

Abstract

In this pieces of notes, we will go through the concepts related to functions and the meaning of graphs of the functions.

Contents

1	Function and its graph	7
1.1	What is a Function?	7
1.1.1	Function as an input-output pair	8
1.1.2	Defining a function with variables	10
1.1.3	Domain, Co-domain and Range	11
1.2	Graph of a function	15
1.2.1	Blobs-and-arrows diagram for discrete functions	15
1.2.2	The pairing table and xy-coordination	17
1.3	Fundamental properties of a graph	20
1.3.1	points lying on the graph	20
1.3.2	Special intersections: axis intercepts	22
1.4	Solving equations using graphs of functions	23
1.4.1	Homogeneous equations	23
1.4.2	Non-homogeneous equations	23
1.4.3	Simultaneous equations	23
1.5	Transformation of functions	23
1.5.1	Translation	23
1.5.2	Dilation	23
1.6	Challenging questions	23
2	Trigonometric functions	24
2.1	Why are we calling it ‘trigonometry’ ?	24
2.2	The relation of sides of triangle	24
2.2.1	Triangle inequality and Pythagoras theorem	24
2.2.2	The Sine function	25
2.2.3	The Cosine function	26
2.2.4	The tangent function	26
2.2.5	Fundamental identities for trigonometric functions	27
2.3	The compound angle formulae	28

2.4	The trigonometric laws	30
2.4.1	The area formula of a triangle	30
2.4.2	The Sine Law	31
2.4.3	The Cosine Law	33
2.4.4	The Tangent Law	34
2.4.5	Heron's formula	36
2.5	Extending trigonometric functions to describing circles	39
2.5.1	The relation between rectangular coordinates and polar coordinates	40
2.5.2	Defining trigonometry as coordinate function	40
2.5.3	Reduction formula for trigonometric functions	40
2.5.4	The C-A-S-T graph and reduction strategy	40
2.6	Challenging Questions	40
3	Linear functions	41
3.1	Fundamental concepts of points	41
3.1.1	Distance between two points	41
3.1.2	Mid-point and Division points	43
3.1.3	Slope of two points	45
3.2	Different forms of a linear function	48
3.2.1	Segment	49
3.2.2	Line as extension of a segment	49
3.2.3	Point-slope form	50
3.2.4	Two-point form	51
3.2.5	Slope-intercept form	51
3.2.6	Intercepts form	52
3.2.7	General form	53
3.3	Rotational geometry	54
3.3.1	Angle of elevation from horizon	55
3.3.2	Angle of depression from horizon	55
3.3.3	Angle between two arbitrary lines	55
3.3.4	Rotation of points and lines	55
3.4	Relation between lines	55
3.4.1	Point of intersection	55
3.4.2	Parallel lines	55
3.4.3	Perpendicular lines	55

3.4.4	Number of intersections	55
3.4.5	Angle bisector	55
3.5	Additional content: Point-line distance	55
3.6	Linear inequalities	56
3.6.1	One variable inequality	56
3.6.2	Two variable inequality	56
3.6.3	Linear programming	56
3.7	Challenging questions	56
4	Quadratic functions	57
4.1	Solving Quadratic Equations	57
4.1.1	The square-root method	57
4.1.2	General form and Completing the square method	57
4.1.3	The quadratic formula	57
4.2	Solvability of Quadratic Equations	57
4.2.1	Discriminant: the factor affecting solvability	57
4.2.2	Application of solvability	57
4.3	Relation between coefficients and roots	57
4.3.1	What does it mean by a root?	57
4.3.2	The Vieta formula	57
4.3.3	Forming quadratic equations with given roots	57
4.4	Vertex form of a quadratic function	57
4.4.1	Obtaining vertex form using Transformation	57
4.4.2	Relation between vertex form and general form	57
4.4.3	Opening direction	57
4.4.4	The axis of symmetry	57
4.4.5	The extremum of quadratic functions	57
4.5	Quadratic inequalities	57
4.5.1	Boolean algebra	57
4.5.2	Solving quadratic inequalities	57
4.6	Challenging questions	57
5	Polynomial functions	58
5.1	Arithmetic rules for polynomials	58
5.1.1	Addition, Subtraction and Multiplication	58

5.1.2	Division	58
5.2	Divisibility of polynomials	58
5.2.1	Division algorithm	58
5.2.2	Remainder Theorem	58
5.2.3	Factor Theorem and General Vieta formula	58
5.3	G.C.D. and L.C.M.	58
5.3.1	Greatest Common Divisor	58
5.3.2	Least Common Multiple	58
5.4	Rational Function	58
5.4.1	Simplification of Rational Functions	58
5.4.2	Extended: Partial Fractions	58
5.5	Positional notation	58
5.5.1	Binary, Octal, Decimal and Hexadecimal	58
5.5.2	Conversion between different bases	58
5.6	Challenging questions	58
6	Exponential and Logarithmic functions	59
6.1	Index notation	59
6.1.1	From multiplication to power	59
6.1.2	negative integral power	59
6.1.3	Law of integral index	59
6.2	Rational power	59
6.2.1	The n-th root notation	59
6.2.2	Fractional power	59
6.2.3	Law of fractional index	59
6.3	Exponential functions	59
6.3.1	Law of Exponential algebra	59
6.3.2	Seeing surds as fractional exponents	59
6.3.3	Solving equations using exponentiation method	59
6.4	Logarithmic functions	59
6.4.1	What is Logarithm?	59
6.4.2	Law of Logarithmic algebra	59
6.4.3	Solving equations using Logarithmic method	59
6.5	Challenging questions	59

7	Sequence as a function of natural numbers	60
7.1	What is a sequence?	60
7.2	Arithmetic sequence	60
7.2.1	General form	60
7.2.2	Summation of arithmetic sequence	60
7.3	Geometric sequence	60
7.3.1	General form	60
7.3.2	Summation on Geometric sequence	60
7.3.3	Infinite sum of Geometric sequence	60
7.4	Additional content: Arithmetic-Geometric sequence	60
7.5	Challenging questions	60
8	Probability functions	61
8.1	Counting Principle	61
8.1.1	Addition rule	62
8.1.2	Multiplication rule	63
8.2	Pigeonhole Principle	63
8.2.1	Fundamental version	64
8.2.2	General version	64
8.3	Counting functions	64
8.3.1	Factorial	65
8.3.2	Permutation	66
8.3.3	Combination	67
8.4	Fundamental Probability	67
8.4.1	Fair dice intuition	67
8.4.2	Outcomes/Events as sets	67
8.4.3	Probability in terms of outcomes	67
8.5	Probability of independent events	67
8.5.1	Independent events	67
8.5.2	Venn-diagram	67
8.5.3	Supplementary event	67
8.5.4	Rules of coherence	67
8.6	Probability of dependent events	67
8.6.1	Independence decomposition	67
8.6.2	Conditional probability	67

