,3)$. We get the corresponding coefficients $(1)(1)+(0)(0)+(1)(1)=2$.

The reader should verify the rest.

- **Example 2:** Use the above technique to answer the following:

$$\begin{array}{rl} \text{Assume} & (x^{20}+2x^{19}-x^{17}+\cdots+3x+5) \\ & \times(3x^{20}+6x^{19}-4x^{18}+\cdots+5x+6) \\ & = \ ax^{40}+bx^{39}+cx^{38}+\cdots+dx+e \end{array}$$

Determine a, b, c, d, e.

Note that the middle terms are not even given and not needed for the required answers. The only term contributing to x^{40} is $(x^{20}) \cdot (3x^{20})$, so $a=3$.

The only terms producing x^{39} are

$$(x^{20}) \cdot (6x^{19}) \text{ and } (2x^{19}) \cdot (3x^{20})$$

and hence, the coefficient for x^{39} is $(1)(6)+(2)(3)=12$, so $b=12$.

Similarly, check that $c=(1)(-4)+(2)(6)+(0)(3)=8$. Calculate $d=(3)(6)+(5)(5)=43$ and $e=(5)(6)=30$.

- **Example 3:** The technique of polynomial multiplication can often be used to help with integer calculations. Remember that a number like 5265 is really a polynomial $5d^3+2d^2+6d+5$ where $d=10$.

Now, we can use our technique of polynomial operations to add or multiply numbers. The only thing to watch out for is that since d is a number and not really a variable, there are “carries” to worry about.

Here we derive the well known formula for squaring a number ending in 5. Before describing the formula, let us give an example.

Say, you want to square the number 25. Split the number as $2|5$. From the left part 2 construct the number $(2)(3)=6$. This is obtained by multiplying the left part with itself increased by 1. Simply write 25 next to the current calculation, so the answer is 625.

The square of 15 by the same technique shall be obtained thus:

Split it as $1|5$. $(1)(1+1)=2$. So the answer is 225.

The general rule is this:

Let a number n be written as $p|5$ where p is the part of the number after the units digit 5.

Then $n^2 = (p)(p+1)25$, i.e. $(p)(p+1)$ becomes the part of the number from the 100s digit onwards.

To illustrate, consider 45^2. Here $p = 4$, so $(p)(p+1) = (4)(5) = 20$ and so the answer is 2025. Also $105^2 = (10)(11)25 = 11025$. Similarly $5265^2 = (526)(527)25$, i.e. 27720225.

What is the proof? Since n splits as $p|5$, our $n = pd + 5$. Using that $d = 10$ we make the following calculations.

$$\begin{array}{rcl} n^2 & = & (pd+5)(pd+5) \\ & = & p^2d^2 + (2)(5)(pd) + 25 \\ & = & p^2(100) + p(100) + 25 \\ & = & (p^2+p)100 + 25. \end{array}$$

Thus the part after the 100s digits is $p^2 + p = p(p+1)$.

You can use such techniques for developing fast calculation methods.

- **Example 4:** Let us begin by $f(x) = x^n$ for $n = 1, 2, \cdots$ and let us try to calculate the expansions of $f(x+t)$. For $n = 1$ we simply get $f(x+t) = x + t$.

  For $n = 2$ we get $f(x+t) = (x+t)^2 = x^2 + 2xt + t^2$.

  For $n = 3$ we get $f(x+t) = (x+t)^3$. Let us calculate this as follows:

$$\begin{array}{rcl} (x+t)(x+t)^2 & = & (x+t)(x^2+2xt+t^2) \\ & = & x(x^2+2xt+t^2) \\ & & +t(x^2+2xt+t^2) \\ & = & x^3 + 2x^2t + xt^2 \\ & & \quad + x^2t + 2xt^2 + t^3 \\ & = & x^3 + 3x^2t + 3xt^2 + t^3 \end{array}$$

  For $n = 4$ we invite the reader to do a similar calculation and deduce[12]

$$\begin{array}{rcl} (x+t)^4 & = & (x+t)(x^3+3x^2t+3xt^2+t^3) \\ & = & x^4 + (3x^3t + x^3t) + (3x^2t^2 + 3x^2t^2) + (xt^3 + 3xt^3) + t^4 \\ & = & x^4 + 4x^3t + 6x^2t^2 + 4xt^3 + t^4. \end{array}$$

[12] An alert reader may see similarities with the process of multiplying two integers by using one digit of the multiplier at a time and then adding up the resulting rows of integers. Indeed, it is the result of thinking of the numbers as polynomials in 10, but we have to worry about carries. Here is a sample multiplication for your understanding:

$$\begin{array}{cccc} & 4 & 3 & 1 \\ & \times & 1 & 2 \\ \hline & 8 & 6 & 2 \\ 4 & 3 & 1 & \\ \hline 5 & 1 & 7 & 2 \end{array}$$

Let us step back and list our conclusions.

n	Coefficient list							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	35	35	21	7	1

We note a pattern from first few rows, as well as recalling how we collected the coefficients for small values of n. The pattern is that a coefficient is the sum of the one above and the one in its northwest corner (above-left), where any coefficient which is missing is treated as zero.

We have then boldly gone where we had not gone before and filled in the rows for $n = 5, 6, 7$. What do they tell us?

The row for $n = 5$ says that

$$(x+t)^5 = (1)x^5 + (5)x^4t + (10)x^3t^2 + (10)x^2t^3 + (5)xt^4 + t^5.$$

Notice that we start with the highest power of x, namely x^5 and using the coefficients write successive terms by reducing a power of x while increasing a power of t, until we end up in t^5.

Clearly, we can continue this pattern as long as needed. This arrangement of coefficients is often referred to as the Pascal (17th Century) triangle, but it was clearly known in China (13th century) and India (10th century or earlier). A better name would be the Halāyudha arrangement to be named after the commentator who gave it in his tenth century commentary on a very old work by Pingala (from the second century BC).

The general formula called **The Binomial Theorem** is this:

$$(x+t)^n = {}_nC_0x^n + {}_nC_1x^{n-1}t + {}_nC_2x^{n-2}t^2 + \cdots + {}_nC_{n-2}x^2t^{n-2} + {}_nC_{n-1}xt^{n-1} + {}_nC_nt^n.$$

The binomial coefficients ${}_nC_r$ (read as "en cee ar") will now be described.

Recall that for any positive integer m, we write $m!$ for $(1)(2)\cdots(m)$. We also make a special convention that $0! = 1$.

The formula for the binomial coefficients is: [13]

[13]Why the word "binomial"? These are the coefficients coming from powers of a "two term expression" which is called a binomial, where "bi" means two. Indeed, "poly" means many, so our polynomial means a many term expression.

$$_nC_r = \frac{n!}{r!(n-r)!} \text{ or equivalently } \frac{n(n-1)\cdots(n-r+1)}{r!}.$$

(**Hint:** To get the second form, write out the factors in the first formula and cancel terms.)

We shall not worry about a formal proof, but the reader can use the idea of Induction to prove the formula. [14]

Note. The notation, $_nC_r$ is convenient, but many variants of the same are common. Many people use $\binom{\mathbf{n}}{\mathbf{r}}$.

If the complicated format of this notation is not easy to implement, a simpler $\mathbf{C}(\mathbf{n},\mathbf{r})$ is also used.

Why are we giving two different looking formulas, when one would do? The first formula is easy to write down, but the second is easier to evaluate, since it has already done many of the cancellations.

There is a more subtle reason as well; the second formula can be easily evaluated when n is not even an integer and indeed there is a famous Binomial Theorem due to Newton which will let you compute $(1+x)^n$ for any power n, provided you take x small (i.e. between -1 and 1) and provided you get clever enough to add infinitely many terms. We shall discuss this later in the Appendix.

First let us calculate a few binomial coefficients:

$$_3C_2 = \frac{3!}{2!1!} = \frac{1\cdot 2\cdot 3}{1\cdot 2\cdot 1} = 3.$$

Similarly,

$$_5C_3 = \frac{5!}{3!2!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{1\cdot 2\cdot 3\cdot 1\cdot 2} = \frac{3\cdot 4\cdot 5}{1\cdot 2\cdot 3} = \frac{2\cdot 5}{1} = 10.$$

How do we use these? The meaning of these coefficients is that when we expand $(x+t)^n$ and collect the terms $x^r t^{n-r}$ then its coefficient is $_nC_r$.

[14] The main point to realize what we are doing. Here is a nice organization method.

1. Take the known expansion of $(x+t)^2$, say.
2. Multiply the expansion by x and t separately.
3. Add up the two multiplication results by writing the terms below each other while lining up like degree terms vertically.
4. add up the terms and report the answer.
5. Now repeat this idea by replacing 2 by $3,4,5,\cdots,n$. That is the inductive process.

Consider:

$(x+t)^{10} = {}_{10}C_0 x^{10} + {}_{10}C_1 x^9 t + {}_{10}C_2 x^8 t^2 +$ more terms with higher powers of t.

Note that ${}_{10}C_0 = \frac{10!}{0!10!} = 1$. Indeed, ${}_nC_0 = 1 = {}_nC_n$ for all n.

The value of ${}_{10}C_1 = \dfrac{10}{1!} = 10$.

Note that we are using the second form of the formula with $n = 10, r = 1$, so the numerator $n(n-1)\cdots(n-r+1)$ reduces to just 10. Indeed ${}_nC_1 = n$ is easy to see for all n.

$$x^{10} + 10x^9 t + \frac{10\cdot 9}{1\cdot 2}x^8 t^2 = x^{10} + 10x^9 t + 45x^8 t^2.$$

You should verify the set of eleven coefficients:

$$\begin{pmatrix} {}_{10}C_0 & {}_{10}C_1 & {}_{10}C_2 & {}_{10}C_3 & {}_{10}C_4 & {}_{10}C_5 & {}_{10}C_6 & {}_{10}C_7 & {}_{10}C_8 & {}_{10}C_9 & {}_{10}C_{10} \\ 1 & 10 & 45 & 120 & 210 & 252 & 210 & 120 & 45 & 10 & 1 \end{pmatrix}.$$

Note the neat symmetry of the coefficients around the middle. It is fun to deduce it from the formula alone.

Exercise based on the above formula. Expand $(5-x)^4$ and $(2x+3)^5$, using the binomial theorem.

Answer:

For the first expansion, we try to use:

$$(x+t)^4 = x^4 + 4x^3 t + 6x^2 t^2 + 4xt^3 + t^4.$$

We can clearly match the expression $x+t$ with $5-x$ by replacing $x \to 5$ and $t \to -x$.

Thus, after a careful substitution, we get:

$$(5-x)^4 = (5)^4 + 4(5)^3(-x) + 6(5)^2(-x)^2 + 4(5)(-x)^3 + (-x)^4.$$

After simplification, we conclude:

$$(5-x)^4 = 625 - 500x + 150x^2 - 20x^3 + x^4.$$

Try to deduce the same answer by using a different substitution:
$x \to -x, t \to 5$.

For the second exercise, we make the substitution $x \to 2x, t \to 3$ in the formula for $(x+t)^5$.

Thus, we get:

$$(2x+3)^5 = (2x)^5 + 5(2x)^4(3) + 10(2x)^3(3)^2 + 10(2x)^2(3)^3 + 5(2x)(3)^4 + (3)^5.$$

This simplifies to:

$$(2x+3)^5 = 32\,x^5 + 240\,x^4 + 720\,x^3 + 1080\,x^2 + 810\,x + 243.$$

- **Example 5: Substituting in a polynomial.** Given a polynomial $p(x)$ it makes sense to substitute any valid algebraic expression for x and recalculate the resulting expression. It is important to replace every occurrence of x by the substituted expression *enclosed in parenthesis* and then carefully simplify.

  For example, for

$$\begin{aligned} p(x) &= 3x^3 + x^2 - 2x + 5 \text{ we have} \\ p(-2x) &= 3(-2x)^3 + (-2x)^2 - 2(-2x) + 5 \\ & \quad \text{and this simplifies to} \\ &= 3(-8)x^3 + (4)x^2 - 2(-2)x + 5 \\ &= -24x^3 + 4x^2 + 4x + 5. \end{aligned}$$

  Given below are several samples of such operations. Verify these:

Polynomial	Substitution for x	Answer
$(x+5)(x+1)$	$x+2$	$(x+7)(x+3)$
x^2+6x+5	$x-3$	x^2-4
x^2+6x+5	$2x-3$	$4x^2-4$
x^2+6x+5	$-3x$	$9x^2-18x+5$

Polynomial	Substitution for x	Answer
x^2+6x+5	$\frac{1}{x}$	$\frac{1}{x^2}+\frac{6}{x}+5=\frac{5x^2+6x+1}{x^2}$
x^2+6x+5	$\frac{-2}{x}$	$\frac{4-12x+5x^2}{x^2}$
x^2+6x+5	$\frac{1}{x+1}$	$\frac{12+16x+5x^2}{(x+1)^2}$
$(x+5)(x+1)$	-5	0

  As an example of substituting in a rational function, note that substitution of $(x-5)$ for x in $\dfrac{x+2}{(x+5)(x+1)}$ can be conveniently written as $\dfrac{x-3}{(x)(x-4)}$.

  Note that, if convenient, we leave the answer in factored form. Indeed, the factored form is often better to work with, unless one needs coefficients.

  The substitutions by rational expressions are prone to errors and the reader is advised to work these out carefully!

- **Example 6: Completing the square.**

  Consider a polynomial $q(x) = ax^2 + bx + c$ where $0 \neq a, b, c$ are constants.

  Find a substitution $x = u + s$ such that the substituted polynomial $q(u + s)$ has no u-term (i.e. it has no monomial in u of degree 1).

  In a more colorful language, we say that the u-term in $q(u + s)$ is killed!

  Find the resulting polynomial $q(u + s)$.

  Answer. Check out the work:

  $$q(u+s) = a(u+s)^2 + b(u+s) + c = au^2 + (2as+b)u + (as^2 + bs + c).$$

  We want to arrange $2as + b = 0$ and this is true if $2as = -b$, i.e. $s = -b/(2a)$.

  We substitute this value of s in the final form above. [15]

  $$\begin{aligned} q(u - \frac{b}{2a}) &= au^2 + \left(a(-\frac{b}{2a})^2 + b(-\frac{b}{2a}) + c\right) \\ &= au^2 + \frac{b^2}{4a} - \frac{b^2}{2a} + c \\ &= au^2 - \frac{b^2}{4a} + c = au^2 - (\frac{b^2 - 4ac}{4a}). \end{aligned}$$

  Sometimes, it is better not to bring in a new letter u. So, note that $x = u + s = u - \frac{b}{2a}$ means $u = x + \frac{b}{2a}$.

  Thus our final formula can be rewritten as:

  $$q(x) = a(x + \frac{b}{2a})^2 - \frac{b^2 - 4ac}{4a}.$$

  Thus, we have rewritten our expression $q(x)$ in the form
  "(constant)(square) + constant".

  Hence, this process is also called **completing of the square.**

  Exercise based on the above.

  Let $q(x) = 2x^2 + 5x + 7$.

  Find a substitution $x = u + s$ which makes $q(u + s)$ have no u-term.

  Answer: Here $a = 2, b = 5, c = 7$, so $s = -\dfrac{b}{2a} = -\dfrac{5}{4}$.

  Thus

  $$q(u - \frac{5}{4}) = au^2 - \left(\frac{b^2 - 4ac}{4a}\right) = 2u^2 - \frac{5^2 - 4 \cdot 2 \cdot 7}{4 \cdot 2} = 2u^2 + \frac{31}{8}.$$

[15] It might be tempting to memorize the whole final answer. We recommend only memorizing the substitution $u - \frac{b}{2a}$. It is best to carry out the rest of the simplification as needed.

Note that we have $x = u - \frac{5}{4}$ and hence $u = x + \frac{5}{4}$.

Thus, our final answer can also be written as:

$$q(x) = 2\left(x + \frac{5}{4}\right)^2 + \frac{31}{8}.$$

A more general idea: Indeed, the same idea can be used to kill the term whose degree is one less than the degree of the polynomial.

For example, For a cubic $p(x) = 2x^3 + 5x^2 + x - 2$, the reader should verify that $p(u - 5/6)$ will have no u^2 term.

Challenge! We leave the following two questions as a challenge to prove:

- Given a cubic $p(x) = ax^3 + bx^2 + \text{ lower terms }$ with $a \neq 0$, find the value of s such that $p(u+s) = au^3 + (0)u^2 + \text{ lower terms }$. **Answer:** $s = -\frac{b}{3a}$.
- More generally, do the same for a polynomial $p(x) = ax^n + bx^{(n-1)} + \text{ lower terms }$. This means find the s such that $p(u+s) = au^n + (0)u^{(n-1)} + \text{ lower terms }$. **Answer:** $s = -\frac{b}{na}$.

Chapter 2

Solving linear equations.

As explained in the beginning, solving equations is a very important topic in algebra. In this section, we learn the techniques to solve the simplest types of equations, namely the linear equations.

Definition: Linear Equation An equation is said to be linear in its listed variables $x_1, \cdots, x_n$, if each term of the equation is either free of the listed variables or is a monomial of degree 1 involving exactly one of the $x_1, \cdots, x_n$. [1]
Examples Each of these equations are linear in indicated variables:

1. $y = 4x + 5$: Variables x, y.
2. $2x - 4y = 10$: Variables x, y.
3. $2x - 3y + z = 4 + y - w$: Variables x, y, z.

Now we explain the importance of listing the variables.

1. $xy + z = 5w$ is not linear in x, y, z, w. This is because the term xy has degree 2 in x, y.
2. The same expression $xy + z = 5w$ is linear in x, z, w. It is also linear in y, z, w.

   We can also declare it as simply linear in y.

   Indeed, it is linear in any choice of variables from among x, y, z, w, as long as we don't include both x and y!

[1]This may seem unnecessarily complicated. It is tempting to define the equation to be linear by requiring that the maximum degree of its non zero terms is 1.

However, the problem is that this maximum could be zero, since the supposed variables might be absent!

Why would one intentionally write an equation in x which has no x in it?

This can happen if we are moving some of the parameters of the equation around and x may accidentally vanish! For example the equation $y = mx + c$ is clearly linear in x, but if $m = 0$ it has no x in it!

The point is that we don't have to list every visible symbol as a variable.

3. Sometimes, equations which appear as non linear may reduce to linear ones; but technically they are considered non linear. Thus, $(x-1)(x-5)=(x-2)(x-7)$ is clearly non linear as it stands, but simplifies to $x^2-6x+5=x^2-7x+14$ or $-6x+5=-7x+14$ after cancelling the x^2 term.

4. Sometimes, we can make equations linear by creating new names. Thus:
$$\frac{5}{x}+\frac{3}{y}=1$$
is certainly not linear, but we can set $u=\frac{1}{x}$ and $v=\frac{1}{y}$ and say that the resulting equation $5u+3v=1$ is linear in u,v. Such tricks let us solve some non linear equations as if they are linear!

2.1 What is a solution?

We will be both simplifying and solving equations throughout the course. So what is the difference in merely simplifying and actually solving an equation?

Definition: A solution to an equation. Given an equation in a variable, say x, by a solution to the equation we mean a value of x which makes the equation true. For example, the equation:

$$3x+4=-2x+14$$

has a solution given by $x=2$, since substituting $x=2$ makes the equation:

$$6+4=-4+14 \text{ which simplifies to } 10=10$$

and this is a true equation.
On the other hand, $x=3$ is not a solution to this equation since it leads to:

$$9+4=-6+14 \text{ which simplifies to } 13=8$$

which is obviously false.
How did we find the said solution? In this case we simplified the equation in the following steps:

LHS		RHS	Explanation
$3x+4$	$=$	$-2x+14$	Original equation
$3x+2x+4$	$=$	14	Collect x terms on LHS
$3x+2x$	$=$	$-4+14$	Move constants to RHS
$5x$	$=$	10	Simplify
x	$=$	2	Divide by 5 on both sides

Usually, the above process is described by the phrase "By simplifying the equation we get $x = 2$ ". The reader is expected to carry out the steps and get the final answer.
What does it mean to simplify? It is a process of changing the original equation into a sequence of other equations which have the same set of solutions but whose solutions are easier to deduce.
As another example of a solution, consider the equation $3x + 2y = 7$ and check that $x = 1, y = 2$ gives a solution.
As yet another example, consider $x^2 + yz = 16$ with solution $x = 2, y = 3, z = 4$. Note that these last two examples can have many different solutions besides the displayed ones.

Definition: Consistent equation. A single equation in variables $x_1, \cdots, x_r$ is said to be **consistent** if it has at least one solution $x_1 = a_1, \cdots, x_r = a_r$.
It is said to be **inconsistent** if there is no solution.
For example the equation $x^2 + y^2 + 5 = 0$ cannot have any (real) solution, since the left hand side (LHS) is always positive and the right hand side (RHS) is zero. Thus it is inconsistent when we are working with real numbers.
Since we have already studied complex numbers, we note that $x = (\sqrt{5})i, y = 0$ is a complex solution of the same equation. Thus, existence of a solution also dpends on the field used in our work. Luckily, when dealing with linear equations, the solution, if it exists, can always be found in any field which contains all the coefficients of the equation.

Definition: System of equations By a system of equations in variable $x_1, \cdots, x_r$ we mean a collection of equations $E_1, E_2, \cdots, E_s$ where each E_i is an equation in $x_1, \cdots, x_r$.
Thus

$$E_1 : 3x + 4y = -1 \ , E_2 : 2x - y - z = 1 \ , E_3 : x + y + z = 2 \ , E_4 : y + 2z = 3$$

is a system of four equations in three variables x, y, z.

Definition: Solution of a system of equations By a solution of a system of equations $E_1, E_2, \cdots, E_s$ in variables $x_1, \cdots, x_r$ we mean values for the variables $x_1 = a_1, \cdots, x_r = a_r$ which make all the equations true.
For instance, you can check that $x = 1, y = -1, z = 2$ is a solution of the above system.

Definition: Consistent system of equations. A system of equations is said to be consistent if it has at least one solution and is said to be inconsistent otherwise.
For example the system of equations

$$x + y = 3, x + y = 4$$

is easily seen to be inconsistent.
In general, we want to learn how to find and describe all the solutions to a system of linear equations. In this course, we learn about equations in up to three variables. The general analysis is left for an advanced course in linear algebra.
We discuss methods of solution for systems of linear equations in the following sections.

2.2 One linear equation in one variable.

A linear equation in one variable, say x, can be solved by simple manipulation. We rewrite the equation in the form $ax = b$ where all x terms appear on the left and all terms free of x appear on the right. This is often referred to as **isolating the variable.**
Then the solution is simply $x = b/a$, provided $a \neq 0$. We come up with this solution by dividing both the right hand side (RHS) of the equation and the LHS by a. Thus $ax/a = x$ and b/a is simply b/a.
If $a = 0$ and $b \neq 0$, then the equation is **inconsistent** and has no solution.
If both a, b are zero, then we have an **identity** - an equation which is valid for all values of the variable - i.e. any number we put for x will make the equation true. Therefore there are infinitely many solutions. (We are assuming that we are working with an infinite set of numbers like the usual real numbers.)
This is **the principle of $\mathbf{0}, \mathbf{1}, \infty$ number of solutions!**
It states that any system of linear equations in any number of variables always falls under one of three cases: [2]

1. No solution (inconsistent system) or
2. a unique solution or
3. infinitely many solutions.

Example: The equation

$$-5x + 3 + 2x = 7x - 8 + 9x$$

can be rearranged to:

$$x(-5 + 2 - 7 - 9) = -8 - 3$$

or even more simplified

$$-19x = -11$$

[2] Even though we have stated it, we are far from proving this principle; its proof forms the basic investigation in the course on Linear Algebra and it is a very important result useful in many branches of mathematics.

and hence $x = 11/19$ is a **unique** solution
The reader should note the efficient collection of terms, namely, we collected all the x terms by writing x followed by all the coefficients of x inside parenthesis and with adjusted signs if they crossed the sides of the equation. The reader should also note that the solution is really another equation! Thus **to solve an equation really means to change it to a convenient but equivalent form.**
Example: The equation

$$(3x + 4) + (5x - 8) = (5x - 4) + (3x + 8)$$

would lead to

$$x(3 + 5 - 5 - 3) = -4 + 8 - 4 + 8$$

or

$$0 = 8$$

an **inconsistent** equation; hence there is no solution.
Example: The equation:

$$(3x + 4) + (5x - 8) = (5x + 4) + (3x - 8)$$

reduces to

$$8x - 4 = 8x - 4$$

and is an example of an **identity**. It gives all values of x as valid solutions!

2.3 Several linear equations in one variable.

If we are given several equations in one variable, in principle, we can solve each of them and if our answers are consistent, then we can declare the common solution as the final answer. If not, we shall declare them as inconsistent.
Example: Given equations:

$$2x = 4\ ,\ 6x = 12.$$

Solve each equation and determine if they are consistent or inconsistent.
$2x = 4 \to x = 2$ $6x = 12 \to x = 2$ Since each equation gives us $x = 2$ we have a **unique** solution.
Example:Given equations:

$$2x = 4\ ,\ 6x = 8.$$

Working as above, we get: $2x = 4 \to x = 2$ $6x = 8 \to x = 4/3$ Since the equations give us different solutions, they are **inconsistent**.
One systematic way of checking this is to solve the first one as $x = 2$ and substitute it in the second to see if we get a valid equation. In this case, we get $6(2) = 8$ or $12 = 8$ an inconsistent equation!

2.4 Two or more equations in two variables.

Suppose we wish to solve

$$x + 2y = 10 \text{ and } 2x - y = 5 \text{ together.}$$

The way we handle it is to think of them as equations in x first. Solution of the first equation in x is then $x = 10 - 2y$. If we plug this into the second, we get

$$2(10 - 2y) - y = 5$$

or, by simplification

$$20 - 4y - y = 5 \ \ \textbf{and further} \ \ -5y = -15.$$

Thus, unless $y = 3$ we shall have an inconsistent system on our hands.
Further, when $y = 3$, we have $x = 10 - 2y = 10 - 2(3) = 4$.
Thus our unique solution is $x = 4, y = 3$.
Note: We could have solved the above two linear equations $x + 2y = 10$ and $2x - y = 5$ by a slightly different strategy.
We could have solved the first equation for y as

$$y = \frac{10 - x}{2} = 5 - \frac{x}{2}.$$

Substitute this answer in the second to get $2x - (5 - \frac{x}{2}) = 5$, deduce that $\frac{5x}{2} = 10$ or $x = 4$ and then finally plug this back into $y = 5 - \frac{x}{2}$ to give $y = 3$. Thus we get the same solution again.
The choice of the first variable is up to you, but sometimes it may be easier to solve for one particular variable. With practice, you will be able to identify more easily which variable should be solved first.
Example:
Solve $x + 2y = 10$, $2x - y = 5$ and $3x + 4y = 10$ together.
We already know the unique solution of the first two; namely $x = 4, y = 3$. So, all we need to do is plug it in the third to get $3(4) + 4(3) = 10$ or $24 = 10$. Since this is not true, we get inconsistent equations and declare that that there is no solution.
Example:

$$\text{Solve } x + 2y = 10 \ , \ 2x - y = 5 \text{ and } 3x - 4y = 0 \text{ together;}$$

As above, the first two equations give $x = 4, y = 3$. Plugging this into the third, we get:

$$3(4) - 4(3) = 0$$

which reduces to $0 = 0$ a consistent equation; hence the unique answer is
$x = 4, y = 3$.

2.5 Several equations in several variables.

Here is an example of solving three equations in x, y, z.

$$\text{Solve: Eq1: } x + 2y + z = 4\ ,\ \text{Eq2: } 2x - y = 3\ ,\ \text{Eq3: } y + z = 7.$$

We shall follow the above idea of solving one variable at a time.
We solve Eq1 for x to get

*Solution*1. $$x = 4 - 2y - z$$

Substitution into Eq2 gives a new equations in y, z:

$$2(4 - 2y - z) - y = 3 \text{ or } y(-4 - 1) - 2z = 3 - 8 \text{ or } 5y + 2z = 5.$$

The third equation Eq3 has no x anyway, so it stays $y + z = 7$.
Now we solve the new equations:

$$Eq2^* : 5y + 2z = 5\ , Eq3^* :\ y + z = 7.$$

Our technique says to solve one of them for one of the variables. Note that it is easier to solve the equation $Eq3^*$ for y or z, so we solve it for y:

*Solution*2. $$y = 7 - z$$

Finally, using this, $Eq2^*$ becomes $5(7 - z) + 2z = 5$ or $z(-5 + 2) = 5 - 35$.
Thus $z = -30/(-3) = 10$. So we have:

*Solution*3. $$z = 10$$

Notice our three solutions. The third pins z down to 10.
Then the second says $y = 7 - z = 7 - 10 = -3$.
Finally, the first one gives $x = 4 - 2y - z = 4 - 2(-3) - (10) = 0$ giving the: [3]

Final solution. $$x = 0\ ,\ y = -3\ ,\ z = 10$$

This gives the main idea of our solution process. We summarize it formally below, as an aid to understanding and memorization.

- Pick some equation and a variable, say x, appearing in it.

[3]The three solutions which give solutions for each of x, y, z in terms of later variables can be considered an almost final answer. Indeed, in a higher course on linear algebra, this is exactly what done!

Given a set of equations, we order our variables somehow and try to get a sequence of solutions giving each variable in terms of the later ones. The resulting set of solution equations is often called the row echelon form.

- Solve the chosen equation for x and save this as the first solution. Substitute the first solution in all the other equations.

  These other equations are now free of x and represent a system with fewer variables.

- Continue to solve the smaller system by choosing a new variable etc. as described above.

- When the smaller system gets solved, plug its solution into the original value of x to finish the solution process.

2.6 Solving linear equations efficiently.

- Example 1. The process of solving for a variable and substituting in the other can sometimes be done more efficiently by manipulating the whole equations. Here is an example.

  $$\text{Solve E1: } 2x + 5y = 3 \text{ , E2: } 4x + 9y = 5.$$

  We note that solving for x and plugging it into the other equation is designed to get rid of x from the other equation. So, if we can get rid of x from one of the equation with less work, we should be happy.

  One permissible operation (which still leads to the same final answers) is to change one equation by adding a suitable multiple of one (or more) other equation(s) to it.

  There are two more natural operations which are very familiar. One is to add equal quantities to both sides of the equation and the other is to multiply the whole equation by a non zero number. [4]

  These three types of operations can indeed lead to a complete analysis of solutions for any system of linear equations.

  By inspection, we see that E2-2E1 will get rid of x. So, we replace E2 by E2-2E1. We calculate this efficiently by collecting the coefficients as we go:

  $E2-2E1: \quad x(4-2(2))+y(9-2(5)) = 5-2(3)$ or, after simplification $-y = -1$.

[4] You should note that these familiar operations are special cases of our permissible operation. Thus, to add a quantity "p" to both sides can be described as adding the true equation $p = p$ to the given equation. Similarly, multiplying by a non zero number "a" can be explained as writing the same equation twice and adding $(a-1)$-times the second to the first. Why do we want a to be non zero?. If we take $a = 0$, then we get a true equation, but its solutions have nothing to do with the given equation!

So $y = 1$. The value of x is then easily determined by plugging this into either equation:

$$2x + 5(1) = 3 \text{ so } x = (1/2)(3 - 5) = -1.$$

- Example 2. There is an **even more efficient calculation** for two equations in two variables using what is known as the Cramer's Rule.

  First we need a definition of a 2×2 determinant.

  Definition: Determinant of a 2 by 2 matrix An array of four numbers: $M = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$ arranged in two rows and two columns is said to be a **2 by 2 matrix.**

  Its determinant is denoted as $\det(M) = \begin{vmatrix} x & y \\ z & w \end{vmatrix}$ and is **defined to be** equal to the number $xw - yz$.

  A way to recall this value is to remember it as "the product of the entries on the main diagonal minus the product of the entries on the opposite diagonal".

  Now we are ready to state the Cramer's Rule to solve two linear equations in two variables.

  Cramer's Rule The solution to

  $$ax + by = c \text{ and } px + qy = r \text{ is given by}$$

  $$x = \frac{\begin{vmatrix} c & b \\ r & q \end{vmatrix}}{\begin{vmatrix} a & b \\ p & q \end{vmatrix}}, \; y = \frac{\begin{vmatrix} a & c \\ p & r \end{vmatrix}}{\begin{vmatrix} a & b \\ p & q \end{vmatrix}}.$$

  This looks very complicated, but there is an easy trick to remember it!

  The denominator is the determinant formed by the coefficients of the two variables in the two equations. For convenience, let us call it

  $$\Delta = \begin{vmatrix} a & b \\ p & q \end{vmatrix}.$$

  To get the value of x we start with Δ and replace the coefficients of x by the right hand sides. We get the determinant which we call

  $$\Delta_x = \begin{vmatrix} c & b \\ r & q \end{vmatrix}.$$

Then the value of x is $\frac{\Delta_x}{\Delta}$.

To get the y-value, we do the same with the y coefficients. Namely, we get

$$\Delta_y = \begin{vmatrix} a & c \\ p & r \end{vmatrix}$$

after replacing the coefficients of y by the right hand side.

Then the value of y is $\frac{\Delta_y}{\Delta}$.

To illustrate, let us solve the above equations again:

$$\begin{aligned} 2x + 5y &= 3 \\ 4x + 9y &= 5 \end{aligned}$$

Here the denominator is

$$\Delta = \begin{vmatrix} 2 & 5 \\ 4 & 9 \end{vmatrix} = (2)(9) - (5)(4) = -2.$$

The numerator for x is

$$\Delta_x = \begin{vmatrix} 3 & 5 \\ 5 & 9 \end{vmatrix} = (3)(9) - (5)(5) = 2.$$

The numerator for y is

$$\Delta_y = \begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = (2)(5) - (3)(4) = -2.$$

So

$$x = \frac{\Delta_x}{\Delta} = \frac{2}{-2} = -1 \text{ and } y = \frac{\Delta_y}{\Delta} = \frac{-2}{-2} = 1.$$

Exceptions to Cramer's Rule. Our answers $x = \frac{\Delta_x}{\Delta}$ and $y = \frac{\Delta_y}{\Delta}$ are clearly the final answers as long as $\Delta \neq 0$.

Thus we have a special case if and only if $\Delta = aq - bp = 0$. [5]

Further, if one of the numerators is non zero, then there is no solution, while if both are zero, then we have infinitely many solutions.

The reader can easily verify these facts.

We illustrate them by some examples.

[5]Thus Δ lets us "determine" the nature of the solution. This is the reason for the term "determinant".

- Example 3.

  Solve the system of equations:

  $$2x + y = 5 \ , \ 4x + 2y = k$$

  for various values of the parameter k.

  If we apply the Cramer's Rule, then we get

  $$x = \frac{\Delta_x}{\Delta} = \frac{\begin{vmatrix} 5 & 1 \\ k & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix}} = \frac{10-k}{0}.$$

  Similarly, we get

  $$y = \frac{\Delta_y}{\Delta} = \frac{\begin{vmatrix} 2 & 5 \\ 4 & k \end{vmatrix}}{\begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix}} = \frac{2k-20}{0}.$$

  Technically, what we wrote above is illegal! We are not supposed to write a zero in the denominator. How do we avoid this problem?

  Write the original Cramer's Rule as:

  $$(\Delta)x = \Delta_x \text{ and } (\Delta)y = \Delta_y.$$

  Then our current equations become:

  $$(0)x = 10 - k \text{ and } (0)y = 2k - 20.$$

  We see that when $k \neq 10$ then there are no values for x and y, while if $k = 10$, then we get the true but useless equations $0 = 0$ twice.

  So we need to analyze the situation further.

  A little thought shows that for $k = 10$ the second equation $4x + 2y = 20$ is simply twice the first equation $2x + y = 5$ and so it can be forgotten as long as we use the first equation!

  Thus the solution set consists of all values of x, y which satisfy $2x + y = 5$. It is easy to describe these as follows:

  Take any value, say t for y and then take $x = \frac{5-t}{2}$. Some of the concrete answers are $x = \frac{5}{2}, y = 0$ using $t = 0$, or $x = 2, y = 1$ using $t = 1$, or $x = 3, y = -1$ using $t = -1$ and so on. Indeed it is customary to declare:

$$x = \frac{5-t}{2}\ ,\ y = t \text{ where } t \text{ is arbitrary.}$$

Thus, we have infinitely many solutions.

We wish to emphasize that we have indeed described **"all possible solutions"** to our system of equations. This is said to be a **parametric solution** to our system and is needed when the system has infinitely many solutions.

We shall study such answers when we study the parametric forms of equations of lines later.

Many linear equations in many variables.

There is a similar theory of solving many equations in many variables using higher order determinants. The interested reader may look up books on determinants or linear algebra.

- Example 4. Here, we shall only illustrate how the Cramer's Rule mechanism works beautifully for more variables, but we shall refrain from including the details of calculations. The reader is encouraged to search for the details from higher books.

  We shall include a quick outline of the definition and calculation of higher order determinants in the appendix.

  Solve the three equations in three variables:

$$x + y + z = 6\ ,\ x - 2y + z = 0 \text{ and } 2x - y - z = -3.$$

  As before, we make a determinant of the coefficients

$$\Delta = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 2 & -1 & -1 \end{vmatrix}.$$

  As before, we get Δ_x by replacing the x coefficients by the right hand sides:

$$\Delta_x = \begin{vmatrix} 6 & 1 & 1 \\ 0 & -2 & 1 \\ -3 & -1 & -1 \end{vmatrix}.$$

  Once you get the right definition for a 3 by 3 determinant, it is easy to calculate $\Delta = 9 = \Delta_x$, so $x = \frac{9}{9} = 1$.

  Also, we get

$$\Delta_y = \begin{vmatrix} 1 & 6 & 1 \\ 1 & 0 & 1 \\ 2 & -3 & -1 \end{vmatrix} = 18.$$

Finally, we have

$$\Delta_z = \begin{vmatrix} 1 & 1 & 6 \\ 1 & -2 & 0 \\ 2 & -1 & -3 \end{vmatrix} = 27.$$

This gives us $y = \frac{18}{9} = 2$ and $z = \frac{27}{9} = 3$. The reader can at least verify the answers directly by known methods.

Chapter 3

The division algorithm and applications

3.1 Division algorithm in integers.

Given two integers u, v where v is positive, we already know how to divide u by v and get a quotient (a division) and a remainder.
This was taught as the long division of integers in schools.
For instance, when we divide 23 by 5 we get the quotient 4 and remainder 3. In equation form, we write this as

$$23 = (4)5 + 3 \text{ or } 23 - (4)5 = 3.$$

For the readers who prefer to **use their calculators,** here is a simple recipe for finding the quotient and the remainder.

- Have the calculator evaluate u/v. the quotient q is simply the integer part of the answer. Thus $23/5 = 4.6$, so $q = 4$.
- Then calculate $u - qv$. This is the remainder r. Thus, we get $23 - (4)5 = 3$, the remainder.

While useful, the above recipe can become unwieldy for large or complicated integers, so we urge the readers to not get dependent on it.
Here is an example:
Let $u = 1 + 3 + 3^2 + 3^3 + \cdots + 3^{1000}$ and $v = 3$. Divide u by v and find the division q and the remainder r.
The person with a calculator might be working a long time, just to determine what u is and may get overflow error messages from the calculator.
A person who looks at the form can easily see that we want

$$u = 1 + 3(1 + 3 + \cdots + 3^{999}) = r + 3q.$$

Note that $0 < 1 < 3$, so we can happily declare: $r = 1$ and $q = (1 + 3 + \cdots + 3^{999})$. First, we make a formal

Definition: Division Algorithm in Integers. Given two integers u, v with $v \neq 0$, there are unique integers q, r satisfying the following conditions:

1. $u = qv + r$ and
2. $0 \leq r < |v|$ where, as usual, $|v|$ denotes the absolute value of v.

The integer q is called **the quotient** and r is called **the remainder.** These are often described more casually as follows.
We may say that **u is equal to r modulo v** or that the remainder of u **modulo** v is r.
Here are some examples to illustrate how the definition works for negative integers.

- Let $u = -23, v = 5$. Then we get:

  $$-23 = (-5)5 + 2$$

  so now the quotient is -5 and remainder is 2!

  The calculator user has to note that now $u/v = -4.6 = -5 + 0.4$. So the division is now taken as -5, since we want our remainder to be bigger than or equal to zero.

  Then $u - qv = -23 - (-5)5 = -23 + 25 = 2$ so $r = 2$.

- Now take $u = 23, v = -5$ and note that $|v| = |-5| = 5$. Thus the remainder has to be one of $0, 1, 2, 3, 4$.

  We get:

  $$23 = (-4)(-5) + 3 \text{ so } q = -4 \text{ and } r = 3.$$

  The calculator work will give:

  The ratio
  $$u/v = \frac{23}{-5} = -4.6 = -4 - 0.6.$$

  If we take $q = -4$, then we get $u - qv = 23 - (-4)(-5) = 23 - 20 = 3$ so $r = 3$.

  Note that if we try taking $q = -5$, we would get a negative remainder, hence we need $q = -4$.

- Let $u = -23, v = -5$. Now again $|v| = 5$. We get:

  $$-23 = (5)(-5) + 2 \text{ so } q = -5 \text{ and } r = 2.$$

  The calculator work will give:

  The ratio [1]
  $$u/v = \frac{-23}{-5} = 4.6 = 5 - 0.4.$$

  If we take $q = 5$, then we get $u - qv = -23 - (5)(-5) = -23 + 25 = 2$ so $r = 2$.

Definition: Divisibility of integers. We shall say that **v divides u** if $u = kv$ for some integer k. We write this in **notation as v|u.** Sometimes, this is also worded as **"u is divisible by v"**.
Thus, the statement "u is equal to r modulo v " can be alternatively phrased as "$u - r$ is divisible by v."
Thus, if $v \neq 0$ then we see that $v|u$ exactly means that, when u is divided by v, the remainder r becomes 0.
It is clear that the only integer divisible by 0 is 0.
Let us now apply the division algorithm to find the GCD of two integers. For completeness, we record the definition:

Definition: GCD of integers. Given two integers u, v we say that an integer d is their **G(reatest) C(ommon) D(ivisor) or GCD** if the following conditions hold:

1. d divides both u, v.
2. If x divides both u and v then x divides d.
3. Either $d = 0$ or $d > 0$.

We define:

Definition: LCM of integers. Given two integers u, v we say that an integer L is their **L(east) C(ommon) M(ultiple) or LCM** if the following conditions hold:

1. Each of u, v divides L.
2. If each of u, v divide x, then L divides x.
3. Either $L = 0$ or $L > 0$.

[1] You should verify why $q = 4$ does not work.

Calculation of LCM:
Actually, it is easy to find LCM of two non zero integers u, v by a simple calculation. The simple formula is:

$$\mathrm{LCM}(u, v) = c\frac{uv}{\mathrm{GCD}(u, v)}.$$

Here c is equal to 1 or -1 and is chosen such that the LCM comes out positive.
Remarks. First note that the definitions of GCD and LCM can be easily extended to include several integers rather than just two u, v. The reader is encouraged to do this. Next note that the GCD becomes zero exactly when $u = v = 0$.
Important Fact. If we had to check these definitions every time, then the GCD would be a complicated concept. However, we show how to effectively find the GCD using a table based on the division algorithm defined above.
Example 1: GCD calculation in Integers. Find the GCD of 7553 and 623.
Answer:
We write the two given integers on top, divide the bottom into the top and write the negative of the quotient on the left and finally write the remainder below the top two integers.

Step no.	Minus Quotients	Integers
0	Begin	7553
1	-12	623
2	End of step 1	77

You could, of course, **do this with a calculator:**

$$7553/623 = 12.12 \text{ so } q = 12 \text{ and } r = 7553 - (12)(623) = 77.$$

Don't forget to enter the negative of q, i.e. -12 in the column.
Then we treat the next two numbers the same way; i.e. we divide 77 into 623 and get the quotient 8 and remainder 7.
Finally, when we divide 7 into 77, we have a quotient 11 and remainder 0, ending the process.

Step no.	Minus Quotients	Integers
0	Begin	7553
1	-12	623
2	-8	77
3	-11	7
4	End of process.	0

As before, we argue that 7 must be their GCD. [2]

[2] Any integer dividing the top two integers in our "Integers" column must divide the next one below and continuing must divide the bottom non zero integer 7. Also, the bottom 7 clearly divides everything above and is positive by construction! So it is the answer.

3.2 Āryabhaṭa algorithm: Efficient Euclidean algorithm.

A problem often encountered is this:
Problem. Suppose that you are given two non zero integers u, v and let d be their GCD.
Write a given integer z as a combination $z = au - bv$ of the two integers.
Of course, if z can be written thus, then we must have z divisible by the GCD d.
Thus the solution of the problem has two steps, first to find the GCD d to decide if the given z is a combination and then to actually find the combination.
Āryabhaṭa algorithm gives us a way to solve such a problem completely and efficiently. [3]
There are two traditional ways of stating the problem which we first describe.
The Kuttaka problem asks us to write any multiple of d as a combination of u, v.
In symbols, we need to write an integer multiple sd as $sd = xu + yv$ for some integers x, y.
The problem known as **Chinese Remainder Problem** asks us to find an integer n which has assigned remainders p, q when divided by u, v respectively. This gives a pair of equations:

$$n = au + p \text{ and } n = bv + q \text{ for some integers } a, b.$$

The Chinese Remainder Problem can always be solved by reducing to a Kuttaka problem as follows:

- Subtract the second from the first, then we get:

  $$0 = au + p - bv - q \text{ or } q - p = au - bv.$$

- Solve the Kuttaka problem of writing $q - p$ as a combination of u, v.

  Of course, it **must satisfy the GCD condition** that $q - p$ must be divisible by $d = \mathrm{GCD}(u, v)$.

[3]Indeed, this algorithm is usually known as the Euclidean algorithm. Our version is a more convenient arrangement of the algorithm given by an Indian Astronomer/Mathematician in the fifth century A.D., some 800 years after Euclid. It was, however, traditionally given in a much more complicated form by commentators. The original text is only one and a half verses long!

We also note that the usual Euclidean Algorithm has two components: one is how to find the GCD of two given numbers and the second is how to express the GCD as a combination of the two numbers. Euclid did not state the second part at all. Indeed his geometric terminology did not even provide for a natural statement of this property since the combination always includes a negative number and since for Euclid all numbers were lengths of lines, a negative number was not a valid object.

Thus, the Āryabhaṭa algorithm represents the first complete formulation of what is now called the Eucliden algorithm!

If the condition fails, then the original Chinese Remainder Problem is also unsolvable.

- Once we have $q - p = au - bv$ our n is equal to $au + p$ which is also equal to $bv + q$.

We begin by illustrating the technique with our solved example with $u = 7553, v = 623, d = 7$ above. We shall show how to write 7 as a combination of $u = 7553$ and $v = 628$.
We shall start with a similar table, but drop the step numbers and put a pair of blank columns for answers next to the numbers. We fill in the numbers as shown in the answers' columns.

$$\left[\begin{array}{r|r|r|r} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Integers} \\ \text{Begin} & 1 & 0 & 7553 \\ -12 & 0 & 1 & 623 \end{array}\right]$$

Now, here is how we fill in the next entries in the Answers' and Integers columns. The entry in a slot is obtained by multiplying the entry above it by the "minus quotient" on its left and then adding in the entry further above it.
Thus the entry in the Answer 1 column shall be $(-12)(0) + 1 = 1$. Similarly, the entry in the Answer 2 column shall be $(-12)(1) + 0 = -12$. The entry in the Integers column shall be $(-12)(623) + 7553 = 77$ the remainder that we had already figured out.
Thus we get:

$$\left[\begin{array}{r|r|r|r} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Integers} \\ \text{Begin} & 1 & 0 & 7553 \\ -12 & 0 & 1 & 623 \\ & 1 & -12 & 77 \end{array}\right]$$

Now the entry in the "Minus Quotients" column is the negative of the quotient when we divide the last integer 77 into the one above it, namely 623. We already know the quotient to be 8 and so we enter -8.

$$\left[\begin{array}{r|r|r|r} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Integers} \\ \text{Begin} & 1 & 0 & 7553 \\ -12 & 0 & 1 & 623 \\ -8 & 1 & -12 & 77 \end{array}\right]$$

We now continue this process to make the next row whose entries shall be as follows.
Answer 1: $(-8)(1) + 0 = -8$.
Answer 2: $(-8)(-12) + 1 = 97$.
Integer: $(-8)(77) + 623 = 7$.

