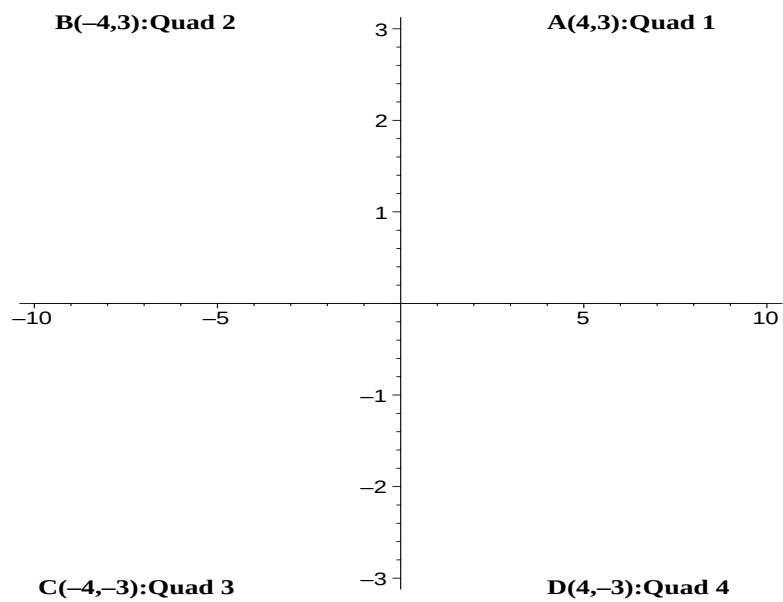


College Algebra

by Avinash Sathaye, Professor of Mathematics
Department of Mathematics, University of Kentucky



August 22, 2005

Contents

1	Polynomials, the building blocks of algebra	1
1.1	Underlying field of numbers.	1
1.2	Indeterminates, variables, parameters	2
1.3	Basics of Polynomials	3
1.4	Working with polynomials	6
1.5	Examples of polynomial operations.	8
	Example 1. Polynomial operations.	8
	Example 2. Collecting coefficients.	9
	Example 3. Using algebra for arithmetic.	9
	Example 4. Expanding powers.	10
	Example 5. Substituting in a polynomial.	11
	Example 6. Completing the square.	12
2	Solving linear equations.	15
2.1	What is a solution?	16
2.2	One linear equation in one variable.	16
2.3	Several linear equations in one variable.	18
2.4	Two or more equations in two variables.	18
2.5	Several equations in several variables.	19
2.6	Solving linear equations efficiently.	20
	Example 1. Manipulation of equations.	20
	Example 2. Cramer's Rule.	21
	Example 3. Exceptions to Cramer's Rule.	22
	Example 4. Cramer's Rule with many variables.	23
3	The division algorithm and applications	27
3.1	Repeated Division.	30
3.2	The GCD and LCM of two polynomials: meaning and calculations. .	31
3.3	Āryabhata algorithm: Efficient Euclidean algorithm.	32
	Example 1. Kuttaka or Chinese Remainder Theorem.. . . .	35
3.4	Division algorithm in integers.	38
	Example 2. GCD calculation in Integers.	40

	Example 3. Efficient division by a linear polynomial.	42
	Example 4. Division by a quadratic polynomial.	44
4	Introduction to analytic geometry.	49
4.1	Coordinate systems.	49
4.2	Geometry: Distance formulas.	50
4.3	Change of coordinates on a line.	52
4.4	Change of coordinates in the plane.	53
4.5	General change of coordinates:	54
	Examples. Changes of coordinates.	55
5	Equations of lines in the plane.	59
	Examples. Parametric equations of lines.	59
5.1	Meaning of the parameter t :	61
	Examples. Special points on parametric lines.	62
5.2	Details of the equation of a line.	64
5.3	Examples of equations of lines.	68
	Example 1. Points equidistant from two given points.	68
	Example 2. Right angle triangles.	69
6	The Circle	71
6.1	Parametric form of a circle.	72
6.2	Trigonometric parametrization.	73
	Definition. Trigonometric Functions.	75
6.2.1	Yet another take on parametrizing the circle.	79
	Pythagorean Triples. Generation of.	79
6.3	Examples of equations of a circle.	82
	Example 1. Intersection of two circles.	82
	Example 2. Line joining through the intersection of two circles. . . .	83
	Example 3. Circle through three given points.	83
	Example 4. Exceptions to a circle through three points.	84
	Example 5. Smallest circle with a given center meeting a given line. .	84
	Example 6. The distance between a point and a line.	85
	Example 7. Half plane defined by a line.	86
7	Special study of Linear and Quadratic Polynomials.	89
7.1	Linear Polynomials.	89
7.2	Factored Quadratic Polynomial.	90
	Interval notation. Intervals on real line.	90
7.3	The General Quadratic Polynomial.	92
7.4	Examples of quadratic polynomials.	93

8 Functions	95
8.1 Plane algebraic curves	95
8.2 What is a function?	96
8.3 Modeling a function.	97
9 Looking closely at a function	101
Parabola. Analysis near its points.	101
Circle. Analysis near its points.	104
9.1 Analyzing a general curve $y = f(x)$ near a point $(a, f(a))$	105
9.2 The slope of the tangent, calculation of the derivative.	107
9.3 Derivatives of more complicated functions.	110
9.4 More complicated derivatives.	111
9.5 Using the derivatives for approximation.	114
Linear Approximation. Examples.	114
9.6 Efficient expansions of polynomials, the Binomial Theorem.	116
10 Root finding	119
10.1 Newton's Method	119
10.2 Limitations of the Newton's Method	121
11 Summation of series.	125
11.1 Application of polynomials	125
11.2 Examples of summation of series.	126
12 More on the Trigonometric Functions.	129
12.1 Basic definitions of Trigonometric Functions.	129
12.2 Connection with the usual Trigonometric Functions.	131
12.3 Addition Formula.	132
12.4 Identities Galore.	134
12.5 Using trigonometry.	139
Index	143

Introduction.

This book on College Algebra is designed with a different philosophy from the other textbooks on the same subject.

The core material of a course on College Algebra is usually a subset of the recommended material expected to be taught in a high school course. The material is usually taught at a slow pace with extensive practice and lots of routine exercises. The College Algebra course usually attempts to get the student prepared for extensive calculations using a calculator and discusses several real life examples for motivation as well as practice. Due to short amount of time available and many specific topics to study, the students end up learning how to answer specific questions.

Often there is not enough time or motivation to get an overall view of any specific subject and due to the emphasis on solving specific practical problems with explicit decimal answers, the student has no time or motivation to study how the answers are derived. A typical student often insists that "don't tell me too many ways of doing something; don't tell me how the formula is derived; just show me how to do the problems which will appear on the test!".

When a trained student goes to a higher level course, he still needs such an explicit help, since he has not learned how to create his own theory and set up his own equations.

Our hope is to offer a course where the emphasis is on mastering the Algebraic technique. Algebra consists of manipulating expressions to put them in convenient form for getting at desired answers. Algebra consists of looking for ways of finding information about various quantities even though it is difficult to actually solve for them. Algebra consists of

finding multiple expressions for the same quantities, for the comparison of different expressions often leads to new discoveries.

We urge the readers to approach this book with an open mind. You may find new perspective on known topics.

We urge the reader to carefully study and *memorize* the definitions. A majority of mistakes are caused by forgetting what a certain term means.

We urge the reader to be bold. Don't be afraid of a long involved calculation.

Answers which pop out in a step or two hide the true meaning of what is going on. As far as possible, try to do the derivations yourself. If you get stuck, look up or ask. The derivations are not to be memorized, they should be done as a fresh exercise in Algebra; a regular exercise is good for you!

We urge the reader to be inquisitive. Don't take anything for granted, until you understand it. Don't be ever satisfied by a single way of doing things; look for alternative shortcuts.

We also urge the reader to be creatively lazy. Look for simpler (yet correct, of course) ways of doing the same calculations. If there is a string of numerical

calculations, don't just do them. Try to build a formula of your own; perhaps something that you could then feed into a computer some day.

A warning about graphing. Graphs are a big help in understanding the problem and they help you set up the right questions. They are also notorious in misleading you into wrong configurations or suggesting possible wrong answers. Never trust an answer until it is verified by theory or straight calculations.

Calculators are useful for getting answers but in this course, most questions are designed for precise algebraic answers. A computer system which is capable of infinite precision calculations can be used for study and is recommended. But make sure that you understand the calculations well.

What is Algebra?

Algebra is the part of mathematics dealing with manipulation of expressions and solutions of equations. Since these operations are needed in all branches of mathematics, algebraic skill is a fundamental need for doing mathematics.

The name algebra itself is a shortened form of the title of an old Arabic book on algebra entitled *al jabr wa al mukābalā* which roughly means manipulation (of expressions) and comparison (of equations).

It is the foundation and the core of higher mathematics. A strong foundation in Algebra will help students to become better mathematicians, analyzers, and thinkers.

The fundamental operations of algebra are addition, subtraction, multiplication and division (except by zero). An extended operation derived from the idea of repeated multiplication is the exponentiation (raising to a power).

We start Algebra with the introduction of variables. **Variables** are simply symbols used to represent unknown quantities. Next we combine these **variables** into mathematical **expressions** using algebraic operations and numbers. In other words, we learn to handle them just like numbers. The real power of Algebra comes in when we learn tools and techniques to solve equations involving various kinds of expressions for appropriate values.

Often, the expected solutions are just numbers, but at higher levels, we need to develop a finer idea of what we can accept as a solution. This equation solving is **the art of Algebra!** ¹

¹Here is thought for the philosophical readers. Think what you mean when you say that $\sqrt{2}$ is a solution to the equation $x^2 - 2$. Do we really have an independent evidence that $\sqrt{2}$ makes sense, other than saying that it has square equal to 2? And is it not just an alternate way of saying that it is a solution of the desired equation? Now having thought of a name $\sqrt{2}$ for the answer, we can proceed to compare it with other known numbers, find various decimal digits of its expansion and so on.

In higher mathematics, we find a way of turning this abstract thought into solutions of equations by declarations and into the art of finding the properties of solutions of equations without ever

A student in a College Algebra course is expected to be familiar with such operations and sufficiently skilled in performing them with ease and speed. The first homework and diagnostic test will help evaluate the level of these skills.

We next discuss a few generalities which will be useful for a general understanding. If things begin to become frustrating, remember that learning Algebra is much like learning a new language.

It is important to first have all the letters, sounds, words, and punctuation down before you can make a correct sentence, or in our case, write and solve an equation.

Of course, there are people who can pick up a language without ever learning grammar, by being good at imitating others and picking up on what is important. You may be one of these few lucky ones, but be sure to verify your feeling with the correct rules. Unlike a spoken language, mathematics has a very precise structure!

solving them! You will see some examples of this later.

Chapter 1

Polynomials, the building blocks of algebra

1.1 Underlying field of numbers.

To begin our work in Algebra, we must first each use the same notations to communicate. Mathematicians do this by agreeing to a **field** of numbers to work in. In our **field** we will want numbers which allow all of the four basic operations: addition, subtraction, multiplication, and division (except by 0). For our purpose, we usually include the set of all real numbers denoted by $\mathbb{R}$.

The set of **rational numbers** consists of all numbers of the form $\pm a/b$ where a, b are non negative integers with a non zero b . It is denoted by $\mathbb{Q}$. A more familiar set for you is probably $\mathbb{R}$ the set of **real numbers** which may be thought of as the set of all possible decimal numbers (with expansions of unlimited length).

You might have already seen that a rational number may be expressed as a finite decimal (for example $\frac{3}{4} = 0.75$ or a repeating decimal $\frac{4}{3} = 1.333\overline{3}$)¹

In a slightly more advanced setting you may find the set of **complex numbers** $\mathbb{C}$. It is obtained from $\mathbb{R}$ by throwing in one extra number denoted as i . This number i is defined by the basic equation $i^2 = -1$. By the known algebraic operations, it is easy to see that every number of the form $u + iv$ where u, v are real numbers must be in $\mathbb{C}$. Conversely, it is not hard to convince yourself that all the four algebraic operations on numbers of the form $u + iv$ result in numbers of the same form.²

In your higher level math classes you will also encounter the set of complex numbers denoted by $\mathbb{C}$. However, we will use the complex numbers only on special occasions.

¹Recall that the line over 3 indicates the digit 3 repeated indefinitely.

²An enthusiastic reader only needs to verify the formulas:

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

It is easy to see that now all algebraic operations can be easily performed.

To verify your understanding, assign yourself the task of verifying the following:

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. $F = \mathbb{R}$.

1.2 Indeterminates, variables, parameters

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

As you may recall from Chapter 1, 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 it is typically a matter of declaration in order to decipher which is a variable and which is a constant.

Take this familiar expression for example:

$$mx + c.$$

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.

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 thru 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.

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

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! ³

1.3 Basics of Polynomials

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

Monomials

The basic definition of a monomial is this:

- Given a variable x and a constant a , let n be a non negative integer. Then an expression of the form ax^n is said to be a **monomial**.
- The constant a is said to be its **coefficient**. (Sometimes we call it the coefficient of x^n).
- The **degree** of the monomial is said to be n .
- 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. We will always make it clear if we are using such general exponents.

³In short, the distinction between an indeterminate, variable and parameter, is, like beauty, in the eye of the beholder!

For example, consider the monomial $4x^3$. Its variable is x , its coefficient is 4 and its degree is 3.

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 more than one variables and write a monomial of the form ax^py^q . This is a monomial in x, y with **coefficient** a . 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. ⁴

Example.

Consider this **example of a monomial in x, y** .

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

What are the exponents, degree etc.? First, we **must simplify** the expression to:

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

Thus 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 $-7x^2y^3$ 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.

⁴For the reader who likes formalism, here is a general definition. A monomial in variables $x_1, \dots, x_r$ is an expression of the form $ax^{n_1} \cdots x^{n_r}$ where $n_1, \dots, n_r$ are non negative integers and a is any expression which is free of the variables $x_1, \dots, x_r$. The coefficient of the monomial is a , the degree is $n_1 + \cdots + n_r$ and the exponents are said to be $n_1, \dots, n_r$.

In higher mathematics, $n_1, \dots, n_r$ may be allowed to be more general.

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 2. Its exponents are $(1, -1)$ respectively.

Binomials, Trinomials, $\dots$ and Polynomials

Recall what you learned in the previous section regarding monomials. 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.⁵

In general “poly” means many so we define:

Definition: A **polynomial** A polynomial is defined as a sum of monomials whose exponents are all non negative. Thus $x^5 + 2x^3 - x + 5$ is a polynomial.

Usually, a polynomial is written in a simplified form by combining all terms with the same exponents. Thus the same polynomial is also equal to $x^5 + x^3 + x^3 + 3x - 4x + 5$, but 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.

A polynomial with no terms has to be admitted for algebraic reasons and it is denoted by just 0 and called the **zero polynomial**.

About the notation for a polynomial. If we wish to identify the main variable of a polynomial, we may 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. 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 $w(x) = x^5 + 2x^3 - x + 5$ is 2. The coefficient of the missing monomial x^4 in $w(x)$ is declared to be 0. The coefficient of x^{100} in $w(x)$ is also 0 by the same reasoning.

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 .

⁵Can you find or make up terms for a polynomial with four or five or six terms?

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.⁶

As a simple example, let $f(x) = 3x^5 + x$, $g(x) = 2x^5 - 2x^2$, $h(x) = 3$, $w(x) = 2$. Then $f(x) + g(x) = 3x^5 + x + 2x^5 - 2x^2$ and we need to collect like terms to write the answer as $f(x) + g(x) = 5x^5 - 2x^2 + x$. 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 polynomial (i.e. any monomial ax^m with $a \neq 0$ in 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, other 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 also have use for **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 $u = ax^n + \dots$, $v = bx^m + \dots$ are **non zero polynomials** in x of degrees n, m respectively.

We shall need the following **useful facts**.

1. Note that by saying that u has degree n we are announcing 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, by saying that v has degree m , we mean that $b \neq 0$ and no other terms in v are of degree as high as m .

2. For a non zero number c we have that cu has the same degree as u and indeed its leading coefficient is ca .

⁶The 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. You might have also seen the idea of division with remainder which says divide 7 and 2 and you get a division of 3 and a remainder of 1, i.e. $7 = (2)(3) + 1$.

As promised, we shall take this up later.

For example, if c is a constant, then the degree of $c(2x^5 + x - 2)$ is always 5, except when $c = 0$. For $c = 0$ it becomes undefined! Note that for non zero c , the leading coefficient of cu is simply ca .

3. Suppose that the degrees n, m of u, v are unequal. Then for non zero constants c, d we can decide the degree of $cu + dv$ to be the maximum of n, m . It is easy to figure out the degree if one or more of c, d have special values.

For example, if $u = 2x^5 + x - 2$ and $v = -x^3 + 3x - 1$, then the degree of $cu + dv$ is calculated 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, the degree will be 5 as long as $c \neq 0$. It will drop to 3 if $c = 0$ but $d \neq 0$. If c, d are both zero then it becomes undefined. It is useful to understand this using concrete values of c, d .

The reader should verify that the leading coefficients are respectively, $2c$ or $-d$ or undefined.

4. Let c, d be non zero numbers and consider $h = cu + dv$. If the degrees of u, v are the same, i.e. $n = m$ then the degree of h is not so easy to tell without careful analysis.

However, we can at least say that

- 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.

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

We determine the degree of the expression thus:

- If $-c + d \neq 0$, i.e. $c \neq d$ then the degree is 3.
- If $c = d$, the 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.