8.7	Discrete random variables	67
8.7.1	Finite domain and infinite domain	67
8.7.2	Mean and expected value	67
8.7.3	Standard deviation and variance	67
8.8	Distributions	67
8.8.1	Binomial distribution	67
8.8.2	Geometric distribution	67
8.8.3	Poisson distribution	67
8.8.4	Normal distribution and Standard distribution	67
8.9	Sampling	67
8.9.1	Point sampling	67
8.9.2	Central limit theorem	67
8.9.3	Confidence interval	67
8.10	Challenging questions	67

1 Function and its graph

‘Functions describes the world!’, one Professor in Mathematics of Massachusetts Institute of Technology (a.k.a. MIT) said that. His speech was greatly influential, as I have never heard such conclusive thinking about functions. In fact, in the past few years, whenever I was studying in schools, my thought about functions is always only about projecting elements from one set to another set, but what he said had a big impact to my knowledge about functions.

What function talks about, is a subjection of one elements to one another. It can be thought of as a pointing action started from element A to element B , which not so far away, if we could think of subjecting a lot of, a bunch of, or a list of, whatever, objects from one collection to another collection, and they can all be matched under this pointing action, then it is a so-called function.

For example, in an Indian factory, the production of food undergoes many different process. Those can all be called functions. Let say a raw material comes to the factory first, it then be put to a machine to chop into many small pieces. It is the chopping function inside the factory. Next, the chopped material will be put into a pool of yellowish-brownish liquid and be stirred by dirty hands. It is the Mixing function in the factory. After that, the liquid will be drained on the dirty floor and be stepped on by Indian workers so that they can be smell freshed. It is the flavouring function in the factory. Finally, it will be sold to stores, which is the selling function.

Another example is what our body does. We eat and drink, going down the digestive system, and we sit on a toilet. Although we stupid human knows nothing about how the digestive system works, we could still name the conversion from food to poops a digestion, which means the digestive function representing the process in our body.

So we know that function as an english word represents the naming of a process of conversion, it is the time to explore how Math functions works.

1.1 What is a Function?

A function is defined as follow:

Definition 1.1 (Function). *Given an input x and an output y , a **function** is a relation between x and y so that we can write $y = f(x)$ to represent the relationship.*

Essential Practice 1.1.0.1. *Write down functions for the following input-output variables:*

1. u as input and v as output;
2. b as input and a as output;

3. n as input and 1 as output (which we call it a constant function);
4. x^2 as input and y as output;
5. xy as input and z as output;
6. 2^x as input and k as output;
7. \sqrt{p} as input and q as output;

Remark. It is notable that we may write functions as $y = g(x)$, $y = h(x)$, $y = d(x)$, ... as we want. The 'naming' of a function is always definitive and up to user's construction.

We can also apply functions after functions. To do so, we have to talk about the following:

Definition 1.2 (Composite functions). Let f and g be functions such that f takes x as input and y as output, and g takes y as input and z as output. Then we can say there is a function h that takes x as input and z as output. In other words, h is a **composite function** such that $z = h(x) = g(f(x))$.

Essential Practice 1.1.0.2. Let f and g be functions such that $b = f(a)$, $c = g(b)$. Write a function for a as input and c as output.

1.1.1 Function as an input-output pair

We may now consider how functions carry things to things using arrow notations. We may use a stroked arrow \mapsto to emphasis the carrying process. Let's take a look at the following examples.

Example. Given a function that undergoes the process of adding one to the given element. We can say

$$\dots, 1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 4, \dots$$

In other words, we may write

$$x \mapsto x + 1$$

to generalize the process of adding one to the given element, which is how we usually write to describe any functions.

The example shows that we can generalize the process using algebraic notation, in which it is usually in the form of

$$x \mapsto f(x)$$

with the function f . It is equivalent to say that

$$\dots, 1 \mapsto f(1), 2 \mapsto f(2), \dots$$

but what's important is how explicit the mapping process is done. More examples could be viewed to familiarize with it.

Example. *To describe a function undergoes the process ‘multiplying the element by two and add one to it’, we may examine that*

$$\dots, 0 \mapsto 1, 1 \mapsto 3, 2 \mapsto 5, \dots$$

so that it is equivalent to write

$$x \mapsto 2x + 1$$

as a generalization of the function. It is equivalent to write $f(x) := 2x + 1$ to emphasize that we shall call the function f as a naming for the given process, as long as we can write

$$x \mapsto f(x), f(x) := 2x + 1$$

Essential Practice 1.1.1.1. *Examine the following functions from 0 to 3, and generalize the function using algebraic notation. You may either choose writing $x \mapsto \square$ directly or $x \mapsto f(x), f(x) := \square$.*

1. *Multiply the element by 3 and then add 2 to it.*
2. *Divide the element by 10 and then Subtract 7 from it.*
3. *Multiply the element by a and then add b to it.*
4. *Squaring the element.*
5. *Multiply 3 to the square of the element, an then add 5 to it.*
6. *Add 1 to the element first, then multiply the square of the result by 6, and then add 1 to it.*
7. *Subtract 8 from the element first, then divide the square of the result by 2, and then add 4 to it.*
8. *Subtract h from the element first, then multiply the square of the result by a , and then add k to it.*

It is notable that a function can only give one output for each input, which makes the next page a fruitful discussion.

1.1.2 Defining a function with variables

So far, we have learned how writing a generalization of a function is, and have we used algebraic notation to shorten the examination, one suggest we can always write algebraic notation for a function definition. In addition, we also find that a function cannot have more than one output, as we see functions as a pointing process from one element to another element. So we have the following definition:

Definition 1.3 (Function). *A **function** f of x defines the pointing process $x \mapsto f(x)$ is a one-to-one pointing process, which can have only one output.*

It is important to note that $f(x)$ is a function of x if and only if one x produce one $f(x)$. The x is called a *dummy variable*, which is a variable that can be changed all the time. For example, writing $f(y)$ or $f(z)$ still makes sense to say a function f , but not of x .

From now on, we can determine whether a given relation is a function or not.