Finally, the new "Minus Quotient" is the negative of the quotient when we divide 7 into 77, i.e. -11.
Thus the new table is:

$$\begin{bmatrix} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Integers} \\ \text{Begin} & 1 & 0 & 7553 \\ -12 & 0 & 1 & 623 \\ -8 & 1 & -12 & 77 \\ -11 & -8 & 97 & 7 \end{bmatrix}$$

Now we carry out one more step of similar calculation. We leave it for the reader to verify the calculations.

$$\begin{bmatrix} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Integers} \\ \text{Begin} & 1 & 0 & 7553 \\ -12 & 0 & 1 & 623 \\ -8 & 1 & -12 & 77 \\ -11 & -8 & 97 & 7 \\ \text{End} & 89 & -1079 & 0 \end{bmatrix}$$

We wrote "End" because we reached a zero in the last column and we don't apply division algorithm by 0.
We claim that we have all the answers we need. How?
We note the important property that in any of the rows, we have:

$$7553(\text{ Answer 1}) + (623)(\text{ Answer 2 }) = \text{ The entry in the Integers column.}$$

It is instructive to verify this in each row:

$$\begin{array}{rcrr} 7553 \cdot (1) & + & 623 \cdot (0) & = 7553 \\ 7553 \cdot (0) & + & 623 \cdot (1) & = 623 \\ 7553 \cdot (1) & + & 623 \cdot (-12) & = 77 \\ 7553 \cdot (-8) & + & 623 \cdot (97) & = 7 \\ 7553 \cdot (89) & + & 623 \cdot (-1079) & = 0 \end{array}$$

The second row from the bottom in Answers columns gives us the desired expression of the GCD:

$$7 = (97)623 - (8)7553.$$

The last row is used to find all possible expressions of the GCD using the one obtained above.
Thus from the last two equations, we see that for any integer n, we have:

$$7553 \cdot (-8 + 89n) + 623 \cdot (97 - 1079n) = 7$$

Taking $n = 0$ we get the main answer:

$$7 = 623(97) - 7553(8).$$

By taking $n = 1$ we also see that:

$$7 = 7553(81) - 623(982).$$

We need both these forms of expressions depending on the intended use.
Note on the algorithm. Even though we have called the above algorithm as "Āryabhaṭa algorithm", it is a changed version of the original.
We explain the changes made for purposes of historical accuracy.
The original algorithm was a two step process. It proceeded as in the original GCD calculations where we had only two columns: Minus quotients and Integers.
Actually, it had the quotients rather than their negatives. (The negative integers, though known, were not popular.)
The necessary column of answers was then created by starting with $1, 0$ at the bottom and lifting the answers up by using the quotients. To read off the final answer, a careful adjustment of sign was needed.
The algorithm also provided for a shortcut. The aim of the algorithm was to write a given number as the combination of the chosen numbers. So, it was not necessary to "find" the GCD. If at some point during the calculations, we can write the given number as a combination of the bottom two entries of the "Integers" column, then we could record the multipliers next to them and then lift them up using the quotients using the algorithm.
If the given number is not a multiple of the GCD and then the problem is unsolvable. This becomes evident when we reach the GCD and the lifting work is no longer needed.
We have chosen to add the two columns of "Answers" to make the process fully automatic and avoid any "lifting work" at the end. This method has more in common with the more modern computer implementations of the Euclidean algorithm (especially the one known as Motzkin algorithm).
In this formulation, the record of the quotients is mostly for double checking the work, since they are not needed once we reach the end of the table.
Example 2: Kuttaka or Chinese Remainder Problem.
Find a smallest positive integer n such that when n is divided by 7553 it gives a remainder of 11 and when divided by 623 it gives a remainder of 4.
Answer: First, let us note that we want

$$n = a(623) + 4 \text{ and } n = b(7553) + 11$$

for some integers a, b.
Subtracting the second equation from the first, we get:

$$a(623) - b(7553) + 4 - 11 = 0 \text{ or } a(623) - b(7553) = 7.$$

We already know an answer to this, namely $a = 97$ and $b = 8$.
This gives us:

$$n = 97(623) + 4 = 8(7553) + 11 = 60435.$$

Is this indeed the smallest positive integer? **Indeed it is!** How do we prove this? Suppose, we have another possible answer n_1 such that:

$$n_1 = a_1(623) + 11 \text{ and } n_1 = b_1(7553) + 4.$$

Then clearly

$$n_1 - n = (a_1 - a)623 = (b_1 - b)7553.$$

Hence $n_1 - n$ is divisible by both $623, 7553$ and hence by their LCM. From the known formula, the LCM is:

$$\mathrm{LCM}(7553, 623) = \frac{(7553)(623)}{7} = 672217.$$

Thus all other answers n_1 differ from n by a multiple of 672217, thus, our n must be the smallest positive answer!

Example 3: More Kuttaka problems.

Suppose we try to solve a problem very similar to the above, except we change the two remainders.

We illustrate possible difficulties and their resolution.

Problem 1:

Find a smallest positive integer n such that when n is divided by 7553 it gives a remainder of 11 and when divided by 623 it gives a remainder of 4.

Problem 2:

Find a smallest positive integer n such that when n is divided by 7553 it gives a remainder of 4 and when divided by 623 it gives a remainder of 10.

Answer: Problem 1. We can work just as before and get:

$$n = a(623) + 11 \text{ and } n = b(7553) + 4 \text{ so } a(623) - b(7553) = -7.$$

Multiplying by -1 we get:

$$7 = b(7553) - a(623).$$

If we recall our alternate answer $7 = 81(7553) - 982(623)$, we see that we could simply take $a = 982$ and $b = 81$.

Thus, we have, $n = 982(623) + 11 = 611797$. We could get the same answer as $n = 81(7553) + 4$. [4]

[4] It is well worth checking all such answers again: Using the calculator verify: $611797/623 = 982.02$ and $611797 - (982)(623) = 11$. Similarly, $611797/7553 = 81.0005$ and $611797 - (7553)(81) = 4$.

Answer: Problem 2. Proceeding as before, we get equations:

$$n = a(623) + 10 \text{ and } n = b(7553) + 4.$$

By subtraction we get

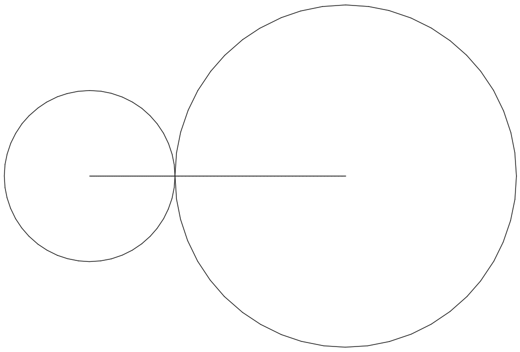
$$a(623) - b(7553) = -6.$$

Since the GCD 7 does not divide the right hand side -6, we conclude that there is no such integer n.

Example 4: Disguised problems.

Sometimes, our problem comes well disguised as an astronomical or a mechanical problem.

We give one instance of this as a sample.



A wheel of radius 546 mm drives a wheel of radius 798 mm. When the wheels first start turning, the marked spokes are lined up as shown.

How many revolutions of the smaller wheel will be required before the two spokes line up again?

When the two spokes are lined up again, how many revolutions should the larger wheel have turned through?

Answer. An unsaid assumption here is that the wheels are turning against each other without slipping. Suppose that the smaller wheel makes n revolutions and the bigger wheel makes m revolutions in the same time.

As our standard strategy of problem solving, we assume that the radii of the two wheels are r mm and s mm respectively so that we have:

$$r = 546 \ , \ s = 798.$$

Keeping our attention to the touching point, the original point on the small wheel must have gone through a distance of $2\pi rn$ mm. The corresponding touching point on the large wheel must similarly have gone through a distance of $2\pi sm$ mm.

Because of non slipping condition, we must have:

$$2\pi rn = 2\pi sm \text{ or } rn = sm.$$

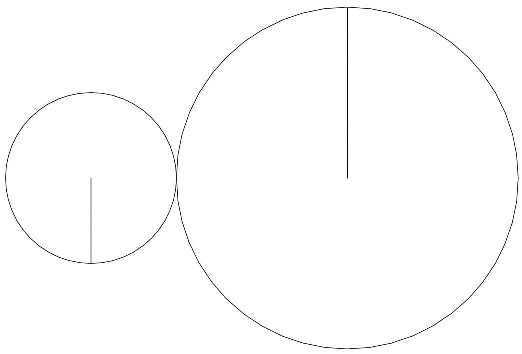
Thus, this common number is a multiple of the LCM(r, s).
By using our technique, we can easily deduce that $\text{GCD}(546, 798) = 42$ and hence

$$\text{LCM}(546, 798) = \frac{546 \cdot 798}{42} = 10374.$$

Thus, for the first time when the marks on the two wheels are lined up, we should have

$$r = \frac{10374}{546} = 19 \text{ and } s = \frac{10374}{798} = 13.$$

Note. The above problem only depended on the knowledge of the LCM. It becomes a true Kuttaka problem if we consider the possibility that the marks on the two wheels are lined in a certain configuration. Thus, for example, if we look for the first time when the marks are pointing as shown, we get the following.



We assume that the left wheel is turning counterclockwise and the right one must naturally turn clockwise.
Suppose we let p, q be the respective number of revolutions of the small and large wheel at the desired location. We allow these numbers to be fractions.
Then we have: [5]

(1) $$2\pi r p = 2\pi s q \text{ which reduces to } 13p = 19q.$$

Note that the number p represents a certain number of complete revolutions followed by a three quarter rotation, so

(2). $$p = a + \frac{3}{4} \text{ for some integer } a.$$

[5]Here are the details of the reduction. Plug in the values of r, s and write $2\pi(546)p = 2\pi(798)q$. Cancel the common factor $2\pi(42)$ to get

$$\frac{546p}{42} = \frac{798q}{42} \text{ or } 13p = 19q.$$

Similarly, we deduce that

(3). $$q = b + \frac{1}{4} \text{ for some integer } a.$$

Thus we see that $13(a + \frac{3}{4}) = 19(b + \frac{1}{4})$ and multiplying by 4 on both sides we see that

(4) $$13(4a + 3) = 19(4b + 1).$$

Now that all the factors are integers and again since 13 and 19 have GCD equal to 1, we must have 13 divides $4b + 1$, so $4b + 1 = 13x$ for some integer x.
This gives

$$13(4a + 3) = 19(13x) \text{ or after canceling 13 we get } 4a + 3 = 19x.$$

Thus we have proved:

(5) $$19x = 4a + 3 \text{ and } 13x = 4b + 1.$$

Thus, the first equation asks us to write the number 3 as a combination of the numbers 19 and 4 whose GCD is 1. This is a Kuttaka problem.
So is $13x = 4b + 1$.
However, the solution is easy to see here!
It appears that $x = 1$ gives the smallest solution, i.e. the number of revolutions for the small wheel is $p = \frac{19}{4}$ and for the large wheel it is $q = \frac{13}{4}$.

3.3 Division algorithm in polynomials.

Earlier, we have discussed addition, subtraction and multiplication of polynomials.
As indicated, we have a problem if we wish to divide one polynomial by another, since the answer may not be a polynomial again.
Let us take an example of $u(x) = x^3 + x, v(x) = x^2 + x + 1$. It seems obvious that the ratio $\frac{u(x)}{v(x)}$ is not a polynomial.
But we want to really prove this! Here is how we argue. Suppose, if possible, it is equal to some polynomial $w(x)$ and write the equation:

$$\frac{u(x)}{v(x)} = w(x) \text{ or } u(x) = w(x)v(x).$$

We write this equation out and compare both sides.

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

By comparing the degrees of both sides, it follows that $w(x)$ must have degree 1 in x. [6]
Let us write $w(x) = ax + b$. Then we get

$$x^3 + x = (ax + b)(x^2 + x + 1) = ax^3 + (a + b)x^2 + (a + b)x + b.$$

Comparing coefficients of x on both sides, we see that we have:

$$a = 1\ ,\ a + b = 0\ ,\ (a + b) = 1\ ,\ b = 0.$$

Obviously these equations have no chance of a common solution, they are inconsistent!
This proves that $w(x)$ is not a polynomial!
But perhaps we are being too greedy. What if we only try to solve the first two equations?
Thus we solve $a = 1$ and $a + b = 0$ to get $a = 1, b = -1$. Then we see that using $a = 1, b = -1$:

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

Let us name $x - 1$ as $q(x)$ and $x + 1$ as $r(x)$. Then we get:

$$u(x) = v(x)q(x) + r(x).$$

where $r(x)$ has degree 1 which is smaller than the degree of $v(x)$ which is 2.
Thus, when we try to divide $u(x)$ by $v(x)$, then we get the best possible division as $q(x) = x - 1$ and a remainder $x + 1$.
We now formalize this idea.

Definition: Division Algorithm. Suppose that $u(x) = ax^n + \cdots$ and $v(x) = bx^m + \cdots$ are polynomials of degrees n, m respectively and that $v(x)$ is not the zero polynomial. Then there are **unique polynomials** $q(x)$ and $r(x)$ which satisfy the following conditions:

1. $u(x) = q(x)v(x) + r(x)$.
2. Moreover, either $r(x) = 0$ or $\deg_x(r(x)) < \deg_x(v(x))$.

When the above conditions are satisfied, we declare that $q(x)$ is the **quotient** and $r(x)$ is the **remainder** when we **divide u(x) by v(x). Some people use the word division, in place of the word quotient.**
We may rephrase the same by saying "$u(x)$ is equal to $r(x)$ modulo $v(x)$" or "$u(x)$ has remainder $r(x)$ modulo $v(x)$".

[6]Note that $\deg_x(u(x)) = \deg_x(w(x)) + \deg_x(v(x))$ or $3 = \deg_x(w(x)) + 2$. Hence the conclusion.

The quotient $q(x)$ and the remainder $r(x)$ always exist and are uniquely determined by $u(x)$ and $v(x)$. [7]

In analogy with integers, we make:

Definition: Divisibility of polynomials. We say that a polynomial $v(x)$ divides a polynomial $u(x)$, if the remainder $r(x)$ obtained by dividing $u(x)$ by $v(x)$ becomes zero.

As before, we may also express the same idea by the symbolism "$v(x)|u(x)$ " or by saying $u(x)$ is 0 modulo $v(x)$. The zero polynomial can divide only a zero polynomial.

First, we explain how to find $q(x)$ and $r(x)$ systematically.

Let us redo the problem with $u(x) = x^3 + x$ and $v(x) = x^2 + x + 1$ again. Since u has degree 3 and v has degree 2, we know that q must have degree $3 - 2 = 1$.

For convenience, we often drop the x from our notation and simply write u, v, q, r in place of $u(x), v(x), q(x), r(x)$.

What we want is to arrange $u - qv$ to have a degree as small as possible; this means that either it is the zero polynomial or its degree is less than 2.

Start by guessing $q = ax$. Calculate

$$u - (ax)v = x^3 + x - (ax)(x^2 + x + 1) = (1 - a)x^3 + (-a)x^2 + (1 - a)x.$$

This means $a = 1$ and then we get

$$u - (x)v = (-1)x^2.$$

Our right hand side has degree 2 which is still not small enough! So we need to improve our q. Add a next term to the current $q = x$ and make it $q = x + b$.

Recalculate: [8]

$$\begin{aligned} u - (x + b)v &= u - xv - bv \\ &= (-1)x^2 - b(x^2 + x + 1) \\ &= (-1 - b)x^2 + (-b)x + (-b) \end{aligned}$$

Thus, if we make $-1 - b = 0$ by taking $b = -1$, we get $q = x - 1$ and

$$u - qv = (-(-1))x + (-(-1)) = x + 1 = r.$$

Let us **summarize this process**:

---

[7] Hint for the proof: Suppose that $u - q_1 v = r_1$ and $u - q_2 v = r_2$ where r_1, r_2 are either 0 or have degrees less than m.

Subtract the second equation from the first and consider the two sides $(q_2 - q_1)v$ and $(r_1 - r_2)$.

If both are 0, then you have the proof that $q_2 = q_1$ and $r_2 = r_1$. If one of them is zero, then the other must be zero too, giving the proof.

Finally, if both are non zero, then the left hand side has degree bigger than or equal to m while the right hand side has degree less than m; a clear contradiction!

[8] Note that we use the already done calculation of $u - xv$ and don't waste time redoing the steps. We only figure out the contribution of the new terms.

1. Suppose $u(x), v(x)$ have degrees n, m respectively where $n \geq m$. (Note that we are naturally assuming that these are non zero polynomials.)

   As above, for convenience we shall drop x from our notations for polynomials.

2. Start with a guess $q = ax^{(n-m)}$ and choose a such that $u - qv$ has degree less than n.

3. Note that the whole term $ax^{(n-m)}$ can simply be thought of as:

$$ax^{(n-m)} = \frac{\text{the leading term of } u}{\text{the leading term of } v}.$$

   This formula is useful, but often it is easier to find the term by inspection.

   In our example above, this first term of q came out to be $\frac{x^3}{x^2} = x$. Thus, the starting guess for q is $q = x$.

4. Using this current value of q, if $u - qv$ is zero or if its degree is less than m, then stop and set $u - qv$ as the final remainder r.

5. If not, add a next term to q to make the degree of $u - qv$ even smaller. This next term can be easily found as:

$$\frac{\text{leading term of the current } u - qv}{\text{leading term of } v}.$$

   In our example this was $\frac{-x^2}{x^2} = -1$. Thus, the new guess for q is $q = x - 1$.

6. Continue until $u - qv$ becomes 0 or its degree drops below m.

   In our example, $q = x - 1$ gives the remainder $x + 1$ whose degree is clearly less than 2 and we stop.

   What we are describing is the process of long division that you learned in high school. We are going to learn more efficient methods for doing it below.

   For comparison, we present the long division process for the same polynomials. Compare the steps for better understanding:

$$\begin{array}{rr|rrr}
 & & x & & \\ \cline{3-5}
x^2 + x + 1 &) & x^3 & & x \\
 & & -x^3 & -x^2 & -x \\ \cline{3-5}
 & & & -x^2 &
\end{array}$$

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

There are two **lucky situations** where we get our q, r without any further work. Let us record these for future use.

1. In case u is the zero polynomial, we take $q(x) = 0$ and $r(x) = 0$. Check that this has the necessary properties!
2. In case u has degree smaller than that of v (i.e. $n < m$), we take $q(x) = 0$ and get $r(x) = u(x)$. Check that this is a valid answer as well.

One important principle is to **"let the definition be with you!!"** If you somehow see an answer for q and r which satisfies the conditions, then don't waste time in the long division.
We will see many instances of this later.
We now begin with an extension of the division algorithm.

3.4 Repeated Division.

For further discussion, it would help to define some terms.

Definition: Dividend and Divisor. If we divide $u(x)$ by $v(x)$, then we shall call $u(x)$ the **dividend** and $v(x)$ the **divisor** . We already know the meaning of the terms quotient and the remainder.
We wish to organize the work of division algorithm in a certain way, so we can get more useful information from it.
We assign ourselves some simple tasks:
Task 1. Given
$$u(x) = x^3 + 2\,x - 1 \text{ and } v(x) = x^2 + x + 2,$$
calculate the quotient $q(x)$ and the remainder $r(x)$ when you divide $u(x)$ by $v(x)$.
Answer:
As before, we shall drop the x from the notation for convenience.
We need to make $u - qv$ to be zero or have degree less than 2.
We see $u - (x - 1)v = x + 1$, so $r = x + 1$ and $q = x - 1$.

We recommend the following arrangement of the above work, for future use.
We shall make a table:

$$\left[\begin{array}{l|l|l} \text{Step no.} & \text{Minus Quotients} & \text{polynomials} \\ 0 & \text{Begin} & u \\ 1 & -q & v \\ 2 & \text{End of step 1} & r \end{array}\right]$$

Thus after writing the entries dividend and divisor u, v in a column, we write the negative of the quotient in the "Minus Quotients" column. The entry below u, v is the remainder r obtained by adding $-q$ times v to u.
Thus, our example becomes:

$$\left[\begin{array}{l|l|l} \text{Step no.} & \text{Minus Quotients} & \text{polynomials} \\ 0 & \text{Begin} & x^3 + 2\,x - 1 \\ 1 & -x + 1 & x^2 + x + 2 \\ 2 & \text{End of step 1} & x + 1 \end{array}\right]$$

Note that the actual long division process is not recorded and is to be done on the side.
Task 2. Now repeat the steps of task 1 and continue the table by treating the last two polynomials as the dividend and the divisor.
Thus our new dividend shall be $x^2 + x + 2$ and the new divisor shall be $x + 1$.
Answer: We easily see that

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

i.e. the quotient is x and the remainder is 2.
Thus our table extends as:

$$\left[\begin{array}{l|l|l} \text{Step no.} & \text{Minus Quotients} & \text{polynomials} \\ 0 & \text{Begin} & x^3 + 2\,x - 1 \\ 1 & -x + 1 & x^2 + x + 2 \\ 2 & -x & x + 1 \\ 3 & \text{End of step 2} & 2 \end{array}\right]$$

Task 3. Continue the tasks as far as possible, i.e. until the last remainder is zero.
Answer: Now the last two polynomials give the dividend $x + 1$ and the divisor 2.
Thus

$$x + 1 - (\frac{x + 1}{2})(2) = 0.$$

Thus quotient is $\frac{x+1}{2}$ and the remainder is 0. The zero remainder says our tasks are over!
Our table extends as:

$$\left[\begin{array}{l|l|l} \text{Step no.} & \text{Minus Quotients} & \text{polynomials} \\ 0 & \text{Begin} & x^3+2\,x-1 \\ 1 & -x+1 & x^2+x+2 \\ 2 & -x & x+1 \\ 3 & -\frac{x+1}{2} & 2 \\ 4 & \text{End of steps!} & 0 \end{array}\right]$$

The step numbers are not essential and we shall drop them in future.
The steps carried out above are for a purpose. These steps are used to find the GCD or the greatest common divisor (factor) of the original polynomials $u(x), v(x)$. The very definition of the division algorithm says that

$$u(x) - q(x)v(x) = r(x).$$

Thus, any common factor of $u(x)$ and $v(x)$ must divide the remainder. Using this repeatedly, we see that such a factor must divide each of the polynomials in our column of polynomials in the calculation table.
Since one of these polynomials is 2, we see that $u(x)$ and $v(x)$ have no non constant common factor!
Indeed, it is not hard to show that the last polynomial in our column, just above the zero polynomial must divide every polynomial above it, and in turn any polynomial which divides the top two, must divide it!
Thus, in some sense, it is the largest degree polynomial dividing both the top polynomials - or it is their **G(reatest) C(ommon) D(ivisor)- or GCD for short.**

3.5 The GCD and LCM of two polynomials.

Indeed, this last polynomial (or any constant multiple of it) can be defined as the GCD.
To make the idea of GCD precise, most algebraists would choose a multiple of the last polynomial which makes the highest degree coefficient equal to **1**.
To make this idea clear, we start with:

Definition: Monic polynomial A nonzero polynomial in one variable is said to be monic if its leading coefficient is 1. Note that the only monic polynomial of degree 0 is 1.

Definition: GCD of two polynomials
The GCD of two polynomials $u(x), v(x)$ is defined to be a monic polynomial $d(x)$ which has the property that $d(x)$ divides $u(x)$ as well as $v(x)$ and is the largest degree monic polynomial with this property.

We shall use the notation:

$$d(x) = \text{GCD}(u(x), v(x)).$$

An alert reader will **note that the definition gets in trouble** if both polynomials are zero. Some people make a definition which produces the zero polynomial as the answer, but then we have to let go of the 'monicness' condition. In our treatment, this case is usually excluded!

Definition: LCM of two polynomials The LCM (least common multiple) of two polynomials $u(x), v(x)$ is defined as the smallest degree monic polynomial $L(x)$ such that each of $u(x)$ and $v(x)$ divides $L(x)$.
As in case of GCD, we shall use the notation:

$$L(x) = \text{LCM}(u(x), v(x)).$$

If either of $u(x)$ or $v(x)$ is zero, then the only choice for $L(x)$ is the zero polynomial and the degree condition as well as the monicness get in trouble!
As in the case of GCD, we can either agree that the zero polynomial is the answer or exclude such a case!
Calculation of LCM.
Actually, it is easy to find LCM of two non zero polynomials $u(x), v(x)$ by a simple calculation. The simple formula is:

$$\text{LCM}(u(x), v(x)) = c\frac{u(x)v(x)}{\text{GCD}(u(x), v(x))}.$$

Here c is a constant which is chosen to make the LCM monic.
There are occasional shortcuts to the calculation and these will be presented in special cases.
The most important property of GCD. It is possible to show that the GCD of two polynomials $u(x), v(x)$ is also the **smallest degree monic polynomial** which can be expressed as a combination $a(x)v(x) - b(x)u(x)$.
Such a presentation of the GCD of two integers was already explained as the Āryabhaṭa algorithm. For polynomials, a similar scheme works, however, the calculation can get messy due to the number of terms in a polynomial. We shall leave it for the reader to investigate further as independent work. We will, however, show the details of the work for the same two polynomials analyzed above. See the worked example below.
For our polynomials above, the GCD will be the monic multiple of 2 and so the GCD of the given $u(x), v(x)$ is said to be 1. [9]

[9] If you are confused about how 2 becomes 1, remember that the only monic polynomial of degree 0 is 1.

You may have seen a definition of the GCD of polynomials which tells you to factor both the polynomials and then take the largest common factor.
This is correct, but requires the ability to factor the given polynomials, which is a very difficult task in general.
Our GCD algorithm is very general and useful in finding common factors without finding the factorization of the original polynomials.
Example 5: Āryabhaṭa algorithm for polynomials.
Let us start with the polynomials $u(x) = x^3 + 2x - 1$ and $v(x) = x^2 + x + 2$ as before. The only twist is to add the two columns of "Answers" and fill them in as we go along. Since the steps for the division, including the remainders and the quotients are the same as above, we just display the steps without comment.

$$\left[\begin{array}{r|r|r|r} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Polynomials} \\ \text{Begin} & 1 & 0 & x^3+2x-1 \\ -x+1 & 0 & 1 & x^2+x+2 \\ \text{Step 1} & & & x+1 \end{array}\right]$$

The only work done is that of finding the minus quotient $-(x-1) = -x+1$ and the remainder $x+1$.

$$\left[\begin{array}{r|r|r|r} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Polynomials} \\ \text{Begin} & 1 & 0 & x^3+2x-1 \\ -x+1 & 0 & 1 & x^2+x+2 \\ -x & 1 & -x+1 & x+1 \\ \text{Step 2} & & & 2 \end{array}\right]$$

First the entries $1, -x+1$ are filled in by multiplying the entries $0, 1$ by $-x+1$ and adding to the corresponding entries $1, 0$ above.
The next minus quotient when we divide x^2+x+2 by $x+1$ is $-(x)$. The remainder is 2. All these are filled in.

$$\left[\begin{array}{r|r|r|r} \text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Polynomials} \\ \text{Begin} & 1 & 0 & x^3+2x-1 \\ -x+1 & 0 & 1 & x^2+x+2 \\ -x & 1 & -x+1 & x+1 \\ -\frac{x+1}{2} & -x & x^2-x+1 & 2 \\ \text{End of Steps.} & & & 0 \end{array}\right]$$

The entries $-x, x^2-x+1$ are filled in by multiplying the entries $1, -x+1$ by $-x$ and adding to the corresponding entries $0, 1$ above.
We divide 2 into $x+1$ and record the minus quotient $-\frac{x+1}{2}$ and the remainder 0. This signals the end.

$$\left[\begin{array}{r|r|r|r}
\text{Minus Quotients} & \text{Answer 1} & \text{Answer 2} & \text{Polynomials} \\
\text{Begin} & 1 & 0 & x^3+2\,x-1 \\
-x+1 & 0 & 1 & x^2+x+2 \\
-x & 1 & -x+1 & x+1 \\
-\frac{x+1}{2} & -x & x^2-x+1 & 2 \\
\text{End of steps!} & \frac{x^2+x+2}{2} & \frac{-x^3-2x+1}{2} & 0
\end{array}\right]$$

We have just filled in the last row entries by multiplying the entries $-x, x^2-x+1$ by $-\frac{x+1}{2}$ and adding to the corresponding entries $1, -x+1$ above.
Now it is time to read off the necessary solutions.
Recall that our u, v are the top two entries of the Polynomials' column.
In any row of the final table, we have the relation:

$$(\text{Answer 1})\cdot u + (\text{Answer 2 })\cdot v = \text{ The Polynomial at the right end.}$$

In particular, we have

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

Dividing by 2, we see how to write the GCD$(u, v) = 1$ as a combination:

$$(-\frac{x}{2})\cdot u + (\frac{x^2-x+1}{2})\cdot v = 1.$$

The last entry above the zero in the Polynomials' column is always the GCD, except it may need to be divided by some constant to make it monic.
We may call it the "temporary GCD".
The entries of the last row in the answers' columns are easily seen to be
Answer 1 $= \pm\frac{v}{\text{temporary GCD}}$ and Answer 2 $= \pm\frac{u}{\text{temporary GCD}}$.
Thus, the LCM can be computed by multiplying Answer 1 in the last row by u and making it monic by a constant multiple.
Similarly, it can also be obtained by multiplying Answer 2 in the last row by v and making it monic by a constant multiple. [10]

Application to partial fractions.

Given a rational function $\frac{p(x)}{q(x)}$, it is often important to write it as a sum of rational functions with simpler denominators. This simply means that the allowed denominator is a power of an irreducible polynomial and naturally, these denominators must be factors of the original denominator $q(x)$.
These are important in Calculus and Differential equations.
In actual applications, the polynomials have real coefficients and the simpler denominators end up being powers of linear or quadratic polynomials.[11]

[10] It is an important consequence of the nature of our steps that the 2×2 determinant formed by the last two rows in the Answers' columns is always ± 1. It is interesting to figure out why! At any rate, it is a good check on the correctness of the work.

[11] This is not obvious, but is a consequence of the Fundamental Theorem of Algebra.

We now indicate a crucial first step to this process of simplification.
Suppose that $q(x) = u(x)v(x)$ such that $\mathrm{GCD}(u(x), v(x)) = 1$. Then we have a simplification:

$$\frac{p(x)}{q(x)} = \frac{g(x)}{u(x)} + \frac{h(x)}{v(x)}.$$

Here is how this works. Using our algorithm, we can write

$$1 = a(x)u(x) + b(x)v(x) \text{ or } p(x) = p(x)a(x)u(x) + p(x)b(x)v(x).$$

Dividing by $q(x) = u(x)(v(x)$ we get:

$$\frac{p(x)}{q(x)} = \frac{p(x)a(x)}{v(x)} + \frac{p(x)b(x)}{u(x)}.$$

Some observations about **further simplification:**

- It is possible that the rational functions on the right hand side are "vulgar", meaning the numerator has a degree bigger than or equal to the denominator. In this case, it is customary to divide the numerator by the denominator and write the rational function as a polynomial plus a proper rational function (meaning one whose numerator has a smaller degree than the denominator). [12]

- In case, $\deg_x(p(x)) < \deg_x(q(x))$, i.e. the original rational function is proper, it can be shown that after rewriting the simplified expressions as polynomials plus proper rational functions, the two polynomial parts will simply cancel out.

  In other words, a proper rational function gets written as a sum of proper rational functions with simpler denominators.

- In the special case when $p(x) = 1$ it can be shown that

  $$\frac{1}{q(x)} = \frac{1}{u(x)v(x)} = \frac{a(x)}{v(x)} + \frac{b(x)}{u(x)}$$

  is already a sum of proper rational functions!

  This is a consequence of the algorithm which can be proved by a careful analysis of degrees of the polynomials at each stage.

[12] This is another convenient way to think about the division algorithm. Given any rational function, $\frac{u(x)}{v(x)}$ when we do the division algorithm to write $u(x) = q(x)v(x) + r(x)$ we get:

$$\frac{u(x)}{v(x)} = q(x) + \frac{r(x)}{v(x)}.$$

Thus, $q(x)$ is the polynomial part and $\frac{r(x)}{v(x)}$ is the proper fraction. If the proper fraction turns out to be zero, then $v(x)$ divides $u(x)$.

Example 6: Efficient division by a linear polynomial. Let a be a constant. Find the division and remainder when $(x-a)$ divides various powers of x, like $x, x^2, x^3, x^4, \cdots$.
Using these results deduce a formula to compute the remainder when a polynomial $f(x)$ is divided by $(x-a)$.
Answer: The main point to note is this:
Suppose that we wish to divide a polynomial $u(x)$ by a non zero polynomial $v(x)$ to find the quotient and remainder.
If we can **somehow guess a polynomial** $r(x)$ such that

- $$u(x) - r(x) = q(x)v(x),$$
  i.e. $v(x)$ divides $u(x) - r(x)$
- and either $r(x) = 0$ or the degree of $r(x)$ is less than that of $v(x)$,

then we have found our remainder $r(x)$. If needed, we can find the quotient by dividing $u(x) - r(x)$ by $v(x)$ if needed!
No long division is actually needed.
For $u(x) = x$, we guess $r(x) = a$ and note that

$$u(x) - r(x) = x - a = (1)(x - a)$$

so obviously $q(x) = 1$ and $r(x) = a$.
Note that $r(x) = a$ is either 0 or has degree zero in x and hence is a valid remainder.
For $u(x) = x^2$ we guess that $r(x) = a^2$ and verify it from the identity:

$$x^2 - a^2 = (x + a)(x - a) \text{ which means } x^2 = (x + a)(x - a) + a^2.$$

Thus, dividing x^2 by $x - a$ gives the quotient $(x + a)$ and the remainder a^2.
To divide x^3 by $(x - a)$, we start with the above relation and multiply it by x, so:

$$x^3 - a^2x = x(x + a)(x - a) \text{ or } x^3 = x(x + a)(x - a) + a^2x.$$

Rearranging, we get:

$$x^3 - a^3 = x(x+a)(x-a) + a^2x - a^3 = (x(x+a) + a^2)(x-a) = (x^2 + ax + a^2)(x-a).$$

So x^3 divided by $x - a$ gives $r(x) = a^3$ and $q(x) = x^2 + ax + a^2$.
The reader is invited to repeat this idea and deduce

$$x^4 - a^4 = (x)(x^2 + ax + a^2)(x - a) + a^3(x - a)$$

or

$$x^4 - a^4 = (x^3 + ax^2 + a^2x + a^3)(x - a) \text{ giving } q(x) = x^3 + ax^2 + a^2x + a^3 \text{ and } r(x) = a^4.$$

Inspired by this let us guess and prove: For $n = 1, 2, \cdots$:

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

Proof. Here is a simple arrangement of multiplication on the right hand side which makes a convincing proof.

$$\begin{array}{llllll}
 & x^{n-1} & +ax^{n-2} & \cdots & +a^{n-2}x & a^{n-1} \\
 & & & & \times & (x-a) \\
\hline
x^n & +ax^{n-1} & \cdots & +a^{n-2}x^2 & a^{n-1}x & \\
 & -ax^{n-1} & \cdots & -a^{n-2}x^2 & -a^{n-1}x & -a^n \\
\hline
x^n & & & & & -a^n
\end{array}$$

We conclude that the remainder of x^n is a^n with quotient $x^{n-1} + ax^{n-2} + \cdots + a^{n-1}$. It is easy to see that the remainder of cx^n will be ca^n for a constant c and generally for any polynomial $f(x)$ the remainder when divided by $(x - a)$ shall be $f(a)$.
For proof, consider this:
The polynomial $f(x)$ is a sum of monomials of the form cx^n and we have already seen that each such monomial has remainder ca^n.
This means that $cx^n - ca^n$ is divisible by $(x - a)$. Adding up various such differences, we see that $f(x) - f(a)$ is indeed divisible by $(x - a)$. Hence $f(x)$ has remainder $f(a)$. [13]
We have thus proved the well known
Remainder theorem The linear expression $(x - a)$ divides a polynomial $f(x)$ if and only if $f(a) = 0$.
We have also proved, in view of our concrete steps:
The geometric series formula
For any positive integer n we have

$$1 + x + \cdots + x^{n-1} = \frac{x^n - 1}{x - 1}.$$

[13] The informal proof above can me made precise in the following way. Some of you might find it easier to understand this version.

We have proved that for each $n = 0, 1, \cdots$ we have polynomials

$$q_n(x) = x^{n-1} + ax^{n-2} + \cdots + a^{n-2}x + a^{n-1}$$

so that

$$x^n - a^n = q_n(x)(x - a).$$

Now let $f(x) = c_0 + c_1x + \cdots c_dx^d$ be a polynomial of degree d. Then we get:

$$f(x) - f(a) = (c_0 - c_0) + c_1(x - a) + c_2(x^2 - a^2) + \cdots + c_d(x^d - a^d)$$

and this simplifies to

$$f(x) - f(a) = (c_1q_1(x) + c_2q_2(x) + \cdots + c_dq_d(x))(x - a).$$

Hence the remainder of $f(x)$ when divided by $(x - a)$ is $f(a)$.

For proof, proceed as follows:

- Start with the established identity:

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

- Set $a = 1$ and deduce:

$$(x^{n-1} + x^{n-2} + \cdots + 1)(x - 1) = x^n - 1$$

- Rewrite as:

$$(x^{n-1} + x^{n-2} + \cdots + 1) = \frac{x^n - 1}{x - 1}.$$

Example 7: Division by a quadratic polynomial. Let a be a constant. Find the division and the remainder when $(x^2 - a)$ is divided into various powers of $x, x^2, x^3, x^4, \cdots$.
Using these results, deduce a formula for the remainder when a polynomial $f(x)$ is divided by $(x^2 - a)$.
Answer: As before we do some initial cases for understanding and experience: We wish to divide various x^n by $x^2 - a$ and find the division and remainder.

Case $n = 0$

$$x^0 = 1 = (0)(x^2 - a) + 1 \text{ so } q(x) = 0, r(x) = 1.$$

Case $n = 1$

$$x^1 = x = (0)(x^2 - a) + x \text{ so } q(x) = 0, r(x) = x.$$

Case $n = 2$

$$x^2 = (x^2 - a) + a \text{ so } q(x) = 1, r(x) = a.$$

Case $n = 3$

$$x^3 = x(x^2 - a) + ax \text{ so } q(x) = x, r(x) = ax.$$

Case $n = 4$

$$x^4 = x^2(x^2 - a) + a(x^2 - a) + a^2 \text{ so } q(x) = x^2 + a, r(x) = a^2.$$

Case $n = 5$

$$x^5 = x(x^2 + a)(x^2 - a) + a^2x \text{ so } q(x) = x^3 + ax, r(x) = a^2x.$$

Case $n = 6$

$$x^6 = x^2(x^3 + ax)(x^2 - a) + a^2(x^2 - a) + a^3 \text{ so } q(x) = x^4 + ax^2 + a^2, r(x) = a^3.$$

This is enough experience!
Look at the answers for even powers:

$n =$	$q(x)$	$r(x)$
2	1	a
4	$x^2 + a$	a^2
6	$x^4 + ax^2 + a^2$	a^3

We now guess and prove the formula for x^8. The guess is
$q(x) = x^6 + ax^4 + a^2x^2 + a^3$ and $r(x) = a^4$.
The proof is this:

$$\begin{aligned} x^8 &= x^2(x^6) \\ &= x^2(x^4 + ax^2 + a^2)(x^2 - a) + x^2(a^3) \\ &= x^2(x^4 + ax^2 + a^2)(x^2 - a) + (x^2 - a)(a^3) + a^4 \\ &= (x^6 + ax^4 + a^2x^2 + a^3)(x^2 - a) + a^4. \end{aligned}$$

Here is a challenge to your imagination and hence a quick road to the answer!
Name the quantity x^2 as y and note that then we are trying to divide it by $(y - a)$.
We already learned how to find the answers as polynomials in y. use them and the plug back $y = x^2$ to get our current answers!
But what do we do for an odd power? Suppose that we know the answer for an even power:

$$x^{2m} = q(x)(x^2 - a) + r(x).$$

We now know that $r(x) = a^m$.
Then the answer for x^{2m+1} shall be given by:

$$x^{2m+1} = x(q(x)(x^2 - a) + r(x)) = xq(x)(x^2 - a) + a^m x.$$

Since $a^m x$ is zero or has degree less than 2, it is the remainder.
Combining the above work, we see that we at least have the complete formula for the remainders of x^n. It is this: Write $n = 2m + s$ where $s = 0$ or 1. (In other words, s is the remainder when we divide n by 2.)
Then the remainder of x^n is $a^m x^s$!
For a general polynomial, rather than writing a general formula, let us illustrate the method of calculation.
Say $f(x) = x^5 + 2x^4 + x^3 - x + 3$. Find the remainder when it is divided by $x^2 - 2$.
The remainders of the five terms are as follows:

term	Even or odd	remainder
x^5	odd	$2^2x = 4x$
$2x^4$	even	$2(2^2) = 8$
x^3	odd	$2x$
$-x$	odd	$(-1)x = -x$
3	even	3

So finally the remainder of $f(x)$ is

$$4x + 2x - x + 8 + 3 = 5x + 11.$$

Note that we have not kept track of $q(x)$. If we need it, we can use the definition and calculate by dividing $f(x) - r(x)$ by $x^2 - 2$.

Enhanced remainder theorem!

Here is an even simpler idea to calculate the remainder. Note that we ultimately have $f(x) = q(x)(x^2 - a) + r(x)$. If we agree to replace every "x^2" by a in $f(x)$, then we notice that on the right hand side the part $q(x)(x^2 - a)$ will become zero. So, $f(x)$ will eventually reduce to the remainder!

Here is how this wonderful idea works for our problem solved above:

Current polynomial	Explanation
$x^5 + 2x^4 + x^3 - x + 3$	Start
$2x^3 + 4x^2 + 2x - x + 3$	x^2 in each term is replaced by 2
$4x + 8 + 2x - x + 3$	two more reductions done
$5x + 11$	collected terms

Naturally, this is much easier than the actual division.

Chapter 4

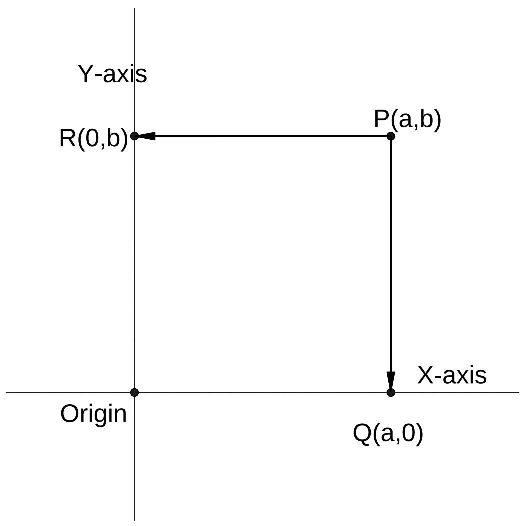
Introduction to analytic geometry.

While algebra is a powerful tool for calculations, it is too abstract for visualization. It is also hard to figure out what to prove when only algebraic calculations are available.
Hence, we next introduce the algebraic version of geometric concepts.
Here is a quick summary of our plan. We think of the geometric objects called lines, planes, three space, etc.
These are made up of points. We associate numbers (single or multiple) to each point so that we can express various geometric concepts and calculations as algebraic concepts and calculations.
In geometry, we first study points in a line, then a plane and then a three (or higher) dimensional space.
In most of this chapter, we will use $\Re$ the field of real numbers as our number field. The definitions extend to other fields, but the pictures are either too complicated or not helpful!
Much of this is a review of what you might have already seen in High School. The main difference is that we present some new useful formulas and work extensively with parametric descriptions, which may be a new concept.

4.1 Coordinate systems.

To work with **a line,** we agree to choose a special point called **the "origin"** and a scale so that a certain other point is at a distance of 1 unit from the origin - to be called **the unit point.** This sets up a system of coordinates (associated real numbers) to every point of the line and we can handle the points by using the algebraic operations of real numbers.

If a point named P has coordinate x, then we can also indicate this by calling it the point $P(x)$.
To work with **a plane,** we agree to set up a special point called **the origin** and a pair of perpendicular lines called the x and y axes.
On each of the axes, we agree to a scale so that every axis has a coordinate system of its own. Then every point in the plane can be associated with a pair of real numbers called coordinates and these are described in the following way:
Given any point P in the plane, draw a line from it parallel to the y axis. Suppose that it hits the x-axis at a point Q. Then the coordinate Q on the x-axis is called the x coordinate of P. Similarly a line is drawn through P parallel to the x-axis. Suppose that it hits the y-axis at a point R. Then the coordinate of R on the y-axis is called the y coordinate of P.
If these coordinates are (a, b) respectively, then we simply write $P(a, b)$ or say that P is the point (with coordinates) (a, b).



Thus the chosen origin has coordinates $(0, 0)$, the x-axis consists of all points with coordinates of the form $(a, 0)$ with $(1, 0)$ being the unit point on the x axis.
Similarly, the y-axis consists of all points with coordinates $(0, b)$ with its unit point at $(0, 1)$.
An algebraically minded person can turn this all around and simply declare a plane as a set of pairs of real numbers (a, b) and declare what is meant by the origin and lines and other various geometric objects contained in the plane. [1]
This has the advantage of conveniently defining the three space as a set of all triples of real numbers (a, b, c), where each triple denotes a point. The x-axis is now all

[1] When we take this algebraic viewpoint, then we don't need pictures and we can take more general numbers. We can still use the pictures based on real numbers for inspiration.

points of the form $(a, 0, 0)$, the y -axis is full of points of the form $(0, b, 0)$ and the z-axis is full of points of the form $(0, 0, c)$.
Evidently, now it is very easy to think of $4, 5, 6$ or even higher dimensions!

4.2 Geometry: Distance formulas.

Distance on a line.
For points $A(a), B(b)$ on a line, the distance between them is given by $|b - a|$ and we may write $d(A, B)$ to denote it.
For example, given points $A(5), B(3), C(-1)$, the distance between A and B is

$$d(A, B) = |3 - 5| = | - 2| = 2.$$

The distance between A and C is

$$d(A, C) = |(-1) - (5)| = | - 6| = 6.$$

Finally, the distance between C and B is

$$d(C, B) = |(3) - (-1)| = |4| = 4.$$

We also have reason to use a signed distance, which we shall call the **shift** from A to B as $b - a$. Sometimes, we may wish do denote this formally as $\overrightarrow{AB}$.[2]
Warning: The notation $\overrightarrow{AB}$ is often replaced by a simpler $\overline{AB}$ where the direction from A to B is understood by convention.
There are also some books on Geometry which use the notation $\overrightarrow{AB}$ to denote the *ray* from A towards B, so it includes all points in the direction to B, both between A, B and beyond B. So, when you see the notation, carefully look up its definition.
Our convention for a ray. We shall explicitly write "ray AB " when we mean to discuss all points of the line from A to B extended further out from B.
Thus, if you consider our picture above, the ray AB will refer to all points on the line with coordinates less than or equal to 5, the coordinate of A (i.e. to the left of A). The ray BA instead, will be all points with coordinate bigger than or equal to 3, the coordinate of B, or points to the right of B.

[2]What we are doing here is informally introducing the concept of vectors. One idea about vectors is that they are line segments with directions marked on them. This is a useful concept leading to the theory known as Linear Algebra. You may have occasion to study it later in great detail. The discussion of matrices in the Appendix is already a step in that direction.

This shift gives the distance when we take its absolute value but it tells more. If the shift is positive, it says B is to the right of A and when negative the opposite holds. [3]

Thus, for our points A, B, C, we get

$$\overrightarrow{AB} = 3 - 5 = -2\ ,\ \overrightarrow{AC} = -1 - 5 = -6 \text{ and } \overrightarrow{CB} = 3 - (-1) = 4.$$

Thus we see $\overrightarrow{AB}$ is negative, so A is to the right of B. On the other hand, $\overrightarrow{CB}$ is positive and hence B is to the right of C. This is clear from the picture as well.

Distance in the plane.

For points $P_1(a_1, b_1), P_2(a_2, b_2)$ in a plane, we have two shifts which we conveniently write as a pair $(a_2 - a_1, b_2 - b_1)$ and call it the shift from P_1 to P_2. As before, we may use the notation $\overrightarrow{P_1P_2}$ which is clearly very convenient now, since it has both the components built into it!

Alert! When we write $\overrightarrow{P_1P_2}$, be sure to subtract the coordinates of the *first* point from the *second*. It is a common mistake to subtract the second point instead, leading to a wrong answer.

Now, the distance between the two points has to be computed by a more complicated formula using the Pythagorean Theorem [4] and we have:

$$\textbf{The distance formula } d(P_1, P_2) = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2} = |\overrightarrow{P_1P_2}|.$$

To make it look similar to a line, we have added a more suggestive notation for it: $|\overrightarrow{P_1P_2}|$.

The signs of the two components of the shift also tell us of the relative position of P_2 from P_1. Basically it says which quadrant the point P_2 lives in when viewed from P_1. Because of this, we shall also call this shift by the term **direction numbers** from the first point to the second. [5]

For example, now consider the points $P(3, 2), Q(2, 3), R(-1, -1)$. Study their display in the plane:

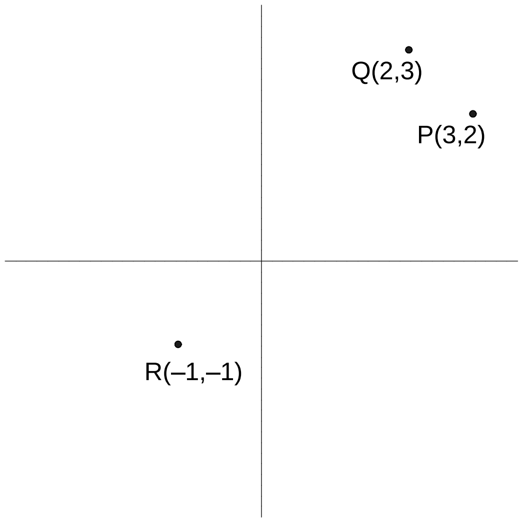
---

[3] Actually, "right" is not the right word here! We should really say that the direction from A to B is the same as direction from the origin to the unit point. Typically, the number line is drawn with the unit point to the right of the origin and hence we used the word "right".

[4] The well known Pythagorean Theorem states that the square of the length of the hypotenuse in a right angle triangle is equal to the sum of the squares of the lengths of the other two sides.

This theorem has arguably been discovered and proved by mathematicians in different parts of the world both before and after Pythagoras!

[5] Some books may define the direction numbers as $\frac{\overrightarrow{P_1P_2}}{|\overrightarrow{P_1P_2}|}$. This makes them less dependent of the chosen points (except for their order). We do not need this convention.



The shift from P to Q is

$$\overrightarrow{PQ} = (2-3, 3-2) = (-1, 1)$$

and the distance from P to Q is

$$d(P, Q) = \sqrt{(-1)^2 + (1)^2} = \sqrt{2}.$$

Notice that from the position P, the point Q does appear to be one unit to the left and one unit up!
Verify that for the points Q, R, we get $\overrightarrow{QR} = (-1-2, -1-3) = (-3, -4)$ and $d(Q, R) = \sqrt{(-3)^2 + (-4)^2} = \sqrt{25} = 5$.
Similarly for the points P, R, we get $\overrightarrow{PR} = (-1-3, -1-2) = (-4, -3)$ and hence the distance $d(P, R) = 5$ also.
We can similarly handle points in a three dimensional space and set up shifts and distance formulas. The higher dimensions can be handled as needed.

4.2.1 Connection with complex numbers.

Recall that each complex number is expressed as $z = x + iy$ where x and y are respectively its real and imaginary parts.
When we identify it with a point $P(x, y)$ in the plane, we get a resulting Argand diagram.[6] Sometimes, we may find it convenient to just write $P(z)$ in place of $P(x, y)$; but we will avoid this abuse of notation as far as possible.
Here are the geometric meanings of some of the concepts of complex numbers.

[6]Recall that we had introduced the Argand diagram in the introduction of complex numbers. We shall be referring to it repeatedly in various places.

1. The complex conjugate $\overline{z}$ is the point $(x, -y)$ and can be thought of as a reflection in the x-axis of the point corresponding to z.

2. The absolute value satisfies $|z|^2 = x^2 + y^2$. So, if O is the origin $(0, 0)$, then we have:
$$d(O, P)^2 = |z|^2 \text{ or } |z| = d(O, P).$$

3. More generally, if $z_1 = x_1 + iy_1, z_2 = x_2 + iy_2$ are two complex numbers and P_1, P_2 are their respective points in the Argand diagram, then
$$d(P_1, P_2) = |z_2 - z_1|.$$

4.3 Change of coordinates on a line.

In the case of the coordinates of a line, we had to choose an origin and then a scale and indeed we inadvertently chose a direction as well, by deciding where the unit point was.
What happens if we choose a different origin and a different scale?
Clearly the coordinates of a point will change. We would like to create an algebraic recipe for the resulting change.
We first illustrate this by an example.
Let us say that we have a coordinate system so that each point has an assigned coordinate. We shall call these the original x-coordinates.
We would like to choose a new origin at the point $Q(3)$ - the point with original x-coordinate 3. Let us choose the new unit point to the right of Q at distance 1, so it would be at $x = 4$.
Let x be the value of the original coordinate of some point and let x' be its new coordinate. Then we claim the formula:

$$x' = x - 3$$

We shall say that under this coordinate change the point $P(x)$ goes to the point $P'(x - 3)$.
Thus the point $Q(3)$ goes to the point $Q'(3 - 3) = Q'(0)$. This is the new origin!
Also, it is easy to see that the distance between any two points stays unchanged whether we use the x or the x'coordinates. [7]
A most general change of coordinates can be described by a substitution $x' = ax + b$ where a is non zero. Its effect can be analyzed as described below.