The reader is encouraged to figure out the formula for the leading coefficients in each case.

The reader 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

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

Thus the leading coefficient of the product is ab , i.e. the product of their leading coefficients.

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:

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

Answers: The degrees of $p(x), q(x)$ are both 3. The remaining answers are:

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

Here is an alternate technique for 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 times x^3 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:
You are given that:

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

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 $x^{20}3x^{20}$, so $a = 3$. 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 $d = 10$.

Thus, 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 $p5$ 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? Our $n = pd + 5$, so

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

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 $u(x) = x + t$. Calculate the powers $u(x)^n$ for $n = 1, 2, 3, 4, 5$. For each answer, just list the coefficients of powers of x starting with x^0 .

Answers:

Polynomial	1	x	x^2	x^3	x^4	x^5
$(x+t)^1$	t	1	0	0	0	0
$(x+t)^2$	t^2	$2t$	1	0	0	0
$(x+t)^3$	t^3	$3t^2$	$3t$	1	0	0
$(x+t)^4$	t^4	$4t^3$	$6t^2$	$4t$	1	0
$(x+t)^5$	t^5	$5t^4$	$10t^3$	$10t^2$	$5t$	1

There is a visible pattern here. To get a certain coefficient, you can multiply the one above it by t and add to it the coefficient in the northwest corner (i.e. the “above left” entry, if present)! Verify this for the given examples. ⁷

Here is a little hint on how to do these calculations efficiently.

We all know that $(x+t)^2 = x^2 + (2t)x + t^2$. Think of $(x+t)^3$ thus:

$$\begin{aligned} (x+t)^3 &= (x+t)(x+t)^2 = (x+t)(x^2 + (2t)x + t^2) \\ &= x(x^2 + (2t)x + t^2) + t(x^2 + (2t)x + t^2) \\ &= x^3 + (2t)x^2 + (t^2)x + (t)x^2 + (2t^2)x + (t^3) \\ &= x^3 + (2t+t)x^2 + (t^2+2t^2)x + (t^3) \\ &= x^3 + (3t)x^2 + (3t^2)x + t^3 \end{aligned}$$

⁷Challenge for the smart readers: Try to prove this for a general power, by thinking of $(x+t)^{n+1}$ as $(x+t)^n(x+t)$ and collecting the coefficients in the product.

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.

Here is the method in action: ⁸

$$\begin{array}{r} x^3 \quad (2t)x^2 \quad (t^2)x \\ \quad (t)x^2 \quad (2t^2)x \quad t^3 \end{array}$$

Now we simply add the terms. To illustrate the power of this thought, let us record the two rows needed to get $(x + t)^6$ from the known $(x + t)^5$.

$$\begin{array}{r} x^6 \quad (5t)x^5 \quad (10t^2)x^4 \quad (10t^3)x^3 \quad (5t^4)x^2 \quad (t^5)x \\ \quad (t)x^5 \quad (5t^2)x^4 \quad (10t^3)x^3 \quad (10t^4)x^2 \quad (5t^5)x \quad t^6 \end{array}$$

By combining, we get the answer:

$$(x + t)^6 = x^6 + (6t)x^5 + (15t^2)x^4 + (20t^3)x^3 + (15t^4)x^2 + (5t^5)x + t^6.$$

- 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

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

⁸An 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 the numbers as polynomials in 10, but we have to worry about carries. Here is a

sample multiplication for your understanding:

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

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
$\frac{x+2}{(x+5)(x+1)}$	$x-5$	$\frac{x-3}{(x)(x-4)}$

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 $x \rightarrow x + s$ such that the transformed polynomial $q(x + s)$ has no x -term (i.e. a monomial in x of degree 1).

Find the resulting polynomial.

Answer. Check out the work:

$$q(x+s) = a(x+s)^2 + b(x+s) + c = ax^2 + (2as+b)x + (as^2+bs+c).$$

What we want is 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.

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

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 $p(x - 5/6)$ will have no x^2 term.

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

- Given a cubic $p(x) = ax^3 + bx^2 + \text{lower terms}$ with $a \neq 0$, find the value of s such that $p(x + s) = ax^3 + (0)x^2 + \text{lower terms}$.
- 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 $p(x + s) = ax^n + (0)x^{(n-1)} + \text{lower terms}$.

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 called linear equations.

Definition: Linear Equation An equation is said to be linear in its listed variables $x_1, \dots, x_n$, 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 $x_1, \dots, x_n$.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 .
3. The same expression is also linear in y, z, w . Indeed, it is linear in any of the set of variables as long as the set does not include both x, y !

We can, however, say that $xy + z = 5w$ is linear in x, z, w . It may also be called linear in y, z, w .

¹A simple way of describing this is to say that the degree of the expressions in the equation is 1; however, an alert reader should note that our linear equation is technically allowed to be degree zero, i.e. the supposed variable 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!

We can also declare it as simply linear in y . What we are trying to emphasize we don't have to list every visible symbol as a variable.

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.

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 simply simplified the equation in the following steps:

$3x + 4$	$=$	$-2x + 14$	original equation
$3x + 2x + 4$	$=$	14	collect x terms on left
$3x + 2x$	$=$	$-4 + 14$	collect constants on right
$5x$	$=$	10	simplify
x	$=$	2	cancel 5 from both sides

Usually, the above process is simply 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.

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 left hand side (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 0, 1, ∞ 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. A unique solution,
2. No solution (inconsistent system) or
3. Infinitely many solutions.

Example:

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

can be rearranged to:

$$x(-5+2-7-9) = -8-3 \text{ or even more simplified } -19x = -11 \text{ and hence } x = 11/19.$$

a **unique** solution

The reader should note the efficient collection of terms. 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** solution.

Example:

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

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

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

²Even 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.

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

Example: $2x = 4 \rightarrow x = 2$ $6x = 8 \rightarrow x = 4/3$ Since the equations give us different answers 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 invalid 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

$$20 - 4y - y = 5$$

collecting like terms we get

$$-5y = -15.$$

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 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 for 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 there is no solution.

Example: If we change our problem to:

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

then we get consistent equations with the known solution $x = 4, y = 3$.

2.5 Several equations in several variables.

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

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

We shall follow the above idea of solving one variable at a time.

We solve Eq1 for x to get

Solution1.
$$x = 4 - 2y - z$$

Substitution into Eq2, Eq3 gives two 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 has no x anyway, so it stays $y + z = 7$.

Now we solve

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

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 to get

Solution2.
$$y = 7 - z$$

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

Solution3.
$$z = 10$$

Notice our three solutions. The third pins z down. 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.
- 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.

Solve E1: $2x + 5y = 3$, 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.

The 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.

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:

$$x(4 - 2(2)) + y(9 - 2(5)) = 5 - 2(3) \text{ or, after simplification } -y = -1.$$

³The 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 them somehow and try to get a sequence of solution giving each variable in terms of the later ones. The resulting set of solution equations is often called the row echelon form.

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

$ax + by = c$ and $px + qy = r$ is given by

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

There is an easy way to remember this formula, which is what makes it most useful.

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 example again.

Here the denominator is

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

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

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

So $x = 2/(-2) = -1$ and $y = (-2)/(-2) = 1$.

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 $\Delta = aq - pb = 0$.⁴

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.

We shall illustrate this by some examples.

- Example 3.

Solve the system of equations:

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

for various values of the parameter k .

⁴Thus Δ lets us “determine” the nature of the solution. This is the reason for the term “determinant”.

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 ⁵

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

We see that when $k \neq 10$ then this leads to illegal values for x and y , while if $k = 10$ then we are left with $x = \frac{0}{0}$ and $y = \frac{0}{0}$.

We should not declare an answer right away, but try 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 by saying:

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: ⁶

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

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.

⁵Technically, what we wrote is illegal! A true algebraist would insist on writing the equations as

$$(\Delta)x = \Delta_x \text{ or } (0)x = 10 - k$$

and

$$(\Delta)y = \Delta_y \text{ or } (0)y = 2k - 20$$

⁶We shall see such answers when we study the parametric forms of equations of lines later.

Solve the three equations in three variables:

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

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 $\Delta = -3 = \Delta_x$, so $x = \frac{-3}{-3} = 1$.

Also, we get

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

Finally, we have

$$\Delta_z = \begin{vmatrix} 1 & 1 & 6 \\ 1 & -2 & 0 \\ 2 & -1 & -3 \end{vmatrix} = -9.$$

This gives us $y = \frac{-6}{-3} = 2$ and $z = \frac{-9}{-3} = 3$. The reader can at least verify the answers directly by known methods.

Chapter 3

The division algorithm and applications

We have discussed addition, subtraction and multiplications of polynomials earlier. 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 is easy to see that the ratio $\frac{u(x)}{v(x)}$ is not a polynomial. To understand this, suppose 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 .¹

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:

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

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 of the equations?

¹Note that $\deg_x(u(x)) = \deg_x(w(x)) + \deg_x(v(x))$ or $3 = \deg_x(w(x)) + 2$. Hence the conclusion.

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 $u(x) = ax^n + \dots$ and $v(x) = bx^m + \dots$ 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
3. $\deg_x(r(x)) < \deg_x(v(x))$.
4. When the above conditions are satisfied, we declare that $q(x)$ is the **division** and $r(x)$ is the **remainder** when we **divide $u(x)$ by $v(x)$** .
5. The division $q(x)$ and the remainder $r(x)$ always exist and are uniquely determined ² by $u(x)$ and $v(x)$.

First, we explain 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 $3 - 2 = 1$.

What we want is to arrange $u - qv$ to have a degree as small as possible; this means a degree of less than 2 or it may become zero!

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

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

²Hint for the proof: Suppose that $u - q_1v = r_1$ and $u - q_2v = r_2$ where r_1, r_2 are either 0 or have degree 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!

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: ³

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

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

$$u(x) - q(x)v(x) = (-(-1))x + (-(-1)) = x + 1 = r(x).$$

Let us **summarize this process**:

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.)
2. Start with a guess $q(x) = ax^{(n-m)}$ and choose a such that $u - qv$ has degree less than n .
3. The value of a will come out to be the leading coefficient of u divided by the leading coefficient of v . This formula is useful, but often it is easier to find a by inspection.
4. If the degree of $u - qv$ is less than m or if it is zero, then stop and call it the final r .
5. If not, add a next term to q to make the degree of $u - qv$ even smaller.
6. Continue until $u - qv$ becomes 0 or its degree drops below m . What we are describing is the process of long division that you learnt in high school. We are going to learn more efficient methods for doing it below.

There are two **lucky situations** where we need no more work! Let us record these for completeness.

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.
3. 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 now begin with an extension of the division algorithm.

³Note 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.

3.1 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 division (or the quotient) and the remainder.

The assignment:

Given

$$u(x) = x^3 + x + 1 \text{ and } v(x) = -x^3 + x^2 - 2x + 2,$$

calculate the division $q(x)$ and the remainder $r(x)$ when you divide $u(x)$ by $v(x)$.

Answer:

We need to make $u(x) - q(x)v(x)$ to have degree less than 3.

We see $u(x) - (-1)v(x) = 3x^2 + 6x + 6$, so $r(x) = 3x^2 + 6x + 6$ and $q(x) = -1$.

We recommend the following arrangement of the above work, for future use.

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

We are writing the two given polynomials one below the other. Then write the result of the division of the top polynomial (dividend) by the bottom polynomial (divisor) on the right of the divisor and finally enter the remainder as a new polynomial below the divisor. Note that the actual simplification process is not recorded and is to be done on the side.

Additional Task: Now we repeat these types of steps as instructed below.

Continue the above division process by dividing a lower polynomial into the one immediately above it, writing the division on the right and entering the remainder as a new polynomial below it.

Continue until the lowest polynomial becomes 0, the zero polynomial. We first show the work and then explain its use.

Answer: Here is a graphic display of our answers. The idea is to divide a polynomial by the one below and write the remainder below it. We record the “divisions” to the right of the divisor.

$$\left[\begin{array}{c|c|c} \text{Step no.} & \text{Polynomials} & \text{Divisions} \\ 0 & x^3 + x + 1 & \text{Begin} \\ 1 & -x^3 + x^2 - 2x + 2 & -1 \\ 2 & x^2 - x + 3 & -x \\ 3 & x + 2 & x - 3 \\ 4 & 9 & \frac{x}{9} + \frac{2}{9} \\ & 0 & \text{End of process} \end{array} \right]$$

The step numbers are not essential and we shall drop them in future.

We are calculating these by the long division.

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 reader is invited to deduce that our polynomials $u(x), v(x)$ do not have a common (non constant) factor by the following steps:

The equation $u(x) - q(x)v(x) = r(x)$ says that any common factor of $u(x), v(x)$ must be a factor of $r(x)$. Thus a common factor of the top two polynomials in our table above must also divide the polynomial just below them (or the remainder). Repeating this idea, we see that the common factor must divide the last polynomial above the zero polynomial.

3.2 The GCD and LCM of two polynomials: meaning and calculations.

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 we have to let go of the 'monicness' condition. 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 gets in trouble! As in the case of GCD, we 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 $a(x)v(x) - b(x)u(x)$.

Below, we present the technique to find such a combination. The proof is left for the inquisitive reader to finish.

For our polynomials above, the GCD will be the monic multiple of 9 and so the GCD of the given $u(x), v(x)$ is said to be 1.⁴

You may have seen a definition of the GCD of polynomials which tells you to factorize 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. Our algorithm can also be adopted for finding the GCD of two given integers. Indeed that was the original formulation and application. We shall explain it below.

3.3 Āryabhata algorithm: Efficient Euclidean algorithm.

Āryabhata algorithm gives us a way to solve the following problem.⁵

Suppose that we are given two polynomial $u(x), v(x)$ and wish to express a third polynomial $h(x)$ as their combination

$$h(x) = a(x)v(x) - b(x)u(x).$$

First let us study a very simple example. If we let $u(x) = x$ and $v(x) = x^2 + x$ then we cannot write

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

⁴If you are confused about how 9 becomes 1, remember that the only monic polynomial of degree 0 is 1.

⁵Indeed, 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!

The simple reason is that if we set $x = 0$ the right hand side becomes 0 while the left hand sides stays 1, so the two sides cannot be equal! ⁶

Thus, there is a condition on $h(x)$. What is it?

Let $d(x)$ be the GCD of $u(x), v(x)$, so using our notation, we shall write

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

It is clear that $d(x)$ divides the right hand side of the above equation for $h(x)$ and hence must divide $h(x)$.

Thus, the necessary condition is that the GCD of the two polynomials divides $h(x)$. Conversely, if $d(x)$ divides $h(x)$, then will soon show how to write $h(x)$ as a combination of $u(x)$ and $v(x)$ as desired.

Illustration of the method.

We shall show how to write $h(x) = 1$ as a combination $b(x)v(x) - a(x)u(x)$ where $u(x) = x^3 + x + 1$ and $v(x) = -x^3 + x^2 - 2x + 2$. We already know that their GCD is 1, so our condition is valid.

The general method is similar and easily deduced from our description.

1. Start with the display of the division algorithm. By assumption, $h(x)$ is the product of some $w(x)$ and the last polynomial in our table just above the 0. For our example we get:

$$h(x) = 1 = \frac{1}{9}(9) - 0(0).$$

We begin a new column of answers next to the polynomials' column by writing $w(x) = \frac{1}{9}$ at the bottom and 0 above it.

Thus we have the beginning of the answers' column:

Answers	Polynomials	Divisions
	$x^3 + x + 1$	Begin
	$-x^3 + x^2 - 2x + 2$	-1
	$x^2 - x + 3$	$-x$
	$x + 2$	$x - 3$
0	9	$\frac{x}{9} + \frac{2}{9}$
$\frac{1}{9}$	0	End of process

⁶This is a very efficient way of testing equality of two polynomials without the tedious task of evaluating and comparing all their coefficients. Just substitute lots of values on both sides and see if they give the same answer. If they ever fail to give the same answer, then they are unequal.

How many values do we need to check before we can declare them equal? It is possible to argue that one more than their maximum degree is the magic number of values. This is left as a challenging exercise!

2. How do we fill in the rest of the answers' column? The simple rule is: **To create a new answer, multiply the answer below it by the corresponding division on the right and add in the answer two steps below.**

It is important to understand and execute this rule without fail. It is the only calculation which can go wrong, if you are careless!

Let us calculate the new answer above 0.

First multiply 0 with the entry with the corresponding "divisions" entry:

$$(0)\left(\frac{x}{9} + \frac{2}{9}\right) = 0.$$

Add to it, the answer $\frac{1}{9}$ below the 0, i.e. the answer two steps below!

Thus the final new answer is:

$$(0)\left(\frac{x}{9} + \frac{2}{9}\right) + \frac{1}{9} = \frac{1}{9}.$$

Now update our answers' column.

Answers	Polynomials	Divisions
	$x^3 + x + 1$	Begin
	$-x^3 + x^2 - 2x + 2$	-1
	$x^2 - x + 3$	$-x$
$\frac{1}{9}$	$x + 2$	$x - 3$
0	9	$\frac{x}{9} + \frac{2}{9}$
$\frac{1}{9}$	0	End of process

Another step gives the new answer as

$$\left(\frac{1}{9}\right)(x - 3) + 0 = \frac{x - 3}{9}.$$

Check how this works out.