Example. *Given the relation $y = mx + c$. Since one x can produce only one y , y is a function of x ; on the other hand, we also see one y produces only one x , so x is also a function of y .*

Example. *Given the relation $y = x^2$. Since one x can produce only one y , y is a function of x ; however, one y may produce more than one x , say if $y = 4$ then x can be 2 or -2 , so x is not a function of y .*

Essential Practice 1.1.2.1. *Determine whether the following given relation between x and y is a function of one another or not. Provide counterexample if it is not a function.*

1. $y = -x$;
2. $y = 4x + 3$;
3. $y = \frac{1}{x}$;
4. $y = \frac{x}{6}$;
5. $y^2 = x$;
6. $y^2 = x^2$;
7. $y^3 = 4x^2 - 3$.

1.1.3 Domain, Co-domain and Range

For explicit definition of a function, we need the following concepts to help with: *domain*, *co-domain* and *range*.

A **domain** is where the input comes from, which is usually half-customized and half-restricted. For example, $f(x) = \frac{1}{x}$ can have input of negative real numbers, positive real numbers, any complex numbers except 0. This means that 0 is naturally restricted by the operation of $\frac{1}{x}$, but other than 0, we can choose freely our input from all complex numbers. Thus, the largest domain of $f(x) = \frac{1}{x}$ is all complex numbers except 0. However, it is not saying that the domain of $f(x) = \frac{1}{x}$ must be all complex numbers except 0, we can still put restrictions on our own, what means by customize, like all real numbers except 0 or all positive real numbers except 0 as its domain, is still a possible choice. Hence, we shall usually talk about the *greatest possible domain of a function* if we need to find the natural restrictions, and the *domain of a function* if we are going to define our source of input.

Essential Practice 1.1.3.1. Find the greatest possible domain of the following functions if (i) the output is restricted to complex numbers and (ii) the output is restricted to be real numbers:

1. $f(x) := x$;
2. $f(x) := ax + b$;
3. $f(x) := \frac{a}{x}$;
4. $f(x) := x^2$;
5. $f(x) := a(x - h)^2 + k$;
6. $f(x) := \sqrt{x}$;
7. $f(x) := \sqrt[3]{x}$;
8. $f(x) := \frac{1}{\sqrt{x}}$;

A **co-domain** is where the output can go to. It is more likely a limitation of the output of the function so that we know where our target is. Similarly, we shall usually talk about the *greatest possible co-domain of the function* if we need to find the natural restrictions, and the *domain of the function* if we are going to define our target output.

Essential Practice 1.1.3.2. Find the greatest possible co-domain of the following functions if the input is unrestricted:

1. $f(x) := x;$
2. $f(x) := ax + b;$
3. $f(x) := \frac{a}{x};$
4. $f(x) := x^2;$
5. $f(x) := a(x - h)^2 + k;$
6. $f(x) := \sqrt{x};$
7. $f(x) := \sqrt[3]{x};$
8. $f(x) := \frac{1}{\sqrt{x}};$

With domain and co-domain, we can now define a function in a more explicit manner. Writing a function with where the inputs are in and where the outputs to go, we have a nice notation - an arrow \rightarrow emphasizing the direction from domain to co-domain. In general, we will write

$$f : D \rightarrow R$$

to specify the function f is a function goes from domain D to co-domain R . The formal way to define a function is like below:

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x)$$

which reads ‘a function f sending a real number x to a real number $f(x)$ ’.

Example. For a function sending a natural number n to a natural number 2^n , we may write its definition as

$$f : \mathbb{N} \rightarrow \mathbb{N}, n \mapsto 2^n$$

Example. For a function sending an integer n to a rational number 2^n , we may write its definition as

$$f : \mathbb{Z} \rightarrow \mathbb{Q}, n \mapsto 2^n$$

We shall see both example shows the same function process $f(n) := 2^n$ but different domain and codomain. Therefore, we acknowledge that although both functions are having the same process, they are indeed representing different things, which yields they are in fact different functions.

From this point of view, we can further examine the so-called **range** of a function, which is the exact target region of the function under the codomain. We shall define some set notation to present its meaning.

Definition 1.4 (Union and Intersection). *Let A and B be sets. The **union** of A and B is defined as*

$$A \cup B := \{x : x \in A \text{ or } x \in B\}$$

*while the **intersection** of A and B is defined as*

$$A \cap B := \{x : x \in A \text{ and } x \in B\}$$

In fact, we say union is the collection of the objects which are either in set A or set B , or simply say it is the joined set of two sets. We can take a look at the following examples:

Example. *Let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5\}$, then*

$$A \cup B = \{1, 2, 3, 4, 5\}$$

Example. *Let $A = \{1, 3, 5, 7, 9, \dots\}$ be the set of all positive odd numbers, $B = \{2, 4, 6, 8, 10, \dots\}$ be the set of positive even numbers, then*

$$A \cup B = \{1, 2, 3, 4, 5, 6, \dots\}$$

which is the set of all positive numbers. Sometimes, we may write the set by specifying the property of the elements as following:

$$A \cup B = \{x : x \text{ is a positive integer}\}$$

For intersection, it is generally speaking the collection of repeated elements in both sets, or we can say the sharing elements. We can take a look at the following examples:

Example. *Let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5\}$, then*

$$A \cap B = \{3, 4\}$$

Example. *Let $A = \{1, 3, 5, 7, 9, \dots\}$ be the set of all positive odd numbers, $B = \{2, 4, 6, 8, 10, \dots\}$ be the set of positive even numbers, then*

$$A \cap B = \emptyset$$

which is the set with no element, an empty set.

We may represent the two definitions by shading regions in a Venn-diagram. Suppose we are calling a set by enclosing the region by a circle, then we have the following representation.

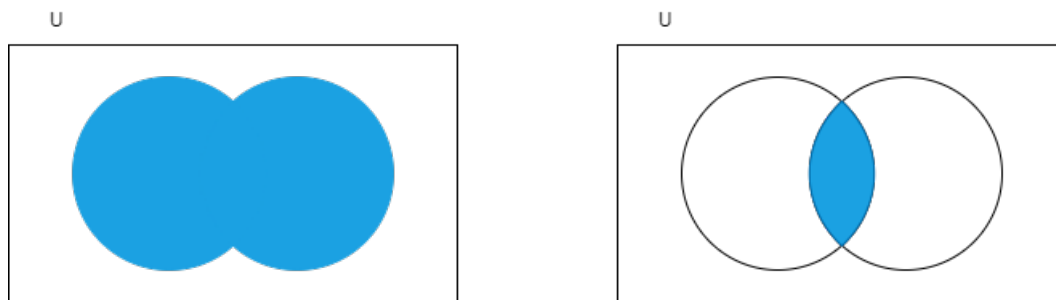


Figure 1: Union(Left) and Intersection(Right) of two sets

Essential Practice 1.1.3.3. Let $A = \{1, 3, 4, 6, 7\}$, $B = \{3, 4, 5, 7, 8\}$, then find $A \cup B$ and $A \cap B$.