[7] An alert reader can verify this thus: The distance between two points $P_1(a_1), P_2(a_2)$ is given by $|a_2 - a_1|$. In new coordinates it becomes $|(a_2 - 3) - (a_1 - 3)| = |a_2 - a_1|$, i.e. it is unchanged. Indeed, the new coordinates also preserve all shifts.

1. Case $a > 0$: The origin is shifted to the point with x-coordinate $-b/a$. All shifts are multiplied by a and hence the direction is preserved. The distances are also scaled (multiplied) by a.

   For example, let $a = 3, b = -1$. Consider the points $P(2), Q(5)$. The new coordinate of P shall be $3(2) - 1 = 5$. Thus $P(2)$ becomes $P'(5)$.

   Similarly for the point Q we get $3(5) - 1 = 14$ and thus $Q(5)$ becomes $Q'(14)$.

   The original shift $\overrightarrow{PQ} = 5 - 2 = 3$. The new shift is $14 - 5 = 9$ which is 3 times the old shift.

   The distance 3 becomes $3(3) = 9$.

2. Case $a < 0$: The origin is still shifted to the point with x-coordinate $-b/a$. All shifts are multiplied by $a < 0$ and hence the direction is reversed. The distances are scaled (multiplied) by $|a| = -a$.

   For example, let $a = -2, b = -5$.

   For the same points $P(2), Q(5)$, the new coordinates become $(-2)2 - 5 = -9$ and $(-2)5 - 5 = -15$. Thus $P(2)$ goes to $P'(-9)$ and $Q(5)$ goes to $Q'(-15)$.

   Thus the new shift $\overrightarrow{P'Q'}$ is $-15 - (-9) = -6$ which is indeed the old shift $\overrightarrow{PQ} = 3$ multiplied by $a = -2$. The distance changes from 3 to 6.

The reader should investigate what happens to the unit points or other specific given points.

4.4 Change of coordinates in the plane.

We now generalize the above formulas and discuss some special cases.
Thus we have:

1. Translation or Change of origin only. A transformation $x' = x - p$, $y' = y - q$ changes the origin to the point with original (x, y) coordinates (p, q). All shifts and distances are preserved.

2. Change of origin with axes flips. A transformation $x' = ux - p, y' = vy - q$ where u, v are ± 1 changes the origin to the point with original (x, y) coordinates $(p/u, q/v)$. Since u, v are both ± 1, it is clear that this simplifies to (up, vq).

   All distances are preserved. However, shifts can change and we list their properties in the form of a table. It is recommended that the reader verifies these by making a picture similar to the worked out problem below.

Value of u	Value of v	Effect
1	-1	y axis is flipped. A reflection about the x axis.
-1	1	x axis is flipped. A reflection about the y axis.
-1	-1	Both axes are flipped. A reflection about the origin.

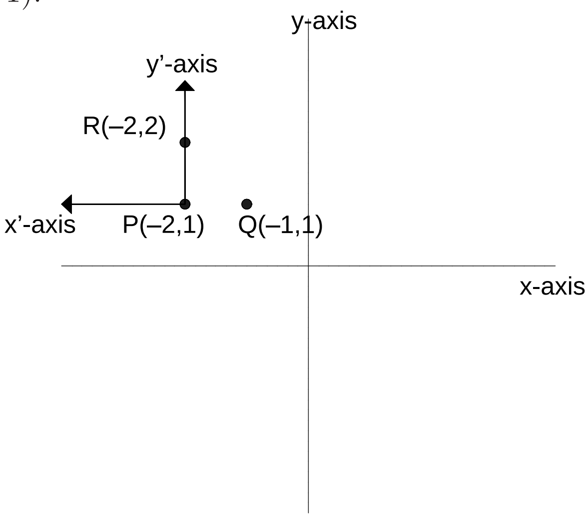
Thus, the transformation consists of first a translation to make (up, uq) as the new origin followed by the indicated flip(s).

The process of identifying the new origin and describing the flips is analyzing the images of a few well chosen points.

As an illustration, let us see what a transformation $x' = -x - 2$, $y' = y - 1$ does. The point $P(-2, 1)$ becomes $P'(0, 0)$ or the new origin.

All points of the form $(a, 1)$ get transformed to $(-a - 2, 0)$ i.e. become the new x' axis. The point $Q(-2 + 1, 1) = (-1, 1)$ which has a shift of $(1, 0)$ from P goes to $(-1, 0)$ showing that there is a flip of the x axis direction.

Similarly the points $R(-2, b)$ will map to the $(0, b - 1)$ to become the new y' axis. There is no flip involved, since the point $R(-2, 2)$ with shift $(0, 1)$ from P goes to $R(0, 1)$.



The general idea consists of three steps.

(a) First determine what becomes the new origin - say it is $P(p, q)$.

(b) Then find the images of $Q(p + 1, q)$ and $R(p, q + 1)$.

(c) Compare the original triangle PQR with its image for any possible flips!

4.5 General change of coordinates.

We can indeed stretch these ideas to analyze a general change of coordinates in the plane. We will, however, leave the analysis for higher courses. We shall present more details in the Appendix after developing the concepts from Trigonometry and complex numbers.
First we state the most important facts for information and inducement to study further.
The most general (linear) coordinate change in the plane can be written as

$$x' = a(x-p) + b(y-q)\ ,\ y' = c(x-p) + d(y-q).$$

where a, b, c, d are constants such that the determinant $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \neq 0$.
It is easy to see that the origin is shifted to the point with original coordinates (p, q), but shifts and distances change in a complicated manner. If you wish to study geometric properties like distances and angles, you must investigate changes which preserve distances. These are the so called isometries ("iso" means same, "metry" is related to measure)!
The changes of coordinates which preserve distances can be described nicely using trigonometry.

4.5.1 Description of Isometies.

Here is a description of all isometries in the plane, presented without proof.
We know how to translate and so to simplify matters, we assume that there is no translation involved.
Distance preserving transformations can be achieved exactly when there is an angle θ and a number $u = \pm 1$ such that

Isometry. $$x' = \cos(\theta)x + \sin(\theta)y\ ,\ y' = -u\sin(\theta)x + u\cos(\theta)y.$$

The transformation can be described as a rotation by θ if $u = 1$ and it is a rotation by θ followed by a flip of the y' axis if $u = -1$.
The trigonometric functions $\cos(\theta), \sin(\theta)$ have been studied from ancient times and here is a quick definition when $0 < \theta < 90$ in degrees. Make a triangle ABC where the angle at B is the desired θ and the one at C is 90 degrees. Then

$$\cos(\theta) = |\overrightarrow{BC}|/|\overrightarrow{BA}| \text{ and } \sin(\theta) = |\overrightarrow{CA}|/|\overrightarrow{BA}|.$$

We have already described the general concept of angles and these functions in our description of the Euler representation of complex numbers.

4.5.2 Complex numbers and plane transformations.

In view of the Argand diagram, we can think of the "real plane" as a "complex line". Naturally, in analogy with the real line, we can expect a change of complex coordinates as given by

$$z' = pz + q \text{ or } z' = \overline{pz + q} \text{ where } p, q \text{ are complex numbers with } p \neq 0 \;.$$

It turns out that what we have described as an "Isometry" in the above section can be described as a transformation setting the multiplier p to be a complex number with $|p| = 1$, i.e. the point corresponding to p is on the unit circle!
We make this explicit. **For convenience,** we are only discussing **the case $\mathbf{q = 0}$**, since we already know how translations behave.
Let $z = x + iy$ be the complex number associated with the original coordinates (x, y). Let $z' = (x', y')$ be the corresponding complex number in terms of the new coordinates.
Then the above equations reduce to the following:

- Set $p = \exp(i\theta), q = 0$, where θ is some angle.
- If $u = 1$ then $z' = pz = \exp(i\theta)z$ for some angle θ.
- If $u = -1$ then $z' = \overline{pz} = \overline{\exp(i\theta)z}$ which is the conjugate of the complex number $pz = \exp(i\theta)z$.

Thus, if we think of points in the plane as complex numbers, then the transformation is either multiplication by a complex number with length 1 or such a multiplication followed by a conjugation!
What does it mean when we use an arbitrary non zero complex number p? We don't get an isometry, but we get **an isometry followed by a scaling** by $|p|$.
Moreover, this characterizes all transformations of the plane (without translation) which **preserve angles** between lines.
Here is a sample of what happens.

4.5.3 Examples of complex transformations.

- **Problem.** Let $p = 1 + i$. Note that $p = \sqrt{2}\exp(i\pi/4)$. Describe the nature of the plane transformation corresponding to $z' = pz$.

  Answer.

  Our transformation leads to the equations:

  $$x' + iy' = z' = (1 + i)z = (1 + i)(x + iy) \text{ or } x' = x - y, y' = x + y.$$

Thus, the unit point $(1, 0)$ on the x-axis goes to $(1, 1)$ which is a rotation by $45°$. Also, its distance from the origin changes from 1 to $\sqrt{1^2 + 1^2} = \sqrt{2}$.

Similarly, the unit point $(0, 1)$ on the y-axis goes to $(-1, 1)$ which is a rotation by $45°$. Also, its distance from the origin changes from 1 to $\sqrt{1^2 + 1^2} = \sqrt{2}$.

More generally, for any number z, we see that

$$|z'| = |1 + i||z| = \sqrt{2}|z|.$$

Also the argument of z increases by $45°$ which the argument of the complex number $1 + i$ in degrees.

- **Problem.** Conversely, consider the transformation

$$x' = 3x - 4y, y' = 4x + 3y.$$

Assuming this to be a complex transformation $z' = pz + q$, determine p, q. Describe the corresponding rotation and scaling.

Answer. Note that $q = 0$, since there is no translation involved. Assuming $p = a + bi$ we see that [8]

$$x' + iy' = (a + bi)(x + iy) \text{ or } x' = ax - by, y' = bx + ay.$$

Comparing with the given transformation, we see that $a = 3, b = 4$, so $p = 3 + 4i$.

$|p| = \sqrt{3^2 + 4^2} = 5$, so this is a scaling by a factor of 5.

The rotation is by an angle θ for which $\cos(\theta) = 3, \sin(\theta) = -4$. Using the calculator, you can decide the angle to be approximately $-53.13°$. This means it is a clockwise rotation by the angle of approximately $53.13°$.

4.5.4 Examples of changes of coordinates:

We illustrate the above discussion of coordinate changes in the line and plane by various examples.

Example 1 Find a transformation on the line which changes the point $A(3)$ to $A'(5)$ and $B(-1)$ to $B'(1)$.

Use your answer to find out the new coordinate of the point $C(3)$. Also, if $D(d)$ goes to $D'(-5)$ what is d?

Answer:

[8] **Note.** The calculation shows that the transformations of the type $z' = pz$ have a very special form. The transformation $z' = \overline{pz}$ is also very special and you can work out its form. The complex number p can be easily read off from the equations of transformation.

- We assume that the transformation is $x' = px + q$ and try to find p, q. From the information about the point A, we see that

$$5 = p(3) + q = 3p + q.$$

The information about B says:

$$1 = p(-1) + q = -p + q.$$

We need to solve these two equations together. Of course, we can use any of the techniques that we know, but here, it is easiest to subtract the second from the first to get:

$$4 = 3p + q - (-p + q) = 4p.$$

So $p = 1$. Plugging the value of p into the first equation, we see that:

$$5 = 3 + q \text{ so } q = 2.$$

Thus the transformation is:

$$x' = x + 2.$$

- To find the image of $C(3)$, we set $x = 3$ and deduce that $x' = 3 + 2 = 5$. So, $C(3)$ becomes $C'(5)$.
- To find the point D, we set $x' = -5$ and solve $-5 = x + 2$ to get $x = -7$. So $d = -7$.

Example 2 Find a plane transformation using only translations and flips which sends $P(1, 1)$ to $P'(4, 1)$, $Q(2, 3)$ to $Q'(5, -1)$.

Determine the new image of the origin. Also determine the point which goes to the new origin.

What point R goes to $R'(3, 3)$?

Answer:

- Assume that the transformation is given by

$$x' = ux - p, y' = vy - q.$$

We have the equations for point P:

$$4 = u - p \ , \ 1 = v - q.$$

Similarly for the point Q, we get:

$$5 = 2u - p\ ,\ \ -1 = 3v - q.$$

Solve the equations

$$u - p = 4 \text{ and } 2u - p = 5$$

to deduce $u = 1, p = -3$.
Similarly, solve the equations

$$v - q = 1 \text{ and } 3v - q = -1$$

to deduce $v = -1, q = -2$.
Thus the transformation is:

$$x' = x + 3\ ,\ y' = -y + 2.$$

- Clearly, the origin $O(0,0)$ goes to $O'(3,2)$.
- If $x' = 0$ and $y' = 0$ then clearly, $x = -3$ and $y = 2$, so the point $S(-3,2)$ becomes the new origin!
- If $x' = 3$ and $y' = 3$, then $x = 0$ and $y = -1$, so $R(0,-1)$ is the desired point.

Chapter 5

Equations of lines in the plane.

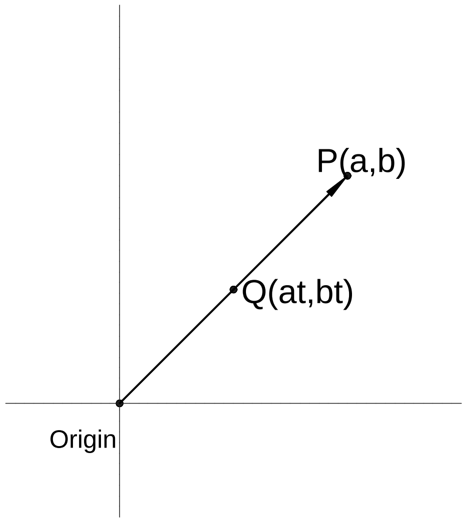
You are familiar with the equations of lines in the plane of the form $y = mx + c$. In this chapter, we introduce a different set of equations to describe a line, called the parametric equations. These have distinct advantages over the usual equations in many situations and the reader is advised to try and master them for speed of calculation and better understanding of concepts.

5.1 Parametric equations of lines.

Consider a point $P(a, b)$ in the plane. If t is any real number, we can see that the point (at, bt) lies on the line joining P to the origin. Define $Q(t)$ to be the point (at, bt). We may simply write Q for $Q(t)$ to shorten our notation.

It is easy to see that the line joining P to the origin is filled by points $Q(t)$ as t takes all possible real values. (See the picture below.) [1]

[1] The formal proof, however, is not so easy since we are assuming that we know what a line is, without ever defining it formally! An algebraist resolves the problem by declaring that a line is simply the set of all points satisfying a linear equation! Then he gets lines defined even for other number fields. A geometer has to simply rely on a geometric intuition and a picture; both can be misleading from time to time. The problem of formalizing and proving the concepts of Geometry are tackled in higher courses on College Geometry.

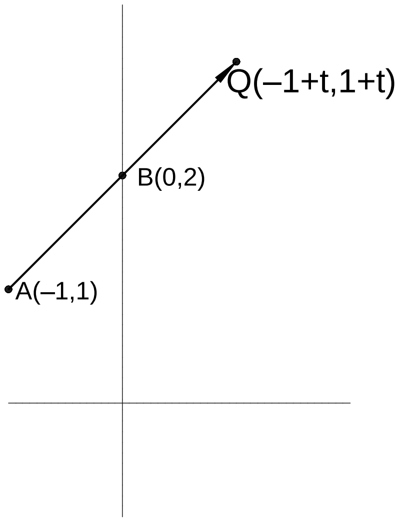


Thus, if we write O for the origin, then we can describe our formula as

$$\overrightarrow{OQ} = (at, bt) = t(a, b) = (t)(\overrightarrow{OP}) \text{ for some real } t \ .$$

More generally, a line joining a point $A(a_1, b_1)$ to $B(a_2, b_2)$ can then be described as all points $Q(x, y)$ such that $\overrightarrow{AQ} = (t)(\overrightarrow{AB})$, i.e. $(x - a_1, y - b_1) = t(a_2 - a_1, b_2 - b_1)$ and this gives us the:

Parametric two point form: $x = a_1 + t(a_2 - a_1)$, $y = b_1 + t(b_2 - b_1)$.



Here is a concrete example where we take points $A(-1, 1)$ and $B(0, 2)$. (See the picture above.)

The parametric equation then comes out as

$$x = -1 + t(0 - (-1)) = -1 + t \text{ and } y = 1 + t(2 - 1) = 1 + t.$$

We have marked this point as Q in the picture.
Even though, we have a good formula in hand, we find it convenient to rewrite it in different forms for better understanding. Here are some variations:

- The shift $\overrightarrow{AB}$ can be conveniently denoted as $B - A$, where we think of the points as pairs of coordinates and perform natural term by term operations on them. Now, our formula can be conveniently written as:

  Compact parametric two point form:

  $$Q(x, y) = A + t(B - A) \text{ or simply } (x, y) = A + t(B - A).$$

  For example, for our $A(-1, 1)$ and $B(0, 2)$ we get:

  $$A+t(B-A) = (-1, 1)+t\,((0, 2) - (-1, 1)) = (-1, 1)+t(1, 1) = (-1+t, 1+t) \text{ as before.}$$

- Note that this formula can also be rewritten as: $Q(x, y) = Q = (1-t)A+t(B)$.

  Here, the calculation gets a bit simpler:

  $$(1 - t)(-1, 1) + t(0, 2) = (-1 + t, 1 - t) + (0, 2t) = (-1 + t, 1 + t).$$

  In this form, the parameter t carries more useful information which we investigate next.

Examples of parametric lines:

1. Find a parametric form of the equations of a line passing through $A(2, -3)$ and $B(1, 5)$.

   Answer: The two point form gives:

   $$x = 2 + t(1 - 2)\ ,\ y = -3 + t(5 - (-3)) \text{ or } x = 2 - t\ ,\ y = -3 + 8t.$$

   The compact form gives the same equations:

   $$(x, y) = (2, -3) + t\,((1, 5) - (2, -3)) = (2, -3) + t(-1, 8) = (2 - t, -3 + 8t).$$

2. Verify that the points $P(1,5)$ and $Q(4,-19)$ lie on the line given above.

   Answer: We want to see if some value of t gives the point P. This means:

   $$\text{Can we solve? } 1 = 2 - t\ ,\ 5 = -3 + 8t.$$

   A clear answer is yes, $t = 1$ does the job. Similarly, $t = -2$ solves:

   $$4 = 2 - t, -19 = -3 + 8t.$$

   Thus P and Q are points of the above line through A, B.

3. Determine the parametric form of the line joining P, Q. Note that we must get the same line! Compare with the original form.

   Answer: The compact form would be given by:

   $$(x,y) = P + t(Q - P) = (1,5) + t((4,-19) - (1,5)) = (1,5) + t(3,-24).$$

   The original compact form of the line was:

   $$(x,y) = (2,-3) + t(-1,8).$$

   An observation.

   Note that the pair of coefficients multiplying t were $(-1,8)$ for the original equation and are $(3,-24)$ for the new one. Thus the second pair is -3 times the first pair.

   This is always the case! Given two parametric forms of equations of the same line, the pair of coefficients of t in one set is always a non zero multiple of the pair of coefficients in the other set.

   Thus, it makes sense to make a:

   Definition: Direction numbers of a line. The pair of multipliers of t in a parametric form of the equation of a line are called (a pair of) direction numbers of that line. Any non zero multiple of the pair of direction numbers is also a pair of direction numbers for the same line.

   The pair of direction numbers can also be understood as the shift $\overrightarrow{PQ}$ for some two points on the line!

   In addition, if we set the parameter $t = 0$, then we must get some point on the line.

   In turn, if we have two lines with one common point and proportional direction numbers, then they must be the same line!

Thus we can easily manufacture many other parametric forms of the same line by starting with any one point on the line and using any convenient direction numbers.

For example, the above line can be also conveniently written as:

$$(x, y) = (2, -3) + t(-2, 16) = (2 - 2t, -3 + 16t).$$

Here, we chose a point $(2, -3)$ on the line and chose the direction numbers as $(-2, 16)$ - a multiple by 2 of our original direction numbers $(-1, 8)$.

If we choose the point $(4, -19)$ instead and take $(5, -40) = (-5)(1, -8)$ as direction numbers, then we get:

$$(x, y) = (4, -19) + t(5, -40) = (4 + 5t, -19 - 40t).$$

You can make many other examples! Thus, the parametric form is good for generating many points on a chosen line (by taking many values of t) but is not so good to decide if it describes the same line that we started with.

5.2 Meaning of the parameter t:

Consider the parametric description of points on a line joining $A(a_1, b_1)$ and $B(a_2, b_2)$ given by

$$(x, y) = (1 - t)A + tB = (a_1 + t(a_2 - a_1), b_1 + t(b_2 - b_1)) .$$

We have the following information about the parameter t. [2]

Value of t	Interpretation or conclusion
$t = 0$	Point A
$0 < t < 1$	Points from A to B
$t = 1$	Point B
$t > 1$	Points past B away from A
$t < 0$	Points past A away from B

In other words, the parameter t can be thought as a coordinate system on the line joining A, B chosen such that A is the origin and B the unit point.
We can obviously get different coordinate systems, if we choose some other two points on the line for the origin and the unit point.
The above facts can be proved by working out the following precise formula.
For any point $Q = (1-t)A + tB$ on the line, let us note that

$$Q - A = (1-t)A + tB - A = t(B - A)$$

and

$$B - Q = B - ((1-t)A + tB) = (1-t)(B - A).$$

Thus the distances $|\overrightarrow{AQ}|$ and $|\overrightarrow{QB}|$ are indeed in the proportion $|t| : |(1-t)|$.
This says that to divide the interval AB in a ratio $t : (1-t)$ we simply take the point $Q = (1-t)A + tB$. For $0 < t < 1$ this has a natural meaning and we note the following:

Useful Formulas

1. **Midpoint formula.** The midpoint of A, B is given by $\dfrac{(A+B)}{2}$.
2. **Division formula.** For $0 < t < 1$ the point $Q = (1-t)A + tB$ divides the interval in the ratio $t : (1-t)$. Thus the two trisection points between A, B are given as follows:

   Take $t = 1/3$ and then $Q = (2/3)A + (1/3)B$ is the trisection point near A.

   Take $t = 2/3$ and then $Q = (1/3)A + (2/3)B$ for the point near B.

[2]Note that the expression $(1-t)A + tB$ describes all points "between A and B" when $0 \leq t \leq 1$ and this condition can also be described by saying that both the coefficients $t, 1-t$ are non negative.

Such a conclusion also works when we allow our points A, B in higher dimensions. This is the **power of the parametric method!**

Note that we could also describe our expression as $aA + bB$ where we assume that both a, b are non negative and add up to 1. This is also called the convex combination of A and B.

A natural extension of this idea can be made to three points A, B, C by writing $lA + mB + nC$ where $l + m + n = 1$ and l, m, n are non negative. The resulting set of points describes **the triangle formed by the three points, in any dimension.**

3. **External divisions formulas.** If $t > 1$ then the point $Q = (1-t)A + tB$ gives an external division point in the ratio $t : (t-1)$ on the B side. If $t < 0$ then $1 - t > 1$ and we get an external division point in the ratio $-t : 1 - t$ on the A side.

   Thus, for example, the point $Q = -A + 2B$ with $t = 2$ has the property that the distance $|\overrightarrow{AQ}| = 2|\overrightarrow{QB}|$.

4. **The distance formula** If $Q = (1-t)A + tB$, then the distance $|\overrightarrow{AQ}|$is simply $|t| \cdot |\overrightarrow{AB}|$ and consequently, for any other point $R = (1-s)A + sB$, we can calculate the distance

$$d(Q,R) = |\overrightarrow{QR}| = |s-t| \cdot |\overrightarrow{AB}| = |s-t| \cdot d(A,B).$$

Examples of special points on parametric lines:

1. Let L be the line (discussed above) joining $A(2,-3), B(1,5)$.

   Let us use the parametric equations found earlier:

$$x = 2 - t\ ,\ y = -3 + 8t$$

   What values of the parameter t give the points A, B and their midpoint M?

   Answer: At $t = 0$ we get the point A and at $t = 1$ we get the point B. At $t = \frac{1}{2}$ we get

$$\left(2 - \frac{1}{2}, -3 + 8(\frac{1}{2})\right) = (\frac{3}{2}, 1).$$

   This is the midpoint M.

2. Find the two trisection points of the segment from A to B.

   Answer: We simply take $t = \frac{1}{3}$ and $t = \frac{2}{3}$ in the parameterization.

   So, the two points are:

$$(2 - \frac{1}{3}, -3 + 8 \cdot \frac{1}{3}) = (\frac{5}{3}, -\frac{1}{3})$$

   and

$$(2 - \frac{2}{3}, -3 + 8 \cdot \frac{2}{3}) = (\frac{4}{3}, \frac{7}{3}).$$

   Warning! Don't forget that for this formula to work, we must take the parameterization for which the original two points are given by $t = 0$ and $t = 1$ respectively.

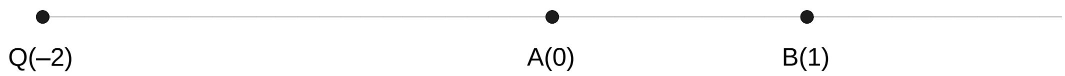
3. Find out where the point $Q(4,-19)$ is situated relative to A, B.

   Answer: First we decide that value of t which gives the point Q.

   Thus we solve:

   $$4 = 2 - t \text{ and } -19 = -3 + 8t.$$

   Clearly (as we already know) this is given by $t = -2$. For understanding, let us make a picture of the line and mark the values of t next to our points.



   The fact that $t = -2 < 0$ says that we have a point on the A side giving an external division in the ratio $\frac{-t}{1-t} = \frac{2}{3}$. Thus it confirms that the point Q is on the A-side and the ratio of the distances $d(Q, A)$ and $d(Q, B)$ is indeed $\frac{2}{3}$.

4. Let U be a point on the same line L joining the points $A(2,-3)$ and $B(1,5)$. Assume that in the usual parameterization which gives $t = 0$ at A and $t = 1$ at B, we have $t = 3$ at U. Assume that V is another point on the same line at which $t = -5$. Find the distance $d(U, V)$.

   Answer: We could, of course, calculate U by plugging in $t = 3$ in the parametric equations $(x, y) = (2 - t, -3 + 8t)$ and get $U(-1, 21)$. Similarly, we get $V(7, -43)$. We can apply the distance formula and declare:

   $$\sqrt{(7-(-1))^2 + (-43-21)^2} = \sqrt{8^2 + 64^2} = \sqrt{4160}.$$

   The final answer comes out to be $8\sqrt{65}$.

   But we recommend using a more intelligent approach! Note that we know:

   $$d(A, B) = \sqrt{(1-2)^2 + (5-(-3))^2} = \sqrt{1 + 64} = \sqrt{65}.$$

   Since we are using parameter values 3 and -5 for our two points, we know that their distance is simply:

   $$d(U, V) = |3 - (-5)| d(A, B) = 8\sqrt{65}.$$

   Clearly, this is far more efficient! [3]

[3] Also note that we did not rush to "evaluate" $\sqrt{65}$!

A decimal answer like 8.062257748 spit out by a calculator is not to be trusted. In fact, the calculator's square root of an integer is precise only when it is an integer.

Also, it tells us very little about the true nature of the number. The decimal approximation should only be used if the precise number is not needed or so complicated that it impedes understanding.

Actually, when the precise number gets so complicated, mathematicians often resort to giving the number a convenient name and using it, rather than an inaccurate value! The most common example of this is π - the ratio of the circumference of a circle and its diameter. It can never be written completely precisely by a decimal, so we just use the symbol!

5. Use the parametric form of the above line to verify that the point $S(-1, 20)$ is not on it.

   Answer:

   We try to solve:
   $$-1 = 2 - t \ , \ 20 = -3 + 8t.$$

   The first equation gives $t = 3$ and when substituted in the second equation, we get:
   $$20 = -3 + 8(3) = 21 \text{ a clearly wrong equation!.}$$

   This shows that the point S is outside the line!

5.3 Comparison with the usual equation of a line.

The parametric description of a line given above is very useful to generate lots of convenient points on a line, but if we wish to understand the position and the orientation of a line as well as its intersections with other lines and curves, we also need the usual equation of a line as a relation between its x, y coordinates.
We start with the parametric equations of a line through points $A(a_1, b_1), B(a_2, b_2)$

$$\text{Equation 1: } x = a_1 + t(a_2 - a_1) \ , \quad \text{Equation 2: } y = b_1 + t(b_2 - b_1)$$

We would like to get rid of this parameter t and we can eliminate it by the following simple trick.
Construct a new equation $(a_2 - a_1)$(Equation 2) $-$ $(b_2 - b_1)$(Equation 1) and we get:

$$(a_2 - a_1)y - (b_2 - b_1)x = (a_2 - a_1)b_1 - (b_2 - b_1)a_1 = (a_2b_1 - a_1b_2).$$

This clearly describes the precise condition that a point (x, y) lies on the line. We record this as:
The two point form of the equation of a line through $\mathbf{A(a_1, b_1), B(a_2, b_2)}$:

$$(a_2 - a_1)y - (b_2 - b_1)x = (a_2b_1 - a_1b_2).$$

If $a_2 = a_1$ then it is easy to see that the equation reduces to $x = a_1$ **a vertical line.**
Similarly, if $b_2 = b_1$, then we get a **horizontal line** $y = b_1$.
Example.
Thus for our points $A(2, -3), B(1, 5)$ we get:

$$(1 - (2))y - (5 - (-3))x = ((1)(-3) - (2)(5) \text{ or } -y - 8x = -13 \text{ or } y + 8x = 13.$$

You can also verify this by the elimination of t from the equations

$$x = 2 - t \ , \ y = -3 + 8t.$$

We now **define the slope of a line** joining any two points $A(a_1, b_1), B(a_2, b_2)$ as the ratio $m = \dfrac{b_2 - b_1}{a_2 - a_1}$ where we agree that it is ∞ in case $a_1 = a_2$ i.e. when we have a vertical line.
We also **define the intercepts of a line** as follows:
We declare:

- The **x-intercept** is a number a if $(a, 0)$ is on the line. From the above equation of the line, the x-intercept is clearly $\dfrac{a_2b_1 - a_1b_2}{b_1 - b_2}$.

- This may fail if $b_2 = b_1$ but $a_1b_2 \neq a_2b_1$, i.e. the line is horizontal but not the x-axis.

  In this case the intercept is declared ∞.

- When the line is $y = 0$ i.e. the x-axis itself, then the x-intercept is considered **undefined**!

- Thus, in general, the x-intercept is $\dfrac{a_2b_1 - a_1b_2}{b_1 - b_2}$ with a suitable agreement when the denominator is zero.

- The **y-intercept** is similarly defined as b, when the point $(0, b)$ is on the line. From the above equation of the line, the y-intercept is clearly $\dfrac{a_2b_1 - a_1b_2}{a_2 - a_1}$.

- This may fail if $a_2 = a_1$ but $a_1b_2 \neq a_2b_1$, i.e. if the line is vertical but not the y axis then the intercept is declared to be ∞.

- When the line is $x = 0$ i.e. the y-axis itself, then the y-intercept is considered **undefined**!

- In general, the formula is: $\dfrac{a_2b_1 - a_1b_2}{a_2 - a_1}$ with a suitable agreement when the denominator is zero.

- **Warning.** It is tempting to memorize the formulas for the intercepts, but notice that there is a subtle change in the form of the denominator. We recommend a fresh calculation rather than a memorized formula.

- For a non vertical line, it is customary to divide the equation by $a_2 - a_1$ and rewrite it as:

$$y - \frac{b_2 - b_1}{a_2 - a_1}x = \frac{a_2b_1 - a_1b_2}{a_2 - a_1}.$$

  This formula is worth memorizing.

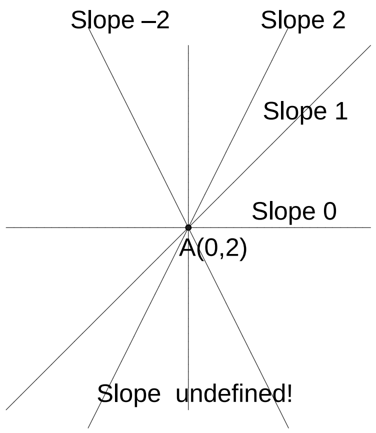
The last form above is often further restated as: **Slope intercept form of a line:** The slope intercept form is the familiar $\mathbf{y = mx + c}$ where:

$$m = \frac{b_2 - b_1}{a_2 - a_1} \textbf{ is the slope and } c = \frac{a_2 b_1 - a_1 b_2}{a_2 - a_1} \textbf{is the y-intercept.}$$

Thus, for our line through $A(2, -3)$,$B(1, 5)$, we have:

$$-y - 8x = -13 \text{ becomes } y + 8x = 13 \text{ and finally } y = -8x + 13.$$

Thus, the slope is -8 and the y-intercept is 13.
Often, one calls it just the intercept, the y being understood in this form.
We can calculate each of these quantities from the formulas and then write the equation directly.
It is convenient to understand the orientation of such a line $y = mx + c$.
If we fix c, then we get the family of lines with a common point $(0, c)$. When $m = 0$ the line is horizontal and as m increases, the line starts to twist upwards (in a counter clockwise direction). Higher and higher values of m bring it closer to vertical, the true vertical being reserved for $m = \infty$ by convention. All these lines appear to rise as the x-coordinates increase.
If we, in turn start to run m through negative values to $-\infty$, then the line twists clockwise with $m = -\infty$ corresponding to a vertical line. These lines appear to fall as x-coordinate increases.



While the above two point formula is good in all situations, sometimes it is useful to find fast clever ways to find the equation of a line.
Consider an equation of the form $ax + by = c$ where a, b, c are constants. This is clearly the equation of some line provided at least one of a, b is non zero.
Hence, if we can somehow construct such an equation satisfying the given conditions, then it must be the right answer. In fact, if we construct an equation which can be evidently manipulated to such a form, then it is a correct answer as well. [4]

[4] We may call this the "duck principle" named after the amusing saying: "If it walks like a duck and talks like a duck, then it is a duck." Thus, if an equation looks like the equation of a line and satisfies the required conditions, then it must be the right equation!

Here are some of the samples of this simple idea.
Another two point form.

$$(y - b_1)(a_2 - a_1) = (x - a_1)(b_2 - b_1).$$

Proof: It has the right form and is clearly satisfied by the two points, just substitute and see!
Point $(\mathbf{P}(\mathbf{a}, \mathbf{b}))$-slope(m) form.

$$(y - b) = m(x - a).$$

Proof: When rearranged, it becomes $y = mx + (b - ma)$.
So, m is the slope. Also, the equation is clearly satisfied, if we plug in $x = a, y = b$.
So, we are done by the "duck principle".
By the way, this work also gives us the formula for the y-intercept as $b - ma$.
The x-intercept, if needed is seen to be $a - \frac{1}{m} \cdot b$ and the evident symmetry in these two formulas is noteworthy!
Example. Find the equation of a line with slope -6 and passing through the point $P(3, 4)$.
Answer: Direct application of the formula gives:

$$(y - 4) = -6(x - 3).$$

This simplifies to:

$$y = -6x + 4 + 18 = -6x + 22.$$

It has slope -6 and intercept 22.
Intercept form. Suppose the x, y intercepts of a line are p, q respectively and these are non zero. Then the equation of the line can be written as:

$$\frac{x}{p} + \frac{y}{q} = 1.$$

This form can be interpreted suitably if one of p, q is zero, but fails completely if both are zero.
Proof. Just check that the two intercept points $(p, 0), (0, q)$ satisfy the equation and apply the duck principle!
Example. Find the equation of a line with x intercept 3 and y intercept 4.
Answer: Apply the formula:

$$\frac{x}{3} + \frac{y}{4} = 1.$$

This simplifies to:

$$4x + 3y = 12 \text{ or } y = -\frac{4}{3}x + 4.$$

Actually, it is debatable if this is really a simplification! One should not get too attached to a single standard form.

Parallel lines.

Two lines in the real plane are said to be parallel, if they have no point in common. From our solution of two linear equations in two variables, it is easy to see that two lines with different slope will always meet in some point and thus are not parallel. [5]

Thus it is clear that **parallel lines** have the **same slope.** [6] Thus a line parallel to a given $y = mx + c$ is given by $y = mx + k$ where k can be chosen to satisfy any further conditions. More generally, the line parallel to any given $ux + vy = w$ can also be written as $ux + vy = k$ for a suitable k. The point is that we don't have to put the equation in any standard form to get it right!

Example. Find a line parallel to $y+8x = 13$ which passes through the point $(3, 10)$.

Answer: We know that the answer can be of the form $y + 8x = k$. Impose the condition of passing through $(3, 10)$, i.e.

$$10 + 8(3) = k$$

so we deduce that $k = 34$ and hence the answer is

$$y + 8x = 34.$$

Perpendicular lines. It is possible to prove that a line **perpendicular to a line with slope m has slope $-1/m$,** where by convention we take $1/\infty = 0$ and $1/0 = \infty$.

The needed concept of angles will be discussed after the introduction of Trigonometry.

Thus, to make a perpendicular line to a given line of slope m, we can simply choose any line with slope $-1/m$.

But should we find the slope first? No! We give an alternative way using the duck principle.

---

[5]Without loss of generality, we may assume that one line is $y = mx + c$ and the other line is either $y = m'x + c'$ with $m \neq m'$ or $x = p$ for some p.

In the first case, the common point is:

$$x = -\frac{c - c'}{m - m'}, \quad y = -m\frac{c - c'}{m - m'} + c$$

is a common point, while in the second case:

$$x = p, \quad y = mp + c$$

is a common point.

[6]For an algebraist, this is the definition of a parallel line, and thus it can be extended to any number system, where a picture and an "easy" visualization may be out of question!

A line perpendicular to $ux + vy = w$ is given by $vx - uy = k$ where, as before, k is to be chosen by any further conditions. For proof, it is easy to verify the slopes for the two equations ($-\frac{u}{v}$ and $\frac{v}{u}$ respectively) and verify that their product is indeed -1.[7]
Example. Find a line perpendicular to $8x + y = 13$ which passes through the point $(3, 10)$.
Answer: We know that the answer can be of the form $x - 8y = k$. Impose the condition of passing through $(3, 10)$, i.e.

$$3 - 8(10) = k.$$

So we deduce that $k = -77$ and hence the answer is

$$x - 8y = -77 \text{ or } 8y - x = 77.$$

5.4 Examples of equations of lines.

1. **Intersecting Lines.** Find the point of intersection of the line L given by parametric equations:

$$L : x = 2 + 3t,\ y = -1 + 2t$$

with the line M given by a usual equation:

$$M : 4x - 5y = 7.$$

Answer: Note that at a common point, we can substitute the parametric equations of L into the usual equation of M to get:

$$4(2 + 3t) - 5(-1 + 2t) = 7 \text{ or } (12 - 10)t + 8 + 5 = 7.$$

Simplification leads to:

$$2t = 7 - 13 = -6 \text{ or } t = -3.$$

[7] An alert reader might worry about getting different looking expressions, depending on the starting equation. Thus, for a starting equation of $2x + 3y = 1$, we can have $3x - 2y = k$, but $6x - 4y = k$ seems like a valid answer as well, since our original equation could have been $4x + 6y = 2$.

This is not an error. The point is that the equation of a line is never unique, since we can easily multiply it by any non zero number to get an equivalent equation.

One way of pinning down the equation of a line is to demand that it is in slope intercept form $y = mx + c$ or is a vertical line $x = p$. In this form, the coefficients are uniquely determined by the line. This is why you were always required to get the answer to this final form.

We are taking a more flexible view which helps with faster calculation, at the price of a "canonical form".

Plugging back into the equations for L, we get:

$$\text{The common point is: } x = 2 + 3(-3) = -7,\ y = -1 + 2(-3) = -7.$$

Observation. You will find that for intersecting two lines, it is best to have one in parametric form and one in usual form.

2. **Conversion between the parametric and the usual form of a line.** Find the point of intersection of the lines given by parametric equations:

$$L : x = 2 + 3t,\ y = -1 + 2t$$

and

$$N : x = 3 + 5t,\ y = 1 + 4t.$$

Answer. We could try to equate the corresponding x, y parameterizations, but it would be a mistake to equate them using the same parameter t. After all, at a common point, we could have two different values of the parameter, depending on the line being used!

So one way is to solve the two equations:

$$2 + 3t = 3 + 5s \text{ and } -1 + 2t = 1 + 4s.$$

These are two linear equations in two variables, and are easily solved. Using the values of the parameter, we can read of the points from the parametric equations of either line.

But we recommend the following trick! Convert the second parametric equations to the usual form by eliminating t.

Thus, from the equations $x = 3 + 5t,\ y = 1 + 4t$ we see that

$$4x - 5y = 4(3 + 5t) - 5(1 + 4t) = 12 - 5 = 7.$$

So we discover that the second line N is the same as the old line M of the above problem.

Now intersect $L, N = M$ as before to find the common point $(-7, -7)$ with the parameter $t = -3$ on line L. We did not find the parameter value on the line N, but can find it if challenged! On the line N at the common point, we have $-7 = 3 + 5t$ so $t = \frac{-10}{5} = -2$.

The moral is that when intersecting two lines, try to put one in parametric form while the other is in usual form. In fact, this is exactly what we do when we solve one equation for y and plug into the other; we just did not mention the word parameter when we learned this.

For example, if we solve $y + 2x = 3$ for y, we are making a parametric form by setting $x = x, y = 3 - 2x$, i.e. we are using the letter x itself as a parameter name!

3. **Points equidistant from two given points.** Suppose that you are given two points $A(a_1, b_1)$ and $B(a_2, b_2)$. Determine the equation satisfied by the coordinates of a point $P(x, y)$ equidistant from both.

   Answer: We use the distance formula to write the equality of the square of distances from P to each of A, B.

$$(x - a_1)^2 + (y - b_1)^2 = (x - a_2)^2 + (y - b_2)^2.$$

   Now expand to get:

$$x^2 + y^2 - (2a_1x + 2b_1y) + a_1^2 + b_1^2 = x^2 + y^2 - (2a_2x + 2b_2y) + a_2^2 + b_2^2.$$

   Simplification leads to:

$$x(a_1 - a_2) + y(b_1 - b_2) = (1/2)\left((a_1^2 - a_2^2) + (b_1^2 - b_2^2)\right),$$

   where we have cancelled the terms x^2, y^2 and then thrown away a factor of -2 before collecting terms.

   This is clearly a line and it is easy to check that it passes through the midpoint of A, B, namely, $(\frac{a_1 + a_2}{2}, \frac{b_1 + b_2}{2})$.

   The reader can also see that the slope of the line is $-\frac{a_1 - a_2}{b_1 - b_2}$.

   Since the slope of the line joining A, B is $\frac{b_2 - b_1}{a_2 - a_1}$, it is obviously perpendicular to the line joining A, B.

   Thus, we have proved that the set of points equidistant from two given points A, B is the perpendicular bisector of the segment joining them.

   Often, it is easy to use this description rather than develop the equation directly. Thus the locus of points equidistant from $A(1, 3), B(3, 9)$ is easily found thus:

   The slope of the given line is $\frac{9-3}{3-1} = 6/2 = 3$. The midpoint is

$$\left(\frac{(1+3)}{2}, \frac{(3+9)}{2}\right) = (2, 6).$$

   Therefore, perpendicular line has slope $-1/3$ and hence the answer using the point-slope form is:

$$(y - 6) = -(1/3)(x - 2) \text{ or } y = -x/3 + 20/3.$$

4. **Right angle triangles.** Given three points A, B, C, they form a right angle triangle if the Pythagorean theorem holds for the distances or equivalently two of the sides are perpendicular.

   Consider the following

   Problem:

   Say, we are given points $A(2,5), B(6,4)$ and wish **to choose** a third point C somewhere on the line $x = 4$ such that ABC is **a right angle triangle.**

   How shall we find the point C?

   Answer: We can assume that the point C has coordinates $(4, t)$ for some t. The right angle can occur at A or B or C. We handle each case separately.

Case A **Suppose that we have a right angle at A.** Then then the slope of the line AC times the slope of the line AB must be -1. [8] This gives:

$$\frac{t-5}{4-2} \cdot \frac{4-5}{6-2} = -1 \text{ i.e. } t - 5 = (-1)(4)(2)/(-1) = 8.$$

This gives $t = 13$ and C is at $(4, 13)$. This is marked as C in the picture below.

Case B **Suppose that we have a right angle at B.** Then our equation becomes:

$$\frac{t-4}{4-6} \cdot \frac{4-5}{6-2} = -1 \text{ i.e. } t - 4 = (-1)(4)(-2)/(-1) = -8.$$

So we get C at $(4, -4)$. This is noted as P in the picture below.

Case C **Suppose that we have a right angle at C.**

Then we have the equation:

$$\frac{t-5}{4-2} \cdot \frac{t-4}{4-6} = -1 \text{ or } (t-5)(t-4) = (-1)(2)(-2) \text{ i.e. } t^2 - 9t + 20 = 4.$$

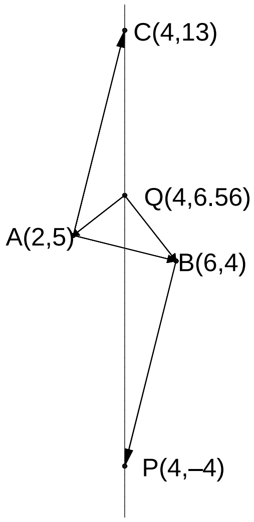
This leads to two solutions $\frac{9 \pm \sqrt{17}}{2}$.

To avoid cluttering our picture, we have shown only one of these and have marked it as a Q and we have noted a decimal approximation to its coordinate.

[8]The alert reader has to observe our convention that when we say that the product of slopes of perpendicular lines is -1, we have to make a special exception when one slope is 0 (horizontal line) and the other is ∞ (vertical line). For our problem, clearly neither slope has a chance of becoming infinite, so it is correct to equate the product of slopes to -1 and finish the solution.

In general, it is possible to set up such an equation and solve it, as long as we are sensitive to cases where some denominator or numerator becomes zero.

You should figure out where the other solution lies.



Food for thought. Should we always get four possible solutions, no matter what line we choose to lay our point C on? It is interesting to experiment. Can you think of examples where no solution is possible?

Chapter 6

Special study of Linear and Quadratic Polynomials.

We now pay special attention to the polynomials which occur most often in applications, namely the
linear polynomials ($ax + b$ with $a \neq 0$) and the
quadratic polynomials ($ax^2 + bx + c$ with $a \neq 0$).

6.1 Linear Polynomials.

First, as an example, consider $L(x) = 3x - 5$. This can be rearranged as

$$L(x) = 3(x - 5/3).$$

At $\mathbf{x = 5/3}$ we have $L(x) = 0$, so the expression **is zero! If $\mathbf{x > 5/3}$** then $L(x) = 3(x - 5/3) > 0$, so the expression **is positive**, since it is a product of positive factors. Similarly, **if $\mathbf{x < 5/3}$,** then $L(x) = 3(x - 5/3)$ **is negative** since it is a product of a positive and a negative number.
Thus, we see that the linear expression $L(x)$ is zero at a single point, positive on one side of it and negative on the other.
If the expression were $-3x + 5$ then the argument will be similar, except now it will be positive when $x < 5/3$ and negative when $x > 5/3$.
We summarize these arguments below.
The linear polynomials $L(x) = ax + b$ can be written as $a(x + b/a)$ and their behavior is very easy to describe.

1. **Case** $a > 0$

   If x has value less than $-b/a$ then $x + b/a$ is negative and hence $L(x) = a(x + b/a)$ is also negative. At $x = -b/a$, the expression becomes 0 and it becomes positive thereafter. **The expression steadily increases**

from large negative values to large positive values, hitting the zero exactly once!

2. **Case** $a < 0$

   This is just like above, except for $x < -b/a$ the value of $a(x + b/a)$ now becomes positive, it becomes 0 at $x = -b/a$ just like before and then becomes negative for $x > -b/a$. Thus, **the expression steadily decreases from large positive values to large negative values, hitting zero exactly once!**

3. **Recommended method.**

   In spite of the theory above, the best way to handle a concrete expression goes something like this.

   Say, we have to analyze the behavior of $L(x) = -4x + 8$. Find out where it is zero, i.e. solve $-4x + 8 = 0$. This gives $x = 8/4 = 2$. Thus it will have a certain sign for $x < 2$ and a certain sign for $x > 2$.

   How do we find these signs? Just test any convenient points.

   To analyze the cases $x < 2$, we try $x = 0$.

   The evaluation $L(0) = -4(0) + 8 = 8$ gives a positive number. So we conclude that

   $$L(x) \text{ is positive for } x < 2.$$

   Similarly, to test $x > 2$, try $x = 3$.

   Note that $L(3) = -4(3) + 8 = -4 < 0$. So we conclude that

   $$L(x) \text{ is negative for } x > 2.$$

   You may be wondering about why we bothered developing the theory if all we have to do is to check a few points.

   The reason is that without the theory, we would not know which and how many points to try and why should we trust the conclusion. If we understand the theory, we can handle more complicated expressions as well. We illustrate this with the quadratic polynomials next.

6.2 Factored Quadratic Polynomial.

Suppose that our quadratic expression $Q(x) = ax^2 + bx + c$ is written in [1]

$$\textbf{factored form } a(x-p)(x-q)$$

where $p < q$.
We describe how to make a similar analysis depending on the sign of a. First, we need a good notation to report our answers efficiently.
Interval notation.
Given real numbers $a < b$ we shall use the following standard convention. [2]

Values of x	Notation
$a < x$ and $x < b$	(a, b)
$a \leq x$ and $x < b$	$[a, b)$
$a < x$ and $x \leq b$	$(a, b]$
$a \leq x$ and $x \leq b$	$[a, b]$

A simple way to remember these is to note that we use open parenthesis when the end point is not included, but use a closed bracket if it is!
The notation is often extended to allow a to be $-\infty$ or b to be ∞. **The only warning is** that $\pm\infty$ are not real numbers and hence we never have the interval closed (i.e. the bracket symbol) at $\pm\infty$. [3]
Thus $(-\infty, 4)$ denotes all real numbers x such that $-\infty < x < 4$ and it clearly has the same meaning as $x < 4$, since the first inequality is automatically true!
As an extreme example, the interval $(-\infty, \infty)$ describes the set $\Re$ of all real numbers!
The reader should verify the following:

1. Case $a > 0$ Remember that

$$Q(x) = a(x-p)(x-q) \text{ with } p < q.$$

Then we have:

[1] In general this cannot be done with p and q real. However solving hard mathematical problems almost always turns on finding the right special cases to study for insight. It is a maxim among mathematicians that "If there is a hard problem you can't solve then there is an easy problem you can't solve".