Update the answers' column:

Answers	Polynomials	Divisions
	$x^3 + x + 1$	Begin
	$-x^3 + x^2 - 2x + 2$	-1
	$x^2 - x + 3$	$-x$
$\frac{x-3}{9}$	$x + 2$	$x - 3$
$\frac{1}{9}$	9	$\frac{x}{9} + \frac{2}{9}$
0	0	End of process

We show the result of next two steps at once.⁷

Answers	Polynomials	Divisions
$\frac{x^2-2x-4}{9}$	$x^3 + x + 1$	Begin
$\frac{-x^2+3x+1}{9}$	$-x^3 + x^2 - 2x + 2$	-1
$\frac{x-3}{9}$	$x^2 - x + 3$	$-x$
$\frac{1}{9}$	$x + 2$	$x - 3$
0	9	$\frac{x}{9} + \frac{2}{9}$
$\frac{1}{9}$	0	End of process

3. Now comes the reward for our calculations. Let

$$a(x) = \frac{x^2 - 2x - 4}{9} \text{ and } b(x) = \frac{-x^2 + 3x + 1}{9}.$$

Recall that our original polynomials are $u(x) = x^3 + x + 1$ and $v(x) = -x^3 + x^2 - 2x + 2$ and we wished to write $h(x) = 1$ as their combination.

Then our answer is

$$a(x)v(x) - b(x)u(x) = \pm 1 = \pm h(x).$$

It is easy to figure out whether the sign is $+$ or $-$ by checking some term. In our example, if we compute the constant term of $a(x)v(x) - b(x)u(x)$, we see $(-4/9)(2) - (1/9)(1) = -1$, so we get $-h(x)$. The correct answer to the problem is then:⁸

$$\begin{aligned} h(x) = 1 &= -(a(x)v(x) - b(x)u(x)) \\ &= b(x)u(x) - a(x)v(x) \\ &= \frac{-x^2+3x+1}{9}(u(x)) - \frac{x^2-2x-4}{9}(v(x)) \end{aligned}$$

Example 1: Determine a polynomial $P(x)$ such that when divided by $x^2 - 1$ it gives a remainder x and when divided by $x^3 - 2$ it gives a remainder $1 - x$.

Answer: From the given conditions we get:

⁷For the reader's convenience, here is the explanation of the very top answer $\frac{x^2-2x-4}{9}$.

Answer below is $\frac{-x^2+3x+1}{9}$ and its corresponding divisions' entry is -1 . The answer two steps below is $\frac{x-3}{9}$, so we calculate:

$$\left(\frac{-x^2+3x+1}{9}\right)(-1) + \frac{x-3}{9} = \frac{x^2-3x-1+x-3}{9} = \frac{x^2-2x-4}{9}.$$

⁸An alert reader should note the determinant popping up in our expression. It is the topmost two by two determinant. **Rule of the sign.** It is possible to make a rule for the sign and it is this: The sign is $+$ if the columns have an odd number of terms and $-$ if they have even number of terms. It is, however, more reliable to actually check the sign.

1.

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

and

2.

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

A simple subtraction gives:

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

Thus we wish to write $h(x) = 2x - 1$ as a combination of the polynomials

$$u(x) = x^3 - 2 \text{ and } v(x) = x^2 - 1.$$

We apply the above Āryabhata-Euclid algorithm. The reader should check these steps:

Step no.	Polynomials	Divisions
0	$x^3 - 2$	Begin
1	$x^2 - 1$	x
2	$x - 2$	$x + 2$
3	3	$\frac{x-2}{3}$
	0	End of process

Thus, we see that the GCD of the two polynomial divides 3 and hence is 1.

Now we do the algorithm to write our $h(x) = 2x - 1$ as a combination as explained by the Āryabhata algorithm.

This means, we begin by writing the desired $2x - 1$ as a combination of the last two polynomials 3 and 0 in the simplest possible manner. Namely

$$2x - 1 = \frac{2x - 1}{3}(3) - 0(0).$$

We write the coefficients 0 and $\frac{2x-1}{3}$ as the last two answers and lift them in the answers' column as before. ⁹

⁹As earlier, we explain the top entry $\frac{2x^3 + 3x^2 - 1}{3}$. It is obtained by calculating:

$$\text{multiply } \frac{2x^2 + 3x - 2}{3} \text{ with } x \text{ and add } \frac{2x - 1}{3}.$$

Note the evaluation:

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

Answers	Polynomials	Divisions
$\frac{2x^3+3x^2-1}{3}$	$x^3 - 2$	Begin
$\frac{2x^2+3x-2}{3}$	$x^2 - 1$	x
$\frac{2x-1}{3}$	$x - 2$	$x + 2$
0	3	$\frac{x-2}{3}$
$\frac{2x-1}{3}$	0	End of process

Check that by setting

$$a(x) = \frac{2x^3 + 3x^2 - 1}{3} \text{ and } b(x) = \frac{2x^2 + 3x - 2}{3}$$

we actually get

$$a(x)v(x) - b(x)u(x) = 2x - 1 = h(x).$$

What is the desired polynomial $P(x)$?

Recall that we had $P(x) = a(x)(x^2 - 1) + x$ so we get

$$P(x) = \frac{2x^3 + 3x^2 - 1}{3}(x^2 - 1) + x = \frac{3x^4 - 4x^2 + 2x^5 - 2x^3 + 1 + 3x}{3}.$$

The Chinese Remainder Theorem or Kuttaka Our work above is the basis of what is known as the Kuttaka in ancient Indian Mathematics. It is commonly known as the Chinese Remainder Theorem in modern mathematics books.

The idea is to find a polynomial $f(x)$ which has desired remainders $r(x), s(x)$ respectively when divided by two given polynomials $u(x)$ and $v(x)$.

We work as before, writing two equations:

$$f(x) = b(x)u(x) + r(x).$$

and

$$f(x) = a(x)v(x) + s(x)$$

After subtraction of the first equation from the second and a rearrangement, we get the equation

$$a(x)v(x) - b(x)u(x) + s(x) - r(x) = 0 \text{ or } r(x) - s(x) = a(x)v(x) - b(x)u(x).$$

You can see how the above algorithm can now be conveniently applied. We can also extend the theorem to several polynomials, but that would take us too far.

Before turning to another topic, let us give an easier and older application of this process for working with integers. Indeed, this is the more common application, rather than the polynomial version.

3.4 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 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 division 4 and remainder 3. In equation form, we write this as

$$23 = (4)5 + 3.$$

For the readers who are very fond of their calculators, here is a simple recipe for finding the division and the remainder.

- Have the calculator evaluate u/v . the division 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 $u = 1 + 3 + 3^2 + 3^3 + \cdots + 3^{100}$ 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 and may get error messages from the calculator.

A person who looks at the form can easily see that

$$u = 1 + 3(1 + 3 + \cdots + 3^{99}) = r + 3q.$$

So we can happily declare: $r = 1$ and $q = (1 + 3 + \cdots + 3^{99})$.

You may not have seen much of the variations which we discuss next.

Our choice u can be negative, say -23 and then we get:

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

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

The calculator person 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 bigger than or equal to zero.

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

For better understanding, we record the definitions.

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 the division and r is the remainder.

To continue our example above, let us now take $v = -5$ and note that $|v| = |-5| = 5$, so we get the following:

If $u = 23$ then

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

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 taking $q = -5$ makes a negative r in this case!

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, if $v \neq 0$ then we see that $v|u$ exactly means that the remainder r when u is divided by v becomes 0.

It is clear that the only integer divisible by 0 is 0.

Exercise. We leave it as an exercise to make a precise definition for divisibility of polynomials and provide suitable examples.

Definition: Divisibility of polynomials. *Definition and examples to be supplied by the reader!*

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 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 $d > 0$.

As before, for convenience, we define:

Definition: LCM of integers. Given two integers u, v we say that an integer L is their 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 $L > 0$.

As before, we have a simple fact:

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:

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

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 $d = 0$ 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 the Āryabhata-Euclid algorithm. Moreover, we get it in the form of a combination of u and v .

We show this by the following:

Example 2: Find the GCD of 18 and 14 and write it as a combination of 18 and 14.

Moreover, find an integer n such that n gives remainder 1 when divided by 18 and a remainder of 5 when divided by 14.

Answer: We apply the same algorithm as we did in polynomials. The reader is encouraged to fill in the justification.

1. Finding the GCD. Just imitate the Āryabhata algorithm!

Step no.	Integers	Divisions
0	18	Begin
1	14	1
2	4	3
3	2	2
	0	End of process

The GCD is the number 2 above the 0.

2. Writing the GCD as a combination of the given integers. Again, imitate the same steps as in the polynomial case.

Answers	Integers	Divisions
4	18	Begin
3	14	1
1	4	3
0	2	2
1	0	End of process

Thus, as before we have $(4)14 - (3)18 = \pm 2$. By direct evaluation, we see that indeed we have a plus sign, i.e.

$$2 = (4)14 - (3)18.$$

3. For the last problem, we write equations as before:

$$n = (b)18 + 1 \text{ and } n = (a)14 + 5.$$

Subtracting the first from the second and rearranging, we get

$$(a)14 - (b)18 = -4.$$

First we note that this -4 is divisible by our GCD, so we are going to succeed. We start with:

$$\left[\begin{array}{c|c|c} \text{Answers} & \text{Integers} & \text{Divisions} \\ \hline & 18 & \text{Begin} \\ & 14 & 1 \\ & 4 & 3 \\ 0 & 2 & 2 \\ -2 & 0 & \text{End of process} \end{array} \right]$$

So the desired -4 is written as a combination of the last two numbers $2, 0$ in the form

$$-4 = (-2)2 - (0)0.$$

We lift the answer column by the same recipe as before:

$$\left[\begin{array}{c|c|c} \text{Answers} & \text{Integers} & \text{Divisions} \\ \hline -8 & 18 & \text{Begin} \\ -6 & 14 & 1 \\ -2 & 4 & 3 \\ 0 & 2 & 2 \\ -2 & 0 & \text{End of process} \end{array} \right]$$

Thus we see that our final answer is likely to be

$$(-8)14 - (-6)18 = \pm(-4).$$

A simple check shows that

$$(-8)14 - (-6)18 = -4.$$

This gives us $a = -8, b = -6$ and hence

$$n = (b)18 + 1 = (-6)18 + 1 = -107.$$

We found an answer, but perhaps we want a positive answer. Is there one?

Indeed, there is! We can add any multiple of the $\text{LCM}(14, 18)$ to our answer and it stays a valid answer.

Since the LCM is $(14)(18)/2 = 126$ we use it. (Note that we are using the LCM formula given above.)

So $126 - 107 = 19$ is a possible answer as well! Indeed 19 does give the remainder 1 when divided by 18 and remainder 5 when divided by 14. So, here a trial and error method could have been more efficient! However, we cannot, however, count on trial and error!

Example 3: Let a be a constant. Find the division and remainder when $(x - a)$ divides various powers of x , like $x, x^2, x^3, x^4, \dots$.

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:

If we can *somehow guess answers* $q(x), r(x)$ such that

•

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

• and either $r(x) = 0$ or the degree of $r(x)$ is less than that of $v(x)$,

then we have found our division and remainder! No long division is actually needed.

If we divide $(x - a)$ by $(x - a)$, we clearly have $q(x) = 1, r(x) = 0$.

In view of the identity

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

we get by rearrangement:

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

so, dividing x^2 by $x - a$ gives division $(x + a)$ and remainder a^2 .

To divide x^3 by $(x - a)$, we start with the above relation and multiply it by x , so:

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

i.e. we get

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

This simplifies to:

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

The reader is invited to repeat this idea and deduce

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

or

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

It is not hard to imagine and prove the identity:

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

Thus in general the remainder of x^n is a^n . 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 $f(a)$.¹⁰

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

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

For proof, proceed as follows:

- Start with the known identity:

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

- Set $a = 1$ and deduce:

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

- Rewrite as:

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

Divide by $(x - 1)$ and finish!

¹⁰For proof, the reader may try this idea.

Given a divisor $v(x)$, a general remainder of a polynomial $u(x)$ shall be defined as any polynomial of the form $u(x) - h(x)v(x)$ and a good remainder is one which is either 0 or of degree smaller than that of $v(x)$. Thus the division algorithm simply asks us to find the unique good remainder! Note that by definition, a constant times a good remainder is again a good remainder and so is a sum of two good remainders. It is easy to show that this finishes the proof.

Example 4: Let a be a constant. Find the division and the remainder when $(x^2 - a)$ divides various powers of $x, x^2, x^3, x^4, \dots$.

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 $n = 2$

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

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 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 as:

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

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.

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 of calculations, it is not so helpful for visualization and intuition for a beginning student. Also, many important motivations for algebra originally came from geometric ideas. 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 points 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.

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 a 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.

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 and the coordinate of the point where it hits the x axis is called the x coordinate of P .

Similarly a line is drawn through P parallel to the x axis and its intersection with the y axis is 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.

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 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 $A(5), B(2), C(-1)$, the distance between A and B is

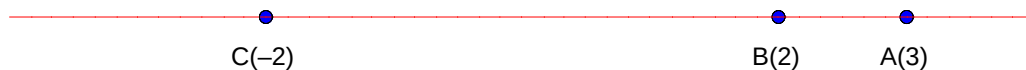
$$d(A, B) = |3 - 5| = |-2| = 2.$$

The distance between A and C is

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

Finally, the distance between C and B is

$$d(C, B) = |(2) - (-1)| = |3| = 3.$$



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 to denote this formally as $\overrightarrow{AB}$.¹

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.

¹What 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.

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

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

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.

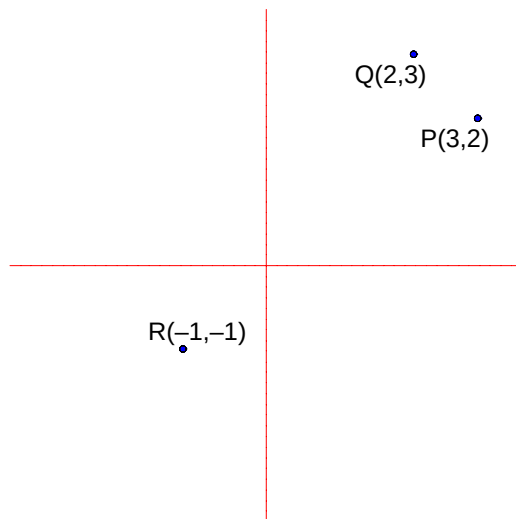
Distance in the plane.

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 $(a_2 - a_1, b_2 - b_1)$ 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 in it! Now, the distance between the two points has to be computed by a more complicated formula using the Pythagorean Theorem and we have:

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

To make it look similar to a line, it is customary to denote this distance by $|\overrightarrow{P_1P_2}|$. 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.

For example, now consider the points $P(3, 2), Q(2, 3), R(-1, -1)$. Study their display in the plane:



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 R, S , we get $\overrightarrow{RS} = (-1 - 3, -1 - 2) = (-4, -3)$ and hence the distance $d(R, S) = 5$ also.

We can similarly handle points in a three dimensional space and set up shifts and distance formulas. The third and higher dimensions can be handled as needed.

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 is.

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 illustrate this by an example.

Let us say that we have a coordinate system so that each point $P(a)$ has an assigned coordinate. We would like to choose a new origin at the point $Q(3)$ - the point with current coordinate 3. Let x be the value of the original coordinate of a point and let x' be its new coordinate. Then we claim that

$$x' = x - 3$$

We shall say that under this coordinate change the point $A(x)$ goes to the point $A'(x - 3)$.

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.²

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.

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 .

²An alert reader should observe that the distance between two points $P_1(a_1), P_2(a_2)$ is given by $|a_2 - a_1|$. So, in new coordinates it becomes $|(a_2 - 3) - (a_1 - 3)| = |a_2 - a_1|$, or is unchanged.

Indeed, the new coordinate also preserves all shifts.

For example, let $a = 3, b = -1$. Consider the points $P(2), Q(5)$. The new coordinate of P shall be $3(2) - 1 = 5$.

Similarly for the point Q we get $3(5) - 1 = 14$.

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 can easily generalize the above formulas and get similar results. 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 (up, vq) .

All distances are preserved. However, shifts can change and we list their properties:

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, vq) as the new origin followed by the indicated flip(s).

The proof simply consists of comparing effects on points. This is best done by plotting the original points and their images on the same graph paper.

As an illustration, let us see what a transformation $x' = -x - 5$, $y' = y - 4$ does. The point $P(-5, 4)$ becomes $(0, 0)$ or the new origin.

All points of the form $(a, 4)$ get transformed to $(-a - 5, 0)$ i.e. become the new x' axis. The point $(-5 + 1, 4) = (-4, 4)$ 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 $(-5, b)$ will map to the $(0, b - 4)$ to become the new y' axis. There is no flip involved, since the point $(-5, 5)$ with shift $(0, 1)$ from P goes to $(0, 1)$.

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 $P_1(p + 1, q)$ and $P_2(p, q + 1)$.
- (c) Compare the original triangle PP_1P_2 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. We will, however, leave the analysis for higher courses. Here we only state the most important facts for information and as a challenge for the super smart ones.

The most general coordinate change in the plane can be written as

$$x' = a(x - p) + b(y - q) , \quad y' = c(x - p) + d(y - q).$$

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. We, however, choose to postpone these discussions to a higher course. We will give a brief sample of this in a later chapter.³

³Here are the details for the inquisitive.

We know how to translate and so for this part, 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

$$x' = \cos(\theta)x + \sin(\theta)y , \quad y' = -u \sin(\theta)x + u \cos(\theta)y.$$

Examples of changes of coordinates:

We illustrate the above discussion 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:

- 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 $d = -7$.