Essential Practice 1.1.3.4. Prove the following identities:

1. $A \cap (B \cup C) \equiv (A \cap B) \cup (A \cap C)$;
2. $A \cup (B \cap C) \equiv (A \cup B) \cap (A \cup C)$.

Definition 1.5 (Range). The **range** of a function $f : D \rightarrow R$, denoted by $\mathbf{Ran}(f)$, is the set $f(D) \cap R$, where $f(D) := \{f(x) : x \in D\}$.

Usually, a teacher in high school aims to discuss the range of a function rather than the co-domain of a function, as what he shall teach is the size of the output, but not where the output shall be in. However, the difference between the co-domain of a function and the range of a function is we can further define the concept of a well-defined function with the concept of range.

Example. For a function f defined as

$$f : \mathbb{N} \rightarrow \mathbb{N}, n \mapsto 2^n$$

The range of f , denoted by $\mathbf{Ran}(f)$, is the set of all possible outcomes of $f(n)$ with natural numbers n . That is, the set

$$\{2^n : n \in \mathbb{N}\} = \{2, 4, 8, 16, 32, \dots\}$$

Example. For a function f defined as

$$f : \mathbb{Z} \rightarrow \mathbb{Q}, n \mapsto 2^n$$

$\mathbf{Ran}(f)$ is the set of all possible outcomes of $f(n)$ with integers n . That is, the set

$$\{2^n : n \in \mathbb{Z}\} = \{\dots, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, \dots\}$$

We must observe that the two ranges of the same function process f are different when their co-domains are different. This shows the importance of discussion of co-domain when we are defining ranges of functions.

Essential Practice 1.1.3.5. *Find the range of the following functions:*

1. $f : \mathbb{N} \rightarrow \mathbb{N}, n \mapsto n;$
2. $f : \mathbb{N} \rightarrow \mathbb{N}, n \mapsto 3^n;$
3. $f : \mathbb{N} \rightarrow \mathbb{N}, n \mapsto 3^n - n^3;$
4. $f : \mathbb{Z} \rightarrow \mathbb{Z}, n \mapsto n;$
5. $f : \mathbb{Z} \rightarrow \mathbb{Z}, n \mapsto n^2;$
6. $f : \mathbb{Z} \rightarrow \mathbb{Q}, n \mapsto 5^n;$
7. $f : \mathbb{Q} \rightarrow \mathbb{Q}, n \mapsto n;$
8. $f : \mathbb{Z} \rightarrow \mathbb{Q}, n \mapsto n/2;$
9. $f : \mathbb{N} \rightarrow \mathbb{Q}, n \mapsto n/10;$
10. $f : \mathbb{Z} \rightarrow \mathbb{R}, n \mapsto \pi n^2;$

Essential Practice 1.1.3.6. *Are the functions in previous practice all well-defined? Which of them are not?*

1.2 Graph of a function

Graphing has been an essential skill in understanding mathematical objects. We usually say it is a mathematical modeling technique. Through graphing, we can see the relationship between the input and output clearly, whether they are related, increasing or decreasing. It is also easier to draw conclusion to estimations by valid graphs. This section aims to build the concept of representing functions by graphs.

1.2.1 Blobs-and-arrows diagram for discrete functions

We shall first take a step backward to a simpler function - a *discrete function* with finite inputs. This will help the construction very much.

Definition 1.6 (Discrete functions). A **discrete function** is a function with direct indication of the function process for each element in domain. That is, for each element $x \in D$, the output $f(x)$ is assigned to co-domain directly.

A discrete function is usually with a discrete domain, and random assignment.

Example. Let $D := \{1, 2, 3, 4, 5\}$ be the domain of a function f . Suppose f is defined by

$$f(x) := \begin{cases} 1 & , x = 1 \\ 4 & , x = 2 \\ 3 & , x = 3 \\ 2 & , x = 4 \\ 5 & , x = 5 \end{cases}$$

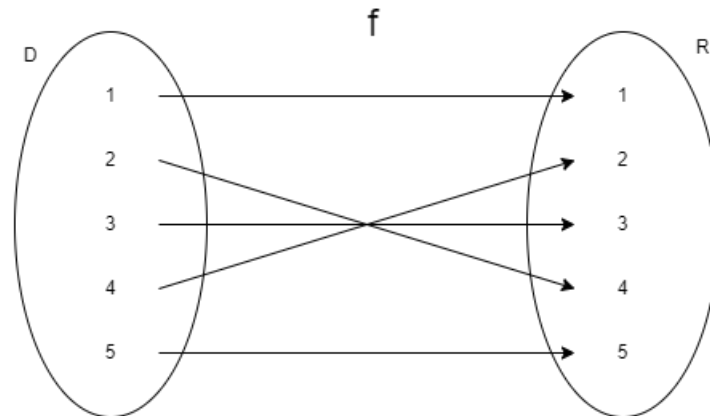
In this case, f is a discrete function.

To represent the above function using a graph, it is recommended to use a *blobs-and-arrows* diagram.

Definition 1.7 (Blobs-and-arrows diagram). A **blobs-and-arrows** diagram is a diagram representing a function by denoting the domain and co-domain by two circles, and for each element in the circle of domain, a pointer arrow over-set with f pointing to one of the element in the co-domain, to emphasize the relation between the two elements are input and output pair.

It is undesired to read through such complicated definition of the diagram. Let's see the following example.

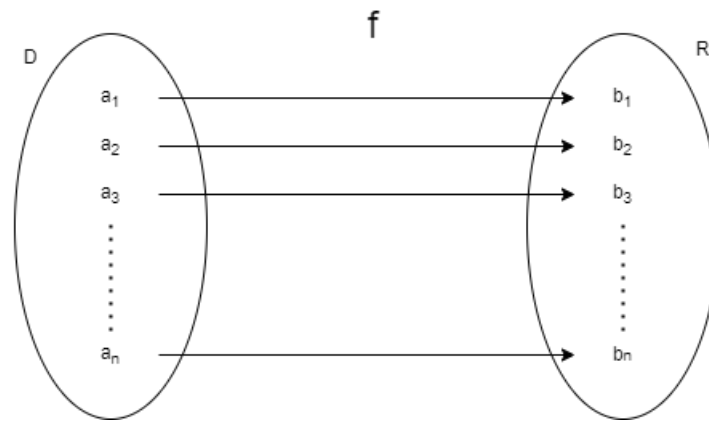
Example. Recall the function in previous example, we could draw the blobs-and-arrows diagram as shown.