[2] The a and b used in this definition have nothing to do with our coefficients of the quadratic! They are temporary variable names.

[3] In Calculus, we may take limit of numbers as x goes to ∞ or $-\infty$, but never set $x = \infty$ or $x = -\infty$. Algebraists and Algebraic Geometers do find a way of handling infinite points and still keeping things precise, but the calculations get a lot more elaborate!

Values of x	behavior of $Q(x)$
$(-\infty, p)$	$Q(x) > 0$
$x = p$	$Q(x) = 0$
(p, q)	$Q(x) < 0$
$x = q$	$Q(x) = 0$
(q, ∞)	$Q(x) > 0$

See the graphic display below showing the results when $a > 0$.

How do you verify something like this? Here is a sample verification.

Consider the third case of the interval (p, q). Since $p < x < q$ in this interval, we note that:

$$x - p > 0 \text{ and } q - x > 0 \text{ which means the same as } x - q < 0.$$

Since $Q(x) = (a)(x - p)(x - q)$ and since $a > 0$, we see that $Q(x)$ is a product of two positive terms a and $(x - p)$ and one negative term $(x - q)$.

Hence it must be negative!

In general, if our expression is a product of several pieces and we can figure out the sign for each piece, then we know the sign for the product!

2. Case $a < 0$

   Everything is as in the above case, except the signs of $Q(x)$ are reversed!

To simplify our discussion, let us make a few formal definitions.

Definition: Absolute Maximum value Given an expression $F(x)$ we say that it has an absolute maximum at $x = t$ if $F(t) \geq F(x)$ for every $x \in \Re$. We will say that $F(t)$ is the absolute maximum value of $F(x)$ and that it occurs at $x = t$.
We may alternatively say that $x = t$ is an absolute maximum for $F(x)$ and the absolute maximum value of $F(x)$ is $F(t)$.
We similarly make a

Definition: Absolute Minimum value Given an expression $F(x)$ we say that it has an absolute minimum at $x = t$ if $F(t) \leq F(x)$ for every $x \in \Re$. We will say that $F(t)$ is the absolute minimum value of $F(x)$ and that it occurs at $x = t$.

We may alternatively say that $x = t$ is an absolute minimum for $F(x)$ and the absolute minimum value of $F(x)$ is $F(t)$.

Definition: Absolute Extremum value A value $x = t$ is said to be an absolute extremum for $F(x)$, if $F(x)$ has either an absolute maximum or an absolute minimum at $x = t$.
Warning: Our definition of absolute max/min are designed for expressions $F(x)$ which are well defined for all real numbers x. In practice, $F(x)$ may be defined in some convenient domain and the definitions of absolute max/min are appropriately restricted to values of x in the same domain.
Thus, you should always decide what the domain under discussion is, before interpreting absolute max/min values.
Now we return to our analysis of a quadratic polynomial.
Set

$$m = \frac{(p+q)}{2} \text{ and } u = x - m.$$

Note that m is the average value of p and q and also $x = u + m$. Note the following algebraic manipulation:

$$\begin{aligned} Q(x) &= a(u + \tfrac{p+q}{2} - p)(u + \tfrac{p+q}{2} - q) \\ &= a(u - \tfrac{p-q}{2})(u + \tfrac{p-q}{2}) \\ &= a(u^2 - T^2) \text{ where } T = \tfrac{p-q}{2} \end{aligned}$$

Since T^2 is a positive number and u^2 is always greater than or equal to zero, the smallest value that $u^2 - T^2$ can possibly have is when $u = 0$. However this happens exactly when $x = m$. This shows that $u^2 - T^2 = (x - m)^2 - T^2$ has an absolute minimum value at $x = m$. If we take x large enough then u^2 can be made as large as we want. This means that there is no upper bound to the values of $u^2 - T^2$
It follows that if $Q(x)$ that can be put in factored form $Q = (a)(x - p)(x - q)$ with **positive** a, then $Q(x)$ has an absolute minimum at $x = m$, where m is the average of p, q.
In case $Q(x) = a(x - p)(x - q)$ with a **negative** a, the situation reverses and $Q(x)$ has an absolute maximum (and no absolute minimum) at $x = m$, the average of p, q.
Thus, in either case, we can say that $Q(x)$ has an extremum at $x = m$ the average of p, q.
We shall show below that indeed, this value $x = m$ is actually $x = -b/(2a)$.
Thus **the important point is** that we need not find the values p, q to find this extremum value!
Remark. Note that for a general expression, the absolute maximum or minimum value may not exist and may occur at several values of x.
Here are some examples to clarify these ideas.

1. As already observed, a linear expression like $3x - 5$ does not have a absolute maximum or minimum; its values can be arbitrarily small or large!

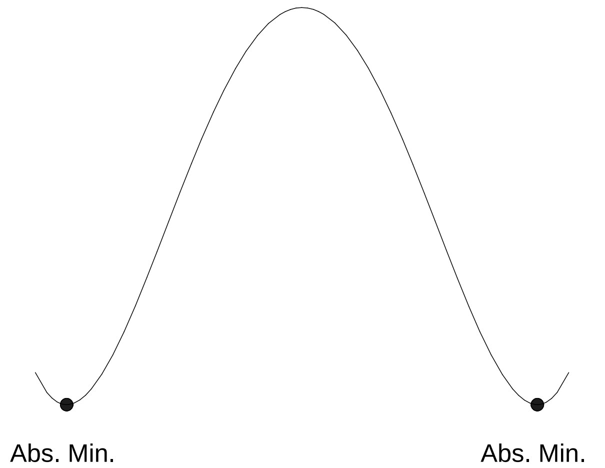
2. Consider the expression

$$F(x) = (x^2 - 1)(2x^2 - 7).$$

We invite the reader to use a graphing calculator or a computer to sketch a graph and observe that it has value $-25/8 = -3.125$ at $x = -1.5$ as well as $x = 1.5$.

It is easy to see from the graph that it has an absolute minimum value at these values of x. It is also clear that there are no absolute maximum values.

The complete argument to prove this without relying on the picture requires a further development of calculus and has to be postponed to a higher level course.



6.3 The General Quadratic Polynomial.

Recall that in the example 6 of (1.5), we proved that:

$$Q(u - \frac{b}{2a}) = au^2 - (\frac{b^2 - 4ac}{4a}) = a\left(u^2 - \frac{b^2 - 4ac}{4a^2}\right).$$

For convenience, let us set

$$u = x + \frac{b}{2a} \text{ and } H = \frac{b^2 - 4ac}{4a^2}.$$

Our equation then becomes:

$$Q(x) = a\left(u^2 - H\right)$$

It is now very easy to describe its behavior.
We have exactly three cases:

1. Case of $H < 0$ or **No real roots.** The quantity $u^2 - H$ is always positive and hence there are no real values of x which make $Q(x) = 0$.

   In this case either $a > 0$ and $Q(x)$ is always positive or $a < 0$ and $Q(x)$ is always negative.

   In case $a > 0$, the value $Q(x) = a(u^2 - H)$ is going to be the least exactly when $u = 0$, i.e. $x = u - \frac{b}{2a} = -\frac{b}{2a}$.

   Thus **the absolute minimum value is** $-aH = \frac{4ac - b^2}{4a}$ and **it is reached at** $x = -\frac{b}{2a}$.

   In case $a < 0$, a similar analysis says that **the absolute maximum value is** $-aH = \frac{4ac - b^2}{4a}$ and **it is reached at** $x = -\frac{b}{2a}$.

   Thus we have established that $\mathbf{x = -\frac{b}{2a}}$ **gives an extremum value regardless of the sign of a. Moreover, the extremum value is always** $\mathbf{\frac{4ac - b^2}{4a}}$

   Since we know about the complex numbers, we can also observe the following.

   If we solve $u^2 - H = 0$, then we get two complex solutions: $u = \pm i\sqrt{-H}$. This leads to two complex values of x, namely:

   $$x = -\frac{b}{2a} \pm i\sqrt{-H}.$$

   Note that the average of the two solutions is still:

   $$\frac{1}{2}\left((-\frac{b}{2a} + i\sqrt{-H}) + (-\frac{b}{2a} - i\sqrt{-H}) \right) = -\frac{b}{2a}.$$

   Thus, our extremum value is the average of the two complex roots of $Q(x) = 0$.

2. Case of $H = 0$ or **a double root.** In this case $Q(x) = au^2 = a(x + \frac{b}{2a})^2$ is a times a complete square and becomes zero at $x = -\frac{b}{2a}$.

   By an analysis similar to the above, we see that $\mathbf{x = -\frac{b}{2a}}$ **again gives an extremum value and the value as before is** $\mathbf{-aH = 0}$**.** Indeed, it is a minimum for $a > 0$ and maximum for $a < 0$.

3. Case of $H > 0$ or **two real roots** In this case $Q(x) = 0 = a(u^2 - H)$ leads to a pair of solutions [4]

[4] The reader should note that we have already done this case when we assumed a factored form $a(x - p)(x - q)$ for $Q(x)$. We are redoing it just to match the new notation and get the famous quadratic formula established!

$$u = \pm\sqrt{H} = \pm\frac{\sqrt{b^2-4ac}}{\sqrt{4a^2}} = \pm\frac{\sqrt{b^2-4ac}}{2|a|}.$$

Watch the absolute value sign in $|a|$. The square root of $4a^2$ is $2a$ if $a > 0$ and $-2a$ if $a < 0$. Failure to note this can lead to serious errors! However, in this case, since we have $\pm$ as a multiplier, we can replace $|2a|$ by simply $2a$.

So, the solutions for u are simply:

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

This leads to the two roots of the original equation, namely:

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

To put it in the more familiar form, we write it with a common denominator $2a$ as: **The quadratic formula for solutions of** $\mathbf{ax^2+bx+c=0, a\neq 0}$

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

If we call the two roots p, q for convenience, then they are

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

in some order and their average $\frac{p+q}{2}$ is clearly

$$\frac{1}{2}\cdot\left(\frac{-b+\sqrt{b^2-4ac}}{2a}+\frac{-b-\sqrt{b^2-4ac}}{2a}\right) = \frac{1}{2}\cdot\left(\frac{-2b}{2a}\right) = -\frac{b}{2a}.$$

4. We note that the extremum value has the property that it always occurs at the average of the two roots of $Q(x) = 0$. We have shown this when the two roots are real or complex or a single root counted twice!

   With a little thought, we can even see the following more general **Average Value Theorem.**

   Consider the equation

   $$Q(x) = k \text{ or } ax^2+bx+c = k \text{ or } ax^2+bx+(c-k) = 0.$$

   where k is any chosen number.

From the quadratic formula, we know that the sum of roots of even this equation is always $-\frac{b}{a}$ and hence the average of the roots is again $-\frac{b}{2a}$.

Thus, if we were to plot the graph of $y = Q(x) - k$, we get a parabola and its extremum value occurs at the x-coordinate of the vertex of the parabola, namely $x = -\frac{b}{2a}$.

Thus, the extremum value of the function $Q(x) - k$ is the y-coordinate of the vertex of the parabola $y = Q(x) - k$. Thus:

The extremum value of $Q(x) - k$ occurs at $x = -\dfrac{b}{2a}$ and is equal to $Q(-\dfrac{b}{2a})-k$.

This is sometimes called the **Vertex Theorem.** .

6.4 Examples of Quadratic polynomials.

We discuss a few exercises based on the above analysis.

1. Find the extremum values for $Q(x) = 2x^2 + 4x - 6$ and determine interval(s) on which $Q(x)$ is negative.

   Answer: We have $a = 2, b = 4, c = -6$. Thus $-\dfrac{b}{2a} = -\dfrac{4}{4} = -1$. So, if we replace $x = u - \dfrac{b}{2a} = u - 1$, then we have:

$$\begin{aligned} Q(x) &= 2(u-1)^2 + 4(u-1) - 6 \\ &= 2u^2 - 4u + 2 + 4u - 4 - 6 \\ &= 2u^2 - 8 = 2(u^2 - 4) \end{aligned}$$

   We have shown the simplification by hand, but if you remember the theory, you could just write down the last line from $Q(x) = a(u^2 - H)$ thus:

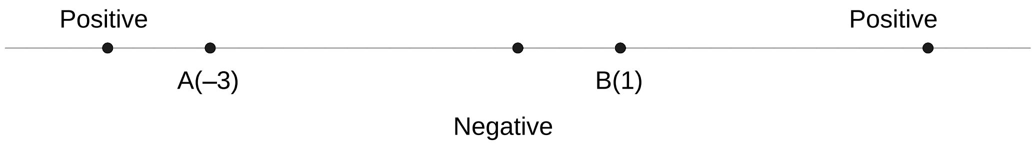
$$a = 2 \text{ and } H = \frac{b^2 - 4ac}{4a^2} = \frac{16 - 4(2)(-6)}{4(2^2)} = \frac{64}{16} = 4.$$

   At any rate, the value of $Q(x)$ is clearly bigger than or equal to $2(-4) = -8$ and this

   absolute minimum -8 is attained when $\mathbf{u = 0}$ or $\mathbf{x = u - 1 = -1}$.

   Finally, the quadratic formula (or straight factorization) will yield two roots $p = -3, q = 1$.

If you plot these two values on the number line you get three intervals. To check the sign of $Q(x)$, the easiest way is to test a few strategic points as we did in the linear case.



Intervals	$(-\infty,-3)$	-3	$(-3,1)$	1	$(1,\infty)$
Test points	-4	-3	0	1	4
Values	10	0	-6	0	42
Conclusion	Positive	Zero	Negative	Zero	Positive

Thus, **the expression is negative exactly on the interval** $(-3,1)$.

2. Find the extremum values for $Q(x) = 2x^2 + 4x + 6$ and determine interval(s) on which $Q(x)$ is negative.

   Answer: Now we have $a = 2, b = 4, c = 6$, so

$$H = \frac{b^2 - 4ac}{4a^2} = \frac{16 - 4(2)(6)}{4(4)} = -2.$$

This is negative, so the theory tells us that $Q(x)$ will have its extremum at $x = -\dfrac{b}{2a} = -1$ and this extremum value of $Q(x)$ is $2(-1)^2 + 4(-1) + 6 = 4$. The theory says that the expression will always keep the same positive sign and so 4 is actually the **minimum** value and there is no maximum.

Hence **there are no points** where $Q(x) < 0$.

Chapter 7

Functions

7.1 Plane algebraic curves

The simplest notion of a plane algebraic curve is the set of all points satisfying a given *polynomial equation* $f(x, y) = 0$.
We have seen that a parametric form of the description of a curve is useful for generating lots of points of the curve and to understand its nature.
We have already learned that all points of a line can be alternately described by linear parameterization $x = a + ut, y = b + vt$ for some suitable constants a, b, u, v where at least one of u, v is non zero.
In case of a circle (or, more generally a conic) we will find such a parameterization, but with rational functions of the parameter t, rather than polynomial functions.

Definition: A rational Curve. We say that a plane algebraic curve $f(x, y) = 0$ is rational if it can be parameterized by rational functions.
This means that we can find two rational functions $x = u(t), y = v(t)$ such that at least one of $u(t), v(t)$ is not a constant and $f(u(t), v(t))$ is identically zero.
Actually, to be precise, we can only demand that $f(u(t), v(t))$ is zero for all values of t for which $u(t), v(t)$ are both defined.
Examples. Here are some examples of rational algebraic curves. You are advised to verify the definition as needed.

1. **A line.** The line $y = 3x + 5$ can be parameterized as $x = t - 2, y = 3t - 1$. Check:
$$(3t - 1) = 3(t - 2) + 5.$$
2. **A Parabola.** The parabola $y = 4x^2 + 1$ can be parameterized by $x = t, y = 4t^2 + 1$.
3. **A Circle.** The circle $x^2 + y^2 = 1$ can be parameterized by
$$x = \frac{1 - t^2}{1 + t^2}, y = \frac{2t}{1 + t^2}.$$

You need to verify the identity:

$$(\frac{1-t^2}{1+t^2})^2 + (\frac{2t}{1+t^2})^2 = 1$$

for all values of t for which $1+t^2 \neq 0$. Since we are working with real numbers as values, this includes all real numbers.

4. **A Singular Curve.** The curve $y^2 = x^3$ is easily parameterized by $x = t^2, y = t^3$. Note that a very similar curve $y^2 = x^3 - 1$ is declared not to be parameterizable below.

5. **A Complicated Curve.** Verify that the curve:

$$xy^3 - x^2y^2 + 2\,xy \;=\; 1$$

can be parameterized by:

$$x = \frac{t^3}{t^2-1}, y = \frac{t^2-1}{t}.$$

This is simply an exercise in substitution. The process of deciding when a curve can be parameterized and then the finding of the parameterization, both belong to higher level courses.

It is tempting to think that all plane algebraic curves could be parameterized by rational functions, but alas, this is not the case. The simplest example of a curve which cannot be parameterized by rational functions is the curve given by $y^2 = x^3 - 1$.[1]

Since we cannot hope to have all algebraic curves parameterized, we do the next best thing; restrict our attention to the ones which are! In fact, to get a more useful theory, we need to generalize the idea of an acceptable function to be used for the parameterization. We may see some instances of this later.

[1]The proof that a given curve is not rational is not elementary and requires new ideas, found in advanced books on Algebraic Geometry. Pursuing this example is a good entry point into this important branch of mathematics.

7.2 What is a function?

As we have seen above, a line $y = 3x + 4$ sets up a relation between x, y so that for every value of x there is an associated well defined value of y, namely $3x + 4$. We may give a name to this expression and write $y = f(x)$ where $f(x) = 3x + 4$.
We wish to generalize this idea.
Definition of a function. Given two variables x, y we say that y is a function of x on a chosen domain D if for every value of x in D, there is a well defined value of y. Often this is rephrased as "for every value of x in D, there is a well defined value of y" which means the same thing.
Sometimes we can give a name to the procedure or formula which sets up the value of y and write $y = f(x)$.
We often describe x as the **independent variable** and y as the **dependent variable** related to it by the function.
The chosen set of values of x, namely the set D is called the **domain** of the function. [2]
The **range** of the function is the set of all "values" of the function. Thus for the the function $y = x^2$ with domain $\Re$, the range is the set of all non negative real numbers or the interval $[0, \infty)$. Some books use the word "image" for the range.
Be aware that some books declare this range as the intended set where the y-values live; thus, for our $y = x^2$ function, the whole set of reals could be called the range. **We prefer** to call such a set "target" and reserve the term "range" for the actual set of values obtained.
In other words, a target of a function is like a wish list, while the range is the actual success, the values taken on by the function as all points of the domain are used. It is often difficult to evaluate the range.
We will not, however, worry about these technicalities and explain what we mean if a confusion is possible.

1. Polynomial functions. Let $p(x)$ be a polynomial in x and let D be the set of all real numbers $\Re$. Then $y = p(x)$ is a function of x on $D = \Re$.

   It is known that for any odd degree polynomial,the range is $\Re$, while for any even degree polynomial, the range is an interval of the form $[a, \infty)$ or $(-\infty, a]$ for some $a \in \Re$. We don't have enough tools yet to prove it.

   The polynomials $p(x) = x + 1$ and $p(x) = x^2 + 1$ respectively illustrate this.

   If we take a polynomial with real or complex coefficients we can also take a larger domain for it, namely $\mathbb{C}$ the field of complex numbers. The target is

[2] Technically, this domain must be a part of our concept. We get a different function if we use a different domain.
Thus, the function defined by $y = 3x+4$ where $D = \Re$ -the set of all reals and the function defined by $y = 3x + 4$ where the domain is $\Re_+$ -the set of positive real numbers are two different functions. **Often, this subtle point is ignored!**

then taken as the field of complex numbers.

It is an amazing fact that the range for a complex polynomial function can be completely described. If the polynomial is just a constant c (i.e. has degree zero) then the range is just the single number c, but it is always equal to $\mathbb{C}$ in all other cases. In other words, if $f(x)$ is a polynomial of degree at least 1 with complex coefficients and k is any complex number, then k is in the range of the function f or explicitly the equation $f(x) = k$ has at least one solution!

This is known as the **Fundamental Theorem of Algebra.**

2. Rational functions. Let $h(x) = \dfrac{x-1}{(x-2)(x-3)}$. Then $y = h(x)$ has a well defined value for every real value of x other than $x = 2, 3$. There is a convenient notation for this set, namely $\Re \setminus \{2, 3\}$, which is read as "reals minus the set of $2, 3$". Let $D = \Re \setminus \{2, 3\}$. Then $y = h(x)$ is a function of x on D.

3. Piecewise or step functions. Suppose that we define $f(x)$ to be the integer n where $n \le x < n + 1$. It needs some thought to figure out the meaning of this function. The reader should verify that $f(x) = 0$ if $0 \le x < 1$; $f(x) = 1$ if $1 \le x < 2$ and so on. For added understanding, verify that $f(-5.5) = -6$, $f(5.5) = 5$, $f(-23/7) = -4$ and so on.

   This particular function is useful and is denoted by the word **floor**, so instead of $f(x)$ we may write $floor(x)$.

   There is a similar function called ceiling or **ceil** for short. It is defined as $ceil(x) = n$ if $n - 1 < x \le n$.

   In general, **a step function** is a function whose domain is split into various intervals over which we can have different definitions of the function.

   A natural practical example of a step function can be something like the following:

   Shipping Charges Example.

   A company charges for shipping based on the total purchase.

   For purchases of up to \$50, there is a flat charge of \$10. For every additional purchase price of \$10 the charge increases by \$1 until the net charge is less than \$25. If the charge calculation becomes \$25 or higher, then shipping is free!

   Describe the shipping charge function $S(x)$ in terms of the purchace price x.

Purchase price	Shipping charge	Comments
$x \leq 0$	0	No sale!
$0 < x \leq 50$	10	Sale up to \$50
$50 < x < 200$	$10 + ceil(\frac{x-50}{10})$	Explanation below.
$200 \leq x$	0	Free shipping for big customers!

The only explanation needed is that for $x > 50$ the calculation of shipping charges is 10 dollars plus a dollar for each additional purchase of up to 10 dollars.

This can be analyzed thus: For $x > 50$ our calculation must be \$10 plus the extra charge for $(x - 50)$ dollars. This charge seems to be a 10-th of the extra amount, but rounded up to the next dollar. This is exactly the ceil function evaluated for $(x - 50)/10$.

This calculation becomes $25 = 10 + 15$ when $x = 200$ and $ceil((200 - 50)/10) = 15$.

Finally, according to the given rule we set $S(x) = 0$ if $x \geq 200$.

About graphing. It is often recommended and useful to make a sketch of the graph of a function. If this can be done easily, then we recommend it. The trouble with the graphical representation is that it is prone to errors of graphing as well as calculations. The above example can easily be graphed, but would need at least a dozen different pieces and then reading it would not be so easy or useful. On the other hand, curves like lines and circles and other conics can be drawn fairly easily, but if one relies on the graphs, their intersections and relative positions are easy to misinterpret.

In short, we recommend relying more on calculations and less on graphing!

4. Real life functions. Suppose that we define $T(t)$ to be the temperature of some chemical mixture at time t counted in minutes. It is understood that there is some sensor inserted in the chemical and we have recorded readings at, say 10 minute intervals for a total of 24 hours.

   Clearly, there was a definite function with well defined values for the domain $0 \leq t \leq (24)(60) = 1440$. The domain might well be valid for a bigger set.

   We have, however, no way of knowing the actual function values, without access to more readings and there is no chance of getting more readings afterwards.

   What do we do?

   Usually, we try to find a model, meaning an intelligent guess of a mathematical formula (or several formulas in steps), based on the known data and perhaps known chemical theories. We could then use Statistics to make an assertion about our model being good or bad, acceptable or unacceptable and so on.

In this elementary course, we have no intention (or tools) to go into any such analysis. In the next few sections, we explain the beginning techniques of such an analysis; at least the ones which can be analyzed by purely algebraic methods.

7.3 Modeling a function.

We assume that we are given a small number of points in the plane and we try to find a function $f(x)$ of a chosen type so that the given points are on the graph of $y = f(x)$. Then the function $f(x)$ is said to be a fit for the given (data) points. Here are the simplest techniques:

1. Linear fit. Given two points $P_1(a_1, b_1), P_2(a_2, b_2)$ we already know how to find the equation of a line joining them, namely $(y - b_1)(a_2 - a_1) = (x - a_1)(b_2 - b_1)$, thus

   $$y = \frac{b_2 - b_1}{a_2 - a_1}(x) + \frac{b_1 a_2 - a_1 b_2}{a_2 - a_1}.$$

   is the desired fit.

   Of course, we know the problem when $a_1 = a_2$. Thus, if our data points are $(2, 5)$ and $(2, 7)$, then we simply have to admit that finding a *function* $f(x)$ with $f(2) = 5$ and $f(2) = 7$ is an impossible task and no exact fitting is possible.

   Now, if there are more than two points given and we still want to make a best possible linear function fit, then there is a well known formula. A full development will take us afar, but here is a brief summary:

   Optional Explanation of General Linear Fit. Here are the details, without proof.

   Suppose that you are given a sequence of distinct data points $(a_1, b_1), \cdots (a_n, b_n)$ where $n \geq 2$, then the best fit linear function is given by $y = mx + c$ where m and c are calculated by the following procedure.

   As will become apparent below, we have to exclude the case when all the x-coordinates take on the same value, say a. In this case, the common sense answer $x = a$ is the only possible fit!

   Define and evaluate:

   $$\begin{aligned} p &= \sum a^2 = a_1^2 + \cdots + a_n^2 \ , \ q = \sum a = a_1 + \cdots + a_n \\ r &= \sum ab = a_1 b_1 + \cdots + a_n b_n \\ s &= \sum b = b_1 + \cdots b_n. \end{aligned}$$

Here the suggestive notations like $\sum a^2$ are presented as an aid to memory, the explicit formulas follow them. The greek symbol $\sum$ indicates a "sum" and the expression describes a typical term.

Our idea is to have an equation of the form

$$y = mx + c$$

which is closest the set of the given data points.

If we take a typical pair of values $x = a_i, y = b_i$ for $i = 1, 2, \cdots, n$ then we get a "wished for" equation for each:

$$b_i = ma_i + c$$

and if we add these for all values of i then we get the equation $s = mq + cn$.

If we multiply our starting wished for equation by a_i, we get

$$a_i b_i = ma_i^2 + ca_i$$

and adding these for all values of i we get $r = mp + cq$.

Thus, we may think of the equations $r = mp + cq$ and $s = mq + cn$ as the average "wish" for all the data points together.

Then we solve $mp + cq = r$ and $mq + cn = s$ for m, c. Note that the Cramer's Rule or a direct verification will give the answer:

$$m = \frac{nr - sq}{np - q^2}\ ,\ c = \frac{sp - qr}{np - q^2}.$$

It is instructive to try this out for some concrete set of points and observe that sometimes, none of the points may lie on the resulting line and the answer can be somewhat different from your intuition. However, it has a statistical and mathematical justification of being the best possible fit!

Example.

Suppose that in a certain experiment we have collected the following ten data points:

$(1, 1)$, $(2, 3)$, $(3, 1)$, $(4, 2)$, $(5, 4)$, $(6, 3)$, $(7, 4)$, $(8, 7)$, $(9, 10)$, and $(10, 8)$. Find the best fit linear function for this data.

Answer: Noting that we have a collection of points (a_1, b_1),...,(a_{10}, b_{10}), let us calculate the quantities p, q, r, and s as explained in the theory above.

$$\begin{array}{lcll}
p &=& 1+4+9+16+25+36+49+64+81+100 &= 385\\
q &=& 1+2+3+4+5+6+7+8+9+10 &= 55\\
r &=& 1+6+3+8+20+18+28+56+90+80 &= 310\\
s &=& 1+3+1+2+4+3+4+7+10+8 &= 43
\end{array}$$

Now we write a proposed linear equation $y = mx + c$ using the facts that $n = 10$, $m = \dfrac{nr - sq}{np - q^2}$, and $c = \dfrac{sp - qr}{np - q^2}$.

So we have that

$$\begin{aligned} m &= \frac{(10)(310) - (43)(55)}{(10)(385) - 55^2} &= \frac{49}{55} \\ c &= \frac{(43)(385) - (55)(310)}{(10)(385) - 55^2} &= -\frac{3}{5} \end{aligned}$$

The linear equation of best fit is $y = \frac{49}{55}x - \frac{3}{5}$. The reader should plot the ten points along with the line and see how best the fit looks.

Another Example.

Consider the case of just two data points $(2, 5), (2, 7)$ that we considered above.

What do our formulas produce as a best fit here?

Answer.

Let us calculate the quantities p, q, r, and s as explained in the theory above.

$$\begin{aligned} p &= 4 + 4 &= 8 \\ q &= 2 + 2 &= 4 \\ r &= 10 + 14 &= 24 \\ s &= 5 + 7 &= 14 \end{aligned}$$

Using our formulas with $n = 2$

$$m = \frac{nr - sq}{np - q^2} \text{ and } c = \frac{sp - qr}{np - q^2}$$

we see that the denominator $np - q^2 = (2)(8) - (4)^2 = 0$. This again says the the slope is infinite and our line is vertical.

Thus, our formulas fail and we must declare the best fit as the vertical line $x = 2$.

What is a way out? Perhaps, we should interchange our notion of who the x and y are.

Thus, interchanging them, we may think of our data points as $(5, 2), (7, 2)$. What fit shall we get?

Let us calculate the quantities p, q, r, and s with the new data points.

$$\begin{aligned} p &= 25 + 49 &= 74 \\ q &= 5 + 7 &= 12 \\ r &= 10 + 14 &= 24 \\ s &= 2 + 2 &= 4 \end{aligned}$$

It is easy to check that the formulas now give $m = 0$, $c = 2$, so the new line is $y = 2$, as expected!

We leave the reader with a

Challenge:

Consider data points $(a_1, b_1), \cdots, (a_n, b_n)$. Define the quantities p, q as above.

It is easy to see that if we take all a_i equal to a common value a, then the denominator $np - q^2$ in our fitting formulas will vanish. Also, note that this denominator is independent of the b_i's.

It is also clear that in this situation, we expect the fitting line to be the vertical $x = a$.

Prove that

$$np - q^2 = 0 \text{ if an only if } a_1 = a_2 = \cdots = a_n.$$

This ends the optional explanation of the General Linear Fit.

2. **Quadratic fit.** In general three given points will not lie on any common line. So, for a true fit, we need to use a more general function. We use a quadratic fit or a quadratic function $y = q(x) = px^2 + qx + r$. We simply plug in the given three points and get three equations in the three unknowns p, q, r. We then proceed to solve them.

   Example: Find the quadratic fit for the three points $(2, 5)$, $(-1, 1)$ and $(3, 7)$.

   Answer: Let the quadratic function be

   $$y = f(x) = px^2 + qx + r.$$

   Don't confuse these p, q, r with the expressions used above! They are new notations, just for this example.

   Since we know that $f(2) = 5$, $f(-1) = 1$, and $f(3) = 7$, we can write a system of three equations with three unknowns p, q, r as follows:

   $$\begin{aligned} 5 &= 4p + 2q + r \\ 1 &= p - q + r \\ 7 &= 9p + 3q + r. \end{aligned}$$

   Using methods described in previous chapters, we solve:

   $$p = \frac{1}{6}\ ,\ q = \frac{7}{6} \text{ and } r = 2.$$

   So the quadraric function that fits the three given points is

   $$y = f(x) = \frac{1}{6}x^2 + \frac{7}{6}x + 2.$$

It is good to double check that its graph indeed passes thru the three points. Thus verify that $f(2) = 5$, $f(-1) = 1$, and $f(3) = 7$.

Don't forget that if someone gives four or more points for such a problem, then usually the precise answer may not exist. We can have a statistical fitting answer like above, but we will not developing such general formulas here. Look up the topic of **Linear Regression** in appropriate sources for Statistics.

7.3.1 Inverse Functions.

Let is recall the various terms used in connection with a function.
Suppose that we have a function $f : A \to B$ where A, B are certain sets of numbers (real or complex.)
We call A the **domain** of the function and call B its **target**.
The **range** of the function is defined as the set

$$\{y \in B \mid y = f(x) \text{ for some } x \in A\}.$$

In general, determining the range of a function involves extra calculations.
We say that the function f is **"one to one"** if different values of x give different values of $f(x)$.
This can be reworded as

One to one property. If $f(x_1) = f(x_2)$ then, $x_1 = x_2$

We say that the function f is **onto B** if B is the same as the range. The set B is not mentioned, if clear from the context.
Explicitly, this means:

Onto property. For every $y \in B$ there is some $x \in A$ such that $f(x) = y$

A function $g : B \to A$ is said to be the **inverse of the function** f if it satisfies:

Inverse conditions. $f(g(y)) = y$ for all $y \in B$ and $g(f(x)) = x$ for all $x \in A$

[3]

Finally, we have the **test for existence of an inverse function.**
A function $f : A \to B$ has an inverse if and only if f is both **one to one and onto**.
[4]

If a graph of a real $y = f(x)$ available, then there is a graphical way of deciding its "one to one" and "onto" properties.
If we assume that both the target and the domain of the function are the same $\Re$-the set of all real numbers, then the simple test is this:

[3]Why do we use two different letters x, y in this definition? It is meant as a help in calculation. We are thinking of the function as $y = f(x)$, so elements of the domain are denoted by x and elements of the target are denoted by y. Often both the domain and the target are the same, but this convention keeps us thinking of the correct set.

[4]Here is a sketch of the proof of the test.

- The function is onto a set $A \subset \Re$ if and only for every $a \in A$, the horizontal line $y = a$ meets the graph at least once.

- The function is one to one if and only for every $b \in \Re$, the horizontal line $y = b$ meets the graph at most once.

Here is the modification of the test for more special targets and domains.
If the domain is a set D smaller than $\Re$, then in the " one to one " test, we require a horizontal line to meet the graph at most once over the domain D.
Similarly, if the target is smaller than $\Re$ then, the horizontal lines $y = b$ need to be chosen only with b in the target.

Special notation for inverse function.

Suppose $f : A \to B$ is a given function which has an inverse, say g. Then it is customary to denote g by the symbol f^{-1}. Thus, we have

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

This notation can lead to confusion and one needs to be careful with it. Thus $f^{-1}(x)$ is **not the same** as $f(x)^{-1} = \dfrac{1}{f(x)}$. Indeed, it is usually very different from it.
We will avoid this notation, unless dictated by tradition.
Let us illustrate the above concepts by a few examples.

1. **Analyzing a real function.** Consider $f : \Re \to \Re$ defined by $f(x) = x^2$.

    - Determine the range of f.
    - Determine if the function f is one to one.
    - Determine the inverse of f, if it exists.

    Answer.

    Set $y = f(x) = x^2$. We want to find those values of y for which we can solve the equation $y = x^2$. We already know that the solution of this equation is:

    $$x = \pm\sqrt{y} \text{ if } y \geq 0 \text{ and no real solution if } y < 0.$$

- Suppose that f has an inverse g. If $f(x) = y$, then applying g to both sides, we see that $g(f(x)) = g(y)$ and by definition of the inverse function, we get $x = g(y)$.

  Now, if $f(x_1) = f(x_2) = y$, then we see that $x_1 = g(y) = x_2$, and this proves f is one to one.

  Also, given any $y \in B$ we see by definition that $y = f(g(y))$ so y is the image of $g(y)$ by f. This proves that f is onto B.

- Conversely, if f is onto B, we see that for every $y \in B$, there is at least one $x \in A$ such that $y = f(x)$. Moreover, by one to one property, there is a single x such that $y = f(x)$. We define $g(y) = x$ and verify the two conditions easily.

Thus the range is

$$\Re_+ = \{y \in \Re \mid y \geq 0\} = [0, \infty).$$

Since both $x = 1$ and $x = -1$ lead to $y = f(x) = 1$, we see that the function is not one to one.

By the known test for inverse functions, the inverse function does not exist!

2. **Analyzing a complex function.** Consider $f : \mathbb{C} \to \mathbb{C}$ defined by $f(x) = x^2$.

   - Determine the range of f.
   - Determine if the function f is one to one.
   - Determine the inverse of f, if it exists.

   Answer. When working with complex numbers, we see that the equation $y = x^2$ always has solutions for any y.

   Indeed, if $y = 0$ then take $x = 0$. If $y \neq 0$, then write $y = r\exp(i\theta)$ using Euler's representation, where $r = |y| > 0$ and θ is some real number which is its argument $Arg(y)$.

   Then it is easy to show that $x_1 = r\exp(i\theta/2)$ and $x_2 = r\exp(i(\theta + \pi)/2)$ satisfy $f(x_1) = f(x_2) = y$. Thus the function is onto but not one to one.

   In particular it does not have an inverse.

3. **Changing the domain or the target.** We remark that a function can be made "onto" by trimming its target down to its range. We also note that a function can be made "one to one" by trimming down its domain, but this process can be rather complicated and may lead to unwieldy domains. Here are a few examples.

   The same real function $y = x^2$ is onto if we set our target as $\Re_+$ the set of non negative reals. If we consider the same formula but take our domain and target as both $\Re_+$, then it is one to one and onto. Its inverse is given by $g(x) = \sqrt{x}$ which is defined on the domain $\Re_+$ and by $\sqrt{x}$, we mean the non negative square root.

4. **Finding the inverse.** In general, for a complicated function, verifying the conditions for an inverse is difficult. You will see some of the hard calculations in the Appendix.

   There is a simple procedure which may sometimes succeed and we describe it next.

Given a function $f(x)$, try to solve the equation $y = f(x)$ for x. The range would be those values of y for which the equation has a solution and the function would be one to one if the solution is unique.

Thus, the solution, written as $x = g(y)$ would give the necessary inverse function, namely $g(x)$ (i.e. the function obtained by replacing y by x in $g(y)$).

Here is an example of a function $f(x)$ for which this simple idea works.

Consider $f(x) = 3x + 5$. Set $y = 3x + 5$ and solve for x. This gives $x = \dfrac{y-5}{3}$. This led to a well defined answer, so the inverse function exists. It is defined so that $g(y) = x$, i.e. $g(y) = \dfrac{y-5}{3}$.

Now we change the variable to x, and write $g(x) = \dfrac{x-5}{3}$.

In general, if $y = f(x) = ax + b$ where $a \neq 0$, then the the inverse function is given by $x = \dfrac{y-b}{a}$. Thus $g(y) = \dfrac{y-b}{a}$ or $g(x) = \dfrac{x-b}{a}$.

5. **Inverse functions on a calculator.** Many routine functions appear to have inverse keys on a calculator. The existence of the keys does not mean the functions have an inverse. It only means that under a suitably chosen domain and target they have an inverse known to the calculator.

   The simplest example of this is the square root function.

   Recall that the square root function is defined as the inverse of the square function which may be defined as

$$f := \Re \to \Re \text{ where } f(x) = x^2.$$

   It is known that the range is the set of non negative reals and the function is not one to one. (For example, $f(2) = f(-2) = 4$.)

   The calculator will declare an error if you ask for $\sqrt{x}$ when x is outside the range , i.e. $x < 0$.

   It will, however, give $\sqrt{4} = 2$, not bothering to remind you that -2 is also a possible number whose square is 4. Thus, it is acting as if the domain of the function f is also the set of non negative reals.

   It is your task to think about which solution of $f(x) = 4$ is needed. This problem becomes more acute when trigonometric functions are used, since they have many values of x leading to the same y.

Chapter 8

The Circle

8.1 Circle Basics.

A circle is one of the most studied geometric figures. It is defined as the locus of all points which keep a fixed distance r (called the radius) from a given point $P(h,k)$ called the center. [1]

The distance formula leads to a simple equation: $\sqrt{(x-h)^2+(y-k)^2}=r$, but since it is not easy to work with the square root, we square the equation and simplify:

$$\text{Basic Circle } (x-h)^2+(y-k)^2=r^2 \text{ or } x^2+y^2-2hx-2ky=r^2-h^2-k^2.$$

For example, the circle with center $(2,3)$ and radius 5 is given as:

$$(x-2)^2+(y-3)^2=5^2 \text{ or } x^2-4x+4+y^2-6y+9=25.$$

The equation is usually rearranged as:

$$x^2+y^2-4x-6y=12.$$

Recognizing a circle.

We now show how to guess the center and the radius if we see the rearranged form of the equation.

Assume that the equation of a circle appears as: $x^2+y^2+ux+vy=w$. Compare with the basic circle equation and note:

$$u=-2h\ ,\ v=-2k\ ,\ w=r^2-(h^2+k^2).$$

[1]We used the traditional word locus which is a dynamic idea - it signifies the path traced by a point moving according to the given rules; but in practice it is equivalent to the word set instead. Somehow, the locus conveys the ideas about the direction in which the points are traced and may include the idea of retracing portions of the curve.

This gives us that the center must be $(h, k) = (-u/2, -v/2)$.
Also, $r^2 = w + (h^2 + k^2)$, so $r^2 = w + \frac{u^2}{4} + \frac{v^2}{4}$ and hence the radius must be $r = \sqrt{w + \frac{u^2}{4} + \frac{v^2}{4}}$.
Thus, for example, if we have a circle given as:

$$x^2 + y^2 + 4x - 10y = 20$$

then we have:

$$u = 4, \ v = -10, \ w = 20.$$

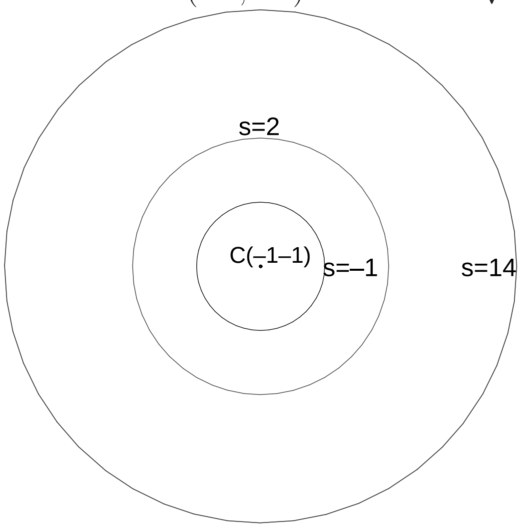
Thus, we get:

$$\text{Center is: } (-u/2, -v/2) = (-\frac{4}{2}, -\frac{-10}{2}) = (-2, 5).$$

Also,

$$\text{The radius is: } \sqrt{w + \frac{u^2}{4} + \frac{v^2}{4}} = \sqrt{20 + \frac{(4)^2}{4} + \frac{(-10)^2}{4}} = \sqrt{20 + 4 + 25} = \sqrt{49} = 7.$$

What happens if the quantity $w + \frac{u^2}{4} + \frac{v^2}{4}$ is negative or zero? We get a circle which has no points (i.e. is imaginary) or reduces to a single point $(-u/2, -v/2)$. [2]
Consider, for example the equation $x^2 + y^2 + 2x + 2y = k$ and argue that for $k < -2$ we get a circle with no point, for $k = -2$ we have just the point $C(-1, -1)$ and for $k > -2$ we get a circle centered at $C(-1, -1)$ with radius $\sqrt{k + 2}$.



[2]Of course, with our knowledge of complex numbers, the circle is never really empty, it may only have a limited number of real points! Even the circle $x^2 + y^2 = -1$ has the point $(0, i)$ and then, if we follow our parameterization given below, it would have an infinitely many complex points, none of them real!

8.2 Parametric form of a circle.

For the equation of a line, say something like $y = mx + c$, it is easy to get a parametric form, say $x = t, y = mt + c$ and thus it is easy to create many points on the line by choosing different values of t.
This is not so easy for the complicated equation of a circle, say something like

$$x^2 + y^2 = 12.$$

Here, if we try to set $x = t$, the value of y is not unique and we need to solve a quadratic equation $y^2 = 12 - t^2$ to finish the task.
We now construct a parametric form for the circle which also illustrates a new technique of analyzing an equation.
Consider the general equation of a circle centered at $(0, 0)$ and having radius $r > 0$, given by

$$x^2 + y^2 = r^2.$$

We shall show below that every point $P(x, y)$ of this circle can be uniquely described as

$$x = r\frac{1 - m^2}{1 + m^2} \text{ and } y = r\frac{2m}{1 + m^2}.$$

Here m is a parameter which is allowed to be any real number as well as ∞ and has a well defined geometric meaning as the slope of the line joining joining the point (x, y) to the special point $A(-r, 0)$.

Proof of the parameterization of a circle.

The following derivation of this parameterization can be safely omitted in a first reading.
Note that $A(-r, 0)$ is clearly a point on our circle. Let us take a line of slope m through A in a parametric form, say $x = -r + t, y = 0 + mt$. Let us find where it hits the circle by plugging into the equation of the circle.
We get the following sequence of simplifications, to be verified by the reader.

$$\begin{array}{lcl}
(t-r)^2 + m^2t^2 & = & r^2 \\
t^2 - 2rt + r^2 + m^2t^2 & = & r^2 \\
(1+m^2)t^2 - 2rt & = & 0 \\
t\left((1+m^2)t - 2r\right) & = & 0
\end{array}$$

This gives two easy solutions, $t = 0$ or $t = \dfrac{2r}{(1 + m^2)}$.
Clearly $t = 0$ gives the point A and the other value gives the other intersection of the circle with the line.

This other point has

$$x = -r + t = -r + \frac{2r}{(1+m^2)} = r\frac{1-m^2}{1+m^2}$$

and

$$y = mt = m\frac{2r}{1+m^2} = r\frac{2m}{1+m^2}.$$

Thus:

$$(x, y) = \left(r\frac{1-m^2}{1+m^2} \ , \ r\frac{2m}{1+m^2} \right).$$

This gives a nice parametric form for the circle with parameter m. As we take different values of m we get various points of the circle.
What happens to the limiting case when m is sent to infinity. Indeed for this case, the line should be replaced by the vertical line $x = -r$ and the equation of substitution gives us:

$$r^2 + y^2 = r^2 \text{ which leads to } y^2 = 0.$$

This says that the only point of intersection is $A(-r, 0)$, so it makes sense to declare that the limiting value of the parameterization gives $A(-r, 0)$.
There is an interesting algebraic manipulation of the formula which helps in such limiting calculations often. We present it for a better understanding.
We rewrite the expression $r\frac{1-m^2}{1+m^2}$ by dividing both the numerator and the denominator by m^2:

$$r\frac{1-m^2}{1+m^2} = r\frac{\frac{1}{m^2} - 1}{\frac{1}{m^2} + 1}$$

Now with the understanding that $1/\infty$ goes to 0, we can deduce the result to be

$$r\frac{-1}{1} = -r.$$

A similar calculation leads to:

$$r\frac{2m}{1+m^2} = r\frac{\frac{2}{m}}{\frac{1}{m^2} + 1} \to r\frac{0}{1} = 0.$$

Remark.
The above rational parameterization of a circle is indeed nice and useful, but not as well known as the trigonometric one that we present later in greater detail. It was already introduced indirectly as Euler's representation.
Even though rational functions are easier to work with, the trigonometric functions have a certain intrinsic beauty and have been well understood over a period of well over two thousand years.

8.3 Application to Pythagorean Triples.

The Simple Identity: $(a-b)^2 + 4ab = (a+b)^2$

has been used for many clever applications.
We now show how a clever manipulation of it gives us yet another proof of the circle parameterization.
First we divide it by $(a+b)^2$ to get

$$\frac{(a-b)^2}{(a+b)^2} + \frac{4ab}{(a+b)^2} = 1.$$

Replacing a by 1 and b by m^2 we see that

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

We now show how it can lead to a new explanation of the parameterization of a circle.
Consider, as above the circle

$$x^2 + y^2 = r^2.$$

Rewrite it after dividing both sides by r^2 to get:

$$\left(\frac{x}{r}\right)^2 + \left(\frac{y}{r}\right)^2 = 1.$$

Using our identity, we could simply set:

$$\frac{x}{r} = \frac{1-m^2}{1+m^2} \text{ and } \frac{y}{r} = \frac{2m}{1+m^2}$$

and clearly our circle equation is satisfied!
It follows that

$$x = r\frac{1-m^2}{1+m^2} \ , \ y = r\frac{2m}{1+m^2}.$$

is a parameterization! [3]
We now use the above to discuss:
Pythagorean triples.
A Pythagorean triple is a triple of integers x, y, z such that

$$x^2 + y^2 = z^2.$$

[3] What is not clear is that every point of the circle corresponds to one and only one value of the parameter m, provided we make some reasonable assignment for $m = \infty$.
This can be deduced, but does not follow immediately from the equation being satisfied.

For convenience, one often requires all numbers to be positive. If x, y, z don't have a common factor, then the triple is said to be **primitive.**
The same original (simple) identity can also be used to generate lots of examples of Pythagorean triples.
We divide the original identity by 4 and plug in $a = s^2, b = t^2$. Thus, we have:

$$\frac{(s^2 - t^2)^2}{4} + s^2t^2 = \frac{(s^2 + t^2)}{4}$$

and this simplifies to:

$$\left(\frac{s^2 - t^2}{2}\right)^2 + (st)^2 = \left(\frac{s^2 + t^2}{2}\right)^2.$$

Then we easily see that

$$x = \frac{s^2 - t^2}{2}\ ,\ y = st\ ,\ z = \frac{s^2 + t^2}{2}$$

forms a Pythagorean triple!
What is more exciting is that with a little number theoretic argument, this can be shown to give all possible primitive positive integer solutions (x, y, z). [4]
These primitive Pythagorean triples have been rediscovered often, in spite of their knowledge for almost 2000 years. It is apparent that they have been known around the world for a long time.
Try generating some triples by taking suitable values for s, t. (For best results, take odd values $s > t$ without common factors!) For instance, we can take:

s	t	x	y	z
3	1	4	3	5
5	1	12	5	13
5	3	8	15	17
7	5	12	35	37
7	3	20	21	29
7	1	24	7	25

What is the moral? A good identity goes a long way!

Optional section: A brief discussion of the General conic.

A conic is a plane curve which is obtained by intersecting a cone with planes in different position. Study of conic sections was an important part of Greek Geometry and several interesting and intriguing properties of conics have been developed over two thousand years.

[4] We leave the proof as a challenge to the reader.