The transformation can be described as a rotation by θ if $u = 1$ and this is 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 $0 < \theta < 90$. Make a triangle ABC where the angle at B is the desired θ and the one at C is 90 degrees. Then

$$\cos(\theta) = |\overrightarrow{BC}|/|\overrightarrow{BA}| \text{ and } \sin(\theta) = |\overrightarrow{CA}|/|\overrightarrow{BA}|.$$

For other values of angles, these are defined by suitable adjustments. We will give a more general modern definition during the discussion of the circles.

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 :

$$4 = u - p, \quad 1 = v - q.$$

Similarly for the point Q , we get:

$$5 = 2u - p, \quad 1 = 3v - q.$$

The reader can verify the solution:

$$u = 1, p = -3, v = -1, q = -2.$$

Thus the transformation is:

$$x' = x + 3, \quad y' = -y + 2.$$

- 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.

Given a point $P(a, b)$ in the plane, it is easy to see that the line joining it with the origin consists of all points of the form (at, bt) as t varies over the real numbers. We can conveniently call such a point $Q(t)$.

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 $P_1(a_1, b_1)$ to $P_2(a_2, b_2)$ can then be described as all points $Q(x, y)$ such that $\overrightarrow{P_1Q} = (t)(\overrightarrow{P_1P_2})$, 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: $x = a_1 + t(a_2 - a_1)$, $y = b_1 + t(b_2 - b_1)$.

The shift $\overrightarrow{P_1P_2}$ can be conveniently denoted as $P_2 - P_1$, 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: $Q(x, y) = P_1 + t(P_2 - P_1)$.

Note that this formula can also be rewritten as: $Q(x, y) = Q = (1 - t)P_1 + t(P_2)$.

In this form, the parameter t carries more useful information which we now investigate:

Examples of parametric lines:

1. Find a parametric form of the equations of a line passing thru $P_1(2, -3)$ and $P_2(1, 5)$.

Answer: The two point form gives:

$$x = 2 + t(1 - 2) , \quad y = -3 + t(5 - (-3)) \text{ or } x = 2 - t , \quad y = -3 + 8t.$$

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:

$$\text{Can we solve? } 1 = 2 - t, \quad 5 = -3 + 8t.$$

A clear answer is yes, $t = 1$ does the job. Similarly, $t = -2$ solves:

$$4 = 2 - t, \quad -19 = -3 + 8t.$$

Thus P, Q are points of the above line thru P_1, P_2 .

3. Determine the parametric form of the line joining P, Q . Compare with the original form.

Answer: In compact form, we get:

$$(x, y) = (1, 5) + t(3, -24).$$

An observation.

Note that the pair of coefficients multiplying t were $(-1, 8)$ for the original equation and $(3, -24)$ 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 $t =$, then we must get some point on the line.

In turn, these two condition will guarantee that we have 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 $(5, -40) = (-5)(1, -8)$ 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.

5.1 Meaning of the parameter t :

Consider the parametric description of points on a line joining $P_1(a_1, b_1)$ and $P_2(a_2, b_2)$ given by

$$(x, y) = (1 - t)P_1 + tP_2 = (a_1 + t(a_2 - a_1), b_1 + t(b_2 - b_1)).$$

We have the following information about the parameter t .

Value of t	Interpretation or conclusion
$t = 0$	Point P_1
$0 < t < 1$	Points from P_1 to P_2
$t = 1$	Point P_2
$t > 1$	Points past P_2 away from P_1
$t < 0$	Points past P_1 away from P_2

In other words, the parameter t can be thought as a coordinate system on the line joining P_1, P_2 chosen such that P_1 is the origin and P_2 the unit point.

We can obviously get different coordinate systems, if we choose some other two points on the line.

The above facts can be proved by working out the following precise formula.

For any point $Q = (1 - t)P_1 + t(P_2)$ on the line, let us note that

$$Q - P_1 = (1 - t)P_1 + t(P_2) - P_1 = t(P_2 - P_1)$$

and

$$P_2 - Q = P_2 - (1 - t)P_1 + t(P_2) = (1 - t)(P_2 - P_1).$$

Thus the distances $|\overrightarrow{P_1Q}|$ and $|\overrightarrow{QP_2}|$ are indeed in the proportion $|t| : |(1-t)|$.

This says that to divide the interval P_1P_2 in a ratio $t : (1-t)$ we simply take the point $Q = (1-t)P_1 + tP_2$. For $0 < t < 1$ this has a natural meaning and we note the following corollaries:

Midpoint formula. The midpoint of P_1, P_2 is given by $(P_1 + P_2)/2$. **Division formula.** For $0 < t < 1$ the point $Q = (1-t)P_1 + tP_2$ divides the interval in the ratio $t : (1-t)$. Thus the two trisection points between P_1, P_2 are given as $(2/3)P_1 + (1/3)P_2$ for the point near P_1 and $(1/3)P_1 + (2/3)P_2$ for the point near P_2 .

External divisions formulas. If $t > 1$ then the point $Q = (1-t)P_1 + tP_2$ gives an external division point in the ratio $t : (t-1)$ on the P_2 side. If $t < 0$ then $1-t > 1$ and we get an external division point in the ratio $-t : 1-t$ on the P_1 side.

Thus, for example, the point $Q = -P_1 + 2P_2$ has the property that the distance $|\overrightarrow{P_1Q}| = 2|\overrightarrow{QP_2}|$.

The distance formula If $Q = (1-t)P_1 + tP_2$, then the distance $|\overrightarrow{P_1Q}|$ is simply $|t| \cdot |\overrightarrow{P_1P_2}|$ and consequently, for any other point $R = (1-s)P_1 + sP_2$, we can calculate the distance

$$|\overrightarrow{QR}| = |s-t| \cdot |\overrightarrow{P_1P_2}|.$$

Examples of special points on parametric lines:

1. Let L be the line (discussed above) joining $P_1(2, -3), P_2(1, 5)$.

Let us use the parametric equations found earlier:

$$x = 2 - t, \quad y = -3 + 8t$$

What values of the parameter t give the points P_1, P_2 and their midpoint M ?

Answer: At $t = 0$ we get the point P_1 and at $t = 1$ we get the point P_2 . At $t = \frac{1}{2}$ we get

$$(2 - \frac{1}{2}, -3 + 8\frac{1}{2}) = (\frac{3}{2}, 1).$$

This is the midpoint M .

2. Find the two trisection points of the segment from P_1 to P_2 .

Answer: We simply take $t = \frac{1}{3}$ and $t = \frac{2}{3}$ in the parametrization.

So, the two points are:

$$(2 - \frac{1}{3}, -3 + 8\frac{1}{3}) = (\frac{5}{3}, -\frac{1}{3})$$

and

$$(2 - \frac{2}{3}, -3 + 8\frac{2}{3}) = (\frac{4}{3}, \frac{7}{3}).$$

Warning! Don't forget that for this formula to work, we must take the parametrization for which the two points are given by $t = 0$ and $t = 1$ respectively.

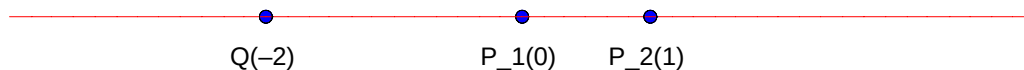
3. Find out where the point $Q(4, -19)$ is situated relative to P_1, P_2 .

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. ¹



The fact that $t = -2 < 0$ says that we have a point on the P_1 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 P_1 -side and the ratio of the distances $d(Q, P_1)$ and $d(Q, P_2)$ is indeed $\frac{2}{3}$.

4. Let U be the point with $t = 3$ and V be the point with $t = -5$. 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:

$$d(P_1, P_2) = \sqrt{(1 - 2)^2 + (5 - (-3))^2} = \sqrt{1 + 64} = \sqrt{65}.$$

Since we are using parameter values 3 and -5 for our two points, we know that their distance is simply:

$$d(U, V) = |3 - (-5)|d(P_1, P_2) = 8\sqrt{65}.$$

Clearly, this is far more efficient! ²

¹Please think of the P_1, P_2 in the picture as P_1, P_2 in our text. It is a software problem with Maple that the correct symbol cannot be introduced in the picture!

²Also note that we did not rush to “evaluate” $\sqrt{65}$! A decimal answer like 8.062257748 spit out by a calculator is never accurate and 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:

$$-1 = 2 - t, \quad 20 = -3 + 8t.$$

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!

5.2 Details of the 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 need the usual equation of a line as a relation between its x, y coordinates.

We start with the parametric equations of a line thru points $P_1(a_1, b_1), P_2(a_2, b_2)$

$$\text{Equation 1: } x = a_1 + t(a_2 - a_1), \quad \text{Equation 2: } y = b_1 + t(b_2 - b_1)$$

and we construct a new equation $(a_2 - a_1)(\text{Equation 2}) - (b_2 - b_1)(\text{Equation 1})$.

This eliminates t 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 thru $P_1(a_1, b_1), P_2(a_2, b_2)$:

$$(a_2 - a_1)y - (b_2 - b_1)x = (a_2b_1 - a_1b_2).$$

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.

Example.

Thus for our points $P_1(2, -3), P_2(1, 5)$ 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 the elimination of t from the equations

$$x = 2 - t, \quad y = -3 + 8t.$$

We now **define the slope of a line** joining any two points $P_1(a_1, b_1), P_2(a_2, b_2)$ as the ratio $m = \frac{b_2 - b_1}{a_2 - a_1}$ 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.
- This may fail if the line is horizontal but not the x -axis. In other words, it fails exactly when $b_2 = b_1$ but $a_1 b_2 \neq a_2 b_1$. In this case the intercept is declared ∞ .
- For the x -axis itself, it is considered **undefined!**
- Thus, in general, the x -intercept is $\frac{a_2 b_1 - a_1 b_2}{b_1 - b_2}$ with a suitable agreement when the denominator is zero.

Similarly:

- The **y-intercept** is similarly defined as b , when the point $(0, b)$ is on the line, with a suitable understanding for a vertical line.
- If $a_2 = a_1$ but $a_1 b_2 \neq a_2 b_1$, i.e. if the line is vertical but not the y axis then the intercept is declared to be ∞ .
- For the y -axis itself, it is considered **undefined!**
- In general, the formula is: $\frac{a_2 b_1 - a_1 b_2}{a_2 - a_1}$ with a suitable agreement when the denominator is zero.

For a non vertical line, it is customary to divide the equation by $a_2 - a_1$ and rewrite it as:

$$y - \frac{b_2 - b_1}{a_2 - a_1}x = \frac{a_2 b_1 - a_1 b_2}{a_2 - a_1}.$$

The last form above is often further rearranged as: **Slope intercept form of a line:**

$$y = mx + c \text{ where } m = \frac{b_2 - b_1}{a_2 - a_1} \text{ is the slope and } c = \frac{a_2 b_1 - a_1 b_2}{a_2 - a_1} \text{ y-intercept.}$$

Thus, for our line thru $P_1(2, -3), P_2(1, 5)$, we have:

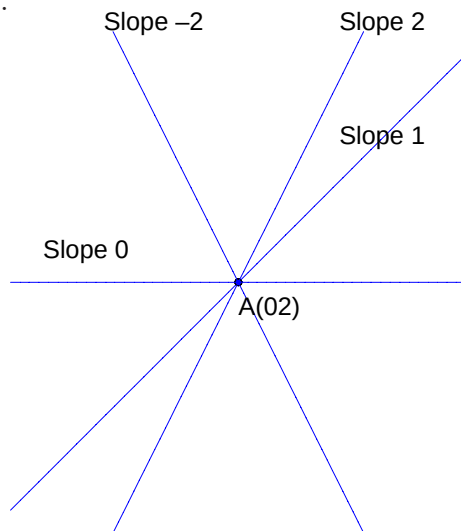
$$-y - 8x = -13 \text{ becomes } y + 8x = 13 \text{ and finally } y = -8x + 13.$$

Thus, the slope is -8 and intercept is 13 . 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 thru 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 an equation of a line.

For this, we note that any equations of the form $ux + vy = w$, or something reducible to it, is the equation of a line provided at least one of u, v is non zero. Hence, if we construct such an equation satisfying the given conditions in some fashion, it is the right answer.³

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 (P(a, b))-slope(m) form.

$$(y - b) = m(x - a).$$

Proof: When rearranged, it becomes $y = mx + (b - ma)$.

³We 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!

So, m is the slope. Also, it clearly is correct if we plug in $x = a, y = b$. 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 $a - b/m$ and the evident symmetry in these two formulas is noteworthy!

Example. Find the equation of a line with slope -6 and passing thru the point $P(3, 4)$.

Answer:

$$(y - 4) = -6(x - 3).$$

This simplifies to:

$$y = -6x + 4 + 18 = -6x + 22.$$

It has slope -6 and intercept 22 .

Intercept form. Suppose the x, y intercepts are p, q respectively and these are non zero. Then:

$$\frac{x}{p} + \frac{y}{q} = 1.$$

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 points $(p, 0), (0, q)$ satisfy the equation.

Example. Find the equation of line with x intercept 3 and y intercept 4 .

Answer: Apply the formula:

$$\frac{x}{3} + \frac{y}{4} = 1.$$

This simplifies to:

$$4x + 3y = 12 \text{ or } y = -\frac{4}{3}x + 4.$$

Parallel lines. It is easy to see that parallel lines have the same slope. 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 thru the point $(3, 10)$.

Answer: We know that the answer can be of the form $y + 8x = k$. Impose the condition of passing thru $(3, 10)$, 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 $1/0 = \infty$.

Thus, to make a perpendicular line to a given line of slope m , we can simply choose any line with slope $-1/m$. But perhaps, we don't wish to calculate the slope first. So, here is a general formula.

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 and compare.⁴

Example. Find a line perpendicular to $y + 8x = 13$ which passes thru the point $(3, 10)$.

Answer: We know that the answer can be of the form $-8y + x = k$. Impose the condition of passing thru $(3, 10)$, i.e.

$$-8(10) + 3 = k$$

so we deduce that $k = -77$ and hence the answer is

$$-8y + x = -77 \text{ or } 8y - x = 77.$$

5.3 Examples of equations of lines.

- Points equidistant from two given points. Suppose you are given two points $P_1(a_1, b_1)$ and $P_2(a_2, b_2)$. 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 P_1, P_2 .

$$(x - a_1)^2 + (y - b_1)^2 = (x - a_2)^2 + (y - b_2)^2.$$

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:

$$x(a_1 - a_2) + y(b_1 - b_2) = (1/2) ((a_1^2 - a_2^2) + (b_1^2 - b_2^2)),$$

where we have cancelled the terms x^2, y^2 and then thrown away a factor of -2 before collecting terms.

⁴An 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 $4x + 6y = 2$. 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.

This is clearly a line and it is easy to check that it passes thru the midpoint $(\frac{a_1 + a_2}{2}, \frac{b_1 + b_2}{2})$.

The reader can also see that its slope is $-\frac{a_1 - a_2}{b_1 - b_2}$. Since the slope of the line joining P_1, P_2 is $\frac{b_2 - b_1}{a_2 - a_1}$, it is obviously perpendicular to the line joining P_1, P_2 .

Thus, we have proved that the set of points equidistant from two given points P_1, P_2 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 $P_1(1, 3), P_2(3, 9)$ is found thus:

The slope of the given line is $\frac{9 - 3}{3 - 1} = 6/2 = 3$. The midpoint is

$$(\frac{1 + 3}{2}, \frac{3 + 9}{2}) = (2, 6).$$

Therefore, perpendicular line has slope $-1/3$ and hence the answer is:

$$(y - 6) = -(1/3)(x - 2) \text{ or } y = -x/3 + 20/3.$$

- 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.

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 it?

Answer: The right angle can occur at A or B or C . We can assume that the point C has coordinates $(4, t)$ for some t .

If we want a right angle at A , then then the slope of the line AC times the slope of the line AB must be -1 . This gives:

$$\frac{t - 5}{4 - 2} \frac{4 - 5}{6 - 2} = -1 \text{ i.e. } t - 5 = (-1)(4)(2)/(-1) = 8.$$

This gives $t = 13$ and C is at $(4, 13)$.

To get a right angle at B our equation becomes:

$$\frac{t - 4}{4 - 6} \frac{4 - 5}{6 - 2} = -1 \text{ i.e. } t - 4 = (-1)(4)(-2)/(-1) = -8.$$

So we get C at $(4, -4)$.

A similar equation for right angle at C leads to:

$$\frac{t-5}{4-2} \frac{t-4}{4-6} = -1 \text{ or } (t-5)(t-4) = (-1)(2)(-2) \text{ i.e. } t^2 - 9t + 20 = 4.$$

This leads to two solutions $\frac{9 \pm \sqrt{17}}{2}$.

Chapter 6

The Circle

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 $P(a, b)$ called the center.¹

The distance formula leads to a simple equation: $\sqrt{(x - a)^2 + (y - b)^2} = r$, but since it is not easy to work with the square root, we square the equation and simplify:

$$(x - a)^2 + (y - b)^2 = r^2 \text{ or } x^2 + y^2 - 2ax - 2by = r^2 - a^2 - b^2.$$

Recognizing a circle. In general, the equation of a circle can be put in the form $x^2 + y^2 + ux + vy = w$ and we can then read off the center and the radius by comparison:

$$u = -2a, \quad v = -2b, \quad w = r^2 - (a^2 + b^2).$$

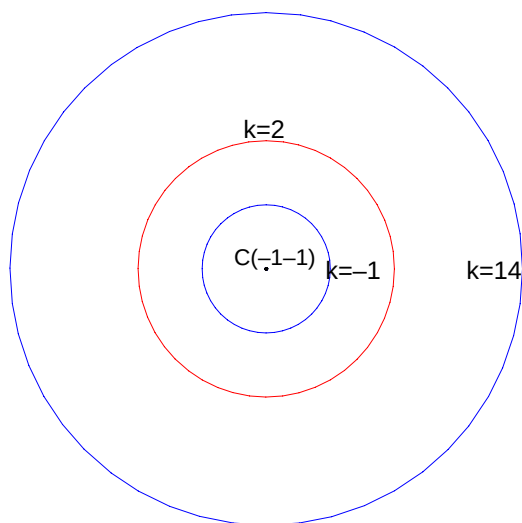
This gives us that the center must be $(a, b) = (-u/2, -v/2)$.