In the figure above, the name of the function is hung over the main diagram, while the left ellipse denote the set of elements of domain D and the right ellipse denote the range of elements of codomain R . The letter R could also be interpreted as the range of f .

For a discrete function with finitely many inputs, where we will take them as a sequence, it can also be used to represent it efficiently. We just need to have some modification.

Axiom. Let $f : D \rightarrow R$ be a function. Let $a_1, a_2, a_3, \dots, a_n$ be a sequence of numbers in D and $b_1, b_2, b_3, \dots, b_n$ be a sequence of numbers in R . Suppose $b_1 = f(a_1), b_2 = f(a_2), \dots, b_n = f(a_n)$ defines the one-to-one correspondence from D to R . Then the following blobs-and-arrows diagram describes f formally:



Essential Practice 1.2.1.1. Draw the blobs-and-arrows diagram for the following discrete functions:

$$1. f(1) = 3, f(2) = 6, f(3) = 7, f(4) = 2, f(5) = 3.$$

$$2. f(x) := \begin{cases} 1 & , x = 1, 2, 3, 4, 5 \\ 2 & , x = 6, 7 \\ 3 & , x = 8, 9, 10 \\ 4 & , x = 11, 12, 13, 14, 15 \\ 5 & , x = 16, 17, 18, 19 \\ 6 & , x = 20 \end{cases}$$

1.2.2 The pairing table and xy-coordination

To extend our concept of function from discrete version to continuous version, that is, infinitely many numbers can be plugged into the function for calculation, we may need to ‘fill up’ the holes of the domain.

Let say we have 1 and 2 in the domain D , then all the real numbers within 1 and 2 also in D . This is the process of **continuation** or **extension**, and the whole set of real numbers within 1 and 2 is called the **interval** between 1 and 2. We can denote the interval by $[1, 2]$ if 1 and 2 are included in the meaning, or $(1, 2)$ if 1 and 2 are excluded from the meaning. To better distinct the difference between $[1, 2]$ and $(1, 2)$, we call the previous one the **closed interval** between 1 and 2, while the latter one the **open interval** between 1 and 2.

From this point of view, a real number line is used to represent the concept of any intervals. Let's define the real number line and discuss intervals using a number line.

Definition 1.8 (Real number line). *A **real number line** is a straight line with ordered property so that one of its direction is increasing, and the other direction is decreasing.*

We usually draw a horizontal real number line for convenience.

For the sake of simplicity and connected-understanding of open and closed intervals, we may put the desired parenthesis in the real number line to denote the meaning. Following that a closed interval is in a block-typed while open interval is in a round-typed, we can directly place them on the real number line.

Example. *The open interval $(0, 1)$.*

Example. *The open interval $(-1.5, 5.5)$.*

Example. *The closed interval $[0.5, 1]$.*

Example. *The closed interval $[-100, -67]$.*

Essential Practice 1.2.2.1. *Draw the mentioned interval on the provided real number line with suitable parenthesis.*

1. $[0, 1]$.
2. $[\alpha, \beta]$.
3. $(0, 1)$.
4. (α, β) .

It is of course we could draw it vertically with the same meaning, by indicating upward as positive direction and downward as negative direction. Consider a function $y = f(x)$ defines the relation between x and y coordinate, we can combine two real number lines in an orthogonal way, like a cross, to present the coordination in a sensible way. This is called the **continuous graph of a function**, or we simply call it the *graph of a function* if there's no ambiguity of what we are discussing.

In order to draw the graph of a function, we need to figure out the valid pairs of x and y first, so that the points could be addressed correctly for the function. We will make use of a **pairing table** so that each value of x in the domain is paired with a unique value of y in the co-domain. We will show its use with some examples.

Example. Let G be the graph of $y = x + 1$. Then we have the table of testing xy -pair

x	0	1	2	3	4
$y = x + 1$	1	2	3	4	5

and thus the graph of G be like

by linking up the points with straight lines.

Example. Let G be the graph of $y = x^2$. Then we have the table of testing xy -pair

x	-2	-1	0	1	2
$y = x^2$	4	1	0	1	4

and thus the graph of G be like

by linking up the points with straight lines. Moreover, by increasing the points of testing value to, say, 100 or 1000, a see-smoothing curve can be drawn by computer.

It is the continuation of the curve under testing value control.

Even higher power polynomial functions can be drawn using this strategy.

Example. Let G be the graph of $y = x^3 + 2x^2 + x - 1$. Then we have the table of testing xy -pair

x	0	1	2	3	4
$y = x^3 + 2x^2 + x - 1$	-1	3	17	47	99

and thus the graph of G be like, with many enough testing points,

with the continuation.

Essential Practice 1.2.2.2. Configure the xy -table with testing values in between -3 to 3 and draw a suitable graph for the following functions.

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

You may also check the graph using WolframAlpha desktop to convince yourself with the plot. Try to conclude the difference between large amount and small amount of points of testing.

With the xy-coordination of a graph, many features and prediction could be done. We will try to analyze the properties that we could make use of in the next subsection.

1.3 Fundamental properties of a graph

A graph shows the relationship between two variables, namely the vertical component (in this case, it is the value of y) and the horizontal component (respectively, the value of x). We will make use of this relationship and consider different situations.

1.3.1 points lying on the graph

Let us reverse the sight of plotting a graph of a function. We plot the graph by linking up the points we get from inputting x into the function to get the corresponding value of y , so that each point (x, y) describes when the function has the input x . We could now see that if a point (x, y) is passed through by a curve then it is the same as the point lies on that curve. We shall say the point (x, y) **satisfies** the equation $y = f(x)$, since this is the definition of a graph.

Theorem. *For any real-valued continuous functions f , any coordinates (x, y) lies on the graph $y = f(x)$ if and only if the xy -pair satisfies the equation $y = f(x)$.*

Example. *For a given function $f(x) = 3x + 1$, the following are true:*

1. *The point $(0, 1)$ lies on $y = f(x)$;*
2. *The point $(3, 10)$ lies on $y = f(x)$;*
3. *The point $(-2, 1)$ does not lay on $y = f(x)$;*
4. *The point $(-10, 2)$ does not lay on $y = f(x)$.*

Example. *For a given function $f(x) = 3x^3 + 2x^2 + x + 1$, the following are true:*

1. *The point $(0, 1)$ lies on $y = f(x)$;*

2. The point $(-1, -1)$ lies on $y = f(x)$;
3. The point $(-2, 0)$ does not lay on $y = f(x)$;
4. The point $(-10, 2)$ does not lay on $y = f(x)$.