In our algebraic viewpoint, a conic can be described as any degree two curve, i.e. a curve of the form

$$\textbf{General conic: } ax^2 + by^2 + 2hxy + 2fx + 2gy + c = 0.$$

We invite the reader to show that the above idea of parameterizing a circle works just as well for a conic. We simply take a point $A(p, q)$ on the conic and consider a line $x = p + t, y = q + mt$. The intersection of this line with the conic gives two value of t, namely

$$t = 0\ ,\ t = \frac{-2(bqm + f + gm + ap + hpm + hq)}{a + 2hm + bm^2}.$$

The reader is invited to carry out this easy but messy calculation as an exercise in algebra.
It is interesting to work out the different types of the conics: circle/ellipse, parabola, hyperbola and a pair of lines. We can work out how the equations can tell us the type of the conic and its properties. But a detailed analysis will take too long. Here we simply illustrate how the idea of parameterization can be worked out for a given equation.
Example of parameterization. Parameterize the parabola

$$y^2 = 2x + 2$$

using the point $A(1, 2)$.
Answer: We shall first change coordinates to bring the point A to the origin. This is not needed, but does simplify our work.
Thus we set $x = u + 1, y = v + 2$, then the point $A(1, 2)$ gets new coordinates $u = 0$, $v = 0$. Our equation becomes:

$$(v + 2)^2 = 2(u + 1) + 2 \text{ or } v^2 + 4v = 2u.$$

Now a line through the origin in (u, v) coordinates is given by $u = t, v = mt$ where m is the slope.
Substitution gives

$$m^2t^2 + 4mt = 2t \text{ or, by simplifying } t(m^2t + (4m - 2)) = 0.$$

Thus there are two solutions $t = 0, t = \dfrac{2 - 4m}{m^2}$. We ignore the known solution $t = 0$ corresponding to our starting point A.
Note that when $m = 0$, we end up with a single solution $t = 0$.
We use the second solution for the parameterization:

$$u = t = \frac{2 - 4m}{m^2} \text{ and } v = mt = \frac{2m - 4m^2}{m^2}.$$

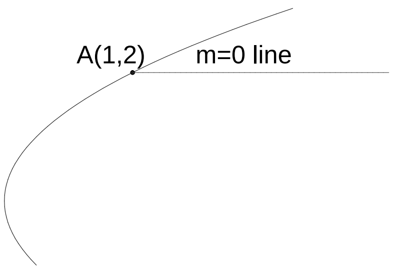
When we go back to the original coordinates, we get:

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

and

$$y = 2 + \frac{2m - 4m^2}{m^2} = \frac{2m - 2m^2}{m^2}.$$

It is an excellent idea to plug this into the original equation and verify that the equation is satisfied *for all values of* m. This means that the final equation should be free of m and always true!
Actually, the value $m = 0$ *is a problem, since it appears in the denominator. However, when plugged into the equation of the curve, the* m *just cancels! When we take the value* $m = 0$ *the other point of intersection of the line simply runs off to infinity. Indeed, this is a situation where a graph is very useful to discover that the line corresponding to* $m = 0$ *will be parallel to the axis of the parabola!*



8.4 Examples of equations of a circle.

1. Equation of a circle with given properties.

   Problem 1. Find the equation of a circle with center $(2, 3)$ and radius 6.

   Answer: We just apply the basic definition and then simplify if necessary.

   The equation must be:

   $$(x-2)^2 + (y-3)^2 = 6^2 \text{ or } x^2 - 4x + 4 + y^2 - 6y + 9 = 36.$$

   It is customary to rearrange the final equation as:

   $$x^2 + y^2 - 4x - 6y = 23.$$

   Problem 2. Find the equation of a circle with $A(1, 5)$ and $B(7, -3)$ as the ends of a diameter.

   Answer. First the long way:

   We know that the center must be the midpoint of AB so it is:

   $$\frac{A+B}{2} = (\frac{1+7}{2}, \frac{5-3}{2}) = (4, 1).$$

   Also, the radius must be half the diameter $(= d(A, B))$, i.e.

   $$\text{radius is } \frac{1}{2}\sqrt{(7-1)^2 + (-3-5)^2} = \frac{1}{2}\sqrt{36+64} = \frac{1}{2}\sqrt{100} = 5.$$

   So the answer is

   $$(x-4)^2 + (y-1)^2 = 25.$$

   Simplified, we get:

   $$x^2 + y^2 - 8x - 2y = 8.$$

   Now here is an elegant short cut. We shall prove that given two points $A(a_1, b_1)$ and $B(a_2, b_2)$ the equation of the circle with diameter AB is:

   Diameter form of circle $(x - a_1)(x - a_2) + (y - b_1)(y - b_2) = 0.$

   This formula will make a short work of the whole problem, since it gives the answer:

   $$(x-1)(x-7) + (y-5)(y+3) = 0.$$

   The reader should simplify and compare it with the earlier answer.

   We give the following hints, so the reader can prove this formula by the "duck principle".

- First expand and simplify the equation and argue that it is indeed is the equation of a circle.
- Notice the coefficients of x, y and argue that the center of the circle is indeed the midpoint of line joining A, B.
- Observe that both the points satisfy the equation trivially!
- So, it must the desired equation by the duck principle.

Extra observation. This particular form of the equation of a circle is developed for yet another purpose, which we now explain.

It is a well known property of circles that if A, B are the ends of a diameter and $P(x, y)$ is any point of the circle, then AP and BP are perpendicular to each other. Conversely, if P is any point which satisfies this condition, then it is on the circle.

In words, this is briefly described as the angle subtended by a circle in a semi circle is a right angle.

We claim that the diameter form of the equation of a circle proves all this by a simple rearrangement. Divide the equation by $(x - a_1)(x - a_2)$ to get:

$$1 + \frac{(y - b_1)(y - b_2)}{(x - a_1)(x - a_2)} = 0$$

or

$$\left(\frac{y - b_1}{x - a_1}\right)\left(\frac{y - b_2}{x - a_2}\right) = -1$$

Note that $\frac{y - b_1}{x - a_1}$ is the slope of AP and similarly $\frac{y - b_2}{x - a_2}$ is the slope of BP. So, the new equation is simply saying that the product of these slopes is -1, i.e. the lines AP and BP are perpendicular!

2. Intersection of two circles. Find all the points of intersection of the two circles:

$$x^2 + y^2 = 5 \ , \ x^2 + y^2 - 3x - y = 6.$$

Answer: Remember our basic idea of solving two equations in two variables x, y was:

- Solve one equation for one variable, say y.
- Plug the value into the second equation to get an equation in x.
- Solve the resulting equation in x and plug this answer in the first solution to get y.

But here, both equations are quadratic in x, y and we will run into somewhat unpleasant calculations with the square roots.

A better strategy is to try the elimination philosophy. It would be nice to get rid of the quadratic terms which make us solve a quadratic. We try this next:

Let us name $f = x^2 + y^2 - 5$ and $g = x^2 + y^2 - 3x - y - 6$.

Note that the quadratic y term is y^2 for both, so we get:

$$f - g = (x^2 + y^2 - 5) - (x^2 + y^2 - 3x - y + 6) = 3x + y + 1.$$

Let us name this linear expression:

$$h = 3x + y + 1.$$

We have at least gotten rid of the quadratic terms. Also, any common points satisfying $f = g = 0$ also satisfy $h = 0$. So, we could solve this $h = 0$ for y and use it to eliminate y from $f = 0$ as well as $g = 0$.

$$h = 0 \text{ gives us } y = -3x - 1.$$

Plugging this into the equation $f = 0$ we get

$$x^2 + (-1 - 3x)^2 - 5 = 0.$$

This simplifies to:

$$x^2 + 1 + 6x + 9x^2 - 5 = 0 \text{ or } 10\,x^2 + 6\,x - 4 = 0.$$

Further, it is easy to verify that it has two roots:

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

Plugging back into the expression for y, we get corresponding y values and hence the intersection points are: [5]

$$P(-1, 2) \text{ and } Q(\frac{2}{5}, -\frac{11}{5}).$$

Actually, for intersecting two circles whose equations have the same quadratic terms ($x^2 + y^2$ for example) the process of subtracting one from the other always gives a nice linear condition.

Indeed, this is a short cut to the solution of the next couple of problems .

[5] Did we forget the g altogether? Not really. If $f = 0$ and $h = f - g = 0$, then automatically, $g = 0$. So, it is enough to solve for $f = 0, h = 0$.

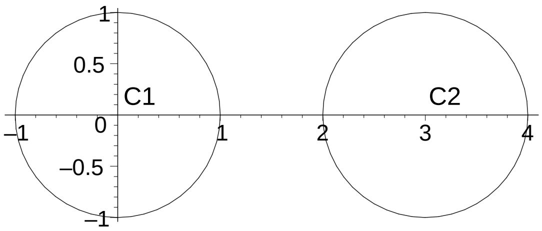
3. Intersecting circles which don't appear to meet.

   Consider circles given by

$$C_1 : \quad x^2 + y^2 = 1 \text{ and } C_2 : \quad x^2 - 6x + y^2 = -8.$$

   Find all their points of intersection.

   Answer. A simple sketch will show that these two circles have no common real point of intersection.



   Let us try to follow our method above anyway. So we set:

$$f = x^2 + y^2 - 1, g = x^2 - 6x + y^2 + 8 \text{ and } h = f - g = -1 + 6x - 8 = 6x - 9.$$

   Thus, the line $6x - 9 = 0$ or $x = \frac{3}{2}$ passes through all the common points of the two circles, whether we see them or not.

   Putting this value of x in the first equation, we get:

$$\left(\frac{3}{2}\right)^2 + y^2 = 1 \text{ or after simplification } y^2 = -\frac{5}{4}.$$

   Thus the y-coordinates are

$$y = \pm\sqrt{-\frac{5}{4}} \text{ or } \pm\frac{\sqrt{5}}{2}i.$$

   Thus, we found the two complex points $(\frac{3}{2}, \frac{\sqrt{5}}{2}i)$ and $(\frac{3}{2}, -\frac{\sqrt{5}}{2}i)$. [6]

[6] It is tempting to think that perhaps any two circles should always meet in two points. Here are two examples which show a couple of exceptions to this idea.

- Consider $x^2 + y^2 = 1, x^2 - 4x + y^2 = -3$. Observe that these meet only at $(1, 0)$ but that point could be imagined as a double intersection. This will become clearer later.
- Consider $x^2 + y^2 = 1, x^2 + y^2 = 4$. Now you cannot find any common points, real or complex. It is possible to resolve this problem by stepping into a branch of Geometry called Projective Geometry. We will not attempt it here, but invite you to study it later.

4. Line joining the intersection points of the two circles. Find the line joining the common points of the two circles:

$$x^2 + y^2 = 5 \ , \ x^2 + y^2 - 3x - y = 6.$$

Answer: We use the "duck principle". As before, we set $f = x^2 + y^2 - 5$ and $g = x^2 + y^2 - 3x - y - 6$ and note that all the common points satisfy $f - g = 3x + y + 1 = 0$.

Thus $3x + y + 1 = 0$ contains all the common points and is evidently a line, in view of its linear form; so we have found the answer!

Note that we do not need to solve for the common points at all. Moral: Avoid work if you can get the right answer without it!

Important Observation. The line $h = 0$ obtained by subtracting one standard circle equation from another actually has one interesting property. It is perpendicular to the line joining the two centers and is actually the perpendicular bisector when the circles have equal radii. This can be checked by writing general equations.

In our current example, the centers are $(0, 0)$ and $(3/2, 1/2)$ and the line $3x + y + 1 = 0$ is seen to be a perpendicular to the line of centers but not their perpendicular bisector.

Also, let us observe that in the above example where we found the complex common points of the two circles

$$C_1 : \quad x^2 + y^2 = 1 \text{ and } C_2 : \quad x^2 - 6x + y^2 = -8,$$

the line passing through their common points was $x = \frac{3}{2}$ and this is actually the perpendicular bisector of the line segment joining the two centers $(0, 0)$ and $(3, 0)$, since both circles have equal radius 1.

Thus, the line passing through the common points of circles of equal radius is always the perpendicular bisector of the line of centers, regardless of the nature of the intersection.

As a further evidence, also review that for the circles:

$$C_1 : \quad x^2 + y^2 = 1 \text{ and } C_2 : \quad x^2 - 4x + y^2 = -3$$

the perpendicular bisector is still $x = 1$ obtained by the same method and indeed passes through the common double point. The fact that this line is tangent to both the circles justifies our idea of calling it a double common point!

5. Circle through three given points. Find the equation of a circle through the given points $A(0,0), B(3,1), C(1,3)$.

   Answer: We begin by assuming that the desired equation is

$$x^2 + y^2 + ux + vy = w$$

   and write down the three conditions obtained by plugging in the given points:

$$0 + 0 + 0 + 0 = w \ , \ 9 + 1 + 3u + v = w \ , \ 1 + 9 + u + 3v = w.$$

   These simplify to:

$$w = 0 \ , \ 3u + v - w + 10 = 0 \ , \quad \text{and } u + 3v - w + 10 = 0.$$

   We note that $w = 0$ and hence the last two equations reduce to:

$$3u + v + 10 = 0 \ , \quad \text{and } u + 3v + 10 = 0.$$

   By known techniques (Cramer's Rule for example) we get their solution and conclude that

$$u = -5/2, v = -5/2, w = 0.$$

   So the equation of the desired circle is:

$$x^2 + y^2 - 5x/2 - 5y/2 = 0.$$

   From what we already know, we get that this is a circle with [7]

$$\text{center: } \left(-\frac{u}{2}, -\frac{v}{2}\right) = \left(\frac{5}{4}, \frac{5}{4}\right)$$

   and

$$\text{radius: } \sqrt{w + \frac{u^2}{4} + \frac{v^2}{4}} = \sqrt{0 + \frac{25}{16} + \frac{25}{16}}\sqrt{50/16} = (5/4)\sqrt{2}.$$

   Done!

6. Failure to find a circle through three given points. The three equations that we obtained above can fail to have a solution, exactly when the three chosen points all lie on a common line. The reader is invited to attempt to find a circle through the points:

$$(0,0), (3,1), (6,2)$$

[7] The center can also be found by using the Geometric fact that it is the intersection of the perpendicular bisectors of the line segments AB and AC. The circle itself can then be determined by finding the radius using the center and any of the given points.

and notice the contradiction in the resulting equations: [8]

$$0+0+0+0=w\ ,\ 9+1+3u+v=w\ ,\ 36+4+6u+2v=w$$

which lead to:

$$w=0\ ,\ 3u+v=-10\ ,\ 6u+2v=-40.$$

7. Smallest circle with a given center meeting a given line. Find the circle with the smallest radius which has center at $A(1,1)$ and which meets the line $2x+3y=6$.

Answer: We assume that the circle has the equation:

$$(x-1)^2+(y-1)^2=r^2 \text{ with } r \text{ to be determined.}$$

We solve the line equation for y:

$$y=(6-2x)/3=2-(2/3)x$$

and plug it into the equation of the circle:

$$(x-1)^2+(2-(2/3)y-1)^2=r^2 \text{ which simplifies to: } \frac{13}{9}x^2-\frac{10}{3}x+2=r^2.$$

Our aim is to find the smallest possible value for the left hand side, so the radius on the right hand side will become the smallest!

From our study of quadratic polynomials, we already know that ax^2+bx+c with $a>0$ has the smallest possible value when $x=-\frac{b}{2a}$.

For the quadratic in x on our left hand side, we see $a=\frac{13}{9}$ and $b=-\frac{10}{3}$ so

$$-\frac{b}{2a}=-\frac{-\frac{10}{3}}{(2)(\frac{13}{9})}=\frac{(10)(9)}{(3)(2)(13)}=\frac{15}{13}.$$

Thus, $x=\frac{15}{13}$ for the desired minimum value. Taking this value for the variable x, our left hand side becomes

$$\frac{13}{9}\left(\frac{15}{13}\right)^2-\frac{10}{3}\left(\frac{15}{13}\right)+2=\frac{1}{13}.$$

[8] If we try to find the center using the Geometric idea explained above, the perpendicular bisectors of the line segments AB and AC would be parallel and hence will not meet. Thus, we will not find a center.

Indeed, this is often turned into a precise statement by observing that a line is nothing but a circle of infinite radius with center at infinity in the direction of a perpendicular line.

Thus $\frac{1}{13} = r^2$ and hence the smallest radius is $\sqrt{1/13}$. Note that if we take $r = \sqrt{1/13}$ then we get a single point of intersection:

$$x = \frac{15}{13} \text{ and } y = 2 - \left(\frac{2}{3}\right) x = \frac{16}{13}.$$

The resulting point $P(\frac{15}{13}, \frac{16}{13})$ is a single common point of the circle with radius $\sqrt{\frac{1}{13}}$ and the given line. In other words, our circle is tangent to the given line.

Finding a circle with a given center and tangent to a given line is a problem of interest in itself, which we discuss next.

8. Circle with a given center and tangent to a given line.

   We can use the same example as above and ask for a circle centered at $A(1,1)$ and tangent to the line
   $$L : 2x + 3y - 6 = 0.$$

   If P is the point of tangency, then we simply note that the segment $\overline{AP}$ must be perpendicular to the line L and the point P is nothing but the intersection point of L with a line passing through A and perpendicular to L.

   Since the line L is
   $$2x + 3y = 6$$
   we already know that any perpendicular line must look like: [9]
   $$L' : -3x + 2y = k \text{ for some } k.$$

   Since we want it to pass through $A(1,1)$ we get:
   $$-3(1) + 2(1) = k$$
   so $k = -1$ and the line is:
   $$L' : -3x + 2y = -1.$$

   We use our known methods to determine that the common point of L and L' is $P(15/13, 16/13)$. The distance $d(A, P)$ is then seen to be $\sqrt{\frac{1}{13}}$ as before.

   Note that this calculation was much easier and we solved a very different looking problem, yet got the same answer. This is an important problem solving principle! Solve an equivalent but different problem which has a simpler formula.

   There is yet another way to look at our work above.

[9] The duck principle in action again: It is clearly a line with correct slope and we can choose k to make it pass through any given point!

9. The distance between a point and a line. Given a line $L : ax + by + c = 0$ and a point $A(p, q)$, show that the point P on the line L which is closest to A is given by the formula

$$P(p - a\frac{w}{a^2 + b^2}, q - b\frac{w}{a^2 + b^2}) \text{ where } w = ap + bq + c.$$

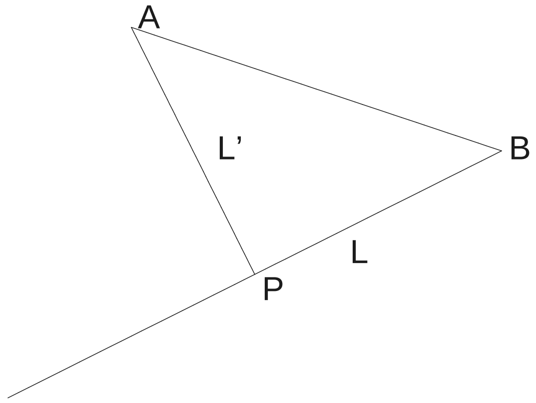
The shortest distance, therefore, is given by

$$\sqrt{\frac{w^2}{a^2 + b^2}} = \frac{|w|}{\sqrt{a^2 + b^2}}.$$

This is often remembered as:

$$\text{The distance from } (p, q) \text{ to } ax + by + c = 0 \text{ is } \frac{|ap + bq + c|}{\sqrt{a^2 + b^2}}.$$

Proof of the formula. We take a point P on L to be the intersection of the line L and a line L' through A perpendicular to L. We claim that this point P must give the shortest distance between any point on L and A.



To see this, take any other point B on L and consider the triangle $\triangle APB$. This has a right angle at P and hence its hypotenuse $\overline{AB}$ is longer than the side $\overline{AP}$; in other words

$$d(A, B) > d(A, P) \text{ for any other point } B \text{ on } L.$$

We leave it as an exercise to the reader to show that the above formula works. Here are the suggested steps to take: [10]

[10] There is another possible route to this answer; namely to write the usual equations for L, L', solving them for the common point P and then finding the distance $d(A, P)$. The reader should try this approach to appreciate the virtue of using the parametric form for L' as suggested!

- The perpendicular line L' should have parametric equations:

$$x = p + at, \ \ y = q + bt.$$

Just verify that this line must be perpendicular to the given line and it obviously passes through the point $A(p, q)$ (at $t = 0$.)

- The common point P of L, L' is obtained by substituting the parametric equations into $ax + by + c = 0$. Thus for the point P, we have:

$$a(p + at) + b(q + bt) + c = 0 \text{ or } t(a^2 + b^2) = -(ap + bq + c).$$

In other words:

$$t = -\frac{w}{a^2 + b^2}.$$

- The point P, therefore is:

$$x = p - a\frac{w}{a^2 + b^2}, y = q - b\frac{w}{a^2 + b^2}.$$

- Now the distance between $A(p, q)$ and P is clearly

$$d(A, P) = \sqrt{(at)^2 + (bt)^2} = \sqrt{(a^2 + b^2)t^2} = \sqrt{(a^2 + b^2) \cdot \frac{w^2}{(a^2 + b^2)^2}}.$$

- The final answer simplifies to:

$$\sqrt{\frac{w^2}{a^2 + b^2}} = \frac{|w|}{\sqrt{a^2 + b^2}}$$

as claimed!

Example of using the above result. Find the point on the line $L : 2x - 3y - 6 = 0$ closest to $A(1, -1)$.

Answer: We have

$$a = 2, b = -3, c = -6, p = 2, q = -1.$$

So $w = ap + bq + c = 2(1) - 3(-1) - 6 = -1$.

Thus the point P is given by:

$$P\left(1 - 2\left(\frac{-1}{4 + 9}\right), -1 - (-3)\left(\frac{-1}{4 + 9}\right)\right) = P\left(\frac{15}{13}, -\frac{16}{13}\right).$$

Moreover, the distance is:

$$|\frac{w}{a^2 + b^2}| = |\frac{-1}{4 + 9}| = \frac{1}{13}.$$

10. Half plane defined by a line.

    Consider a line given by $3x - 4y + 5 = 0$.

    Consider the points $A(3, 4)$ and $B(4, 4)$. Determine if they are on the same or the opposite side of the line, without relying on a graph!

    For any point $P(x, y)$ we can evaluate the expression $3x - 4y + 5$. Obviously, it evaluates to 0 on the line itself. It can be shown that it keeps a constant sign on one side of the line.

    Thus at the point A, the expression evaluates to $3(3) - 4(4) + 5 = -2$, while at B we get $3(4) - 4(4) + 5 = 1$. This tells us that the two points lie on the opposite sides of the line. If we consider the origin $(0, 0)$ which gives the value 5 for the expression, we see that B and the origin lie on the same side, while A is on the other side.

    It is instructive to verify this by a careful graph and also use this idea as a double check while working with graphs.

    How should one prove the above analysis?

    Here is **a formal proof.**

    This can be safely skipped in a first reading.

    Suppose we are given a linear expression $E(x, y) = ax + by + c$ in variables x, y. Consider the line defined by $L : E(x, y) = 0$.

    Suppose we have two points P, Q which give values p, q for the expression respectively.

    Then any point on the segment from P to Q is of the form $tP + (1 - t)Q$ where $0 \leq t \leq 1$ and it is not hard the see that the expression at such a point evaluates to $tp + (1 - t)q$.

    Now suppose that p, q are both positive, then clearly $tp + (1 - t)q$ is also positive for all $0 \leq t \leq 1$.

    We claim that both the points must lie on the same side of the line L.

    Suppose, if possible that the points lie on opposite sides of the line; we shall show a contradiction, proving our claim!

    Since some point in between P, Q must lie on the line, the expression $tp + (1 - t)q$ must be zero at such point. As already observed the expression, however, always stays positive, a contradiction!

    We can similarly argue that when both p, q are negative, the expression $tp + (1 - t)q$ always stays negative for $0 \leq t \leq 1$ and hence there is no point of the line between P, Q.

When p, q have opposite signs, then we can see that that the quantity $\frac{q}{q-p}$ is between 0 and 1 and this value of t makes the expression evaluate to: [11]

$$\frac{q}{q-p}(p) + \left(1 - \frac{q}{q-p}\right) q = \frac{pq}{q-p} + \frac{-pq}{q-p} = 0.$$

Thus, the line joining P, Q must intersect the line given by $E(x, y) = 0$. This proves that the points lie on opposite sides of the line!

We record the above information for future use.

Let L be the line given by the equation $ax + by + c = 0$ as above. We naturally have that a, b cannot be both zero and hence $\sqrt{a^2 + b^2}$ is a positive quantity.

For any point $P(x, y)$ define the expression $f_L(P) = ax + by + c$.

Assume that we have points P, Q outside the line, i.e. $f_L(P) \neq 0 \neq f_L(Q)$.

These points P, Q lie on the same side of the line if $f_L(P)$ and $f_L(Q)$ have the same sign, or, equivalently, $f_L(P)f_L(Q) > 0$.

On the other hand, they lie on the opposite sides if the expressions have opposite signs, i.e $f_L(P)f_L(Q) < 0$.

Moreover as we travel along the segment $\overline{PQ}$ we get every value of the expression from $f_L(P)$ to $f_L(Q)$. In particular, the expression $f_L(A)$ is either a constant or strictly increases or strictly decreases.

This last fact is very important in the subject called Linear Programming which discusses behavior of linear expressions on higher dimensional sets.

[11]The claim that $\frac{q}{q-p}$ is between 0 and 1 is true but needs a careful analysis of cases. The reader should carefully consider the possibilities $p < 0, q > 0$ and $q < 0, p > 0$ separately and verify the claim.

Chapter 9

Trigonometry

9.1 Trigonometric parameterization of a circle.

There is another well known parameterization of a circle which is used in the definition of the trigonometric functions $\sin(t), \cos(t)$. Suppose we take a circle of radius 1 centered at the origin:

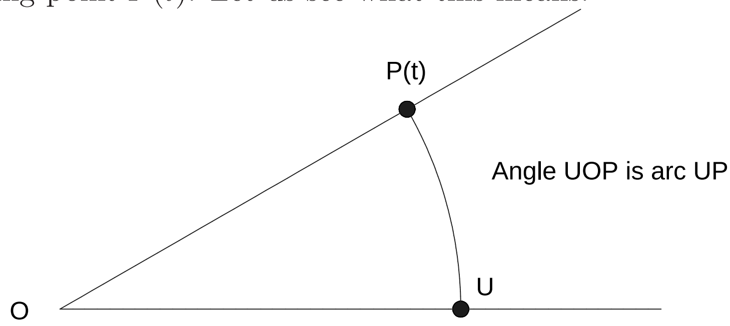
$$x^2 + y^2 = 1$$

A circle of radius 1 is called a unit circle.
We have already discussed complex numbers $z = x + iy$ which can be identified with the point $P(x, y)$ in the plane. Thus, if we view the unit circle as an Argand diagram, then it corresponds to all complex numbers z such that $|z| = 1$.
Caution. We have already seen the assertion (but without formal proof) that such a number z on the unit circle can be written as

$$z = exp(i\theta) = \cos(\theta) + i\sin(\theta)$$

where we had an ad hoc description of all these terms during the discussion of Euler's representation of complex numbers.
Thus, we may pretend that we know the trigonometric functions before formally developing their theory, by invoking Euler's representation.
The reader is advised to treat such references as motivation and as a fact to be established fully in a later course.
We now proceed to give a new explanation of the angle $Arg(z) = \theta$ and the corresponding trigonometric functions. In this explanation, we will not assume much beyond common knowledge and intuition. The virtue of Euler's representation is that it gives a much simpler explanation of the various formulas that we shall be developing in this chapter.
We wish to arrange a parameter t (in memory of θ above) such that given any point P on the unit circle, there is a certain real value of t associated with it and given any real value of the parameter t, there is a unique point $P(t)$ associated with it.

Here is a very simple idea. Start with the point $U(1,0)$ on our circle. Given any real number t, walk t units along the circle in the counter clockwise direction and stop. Call the resulting point $P(t)$. Let us see what this means.



There is a famous number called π which is defined as the ratio of the circumference of a circle to its diameter. Since our circle is a unit circle, its diameter must be 2 and hence by definition, its circumference is 2π.[1]

- Thus, if we take $t = \pi/2$ then we have gone a quarter circle and reach the point $(0, 1)$; we say $P(\pi/2) = (0, 1)$.
- If we go half the circle, we take $t = \pi$ and $P(\pi) = (-1, 0)$.
- If we further go another quarter circle then we reach the point $P(3\pi/2) = (0, -1)$ and finally, if we go a complete circle, we come back to the starting point $P(2\pi) = (1, 0)$.

This means that if we intend to travel a length t around the circle and $t = 2\pi + s$ then the position $P(t)$ at which we ultimately arrive will be the same as if we travel $s = t - 2\pi$. We express this symbolically as $P(t) = P(s) = P(t - 2\pi)$
Thus we conclude that to determine the position $P(t)$ we can always throw away multiples of 2π until we get a "remainder" which lies between 0 and 2π. Formally, we can describe this as follows:
We write our

$$t = 2\pi(n) + s \text{ where } 0 \leq s < 2\pi$$

and n is some integer.
Indeed, this is very much like our old division, except we have now graduated to dividing real numbers by real numbers! [2]

[1]This π has a very interesting history and has occupied many talented mathematicians over the ages. Finding more and more digits of the decimal expansion of π is a challenging mathematical problem. There are many books and articles on the subject of π including a movie with the same title! People have imagined many mysterious properties for the decimal digits of this real number, which is proven to be transcendental. This means that there is no non zero polynomial $p(x)$ with integer coefficients such that $p(\pi) = 0$.

[2]Do note that we still insist that our quotient is an integer, the remainder can be a real number, but is restricted to be in the interval $[0, 2\pi)$.

Then we can march s units along the unit circle counter clockwise from the point $U(1,0)$ and reach a point $P(s)$.

Definition: Locator Point The well defined point $P(s)$ is said to be the locator point for the angle measure t. It may also be denoted as $P(t)$ if convenient.
So for any number t we have an associated point $P(t)$ on the unit circle. As the above figure indicates there is uniquely associated to $P(t)$ an angle made by the ray $\overrightarrow{OP}$ with the starting ray $\overrightarrow{OU}$, where P stands for the point $P(s)$. We **denote this angle** by the symbol $\angle$**UOP**.

Some Observations.

- This allows us to think of the number t as **the measure of an angle** $\angle$**UOP**; namely the resulting angle made by the ray $\overrightarrow{OP}$ with the ray $\overrightarrow{OU}$.

- While we have casually invited you to measure the distance, mathematically we have not created any formula or procedure to calculate it, nor have we defined what a distance is. We have just appealed to common sense.

  In a higher level course, you will find these issues resolved.

- This measure is said to be in units called "radians". The reason for this word is probably as follows.

  If we take a circle of radius r then the circumference of the circle, by definition, is $2\pi r$, so the ratio $\frac{\text{distance along the arc}}{\text{radius}}$ is 2π. Thus the measure t can be thought of as a measure in units of "radiuses" (or should it be radii?).

- By an abuse of the notation, we may simply call t to be the angle $\angle UOP$, in radians. Since it is the number t that is used in our formulas, this should not cause confusion.

  However, if you are working with geometry itself, it is customary to declare $t = m\angle UOP$, i.e. t is the measure of the angle $\angle UOP$.

- The possible "numerical" angles (numbers of radians) for the same locator point are infinitely many, so for a given angle of measure t we usually use the remainder s of t after throwing away enough multiples of 2π as *the* radian measure of the angle. Thus, we take the liberty to say things like "This angle is y radians." when of course we know that the angle itself is a geometric figure and "y radians" is simply its measure.

  Thus for the point $B(0,1)$ the angle of the ray $\overrightarrow{OB}$ is said to be $\pi/2$ radians.

- We reemphasize that we can either march off the t units directly or march off the s units where s is the remainder of t after dividing by 2π as explained above.

  If we choose to march off t units directly, then a negative t requires us to march in the clockwise direction instead of the counter clockwise direction (also called anti clockwise direction.) [3]

- We formally define $P(t) = P(s)$ to be the **locator point** of the "angle" t (in radians).

- In view of the complex number theory, we can simply think of the locator point as corresponding to the complex number $exp(it) = \cos(t) + i\sin(t)$.

- **A more familiar unit: degrees.**

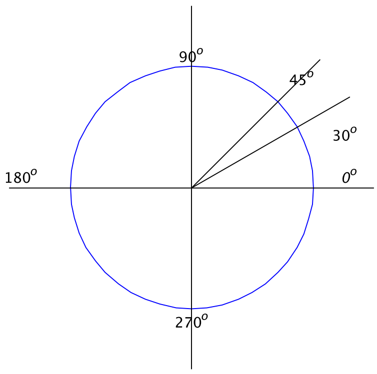
  Even though this description is pleasing, it is difficult to work with, since the number π itself is such a mystery. Even a fancy calculator or a great big computer can only approximate the number π and hence any such calculations are basically imprecise! Then why do we still want to mention and use this number? Because, in mathematical work and especially Calculus, it is important to work with radians, since any other units make π pop into our formulas making new complications!

  However, for our simpler application, it is more convenient to use the usual measure called degrees.

- We simply divide the circle into 360 equal parts and call the angle corresponding to each part a degree. Thus our point $B(0, 1)$ corresponds to the angle of $\frac{360}{4} = 90$ degrees and it is often written as $90°$.

  Here is a picture of the circle showing the degree measures for some important points. The picture shows "90 deg" in place of $90°$ and similarly for all the other angles.

[3] It is not clear why those great old mathematicians wanted "clockwise" to be the negative direction, what did they have against clocks anyway? This convention of positive counter clockwise (anti clockwise) direction, however, is now firmly entrenched in tradition!



- It is well worth making a simple rule to convert between radian and degree measures.

  To convert radians to degrees, simply multiply by $\frac{180}{\pi}$.

  To convert from degrees to radians, reverse the procedure and multiply by $\frac{\pi}{180}$.

  We now review how the locator point $P(t)$ is determined, if t is given in degrees.

  We write:

  $$t = 360n + s \text{ where } 0 \leq s < 360 \text{ and } n \text{ is some integer.}$$

  Thus s is the remainder of t after division by 360 and we mark off s degrees in our circle (in the counter clockwise manner) to locate the point $P(t) = P(s)$.

  Thus, for $t = 765$ we see that $t = 360(2) + 45$ so $s = 45$ and the corresponding point is the point at the tip of the segment making the 45 degree angle.

  Note that for $t = -1035$, we get the same s since $-1035 = 360(-3) + 45$.

  For the reader's convenience we present a table of the corresponding angle measures of the marked points in our picture.

degrees	0	30	45	90	180	270
radians	0	$\pi/6$	$\pi/4$	$\pi/2$	π	$3\pi/2$

- In the remaining work, we shall use the degree notation $^\circ$ when we use degrees, but when we mean to use radians, we may skip any notation or mention of units!

Definition of the Trigonometric Functions.
Given an angle t, let $P(t)$ be its locator point. We know its x and y coordinates. Write

$$P(t) = (x(t), y(t)).$$

Then the **definition of the trigonometric functions** is simply:

$$\sin(t) = y(t) \text{ and } \cos(t) = x(t).$$

Of course, In Euler's representation, this was already stated. So it is consistent. Before proceeding, let us calculate a bunch of these values.

Some trigonometric values.

1. Find the values of $\cos(0), \sin(0)$.

   Answer: When the angle $\angle UOP$ is 0, our point P is the same as $U(1,0)$. So by definition,

   $$\cos(0) = 1, \sin(0) = 0.$$

   This can be seen just as easily by using Euler's representation, since $\exp(0) = 1 + \frac{0}{1!} + \cdots + \frac{0^n}{n!} + \cdots$ is clearly 1 and its real and imaginary parts are clearly $1, 0$ respectively.

2. Find the values of $\cos(90°), \sin(90°)$.

   Answer: We already know this point $P(90°)$ to be $(0,1)$, so we get:

   $$\cos(90°) = 0, \sin(90°) = 1.$$

   It is not so easy to determine that $exp(i\pi/2) = i$. Thus, our geometric definition is more efficient here.

   The reader can verify the values for 180° and 270° just as easily. They are:

   $$\cos(180°) = -1\ ,\ \ \sin(180°) = 0$$

   and

   $$\cos(270°) = 0\ ,\ \ \sin(270°) = -1.$$

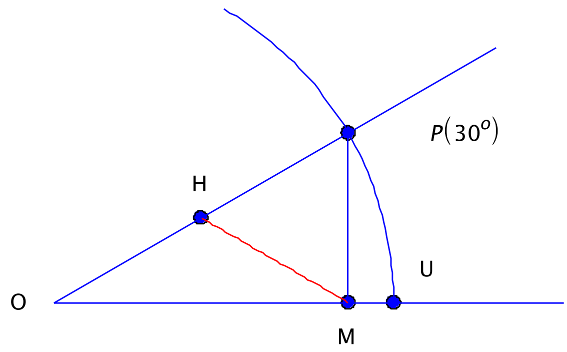
   Again, in terms of Euler's representation, we are claiming that $\exp(i\pi) = -1$ and $\exp(3\pi/2) = -i$. Neither of these is immediately obvious from the definition of exp.

3. Find the values of $\cos(30°), \sin(30°)$.

   Answer: This needs some real work! We shall give a shortcut later.

   Let us make a triangle $\triangle OMP$ where P is the point at 30° and M is the foot of the perpendicular from P onto the segment $\overline{OU}$.

   We show a picture below where we have added the midpoint H of OP and the segment $\overline{HM}$ joining it with M.



A little geometric argument can be used to deduce that [4]

$$d(H,O) = d(H,M) = d(H,P)$$

and then the triangle $\triangle HMP$ is equilateral. This says that

$$d(M,P) = d(H,P) = \frac{1}{2}d(O,P) = \frac{1}{2}.$$

Thus, we have given a brief idea of how to figure out that

$$\sin(30^\circ) = \frac{1}{2}.$$

How to find $\cos(30^\circ)$? We can now use simple algebra using the fact that the point $P(30^\circ)$ is on the circle $x^2 + y^2 = 1$. So,

$$\cos(30^\circ)^2 + (\frac{1}{2})^2 = 1$$

and it is easy to deduce that

$$\cos(30^\circ)^2 = 1 - \frac{1}{4} = \frac{3}{4}.$$

It follows that

$$\cos(30^\circ) = \pm\sqrt{\frac{3}{4}} = \pm\frac{\sqrt{3}}{2}.$$

[4]Here is a hint of the geometric argument. If we draw a circle centered at H and having OP as diameter, then we know from our work with the circles that the point M is on this circle, since the angle between OM and PM is known to be 90°. Therefore, all the distances are equal to the radius of this circle.

Which sign shall we take? Clearly, from the picture, the x-coordinate is positive, so we get

$$\cos(30°) = \frac{\sqrt{3}}{2}.$$

Bonus thought! What do we get if we take the minus sign? After a little thought, you can see that the angle of $180 - 30 = 150$ degrees will have the same sin, namely $\frac{1}{2}$ and its cos will be indeed $-\frac{\sqrt{3}}{2}$.

Similarly, for the angle $-30°$ which is also the angle $330°$, you can see that

$$\sin(-30°) = -\frac{1}{2} \text{ and } \cos(-30°) = \frac{\sqrt{3}}{2}.$$

4. Find the value of $\sin(13°), \cos(13°)$.

   Answer: There is no clever way for this calculation. All we can hope to do is to use a calculator or computer, or develop a formula which can give us better and better approximations.

   A calculator will say:

$$\sin(13°) = 0.2249510544 \text{ and } \cos(13°) = 0.9743700648.$$

   In general, trigonometric functions of a random angle can only be approximated by using a calculator, or we have to work very hard to deduce its mathematical properties.

   The same holds for the exp function.

9.2 Basic Formulas for the Trigonometric Functions.

Introduction.

The aim of this section is to list the basic formulas for the various trigonometric functions and learn their use. The actual theoretical development of these will be done later.

These functions, though easy to define, don't have simple formulas and evaluating their values requires careful approximations. You can use a calculator and get decimal numbers, but that is not our goal.

We shall try to understand how these functions can be evaluated precisely for many special values of the argument and how they can be manipulated formally using algebraic techniques. They form the basis of many mathematical theories and are, in some sense, the most important functions of modern mathematics, next to polynomials. Indeed, the theory of Fourier analysis lets us approximate all practical functions in terms of suitable trigonometric functions and it is crucial to learn how to manipulate them.

It is customary to also learn about exponential and logarithmic functions. Our discussion of Euler's representation also indicates a formula for the exponential function

$$\exp(x) = 1 + \frac{x}{1!} + \cdots + \frac{x^n}{n!} + \cdots .$$

We shall discuss it later.

As far as evaluation at specific points is concerned, trigonometric, exponential or logarithmic functions are equally easy - they are keys on the calculator! You probably have already seen and used them in concrete problems anyway.

We have already described connection between the angle t, its corresponding locator point $P(t)$ and the corresponding trigonometric functions

$$\sin(t) = \text{ the } y\text{-coordinate of } P(t) \text{ and } \cos(t) = \text{ the } x\text{-coordinate of } P(t).$$

Unless otherwise stated, our angles will always be in radian measure, so a right angle has measure $\pi/2$.

Negative angles.

Sometimes it reduces our work if we use negative values of angles, rather that converting them to positive angles. After all, $-30°$ is easier to visualize than $330°$. We can give a natural meaning to negative angles, they represent arc length measure in the *negative or clockwise* direction! Check out that going $30°$ clockwise or $330°$ counterclockwise leads to the same locator point!

Thus, sometimes it is considered more convenient to take the angle measure between $-180°$ to $180°$.

Definition: Other trigonometric functions. Given an angle $t \in \Re$, we know

that its locator point $P(t)$ has coordinates $(\cos(t), \sin(t))$.
We now **define four more functions** which are simply related to the above two functions:

$$\tan(t) = \frac{\sin(t)}{\cos(t)}\ ,\ \cot(t) = \frac{\cos(t)}{\sin(t)}.$$

Also:

$$\csc(t) = \frac{1}{\sin(t)}\ ,\ \sec(t) = \frac{1}{\cos(t)}.$$

These functions have full names which are:

Notation	sin	cos	tan	cot	csc	sec
Full name	Sine	Cosine	Tangent	Cotangent	Cosecant	Secant

We have the following basic identities derived from the fact that $P(t)$ is on the unit circle.

Fundamental Identity 1. $$\sin^2(t) + \cos^2(t) = 1$$

We can divide this identity by $\sin^2(t)$ or $\cos^2(t)$ to derive two new identities for the other functions. The reader should verify these:

Fundamental Identity 2. $$1 + \cot^2(t) = \csc^2(t)$$

Fundamental Identity 3. $$\tan^2(t) + 1 = \sec^2(t)$$

Remark on notation. It is customary to write $\sin^2(t)$ in place of $(\sin(t))^2$. This is not done for other functions. Indeed, many times the notation $f^2(x)$ is used to indicate applying the function twice - i.e. $f \circ f(x) = f(f(x))$. It is wise to verify what this notation means before making guesses!
For trigonometric functions, the convention is as we explained above! If you have any confusion in your mind, it is safer to use the explicit $(\sin(t))^2$ rather than $\sin^2(t)$. This convention works for other powers too. This indeed leads to a confusion later on if you try to use negative powers. Here is how it comes up.

What is that angle?

We have given the definition of the values of trigonometric functions, given an angle either in degrees or radians. In may cases, we have no recourse but to use a calculator to get an approximate answer.
Given a value $a = \sin(t)$ we ask what is t? We already know that there is no unique answer, since adding or subtracting multiples of 2π to any one answer leads to the same value of $\sin(t)$.
If we think of sin as a function from the real numbers to real numbers, we can observe the following:

- The function is not onto. In fact, the range is easily seen to be the interval $[-1,1]$, in view of the definition in terms of the circle.

- The function is not one to one. If we think of tracing the circle from $t = -\pi/2$ to $\pi/2$ counterclockwise, then it is not hard to see that every value of a between -1 and 1 is reached once and exactly once.

  The idea is that if we draw a line $y = a$, then it hits the right half of the circle in exactly one point. We can take t to be the angle corresponding to that point, keeping it between $-\pi/2$ and $\pi/2$, then we get what we want.

- Thus, if we trim the domain of the sin function to $[-\pi/2, \pi/2]$, then it is a one to one and onto function with range $[-1,1]$.

  This sine function has an inverse function, which we denote by $\sin^{-1}$ or arcsin.

  Thus we have defined:

$$t = \sin^{-1}(a) \text{ if } \sin(t) = a \textbf{ and } -\pi/2 \leq t \leq \pi/2.$$

- Now comes the **confusion.** It is tempting to think that $\sin^{-1}(a)$ as the same as $(\sin(a))^{-1} = \dfrac{1}{\sin(a)}$. This is natural since we write $\sin^2(t)$ for $(\sin(t))^2$.

  It is customary to avoid this confusion by using a different notation, **arcsin in place of $\sin^{-1}$**.

  We will systematically use this notation.

- **Complete solution** of $\sin(t) = a$. If $t = \arcsin(a)$, then clearly t is a solution of the equation $\sin(t) = a$. There is, however another equally interesting solution, namely $\pi - t$. We will soon show that $\sin(\pi - t) = \sin(t)$ for all values of t **-the Supplementary Angle Identity.** It corresponds to the other point of intersection of the line $y = a$ with the circle $x^2 + y^2 = 1$.

  Thus, we have two solutions: $t = \sin^{-1}(a), \pi - \sin^{-1}(a)$. It can be shown that all possible solutions are obtained from adding arbitrary multiples of 2π to these two basic answers.

We similarly have inverse function definitions for other trigonometric functions. We list these without further comments.

- **Inverse cosine.**

  We **define** $\cos^{-1}(a) = \arccos(a)$ as the angle t such that $\cos(t) = a$ **and** $0 \leq t \leq \pi$.

In general, there are two solutions to $\cos(t) = a$, namely $t = \arccos(a)$ and $t = -\arccos(a)$. This follows from the evenness of the cosine function that we shall soon prove.

These also represent the two points of intersection of the vertical line $x = a$ with the circle $x^2 + y^2 = 1$.

- **Inverse tangent.**

  We **define** $\tan^{-1}(a) = \arctan(a)$ as the angle t such that $\tan(t) = a$ **and** $-\pi/2 \leq t \leq \pi/2$.

  In general, there are two solutions to $\tan(t) = a$, namely $t = \arctan(a)$ and $t = \pi + \arctan(a)$. This follows from the Fundamental Identity 10 (to be proved below) by taking $s = \pi$.

- The remaining three inverse functions are defined naturally from above, but generally are not specifically needed. Thus $\sin(t) = a$ is the same as $\csc(t) = \frac{1}{a}$. So, $\text{arccsc}(a)$ is simply defined as $\arcsin(\frac{1}{a})$.

  The others are similarly handled.

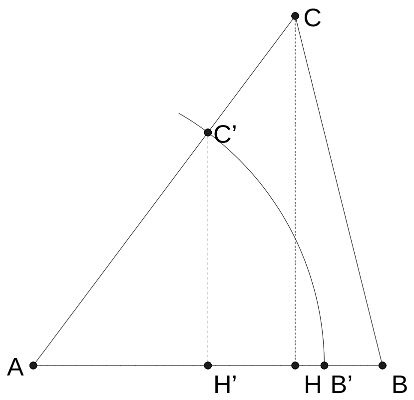
We will give examples of using the inverse trigonometric functions later.

9.3 Connection with the usual Trigonometric Functions.

While the above definitions are precise and general, it is important to connect them with the classical definition.
You may recall that given a triangle $\triangle ABC$ we denoted by A the angle at A and in the modern notation above, it corresponds to the angle $\angle BAC$.
Consider the following picture, where $\overline{CH}$ is the perpendicular from C onto $\overline{AB}$.
The old definition used to be:

$$\sin(A) = \frac{d(C,H)}{d(A,C)} \text{ and } \cos(A) = \frac{d(A,H)}{d(A,C)}.$$

The points B', C', H' are constructed by making the unit circle centered at A and clearly the two triangles $\triangle AHC$ and $\triangle AH'C'$ are similar with proportional sides.



Hence we have:

$$\frac{d(C,H)}{d(A,C)} = \frac{d(C',H')}{d(A,C')} = \frac{d(C',H')}{1} = \sin(A).$$

This proves that the old and the new sines are the same.
Similarly we get the cosine matched!
Thus, we can continue to use the old definitions when needed. The major problem is when the angle goes above $90°$. Then the perpendicular H falls to the left of A and the distance $d(A,H)$ will have to be interpreted as negative to match our new definition!
The old definitions did not provide for this at all. They usually assumed the angles to be between $0°$ and $90°$ and for bigger angles (obtuse angles) you had to make special cases for calculations!
Our new definitions take care of all the cases naturally!

9.4 Important formulas.

In a later section, we shall give proofs for various identities connecting the trigonometric functions. These are very important in working out useful values of these functions and in making convenient formulas for practical problems.
It is useful to understand what these identities are saying and learn to use them, before you understand all the details of their proof.
Using Euler's Representation.
Often, using Euler's representation, we can get a trivial proof of the identities under consideration. We will explicitly indicate such a shortcut when available.

1. **Addition Formulas** We describe how to find the trigonometric functions for the sum or difference of two angles.

   Fundamental Identity 4. $\cos(t-s) = \cos(t)\cos(s) + \sin(t)\sin(s)$

   Fundamental Identity 4-variant. $\cos(t+s) = \cos(t)\cos(s) - \sin(t)\sin(s)$

   It is important to note the changes in sign when we go from cos formulas to the sin formulas.

   Fundamental Identity 5. $\sin(t+s) = \sin(t)\cos(s) + \cos(t)\sin(s)$

   Fundamental Identity 5-variant. $\sin(t-s) = \sin(t)\cos(s) - \cos(t)\sin(s)$

   Examples of use.

   Given the locator point $P(20°) = (0.9397, 0.3420)$ and the locator point $P(30°) = (0.8660, 0.5000)$, answer the following:

   - What is $\cos(50°)$?
     Answer: Note that $50 = 20 + 30$. Using Fundamental Identity 4-variant:
     $$\cos(50°) = \cos(20° + 30°) = (0.9397)(0.8660) - (0.3420)(0.5000) = 0.6427.$$
   - What is $\sin(10°)$?
     Answer: Now we write $10 = 30 - 20$. Using Fundamental Identity 5-variant:
     $$\sin(10°) = \sin(30° - 20°) = (0.5000)(0.9397) - (0.8660)(0.3420) = 0.1737.$$

2. **An important observation.** The exponential function has the following wonderful property:
$$\exp(a)\exp(b) = \exp(a+b)$$
where a, b can be any real or complex numbers. While a precise proof which also explains the meaning of the infinite sum is hard to see, a formal algebraic proof is rather easy.