Also, $r^2 = w + (a^2 + b^2)$, so $r^2 = w + u^2/4 + v^2/4$ and hence the radius must be $r = \sqrt{w + u^2/4 + v^2/4}$.

What happens if the quantity $w + u^2/4 + v^2/4$ is negative or zero? We get a circle which has no points (i.e. is imaginary) or reduces to a single point $(-u/2, -v/2)$.

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 $\sqrt{k + 2}$.

¹We 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.



6.1 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 $x^2 + y^2 = r^2$.

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 in this substitution into the equation of the circle.

We get the following sequence of simplifications, to be verified by the reader.

$$\begin{aligned} (t - r)^2 + m^2 t^2 &= r^2 \\ (1 + m^2)t^2 - 2rt &= 0 \\ t((1 + m^2)t - 2r) &= 0 \end{aligned}$$

This gives two easy solutions, $t = 0$ or $t = \frac{2r}{(1 + m^2)}$.

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

$$y = r \frac{2m}{1+m^2}.$$

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 parametrization 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:

$$r \frac{2m}{1+m^2} = r \frac{\frac{2}{m}}{\frac{1}{m^2} + 1} \rightarrow r \frac{0}{1} = 0.$$

6.2 Trigonometric parametrization.

There is another well known parametrization of a circle which is used in the definition of the trigonometric functions $\sin(t)$, $\cos(t)$ that we briefly discussed earlier. Suppose we take a circle of radius 1 centered at the origin O :

$$x^2 + y^2 = 1$$

We wish to arrange a parameter t such that given any point P on the circle, there is a certain parameter associated with it and given any 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 along the circle in the counterclockwise direction and stop. Call the resulting point $P(t)$. Let us see what this means.

The circumference of the whole circle is known to be the number 2π .²

Thus, if we take $t = \pi/2$ then we have gone a quarter circle and the point $(0, 1)$ is thus the point $P(\pi/2)$. 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 $P(3\pi/2) = (-1, 0)$ and finally, if we go a complete circle, we come back to the starting point $P(2\pi) = (1, 0)$.

Thus we see 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 . Indeed, this is very much like our old division, except we have now graduated to real numbers!

The number t associated with the point $P(t)$ is said to be the angle made by the ray $\overrightarrow{OP}$ with the starting ray $\overrightarrow{OU}$.

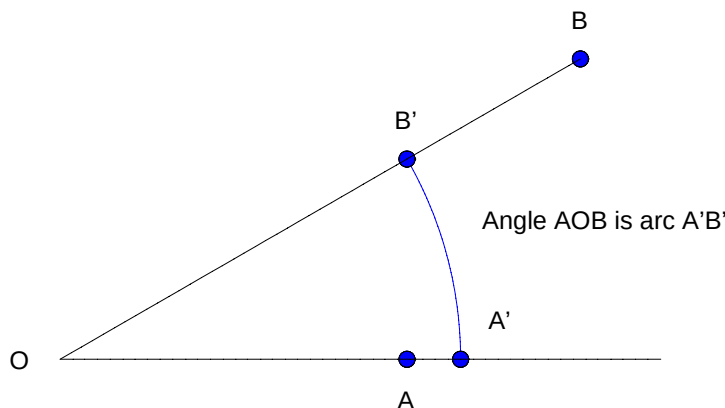
We shall **denote this angle** by the symbol $\angle UOP$.

Formal definition of an angle.

It is easy to define the angle $\angle AOB$ for any two points A, B on the circle of radius 1 centered at the origin by using the distance along the circle measured counterclockwise from A to B .

Finally, given any two line segments $\overrightarrow{OA}$ and $\overrightarrow{OB}$ we define the angle $\angle AOB$ in this way. Draw a unit circle centered at O and let it cut $\overrightarrow{OA}$ at A' . (Extend the ray $\overrightarrow{OA}$ if necessary to intersect the circle.) Similarly, let the circle intersect $\overrightarrow{OB}$ (extended if necessary) at B' .

Define: $\angle AOB = \angle A'OB'$.



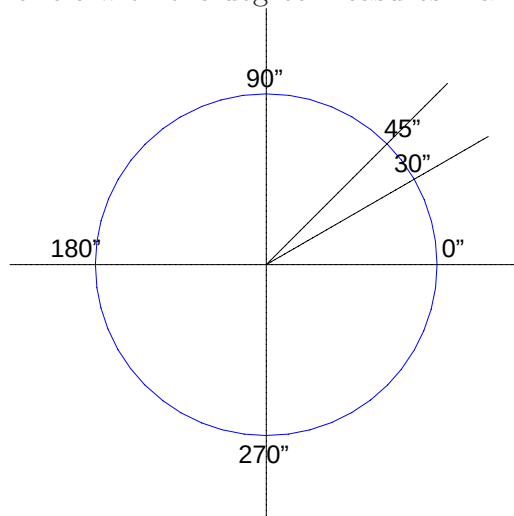
As already explained, the possible angles for the same point are many, so for convenience, we always take the remainder of t after throwing away enough multiples of 2π . This angle is said to be in the units called radians. Thus for the point $B(0, 1)$ the angle of $\overrightarrow{OB}$ is said to be $\pi/2$ radians.

²This π is a famous number with a very interesting history and has occupied many talented mathematicians over the ages. 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.

Even though this description is pleasing, it is difficult to work with, since the number π itself is such a mystery. In mathematical work and especially Calculus, it is important to work with radians. But for our simpler application, it is more convenient to use the more common measure called degrees.

We simply divide the circle into 360 equal parts and call the distance 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 with the degree measures marked at various points.



We now review how $P(t)$ is determined for a given t assumed to be in degrees.

We write:

$$t = 360n + s \text{ where } 0 \leq s < 360.$$

Thus s is the remainder of t after division by 360 and we mark off s degrees in our circle (in the counterclockwise 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 ray 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 , we now know how to find the associated point $P(t)$ and so we know its x and y coordinates. Write

$$P(t) = (x(t), y(t)).$$

Then the definition of the trigonometric functions is:

$$\sin(t) = y(t) \text{ and } \cos(t) = x(t).$$

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,

$$\cos(0) = 1, \sin(0) = 0.$$

2. Find the values of $\cos(90^\circ)$, $\sin(90^\circ)$.

Answer: We already know this point $P(90^\circ)$ to be $(0, 1)$, so we get:

$$\cos(90^\circ) = 0, \sin(90^\circ) = 1.$$

The reader can verify the values for 180° and 270° just as easily. They are:

$$\cos(180^\circ) = -1, \sin(180^\circ) = 0$$

and

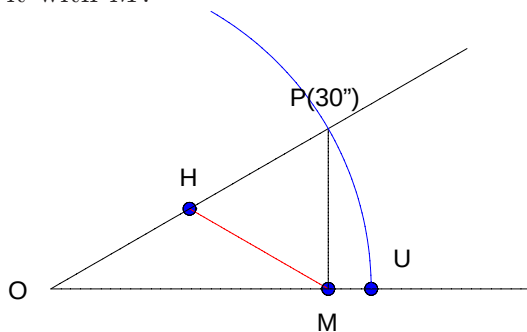
$$\cos(270^\circ) = 0, \sin(270^\circ) = -1.$$

3. Find the values of $\cos(30^\circ)$, $\sin(30^\circ)$.

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 $\overrightarrow{OU}$.

We show a picture below where we have added the midpoint H of OP and the line HM joining it with M .



A little geometric argument can be used to deduce that

$$d(H, O) = d(H, M) = d(H, P)$$

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,

$$\cos(30^\circ)^2 + \left(\frac{1}{2}\right)^2 = 1$$

and it is easy to deduce that

$$\cos(30^\circ)^2 = 1 - \frac{1}{4} = \frac{3}{4}.$$

It follows that

$$\cos(30^\circ) = \pm\sqrt{\frac{3}{4}} = \pm\frac{\sqrt{3}}{2}.$$

Which sign shall we take? Clearly, from the picture, the x -coordinate is positive, so we get

$$\cos(30^\circ) = \frac{\sqrt{3}}{2}.$$

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

$$\sin(-30^\circ) = -\frac{1}{2} \text{ and } \cos(-30^\circ) = \frac{\sqrt{3}}{2}.$$

We will give general formulas like this later.

We shall not pursue this much further here, but want to illustrate that with a more generous idea of parametrization, one can handle difficult calculations, although it may require more ingenuity!

We shall, however, state a fundamental formula (without proof) which can be used for algebraic calculations of sine, cosine functions for most interesting angles.

The reader is encouraged to consult appropriate text on trigonometry for further guidance.

4. Find the value of $\sin(13^\circ)$, $\cos(13^\circ)$.

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:

$$\sin(13^\circ) = 0.2249510544 \text{ and } \cos(13^\circ) = 0.9743700648.$$

6.2.1 Yet another take on parametrizing the circle.

The simple identity

$$(a - b)^2 + 4ab = (a + b)^2$$

has been used for many clever applications.

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 parametrization 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:

$$\left(\frac{x}{r}\right)^2 + \left(\frac{y}{r}\right)^2 = 1$$

.

It follows that

$$x = r \frac{1 - m^2}{1 + m^2}, \quad y = r \frac{2m}{1 + m^2}.$$

is a parametrization!

We now use the above to discuss:

Pythagorean triples.

A Pythagorean triple is a triple of integers x, y, z such that

$$x^2 + y^2 = z^2.$$

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 identity can also be used to generate lots of examples of Pythagorean triples.

We divide the original identity by 4 and plug in $a = s^2, b = t^2$. Then we easily see that

$$x = \frac{s^2 - t^2}{2}, \quad y = st, \quad z = \frac{s^2 + t^2}{2}.$$

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) .

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 lone 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.

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 parametrizing 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

$$t = 0, \quad t = \frac{-2(bqm + f + gm + ap + hpm + hq)}{a + 2hm + bm^2}.$$

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 parametrization can be worked out for a given equation.

Example of parametrization. Parametrize 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 $x = u + 1, y = v + 2$, the 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 thru the origin in (u, v) coordinates is given by $u = t, v = mt$ where m is the slope.

Substitution gives

$$m^2 t^2 + 4mt = 2t \text{ or, by simplifying } t(m^2 t + (4m - 2)).$$

Thus there are two solutions $t = 0, t = \frac{2 - 4m}{m^2}$. We ignore the known solution $t = 0$ corresponding to our starting point A .

We use the second solution for the parametrization:

$$u = t = \frac{2 - 4m}{m^2} \text{ and } v = \frac{2m - 4m^2}{m^2}.$$

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!

6.3 Examples of equations of a circle.

- Intersection of two circles. Find all the points of intersections of the two circles:

$$x^2 + y^2 = 5, \quad x^2 + y^2 - 3x - 4y + 6 = 0.$$

Answer: The general idea of solving two equations is to solve one for one of the variables, say y and then plug in the other to get an equation in the remaining variable x . Then solve the resulting equation for x and then calculate y from the first solution.

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 this: Let $f = x^2 + y^2 - 5$ and $g = x^2 + y^2 - 3x - 4y + 6$. We are looking for common zeros of f, g . It follows that any such common zero will also be a zero of an expression $uf + vg$ where u, v are polynomials in x, y . We try to find suitable expressions which are simpler than f, g themselves.

Thus consider:

$$f - g = (x^2 + y^2 - 5) - (x^2 + y^2 - 3x - 4y + 6) = 3x + 4y - 11.$$

Thus for any of the common points, we have $3x + 4y = 11$ or $y = \frac{11-3x}{4}$.

This is clearly better than the square root expression we would have gotten.

Plugging this into the equation $f = 0$ we get

$$x^2 + \left(\frac{11-3x}{4}\right)^2 - 5 = 0.$$

This simplifies to:

$$\frac{25}{16}x^2 - \frac{33}{8}x + \frac{41}{16} = 0 \text{ or } 25x^2 - 66x + 41.$$

Further, it is easy to verify that it has two roots:

$$x = 1, x = \frac{41}{25}.$$

Plugging back into the expression for y , we get corresponding y values and hence the points are:

$$P(1, 2) \text{ and } Q\left(\frac{41}{25}, \frac{38}{25}\right).$$

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 problem.

- Line joining the intersection points of the two circles. Find the line joining the common points of the two circles:

$$x^2 + y^2 = 5, \quad x^2 + y^2 - 3x - 4y + 6 = 0.$$

Answer: We use the “duck principle”. As before, we set $f = x^2 + y^2 - 5$ and $g = x^2 + y^2 - 3x - 4y + 6$ and note that all the common points satisfy $f - g = 3x + 4y - 11 = 0$.

Thus $3x + 4y - 11 = 0$ 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!

- 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 equations obtained by plugging in the given points:

$$0 + 0 + 0 + 0 = w, \quad 9 + 1 + 3u + v = w, \quad 1 + 9 + u + 3v = w.$$

These simplify to:

$$w = 0, \quad 3u + v = w - 10, \quad \text{and } u + 3v = w - 10.$$

The reader should verify that the solution is

$$u = -5/2, v = -5/2, w = 0 \text{ so the equation is: } x^2 + y^2 - 5x/2 - 5y/2 = 0.$$

From what we already know, we get that this is a circle with

$$\text{center: } (-u/2, -v/2) = (5/4, 5/4)$$

and

$$\text{radius: } \sqrt{w + u^2/4 + v^2/4} = \sqrt{50/16} = (5/4)\sqrt{2}.$$

Here we used the known formulas. What do we do if we have forgotten them? We should learn to derive the formula as needed.

We want to rewrite

$$x^2 + y^2 - 5x/2 - 5y/2 = 0 \text{ as } (x - a)^2 + (y - b)^2 - r^2 = 0.$$

Expanding the desired form, we get:

$$x^2 - 2ax + a^2 + y^2 - 2by + b^2 - r^2 \text{ or by rearrangement } x^2 + y^2 - 2ax - 2by + a^2 + b^2 - r^2.$$

Now it appears in the shape as our equation and a simple comparison of coefficients does the job!

- 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:

$$(0, 0), (3, 1), (6, 2)$$

and notice the contradiction in the resulting equations:

$$0 + 0 + 0 + 0 = w, \quad 9 + 1 + 3u + v = w, \quad 36 + 4 + 6u + 2v = w$$

which lead to:

$$w = 0, \quad 3u + v = -10, \quad 6u + 2v = -40.$$

- 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 will find it convenient to use a parametric form of the line and we set $x = t, y = (6 - 2x)/3 = (6 - 2t)/3 = 2 - (2/3)t$.

Plugging in the circle, we get:

$$(t - 1)^2 + (2 - (2/3)t - 1)^2 = r^2 \text{ which simplifies to: } \frac{13}{9}t^2 - \frac{10}{3}t + 2 = r^2.$$

In a later section, we shall show that a quadratic polynomial $ax^2 + bx + c$ with $a > 0$ has the smallest possible value $-\frac{b^2}{4a} + c$ when $x = -\frac{b}{2a}$, so here we have that the left hand side of our equation has the least value

$$-\frac{\frac{100}{9}}{4 \cdot \frac{13}{9}} + 2 = -\frac{100}{52} + 2 = \frac{-100 + 104}{52} = \frac{1}{13}.$$

Thus the smallest radius is $\sqrt{1/13}$. Note that if we take $r = \sqrt{1/13}$ then $t = \frac{15}{13}$ gives a single point of intersection, while for any bigger value of r we get two points corresponding to the two solution of the quadratic equation in t .

The above calculation can be done more efficiently thus:

A common point of the circle and the line is simply a point on the line whose distance from the center $A(1, 1)$ is equal to the radius r . Thus we simply want the smallest distance between the points on the line given in parametric form as $(t, 2 - (2/3)t)$ and the center $A(1, 1)$. By geometric consideration, we can argue that this minimum must occur at a point B on the line for which $\overrightarrow{AB}$ is perpendicular to the line.³

We now take a line passing thru $A(1, 1)$ perpendicular to the given line $2x + 3y = 6$ and wherever they intersect, we get the point B with minimum distance. The line has slope $-2/3$, so our line $\overrightarrow{AB}$ must have slope $3/2$.

An easy parametric form for the line $\overrightarrow{AB}$ is then $(1 + 2s, 1 + 3s)$ and plugging this into the given equation of the line, we get:

$$2(1 + 2s) + 3(1 + 3s) = 6 \text{ or } 2 + 3 + (4 + 9)s = 6.$$

This leads to $s = 1/13$ and the point $B(1 + 2/13, 1 + 3/13) = (15/13, 16/13)$, the same point as above!

We summarize this as a general result and leave it to the reader to establish:

- 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 B on the line L which is closest to A is given by the formula

$$B\left(p - a\frac{w}{a^2 + b^2}, q - b\frac{w}{a^2 + b^2}\right) \text{ where } w = ap + bq + c.$$

The shortest distance, therefore, is given by

$$\sqrt{\frac{w^2}{a^2 + b^2}} = \frac{|w|}{\sqrt{a^2 + b^2}}.$$

Hint: show that the line through A perpendicular to L can be parametrized as $(p + as, q + bs)$, substitute this in the equation for L and solve for s .