Essential Practice 1.3.1.1. For a given function $f(x) = x^2 - 1$, show whether the following points are lying on the graph of $y = f(x)$:

1. The point $(0, 1)$;
2. The point $(-1, 0)$;
3. The point $(-2, 3)$;
4. The set of points $(-t, t^2 - 1)$ parametrized by the variable t .

We may now examine another feature of a graph, namely **the region of inequality**. In the previous paragraph we find a point satisfying the equation $y = f(x)$ has the same meaning as the point lies on the curve. If we make a small translation of the y coordinate to the upper region separate by the curve, it becomes that all the points have a larger value of y than we could calculate by the function. For that region we will call it the region for $y > f(x)$; Similarly, if we make a small translation of the y coordinate to the lower region, it becomes that all the points in that region have a smaller value of y than we could calculate by the function. For that region, we will call it the region for $y < f(x)$.

Example. For a given function $f(x) = 3x + 1$, the following are true:

1. The point $(0, 2)$ lies above $y = f(x)$;
2. The point $(3, 9)$ lies below $y = f(x)$;
3. The point $(-2, 1)$ lies above $y = f(x)$;
4. The point $(-10, -120)$ lies below $y = f(x)$.

Example. For a given function $f(x) = 3x^3 + 2x^2 + x + 1$, the following are true:

1. The point $(0, 3)$ lies above $y = f(x)$;
2. The point $(-1, -2)$ lies below $y = f(x)$;
3. The point $(-2, 0)$ lies above $y = f(x)$;

4. The point $(10, 2)$ lies below $y = f(x)$.

Essential Practice 1.3.1.2. For a given function $f(x) = x^2 - 1$, determine the position of the following points with respect to the graph of $y = f(x)$:

1. The point $(0, 9)$;
2. The point $(-1, -1)$;
3. The point $(-2, 20)$;
4. The point $(10, 10)$.

1.3.2 Special intersections: axis intercepts

Usually the most important element of a graph is whether it cuts the coordinating axis, namely the y-axis and the x-axis. For practical reason we value them so much and we may give them special names: the y-intercept and the x-intercept.

Definition 1.9 (y-intercept of a graph). Given a function f , the function has its **y-intercept** when it cuts the y-axis. In other words, the y-intercept is defined as

$$y_0 := f(0)$$

whenever the function is well-defined at $x = 0$.

Similarly, we may want the same definition for x-intercept of a graph, however, we do not know whether the inverse function exists or not. This comes to some kind of difficulties to say directly the x-intercept. But still, we can give the possible condition for the discussion.

Definition 1.10 (x-intercept of a graph). Given a function f , the function has its **x-intercept** when it cuts the x-axis. In other words, the x-intercept is defined as the roots of

$$0 = f(x)$$

whenever the given equation is solvable.

We will be familiar with them with some valid examples.

Example. For the graph of the function $f(x) := x + 1$, we have:

- y-intercept: $y_0 = 0 + 1 = 1$;

- *x-intercept:* Set $0 = x + 1$, then $x_0 = -1$ is the root to the given equation, which is the *x-intercept* of the function.

Example. For the graph of the function $f(x) := x^2 - 1$, we have:

- *y-intercept:* $y_0 = 0^2 - 1 = -1$;
- *x-intercept:* Set $0 = x^2 - 1$, then $x_1 = -1$ or $x_2 = 1$ are the roots to the given equation, which are the *x-intercepts* of the function.

Example. For the graph of the function $f(x) := x^2 + 1$, we have:

- *y-intercept:* $y_0 = 0^2 + 1 = 1$;
- *x-intercept:* Set $0 = x^2 + 1$, which has no real solution. Then the graph of the function has no *x-intercepts*.

We see a function has its y-intercept most of the time, but not the same case for x-intercepts. It could have either no x-intercepts, only 1 x-intercept, or more than one x-intercept. It is a consequence of the fundamental theorem of algebra, but we will dig deeper in the future to discuss simple cases we could.

As long as we could plot not only the xy-coordinate but we could pair up something weird, where they could be different functions say $v(y)$ and $u(x)$, the graphs could no longer talked about y-intercept or x-intercept - they are not exactly about x and y! Hence, we turned the naming into something general: we plot the graph by vertical axis and horizontal axis, so let us call them the **vertical intercept** and the **horizontal intercept** respectively.

1.4 Solving equations using graphs of functions

1.4.1 Homogeneous equations

1.4.2 Non-homogeneous equations

1.4.3 Simultaneous equations

1.5 Transformation of functions

1.5.1 Translation

1.5.2 Dilation

1.6 Challenging questions

2 Trigonometric functions

A right-angled triangle is usually a pure triangle we want to study the beauty of geometry. In this section, we are going to examine the usage of trigonometric functions as far as we could.

2.1 Why are we calling it ‘trigonometry’ ?

The word ‘trigonometry’ can be divided into two parts - ‘trigono-’ and ‘-metry’. The word ‘trigono-’ comes from the word ‘trigonal’, meaning ‘the things of trigon’. What trigon means is a triangle: in English, the word ‘tri-’ means the number 3, and what counts to 3 can be named ‘tri-’ things, like tripoid, triangle, tricycle, trisep, etc. So a trigon is equal to a triangle, just a difference in perspective: one tries to describe the number of vertices in a polygon, while one makes a point to number of angles in it. Another word ‘-metry’ comes from the word ‘metric’, which means ‘the method of comparison’. In complete word, ‘trigonometry’ means ‘the method of comparison of triangle’. Therefore, when we learn trigonometry, we will compare many attributes: the length of sides in one triangle, the shapes and sizes of different triangles.

2.2 The relation of sides of triangle

We shall first declare what is a triangle, how it forms and what is special about it.

2.2.1 Triangle inequality and Pythagoras theorem

Given a triangle, if we could draw it on a plane, we must have it with following axiomatic structures:

- There are 3 vertices in a triangle.
- There are 3 edges connecting 3 vertices.
- Given a vertex and its adjacent edges, an angle is formed. There are 3 angles in total.

To emphasize the structure inheriting 3 angles, we call it a **triangle**, with the prefix *tri-* standing for the number 3.

We shall immediate find a property of any given triangle that any edge in a triangle must be shorter than the sum of length of the other two edges. Here we call it **the triangle inequality**.