   Here is how it goes.

   Notice that by multiplying $\exp(a)$ with $\exp(b)$ and collecting together terms whose total degree (as monomials in a, b) is n, we see the following:
$$\begin{aligned}\exp(a)\exp(b) &= \left(1+\tfrac{a}{1!}+\cdots+\tfrac{a^n}{n!}+\cdots\right)\cdot\left(1+\tfrac{b}{1!}+\cdots+\tfrac{b^n}{n!}+\cdots\right)\\ &= 1+\left(\tfrac{a+b}{1!}\right)+\left(\tfrac{a^2}{2!}+\tfrac{a}{1!}\tfrac{b}{1!}+\tfrac{b^2}{2!}\right)+\cdots\\ &\quad \left(\tfrac{a^n}{n!}+\tfrac{a^{n-1}}{(n-1)!}\tfrac{b}{1!}+\cdots+\tfrac{a}{1!}\tfrac{b^{n-1}}{(n-1)!}+\tfrac{b^n}{n!}\right)+\cdots\end{aligned}.$$

   A little manipulation of these terms combined with our knowledge of the Binomial Theorem, we can show that
$$\exp(a)\exp(b) = 1+\frac{(a+b)}{1!}+\cdots+\frac{(a+b)^n}{n!}+\cdots$$
and clearly it follows that
$$\exp(a)\exp(b) = \exp(a+b).$$

   This should remind us of the rules of exponents and suggest that the function $\exp(a)$ must be some number raised to a power a. What number could it be?

   Well, define:
$$\exp(1) = e = 1+\frac{1}{1!}+\cdots+\frac{1}{n!}+\cdots.$$
This number "e" is known as the base of natural logarithms and we can argue that $\exp(a)$ is nothing but e^a. Of course, the point is that it takes a lot of work to make sense out of e^a when a is to be any arbitrary number. Usually, this is done for positive integer values first, then extended to all integers, then to rational numbers, then to real numbers and so on. Our expansion as a power series (long polynomial) gives an immediate definition instead!

3. The above formula coupled with the Euler representation make the trigonometric formulas easy to derive and remember.

   - **Addition formulas at once.** Thus note that
$$\exp(i(t+s)) = \exp(it)\exp(is)$$

and hence

$$\cos(t+s) + i\sin(t+s) = (\cos(t) + i\sin(t))(\cos(s) + i\sin(s).$$

Expanding the RHS and comparing the real and imaginary parts of both sides gives:

$$\cos(t+s) = \cos(t)\cos(s) - \sin(t)\sin(s)$$

and

$$\sin(t+s) = \sin(t)\cos(s) + \cos(t)\sin(s).$$

- **Formulas for negative angles.** If we review the formulas for $\cos(t)$ and $\sin(t)$ it is easy to see that

$$\cos(-t) = 1 - \frac{(-t)^2}{2!} + \frac{(-t)^4}{4!} + \cdots + (-1)^n \frac{(-t)^{2n}}{(2n)!} + \cdots = \cos(t)$$

and

$$\sin(-t) = (-t) - \frac{(-t)^3}{3!} + \cdots + (-1)^{n-1} \frac{(-t)^{2n-1}}{(2n-1)!} + \cdots = -\sin(t).$$

- Thus, we have actually "proved" the above stated fundamental identities and will have a trivial proof of what is coming next. The catch is, of course, that we don't have a complete proof of the Euler representation itself, yet!

4. **General results from the above identities.**

Complementary Angle Identity.

$$\cos(\pi/2 - s) = \sin(s)\ ,\ \ \sin(\pi/2 - s) = \cos(s)$$

Use $t = \pi/2$ in Fundamental Identities 4 and 5-variant.

Proof by Euler representation: Note that $\exp(i\pi/2) = i$ and hence

$$\begin{aligned} \cos(\pi/2 - s) + i\sin(\pi/2 - s) &= \exp(i\pi/2)\exp(-is) \\ &= (i)(\cos(s) - i\sin(s)) \\ &= \sin(s) + i\cos(s). \end{aligned}$$

Supplementary Angle Identity. $\cos(\pi - s) = -\cos(s)\ ,\ \ \sin(\pi - s) = \sin(s)$

Use $t = \pi$ in Fundamental Identities 4 and 5-variant.

Proof by Euler representation.

Calculate $\exp(i(\pi - s))$ as $\exp(i\pi)\exp(-is)$.

We have already established the following from Euler representation.

Evenness of the cosine function. $\cos(-s) = \cos(s)$

Oddness of the sine function. $\sin(-s) = -\sin(s)$

Use $t = 0$ in Fundamental Identities 4 and 5-variant.

Examples of use.

Given the locator point $P(20°) = (0.9397, 0.3420)$ and the locator point $P(30°) = (0.8660, 0.5000)$, answer the following:

- What is $\cos(70°)$?
  Answer: Using Complementary Angle Identity:
  $$\cos(70°) = \cos(90° - 20°) = \sin(20°) = 0.3420.$$
- What is $\cos(250°)$?
  Answer: Note that $250° = 180° + 70°$. So from the addition formulas, we get:
  $$\cos(250°) = \cos(180°)\cos(70°) - \sin(180°)\sin(70°) = 0 - \cos(70°) = -0.3420.$$
  We have, of course used the fact that $P(180°) = (-1, 0)$.
- What is $\sin(60°)$?
  Answer: Using Complementary Angle Identity:
  $$\sin(60°) = \sin(90° - 30°) = \cos(30°) = 0.8660.$$
- What is $\cos(-30°)$?
  Answer: Using Evenness of the cosine function:
  $$\cos(-30°) = \cos(30°) = 0.5000.$$
- What is $\sin(-30°)$?
  Answer: Using Oddness of the sine function:
  $$\sin(-20°) = -\sin(20°) = -0.3420.$$

5. **Double angle formulas.**

Fundamental Identity 6. $\cos(2t) = \cos^2(t) - \sin^2(t)$

Fundamental Identity 7. $\sin(2t) = 2\sin(t)\cos(t)$

Proof by Euler representation. Calculate $\exp(i(2t))$ as $\exp(it)^2$. Thus, we get:

$$\cos(2t) + i\sin(2t) = (\cos(t) + i\sin(t))^2 = \left(\cos^2(t) - \sin^2(t) + 2i\sin(t)\cos(t)\right).$$

Now compare real and imaginary parts on both sides.

The fundamental identity 6 has two variations and they both are very useful, so we record them individually.

Fundamental Identity 6-variant 1. $\cos(2t) = 2\cos^2(t) - 1$

Fundamental Identity 6-variant 2. $\cos(2t) = 1 - 2\sin^2(t)$

Examples of use.

Given the locator point $P(20°) = (0.9397, 0.3420)$ and the locator point $P(30°) = (0.8660, 0.5000)$, answer the following:

- What is $\cos(40°)$?
  Answer: Using Fundamental Identity 6 (or variations) :

  $$\cos(40°) = \cos(2\cdot 20°) = 0.9397^2 - 0.3420^2 = 2(0.9397^2) - 1 = 1 - 2(0.3420^2).$$

  These evaluate to: 0.7660 to a four digit accuracy.
- What is $\sin(40°)$?
  Answer: Using Fundamental Identity 7:

  $$\sin(40°) = \sin(2\cdot 20°) = 2(0.9397)(0.3420) = 0.6428.$$

- What is $\sin(10°)$?

  Answer: Later, we shall use a half angle formula for this, but now we can think $10° = 40° - 30°$.

  So, from Fundamental Identity 5-variant:

  $$\sin(10°) = \sin(40°)\cos(30°) - \cos(40°)\sin(30°).$$

  Using computed values, we get:

  $$(0.6428)(0.8660) - (0.7660)(0.5000) = 0.1737.$$

- What is $\cos(10°)$?

  Answer: We can use the Fundamental Identity 4 as above, or we may use the Fundamental Identity 1 as

  $$\cos(t) = \pm\sqrt{1 - \sin^2(t)}.$$

  Note that this has a plus or minus sign and it needs to be decided by the quadrant that our angle is in.

  Here $10°$ is in the first quadrant, so we have the plus sign and we get:

  $$\cos(10°) = \sqrt{1 - \sin^2(10°)} = \sqrt{1 - 0.1737^2} = 0.9848.$$

6. **Half Angle formulas.**

   Fundamental Identity 8. $$\cos\left(\frac{t}{2}\right) = \pm\sqrt{\frac{1+\cos(t)}{2}}$$

   Fundamental Identity 9. $$\sin\left(\frac{t}{2}\right) = \pm\sqrt{\frac{1-\cos(t)}{2}}$$

   Examples of use.

   Given the locator point $P(20°) = (0.9397, 0.3420)$, answer the following:

   - What is $\cos(10°)$?

     Answer: Using Fundamental Identity 8 :

     $$\cos(10°) = \sqrt{\frac{1+0.9397}{2}} = 0.9848 \text{ again.}$$

     Note that we used the plus sign since the angle is in the first quadrant.

- What is $\sin(10°)$?

  Answer: Using Fundamental Identity 9 :

$$\sin(10°) = \sqrt{\frac{1 - 0.9397}{2}} = 0.1736.$$

  Note that we used the plus since the angle is in the first quadrant.

  Also note that the **accuracy** is lost a bit. A more precise value of $\cos(20°)$ is .9396926208 and a more accurate result for our $\sin(10°)$ comes out 0.1736481776. This should explain why the 4-digit answer can come out 0.1736 or 0.1737 depending on how it is worked out. You have to worry about this when supplying a numerical answer.

Recalculation of a known value.

What is $\sin(\frac{\pi}{4})$ and $\cos(\frac{\pi}{4})$?

Answer:

Since $2\frac{\pi}{4} = \frac{\pi}{2}$, we see two conclusions:

$$\sin(\frac{\pi}{4}) = \cos(\frac{\pi}{4})$$

due to the Complementary Angle Identity and

$$\sin(\frac{\pi}{4}) = \pm\sqrt{\frac{1 - \cos(\frac{\pi}{2})}{2}}.$$

by the Fundamental Identity 9. Using the fact that $\cos(\frac{\pi}{2}) = 0$, we deduce:

$$\sin(\frac{\pi}{4}) = \cos(\frac{\pi}{4}) = \frac{1}{\sqrt{2}}.$$

7. **Functions for related angles.** Given one angle t, there are some angles naturally related to it. These are:

   - **Supplementary angle** $\pi - t$.
   - **Complementary angle** $\pi/2 - t$.
   - **Reflection at the origin** $\pi + t$ which is equivalent to $t - \pi$.
   - **Reflection in the x axis** $-t$.

   The trigonometric functions of all these related angles can be easily calculated in terms of $\sin(t), \cos(t)$. We have shown some of the calculations and we leave the rest as an exercise.

8. **What about the other functions?** You might be wondering about the other four named functions we introduced, namely secant, cosecant, tangent and cotangent. We have not seen them except in the identities 2,3.

   The above results for the sine and cosine functions lead to corresponding results for them. Such results are best derived as needed, rather than memorized.

   Thus, for example, you can deduce for yourself that sec is an even function and that tan, cot and csc are odd functions. As a sample proof, see that:

$$\tan(-t) = \frac{\sin(-t)}{\cos(-t)} = \frac{-\sin(t)}{\cos(t)} = -\tan(t)$$

   where we have used the known properties of sine and cosine functions in the middle.

   There is one formula worth noting:

   Fundamental Identity 10. $$\tan(s+t) = \frac{\tan(s)+\tan(t)}{1-\tan(s)\tan(t)}$$

   The proof consists of using the identities 5 and the variant of the identity 4, together with some easy simplifications. Actually the easiest proof comes by starting from the right hand side and simplifying. We present the steps without explanation:

$$\begin{aligned}
\frac{\tan(s)+\tan(t)}{1-\tan(s)\tan(t)} &= \frac{\frac{\sin(s)}{\cos(s)}+\frac{\sin(t)}{\cos(t)}}{1-\frac{\sin(s)\sin(t)}{\cos(s)\cos(t)}} \\
&= \frac{\sin(s)\cos(t)+\sin(t)\cos(s)}{\cos(s)\cos(t)-\sin(s)\sin(t)} \\
&= \frac{\sin(s+t)}{\cos(s+t)} \\
&= \tan(s+t)
\end{aligned}$$

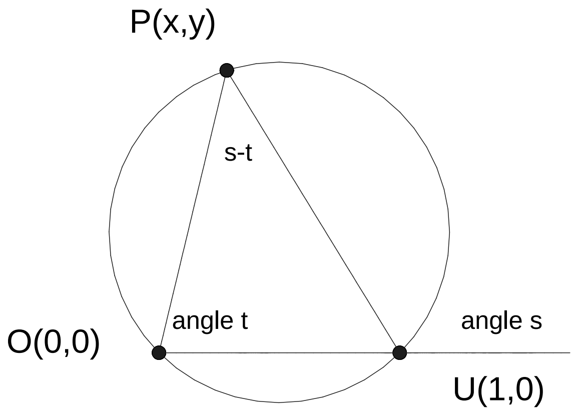
   As a consequence, we also have:

   Fundamental Identity 11. $$\tan(2t) = \frac{2\tan(t)}{1-\tan^2(t)}$$

   The proof is trivial, just take $s = t$ in the identity 10. However, it is very useful in many theoretical calculations.

9. A special application to circles. Consider two points, the origin: $O(0,0)$ and the unit point: $U(1,0)$. Consider the locus of all points $P(x,y)$ such that the angle $\angle OPU$ is fixed, say equal to α.

   What does this locus look like?



   Let us use a variation of the Fundamental Identity 10 derived by changing t to $-t$:

   Fundamental Identity 10-variant.. $$\tan(s-t) = \frac{\tan(s)-\tan(t)}{1+\tan(s)\tan(t)}$$

   Let s be the angle made by line UP with the x-axis, so we know that

$$\tan(s) = \text{ slope of the line UP } = \frac{y}{x-1}.$$

   Let t be the angle made by line OP with the x-axis, so we know that

$$\tan(t) = \text{ slope of the line OP } = \frac{y}{x}.$$

   We know that the angle $\angle OPU$ can be calculated as $s-t$. Thus, by our assumption, $s-t=\alpha$. Now by the above variant of the Fundamental Identity 10 we get:

$$\tan(\alpha) = \tan(s-t) = \frac{\tan(s)-\tan(t)}{1+\tan(s)\tan(t)} = \frac{\frac{y}{x-1}-\frac{y}{x}}{1+\frac{y}{x-1}\cdot\frac{y}{x}}.$$

   This easily simplifies to:

$$\tan(\alpha) = \frac{y}{x^2+y^2-x}$$

and thus we can rearrange it in a convenient form: [5]

$$x^2 + y^2 - x = \cot(\alpha)y.$$

This looks like a circle! Indeed, what we have essentially proved is the following well known theorem from Geometry:

Consider two end points A, B of the chord of a circle and let P be any point on the circle. Then the angle $\angle APB$ has a constant tangent.

In fact, noticing that the points A, B split the circle into two sectors, the angle is constant on each of the sectors and the angle in one sector is supplementary to the angle in the other sector.

But tangent of an angle is the negative of the tangent of the supplementary angle. How does our equation handle both these cases? Luckily, we don't have to worry. If you take the point P below the x-axis, it is easy to see that we are working with the supplementary angles of each of $s, t, (s - t)$ and our equation simply gets multiplied by -1 on both sides. [6]

This is a situation when Algebra is more powerful than the Geometry! We shall make a good use of this formula in our applications.

Observation. Recall that we have done a special case of this earlier in the chapter on circles. There we handled the case of the angle in a semi circle, which, in our current notation corresponds to $\alpha = \pi/2$. In other words, $\cot(\alpha) = 0$. At that time, we did not assume special position for our circle and got a more general formula than our current formula. If necessary, we can redo our current calculations without assuming special positions for the points, but the notation will get messy!

10. **Working with a triangle.** There are two very important formulas for the calculations with a triangle. These are well worth memorizing.

    Notation. For convenience, we assume that the vertices of our triangle are named A, B, C and the corresponding lengths of **opposite** sides are named by the corresponding small letters a, b, c.

    Convention. We shall use the letters A, B, C to also denote the corresponding angles themselves, so instead of $\sin(\angle A)$ we shall simply write $\sin(A)$ and similarly for others.

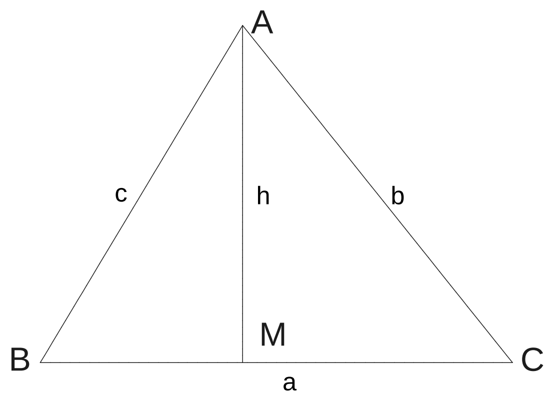
---

[5] This needs a bit of algebra: First cross multiply to get:

$$\tan(\alpha)(x^2 + y^2 - x) = y$$

and then multiply both sides by $\cot(\alpha)$.

[6] There are some missing parts in the argument. We need to argue why our special choice of points O, U still leads to the general result. We also need to argue that the angle is actually constant in each sector, we only proved that it has a constant tangent. All this can be fixed with a careful analysis of the cases.



Our picture also shows the foot of the perpendicular from A onto the side $\overline{BC}$ and its length is marked by h.

Name the part $d(B,M)$ as a_1 and the part $d(M,C)$ as a_2 so $a = a_1 + a_2$.

It is easy to note:

$$c\sin(B) = h = b\sin(C).$$

We note two facts from this:

Fact 1
$$\frac{c}{\sin(C)} = \frac{b}{\sin(B)}$$

Fact 2.
$$h^2 = bc\sin(B)\sin(C)$$

By obvious symmetry in the symbols, we claim from Fact 1, the

The sine law for a triangle. $$\frac{a}{\sin(A)} = \frac{b}{\sin(B)} = \frac{c}{\sin(C)}.$$

Now we develop a formula for a^2.

Note that

$$a^2 = (a_1 + a_2)^2 = a_1^2 + a_2^2 + 2a_1a_2.$$

Also note that

Fact 3
$$a_1 = c\cos(B)\ ,\ a_2 = b\cos(C).$$

Now using the right angle triangles $\triangle ABM$ and $\triangle AMC$, we note:

Fact 4
$$a_1^2 = c^2 - h^2\ ,\ a_2^2 = b^2 - h^2.$$

Using Facts 2,3 and 4 above, we get:

$$
\begin{array}{rcll|l}
a^2 & = & a_1^2 + a_2^2 + 2a_1a_2 & \\
 & = & c^2 - h^2 + b^2 - h^2 + 2a_1a_2 & \text{Fact 4} \\
 & = & b^2 + c^2 - 2h^2 + 2bc\cos(B)\cos(C) & \text{Fact 3} \\
 & = & b^2 + c^2 - 2bc\sin(B)\sin(C) + 2bc\cos(B)\cos(C) & \text{Fact 2} \\
 & = & b^2 + c^2 + 2bc\cos(B+C) & \text{Fundamental Identity} \\
 & & & \text{4-variant.}
\end{array}
$$

Since $B + C = \pi - A$, we see from our supplementary angle identity, $\cos(B + C) = -\cos(A)$ and hence we have:

The cosine law for a triangle. $a^2 = b^2 + c^2 - 2bc\cos(A)$.

9.5 Using trigonometry.

1. **Calculating exact values.** So far, we have succeeded in calculating the trigonometric functions for a few special angles $0°, 30°, 45°, 60°, 90°$ and their related angles.

   The half angle formula lets us evaluate many more angles.

   Thus

   $$\sin(15°) = \sqrt{\frac{1-\cos(30°)}{2}}$$

   where, we have chosen the plus sign since we know the angle is in the first quadrant. The right hand side is

   $$\sqrt{\frac{1-\frac{\sqrt{3}}{2}}{2}} = \sqrt{\frac{2-\sqrt{3}}{4}} = \frac{\sqrt{6}-\sqrt{2}}{4}.$$

   The last answer may be mysterious. When it is given to you, it is not hard to check it; just square it and see that you get the desired $\frac{2-\sqrt{3}}{4}$.

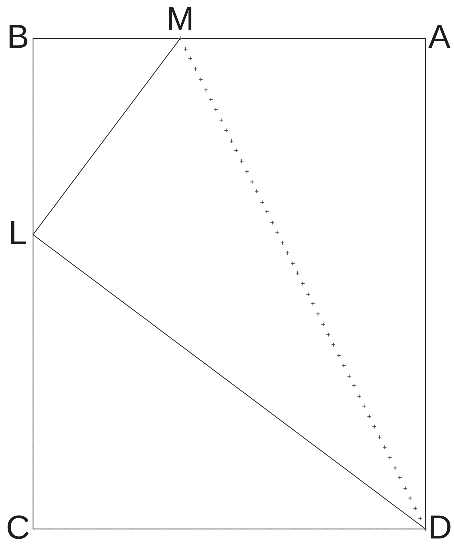
   There is a systematic theory of finding square roots of algebraic expressions like this, called surds. It used to be a standard part of old algebra books; alas, no more!

   It is clear that from this $15°$ angle, you can keep on halving the angle and get down to a small angle. For several hundred years, people did their astronomy using multiples of $7.5°$ angle.

   You may ask why one should bother with this. Your calculator will happily report the value 0.2588190451 for $\sin(15°)$. The calculator will also report the same value for our theoretical calculation $\frac{\sqrt{6}-\sqrt{2}}{4}$. You can and should use the calculator for ordinary practical problems, but not when developing the theory. You should also have a healthy suspicion about the accuracy of the reported answer, especially if it has gone through several steps of calculations.

2. **Folding Paper.** Consider a piece of paper 8 inches wide and 10 inches long.

   Fold it so that the top right corner lands on the left edge and the crease starts at the bottom right corner.



We want to determine the position M where the crease is made.

Note the names of the corners and observe that we must have:

$$d(L, D) = d(A, D) = 10 \text{ and } d(C, D) = 8.$$

By the Pythagorean theorem, we have:

$$d(L, C) = \sqrt{10^2 - 8^2} = \sqrt{36} = 6.$$

Thus, we get:

$$d(L, B) = 10 - 6 = 4.$$

Now a little argument shows that: [7]

$$\angle BLM = \angle CDL.$$

Since $\tan(\angle CDL) = \frac{6}{8} = \frac{3}{4}$ we get that $\tan(\angle BLM) = \frac{3}{4}$. Hence

$$d(B, M) = d(B, L)\tan(\angle BLM) = 4\frac{3}{4} = 3.$$

Thus the crease is made at 3 inches from the top left corner!

[7] Since the sum of the angles $\angle BLM, \angle MLD$ and $\angle DLC$ is $180°$ and since $\angle MLD$ is clearly a right angle, we get

$$\angle BLM + \angle DLC = 90°.$$

Since $\triangle LCD$ is clearly a right angle triangle, we get:

$$\angle CDL + \angle DLC = 90°.$$

The conclusion follows from these two displayed equations!

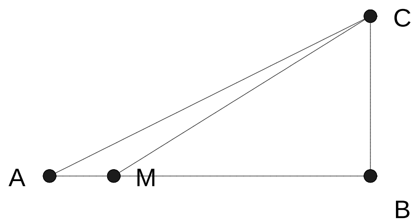
Note: We used the given dimensions. It is instructive to make similar calculations using variable names for the width and length, say w and l respectively. The reader should attempt to prove the formula using similar calculations!

$$d(B,M) = d(B,L)\tan(\angle BLM) = (l - \sqrt{l^2 - w^2})\frac{\sqrt{l^2 - w^2}}{w}.$$

3. **Estimating heights.** Consult the picture below for reference points.

   A tall building BC is observed from the position A on the ground. The top C of the building appears to have an elevation of 26.56°. This means the segment $\overline{AC}$ makes an angle of 26.56° with the horizontal $\overline{AB}$.

   If we observe the top of the building from point M 200 feet closer to the building () then it has an elevation of 32°. Estimate the height of the building.



Answer: Even though, we are given some concrete values, the recommended procedure is to name everything, finish calculations and then plug in values.

Let the angles at A and M be denoted by s and t respectively. Thus $s = 26.55°$ and $t = 32°$. Let h be the height of the building $d(B,C)$ and let k be the unknown distance $d(A,B)$. We have two equations:

$E1$ $$\tan(s) = \frac{h}{k}$$

and

$E2$ $$\tan(t) = \frac{h}{k - 200}$$

We wish to solve for h, k from these two equations.

Rearrangement gives:

$E3.$ $$h - \tan(s)k = 0\ ,\ h - \tan(t)k = -200\tan(t)$$

Subtracting the second equation in E3 from the first, we get

$E4.$ $$(\tan(t) - \tan(s))k = 200\tan(t)$$

Thus

$$k = \frac{200\tan(t)}{(\tan(t) - \tan(s))}.$$

Since we really need h and not k, we use the first equation of E3 to deduce:

$$h = k\tan(s) = \frac{200\tan(s)\tan(t)}{(\tan(t) - \tan(s))}.$$

It remains to plug in the known angles and evaluate in a calculator. You should verify that it gives an answer of 499.88 feet. [8]

4. **Using inverse trigonometric function.**

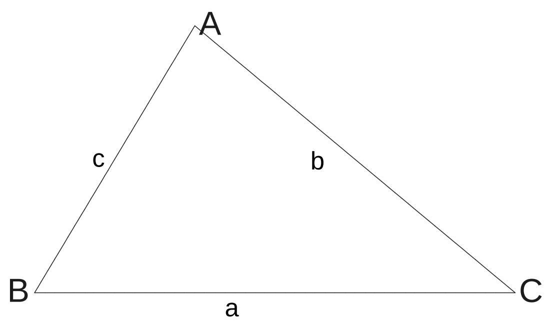
   Calculate the angle M in degree whose sine is 0.5736 and cosine is positive.

   What happens if the cosine is negative?

   Answer. We shall determine an angle in $[0, 2\pi)$. All other answers come by adding multiples of 2π.

   An evaluation of arcsin(0.5736) gives 35.0017°. If the cosine is positive, then it is the answer. Otherwise, $180° - 35.0017°$ is the answer.

   Convention. In the next two problems, we shall use the naming convention from the following picture.



5. **Solving a triangle (SAS).**

   With the usual convention illustrated above, a triangle $\triangle ABC$ has the sides $b = 10, c = 15$. If the angle A between the two sides is $20°$ then determine all the sides and angles of the triangle.

[8] Some of you may be feeling that we are complicating matters too much by naming even the known angles as s, t. You will find that the reverse is true. Try using the explicit values in the equations and then try to solve them and compare the work!

Moreover, this way, you have a wonderful formula at hand. All such height problems can now be solved by changing the various numbers in the final expression!

Answer.

First we use the Cosine Law to find the third side. Using a calculator, we see that:

$$a^2 = b^2 + c^2 - 2bc\cos(A) \;=\; 100 + 225 - 2(10)(15)\cos(20^\circ) = 43.0922.$$

Thus $a = 6.5645$.

Now using the Sine Law, we get:

$$\frac{6.5645}{\sin(A)} = \frac{10}{\sin(B)} = \frac{15}{\sin(15)}.$$

Since we know A, we see that each ratio is the same as the first or 19.1933.

It remains to find B such that

$$19.1933 = \frac{10}{\sin(B)}.$$

Using arcsin function we get

$$B = \arcsin\left(\frac{10}{19.1933}\right) = 31.4004^\circ.$$

We know that the supplementary angle $180^\circ - 31.4004^\circ$ is also a candidate. How do we choose?

Here is the idea. If the supplementary angle were to be the answer, then its cosine would be **negative** and as a result, the Cosine Law would say that $b^2 = a^2 + c^2 - 2ac\cos(B) > a^2 + c^2$.

In our case $a = 6.5645, b = 10, c = 15$, so

$$b^2 = 100 < a^2 + c^2 = 43.0922 + 225.$$

Thus the angle B must be 31.4004°. It follows that the third angle $C = 180^\circ - 20^\circ - 31.4004^\circ = 128.5996^\circ$.

Note that this third angle has negative cosine and this can be verified by noting that $c^2 = 225 > a^2 + b^2 = 43.0922 + 100$.

6. **Solving a triangle (SSS).**

   With the usual convention illustrated above, suppose that we are give three sides of a triangle:

$$a = 10, b = 10, c = 15.$$

Determine all the angles.

Answer. We can use the Cosine Law to find any chosen angle.

Thus for the angle A, we use:

$$a^2 = b^2 + c^2 - 2bc\cos(A) \text{ or } 100 = 100 + 225 - 300\cos(A).$$

This leads to $\cos(A) = \frac{225}{300}$ and $A = \arccos(0.75) = 41.4096°$. [9]

Similarly, the angle B comes out to be $\arccos(\dfrac{a^2+c^2-b^2}{2ac}) = \arccos(\dfrac{225}{300})$ which is the same answer.

We should have noted this from $a = b$ any way.

The third angle C is best calculated as

$$180° - 41.4096° - 41.4096° = 97.1808°.$$

Warning. If you randomly give three positive numbers for a, b, c, then the formulas can collapse telling you that such a triangle does not exist.

For instance $a = 10, b = 15, c = 30$ is not a valid triangle as the calculation of $\cos(A)$ leads to

$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc} = \frac{1025}{900} > 1.$$

Since a cosine cannot be bigger than 1, this is impossible!

These lengths also violate the geometric theorem that the sum of two sides of a triangle must be greater than the third. $(10 + 15 < 30.)$ It is interesting to investigate what conditions on the three side lengths guarantee a valid triangle.

[9] The other possible angle is the negative of this angle and we need not worry about it in view of how the angles of a triangle are defined.

9.6 Proofs I.

Optional section. As we have already seen, there is yet no nice way to calculate sine and cosine values except for what the calculator announces. But these functions have been in use for well over 2000 years. What did people do before the calculators were born?
The technique was to develop an addition formula - a way of computing the functions for a sum of two angles if we know them for each angle. This single formula when used cleverly, lets us create a table of precise values for several angles and then we can interpolate (intelligently guess) approximate values for other angles. Indeed a table of 96 values was extensively used for several hundred years of astronomy and mathematics.
So, let us work on this proof. **You may find some of the arguments repeated from the earlier section, to make this section self contained.**
First, let us recall the

The Fundamental Identity $\sin^2(t) + \cos^2(t) = 1$

which is evident from the definition of the trigonometric functions.
One main idea is this. Given two points $P(s)$ and $P(t)$ on the unit circle, it is clear that the distance between them depends only on the difference of their angles, i.e. $(t - s)$. This is clear since the circle is such a symmetric figure that the distance traveled along the circle should create the same separation, no matter where we take the walk.
Thus, we can safely conclude that

$$(*) \qquad d(P(s), P(t)) = d(P(0), P(t - s)).$$

This is the only main point, the only thing that needs some thought!
The rest is pure algebraic manipulation, as we now show. We shall next evaluate the distance, simplify the expression and get what we want. Before we begin, let us agree to square both sides, since we want to avoid handling the square root in the distance formula.
Note that by definition $P(t) = (\cos(t), \sin(t))$. Also note that

$$P(0) = (\cos(0), \sin(0)) = (1, 0)$$

from the known position of $P(0)$.
Using similar expressions for points $P(s), P(t)$ and $P(t - s)$ in $(*)$, we get:

$$(\cos(t) - \cos(s))^2 + (\sin(t) - \sin(s))^2 = (\cos(t - s) - 1)^2 + (\sin(t - s) - 0)^2.$$

Expansion and rearrangement of the left hand side (LHS) gives:

$$LHS = \cos^2(t) + \sin^2(t) + \cos^2(s) + \sin^2(s) - 2(\cos(t)\cos(s) + \sin(t)\sin(s))$$

Using the fundamental identity, this simplifies to:

$$LHS = 1 + 1 - 2(\cos(t)\cos(s) + \sin(t)\sin(s)).$$

Similarly, the right hand side (RHS) becomes:

$$RHS = \cos^2(t-s) + \sin^2(t-s)) + (1+0) - 2(\cos(t-s)(1) + \sin(t-s)(0)).$$

After simplification, it gives:

$$RHS = 1 + 1 - 2(\cos(t-s)).$$

Comparing the two sides, we have our **addition formula**: [10]

Fundamental Identity 4. $\cos(t-s) = \cos(t)\cos(s) + \sin(t)\sin(s)$

9.7 Proofs II.

Optional section.

Having proved the important addition formula, we now turn attention to proving other formulas using it.
The fourth fundamental identity together with known values leads to a whole set of new identities. Deriving them is a fruitful, fun exercise. We have already learned to use most of these, but we have not seen the proofs. This section is designed to guide you through the proofs.
We are also not allowing the use of the Euler representation here, since we have not established it fully.
Thus, this is an intellectual exercise in proving formulas strictly from what we know so far.

1. Take $t = 0$ in the fourth fundamental identity and write:

   $$\cos(-s) = \cos(0)\cos(s) + \sin(0)\sin(s)$$

   and using the known values $\cos(0) = 1, \sin(0) = 0$ we get that

   Evenness of the cosine function. $\cos(-s) = \cos(s)$

   Note: Recall that function $y = f(x)$ is said to be even if $f(-x) = f(x)$ for all x in its domain. It is said to be an odd function, if, on the other hand $f(-x) = -f(x)$ for all x in the domain. Geometrically, the evenness is recognized by the symmetry of the graph of the function about the y-axis, while the oddness is recognized by its symmetry about the origin.

[10] It appears like a subtraction formula, but it will easily become converted to the addition formula soon!

2. We know that
$$\cos(\pi/2) = 0 \text{ and } \sin(\pi/2) = 1.$$
Take $t = \pi/2$ in the fourth fundamental identity and write:
$$\cos(\pi/2 - s) = (0)\cos(s) + (1)\sin(s) = \sin(s).$$
Replace s by $\pi/2 - s$ in this identity and get: [11]
$$\cos(\pi/2 - (\pi/2 - s)) = \sin(\pi/2 - s).$$
This simplifies to
$$\cos(s) = \sin(\pi/2 - s).$$

Definition: Complementary Angles. The angles s and $\pi/2 - s$ are said to be complementary angles. Indeed, in a right angle triangle, the two angles different from the right angle are complementary. But our definition is more general.

Thus we have established
Complementary Angle Identity.
$$\cos(\pi/2 - s) = \sin(s)\ ,\ \ \sin(\pi/2 - s) = \cos(s)$$

3. **Definition: Supplementary Angles.** The angles s and $\pi - s$ are said to be supplementary angles. We continue as above, except we now note that $P(\pi) = (-1, 0)$ and take $t = \pi$ in the fourth fundamental identity.

We get:
$$\cos(\pi - s) = (-1)\cos(s) + (0)\sin(s) = -\cos(s).$$
This is the first part of the **Supplementary Angle Identity.** Now we want to work out the formula for $\sin(\pi - s)$. This is rather tricky, since we have only a limited number of formulas proved so far. But here is how it goes:

$$\begin{array}{rcl|l} \sin(\pi - s) & = & \sin(\pi/2 + (\pi/2 - s)) & \text{Algebra} \\ & = & \sin(\pi/2 - (s - \pi/2)) & \text{Algebra} \\ & = & \cos(s - \pi/2) & \text{Complementary Angle Identity} \\ & = & \cos(\pi/2 - s) & \text{Evenness of the cosine function} \\ & = & \sin(s) & \end{array}$$

Thus we have established

Supplementary Angle Identity. $\cos(\pi - s) = -\cos(s)\ ,\ \ \sin(\pi - s) = \sin(s)$

[11]If it bothers you to plug in $\pi/2 - s$ for s, you may do it in two steps: First replace s by $\pi/2 - z$ and then replace z by s.

4. Now we replace t by $\pi/2 - t$ in the fourth fundamental identity to get:

$$\begin{aligned}
\cos((\pi/2 - t) - s)) &= \cos(\pi/2 - t)\cos(s) + \sin(\pi/2 - t)\sin(s) \\
\cos(\pi/2 - (t + s)) &= \sin(t)\cos(s) + \cos(t)\sin(s) \\
\sin(t + s) &= \sin(t)\cos(s) + \cos(t)\sin(s)
\end{aligned}$$

The second and third line simplifications are done using the complementary angle identities. We record the final conclusion as the fifth fundamental identity:

Fundamental Identity 5. $\sin(t + s) = \sin(t)\cos(s) + \cos(t)\sin(s)$

5. Now take $s = -t$ in the fifth fundamental identity to write:

$$\sin(t - t) = \sin(t)\cos(-t) + \cos(t)\sin(-t).$$

Notice that the left hand side is $\sin(0) = 0$. Since we know that $\cos(-t) = \cos(t)$ from the evenness of the cosine function, the equation becomes:

$$0 = \cos(t)(\sin(t) + \sin(-t)).$$

We deduce that we must have [12]

Oddness of the sine function. $\sin(-t) = -\sin(t)$

6. It is worth recording the identity:

Fundamental Identity 4-variant. $\cos(t + s) = \cos(t)\cos(s) - \sin(t)\sin(s)$

To do this, simply replace s by $-s$ in the fundamental identity 4 and use the oddness of the sine function.

Armed with the fundamental identities 4 and 5, we can give a whole set of new and useful formulas. We sketch partial argument for some and leave the rest to the reader. Indeed, it is recommended that the reader practices developing more by trying out special cases.
Here is a useful set:

[12] An alert reader might object that our conclusion is not valid since we have not carefully checked the values of t when $\cos(t) = 0$. This is a valid objection, but the only values of t with $\cos(t) = 0$ are given by $t = \pm\pi/2$ up to multiples of 2π. Each of these cases is easily verified by hand.

1. **Double angle formulas** Taking $s = t$ in the fundamental identities 4 and 5 we get two important identities:

   Fundamental Identity 6. $$\cos(2t) = \cos^2(t) - \sin^2(t)$$

   Fundamental Identity 7. $$\sin(2t) = 2\sin(t)\cos(t)$$

   The fundamental identity 6 has two variations and they both are very useful, so we record them individually.

   Fundamental Identity 6-variant 1. $$\cos(2t) = 2\cos^2(t) - 1$$

   Fundamental Identity 6-variant 2. $$\cos(2t) = 1 - 2\sin^2(t)$$

   These are simply obtained using the fundamental identity 1 to replace $\sin^2(t)$ by $1 - \cos^2(t)$ or $\cos^2(t)$ by $1 - \sin^2(t)$.

2. **Theoretical use of the double angle formula.** Remember the complicated proof we had to make to find the sine and cosine of 30°?

   The double and half angle formulas will make a short work of finding the sine and cosine for $\frac{\pi}{6} = 30°$ as well as $\frac{\pi}{3} = 60°$.

   Here is the work.

   For convenience of writing, let us:

   $$\text{use the notation } s \ = \ \frac{\pi}{6} \text{ and } t \ = \ \frac{\pi}{3}.$$

   Notice that
   $$s + t = \frac{\pi}{2} \text{ and } t = 2s.$$

   So by complementary angle identity, we get

   $$\sin(t) = \cos(s) \text{ and } \cos(t) = \sin(s).$$

   Since $t = 2s$, from the fundamental identity 7, we get $\sin(t) = 2\sin(s)\cos(s)$ and from the above we have $\sin(t) = \cos(s)$. Thus we have:

   $$\cos(s) = 2\sin(s)\cos(s).$$

Since $\cos(s)$ clearly is non zero, we cancel it to deduce

$$1 = 2\sin(s) \text{ which means } \sin(s) = \frac{1}{2}.$$

As before, from the fundamental identity 1 we deduce that

$$\cos(s) = \pm\sqrt{1 - \left(\frac{1}{2}\right)^2} = \pm\sqrt{\frac{3}{4}}.$$

And in our case we get

$$\cos(s) = \sqrt{\frac{3}{4}}$$

due to the position of the angle in the first quadrant. We have thus established:

$$\cos(60^\circ) = \sin(30^\circ) = \frac{1}{2} \text{ and } \cos(30^\circ) = \sin(60^\circ) = \frac{\sqrt{3}}{2}.$$

Due to their frequent use, these formulas are worth memorizing!

3. **Half Angle formulas.** These are easy but powerful consequences of the above variants of identity 6.

   Replace t by $\frac{t}{2}$ in variant 1 of the fundamental identity 6 to get

   $$\cos(t) = 2\cos^2\left(\frac{t}{2}\right) - 1$$

   and solve for $\cos(\frac{t}{2})$:

   Fundamental Identity 8. $$\cos\left(\frac{t}{2}\right) = \pm\sqrt{\frac{1+\cos(t)}{2}}$$

   Similarly, replace t by $\frac{t}{2}$ in variant 2 of the fundamental identity 6 to get

   $$\cos(t) = 1 - 2\sin^2\left(\frac{t}{2}\right)$$

   and solve for $\sin(\frac{t}{2})$:

   Fundamental Identity 9. $$\sin\left(\frac{t}{2}\right) = \pm\sqrt{\frac{1-\cos(t)}{2}}$$

   The $\pm$ sign comes due to the square root and has to be determined by the location of the angle $\dfrac{t}{2}$.

   The second one is similarly obtained from variant 2 of the identity 6.

Clearly, there is no end to the kind of formulas that we can develop. There is no point in "learning" them all. You should learn the technique of deriving them, namely convenient substitution and algebraic manipulation.

Chapter 10

Looking closely at a function

We have visited many functions $y = f(x)$ where $f(x)$ is given by a formula based on a rational function. We have also seen plane algebraic curves given by $F(x, y) = 0$ where it is not easy to explicitly solve for y in terms of x or even to express both x and y as nice functions of a suitable parameter.
It is, however, often possible to **approximate a curve near a point** by nice polynomial parameterization. We can even do this with a desired degree of accuracy. This helps us analyze the behavior of a function which may lack a good formula.

10.1 Introductory examples.

We shall begin by discussing two simple examples.

- **Analyzing a parabola.**

  Consider a parabola given by $y = 4x^2 - 2$. We shall analyze it near a point $P(1, 2)$. Like a good investigator, we shall first go near the point by changing our origin to P.

  As we know, this is easily done by the

  Substitution: $x = 1 + u\ ,\ y = 2 + v$ with u, v as the new coordinates.

  We shall then **treat** $\mathbf{u}, \mathbf{v}$ **as our local coordinates** and

  The original equation $\qquad y \ = \ 4x^2 - 2$

  becomes:

  $$(2 + v) = 4(1 + u)^2 - 2 \text{ or } 2 + v = 4(1 + 2u + u^2) - 2$$

  and this simplifies to:

  The local equation. $\qquad v = 4u^2 + 8u$

Since we are studying the equation near the point $u = 0, v = 0$ we rewrite the equation in increasing degrees in u, v as

$$(v - 8u) - 4u^2 = 0.$$

It is clear that when u, v are sufficiently small, the linear part $v - 8u$ is going to be much bigger relative to the quadratic part $-4u^2$. We say that the expression

$$v - 8u = y - 2 - 8(x - 1) = y - 8x + 6$$

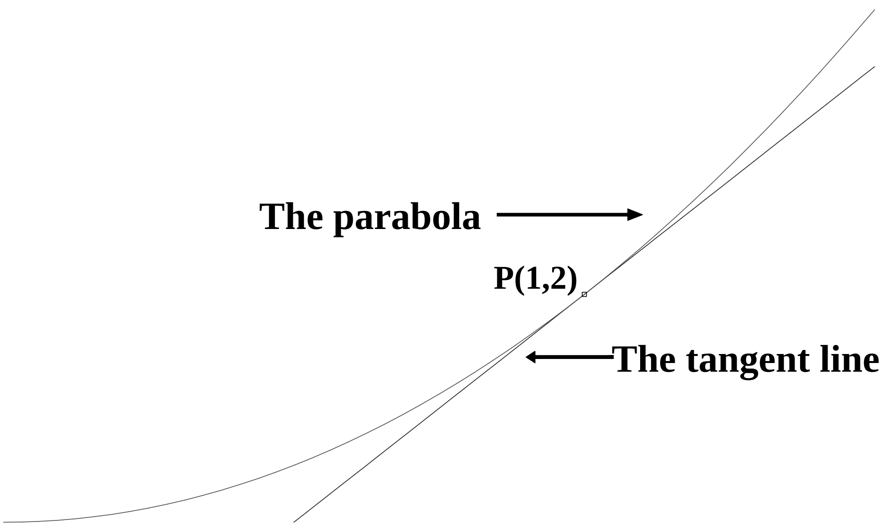
is a **Linear Approximation near the point** $P(1, 2)$ to the expression

$$y - 4x^2 + 2$$

which describes the parabola when set equal to zero.

We can express this alternatively by saying that the equation $y - 8x + 6 = 0$ is a linear approximation to the equation of the parabola near $P(1, 2)$.

The graph below illustrates the closeness between the parabola and its linear approximation near the point $P(1, 2)$.



The linear approximation to the equation of the parabola, namely $y - 8x + 6 = 0$ defines a line and is called the **tangent line** to the parabola at $P(1, 2)$.

Alternative organization of calculations.

There is a different way of doing the above calculation which avoids introducing new variables u, v but requires you to be more disciplined in your work.

For the reader's convenience, we present this alternative technique, which can be used in all the following calculations.

We shall illustrate this by analyzing $y = 4x^2 - 2$ at the same point $(1, 2)$.

We agree to rewrite:

$$x \text{ as } 1 + (x - 1) \text{ and } y \text{ as } 2 + (y - 2).$$

Then our parabola becomes:

$$2 + (y - 2) = 4(1 + 2(x - 1) + (x - 1)^2) - 2$$

which rearranges as

$$(y - 2) = (-2 + 4 - 2) + (4)(2)(x - 1) + 4(x - 1)^2 = 8(x - 1) + 4(x - 1)^2.$$

Thus we again get

$$(y - 2) = 8(x - 1) + \text{ higher degree terms in } (x - 1), (y - 2).$$

and the linear approximation is again:

$$(y - 2) - 8(x - 1) \text{ or } y - 8x - 6.$$

The reader may follow this alternate procedure if desired.

Note: Actually, here there are no higher degree terms involving $(y - 2)$. We have mentioned it sice such terms may need to be dropped in other examples.

Now, let us find the tangent to the same curve at another point while listing the various steps.

- **Choose point.** Choose the point with x-coordinate -1. Its y-coordinate must be $4(-1)^2 - 2 = 2$. Thus the point is $Q(-1, 2)$.
- **Shift the origin.** We move the origin to the point $Q(-1, 2)$ by

  $$x = -1 + u, y = 2 + v.$$

  The equation transforms to: [1]

  $$(2 + v) = 4(-1 + u)^2 - 2 \text{ or } v = (4(1 - 2u + u^2) - 2) - 2 \text{ i.e. } v = 4u^2 - 8u.$$

[1] Note that the constant term cancelled. If the constant term were to survive, it would indicate that either the chosen point is not on the curve or that there is a mistake in the simplification.

– **Get the linear approximation and the tangent line.** Write the equation in terms of increasing degree as

$$v + 8u - 4u^2 = 0.$$

The **linear approximation** is:

$$v + 8u \text{ or } (y - 2) + 8(x + 1) \text{ simplified to } y + 8x + 6.$$

Thus the **tangent line** at this point must be

$$v + 8u = 0 \text{ or } y + 8x + 6 = 0 \text{ or } y = -8x - 6.$$

Now let us do such **calculations** once and for all **for a general point** (a, b).

First make the **origin shift:** $x = a + u, y = b + v$. Then rearrange the equation:

$$y = 4x^2 - 2 \text{ becomes } b + v = 4(a + u)^2 - 2$$

After a little rearrangement, this becomes **simplified local equation:**

$$(b - 4a^2 + 2) + (v - 8au) - 4u^2 = 0.$$

Note that this final form tells us many things:

1. The point (a, b) is on our parabola if and only if the constant term vanishes. This means that $b - 4a^2 + 2 = 0$ or $b = 4a^2 - 2$.
2. If the point is on the parabola, then the **linear approximation is**

$$v - 8au = (y - b) - 8a(x - u) \text{ or } y - 8ax + 8au - b.$$

3. The **tangent line is**

$$v - 8au \ = \ 0 \text{ or } v \ = \ 8au.$$

In original coordinates, it becomes [2]

$$y - b = 8a(x - a) \text{ and is rearranged to } y = 8ax + (b - 8a^2).$$

4. The slope of the tangent line at the point (a, b) is $8a$ and we could simply write down its equation using the point slope formula.

Thus, we really only need to know the slope of the tangent 8a and we could write down the equation of the tangent line.

[2]The line $v = 8au$ is of interest even when the point is not on our curve. It is sometimes called a polar and has interesting properties for some special curves. Interested reader should look these up in books on algebraic geometry!

- **Analyzing a circle.**

  Now consider a circle $x^2 + y^2 = 25$ of radius 5 centered at the origin. Consider the point $P(3,4)$ on it. We shall repeat the above procedure and analyze the equation near P. So, **make the origin shift** $x = 3 + u, y = 4 + v$. The equation becomes:

$$(3+u)^2 + (4+v)^2 = 25 \text{ or } u^2 + 6u + 9 + v^2 + 8v + 16 = 25.$$

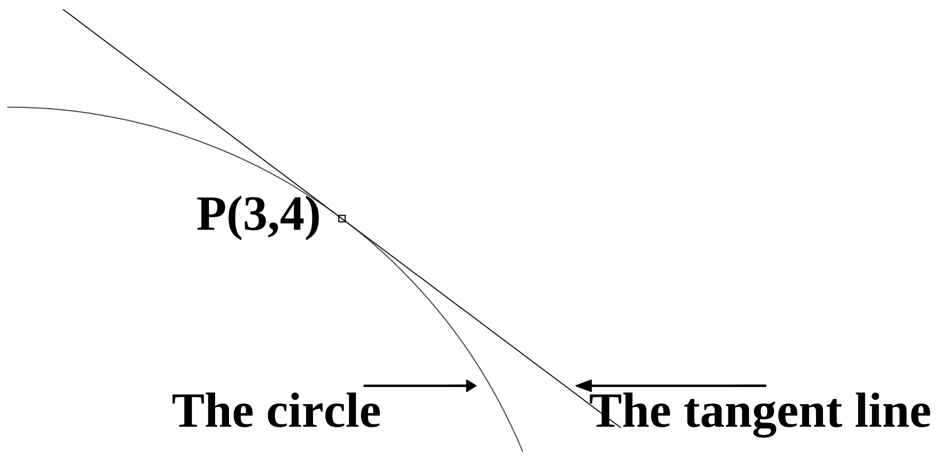
  When simplified and arranged by increasing degrees we get:

$$(6u + 8v) + (u^2 + v^2) = 0.$$

  [3] Thus we know that **the tangent line must be**

$$6u + 8v = 0 \text{ or } 6(x-3) + 8(y-4) = 0 \text{ or } y = -\frac{6}{8}x + \frac{50}{8}.$$

  Thus, the slope of the tangent line is $-\frac{3}{4}$ and the y-intercept is $\frac{25}{4}$.