You could also show that the above point B is on L and that the line joining A with B is perpendicular to L by using slopes. Then we must have the correct answer!

³The argument is simply this:

Suppose that there is a C different from B such that $\overrightarrow{AC}$ is perpendicular to the line. Then we have a right angle triangle $\triangle ABC$ with right angle at C or $\overrightarrow{AB}$ as the hypotenuse. Since hypotenuse is the largest side of a right angle triangle, we have that the length of $\overrightarrow{AB}$ is bigger than the length of $\overrightarrow{AC}$ a contradiction!

- 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 $3(4) - 4(4) + 5 = 1$. Which 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?

Suppose we have two points L, M which give values a, b for the expression respectively.

Then any point on the segment $\overrightarrow{LM}$ is of the form $tL + (1 - t)M$ where $0 \leq t \leq 1$ and it is not hard to see that the expression at such a point evaluates to $ta + (1 - t)b$.

Now if a, b are both positive, then clearly $ta + (1 - t)b$ is also positive for all $0 \leq t \leq 1$. If the points were to lie on opposite sides, then one of the values of t will make the evaluation zero, and this would be a contradiction. Thus the points L, M must lie on the same side of the line.

We can similarly argue when both a, b are negative.

When they have opposite signs, then we can see that the quantity $\frac{b}{b - a}$ is between 0 and 1 and this value of t gives a point on the line. Thus, the line joining L, M must intersect the line. 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 line $\overrightarrow{PQ}$ 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.

Chapter 7

Special study of Linear and Quadratic Polynomials.

We now pay special attention to the polynomials which occur most often in applications, namely the linear ($ax + b$ with $a \neq 0$) and the quadratic ($ax^2 + bx + c$ with $a \neq 0$).

7.1 Linear Polynomials.

First, as an example, consider $L(x) = 3x - 5$. This can be rearranged as $L(x) = 3(x - 5/3)$. At $x = 5/3$ we have $L(x) = 0$. If $x > 5/3$ then $L(x) = 3(x - 5/3) > 0$ since both factors are positive. Similarly, if $x < 5/3$, 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 function 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 analysis. 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 them? Just test any convenient points.

Say the point $x = 0$ has value less than 2. 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 $x = 3$ and 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 functions as well. We illustrate this with the quadratic polynomials next.

7.2 Factored Quadratic Polynomial.

If the quadratic expression $Q(x) = ax^2 + bx + c$ could be written in factored form $a(x - u)(x - v)$ where $u \leq v$ then we can easily make a similar analysis depending on the sign of a .

Interval notation.

Given real numbers $a < b$ we shall use the following standard convention.

Values of x	Notation
$a < x$ and $x < b$	(a, b)
$a \leq x$ and $x < b$	$[a, b)$
$a < x$ and $x \leq b$	$(a, b]$
$a \leq$ 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 ∞ .

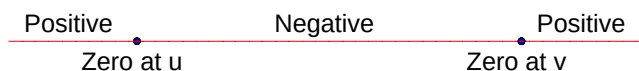
Thus $(-\infty, 4)$ denotes all real numbers x such that $-\infty < x < 4$ and it clearly has the same meaning as $x < 4$, since the first inequality is automatically true!

The reader should verify the following:

1. Case $a > 0$

Values of x	behavior of $Q(x)$
$(-\infty, u)$	$Q(x) > 0$
$x = u$	$Q(x) = 0$
(u, v)	$Q(x) < 0$
$x = v$	$Q(x) = 0$
(v, ∞)	$Q(x) > 0$

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 (u, v) . Since $u < x < v$ in this interval, we note that:

$$x - u > 0 \text{ and } v - x > 0 \text{ which means the same as } x - v < 0.$$

Since $Q(x) = (a)(x - u)(x - v)$ and since $a > 0$, we see that $Q(x)$ is a product of two positive terms $(a, (x - u))$ and one negative term $(x - v)$. Hence it must be negative!

In general, if our function 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!

It follows that for such a quadratic polynomials with positive a , the values have an absolute minimum at some point between $x = u$ and $x = v$ and there is no absolute maximum value. In case $a < 0$, the situation reverses and there is an absolute maximum value instead.

To simplify our discussion, let us

Definition: Absolute Extremum value A value of x to be an absolute extremum if it is either an absolute maximum or an absolute minimum.

Where does this absolute extremum occur? We shall show below that indeed it occurs at $x = (u + v)/2$, and in fact this answer is actually $x = -b/(2a)$. Thus the important point is that we need not find the values u, v to find this extremum value!

7.3 The General Quadratic Polynomial.

Recall that in the example 6 of (1.5), we proved that:

$$Q(x - \frac{b}{2a}) = ax^2 - \frac{b^2 - 4ac}{4a} = a \left(x^2 - \frac{b^2 - 4ac}{4a^2} \right).$$

Replacing x by $x + \frac{b}{2a}$ we see that:

$$\begin{aligned} Q(x) &= a(x + \frac{b}{2a})^2 - \frac{b^2 - 4ac}{4a} \\ &= az^2 - \frac{b^2 - 4ac}{4a} \\ &= a \left(z^2 - \frac{b^2 - 4ac}{4a^2} \right) \end{aligned}$$

where, we have used the symbol z to denote the expression $x + \frac{b}{2a}$. Note that then $x = z - \frac{b}{2a}$.

If we set $\Delta = \frac{b^2 - 4ac}{4a^2}$ for convenience, then this expression becomes much simpler to look at, namely:

$$Q(x) = a(z^2 - \Delta).$$

It is now very easy to describe its behavior.

We have exactly three cases:

1. Case of $\Delta < 0$ or **No real roots**. The quantity $z^2 - \Delta$ 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 $Q(x) = a(z^2 - \Delta)$ is going to be the least exactly when $z = 0$, i.e. $x = z - \frac{b}{2a} = -\frac{b}{2a}$. Thus the minimum value $-a\Delta = \frac{4ac - b^2}{4a}$ is reached at $x = -\frac{b}{2a}$.

In case $a < 0$, a similar analysis says that the maximum value $-a\Delta = \frac{4ac - b^2}{4a}$ is reached at $x = -\frac{b}{2a}$.

Thus we have established that $x = -\frac{b}{2a}$ gives an extremum value regardless of the sign of a . Moreover, the extremum value is $\frac{4ac - b^2}{4a}$.

2. Case of $\Delta = 0$ or **a double root**. In this case $Q(x) = az^2 = a(x + \frac{b}{2a})^2$ is a times a complete square and becomes zero at $x = -\frac{b}{2a}$.

By an analysis similar to the above, we see that $\mathbf{x} = -\frac{\mathbf{b}}{2\mathbf{a}}$ **again gives an extremum value and the value as before is $-\mathbf{a}\Delta = 0$** . Indeed, it is a minimum for $a > 0$ and maximum for $a < 0$.

3. Case of $\Delta > 0$ or **two real roots** In this case $Q(x) = 0 = a(z^2 - \Delta)$ leads to a pair of solutions $z = \pm\sqrt{\Delta}$. This leads to the two roots of the original equation $x = z - \frac{b}{2a} = \pm\sqrt{\Delta} - \frac{b}{2a}$.

To put it in the more familiar form, we write: **The quadratic formula for solutions of $\mathbf{ax}^2 + \mathbf{bx} + \mathbf{c} = 0, \mathbf{a} \neq 0$**

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

We have done some algebraic simplification in the above formula. ¹

Note that our roots u, v are simply

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

in some order and their average $\frac{u+v}{2}$ is clearly

$$(1/2) \left(-\frac{b}{2a} + \sqrt{\Delta} - \frac{b}{2a} - \sqrt{\Delta} \right) = -\frac{b}{2a}.$$

So the extremum value $-b/(2a)$ is indeed the average of the two roots α, β . ²

7.4 Examples of quadratic polynomials.

We discuss a few exercises based on the above analysis.

¹

$$-\frac{b}{2a} \pm \sqrt{\frac{b^2 - 4ac}{4a^2}} = -\frac{b}{2a} \pm \sqrt{b^2 - 4ac}/(2a).$$

This last simplification needs a little thought, since $\sqrt{4a^2}$ is really $2|a|$, which is $2a$ for $a > 0$ and $-2a$ for $a < 0$. However, because of the $\pm$ sign in front, we can just write $2a$.

²Indeed, when one learns about complex numbers to find solutions of $Q(x) = 0$ even for negative Δ , then the roots have the same formula and $-b/(2a)$ shall still be their average. It may no longer have any meaning of maximum or minimum value, but it is still a critical value!

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 $-\frac{b}{2a} = -\frac{4}{4} = -1$. So, if we replace $x = z - \frac{b}{2a} = z - 1$, then we have:

$$\begin{aligned} Q(x) &= 2(z-1)^2 + 4(z-1) - 6 \\ &= 2z^2 - 4z + 2 + 4z - 4 - 6 \\ &= 2z^2 - 8 = 2(z^2 - 4) \end{aligned}$$

If you remember the formula, then you can get the final answer by noticing that

$$\frac{b^2 - 4ac}{4a^2} = \frac{16 - 4(2)(-6)}{4(2^2)} = \frac{64}{16} = 4.$$

At any rate, the value of $Q(x)$ is clearly bigger than or equal to $2(-4) = -8$ and this minimum -8 is attained when $z = 0$ or $x = z - 1 = -1$.

Finally, the quadratic formula (or straight factorization) will yield two roots $u = -3, v = 1$.

If you plot these two values on the number line you get three intervals as in our sample graph above. 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 function 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

$$\Delta = \frac{b^2 - 4ac}{4a^2} = \frac{16 - 4(2)(6)}{4(4)} = -2.$$

This is negative, so the theory tells us that $Q(x)$ will have its extremum at $x = -\frac{b}{2a} = -1$ 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.

There are no points where $Q(x) < 0$.

Chapter 8

Functions

8.1 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 learnt that all points of a line can be alternately described by linear parametrization $x = a + ut, y = b + vt$ 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 found such a parametrization, but with rational functions of the parameter t .

Definition: A rational Curve. We say that a plane algebraic curve $f(x, y) = 0$ is rational if it can be parametrized 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 it is zero for all values of t for which $u(t), v(t)$ are both defined.

After the study of conics, it is tempting to think that all plane algebraic curves could be parametrized in such a way, but alas, this is not the case. The simplest example of a curve which cannot be parametrized by rational functions is the curve given by $y^2 = x^3 - 1$.¹

Since we cannot hope to have all algebraic curves parametrized, we do the next best thing; restrict our attention to the ones which are! In fact, to get a more useful theory, we generalize of idea of an acceptable function.

¹The proof that a given curve is not rational is not elementary and requires new ideas, found in advanced books on Algebraic Geometry.

8.2 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 . Sometimes we can give a name to the procedure or formula which sets up the value of y and write $y = f(x)$.

Let us give some natural examples to illustrate this definition.

1. Polynomial functions. Let $p(x)$ be a polynomial in x and let D be the set of all real numbers $\mathbb{R}$. Then $y = p(x)$ is a function of x on $D = \mathbb{R}$.
2. Rational functions. Let $h(x) = \frac{x-1}{(x-2)(x-3)}$. 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 $\mathbb{R} \setminus \{2, 3\}$, which is read as “reals minus the set of 2, 3”. Let $D = \mathbb{R} \setminus \{2, 3\}$. 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 $n \leq x < n + 1$. It needs some thought to figure out the meaning of this function. The reader should verify that $f(x) = 0$ if $0 \leq x < 1$; $f(x) = 1$ if $1 \leq x < 2$ 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 $\text{floor}(x)$.

There is a similar function called ceiling or ceil for short. It is defined as $\text{ceil}(x) = n$ if $n - 1 < x \leq n$.

In general, a step function is a function whose domain is split into various intervals over which we have definitions which could be different.

A natural practical example of a step function can be something like the following 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 calculation becomes \$25 or higher, then shipping is free!

Describe the shipping charge function $S(x)$ in terms of the purchase price x .

Purchase price	Shipping charge	Comments
$x \leq 0$	0	No sale!
$0 < x \leq 50$	10	Sale up to \$50
$50 < x < 200$	$10 + \text{ceil}(\frac{x-50}{10})$	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 $\text{ceil}((200 - 50)/10) = 15$. Finally, according to the given rule we set $S(x) = 0$ if $x \geq 200$

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.

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.

8.3 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 $a_1 = a_2$ and realize that having a fitting function with $f(a_1) = b_1, f(a_2) = b_2$ is then impossible anyway.

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. For the inquisitive, here are the details, without proof. Suppose that you are given a sequence of distinct data points $(a_1, b_1), \dots, (a_n, b_n)$ 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.

Define and evaluate:

$$\begin{aligned} p &= \sum a^2 = a_1^2 + \dots + a_n^2, \quad q = \sum a = a_1 + \dots + a_n \\ r &= \sum ab = a_1 b_1 + \dots + a_n b_n \\ s &= \sum b = b_1 + \dots + b_n. \end{aligned}$$

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.

If we take a typical pair of values $x = a_i, y = b_i$ for $i = 1, 2, \dots, n$ then we get a “wished for” equation:

$$b_i = ma_i + c$$

and if we add this for all values of i then we get the equation $s = mq + cn$.

If we multiply our starting 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:

$$m = \frac{nr - sq}{np - q^2}, \quad c = \frac{sp - qr}{np - q^2}.$$

It is instructive to try this out for some concrete set of points and observe that 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 $(a_1, b_1), \dots, (a_{10}, b_{10})$, let us calculate the quantities p, q, r , and s as explained in the theory above.

$$\begin{aligned} p &= 1 + 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81 + 100 = 385 \\ q &= 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55 \\ r &= 1 + 6 + 3 + 8 + 20 + 18 + 28 + 56 + 90 + 80 = 310 \\ s &= 1 + 3 + 1 + 2 + 4 + 3 + 4 + 7 + 10 + 8 = 43 \end{aligned}$$

Now we write a proposed linear equation $y = mx + c$ using the facts that $n = 10$, $m = \frac{nr - sq}{np - q^2}$, and $c = \frac{sp - qr}{np - q^2}$.

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. **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 function that passes through the points $(2, 5)$, $(-1, 1)$ and $(3, 7)$.

Answer: Let the quadratic function be ²

$$y = f(x) = px^2 + qx + r.$$

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 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:

$$p = \frac{1}{6}, \quad q = \frac{7}{6} \text{ and } r = 2.$$

So the quadratic 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 it indeed passes thru the three points.

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 are not developing such general formulas here. Look up the topic of Linear Regression elsewhere.

²Don't confuse these p, q, r with the expressions used above! They are new notations.

Chapter 9

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, y as nice functions of a suitable parameter.

It is often possible to describe a curve near a point by nice expressions with a desired degree of accuracy and thus analyze the behavior of a function which may lack a good formula.

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 u, v as our local coordinates and the new equation as

$$(2 + v) = 4(1 + u)^2 - 2 \text{ or } v = (4(1 + 2u + u^2) - 2) - 2 \text{ i.e. } v = 4u^2 + 8u.$$

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 than the quadratic part $-4u^2$. We say that the linear expression

$$v - 8u = y - 2 - 8(x - 1) = y - 8x + 6$$

is a **Linear Approximation** to the parabola near the point $P(1, 2)$.

Alternative organization. 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 the technique, which can be used as an alternative to all the following calculations.

We shall again analyze $y = 4x^2 - 2$ 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:

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

which rearranges as

$$(y - 2) = (-2 + 4 - 2) + (4)(2)(x - 1) + 4(x - 1)^2.$$

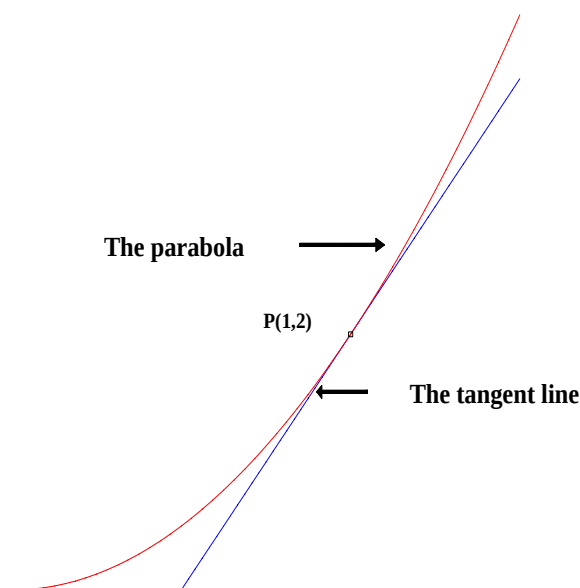
Thus we again get

$$(y - 2) = 8(x - 1) + \text{higher degree terms in } (x - 1).$$

and the linear approximation is again:

$$(y - 2) - 8(x - 1) = y - 8x - 6.$$

The graph below illustrates the closeness between the parabola and its linear approximation near the point $P(1, 2)$.



When we bring it back in the original x, y coordinates, we call the resulting line $y = 8x - 6$ to be the **tangent line** to the original parabola at $P(1, 2)$.

Let us illustrate the same procedure with another point. Consider the point with x -coordinate -1 . Its y -coordinate must be $4(-1)^2 - 2 = 2$. Shift the origin to the point $Q(-1, 2)$ by $x = -1 + u, y = 2 + v$. The equation transforms to:

$$(2 + v) = 4(-1 + u)^2 - 2 \text{ or } v = (4(1 - 2u + u^2) - 2) - 2 \text{ i.e. } v = 4u^2 - 8u.$$

¹

This is rewritten in increasing degree terms as

$$v + 8u - 4u^2 = 0.$$

Thus the tangent line at this point must be

$$v + 8u = 0 \text{ or } (y - 2) + 8(x + 1) = 0 \text{ or } y = -8x - 6.$$

Now let us do such calculations once and for all for a general point (a, b) .