Axiom (Triangle inequality). *Let a, b, c be some positive numbers. If they denote the side lengths of a triangle, then we have the following inequalities:*

- $a + b \geq c$;
- $a + c \geq b$;
- $b + c \geq a$.

We shall take them as axioms, since we do not need to dig too deep to the roots of such geometric constructions. We shall take it to be true with our intuition.

Theorem (Pythagoras theorem). *Let a, b, c be positive numbers. Let a, b be the adjacent sides of a right-angled triangle and c be the hypotenuse of the right-angled triangle. Then*

$$a^2 + b^2 = c^2$$

Corollary. *The following holds for a right-angled triangle with adjacent sides a, b and hypotenuse c .*

- $a < c$;
- $b < c$.

2.2.2 The Sine function

Historically, the sine function is defined for the relation between the opposite side of a given acute angle in a triangle and the hypotenuse of the triangle.

Definition 2.1 (The Sine function). *Let a, b, c be sides of a given right-angled triangle and θ be an acute angle inside the triangle as shown.*

*Then the **sine function** of θ , $\sin \theta$, is defined by*

$$\sin \theta := \frac{a}{c}$$

In some sense, we call the side a the opposite side of θ and c the hypotenuse. We so shorthand opposite side by 'o' and hypotenuse by 'h'. This produces a well known memorization technique of the relation described by sine function, read as 'soh' relation.

Theorem. *The range of the sine function in a right-angled triangle is constrained to be*

$$0 < \sin \theta < 1$$

Proof. By definition, $a < c$, so $\frac{a}{c} < 1$; at the same time, $a > 0$ and $c > 0$, hence $\frac{a}{c} > 0$. □

2.2.3 The Cosine function

The cosine function is defined for the relation between the adjacent side of a given acute angle in a triangle and the hypotenuse of the triangle.

Definition 2.2 (The Cosine function). *Let a, b, c be sides of a given right-angled triangle and θ be an acute angle inside the triangle as shown.*

*Then the **cosine function** of θ , $\cos \theta$, is defined by*

$$\cos \theta := \frac{b}{c}$$

In some sense, we call the side b the adjacent side of θ and c the hypotenuse. We so shorthand adjacent side by 'a' and hypotenuse by 'h'. This produces a well known memorization technique of the relation described by cosine function, read as 'cah' relation.

Theorem. *The range of the sine function in a right-angled triangle is constrained to be*

$$0 < \cos \theta < 1$$

Proof. By definition, $b < c$, so $\frac{b}{c} < 1$; at the same time, $b > 0$ and $c > 0$, hence $\frac{b}{c} > 0$. □

2.2.4 The tangent function

The tangent function is defined for the relation between the opposite side and the adjacent side of a given acute angle in a triangle.

Definition 2.3 (The Tangent function). *Let a, b, c be sides of a given right-angled triangle and θ be an acute angle inside the triangle as shown.*

*Then the **tangent function** of θ , $\tan \theta$, is defined by*

$$\tan \theta := \frac{a}{b}$$

In some sense, we call the side a the opposite side of θ and b the adjacent side of θ . We so shorthand opposite side by 'o' and adjacent side by 'a'. This produces a well known memorization technique of the relation described by tangent function, read as 'toa' relation.

Theorem. *The range of the sine function in a right-angled triangle is constrained to be*

$$0 < \sin \theta$$

Proof. By definition, $a > 0$ and $b > 0$, hence $\frac{a}{b} > 0$. □

2.2.5 Fundamental identities for trigonometric functions

Theorem. $\sin(90^\circ - \theta) = \cos \theta$

Proof. Suppose a, b, c be the opposite side, the adjacent side and the hypotenuse to the acute angle θ in a right-angled triangle respectively. Note that by the law of angle sum of triangle, if we denote the remaining angle other than θ and 90° by ϕ , we get $\phi = 90^\circ - \theta$. Then

$$\begin{aligned}\sin(90^\circ - \theta) &= \sin \phi \\ &= \frac{b}{c} \\ &= \cos \theta\end{aligned}$$

□

Corollary. $\cos(90^\circ - \theta) = \sin \theta$

Theorem. $\sin^2 \theta + \cos^2 \theta = 1$.

Proof. Suppose a, b, c be the opposite side, the adjacent side and the hypotenuse to the acute angle θ in a right-angled triangle respectively. Then

$$\begin{aligned}\sin^2 \theta + \cos^2 \theta &= \left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 \\ &= \frac{a^2}{c^2} + \frac{b^2}{c^2} \\ &= \frac{a^2 + b^2}{c^2} \\ &= \frac{c^2}{c^2} = 1\end{aligned}$$

□

Theorem. $\tan \theta = \frac{\sin \theta}{\cos \theta}$.

Proof. Suppose a, b, c be the opposite side, the adjacent side and the hypotenuse to the acute angle θ in a right-angled triangle respectively. Then

$$\begin{aligned}\frac{\sin \theta}{\cos \theta} &= \sin \theta / \cos \theta \\ &= \frac{a}{c} / \frac{b}{c} \\ &= \frac{a}{c} \cdot \frac{c}{b} \\ &= \frac{a}{b} \\ &= \tan \theta\end{aligned}$$

□

Example. Simplify the expression $\frac{[\sin(90^\circ - \theta) - \cos(90^\circ - \theta)]^2}{1 - (\tan \theta)(2 \cos^2 \theta)}$.

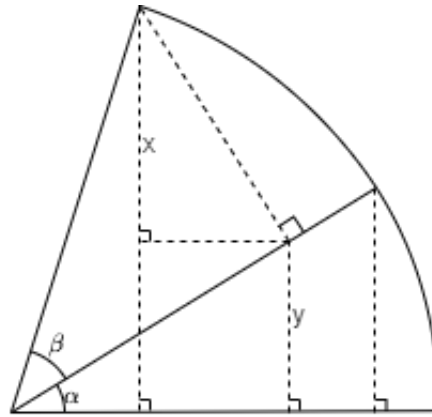
Sol. By proved theorem, we have

$$\begin{aligned} \frac{[\sin(90^\circ - \theta) - \cos(90^\circ - \theta)]^2}{1 - (\tan \theta)(2 \cos^2 \theta)} &= \frac{(\cos \theta - \sin \theta)^2}{1 - \left(\frac{\sin \theta}{\cos \theta}\right)(\cos^2 \theta)} \\ &= \frac{\cos^2 \theta - 2 \sin \theta \cos \theta + \sin^2 \theta}{1 - 2 \sin \theta \cos \theta} \\ &= \frac{1 - 2 \sin \theta \cos \theta}{1 - 2 \sin \theta \cos \theta} \\ &= 1 \end{aligned}$$

2.3 The compound angle formulae

Someone might be interested in computing compound angle formulae, which are important in developing different results of trigonometric relations. We will assume that α and β are both acute angles, so that all the results could be applied to whenever $0 < \alpha + \beta < 180^\circ$.