  As before, **the linear approximation to the circle** near the point $P(3,4)$ is given by

$$6u + 8v = 6(x-3) + 8(y-4) = 6x + 8y - 50.$$

  Now we try to do this for a general point (a,b) of the circle. As before, **make the shift** $x = a + u, y = b + v$. Substitute and rearrange the equation of the circle:

$$x^2 + y^2 - 25 = 0 \text{ becomes } (a^2 + b^2 - 25) + 2(au + bv) + u^2 + v^2 = 0.$$

  As before, we note:

[3] Again, note the vanished constant term as a sign of correctness.

1. The point (a, b) is on our circle if and only if $a^2 + b^2 = 25$.
2. If the point is on the circle, then the tangent line is $2au + 2bv = 0$ or $au + bv = 0$ after dropping a factor of 2.

   In original coordinates, it is $a(x - a) + b(y - b) = 0$ and rearranged to $y = -\frac{a}{b}x + \frac{a^2+b^2}{b}$. This simplifies to: $y = -\frac{a}{b}x + \frac{25}{b}$.
3. The **slope of the tangent line** for the point (a, b) is $-\frac{a}{b}$ and we could simply write down the equation using the point slope formula.
4. Indeed, we note that the slope of the line joining the center $(0, 0)$ to the point (a, b) is $\frac{b}{a}$ and hence this line is perpendicular to our tangent with slope $-\frac{a}{b}$. Thus, we have essentially proved **the theorem** that the radius of a circle is perpendicular to its tangent. This, in fact, is well known for well over 2000 years, since the time of the Greek geometry at least.

10.2 Analyzing a general curve $y = f(x)$ near a point $(a, f(a))$.

Using the experience of the parabola, let us formally proceed to analyze a general function. We follow the same steps without worrying about the geometry.
We formally state the procedure and then illustrate it by a few examples.

1. Consider the point $P(a, b)$ on the curve $y = f(x)$ where, for now, we assume that $f(x)$ is a polynomial in x. Since $P(a, b)$ is on the curve, we have $b = f(a)$.
2. Substitute $x = a + u, y = b + v$ in $y - f(x)$ to get $b + v - f(a + u)$. Expand and rearrange the terms as: [4]

$$(b - f(a)) + (v - mu) + (\text{ higher degree terms in u,v }).$$

3. The constant term $b - f(a)$ is zero by assumption and we declare **the equation of the tangent line** to $y = f(x)$ at $x = a$ to be the linear part set to zero:

$$v - mu = 0 \text{ or } y - b = m(x - a) \text{ or, finally } y = mx + (b - ma).$$

[4] We are clearly implying that $f(a + u)$ expanded in powers of u looks like $f(a) + mu +$ higher u-terms . The fact that the constant term is $f(a)$ is obvious if you notice that the constant term must be obtained by putting $u = 0$ in the expansion and putting $u = 0$ in $f(a+u)$ clearly makes it $f(a)$.

We may also call it the tangent line at the point $(a, f(a))$ and the expression $v - mu$ or $y - b - m(x - a)$ is said to be a linear approximation at this point to the curve $y - f(x)$.

Definition: Linear approximation to a function. We now define the linear approximation of a function $f(x)$ near $x = a$ to be simply

$$f(a) + m(x - a).$$

Let us note that then

$$f(x) = \text{ linear approximation } + \text{ terms involving higher powers of } (x - a)$$

and we could simply take this as a direct definition of a linear approximation of a function without discussing the curve $y = f(x)$.

4. Thus we note that we only need to find the number m which can be defined as follows:

   Definition: The derivative of a polynomial function $\mathbf{f(x)}$ at a point a. A number m is the derivative of $f(x)$ at $x = a$ if $f(a + u) - f(a) - mu$ is divisible by u^2.

   Let us name the desired number m as something related to f and a and the customary notation is $\mathbf{f'(a)}$, **a quantity to be derived from f and a, or the so-called derivative of f at $\mathbf{x = a}$.**

   This gives us a better way of remembering and using our definition of the derivative:

   The derivative $f'(a)$ is defined as that real number for which

$$f(a + u) = f(a) + f'(a)u + \text{ terms divisible by } u^2.$$

Here are some examples of the above procedure.

1. Let $f(x) = p$, where p is a constant. Then we have:

$$f(a + u) = p = f(a) + 0$$

   so our m must be 0.

2. Let $f(x) = px$, where p is a constant. Then we have:

$$f(a + u) = p(a + u) = pa + pu = f(a) + pu$$

   so our m is simply p.

3. $f(x) = px^2$, where p is a constant. Then we have:

$$f(a+u) = p(a+u)^2 = pa^2 + (2pa)u + (p)u^2 = f(a) + 2pau + u^2(p)$$

   so our m is simply $2pa$.

4. $f(x) = px^3$, where p is a constant. Then we have:

$$f(a+u) = p(a+u)^3 = pa^3+(3pa^2)u+(3pa)u^2+(p)u^3 = f(a)+3pa^2u+u^2(3pa+pu)$$

   so our m is simply $3pa^2$.

Clearly we can go on doing this, but it involves more and more calculations. We need to do it more efficiently. The true power of algebra consists of making a good definition and recognizing a pattern!

10.3 The slope of the tangent, calculation of the derivative.

As we know, given a function $f(x)$ we wish to substitute $a+u$ for x and rearrange the expanded terms $f(a+u)$ as $f(a) + mu +$ higher degree terms .
A more convenient notation for the derivative. We note that the parameter a used in the above analysis is just a convenient place holder and $f'(a)$ is really a function of a. Hence, there is no harm in using the same letter x as a variable name again.
Thus, for a function $f(x)$ we will record its derivative $f'(a)$ as simply $f'(x)$. As an example, note that our calculation for the parabola $y = f(x) = 4x^2 - 2$ with $f'(a) = 8a$ can thus be recorded as $f'(x) = 8x$.
Finally, we introduce some shortcuts to our notation.

1. We will find it convenient not to introduce the function name f and would like to denote the derivative directly in terms of the symbol y. We would, thus prefer to write $y' = 8x$. If the mark $'$ gets confusing we shall use the notation: $D_x(y) = 8x$, where the symbol D denotes the word derivative and the subscript x reminds us of the main variable that y is a function of.

   Thus, our earlier results for the derivatives of px^n for $n = 1, 2, 3$ can be restated as

$$D_x(px) = p \ , \ D_x(px^2) = 2px \ , \ D_x(px^3) = 3px^2.$$

   We formally declare that we shall use **the notation:** $D_x(y)$ to denote the derivative $f'(x)$ of the function $y = f(x)$ even though we have no explicit name f declared.

2. We note that there is yet another popular notation $\frac{d}{dx}(y)$ or $\frac{dy}{dx}$ for writing the derivative, but we will mostly use the D_x notation. [5]

3. When we start with a function $f(x)$ and take its derivative, it is a new function $f'(x)$. When we take its derivative, it is **the second derivative** of the original $f(x)$ and can be denoted by $f''(x)$. You can imagine $f'''(x)$, $f''''(x)$.

   We can keep on doing this, but clearly the string of $'$s is hard to read and keep track of.

   So we introduce these notations:

   $$D_x(f)(x) = f'(x), D_x^2(f)(x) = f''(x), D_x^3(f)(x) = f'''(x) \text{ and so on .}$$

   Thus, for example, when $f(x) = x^3$, we have:

   $$D_x(f)(x) = 3x^2, D_x^2(f)(x) = 6x, D_x^3(f)(x) = 6, D_x^4(f)(x) = 0 = D_x^5(f)(x) = \cdots .$$

   The symbol $D_x^n(f)$ is said to be the n-th derivative of the function f.

4. Finally, if the variable x is clear from the context, we may write D in place of D_x, giving us a much simpler notation like $D^n(f(x))$ or equivalently $D^n(f)(x)$.

We begin by making some very simple observations:

Formula 1 **The constant multiplier.**

If c is a constant and $g(x) = cf(x)$, then $g'(a) = cf'(a)$. **In better notation:**

$$D_x(cy) = cD_x(y).$$

Proof: Write $f(a+u) = f(a) + f'(a)u + u^2(\text{ more terms })$. Multiplying by c we see that $cf(a+u) = cf(a) + cf'(a)u + u^2(\text{ more terms })$ which says that $g(a+u) = g(a) + cf'(a)u + u^2(\text{ more terms })$ and hence we get $g'(a) = cf'(a)$.

Formula 2 **The sum rule.** Suppose that for two functions $f(x), g(x)$ we have the derivatives calculated.

Then

$$(f+g)'(x) = f'(x) + g'(x).$$

[5] The definition looks very much like the definition in Calculus, but our case is much simpler.

In Calculus one needs to handle more complicated functions and hence needs a much more sophisticated analysis of the linear approximation involving the concept of limits. We shall advance up to the derivatives of rational functions and some special functions and leave the finer details to higher courses.

Historically, people knew the derivatives for such special functions including the trigonometric functions long before the formal invention of Calculus.

In better notation:

$$D_x(y + z) = D_x(y) + D_x(z).$$

Proof. We have:

$$f(a + u) = f(a) + f'(a)u + u^2(\cdots) \text{ and } g(a + u) = g(a) + g'(a)u + u^2(\cdots).$$

Then we clearly have:

$$(f + g)(a + u) = f(a + u) + g(a + u) = f(a) + g(a) + (f'(a) + g'(a))u + u^2(\cdots).$$

This proves that

$(f + g)'(a) = f'(a) + g'(a)$ or, in compact notation $D_x(y + z) = D_x(y) + D_x(z)$.

Formula 3 **The product rule.** As above, we assume that the derivatives $f'(a), g'(a)$ are known. We now consider the product function $h(x) = f(x)g(x)$.

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

In better notation:

$$D_x(yz) = yD_x(z) + D_x(y)z.$$

Proof. We have:

$$h(a+u) = f(a+u)g(a+u) = \left[f(a) + f'(a)u + u^2(\cdots)\right]\left[g(a) + g'(a)u + u^2(\cdots)\right].$$

The right hand side simplifies to:

$$f(a)g(a) + [f(a)g'(a) + f'(a)g(a)]\, u + u^2(\cdots).$$

Thus we get the product rule:

$$h'(a) = (fg)'(a) = f(a)g'(a) + f'(a)g(a)$$

or, in compact notation:

$$D_x(yz) = yD_x(z) + D_x(y)z.$$

We illustrate the use of this formula by calculating the derivative of $h(x) = x^4$. Write $h(x) = g(x)g(x)$ where $g(x) = x^2$. Note that we have already determined that $g'(a) = 2a$. Then by the product formula we get $h'(a) = g(a)g'(a) + g'(a)g(a) = a^2(2a) + (2a)a^2 = 4a^3$. Remember, we record this conveniently as

$$h'(x) = D_x(x^4) = 4x^3.$$

We generalize this in the next formula.

Formula 4 **The general power of x.** If $f(x) = x^n$ where $n = 1, 2, \cdots$, then $f'(x) = nx^{n-1}$.

In better notation:

$$D_x(x^n) = nx^{n-1} \text{ for } n = 1, 2, \cdots .$$

Proof. We already know this for $n = 1, 2, 3, 4$ from above. We now assume that we have handled all the values of n up to some k and show how to do the next value of $k+1$. [6]

Thus we assume that $f(x) = x^{k+1} = (x)(x^k) = g(x)h(x)$ say. By what we have assumed, we know that $g'(x) = 1$ and $h'(x) = kx^{k-1}$. Then the product formula gives

$$f'(x) = g(x)h'(x) + g'(x)h(x) = x(kx^{k-1}) + (1)x^k = (k+1)x^k.$$

Thus we have proved our formula for the next value of $k+1$.

We are done with the induction!

Formula 5 **The combined formula.** Consider a polynomial in x with constant coefficients $p_0, p_1, \cdots, p_n$:

$$f(x) = p_0x^n + p_1x^{n-1} + \cdots + p_{n-1}x + p_n$$

Then its derivative is given by

$$f'(x) = np_0x^{n-1} + (n-1)p_1x^{n-2} + \cdots + p_{n-1}$$

In compact notation:

$$D_x(p_0x^n + p_1x^{n-1} + \cdots + p_{n-1}x + p_n) = np_0x^{n-1} + (n-1)p_1x^{n-2} + \cdots + p_{n-1}.$$

This is easily seen by combining the formulas 1,2,4.

Thus if

$$f(x) = 3x^4 - 5x^2 + x - 5$$

then

$$f'(x) = (4)(3)x^3 - (2)(5)x + 1 = 12x^3 - 10x + 1.$$

[6] This is the famous method of Mathematical Induction. It gives concrete minimal argument necessary to give a convincing proof that a statement holds for all values of integers from some point on. There are many logically equivalent ways of organizing the argument. These different forms of the argument often reduce the task of proving a formula to just guessing the right answer and applying the mechanical process of induction. All of our formulas for the summation of series in the last chapter, can be easily proved by this technique and the reader is encouraged to try this.

10.4 Derivatives of more complicated functions.

Optional section. The proofs in this section can be omitted in a first reading. Many of the formulas can serve as shortcuts, but can be avoided, if you are willing to work a bit longer!

Even though we defined the derivatives only for a polynomial function so far, we have really worked them out for even more complicated functions, without realizing it!

Just consider our tangent line calculation for the circle where the explicit function would have involved a square root of a quadratic function! ($y = \sqrt{25 - x^2}$).

Inspired by our calculation with the circle, we shall attempt to calculate the derivative of $\frac{1}{x}$.

We begin by setting $y = \frac{1}{x}$ and simplifying it into a polynomial equation $xy = 1$.

Following earlier practice, let us substitute $x = a + u, y = b + v$ and simplify the resulting expression:

$$xy - 1 = (a + u)(b + v) - 1 = (ab - 1) + (av + bu) + (uv).$$

Thus the point $P(a, b)$ is on the curve if and only if $ab - 1 = 0$, i.e. $b = \frac{1}{a}$ and we get:

$$\text{The tangent line } av + bu \ = \ a(y - b) + b(x - a) = ay + bx - 2ab = 0.$$

Thus the tangent line at $x = a, y = \frac{1}{a}$ is given by

$$ay + \frac{x}{a} = 2 \text{ or } y = \frac{2}{a} - \frac{1}{a^2}x.$$

Thus the slope of the tangent is $-\frac{1}{a^2}$ and it makes sense to declare that

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

The reader should note that formally, it follows the same pattern as our earlier power formula.

Would this follow the original definition of the derivative in some new sense? Yes indeed! Let us set $f(x) = \frac{1}{x}$. Consider (and verify) these calculations:

$$f(a + u) = \frac{1}{a + u} = \frac{1}{a} + \frac{-u}{a^2 + au} = -\frac{1}{a^2} \cdot u \cdot \frac{1}{a + u/a}.$$

Thus, we have, after further manipulation:

$$f(a + u) = f(a) + \left(-\frac{1}{a^2}\right) u + \frac{u^2}{(a + u)a^2}.$$

The last term, is u^2 times a rational function (instead of a polynomial) and it does not have u as a factor of the denominator. Thus we simply have to modify our definition of the derivative to allow for such complications:

Enhanced definition of the derivative of a rational function f(x) at a point a. A number m is the derivative of $f(x)$ at $x = a$ if $f(a+u) - f(a) - mu$ is equal to u^2 times a rational function of u which **does not have** u as a factor of its denominator.
Of course, this definition is only needed as a formality; our calculation with the polynomial equation $xy = 1$ had given us the answer neatly anyway!
So, let us make a
Final definition of the derivative of an algebraic function. Suppose we have a polynomial relation $f(x, y) = 0$ between variables x, y. Suppose that $P(a, b)$ is a point of the resulting curve so that $f(a, b) = 0$. Further, suppose that the substitution $x = a + u, y = b + v$ results in a simplified form

$$f(a+u, b+v) = ru + sv + \text{ terms of degree 2 or more in } (u, v) \ ,$$

where r, s are constants such that at least one of r, s is non zero. Then

$$ru + sv = r(x-a) + s(y-b)$$

is said to be a linear approximation of $f(x, y)$ at $x = a, y = b$. The line $r(x-a) + s(y-b) = 0$, is defined to be the tangent line to the curve $f(x, y) = 0$ at $P(a, b)$.
Finally, if $s \neq 0$ then y can be described as a well defined algebraic function of x near the point $P(a, b)$ and we can define $D_x(y) = -\frac{r}{s}$ or the slope of the tangent line. In case $s = 0$, sometimes we may be still able to define an algebraic function, but it will not be well defined.
The above paragraph leaves many details out and the interested reader in encouraged to consult higher books on algebraic geometry to learn them. The aim of the calculations and the definition is to illustrate how mathematics develops. We do some simple calculations which lead to possibly new ideas for functions which we did not have before. In this manner, many new functions are born.

10.5 General power and chain rules..

We give proofs of two powerful rules. The student should learn to use these before worrying about the proof.
Armed with these new ideas, let us push our list of formulas further.

Formula 6 **Enhanced power formula.** Can we handle a function like $y = x^{\frac{2}{3}}$? Yes, indeed.

We shall show: $D_x(x^n) = nx^{n-1}$ when n is any rational number. [7]

[7] In fact, the result is true for any constant exponent, but that proof is well beyond our reach at this point!

Proof:[8] First suppose $n = \frac{p}{q}$ where p, q are positive integers. Let $y = x^n$. Consider a polynomial $f(X, Y) = Y^q - X^p$. Substituting x for X and y for Y, note that:

$$f(x, y) = 0.$$

We know that $D_x(x^p) = px^{p-1}$, so

$$(a + u)^p = a^p + pa^{p-1}u + u^2(\text{ higher degree terms in } u\ )$$

Similarly, we must have $D_y(y^q) = qy^{q-1}$ and hence:

$$(b + v)^q = b^q + qb^{q-1}v + v^2(\text{ higher degree terms } v\ )$$

It follows that

$$\begin{aligned} f(a+u, b+v) = (b+v)^q - (a+u)^p &= (b^q - a^p) + qb^{q-1}(v) - pa^{p-1}(u) \\ &\quad +(\text{ terms of degree 2 or more in } (u, v)\ ). \end{aligned}$$

Thus (a, b) is a point on the curve if and only if $a^p = b^q$ and the tangent line is given by

$$qb^{q-1}(v) - pa^{p-1}(u) = 0 \text{ or } qb^{q-1}(y - b) - pa^{p-1}(x - a) = 0.$$

Its slope is $\dfrac{pa^{p-1}}{qb^{q-1}}$ which we aim to simplify.

Note that since $a^p = b^q$.

$$\frac{pa^{p-1}}{qb^{q-1}} = \frac{p}{q}\frac{a^{-1}}{b^{-1}} = n\frac{b}{a}.$$

Finally, since $b = a^n = a^{\frac{p}{q}}$, we get

$$\frac{b}{a} = a^{\frac{p}{q}-1} = a^{n-1}.$$

Thus the slope of the tangent is: na^{n-1}.

In particular, we get that $D_x(y) = nx^{n-1}$ as required! Our claim is thus proved for all positive fractional powers.

We now think of negative powers, i.e. we assume that $y = x^{-n}$ where n is a positive rational number. Set $z = x^n$ and note that $yz = 1$.

[8] The proof is easy, but messy in appearance. The student is advised to learn its use before worrying about the details!

Our product rule (Formula 3) says that $D_x(1) = yD_x(z) + D_x(y)z$ or from known facts

$$0 = ynx^{n-1} + D_x(y)x^n.$$

When solved for $D_x(y)$, we get [9]

$$D_x(y) = \frac{-ynx^{n-1}}{x^n} = -\frac{x^{-n}nx^n}{x^n} = -\frac{n}{x^n}.$$

Thus, we have

$$D_x(x^{-n}) = -nx^{-n}.$$

We thus have the: **General power rule.**

$$D_x(x^n) = nx^{n-1} \text{ where } n \text{ is any rational number.}$$

We should point out that the function x^n has definition problems when $x \leq 0$ and our derivatives usually work only for positive x unless n is special.

Formula 7 **The chain rule.** We now know how to find the derivative of $g(x) = x^2 + x + 1$ and also we know how to find the derivative of $f(x) = x^5$. Can we find the derivative of $f(g(x)) = (x^2 + x + 1)^5$?

Of course, we could expand it out and use our combined formula, but that is not so pleasant!

It would be nice to have a simple mechanism to write the derivative of a composite function, say $f(g(x))$ where f, g are polynomials. Let us work it out.

Write

$$g(a + u) = g(a) + mu + u^2w$$

where we note that $m = g'(a)$ and w contains remaining terms. For convenience, set

$$h = mu + u^2(w),$$

so we get:

$$f(g(a + u)) = f(g(a) + h)$$

and then by definition of the derivative of $f(x)$ we can write

$$f(g(a + u)) = f(g(a)) + m^*h + h^2r$$

[9] An alert reader should seriously object at this point. We proved the product rule, but the definition of the derivative was only for polynomials at that time! A general product rule can be established, and we are not really going to do it here. We shall leave such finer details for higher courses.

where $m^* = f'(g(a))$ and r contains the remaining terms. Note that h is divisible by u and hence the part $h^2 r$ is divisible by u^2. Thus we can write $h^2 r = u^2 s$ for some s.

Combining terms, we write:

$$f(g(a+u)) = f(g(a)) + m^*(mu + u^2 w) + u^2 s = f(g(a)) + m^* mu + u^2(m^* w + s)$$

and hence we deduce that the derivative of $f(g(x))$ at $x = a$ is

$$m^* m = f'(g(a))g'(a).$$

Thus we write

$$D_x(f(g(x)) = f'(g(x))g'(x).$$

For example, with our $f(x) = x^5$ and $g(x) = x^2 + x + 1$ as stated above we get

$$D_x((x^2 + x + 1)^5) = f'(x^2 + x + 1)g'(x) = 5(x^2 + x + 1)^4(2x + 1).$$

For instance, at $x = 1$ we have the derivative of $(x^2 + x + 1)^5$ equal to

$$5(1 + 1 + 1)^4(2 + 1) = (5)(81)(3) = 1215$$

and hence the tangent line to $y = (x^2 + x + 1)^5$ at $x = 1$ would have the equation

$$y - (1 + 1 + 1)^5 = (1215)(x - 1)$$

simplified to $y = 1215x + 243 - 1215$ or further simplified to $y = 1215x - 972$.

As before, if our functions get more complicated, the chain rule is rather subtle to prove and we need to be careful in providing its proof.

Let us, however, note the following **rule for the derivative of a power of a function** that can be deduced from the above example:

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

When the theory is completely developed, this formula will work for any n for which $f(x)^n$ can make sense.

10.6 Using the derivatives for approximation.

As we discussed in the introductory section on functions, in practice, the calculation of $f(a)$ for a given function $f(x)$ can be difficult and even impossible for real life functions given only by a few data points. If for some reason, we know the function to be linear, then obviously it is easy to calculate its values quickly.

If a function is behaving nicely near a point, we can act as if it is a linear function near the point and hence can take advantage of the ease of calculation (at the cost of some accuracy, of course). In the days before calculators, many important functions were often used as tables of values at equally spaced values of x and for values in between then, one used an "interpolation", i.e. guess based on the assumption that the function was linear in between.
Even when modern calculators or computers give a value of the function to a desired accuracy, they are really doing higher order versions of linear approximations and doing them very fast; thus giving the appearance of a complete knowledge of the function. We will illustrate some of the higher approximations below.
Examples of linear approximations.

1. **Find an approximate value of $\sqrt{3.95}$ using linear approximation.**

   Since we are discussing a square root, consider the function

   $$f(x) = \sqrt{x} = x^{1/2}.$$

   We know $f(4) = 2$ precisely. So we find a linear approximation to $f(x)$ near $x = 4$.

   We have already figured out that

   $$f'(x) = D_x(x^{\frac{1}{2}}) = \frac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}},$$

   using the enhanced power formula.

   Thus **the linear approximation**

   $$L(x) = f(a) + f'(a)(x - a) \text{ becomes}$$

   $$L(x) = \sqrt{4} + (\frac{1}{2\sqrt{4}}(x - 4)) \text{ or } L(x) = 2 + \frac{x - 4}{4}.$$

   Thus the estimated value of $\sqrt{3.95}$ using the linear estimate is

   $$L(3.95) = 2 + \frac{3.95 - 4}{4} = 1.9875.$$

   Note that any reasonable calculator will spit out the answer 1.987460691 which is certainly more accurate, but not as easy to get to! Considering the ease of getting a value correct to 4 decimal places, our method is well justified!

   Technically, we can use the linear approximation to calculate any desired value, but clearly it only makes sense near a known good value of x.

2. **Improving a known approximation.** Suppose we start with the above approximation 1.9875 which is not quite right and its square is actually 3.95015625. Suppose we use $a = 3.95015625$ and then $f(a) = \sqrt{a} = 1.9875$. Then we get a new linear approximation

$$L_1(x) = 1.9875 + \frac{x - 3.95015625}{3.9750}.$$

If we now set $x = 3.95$ and calculate the linear approximation, we get

$$1.9875 + \frac{3.95 - 3.95015625}{3.9750} = 1.987460692,$$

an even better value!

Thus with a repeated application of the linear approximation technique, one can get fairly accurate values. [10]

Indeed, for the square root function, there is a classical algorithm known for over 2000 years to find successive decimal digits to any desired degree of accuracy. This algorithm used to be routinely taught in high schools before the days of the calculators.

A similar but more complicated algorithm exists for cube roots as well.

3. **Estimating polynomials: Introduction to higher approximations.** Suppose that we are trying to analyze a polynomial function $f(x)$ near a point $x = a$. In principle, a polynomial is "easy" to calculate precisely, but it may not be practical for a large degree polynomial. So, we might consider approximation for it.

Let us discuss a concrete example.

Consider $f(x) = 4x^4 - 5x^3 + 3x^2 - 3x + 1$. Let us take $a = 1$ and note that $f(a) = f(1) = 4 - 5 + 3 - 3 + 1 = 0$. What is $f(a + u)$?

The reader should calculate:

$$f(x) = f(a + u) = f(1 + u) = 4u^4 + 11u^3 + 12u^2 + 4u.$$

Thus for value of x near $a = 1$, we have $u = x - 1$ is small and hence the function $f(x)$ can be approximated by various degree polynomials. The Linear approximation will be $L(x) = 4u = 4(x - 1)$. The quadratic approximation will be

$$Q(x) = 12u^2 + 4u = 12(x - 1)^2 + 4(x - 1) = 8 - 20x + 12x^2.$$

[10] There are deeper issues of accumulation of roundoff errors during all calculations and reliability of the final conclusions. This topic is studied in the branch of mathematics called numerical analysis.

The cubic approximation shall be

$$C(x) = 11u^3+12u^2+4u = 11(x-1)^3+12(x-1)^2+4(x-1) = -3+13x-21x^2+11x^3.$$

The quartic (degree four) approximation will become the function itself!

Can we calculate the necessary approximations efficiently, without the whole substitution process?

The answer is yes for polynomial functions, since we have already studied the Binomial Theorem.

4. Let $f(x) = x^{12}$. What is a linear approximation near $x = a$? We set $x = a + u$ and we know that:

$$f(x) = f(a+u) = a^{12} + 12C_1 a^{11} u + \text{ terms with higher powers of } u \ .$$

By our known formula for the binomial coefficient, we get: $12C_1 = \frac{12}{1!} = 12$ and so the linear approximation is

$$L(x) = a^{12} + 12a^{11}u \text{ or alternatively } L(x) = a^{12} + 12(x-a).$$

As above, for a quadratic approximation, we can keep the next term and use:

$$Q(x) = a^{12}+12a^{11}(x-a)+\frac{12 \cdot 11}{2!}a^{10}(x-a)^2 = a^{12}+12a^{11}(x-a)+66a^{10}(x-a)^2.$$

Evidently, we can continue this, if needed.

5. **Higher Approximations.** Given a function $f(x)$ let us analyze the process of approximating it with a polynomial of a convenient degree near $x = a$.

Recall that for a given function $f(x)$ we can talk about its n-th derivative $D_x^n(f)(x)$. Since it is a function of x, we may be able to substitute $x = a$ in it, to get $D_x^n(f)(a)$.

Suppose we expect the approximating polynomial to be a polynomial of degree n. Then we expect it to be of the form:

$$T(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + \cdots + a_n(x-a)^n.$$

In analogy with our linear approximation, we shall say that $T(x)$ is an **approximation of order n** to $f(x)$, if we have:

$$f(a) = T(a), D_x(f)(a) = D_x(T)(a), \cdots, D_x^n(f)(a) = D_x^n(T)(a).$$

In other words, the approximating polynomial $T(x)$ has the same value as $f(x)$ at $x = a$ and also its first n derivatives have the same values as the corresponding n derivatives of $f(x)$ at $x = a$.

Let us find out what it says about the coefficients.

From the first condition, we see that:

$$f(a) = T(a) = a_0.$$

Note that

$$D_x(T)(x) = a_1 + 2a_2(x-a) + \cdots + na_n(x-a)^{n-1}.$$

So, from the second condition, we see that:

$$D_x(f)(a) = D(T)(a) = a_1.$$

Similarly, note that

$$D_x^2(T)(x) = 2a_2 + (3)(2)a_3(x-a)\cdots + (n)(n-1)a_n(x-a)^{n-2}.$$

So, from the third condition, we see that:

$$D_x^2(f)(a) = 2a_2 = 2!a_2.$$

Continuing this way, we can deduce that:

$$a_0 = f(a), a_1 = D_x(f)(a), a_2 = \frac{D_x^2(f)(a)}{2!}, a_3 = \frac{D_x^3(f)(a)}{3!}, \cdots, a_n = \frac{D_x^n(f)(a)}{n!}.$$

This gives us the famous **Taylor Formula** for n-th order approximation:

$$T(x) = f(a)+D_x(f)(a)(x-a)+\frac{D_x^2(f)(a)}{2!}(x-a)^2+\frac{D_x^3(f)(a)}{3!}(x-a)^3+\cdots+\frac{D_x^n(f)(a)}{n!}(x-a)^n.$$

The resulting polynomial expressed in powers of $(x-a)$ is said to be the Taylor Polynomial of n-th order at $x = a$.

Let us redo the case of $f(x) = x^5$ and find the Taylor Formula for its 3-rd order at $x = 2$.

Here are the steps. $f(2) = 2^5$. So, $a_0 = 2^5 = 32$.

$D_x(f)(x) = 5x^4$. So, $a_1 = D_x(f)(2) = 5(2^4) = 80$.

$D_x^2(f)(x) = 5(4)x^3 = 20x^3$. So, $a_2 = \frac{D_x^2(f)(2)}{2!} = \frac{20(2)^3}{2} = 80$.

$D_x^3(f)(x) = 5(4)(3)x^2 = 60x^2$. So, $a_3 = \frac{D_x^3(f)(2)}{3!} = \frac{60(2^2)}{6} = 40$.
Thus the Taylor Formula gives:

$$T(x) = 32 + 80(x-2) + 80(x-2)^2 + 40(x-2)^3.$$

It is instructive to compare the graphs of $y = T(x)$ and $y = x^5$ near $x = 2$ to notice their closeness. We also recommend using the substitution $x = 2 + u$ and expanding $x^5 = (2+u)^5$ by the Binomial Theorem. If we keep the terms of degree at most 3 in u and substitute $u = x - 2$, then we get the same answer.

Chapter 11

Root finding

11.1 Newton's Method

We now find a new use for the idea of linear approximation, namely given a polynomial $f(x)$ we try to find its root, which means a value $x = a$ such that $f(a) = 0$.
For a linear polynomial $f(x) = px + q$ this is a triviality, namely the answer is $x = -\frac{q}{p}$.
Note that for higher degree polynomials, clearly there could be many answers, as the simple polynomial $x^2 - 1 = (x-1)(x+1)$ illustrates. Indeed, if we can factor the polynomial $f(x)$ into linear factors, then it is easy to find the roots; they are simply the roots of the various factors. In general, it is hard to find exact roots of polynomials with real coefficients, unless we are lucky. Moreover, polynomials like $x^2 + 1$ show that there may not even be any roots!
What we are about to describe is a simple idea to find the roots approximately. The idea goes like this:

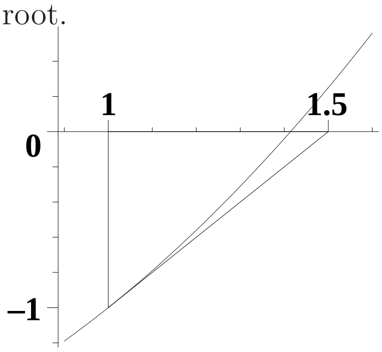
- Choose a starting point $x = a$ suspected to be close to a root. (For instance, you may note $f(a)$ to be a small number.)
- Take the linear approximation $L(x)$ at this $x = a$ for the given function $f(x)$.
- Find the root of $L(x)$, call it a_1.
- Now take $x = a_1$ as a new convenient starting point point and repeat!
- Under reasonable conditions, you expect to land very close to a root, after a few iterations of these steps.

Let us try an example:
Let $f(x) = x^2 - 2$. By algebra, we know that $x = \sqrt{2}$ is a root, but to write $\sqrt{2}$ as a decimal number is not a finite process. Its decimal expansion continues forever,

without any repeating pattern. So, we are going to try and find a good decimal approximation, something like what a calculator will spit out, or may be even better! We know that $\sqrt{2}$ is between 1 and 2. So, let us start with $a = 1$. Note that $f(x) = x^2 - 2$ and hence $f'(x) = 2x$.
Then near $x = 1$ we have:

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

The root of this $L(x)$ is clearly $x = 3/2 = 1.5$ So now we take $a_1 = 1.5$ and repeat the process! Here is a graph of the function, its linear approximation and the new root.



Thus the linear approximation near $x = a_1 = 1.5$ is

$$L(x) = f'(1.5)(x-1.5) + f(1.5) = (2(1.5))(x-1.5) + ((1.5)^2 - 2) = 3x - 4.25.$$

The root of this $L(x)$ is now $\frac{4.25}{3} = \frac{17}{12} = 1.4167$. This is now our a_2. [1]
The linear approximation near a_2 is

$$L(x) = f'(1.4167)(x-1.4167) + f(1.4167) = 2.8334x - 4.007.$$

Thus a_3 will be 1.4142.
If you repeat the procedure and keep only 5 digit accuracy, then you get the same answer back!
Thus, we have found the desired answer, at least for the chosen accuracy!
Let us summarize the technique and make a formula.

1. Given a function $f(x)$ and a starting value a, the linear approximation is:

$$L(x) \;=\; f'(a)(x-a) + f(a) \text{ and hence it has the root } x = a - f(a)/f'(a).$$

2. Hence, if $f'(a)$ can be computed, replace a by $a - \dfrac{f(a)}{f'(a)}$.

3. Repeat this procedure until $f(a)$ or better yet $\dfrac{f(a)}{f'(a)}$ becomes smaller than the desired accuracy.

[1] The reader may try a similar picture. The curve becomes very close to the tangent line and the picture tells you very little, unless it is vastly expanded! It is already a sign that we are very close.

4. When this happens, declare a to be a root of $f(x)$ to within the desired accuracy.

Let us redo our calculations above, using this procedure. Note that for our $f(x) = x^2 - 2$ we have $f'(x) = 2x$ and hence $\frac{f(a)}{f'(a)}$ evaluates to $\frac{a^2-2}{2a}$. Thus it makes sense to calculate a nice table as follows: [2]

$$\begin{bmatrix} a & y = a^2 - 2 & y' = 2a & \frac{y}{y'} = \frac{a^2-2}{2a} & a - \frac{y}{y'} \\ 1.0 & -1.0 & 2.0 & -0.50000 & 1.5000 \\ 1.5000 & 0.25000 & 3.0 & 0.083333 & 1.4167 \\ 1.4167 & 0.0070389 & 2.8334 & 0.0024843 & 1.4142 \\ 1.4142 & -0.00003836 & 2.8284 & -0.000013562 & 1.4142 \end{bmatrix}$$

11.2 Limitations of the Newton's Method

Even though it is easy to understand and nice to implement, the above method can fail for several reasons. We give these below to help the reader appreciate why one should not accept the results blindly.
In spite of all the listed limitations, the method works pretty well for most reasonable functions.

1. **Case of no real roots!**. If we start with a function like $f(x) = x^2 + 2$, then we know that it has no real roots. What would happen to our method?

   The method will keep you wandering about the number line, leading to no value! You should try this!

[2] In this example, the formula which changes a to $a - \frac{a^2-2}{2a}$ simplifies to $\frac{1}{2}(a + \frac{2}{a})$.
In the appendix, you will find this formula developed without using the technique of linear approximation. It was historically known and the invention of Calculus gave a renewed reason for its efficiency.
This has a particularly pleasant description. Whichever side of $\sqrt{2}$ we take our positive a, we can see that $\frac{2}{a}$ lies on the other side and their average is exactly $\frac{1}{2}(a + \frac{2}{a})$. This is easily seen to be closer to $\sqrt{2}$ by observing them as points on the number line. Thus, we know for sure that we are clearly getting closer and closer to our $\sqrt{2}$. We can even estimate the new accuracy as half the distance between a and $\frac{2}{a}$. This particular method was probably well known since ancient times and leads to excellent approximations of any desired accuracy with rational numbers. Techniques for approximating other square roots using rational numbers by special methods are also scattered in ancient mathematical works. Here is a sequence of approximations to $\sqrt{2}$ using this simple formula but keeping the answers as rational numbers, rather than decimals.

$$1, \frac{3}{2}, \frac{17}{12}, \frac{577}{408}, \frac{665857}{470832}, \frac{886731088897}{627013566048}.$$

You will find the second to the sixth approximations commonly used in ancient literature as $\sqrt{2}$.
The fourth is accurate to 5 places. The fifth is accurate to 11 places and the sixth to 23 places. The seventh one, when calculated, will come out correct to 48 places!

2. **Reaching the wrong root**. Usually you expect that if you start near a potential root, you should land into it. It depends on the direction in which the tangent line carries you.

3. **Not reaching a root even when you have them!** Consider $f(x) = x^3 - 5x$. Then $f'(x) = 3x^2 - 5$. Then the starting value of $a = 1$ takes you to
$$1 - \frac{f(1)}{f'(1)} = 1 - \frac{-4}{-2} = -1.$$
Now the value $a = -1$ leads to
$$-1 - \frac{f(-1)}{f'(-1)} = -1 - \frac{4}{-2} = 1.$$
Thus you will be in a perpetual cycle between -1 and 1, never finding any of the roots $0, \sqrt{5}, -\sqrt{5}$. The reader should observe that taking other values of a does the trick. For example $a = 2$ takes you thru:
$$2 \to 2.2857 \to 2.2376 \to 2.2361.$$
This is the $\sqrt{5}$ to 5 decimal places.

   Other starting values will take you to other roots.

4. **Horizontal tangent.** If at any stage of the process, the derivative becomes zero, then the value of the new "a" is undefined and we get stuck. For more complicated functions, the derivative may not even be defined! Often changing the starting value takes care of this.

5. **Accuracy problems.** In addition to the above, there are possible problems with roundoff errors of calculations both in the calculation of the derivative and the function. There are also problems of wild jumps in values if the derivative becomes rather small and so $\frac{y}{y'}$ gets very large!

6. **Conclusion.** We have only sketched the method. An interested reader can pursue it further to study the branch of mathematics called Numerical Analysis, which studies practical calculation techniques to ensure desired accuracy.

Chapter 12

Appendix

12.1 An analysis of $\sqrt{2}$ as a real number.

First, it is a simple exercise in factoring integers to show that **no integers** x, y will satisfy the equation:

$$\left(\frac{x}{y}\right)^2 = 2 \text{ or } x^2 = 2y^2.$$

How do we get the decimal expansion (or the true position on the number line for $\sqrt{2}$)?
Here is an ancient and interesting technique to capture $\sqrt{2}$ in a trap of rational numbers.
The method we are developing here, also appeared as Newton's method of finding roots and used techniques of algebraic derivatives and linear approximations. We present an ad hoc calculation which leads to the same results and it seems to have been used in many ancient mathematical texts.
Start with $1 < \sqrt{2}$. Note that $2 = \frac{2}{1} > \sqrt{2}$. The average of $1, 2$ is $\frac{3}{2}$. We claim it is bigger than $\sqrt{2}$. This is easily checked by comparing $\frac{3}{2}^2 = \frac{9}{4}$ with $\sqrt{2}^2 = 2$.
Thus we have

$$1 < \sqrt{2} < \frac{3}{2}.$$

This locates $\sqrt{2}$ to within an error of 0.5 as it is between 1 and 1.5.
The above process can be generalized thus:

- If $0 < x < \sqrt{2}$ then $\frac{2}{x} > \sqrt{2}$ and

  (1) $$x < \sqrt{2} < \frac{1}{2}\left(x + \frac{2}{x}\right).$$

  Thus $\sqrt{2}$ is trapped between x and $\frac{1}{2}\left(x + \frac{2}{x}\right)$.

- On the other hand, if $\sqrt{2} < x$ then $\frac{2}{x} < \sqrt{2}$ and

$$\frac{1}{2}\left(x + \frac{2}{x}\right) < \sqrt{2} < x. \tag{2}$$

  Thus $\sqrt{2}$ is trapped between $\frac{1}{2}\left(x + \frac{2}{x}\right)$ and x.

Thus taking $x = 1$ in equation (1) above, we get:

$$1 < \sqrt{2} < \frac{3}{2}.$$

Now taking $x = \frac{3}{2}$ we calculate:

$$\frac{1}{2}\left(x + \frac{2}{x}\right) = \frac{1}{2}\left(\frac{3}{2} + \frac{4}{3}\right) = \frac{17}{12}.$$

Using $x = \frac{3}{2}$ in equation (2) above, we get:

$$\frac{17}{12} < \sqrt{2} < \frac{3}{2}.$$

A simple use of calculators will show that now $\sqrt{2}$ is caught between 1.416666667 and 1.5, so a distance of less than 0.1.
The reader should verify these next few steps:
Taking $x = \frac{17}{12}$, we get

$$\frac{17}{12} < \sqrt{2} < \frac{577}{408}.$$

This captures our $\sqrt{2}$ between 1.416666667 and 1.414215686 or a distance of less than 0.01.
Taking $x = \frac{577}{408}$, we get

$$\frac{665857}{470832} < \sqrt{2} < \frac{577}{408}.$$

Sticking in the calculator, you will see that we have captured $\sqrt{2}$ between 1.414215686 and 1.414213562 or we have fixed it to five decimal places, by simple hand calculations.
This particular way of getting to $\sqrt{2}$ can be traced back in history for over $2,000$ years.
While clever and efficient, this does not generalize so easily to solving any arbitrary equation. It does, however, give a basis for methods of modern mathematics for solving equations.

12.2 Idea of a Real Number.

Note the fact that only some rationals have a finite decimal expansion (namely the ones whose denominators divide some power of 10.) All other rationals still have decimals with a repeating pattern.
All other real numbers have infinite decimal expansions without a repeating pattern.
Hence, it becomes clear that one should not really hope that the decimal expansion of a random real number can be written down in any complete sense!
So, how should one think of a real number?
Here is an idea that works in all cases. Indeed our explanation of $\sqrt{2}$ above illustrates this idea very well.
First, we have **a simple observation:**
Given two real numbers $a < b$ we claim that there must be a rational number p such that [1]

$$a < p < b.$$

Proof. This is obvious if a, b have opposite signs, for then 0 is in between.
So, we first consider the case when a, b are both positive. Consider the positive number $b - a$. If it is bigger than 1, then it is clear that some integer is caught between them (for otherwise, both will be between two consecutive integers and hence $b - a < 1$.)
Now we claim that for some positive integer n, the multiple $n(b - a)$ is bigger than 1. We are done if $b - a > 1$.
So assume that $b - a < 1$ and consider the decimal expansion of $b - a$ as:

$$b - a = 0.q_1 q_2 \cdots q_r \cdots$$

where we assume that q_r **is the first non zero digit.**
Then we claim that

$$10^r(b - a) = q_r.q_{r+1} \cdots > q_r.$$

Thus, we can take $n = 10^r$.
The case when a, b are both negative is similar.
Now since $n(b - a) > 1$ we know that there is an integer m such that:

$$na < m < nb \text{ or dividing by } n \text{ we have } a < \frac{m}{n} < b$$

This proves our claim by setting $p = \frac{m}{n}$.
Now we show how to think of a real number.
Imagine a real number x. We must have some rational number L_1 smaller than x and some rational number R_1 bigger than x.

[1]Note that we want p to be rational and thus the simple idea of taking the average $\frac{a+b}{2}$ to be p may not work because it could be irrational.

Thus, we can say that

1. $L_1 < x < R_1$ or x is contained in the interval $I_1 = (L_1, R_1)$

We can certainly find rational numbers L_2, R_2 such that:

2. $L_1 < L_2 < x$ and $x < R_2 < R_1$ or x is contained in the interval $I_2 = (L_2, R_2)$

We can do something **more useful** by choosing L_2 between $\frac{L_1+x}{2}$ and x so we have

$$L_1 < \frac{L_1 + x}{2} < L_2 < x.$$

Similarly, we can choose R_2 such that:

$$x < R_2 < \frac{R_1 + x}{2} < R_1.$$

The advantage of this careful choice is that now the length of the new interval I_2 is clearly seen to be less than half of the length of I_1.
Continuing, we get a sequence of nested intervals $I_1, I_2, \cdots, I_n, \cdots$ so that x is in all of them and each contains the next.
Thus we see that the lengths of I_n approach zero, i.e. can be smaller than any positive number if we take n large enough.
We claim that the only number common to all of I_n's is x.
Here is the simple explanation: If all of them contain a $y \neq x$ then the length of each interval will be at least as big as $|x - y|$.
But we have argued that we can choose n large enough to make the length less than any desired positive number.
This is a contradiction!
Thus, every real number can be identified by a sequence of nested intervals $I_1, I_2, \cdots, I_n, \cdots$ whose lengths go to zero. It can be argued that in turn, any such sequence defines a real number.
Caution. Be aware that many different sequences of intervals can define the same real number. All we need is one sequence to claim that the real number exists.
There is a clever way of avoiding this ambiguity. It is the method of "Dedekind Cuts" introduced by Dedekind.
We observe that every real number x defines a way of splitting the set of all rationals into two sets:

$$L = \{a \in \mathbb{Q} \ |a \leq x\} \text{ and } R = \{a \in \mathbb{Q} \ |a > x\}.$$

These L and R are respectively called the lower and upper cuts for x and the pair is called the Dedekind cut for x. [2]

[2] Actually, the letters L and U would be more suggestive. Our L and R really signify left and right.

We now define a Dedekind cut in general, without having an x in hand.
A pair of subsets $(\mathbf{L}, \mathbf{R})$ of $\mathbb{Q}$ is said to form a Dedekind cut if the following properties hold:

- L, R are disjoint and their union is the set of all rational numbers.
- Given any $a \in L, b \in R$ we have $a < b$.
- The set R does not have a minimum. This means that there is no $b \in R$ such that $b \leq b'$ for all $b' \in R$.

It is possible to develop the necessary properties of number fields for the Dedekind cuts by defining algebraic operations on them. We recommend searching more advanced books on Real Analysis.
It is easy to create the sequence of nested intervals (L_i, R_i) from a Dedekind cut (L, R) by taking $L_i \in L$ and $R_i \in R$. Here is the simple trick. Having chosen an interval (L_i, R_i) let $u = \frac{L_i + R_i}{2}$. If $u \in L$ then take the next interval as (u, R_i) , i.e. $L_{i+1} = u$ and $R_{i+1} = R_i$.
If $u \in R$ then take the next interval as (L_i, u) , i.e. $L_{i+1} = L_i$ and $R_{i+1} = u$.

12.3 Summation of series.

A **series is a sum of numbers.** Suppose f is a function of integers and $a < b$ are two integers. Then by

$$\sum_{i=a}^{i=b} f(i)$$

we mean the sum

$$f(a) + f(a+1) + \cdots f(b).$$

We may shorten the notation to $\sum_a^b f(i)$ if the moving index i is clear.
For example if $f(x) = 3x + 1$ then

$$\sum_{i=3}^{i=7} f(i) = (3\cdot 3+1)+(3\cdot 4+1)+(3\cdot 5+1)+(3\cdot 6+1)+(3\cdot 7+1) = 3(3+4+5+6+7)+5 = 75+5 = 80.$$

Our main aim in this section is to develop formulas for

$$\sum_1^n f(i)$$

for positive integers n, when f is a polynomial function or an exponential function.

12.3.1 A basic formula.

Consider any function f and define a new function g by

$$g(x) = f(x) - f(x-1).$$

Consider the following sequence of substitutions. [3]

x	$g(x)$	$=$	$f(x) - f(x-1)$
1	$g(1)$	$=$	$f(1) - f(0)$
2	$g(2)$	$=$	$f(2) - f(1)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
i	$g(i)$	$=$	$f(i) - f(i-1)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
n	$g(n)$	$=$	$f(n) - f(n-1)$

If we add up the g-column and equate it to the sum of the last column, we see that most values of f cancel leaving us with a very simple formula

$$g(1) + g(2) + \cdots + g(i) + \cdots + g(n) = f(n) - f(0).$$

Thus, we have a simple minded summation formula:

$$\sum_1^n g(i) = f(n) - f(0)$$

where $g(x) = f(x) - f(x-1)$.
The series $\sum_1^n (f(i) - f(i-1))$ is called a **telescoping series** and its sum is always trivially computed by the above formula.
This gives us very efficient and clever ways for adding up a series of terms.

12.3.2 Using the basic formula.

Here are some easy applications of the above formula:

1. Suppose $f(x) = x$. Then $g(x) = (x) - (x-1) = 1$ and we get the trivial result

$$g(1) + \cdots + g(n) = 1 + \cdots + 1 = \sum_{i=1}^{i=n} 1 = f(n) - f(0) = n - 0 = n.$$

[3] Some textbooks will denote g as $\Delta(f)$. It is a kind of "numerical" derivative of f. With care, this process can be made to work like the derivatives.

2. If we take $f(x) = x^2 + x$, then $g(x) = (x^2 + x) - ((x-1)^2 + (x-1)) = 2x$ and we get the identity

   $$g(1) + \cdots + g(n) = 2 + 4 + \cdots + 2n = f(n) - f(0) = (n^2 + n) - (0) = n(n+1).$$

   Division of both sides by 2 gives the well known

   $$\textbf{Arithmetic series formula: } 1 + 2 + \cdots n = \frac{n(n+1)}{2}.$$

3. **A general arithmetic series (A.P.)** [4] A general arithmetic series is defined by taking $g(x) = a + d(x-1)$ so that the beginning term is a and the terms increase by d. Find the formula for $\sum_1^n a + d(i-1)$.