Naturally, we set $x = a + u, y = b + v$ and rearrange the equation:

$$y = 4x^2 - 2 \text{ becomes } b + v = 4(a + u)^2 - 2$$

After a little rearrangement, this becomes:

$$(b - 4a^2 + 2) + (v - 8au) - 4u^2 = 0.$$

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 tangent line is $v - 8au = 0$ or $v = 8au$.

In original coordinates, it becomes ²

$$y - b = 8a(x - a) \text{ and is rearranged to } y = 8ax + (b - 8a^2).$$

¹Note 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.

²The 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!

3. 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 the formula for this slope.

• **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, set $x = 3 + u, y = 4 + v$. The equation becomes:

$$(3 + u)^2 + (4 + v)^2 = 25 \text{ or } u^2 + 6u + 9 + v^2 + 8v + 16 = 25.$$

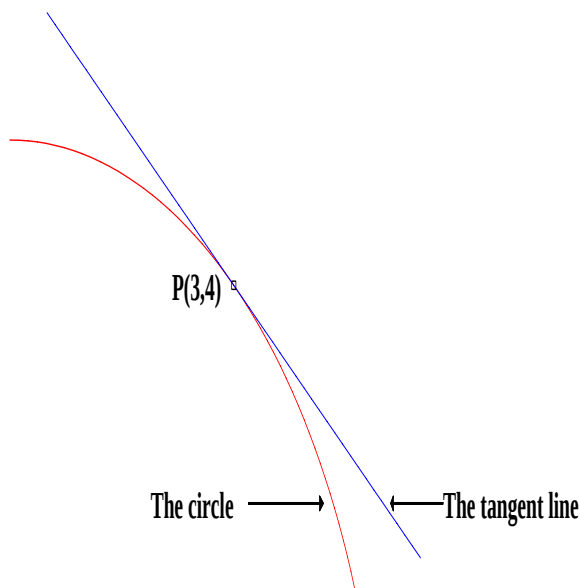
When simplified and arranged by increasing degrees we get: ³

$$(6u + 8v) + (u^2 + v^2) = 0.$$

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

$$6u + 8v = 6(x - 3) + 8(y - 4) = 6x + 8y - 50.$$

³Again, note the vanished constant term as a sign of correctness.

9.1. ANALYZING A GENERAL CURVE $Y = F(X)$ NEAR A POINT $(A, F(A))$.105

Now we try to do this for a general point (a, b) of the circle. As before, set $x = a + u, y = b + v$. 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:

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 $au + bv = 0$ after we drop the common factor 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.

9.1 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 state the procedure followed by a couple of 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 . We note that 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 :

$$(b - f(a)) + (v - lu) + (\text{higher degree terms in } u, v).$$

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:

$$v - lu = 0 \text{ or } y - b = l(x - a) \text{ or, finally } y = lx + (b - la).$$

We may also call it the tangent line at the point $(a, f(a))$ and the expression $v - lu$ or $y - b - l(x - a)$ 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

$$b + l(x - a) = lx + (b - la).$$

Let us note that the quantity

$$f(x) - f(a) = \text{linear approximation} + \text{terms involving higher powers of } (x - a)$$

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 l which can be defined as follows:

Definition: The derivative of a polynomial function $f(x)$ at a point a .

A number l is the derivative of $f(x)$ at $x = a$ if $f(a + u) - f(a) - lu$ is divisible by u^2 .

Let us name the desired number l as something related to f and a and the customary notation is $f'(a)$, a quantity to be derived from f and a , or the so-called derivative of f at a .

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 l 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 l is simply p .

3. $f(x) = px^2$. Then we have:

$$f(a + u) = p(a + u)^2 = pa^2 + 2pa(u) + u^2p = f(a) + 2pau + u^2p \text{ (more terms)}$$

so our l is simply $2pa$.

4. $f(x) = px^3$. Then we have:

$$f(a+u) = p(a+u)^3 = pa^3 + 3pa^2u + 3pau^2 + pu^3 = f(a) + 2pau + u^2(\text{ more terms })$$

so our l 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!

9.2 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 $f(a) + lu + \text{higher degree terms}$.

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 two shortcuts to our notation.

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

$$D_x(px) = p, \quad D_x(px^2) = 2x, \quad D_x(px^3) = 3x^2.$$

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.

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.⁴

We begin by making some very simple observations:

⁴The 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.

Formula 1 The constant multiplier.

If c is a constant and $g(x) = cf(x)$, then $g'(a) = cf'(a)$.

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 $g'(a) = cf'(a)$.

For future understanding, let us rewrite this result as

$$(cf(a))' = c(f'(a)) \text{ or, in compact notation: } D_x(cy) = cD_x(y).$$

Formula 2 The sum of two functions. Suppose that for two functions $f(x), g(x)$ we have the derivatives calculated. Thus 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

$$(f+g)'(a) = f'(a) + g'(a) \text{ or, in compact notation } D_x(y+z) = D_x(y) + D_x(z).$$

Formula 3 The product of two functions. As above, we assume that the derivatives $f'(a), g'(a)$ are known. We now consider the product function $h(x) = f(x)g(x)$. Then we get:

$$h(a+u) = f(a+u)g(a+u) = [f(a) + f'(a)u + u^2(\cdots)] [g(a) + g'(a)u + u^2(\cdots)].$$

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 $n = 1, 2, \dots$, then $f'(a) = na^{n-1}$. In compact notation:

$$D_x(x^n) = nx^{n-1} \text{ for } n = 1, 2, \dots$$

Proof. We already know this for $n = 1, 2, 3, 4$ from the previous section. 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 $k + 1$.⁵

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.** The derivative of a polynomial

$$f(x) = p_0x^n + p_1x^{n-1} + \dots + p_{n-1}x + p_n$$

is given by

$$f'(x) = np_0x^{n-1} + (n-1)p_1x^{n-2} + \dots + p_{n-1}$$

or in compact notation:

$$D_x(p_0x^n + p_1x^{n-1} + \dots + p_{n-1}x + p_n) = np_0x^{n-1} + (n-1)p_1x^{n-2} + \dots + p_{n-1}.$$

Thus if

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

then

$$f'(x) = (4)(3)x^3 - (2)(5)x + 1 = 12x^3 - 10x + 1.$$

⁵This 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 section 18, can be easily proved by this technique and the reader is encouraged to try this.

9.3 Derivatives of more complicated functions.

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, 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

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

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 l is the derivative of $f(x)$ at $x = a$ if $f(a + u) - f(a) - l(u)$ 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 had given 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 is 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. This is how many new functions are born.

9.4 More complicated derivatives.

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. ⁶

Proof:⁷ First suppose $n = \frac{p}{q}$ where p, q are positive integers. Let $y = x^n$. Let $f(x, y) = y^q - x^p$ and note that $f(x, y) = 0$. We know that $D_x(x^p) = px^{p-1}$, so

$$(a + u)^p = a^p + pa^{p-1}u + u^2(\text{ higher degree terms in } u)$$

⁶In fact, the result is true for any constant exponent, but that proof is well beyond our reach at this point!

⁷The proof is easy, but messy in appearance. The student is advised to learn its use before worrying about the details!

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

$$qb^{q-1}(v) - pa^{p-1}(u) = 0 \text{ or } qb^{q-1}(y - b) - pa^{p-1}(x - a) = 0.$$

Note that since $a^p = b^q$,

$$\frac{pa^{p-1}}{qb^{q-1}} = \frac{p a^{-1}}{q b^{-1}} = n \frac{b}{a}.$$

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 $yz = 1$.

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 ⁸

$$D_x(y) = \frac{-ynx^{n-1}}{x^n} = -\frac{x^{-n}nx^n}{x^n} = -\frac{n}{x^n}.$$

Thus, we have

$$D_x(x^{-n}) = -nx^{-n}.$$

⁸An 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.

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. Let us work it out.

Write

$$g(a + u) = g(a) + lu + u^2(w)$$

where we note that $l = g'(a)$ and w contains remaining terms. For convenience, set

$$h = lu + u^2(w),$$

so we get:

$$f(g(a + u)) = f(g(a) + h)$$

and then by definition of the derivative of $f(x)$ we can write

$$f(g(a + u)) = f(g(a)) + mh + h^2(r)$$

where $m = f'(g(a))$ and r contains the remaining terms. Note that h is divisible by u and hence the part h^2r is divisible by u^2 . Thus we can write $h^2r = u^2s$ for some s .

Combining terms, we write:

$$f(g(a + u)) = f(g(a)) + m(lu + u^2w) + u^2(s) = f(g(a)) + m(lu) + u^2(mw + s)$$

and hence we deduce that the derivative of $f(g(x))$ at $x = a$ is

$$ml = f'(g(a))g'(a).$$

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 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, the chain rule is rather subtle to prove as our functions get more complicated and we need to be careful in providing its proof.

9.5 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 with ease.

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.

(a) **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

$$f'(x) = D_x(x^{1/2}) = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}},$$

using the enhanced power formula.

Thus the linear approximation

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

$$\sqrt{4} + \left(\frac{1}{2\sqrt{4}}(x - 4)\right) \text{ or } 2 + \frac{x - 4}{4}.$$

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 .

- (b) **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.⁹

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.

⁹There 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.

(c) **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 $4u = 4(x - 1)$. The quadratic approximation will be

$$12u^2 + 4u = 12(x - 1)^2 + 4(x - 1) = 8 - 20x + 12x^2.$$

The cubic approximation shall be

$$11u^3 + 12u^2 + 4u = 11(x - 1)^3 + 12(x - 1)^2 + 4(x - 1) = -3 + 13x - 21x^2 + 11x^3.$$

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 and all we need to do is to work out some simple examples and observe. We do this next.

9.6 Efficient expansions of polynomials, the Binomial Theorem.

Let us begin by $f(x) = x^n$ for $n = 1, 2, \dots$ and let us try to calculate the expansions of $f(a + u)$. For $n = 1$ we simply get $f(a + u) = a + u$.

For $n = 2$ we get $f(a + u) = (a + u)^2 = a^2 + 2au + u^2$.

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

$$\begin{aligned} (a + u)(a + u)^2 &= (a + u)(a^2 + 2au + u^2) \\ &= a(a^2 + 2au + u^2) \\ &\quad + u(a^2 + 2au + u^2) \\ &= a^3 + 2a^2u + au^2 \\ &\quad + a^2u + 2au^2 + u^3 \\ &= a^3 + 3a^2u + 3au^2 + u^3 \end{aligned}$$

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

$$\begin{aligned}(a + u)^4 &= (a + u)(a^3 + 3a^2u + 3au^2 + u^3) \\ &= a^4 + (3a^3u + a^3u) + (3a^2u^2 + 3a^2u^2) + (au^3 + 3au^3) + u^4 \\ &= a^4 + 4a^3u + 6a^2u^2 + 4au^3 + u^4.\end{aligned}$$

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

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

$$(a + u)^5 = (1)a^5 + (5)a^4u + (10)a^3u^2 + (10)a^2u^3 + (5)au^4 + u^5.$$

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

Clearly, we can continue this pattern as long as needed. This arrangement of coefficients is often referred to as the Pascal 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 commentary on a very old work by Pingala in the second century.

We shall not worry about a formal proof, but the reader can use the idea of Induction and prove the formula by the same technique.

It is more efficient to define a binomial coefficient as follows and use it when convenient: ¹⁰

¹⁰Why 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.

$$nC_r = \frac{n!}{r!(n-r)!} = \frac{n(n-1)\cdots(n-r+1)}{r!}.$$

Here, $m!$ stands for the product $(1)(2)\cdots(m)$ for any positive integer m , with a special understanding that $0! = 1$.

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 don't get into it in this course, but we can still add many terms for a reasonable approximation!

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 $(a+u)^n$ and collect the terms $a^r u^{n-r}$ then its coefficient is nC_r .

Thus we could easily calculate the Linear approximation of $f(x) = x^{10}$ as follows.

We know that we have

$$f(a+u) = (a+u)^{10} = a^{10} + 10C_1 a^9 u + 10C_2 a^8 u^2 + \text{more terms with higher powers of } u.$$

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

Thus the linear approximation is just $a^{10} + 10a^9 u = a^{10} + 10a^9(x-a)$.

But, we knew this from our derivative formula as well. What is new here, is that, in case we would like a quadratic approximation, we simply keep one more term to write:

$$a^{10} + 10a^9 u + \frac{10 \cdot 9}{1 \cdot 2} a^8 u^2 = a^{10} + 10a^9(x-a) + 45a^8(x-a)^2.$$

Chapter 10

Root finding

10.1 Newton's Method

We now find a new use for the idea of linear approximation, namely given a polynomial $f(x)$ we would 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 is simply this:

Given $f(x)$, take its linear approximation $L(x)$ near some convenient point $x = a$ suspected to be close to a root. Find the root of $L(x)$, call it a_1 . Now take $x = a_1$ as a new convenient point and repeat! Under reasonable conditions, you expect to land close to a root.

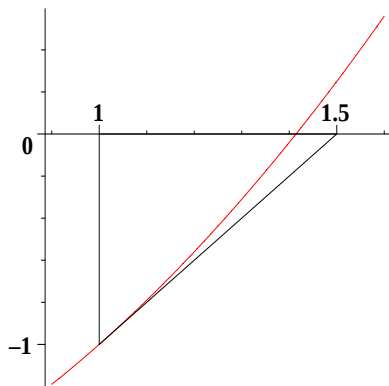
Let us try an example:

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

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

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 $a - \frac{f(a)}{f'(a)}$.
3. Repeat this procedure until $f(a)$ or better yet $\frac{f(a)}{f'(a)}$ becomes smaller than the desired accuracy.
4. When this happens, declare a to be a root of $f(x)$ to within the desired accuracy.

¹The 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.

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

a	$y = f(a)$	$y' = f'(a)$	$\frac{y}{y'}$	$a - \frac{y}{y'}$
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

10.2 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 roots!** If we start with a function like $f(x) = x^2 + 2$, then we know that it has no roots. What would happen to our method?

The method will keep you wandering about the number line, leading to no value!

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

$$1 - \frac{f(1)}{f'(1)} = 1 - \frac{-4}{-2} = -1.$$

²In this example, the formula which changes a to $a - \frac{a^2-2}{2a}$ simplifies to $\frac{1}{2}(a + \frac{2}{a})$. 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.

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:

$$2 \rightarrow 2.2857 \rightarrow 2.2376 \rightarrow 2.2361.$$

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, problems caused by wild jumps because the derivative becomes rather small and so $\frac{y}{y'}$ gets very large!

Chapter 11

Summation of series.

11.1 Application of polynomials

Consider any polynomial expression $f(x)$ and define a new polynomial

$$g(x) = f(x) - f(x - 1).$$

Consider the following sequence of substitutions.

$$\begin{array}{rclcl} x & g(x) & = & f(x) - f(x - 1) \\ 1 & g(1) & = & f(1) - f(0) \\ 2 & g(2) & = & f(2) - f(1) \\ \dots & \dots & \dots & \dots \\ i & g(i) & = & f(i) - f(i - 1) \\ \dots & \dots & \dots & \dots \\ n & g(n) & = & f(n) - f(n - 1) \end{array}$$

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

$$g(1) + g(2) + \dots + g(i) + \dots + g(n) = f(n) - f(0).$$

This gives us very efficient and clever ways for adding up a series of terms. For example, let $f(x) = x$. Then $g(x) = 1$ and we get the trivial result

$$g(1) + \dots + g(n) = 1 + \dots + 1 = f(n) - f(0) = n - 0 = n.$$

If we take $f(x) = x^2 + x$, then $g(x) = 2x$ and we get the identity

$$g(1) + \dots + g(n) = 2 + 4 + \dots + 2n = f(n) - f(0) = (n^2 + n) - (0) = n(n + 1).$$

Division of both sides by 2 gives the well known

$$\text{Arithmetic series formula: } 1 + 2 + \dots + n = \frac{n(n + 1)}{2}.$$

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	$5x^4 - 10x^3 + 10x^2 - 5x + 1$

Indeed, these expressions should be deducible by using our old formulas for $(x+t)^n$ by taking $t = -1$.

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

$$x^2 = (1/3)(3x^2 - 3x + 1) + (1/2)(2x - 1) + (1/6)(1)$$

and this tells us that $f(x) = x^3/3 + x^2/2 + x/6$ will do the trick!

Thus the desired answer is:

$$1^2 + 2^2 + \cdots + n^2 = f(n) - f(0) = n^3/3 + n^2/2 + n/6 = \frac{(n)(n+1)(2n+1)}{6}.$$

Clearly, with little work, you can make your own formulas. ¹

11.2 Examples of summation of series.

- Example 1: 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

$$f(x) - f(x-1) = (3x^2 + x)/2 - ((3x^2 - 5x)/2 - 1) = 3x - 1.$$

¹It is interesting to take up the challenge of redoing the geometric series formula with this new idea. Our $g(x)$ is now $r^{(x-1)}$ and the needed $f(x)$ is to be $r^x/(r-1)$. Verify this.

as desired.²

- Example 2: A general arithmetic series (A.P.)³ with beginning term a and common difference d is a sum of n terms $a, a + d, \dots, a + (n - 1)d$. Find the formula for its sum.

Answer: We can, of course work this out just as above. But to illustrate the power of independent thinking, we shall give two other methods.

First, let us state the result to be proved.