   Answer: First we give the easiest solution. Our $g(x) = (a-d) + dx$ and thus, we see that:

   $$\sum_1^n ((a-d) + di) = (a-d)\sum_1^n (1) + d\sum_1^n i$$

   and from what we already know, we can write our answer as:

   $$\sum_1^n ((a-d)+di) = (a-d)n + d\frac{n(n+1)}{2} = n\frac{2(a-d) + d(n+1)}{2} = n\frac{a + (a + (n-1)d)}{2}.$$

   This last expression is particular easy to remember as :

   $$\text{number of terms } \frac{\text{sum of the first and the last term}}{2}$$

   or more simply as **number of terms times the average of the first and the last terms**. This is known as the **A.P. formula**.

   We shall now derive the same formula by a cuter method, however, it does not generalize very well.

   Let us write the terms of the A.P. in a better notation:

   Let the terms be named thus:

term	description	value
t_1	the first term	a
t_2	the second term	$a + d$
t_3	the third term	$a + 2d$
$\ldots$		
t_n	the n-th term	$a + (n-1)d$

[4] A.P. is a short form of Arithmetic Progression, which stands for the sequence of terms with common differences. As already stated, the term "series" the sum of such a progression of numbers.

Let $S(n)$ denote the sum $t_1 + t_2 + \cdots + t_n$.

Consider the following table: [5]

$$\begin{array}{llllll} S(n) & = & t_1+ & t_2+ & \cdots & +t_n & \text{Forward} \\ S(n) & = & t_n+ & t_{n-1}+ & \cdots & +t_1 & \text{Reverse} \\ 2S(n) & = & (t_1+t_n)+ & (t_2+t_{n-1})+ & \cdots & (t_n+t_1) & \text{Added} \\ 2S(n) & = & (2a+(n-1)d)+ & (2a+(n-1)d)+ & \cdots & (2a+(n-1)d) & \text{Simpler!} \end{array}$$

And this last line gives the desired formula:

$$S(n) = (1/2)n(2a + (n-1)d).$$

For example, consider the A.P. formed by $5, 8, \cdots$. Here $a = 5, d = 3$ and the n-th term is $5 + 3(n-1) = 2 + 3n$. The average of the first and the n-th terms is simply $(1/2)(5 + 2 + 3n) = (7 + 3n)/2$. Thus $f(n) = n(7 + 3n)/2$.

The reader should verify the first 10 values of this function:

$5, 13, 24, 38, 55, 75, 98, 124, 153, 185$.

4. Thus to add up values of a gives $g(x)$ all we have to do is to come up with a cleverly constructed $f(x)$ such that $g(x) = f(x) - f(x-1)$. The best strategy is to try lots of simple $f(x)$ and learn to combine.

   Here is a table from simple calculations.

$f(x)$	$f(x) - f(x-1)$
x	1
x^2	$2x - 1$
x^3	$3x^2 - 3x + 1$
x^4	$4x^3 - 6x^2 + 4x - 1$
x^5	$5x^4 - 10x^3 + 10x - 5x + 1$

   Now suppose we wish to find the sum

$$1^2 + 2^2 + \cdots + n^2.$$

   Here $g(x) = x^2$ and we can manipulate the first three polynomials above and write

$$x^2 = \frac{1}{3}(3x^2 - 3x + 1) + \frac{1}{2}(2x - 1) + \frac{1}{6}(1)$$

   and this tells us that $f(x) = \dfrac{x^3}{3} + \dfrac{x^2}{2} + \dfrac{x}{6}$ will do the trick!

[5] The main idea is that a typical i-th term is $a + (i-1)d$ and so i-th and $(n+1-i)$-th terms added together give: $(a + (i-1)d) + (a + (n+1-i-1)d) = 2a + (n-1)d$.
It is worth thinking about why this gives the right answer!

Thus the desired answer is:

$$1^2+2^2+\cdots+n^2 = \sum_{i=1}^{i=n} i^2 = f(n)-f(0) = n^3/3+n^2/2+n/6 = \frac{(n)(n+1)(2n+1)}{6}.$$

How did we find the magic coefficients $\frac{1}{3}, \frac{1}{2}, \frac{1}{6}$?

Here is a systematic procedure that can be followed:

We want:

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

We try various special values of x to get linear equations in a, b, c and then we solve them.

Thus $x = 0, 1, -1$ respectively gives:

$$0 = a - b + c, \quad 1 = a + b + c, \quad 1 = 7a - 3b + c.$$

Subtracting the first from the second gives $2b = 1$ or $b = \frac{1}{2}$. Subtracting the first from the third gives $6a - 2b = 1$ and the known value of b says, $6a = 2$ or $a = \frac{1}{3}$.

Finally plugging into the first equation gives $\frac{1}{3} - \frac{1}{2} + c = 0$ or

$$c = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

It is clear that one can continue to develop formulas for any polynomial function g with similar work.

5. Example of a more general Arithmetic series. Using the idea for the Arithmetic series formula determine the formula

   for the sum of n terms $2 + 5 + 8 + \cdots$.

   Answer: This looks like adding up terms of $g(x) = -1 + 3x$. (Note that this gives $g(1) = 2, g(2) = 5$ etc. Then $g(n) = 3n - 1$ and we want the formula for $2 + 5 + \cdots + (3n - 1)$.

   By looking up our table, we see that $g(x) = (3/2)(2x - 1) + (1/2)$, so we use $f(x) = (3/2)(x^2) + (1/2)(x)$. Indeed, check that

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

   as desired.[6]

[6] Here is a useful idea for such checking. Instead of straight substitution, first write $f(x) = x(3x + 1)/2$. Then $f(x - 1) = (x - 1)(3(x - 1) + 1)/2 = (x - 1)(3x - 2)/2$. Thus,

$$f(x) - f(x-1) = \frac{x(3x+1) - (x-1)(3x-2)}{2} = \frac{(3x^2 + x) - (3x^2 - 5x + 2)}{2} = \frac{6x - 2}{2} = 3x - 1.$$

6. **Geometric series.**

   A geometric series is of the form

$$\sum_{i=1}^{i=n} ar^{i-1} = a + ar + ar^2 + \cdots ar^{n-1}.$$

   Usually, its summation is derived by ad hoc methods.

   We show how our above technique still applies, although in a rather complicated fashion.

   Note that our $g(x)$ is now $ar^{(x-1)}$ and we claim that

$$f(x) = a\frac{r^x}{(r-1)}$$

   does the job.

   Let us calculate $f(x) - f(x-1)$.

$$f(x) - f(x-1) = a\frac{r^x}{(r-1)} - a\frac{r^{x-1}}{(r-1)} = a\frac{r^{x-1}(r) - r^{x-1}}{r-1} = a\frac{r^{x-1}}{1} = g(x).$$

   Hence, by the basic formula:

$$\sum_{i=1}^{i=n} ar^{i-1} = f(n) - f(0) = a\frac{r^n}{r-1} - a\frac{r^0}{r-1} = a\frac{r^n - 1}{r-1}.$$

   Infinite Geometric Series. One of the most important applications of this formula and also the first significant idea of a limit is to consider the case when $|r| < 1$ and thus as n gets larger and larger r^n gets closer and closer to 0. Thus in the limit, we could state:

$$\sum_{1}^{\infty} ar^{i-1} = a + ar + ar^2 + \cdots = a\frac{-1}{r-1} = \frac{a}{1-r}.$$

12.4 On the exponential and logarithmic functions.

We provide additional properties of the exponential function and the related logarithmic function.

- **A basic inequality**

  Note that by its very definition, the function

  $$\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots > 1 + x$$

  provided $x > 0$. [7]

- Clearly, $\exp(0) = 1$. Thus, $\exp(x) \geq 1 > 0$ if $x \geq 0$.

- We now develop steps to prove that $\exp(x)$ is a one to one function.

  It is enough to prove the following. If $x_1 < x_2$ are real numbers, then $\exp(x_1) < \exp(x_2)$.

  First we make some observations.

  Let $t = x_2 - x_1$, so that $t > 0$.

  From Chapter 9 we know the property of exponential function

  $$\exp(a)\exp(b) = \exp(a + b).$$

  Using this wtih $a = x_1$ and $b = t$ we get:

  $$\exp(x_2) = \exp(x_1 + t) = \exp(x_1)\exp(t).$$

- In particular, it follows that $\exp(x_1) \neq 0$, since otherwise $\exp(x_2) = 0$ and this is incorrect when we take some positive $x_2 > x_1$.

  Thus, $\exp(x) \neq 0$ for any $x \in \Re$.

- Now we claim that $\exp(x) > 0$ for all $x \in \Re$.

  If $x \geq 0$, then this is already proved above. If $x < 0$, then we set $y = -x$ and note that $y > 0$.

  Further,

  $$1 = \exp(0) = \exp(x + y) = \exp(x)\exp(y).$$

  Since $y > 0$, we know that $\exp(y) > 0$. It follows the

  $$\exp(x) = \frac{1}{\exp(y)} > 0.$$

---

[7] The simple reason is that all the terms being added are non negative and so the sum of the whole series is bigger than the sum of any finitely many terms, provided there is at least one more positive term.

Thus, a stronger result is that

$$\exp(x) > 1 + \frac{x}{1!} + \cdots + \frac{x^n}{n!}$$

for all $x > 0$ and positive integers n.

- Thus,

$$\exp(x_2) - \exp(x_1) = \exp(x_1)\exp(t) - \exp(x_1) = \exp(x_1)(\exp(t) - 1).$$

- Since $\exp(x_1) > 0$ for all $x_1 \in \Re$ and $\exp(t) - 1 > 0$ for all $0 < t \in \Re$ we have proved the one to one property!

Thus, the exponential function is one to one and maps $\Re$ to the target $(\mathbf{0}, \infty)$.
We don't have the needed tools to show that it maps onto this set of positive reals, but this will be proved in higher courses.
Assuming it for now, we know that exp has a well defined inverse $\log : (0, \infty) \to \Re$.
From the meaning of an inverse function, we get:

$$\exp(\log(x)) = x \text{ and } \log(\exp(x)) = x.$$

Also, from the known property of the exponential function, we deduce that: [8]

$$\log(x_1 x_2) = \log(x_1) + \log(x_2) \text{ for all } x_1, x_2 \in (0, \infty).$$

It is also customary to use different notations for the exponential and logarithmic functions.
Let $e = \exp(1)$. Then it is easy to see that $\exp(n) = e^n$ for all integers n, in the usual sense of exponents. Thus, by abuse (or enhanced use) of notation, we may write

$$e^x = \exp(x) \text{ for any real } x.$$

This is often how the exponential function is written. The famous number "e" is known to be about 2.718281828 and is a transcendental number like π. It is also known as the base of the natural logarithms.
Since it cannot be written out completely, people often want to use something more convenient like 10 in its place.
Suppose that $a > 0$ with $a \neq 1$ is some chosen number to be used as a base. Write $a = \exp(t)$ where, naturally $t = \log(a)$.
Set $y = tx$.

---

[8] Let $\log(x_1) = y_1, \log(x_2) = y_2$. Then

$$\exp(y_1)\exp(y_2) = \exp(y_1 + y_2)$$

gives, after applying log to both sides:

$$\log(\exp(y_1)\exp(y_2)) = \log(\exp(y_1 + y_2)).$$

Now the RHS is $y_1 + y_2 = \log(x_1) + \log(x_2)$.
Also $\exp(y_1) = \exp(\log(x_1)) = x_1$ and similarly $\exp(y_2) = \exp(\log(x_2)) = x_2$. Therefore the LHS is $\log(x_1 x_2)$. This proves the claimed equation.

Then we have:

$$\exp(y) = \exp(tx) = \exp(t)^x = a^x.$$

Definition: Generalized Exponential and Logarithmic Functions.
Thus, we can define a new exponential function $\exp_a(x)$ with the property that $\exp_a(x) = \exp(tx) = e^{tx}$ where $t = \exp(a)$.
The notation $\exp_a(x)$ is often replaced by the more familiar a^x.
It is popular to use $a = 10$, so at least $\exp_{10}(n) = 10^n$ is easy to compute for integer values of n.
Naturally, we also define $\log_a(x)$ so that

$$a^{\log_a(x)} = x \text{ so that } \log_a(x) = \frac{\log(x)}{\log(a)}.$$

For reference purposes, we record the formulas for changing bases:

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

Notational confusion.
There is a standard confusion of notation between the general Scientific convention and the Mathematical convention for this function.
We now explain this. In Mathematics, the functions $\exp(x), \log(x)$ are exactly as we defined. General scientific community uses e^x more regularly and uses $\ln(x)$ in place of $\log(x)$. The "ln" stands for the natural log.
Also, in the general scientific community and especially Engineering, the function $\log(x)$ stands for what we call $\log_{10}(x)$. Thus, it is crucial to see what the intended symbol means. While using the calculator, it may be necessary to set some appropriate modes to get the correct function.
Why do Mathematicians use the complicated number "e" for a base, rather than the convenient 10? Because, the derivatives of the exponential and logarithmic functions with base e are rather convenient:

$$D_x(\exp(x)) = \exp(x) \text{ and } D_x(\log(x)) = \frac{1}{x}.$$

For other bases, we have an extra factor of $\log(a)$ to tag along and that is inconvenient.

12.5 Infinite series and their use.

We studied the sum of finitely many numbers above (the *finite series*), but often we need to make sense out of adding up infinitely many terms. This leads to both technical and philosophical problems.

In some sense, nobody could ever add up infinitely many terms! No matter how efficiently you can add terms, given that each addition requires some basic positive unit of time, it is easy to see that you would need infinite amount of time to add up all the terms.
This was the basis of one of the ancient paradoxes by Zeno; which we have reformulated slightly here.
Say, the pitcher wants to throw the ball at the batter. We propose to show that it would never reach. The ball must travel the 64 feet (approximately) to the batter's box. Assuming a speed of about 100 miles per hour, it should travel the distance in no more than 0.44 seconds! For this discussion we shall assume it to be exactly 0.44 seconds.
However, think about it this way. The ball must first travel half the distance in 0.22 seconds. Then it must travel half of the remaining distance in 0.11 seconds.
Then it must travel half of the remaining distance in some more time.
Then it must travel half of the remaining distance in some more time.
And so on $\cdots$. In other words, we can imagine infinitely many intervals of time needed to travel successively half of the remaining distances. Every time a ball reaches the batter's box, it has added all the infinitely many time intervals into 0.44 seconds!
Another example of the same phenomenon is this.
Every time you finish off a pizza, you are proving the theorem that:

$$\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \cdots = 1.$$

The first term is half the pizza you eat, the next term is a quarter or half of what you just ate, the third is an eighth or half of what you just ate and so on.
With these examples in mind, let us define an infinite series and analyze when it can have a meaningful value.

Definition: Sequence. Suppose f is a function defined on the set of positive integers $\mathbb{Z}_+ = \{1, 2, \cdots\}$. We often display its values as $(f(1), f(2), \cdots, f(n), \cdots)$ and this is said to be an infinite sequence. Naturally, we can never write out the whole sequence and it is the function value $f(n)$ for a general value of n which defines the sequence.
Often the **sequence function is generalized** by including $f(0)$ or even a few values at negative integers, like $f(-1), f(-2)$ etc. For technical reason, we do not allow infinitely many negative terms in building a sequence, since we always want a well defined first term!
Here are some examples of sequences and creative notations that people use to display them.

- Let the function be $f(n) = 2$ for all n. Then we get the constant sequence $(2, 2, \cdots)$.

- Let the function be $f(n) = (-1)^n)$. Then $f(1) = -1$, $f(2) = 1$, $f(3) = -1$, and so on. The sequence is: $(-1, 1, -1, 1, \cdots, (-1)^n, \cdots)$.
- Let the function be $f(n) = \frac{1}{2^n}$. Then the sequence is $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8} \cdots, \frac{1}{2^n}, \cdots)$.
- Let the function be $f(n) = \frac{x^n}{n!}$ where x is some number. Then we get the sequence $(\frac{x}{1}, \frac{x^2}{2!}, \frac{x^3}{3!}, \cdots, \frac{x^n}{n!}, \cdots)$.

  This can be generalized by throwing in $f(0) = \frac{x^0}{1!} = 1$. [9]
- Often we don't mention a function explicitly but simply write something like this:

  Let the Harmonic sequence be defined by:

$$h = (h_1, h_2, h_3, \cdots, h_n, \cdots) = (1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{n}, \cdots).$$

  The formula $h_n = \frac{1}{n}$ is all we need to make it meaningful.
- Thus, we shall often denote a sequence as $a = (a_n)$ where we specify the value of a_n by some formula or procedure. Moreover we stipulate the domain for n explicitly, unless it is the natural one $(1, 2, 3, \cdots)$.

Definition: Infinite series.
Given an infinite sequence $a = (a_n)$ where $n = 1, 2, 3, \cdots$, we define the associated infinite series

$$\sum_1^\infty a_n = a_1 + a_2 + \cdots + a_n + \cdots .$$

If this sum can be identified with a certain number or expression, then we declare the series **convergent**, otherwise it is declared **divergent.** The identified number or expression is said to be the value of the infinite series. Sometimes this is stated as **the series converges to** the identified number or expression.
The techniques to determine convergence or divergence of a series are beyond the scope of this text, but we shall stipulate the known convergence without proof.
Meaning of the value of a series. We offer a formal definition of the value of a series as a food for thought. This will be developed in greater detail in future courses.
Given a sequence $a = (a_n)$ with $n = 1, 2, 3, \cdots$ form a new sequence of partial sums:

$$s_1 = a_1, s_2 = a_1 + a_2, s_3 = a_1 + a_2 + a_3, \cdots, s_n = a_1 + a_2 + a_3 + \cdots + a_n, \cdots .$$

[9] A subtle point can come up here. If $x = 0$, then there is some technical problem in evaluating $x^0 = 0^0$. We may, however, just declare our f extended by setting $f(0) = 1$ and not make a fuss!

If all these partial sums seem to approach a well defined number b then we say that b is the limit of s_n or $\lim_n s_n = b$ and declare b as the value of the series.
Here is the formal meaning of the limit. We say that the limit of the sequence s_n is b if given any $\epsilon > 0$ we can find some large enough number N such that

$$b - \epsilon < s_n < b + \epsilon \text{ for all } n > N.$$

It is instructive to check that for the series

$$\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^n} + \cdots$$

we get from known formulas for this series:

$$s_n = 1 - \frac{1}{2^n}$$

and thus

$$1 - \epsilon < s_n = 1 - \frac{1}{2^n} < 1 + \epsilon$$

whenever

$$\frac{1}{2^n} < \epsilon \text{ or } 2^n > \frac{1}{\epsilon}.$$

We can, thus take N to be the some integer such that 2^N is bigger than $\frac{1}{\epsilon}$. Using our log function we can say this more explicitly as:

$$\log(2^N) > \log(1/\epsilon) \text{ or equivalently } N\log(2) > \log(1/\epsilon) \text{ or } N > \frac{\log(1/\epsilon)}{\log(2)}.$$

[10]

- **The (infinite) geometric series.** For any real (or complex) number x with $|x| < 1$ the geometric series

  $$1 + x + x^2 + \cdots x^n + \cdots \textbf{is convergent}$$

  and moreover the value associated to the series is $\frac{1}{1-x}$.

  It is instructive to analyze the convergence of this series as shown above for the special case of $x = \frac{1}{2}$.

  It is also instructive to think why the series could not possibly have any meaningful value when $x = \pm 1$ by calculating the sequence (s_n).

[10] Note that we don't really need to evaluate the expression, but the definition of the log lets us formally solve the inequality.

- **The exponential series.**

  For any real (or even complex) number x, the series

  $$\sum_0^\infty \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots$$

  is convergent. and thus **we can define** the exponential function $\exp(x)$ as its value!

  We have already done this without discussion during the explanation of Euler's representation of complex numbers.

- **Power series.** Given any sequence $a = (a_0, a_1, a_2, \cdots, a_n, \cdots)$ of real (or complex) numbers, we can consider an infinite series:

  $$u(x) = \sum_0^\infty a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n.$$

  The resulting series is said to be a **power series.**

  This expression is a polynomial if we discard the terms after some finite number n. Thus a power series is a very long polynomial!

  If we substitute a value of x in the series, we get a series of numbers which may or may not be convergent. The only value that is guaranteed to give a convergent answer is $x = 0$.

  It is possible to have a real number r such that $u(x)$ is convergent for all numbers x with $0 < |x| < r$. The largest possible value of r with this property is said to be the **radius of convergence** of the power series.

  The radius can be 0 if $x = 0$ is the only value of convergence. (It can be shown that the power series $\sum_0^\infty n!x^n$ is divergent for all non zero value of x and hence has radius of convergence 0.)

  The radius can be infinite if any positive value of r works. The exponential function $\exp(x)$ has radius of convergence ∞.

  Our geometric series $\sum_0^\infty x^n$ has radius of convergence 1.

- **Generalized Binomial Series.**

  We have seen the Binomial Theorem which gives as a special case:

  $$(1+x)^n = 1 + {}_nC_1 x + {}_nC_2 x^2 + \cdots + {}_nC_n x^n.$$

  Originally, the theorem was only demonstrated with positive integral values of n. It was also stated that if we use the second formula for ${}_nC_r$, namely:

$$ {}_nC_r = \frac{n(n-1)\cdots(n-r+1)}{r!} $$

then the formula makes sense for any real value of n and positive integer r, but now the expression need not stop at ${}_nC_n$.

Indeed, we can establish this formula (except for the convergence details, of course) as follows.

Let us assume that we have a power series expansion: [11]

$$ (1) \qquad (1+x)^n = a_0 + a_1\frac{x}{1!} + a_2\frac{x^2}{2!} + \cdots + a_i\frac{x^i}{i!} + \cdots . $$

Setting $x = 0$ on both sides, we see that $1 = a_0$. Using our enhanced power formula for the derivative, we can see that

$$ (2) \qquad n(1+x)^{n-1} = a_1 + a_2\frac{x}{1!} + \cdots + a_i\frac{x^{i-1}}{(i-1)!} + \cdots . $$

Now multiply the Equation (2) by $(1+x)$ and rearrange to collect the coefficients of various powers of x as follows:

$$ (3). \qquad \begin{array}{rcl} n(1+x)^n & = & a_1 + a_2\frac{x}{1!} + \cdots + a_i\frac{x^{i-1}}{(i-1)!} + \cdots . \\ & & +0 + a_1x + a_2\frac{x^2}{1!} + \cdots + a_{(i-1)}\frac{x^{i-1}}{(i-2)!} + \cdots \end{array} $$

The coefficient of x^{i-1} in the RHS of Equation (3) is

$$ (*) \qquad a_i\frac{1}{(i-1)!} + a_{(i-1)}\frac{1}{(i-2)!}. $$

Also, if we multiply the Equation (1) by n on both sides, then its LHS matches with Equation (3) and the coefficient of x^{i-1} is:

$$ (**) \qquad na_{(i-1)}\frac{n}{(i-1)!}. $$

The quantities in $(*)$ and $(**)$ must be equal and this gives:

$$ (4) \qquad a_i\frac{1}{(i-1)!} + a_{(i-1)}\frac{1}{(i-2)!} = na_{(i-1)}\frac{n}{(i-1)!}. $$

[11]The factorials $i!$ in the denominator are a technical convenience. We can work without introducing them, but then they appear as factors of our terms at the end and make the calculations more difficult.

If we multiply both sides of this equation by $(i-1)!$ and simplify we get $a_i + a_{(i-1)}(i-1) = a_{(i-1)}(n)$ and this gives

$$a_i = a_{i-1}(n-i+1).$$

Applying this formula repeatedly, we see that for $i = 1$ we have $a_1 = a_0(n) = n$.

Then with $i = 2$ we have $a_2 = a_1(n-2+1) = a_1(n-1) = n(n-1)$.

Then with $i = 3$ we have $a_3 = a_2(n-3+1) = a_2(n-2) = n(n-1)(n-2)$.

Continuing, it is not hard to see that $a_i = n(n-1)(n-2)\cdots(n-i+1)$. Thus the coefficient of x^i in Equation (1) is

$$\frac{a_i}{i!} = \frac{n(n-1)(n-2)\cdots(n-i+1)}{i!} = nC_i.$$

This establishes the Binomial Theorem and the power series expansion of $(1+x)^n$ for all values of n for which the derivative formula holds. The resulting series gives meaningful (convergent) values when $|x| < 1$.

- **Shifted Power series.** It is easy to develop the theory of a more general power series

$$v(x) = \sum_0^\infty a_n(x-p)^n$$

where we are using the shifted $(x-p)$ in place of x. Naturally, this is guaranteed convergent at $x = p$ and we should look for values of x near p for its radius of convergence. The theory of such series is very similar to the usual power series, we just think of $(x-p)$ as a new variable.

As an example, consider the series

$$\sum_0^\infty (x-1)^n = 1 + (x-1) + (x-1)^2 + \cdots + (x-1)^n + \cdots .$$

Using the theory of Geometric series, we can show that this is convergent when

$$|(x-1)| < 1 \text{ which means } -1 < x-1 < 1 \text{ or } 0 < x < 2.$$

It converges to the value $\frac{1}{1-(x-1)} = \frac{1}{2-x}$.

- **Derivatives of Power series.**

Given a shifted power series

$$v(x) = \sum_0^\infty a_n(x-p)^n$$

with radius of convergence $r > 0$, it defines a function of x for all x with $|x - p| < r$. This works even in complex numbers.

Such a function is said to be "the germ of an analytic function" in the domain $|x - p| < r$. An analytic function is usually defined by collecting a set of such germs such that two germs agree in value where their domains may intersect. In addition to being one of the major foundations of Mathematics, The study of such analytic functions is vital in Physics and many branches of Engineering. It is known as the theory of Complex Variables.

The theory leads to an easy calculation of the derivative (as if it were just a polynomial), namely:

$$D_x(v(x)) = v'(x) = \sum_0^\infty (n+1)a_{n+1}(x-p)^n$$

which is convergent in the same domain $|x - p| < r$. [12]

We record, without much explanation, the derivatives of various functions defined by power series. Often, we don't need the series expansions, but just the derivatives. Thus, using the derivatives of $\sin(x), \cos(x)$ from below, we can develop the derivative for $\tan(x)$ by the quotient rule:

$$D_x(\tan(x)) = D_x\left(\frac{\sin(x)}{\cos(x)}\right) = \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{\cos^2(x)} = \sec^2(x).$$

Exponential series

$$D_x(\exp(x)) = \exp(x).$$

Sine series

$$D_x(\sin(x)) = \cos(x).$$

Idea of proof.

$$\begin{aligned} D_x(\sin(x)) &= D_x\left(x - \tfrac{x^3}{3!} + \tfrac{x^5}{5!} + \cdots + (-1)^{n-1}\tfrac{x^{2n-1}}{(2n-1)!}\right) + \cdots \\ &= 1 - \tfrac{x^2}{2!} + \tfrac{x^4}{4!} + \cdots + (-1)^{n-1}\tfrac{x^{2n-2}}{(2n-2)!} + \cdots \\ &= \cos(x). \end{aligned}$$

[12] There is a bit of mystery here, since you would expect the derivative terms to look like

$$na_n(x-p)^{n-1}.$$

This is not wrong, but we have rearranged our terms so that the displayed term has $(x - p)^n$ in it. In case of confusion, we recommend writing out the answer term by term and verifying that it matches your calculation.

Cosine series

$$D_x(\cos(x)) = -\sin(x).$$

- **Using Power series.** Many functions have nice expressions in terms of power series and the main advantage of power series is that they can be handled as conveniently as ordinary polynomials; provided we calculate enough terms **to handle accuracy issues.**

  In some sense, the main ideas of Calculus can be described thus. [13]

  - Study the polynomials and their behavior near a point. Develop formulas for their derivatives and thus approximations near a point.
  - Develop higher and higher degree polynomial approximations (usually called Taylor or McLaurin polynomials.)
  - When possible, further develop full power series which become precise formulas, rather than approximations.
  - Handle the resulting power series like long polynomials and deduce more properties of function. Since power series are valid only within the radius of convergence, we often have to develop different power series near different points. But at least near the point, we have the luxury of using a power series rather than the complicated function!
  - If we are lucky enough to have functions with infinite radius of convergence, then we can use the same power series everywhere. We have seen this for the exponential function $\exp(x)$, $\sin(x)$ and $\cos(x)$.

12.6 Inverse functions by series

Given a function, we know what is meant by an inverse function. Most interesting functions don't have an inverse, but usually, by trimming the domain and the target down to a convenient size, we can arrange the resulting function to have an inverse. It is, however, rarely easy to "find" the inverse of such a function as a concrete formula.

In this section we show how many classical functions defined by power series can yield explicit power series formulas using our notion of derivatives.

Suppose we are given a power series

$$p(x) = p_0 + p_1 x + \cdots + p_n x^n + \cdots$$

and we know that it is the derivative of some power series

$$q(x) = q_0 + q_1 x + \cdots + q_n x^n + \cdots \text{ so that } q'(x) = p(x).$$

[13] Note that we are covering the first two steps with hints for some of the rest.

Then we can conclude by taking a term by term derivative:

$$q_1 = p_0, 2q_2 = p_1, \cdots, (n+1)q_{n+1} = p_n \text{ with } q_0 \text{ free.}$$

Warning: The hardest part of the analysis is to figure out the values of x for which the $q(x)$ exists and the derivative formula holds.
This involves the analysis of the convergence of $p(x)$. This will be done in higher courses. Here, we only prepare the algebraic manipulation needed to determine $q(x)$.

1. **The inverse of** $\exp(x)$**.** As explained above, the exponential function has an inverse. What is a formula for it?

   If we set $y = \exp(x)$ Then we know:

   $$\frac{dy}{dx} = D_x\left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots\right).$$

   The RHS becomes

   $$0 + \frac{1}{1!} + \frac{2x}{2!} + \frac{3x^2}{3!} + \cdots + \frac{nx^{n-1}}{n!} + \cdots.$$

   It is not hard to see that the RHS is again $\exp(x)$ and we have thus proved that $\frac{dy}{dx} = \exp(x) = y$.

   It follows that when we consider the inverse function $x = \log(y)$ then we get:

   $$\frac{dx}{dy} = \frac{1}{y}.$$

   When we use the variable x for the inverse function, this says that $D_x(\log(x)) = \frac{1}{x}$.

   Of course, $\log(x)$ cannot have any power series expression, since $\frac{1}{x}$ can never appear in the derivative of a power series! [14]

   But the situation is not hopeless. After all, we usually study a function near a point and the function $\log(x)$ may have a good expansion near some other point.

   For example, study it near $x = 1$, noting that $\log(1) = 0$. As usual, we set $x = 1 + u$ and note that

   $$D_u \log(1+u) = \frac{1}{1+u} \text{ using a natural Chain Rule.}$$

[14] An indirect reason for this failure in getting a power series is that a power series in x can always be evaluated at $x = 0$, but $\frac{1}{x}$ cannot be so evaluated.

The RHS is a known geometric series

$$\frac{1}{1+u} = 1 - u + u^2 - u^3 + \cdots + (-1)^n u^n + \cdots .$$

This is valid when $|u| < 1$.

So, we deduce that

$$\log(1+u) = c + u - \frac{u^2}{2} + \frac{u^3}{3} + \cdots + (-1)^n \frac{u^{n+1}}{n+1} + \cdots .$$

Setting $u = 0$ on both sides, we get $\log(1) = c$, so $c = 0$.

Thus, we conclude that

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots + (-1)^n \frac{x^{n+1}}{n+1} + \cdots \text{ valid for } |x| < 1.$$

Taking $x = 1$, the series still happens to be convergent and we get:

$$\log(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots (-1)^n \frac{1}{n} + \cdots .$$

This series converges, but too slowly to be useful. A more clever way of computing is this:

Note that due to cancellation of terms:

$$\log(1+x) - \log(1-x) = 2x + 2\frac{x^3}{3} + 2\frac{x^5}{5} + \cdots + \frac{x^{2n+1}}{2n+1} + \cdots .$$

Also, the LHS is nothing but $\log(\frac{1+x}{1-x})$. If we take $x = \frac{1}{3}$, we get

$$\log(\frac{1+x}{1-x}) = \log(\frac{4/3}{2/3}) = \log(2).$$

This gives a very efficient formula:

$$\log(2) = 2\left(\frac{1}{3} + \frac{1}{3 \cdot 3^3} + \frac{1}{5 \cdot 3^5} + \frac{1}{7 \cdot 3^7} + \cdots\right).$$

Even the first four terms give a sum 0.6931347573 which has seven correct digits!

It is instructive to compute values of $\log(\frac{1+x}{1-x})$ for various values of x near 0, using a few terms of this series and comparing them to the straight value from the calculator.

2. **The inverse of** $\tan(x)$**.** The tan function is known to be one to one and onto when the domain is taken to be $(-\pi/2, \pi/2)$ and the range is $\Re$. The function arctan is defined to be the inverse for this range. As before, we calculate the derivative of this function.

   Thus, we set $y = \tan(x)$ and the derivative gives
$$\frac{dy}{dx} = \sec^2 x = 1 + \tan^2(x) = 1 + y^2.$$

   Thus:
$$\frac{dx}{dy} = \frac{1}{1+y^2} = 1 - y^2 + y^4 + \cdots + (-1)^n y^{2n} + \cdots .$$
   As before, we deduce that the expansion of x as a power series in y is
$$\arctan(y) = x = c + y - \frac{y^3}{3} + \frac{y^5}{5} + \cdots + (-1)^n \frac{y^{2n+1}}{2n+1} + \cdots .$$

   Setting $y = \tan(x) = 0$, we see that $x = 0$, when restricted to the given domain $(-\pi/2, \pi/2)$. Hence $c = 0$. Indeed, this convenience is the reason for choosing such a domain. Changing variables, we write: [15]
$$\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} + \cdots + (-1)^n \frac{x^{2n+1}}{2n+1} + \cdots .$$
   This series is also defined for $|x| < 1$.

   If we take $x = 1$, then we get the famous expansion for the mysterious number π: [16]

   Leibnitz' Series
$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} + \cdots + (-1)^n \frac{1}{2n+1} n + \cdots .$$

   Despite the beauty of the formula, this series does not come to the correct value, unless we use a lot of terms. So, people often devise series with faster growing denominators which converge to the final answer more quickly.

   Here is one simple trick!

   Prove that $\arctan(\frac{1}{2}) + \arctan(\frac{1}{3}) = \frac{\pi}{4}$. (Use addition formula for the tangent function.)

   Now use the arctan series for $x = \frac{1}{2}$ and $x = \frac{1}{3}$ and add. Even just 8 terms of each give a good enough value of $\pi/4$ so π comes out correct to 5 places.

[15] This series is often ascribed to Gregory who discovered it in late 17-th century. It was, however, discovered in the 14-th century by Mādhava a mathematician from Kerala, India. His original work is not available but we know of his results from the work of another Kerala mathematician Nīlakaṇṭha in 15-th century.

[16] Leibnitz probably discovered it from geometric arguments and not from arctan series.

3. **The inverse of** $\sin(x)$**.**

   As before, we shall first find the derivative of the inverse function arcsin as a series and then deduce the series for arcsin.

   Writing $y = \sin(x)$ we get that

$$\frac{dy}{dx} = \cos(x) \text{ so } \frac{dx}{dy} = \frac{1}{\cos(x)} = \frac{1}{\sqrt{1-\sin^2(x)}} = \frac{1}{\sqrt{1-y^2}}.$$

   Thus the derivative is given by

$$\frac{dx}{dy} = (1-y^2)^{-\frac{1}{2}}.$$

   This gives the formula:

$$\frac{d\,\mathrm{arcsin}(x)}{dx} = (1-x^2)^{-\frac{1}{2}}.$$

   Using our Binomial Theorem in general form, we get:

$$\frac{d\,\mathrm{arcsin}(x)}{dx} = 1 + (-\frac{1}{2})\frac{-x^2}{1!} + (-\frac{1}{2})(-\frac{1}{2}-1)\frac{(-x^2)^2}{2!} + \cdots .$$

   We leave the task of verifying that the n-th term after simplification is:

$$\frac{(1)(3)(5)\cdots(2n-1)}{2^n n!}x^{2n} = \frac{(1)(3)(5)\cdots(2n-1)}{(2)(4)\cdots(2n)}x^{2n}$$

   Then the corresponding term in arcsin expansion becomes:

$$\frac{(1)(3)(5)\cdots(2n-1)}{(2)(4)\cdots(2n)}\frac{x^{2n+1}}{2n+1}.$$

   Thus we deduce

$$\mathrm{arcsin}(x) = x + \frac{1}{2}\frac{x^3}{3} + \frac{(1)(3)}{(2)(4)}\frac{x^5}{5} + \frac{(1)(3)(5)}{(2)(4)(6)}\frac{x^7}{7} + \cdots .$$

   Here, we have decided that the constant term is zero by comparing both sides for $x = 0$.

   Taking $x = \frac{1}{2}$ and noting $\mathrm{arcsin}(x) = \pi/6$ we get an estimate for π as 3.1415 with only 4 terms!

12.7 Decimal expansion of a Rational Number.

Recall that (a possibly infinite) non zero decimal number appears as:

$$a_0 \cdot a_1 a_2 \cdots a_n \cdots$$

where a_0 is an integer (positive of negative) and each of $a_1, \cdots, a_n, \cdots$ is an integer between 0 and 9.
If all the digits are zero from some point on, then we simply drop them and write a finite decimal.
We shall agree to use a compact notation as

$$a_0 \cdot a_1 \cdots a_r \overline{x}$$

in place of

$$a_0 \cdot a_1 \cdots a_r xxx \cdots$$

where x is a finite string of digits between 0 and 9 and is intended to be repeated indefinitely.
Moreover, we shall agree that

$$a_0 \cdot a_1 \cdots a_{r-1} a_r \overline{9}$$

where $a_r \neq 9$ may be replaced by

$$a_0 \cdot a_1 \cdots a_{r-1}(a_r + 1)\overline{0} = a_0 \cdot a_1 \cdots a_{r-1}(a_r + 1).$$

Thus, for example

$$0.9999 \cdots = 0.\overline{9} = 1.0.$$

With this convention, it is easy to show that two numbers are identical if and only if they have the exact same decimal expansion.
We now show how to write an arbitrary fraction as a decimal number and how to argue that the result is either a finite decimal or a decimal with a repeating pattern.
Consider a rational number $\frac{m}{n}$ where $n \neq 0$. Clearly, it is enough to consider a positive number, so we assume $m > 0, n > 0$.
Here are some simple cases worthy of attention.

$$\frac{2}{5} = \frac{4}{10} = 0.4.$$

$$\frac{1}{3} = 0.3333 \cdots = 0.\overline{3}.$$

$$\frac{1}{7} = 0.142857142857 \cdots = 0.\overline{142857}.$$

1. If either m or n is divisible by a power of 10, then we can take it out and after finding the decimal expansion of the rest, put it back in by shifting the decimals.

   Thus
$$\frac{1}{70} = \frac{1}{10}\frac{1}{7} = (0.1) \cdot 0.\overline{142857} = 0.0142857142857 \cdots = 0.0\overline{142857}.$$

2. Next we may arrange that the denominator n is not divisible by 2 or 5. The reason is that in such cases we can multiply both numerator and denominator by 5 or 2 respectively and make a power of 10 as a factor in the denominator. This power of 10 can then be forgotten as shown above.

   Thus:
$$\frac{5}{6} = \frac{25}{30} = \frac{1}{10}\frac{25}{3} = 8.\overline{3} = 0.8\overline{3}.$$

3. Thus, we may now assume that our fraction is $\frac{m}{n}$ with n having no common factors with 10.

   We shall illustrate our procedure for the number $\frac{1}{7}$.

   We take the powers of 10 and take remainders modulo 10. (**Recall** that this means the remainder after dividing by 10.) It is guaranteed that some power will give remainder 1 and assume that this power is 10^b.

   Thus for our $n = 7$, we calculate the remainders of powers of 10^r modulo 7.

   For $r = 1$ we get 3. For $r = 2$, we get the remainder of 100 modulo 7 to be 2, by writing $100 = 14 \cdot 7 + 2$. **Important observation.** Note that we could also have gotten the same answer by taking the remainder of 10, namely 3, and then finding the remainder of $10 \cdot 3 = 30$ which is also 2 since $30 = 4 \cdot 7 + 2$.

   Now for $r = 3$ we take the remainder of 1000 or using the above observation, remainder of $10 \cdot 2 = 20$ which comes to 6.

   Continuing, for $r = 4$, we get the remainder of $10 \cdot 6 = 60$ which is 4.

   For $r = 5$, we get the remainder of $10 \cdot 4 = 40$ to be 5.

   For $r = 6$, we get the remainder of $10 \cdot 5 = 50$ to be 1. Hence, $b = 6$.

   Thus, 10^6 has remainder 1 modulo 7. In fact, we see that:
$$10^6 - 1 = 142857 \cdot 7.$$

   Thus we can write:
$$\frac{1}{7} = \frac{142857}{10^6 - 1}.$$

   In general, we can write:
$$10^b - 1 = nd$$

and using this:

$$\frac{m}{n} = \frac{md}{nd} = \frac{md}{10^b - 1} = u + \frac{v}{10^b - 1}$$

where we have divided md by $10^b - 1$ to give the quotient u and remainder v.

Thus, for example,

$$\frac{2}{7} = \frac{2 \cdot 142857}{10^6 - 1} = \frac{285714}{10^6 - 1}.$$

4. Now we are ready for the finish. For convenience, write p for the number 10^{-b}.

   Clearly $0 < p < 1$ and by the infinite Geometric series formula, we can write

$$\frac{1}{1-p} = \frac{1}{1 - \frac{1}{10^b}} = \frac{10^b}{10^b - 1} = 1 + p + p^2 + \cdots.$$

   A little thought shows that the right hand side is

$$1 \cdot 0 \cdots 010 \cdots 01 \cdots$$

   where the part past the decimal point consists of $b - 1$ zeros followed by 1 and this pattern of b digits repeated indefinitely.

   Thus, we see that:

$$\frac{1}{10^b - 1} = 0.\overline{h}$$

   where h is the sequence of $(b - 1)$ zeros followed by the digit 1.

   It is then easy to see that for our positive integer $v < 10^b - 1$ we can write:

$$\frac{v}{10^b - 1} = 0.\overline{v^*}$$

   where v^* is obtained from v by filling in enough zeros to the left to make a block of b digits.

   Thus for our number

$$\frac{2}{7} = \frac{285714}{10^6 - 1}$$

   we simply get:

$$\frac{2}{7} = 0.\overline{285714}.$$

12.8 Matrices and determinants: a quick introduction.

Let F be any convenient field of numbers. [17]
A matrix of size $m \times n$ is a rectangular array of numbers in F with m rows and n columns.
Here are two matrices A, B of sizes 3×2 and 2×4 respectively.

$$A = \begin{pmatrix} 1 & 2 \\ 4 & 7 \\ -1 & 2 \end{pmatrix}, \ B = \begin{pmatrix} 1 & 2 & 5 & -2 \\ 1 & 1 & 3 & 3 \end{pmatrix}.$$

For any matrix M, by its (i, j)-th entry, we mean the entry in its i-th row and j-th column. We will conveniently denote it as M_{ij}.
Thus, $A_{31} = -1$ and $B_{13} = 5$. Note that A_{13} is not defined. Thus, one has to be careful in using these expressions.
The basic operations on matrices are as follows:

1. **Addition.** Two matrices P, Q can be added **only if** they have the same size. Then the resulting matrix $P + Q$ is simply obtained by adding the corresponding entries. This is conveniently written as:

$$(P + Q)_{ij} = P_{ij} + Q_{ij}.$$

Thus, if B is as above and $C = \begin{pmatrix} 2 & 1 & 1 & 3 \\ 0 & 3 & 1 & 5 \end{pmatrix}$ then we have:

$$\begin{aligned} B + C &= \begin{pmatrix} 1 & 2 & 5 & -2 \\ 1 & 1 & 3 & 3 \end{pmatrix} + \begin{pmatrix} 2 & 1 & 1 & 3 \\ 0 & 3 & 1 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 1+2 & 2+1 & 5+1 & -2+3 \\ 1+0 & 1+3 & 3+1 & 3+5 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 3 & 6 & 1 \\ 1 & 4 & 4 & 8 \end{pmatrix} \end{aligned}$$

2. **Scalar multiplication.** Given any number $c \in F$ and any matrix P we define the matrix cP to be the matrix where every entry of P is multiplied by c. Thus, $(cP)_{ij} = cP_{ij}$.

Thus, for A as above, we see that

$$5A = \begin{pmatrix} 5 & 10 \\ 20 & 35 \\ -5 & 10 \end{pmatrix}.$$

[17]It is often enough to think of the real or complex numbers.

3. **Product of two matrices.** This is the most complicated of matrix operations.

   Product of two matrices P, Q is defined **only if** the number of columns of the first matrix P is equal to the number of rows of the second matrix Q.

   If this condition is satisfied, then PQ_{ij} is calculated by multiplying the corresponding entries of the i-th row of P and the j-th column of Q.

   Thus for our A, B above, we have:

$$\begin{aligned} AB &= \begin{pmatrix} 1 & 2 \\ 4 & 7 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 5 & -2 \\ 1 & 1 & 3 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 4 & 11 & 4 \\ 11 & 15 & 41 & 13 \\ 1 & 0 & 1 & 8 \end{pmatrix} \end{aligned}$$

   We illustrate a few of these entries. Thus, the entry AB_{23} in the second row third column is calculated by taking the two entries $4, 7$ of the second row of A and multiplying by the corresponding entries of the third column of B, namely $5, 3$ to produce

$$(4)(5) + (7)(3) = 20 + 21 = 41.$$

   Similarly, the entry in the third row second column is:

$$AB_{32} = (-1)(2) + (2)(1) = -2 + 2 = 0.$$

   You should verify the rest.

   Note that the product BA is not even defined since the number 2 of columns of B does not match the number 3 of rows of A. Thus, the matrix product is highly non commutative!

   It is natural to ask why we make such a messy product. The main reason is the application. We use matrices to convert several linear equations in several variables into something which looks like a constant times a variable, but we have to allow both the constant and the variable to be an array or a matrix.

   Thus the equations $2x + 5y = 3, x + 3y = 2$ can be written as:

$$MX = N$$

   where we have:

$$M = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}, X = \begin{pmatrix} x \\ y \end{pmatrix} \text{ and } N = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

The product MX using our definition above produces the correct left hand sides of the two equations and the matrix N produces the two right hand sides.

This clearly has the appearance of one linear equation in one variable (matrix X). If M, N were ordinary numbers, the answer would be N/M, but alas, this simpleminded ratio does not make sense!

The theory of matrices is designed to answer the question of how to get rid of M, if possible and thus solve all linear equations as efficiently as possible.

The Cramer's Rule that we have already studied can be shown to be equivalent to the following operation:

Multiply both sides of the equation $MX = N$ by the matrix

$$M^* = \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix}$$

on the left and write and simplify the resulting equations:

$$\begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix} X = \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

The simplification yields:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} X = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

leading to the answer $X = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ or $x = -1, y = 1$.

More generally, here is how Cramer's Rule operates.

Consider the matrix equation:

$$MX = N \text{ where } M = \begin{pmatrix} a & b \\ p & q \end{pmatrix}, N = \begin{pmatrix} c \\ r \end{pmatrix} \text{ and } X = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Let

$$M^* = \begin{pmatrix} q & -b \\ -p & a \end{pmatrix}$$

and calculate the equation $M^*MX = M^*N$.

It is easy to verify that the left hand side evaluates to:

$$\begin{pmatrix} q & -b \\ -p & a \end{pmatrix} \begin{pmatrix} a & b \\ p & q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \Delta & 0 \\ 0 & \Delta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

where $\Delta = aq - bp$.

Calculation of M^*N gives:

$$\begin{pmatrix} q & -b \\ -p & a \end{pmatrix} \begin{pmatrix} c \\ r \end{pmatrix} = \begin{pmatrix} cq - br \\ ar - cp \end{pmatrix}.$$

Thus we get the equation rewritten as

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} cq - br \\ ar - cp \end{pmatrix}$$

or

$$\Delta x = cq - br, \Delta y = ar - cp \text{ or } x = \frac{cq - br}{\Delta}, y = \frac{ar - cp}{\Delta}.$$

This was our Cramer's Rule!

Generalized Cramer's Rule.

When we have d equations in the same number d of unknowns, then we get a similar theory.

The equations become $MX = N$, where we have X as the column of d variables $x_1, \cdots, x_d$, M is a $d \times d$ matrix of coefficients of the variables and N is the $d \times 1$ matrix of the right hand sides.

We can define the matrix M^* called the adjoint (or adjugate) matrix of M thus:

The entry M^*_{ij} in the i-th row and j-th column of the adjoint matrix is defined as:

$(-1)^{i+j}$ (determinant obtained by dropping the j-th row and i-th column of M.)

It is important to note the switch between the row and column numbers. Also the first power of -1 is simply ± 1 and keeps on alternating as we walk across a row or a column.

It is fairly easy to calculate the adjoint if we can calculate determinants efficiently.

For example, let us revisit the old equations from examples at the end of Cramer's Rule discussion:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 2 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 3 \end{pmatrix}.$$

We claim that the adjoint of the coefficient matrix $M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 2 & -1 & -1 \end{pmatrix}$ is simply:

$$M^* = \begin{pmatrix} 3 & 0 & 3 \\ 3 & -3 & 0 \\ 3 & 3 & -3 \end{pmatrix}.$$

Here are some sample calculations:

$M^*_{22} = (-1)^{2+2} \begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} = (1)(-1-2) = -3.$

$M^*_{12} = (-1)^{1+2} \begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix} = (-1)(-1+1) = 0.$

$M^*_{31} = (-1)^{3+1} \begin{vmatrix} 1 & -2 \\ 2 & -1 \end{vmatrix} = (1)(-1+4) = 3.$

It is easy to see that $M^*M = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix}$ and

$M^*N = \begin{pmatrix} 9 \\ 18 \\ 27 \end{pmatrix}.$

Now calculate the new equation $M^*MX = M^*N$ to give:

$$\begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix} X = \begin{pmatrix} 9 \\ 18 \\ 27 \end{pmatrix}.$$

This leads to the known answers,

$$x = \frac{9}{9} = 1, y = \frac{18}{9} = 2, z = \frac{27}{9} = 3.$$

Note that once M^* is computed, the solutions are quickly found for any give right hand side matrix N. That is the power of this method.

Also note that M^*M is always going to have the special form of the determinant of M on the diagonal and we only need to compute only its $(1,1)$ entry.

It also shows how the determinant works in the general case. The only thing unproved here is the form of M^*M and the reader is encouraged to learn it from a book on determinants or (old fashioned) Linear Algebra.

Index