The first method is to guess that the answer is a quadratic expression:

$$f(x) = px^2 + qx + r.$$

We just try to find suitable p, q, r so the answer is correct for first few values of n . Here is the work.

If we take $n = 0$, i.e. no terms at all, then the sum must be 0, so

$$f(0) = p(0)^2 + q(0) + r.$$

This says that $r = 0$, so $f(x) = px^2 + qx$.

For $n = 1$ the sum is a and hence we want

$$\text{Equation 1 :} \quad f(1) = p + q = a.$$

For $n = 2$ the sum is $a + (a + d) = 2a + d$, so we want

$$\text{Equation 2 :} \quad f(2) = p(2)^2 + q(2) + r = 4p + 2q = 2a + d$$

To solve equations 1 and 2, we subtract twice the equation 1 from equation 2 and get:

$$(4p + 2q) - 2(p + q) = (2a + d) - 2(a) \text{ i.e. } 2p = d \text{ or } p = d/2.$$

Using equation 1, we deduce that $q = a - p = a - d/2$.

²Here 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,

$$f(x) - f(x - 1) = \frac{x(3x + 1) - (x - 1)(3x - 2)}{2} = \frac{(3x^2 + x) - (3x^2 - 5x + 2)}{2} = \frac{6x - 2}{2} = 3x - 1.$$

³A.P. is a short form of Arithmetic Progression, which stands for the sequence of terms with common differences. We are using a more technical term “series” which indicates the idea of the sum of such a progression of numbers.

Thus, we have

$$f(x) = dx^2/2 + (a - d/2)x = x(a + \frac{x(d-1)}{2}).$$

Hence our final answer is going to be $f(n) - f(0) = f(n)$. This gives us:

A.P. formula An A.P. with first term a and common difference d has n -th term as $a + (n-1)d$. The sum of the first n terms is given by the expression “number of terms multiplied by the average of the first and the last terms”, i.e.

$$f(n) = n \frac{(a) + (a + (n-1)d)}{2}.$$

Our second method is much cuter and easy to see, 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	$a + 2d$
$\dots$		
t_n	the n -th term	$a + (n-1)d$

Let $S(n)$ denote the sum $t_1 + t_2 + \dots + t_n$.

Consider the following table:⁴

$S(n)$	$=$	$t_1 +$	$t_2 +$	$\dots$	$+t_n$	Forward
$S(n)$	$=$	$t_n +$	$t_{n-1} +$	$\dots$	$+t_1$	Reverse
$2S(n)$	$=$	$(t_1 + t_n) +$	$(t_2 + t_{n-1}) +$	$\dots$	$(t_n + t_1)$	Added
$2S(n)$	$=$	$(2a + (n-1)d) +$	$(2a + (n-1)d) +$	$\dots$	$(2a + (n-1)d)$	Simpler!

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 $5, 8, \dots$. 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.

⁴The 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!

Chapter 12

More on the Trigonometric Functions.

We now study more details of the trigonometric functions introduced earlier. The aim of this section is to understand the basic definition of the various trigonometric functions and learn some fundamental relations governing them.

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, but we have chosen to postpone it to a higher level course. Part of the reason is that their theoretical treatment requires a lot more sophistication.

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.

12.1 Basic definitions of Trigonometric Functions.

We shall redefine the necessary terms and make some new definitions. It would be useful to consult the treatment of trigonometric functions in the circle parameterizations.

Definition: Angle Given two rays $\overrightarrow{OA}$ and $\overrightarrow{OB}$ starting from O we define the angle from $\overrightarrow{OA}$ to $\overrightarrow{OB}$ thus:

Construct a unit circle (i.e. a circle with radius 1) centered at O . Find points A' and B' on the rays $\overrightarrow{OA}$ and $\overrightarrow{OB}$ respectively, which are also on the unit circle. *You may have to extend the rays away from O so that they hit the circle.*

The length of the arc from A' to B' measured *counterclockwise* is the measure of the angle $\angle AOB$. This is the **radian measure** and we assume the angle to be in radians unless otherwise stated. In radian measures, the whole circle has length 2π . It is customary to use the degree measure where we let the circle have a measure of 360° .

Thus to convert radians to degrees we multiply by $\frac{360}{2\pi} = \frac{180}{\pi}$. Naturally, to change degrees to radians we multiply by $\frac{\pi}{180}$.

The angle is naturally measured as a number in the interval $[0, 2\pi)$. For any real number t , we associate an angle s which is the remainder when the angle t is divided by 2π .

This means we write our

$$t = 2\pi(n) + s \text{ where } 0 \leq s < 2\pi$$

and n is some integer.

Thus an angle of $t = 9\pi/2$ is written as

$$9\pi/2 = 4\pi + \pi/2 = 2\pi(2) + \pi/2.$$

Thus, even though the angle is $9\pi/2$ for all calculations of trigonometric functions with it, it may be safely replaced by $\pi/2$.

We can do the same thing with degrees, except we divide by 360° .

Thus an angle of $t = -30^\circ$ is reduced by:

$$-30^\circ = 360^\circ(-1) + 330^\circ$$

so it is declared to be equivalent to 330° .

Negative angles.

Sometimes it reduces our work if we use negative values of angles, rather than converting them to positive angles. After all, -30° is easier to handle than 330° . We can give a natural meaning to negative angles, they represent arc length measure the *negative or clockwise* direction! Check out that going 30° clockwise or 330° counterclockwise leads to the same position!

Thus, sometimes it is considered more convenient to take the angle measure between -180° to 180° .

Definition: Trigonometric functions. Given an angle $t \in \Re$ we define $P(t)$ as the corresponding point on the unit circle obtained by laying off the angle t in the counterclockwise direction as explained.

We may call $P(t)$ the **locator point** of the angle t . Remember that $P(t)$ is on the unit circle centered at O and let its coordinates be (x, y) .

We **define**:

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

It is customary to define four more functions which are simply related to the above two functions:

$$\tan(t) = \frac{\sin(t)}{\cos(t)}, \quad \cot(t) = \frac{\cos(t)}{\sin(t)}.$$

Also:

$$\csc(t) = \frac{1}{\sin(t)}, \quad \sec(t) = \frac{1}{\cos(t)}.$$

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. $f(f(x))$. It is wise to verify what this notation means before making guesses! For trigonometric functions, this is the accepted convention!

If you have any confusion in your mind, it is safer to use the explicit $(\sin(t))^2$. The convention works for other powers too. This indeed leads to a confusion later on if you try to use negative powers. We will explain it when it comes up.

12.2 Connection with the usual Trigonometric Functions.

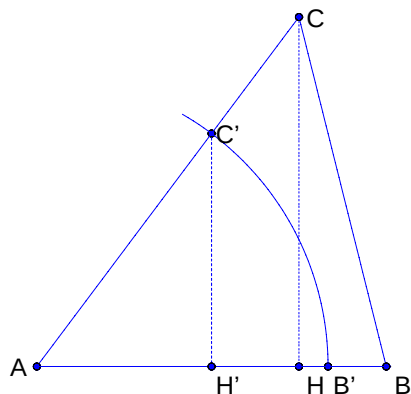
While the above definitions are precise and general, it is important to connect them with the classical definition.

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 $\angle BAC$.¹

Consider the following picture, where $\overrightarrow{CH}$ is the perpendicular from C onto $\overrightarrow{AB}$. 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! 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 conventions for calculations! Our new definitions will take care of it naturally!

12.3 Addition Formula.

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?

¹If we were to rename the triangle $\triangle ACB$ then $\angle A$ will become $\angle CAB$ and with our conventions, it will evaluate to the negative of the other angle. Such delicate analysis did not arise in the old setup since we always arranged a suitable angle *less than* 180° .

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.

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

$$d(P(s), P(t)) = d(P(0), P(t - s)).$$

This is the only main point. 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 all points, 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 first 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**:²

Fundamental Identity 4. $\cos(t - s) = \cos(t)\cos(s) + \sin(t)\sin(s)$

²It appears like a subtraction formula, but it will easily become converted to the addition formula soon!

12.4 Identities Galore.

The fourth fundamental identity together with known values leads to a whole set of new identities. Deriving them is a fruitful fun exercise.

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.

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: ³

$$\cos(\pi/2 - (\pi/2 - s)) = \sin(\pi/2 - s).$$

This simplifies to

$$\cos(s) = \sin(\pi/2 - s).$$

Thus we have established

Supplementary Angle Identity.

$$\cos(\pi/2 - s) = \sin(s) \text{ , } \sin(\pi/2 - s) = \cos(s)$$

The angles s and $\pi/2 - s$ are said to be supplementary angles, hence the name. Indeed, in a right angle triangle, the two angles different from the right angle are supplementary and two positive supplementary angles arise exactly in this way.

³If 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 .

3. We proceed as above, except we now note that $P(\pi) = (-1, 0)$ and take $t = \pi$ in the fourth fundamental identity.

We get:

$$\cos(\pi - s) = (-1) \cos(s) + (0) \sin(s) = -\cos(s).$$

4. Now we replace t by $\pi/2 - t$ in the fourth fundamental identity to get:

$$\begin{aligned} \cos((\pi/2 - t) - s) &= \cos(\pi/2 - t) \cos(s) + \sin(\pi/2 - s) \sin(s) \\ \cos(\pi/2 - (t + s)) &= \sin(t) \cos(s) + \cos(s) \sin(s) \\ \sin(t + s) &= \sin(t) \cos(s) + \cos(t) \sin(s) \end{aligned}$$

The second and third line simplifications are done using the supplementary 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 ⁴

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.

⁴An 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.

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:

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. **Half Angle formulas.** These are easy but powerful consequences of the above variants of identity 6.

Fundamental Identity 8. $\cos\left(\frac{t}{2}\right) = \pm \sqrt{\frac{1 + \cos(t)}{2}}$

Fundamental Identity 9. $\sin\left(\frac{t}{2}\right) = \pm \sqrt{\frac{1 - \cos(t)}{2}}$

The first one is obtained by replacing t by $\frac{t}{2}$ in variant 1 of identity 6. The $\pm$ sign comes due to the square root and has to be determined by the location of the angle $\frac{t}{2}$.

The second one is similarly obtained from variant 2 of the identity 6.

3. **Using 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 $\frac{\pi}{6} = 30^\circ$ as well as $\frac{\pi}{3} = 60^\circ$.

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 supplementary angle identity, we get

$$\sin(t) = \cos(s) \text{ and } \cos(t) = \sin(s).$$

From the 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

$$\cos(s) = \pm\sqrt{1 - \frac{1}{2}^2} = \pm\sqrt{\frac{3}{4}}.$$

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!

Here is another easy but important use of the formulas.

Compute the sine and cosine for $\frac{\pi}{4} = 45^\circ$.

Since $2\frac{\pi}{4} = \frac{\pi}{2}$, we see two conclusions:

$$\sin\left(\frac{\pi}{4}\right) = \cos\left(\frac{\pi}{4}\right) \text{ and } \sin\left(\frac{\pi}{2}\right) = 1 = 2\sin^2\left(\frac{\pi}{4}\right).$$

The reader is invited to provide the reasons. At any rate, we deduce that

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

This formula is also worth memorizing.

4. **Functions for related angles.** Given one angle t , there are some angles naturally related to it. These are:

- **Supplementary angle** $\pi/2 - t$.
- **Complementary angle** $\pi - t$.
- **Reflection at the origin** $\pi + t$ which is equivalent to $t - \pi$.
- **reflection in the x axis** $-t$.

We have enough tools to make our own formulas for the sine and cosine functions of the related angles, given those of the original angle. We leave them as exercise.

5. **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:

$$\begin{aligned}
 \frac{\tan(s) + \tan(t)}{1 - \tan(s)\tan(t)} &= \frac{\frac{\sin(s)}{\cos(s)} + \frac{\sin(t)}{\cos(t)}}{1 - \frac{\sin(s)\sin(t)}{\cos(s)\cos(t)}} \\
 &= \frac{\frac{\sin(s)\cos(t) + \sin(t)\cos(s)}{\cos(s)\cos(t)}}{\frac{\cos(s)\cos(t) - \sin(s)\sin(t)}{\cos(s)\cos(t)}} \\
 &= \frac{\sin(s+t)}{\cos(s+t)} \\
 &= \tan(s+t)
 \end{aligned}$$

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.

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.

12.5 Using trigonometry.

We shall show some of the applications of trigonometry. Like the identities, there is no end to the topics that can be added here. WE only illustrate some chosen topics.

1. **Calculating exact values.** So far, we have succeeded in calculating the trigonometric functions for a few special angles $0^\circ, 30^\circ, 45^\circ, 60^\circ, 90^\circ$ and their related angles.

The half angle formula lets us evaluate many more angles.

Thus

$$\sin(15^\circ) = \sqrt{\frac{1 - \cos(30^\circ)}{2}}$$

where, we have chosen the plus sign since we know the angle is in the first quadrant. The right hand side is

$$\sqrt{\frac{2 - \sqrt{3}}{4}} = \frac{\sqrt{6} - \sqrt{2}}{4}.$$

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!

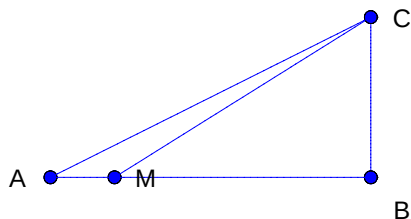
It is clear that from this 15° angle, you can keep on halving the angle and get down to a small angle. For several hundred years, people did their astronomy using multiples of 7.5° angle.

You may ask why one should bother with this. Your calculator will happily report the value 0.2588190451 for $\sin(15^\circ)$ as well as our theoretical value of $\frac{\sqrt{6} - \sqrt{2}}{4}$. 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 thru several steps of calculations.

2. 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 line $\overrightarrow{AC}$ is measured to make an angle of 26.56° with the horizontal $\overrightarrow{AB}$.

If we observe the top of the building after walking 200 feet closer to the building (from point M) 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 $s = 26.55^\circ$ and $t = 32^\circ$. 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:

$$E1 \qquad \qquad \qquad \tan(s) = \frac{h}{k}$$

and

$$E2 \qquad \qquad \qquad \tan(t) = \frac{h}{k - 200}$$

We wish to solve for h, k from these two equations.

Rearrangement gives:

$$E3. \qquad \qquad h - \tan(s)k = 0, \quad h - \tan(t)k = -200 \tan(t)$$

Subtracting second equation in E3 from the first, we get

$$E4. \qquad \qquad (\tan(t) - \tan(s))k = 200 \tan(t)$$

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. ⁵

⁵Some 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!

Index

- Āryabhata algorithm, 32
- absolute extremum, 91
- angle, 74
- arithmetic series, 126
- axes, 49
- axis, 49
- binomial, 5
- binomial coefficients, 118
- Binomial Theorem, 116
- center, 71
- chain rule, 114
- Chinese remainder theorem, 35
- circle, 71
- coefficient, 3, 27
- complex numbers, 1
- conic, 80
- consistent, 27
- coordinate system, 49
- cos, 76
- cot, 131
- Cramer's Rule, 21
- Cramer's rule, 35
- csc, 131
- degree, 3, 27, 75
- degree of a polynomial, 6
- denominator, 65
- derivative, 106
- determinant, 35
- determinant of 2 by 2 matrix, 21
- direction numbers, 51
- distance, 50
- distance formula, 62
- dividend, 30
- divisibility of integers, 39
- divisibility of polynomials, 39
- division, 28
- division algorithm, 27, 33
- division formula, 62
- divisor, 30
- domain, 96
- Efficient Euclidean algorithm, 32
- ellipse, 80
- enhanced remainder theorem, 45
- Euclidean algorithm, 32
- exponential, 129
- fast calculation methods, 10
- field, 1
- Fourier analysis, 129
- function, 96
- GCD, 31
- geometric series, 43, 126
- greatest common divisor, 31
- hyperbola, 80
- identity, 17
- inconsistent, 17, 27
- indeterminant, 3
- intercept form, 67
- isolating the variable, 17
- LCM, 31
- leading coefficient of a polynomial, 6
- least common multiple, 31
- linear approximation, 102
- linear equation, 15

- linear fit, 98
- linear polynomial, 43
- logarithmic, 129
- midpoint formula, 62
- monic, 31
- monomial, 3
- negative angles, 130
- Newton's Method, 119
- numerator, 65
- origin, 52
- parabola, 80
- parallel lines, 67
- parameter, 2, 60
- parametric form, 59
- parametric form for the circle, 73
- Pascal triangle, 117
- perpendicular lines, 67
- pi, 64
- plane algebraic curve, 95
- polynomial, 5, 27
- polynomial functions, 96
- power rule, 114
- Pythagorean Theorem, 51
- Pythagorean triple, 79
- quadrant, 51
- quadratic, 82
- quadratic fit, 99
- quadratic formula, 93
- quadratic polynomials, 89
- radian, 75
- radius, 71
- range, 96
- rational curve, 95
- rational functions, 96
- rational numbers, 1
- ray, 74
- real number, 74
- real numbers, 1
- remainder, 28
- Remainder theorem, 43
- root, 93
- row echelon form, 20
- sec, 131
- shift, 50
- sin, 76
- slope, 65, 81
- slope intercept, 65
- solution to an equation, 16
- step function, 96
- supplementary angles, 134, 135
- surds, 140
- tan, 131
- tangent line, 103
- the principle of $0, 1, \infty$ number of solutions, 17
- transcendental, 74
- transformation, 53
- trigonometric functions, 130
- trinomial, 5
- unique solution, 17
- unit circle, 74
- Variables, vii
- vectors, 50
- zero polynomial, 5