Consider the following sector with radius 1:



Theorem. $\sin(\alpha + \beta) = \sin \alpha \cos \beta + \sin \beta \cos \alpha$.

Proof. Referring to the figure, we have $\sin(\alpha + \beta) = x + y$.

On the other hand, we have

$$\begin{cases} x = \sin \beta \cos \alpha \\ y = \cos \beta \sin \alpha \end{cases}$$

Therefore, $\sin(\alpha + \beta) = \sin \alpha \cos \beta + \sin \beta \cos \alpha$. □

For now we have the result for $\sin(\alpha + \beta)$. Then the remaining results could be proven under substitution of variables:

Theorem. $\sin(\alpha - \beta) = \sin \alpha \cos \beta - \sin \beta \cos \alpha$.

Proof.

$$\begin{aligned}\sin(\alpha - \beta) &= \sin(\alpha + (-\beta)) \\ &= \sin \alpha \cos(-\beta) + \sin(-\beta) \cos \alpha \\ &= \sin \alpha \cos \beta - \sin \beta \cos \alpha\end{aligned}$$

□

Theorem. $\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$.

Proof.

$$\begin{aligned}\cos(\alpha + \beta) &= \sin\left(\frac{\pi}{2} - (\alpha + \beta)\right) \\ &= \sin\left(\frac{\pi}{2} - \alpha\right) \cos(-\beta) + \sin(-\beta) \cos\left(\frac{\pi}{2} - \alpha\right) \\ &= \cos \alpha \cos \beta - \sin \alpha \sin \beta\end{aligned}$$

□

Theorem. $\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$.

Proof.

$$\begin{aligned}\cos(\alpha - \beta) &= \cos(\alpha + (-\beta)) \\ &= \cos \alpha \cos(-\beta) - \sin \alpha \sin(-\beta) \\ &= \cos \alpha \cos \beta + \sin \alpha \sin \beta\end{aligned}$$

□

Theorem. $\tan(\alpha + \beta) = \frac{\tan \alpha + \tan \beta}{1 - \tan \alpha \tan \beta}$.

Proof.

$$\begin{aligned}\tan(\alpha + \beta) &= \frac{\sin(\alpha + \beta)}{\cos(\alpha + \beta)} \\ &= \frac{\sin \alpha \cos \beta + \sin \beta \cos \alpha}{\cos \alpha \cos \beta - \sin \alpha \sin \beta} \\ &= \frac{\frac{\sin \alpha}{\cos \alpha} + \frac{\sin \beta}{\cos \beta}}{1 - \frac{\sin \alpha}{\cos \alpha} \frac{\sin \beta}{\cos \beta}} \\ &= \frac{\tan \alpha + \tan \beta}{1 - \tan \alpha \tan \beta}\end{aligned}$$

□

Theorem. $\tan(\alpha - \beta) = \frac{\tan \alpha - \tan \beta}{1 + \tan \alpha \tan \beta}$

Proof.

$$\begin{aligned}\tan(\alpha - \beta) &= \frac{\tan \alpha + \tan(-\beta)}{1 - \tan \alpha \tan(-\beta)} \\ &= \frac{\tan \alpha - \tan \beta}{1 + \tan \alpha \tan \beta}\end{aligned}$$

□

All the results shown are based on $\alpha + \beta < 90^\circ$, so the case where $\alpha + \beta > 90^\circ$ are left to students.

2.4 The trigonometric laws

To understand the trigonometric functions better, we will try to extend the meaning of the trigonometric functions. The first step is to construct some laws with them so that the meaning could be seen.

In our discussion, we will look at two different triangles: one with acute angles only, while another with one obtuse angle. We will refer to the following triangles by T_1 and T_2 at anytime.

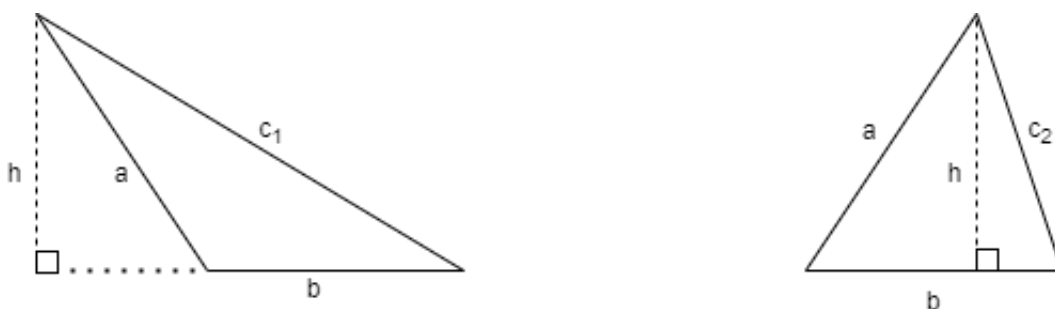


Figure 2: T_1 (left) and T_2 (right)

In convenience, we will denote the angle opposite to c_i to be γ_i . α to a and β to b .

2.4.1 The area formula of a triangle

Consider the area of T_1 and T_2 , if we are taken their height to be the same, we will have the following interesting results:

Consider the area of T_2 , we have $h = a \sin(\gamma_2)$, so its area is

$$\frac{1}{2}ab \sin(\gamma_2)$$

Consider the area of T_1 , we have $h = a \sin(180^\circ - \gamma_1)$. The value of $\sin(180^\circ - \gamma_1)$ can be obtained by consiering

$$\begin{aligned}
 \gamma_1 &= 180^\circ - \alpha - \beta \\
 \sin \gamma_1 &= \sin(180^\circ - \alpha - \beta) \\
 &= \sin((90^\circ - \alpha) + (90^\circ - \beta)) \\
 &= \sin(90^\circ - \alpha) \cos(90^\circ - \beta) + \cos(90^\circ - \alpha) \sin(90^\circ - \beta) \\
 &= \cos \alpha \sin \beta + \sin \alpha \cos \beta \\
 &= \sin(\alpha + \beta) \\
 &= \sin(180^\circ - \gamma_1)
 \end{aligned}$$

Hence, we have its area to be

$$\frac{1}{2}ab \sin(180^\circ - \gamma_1) = \frac{1}{2}ab \sin(\gamma_1)$$

In conclusion, the area of a triangle with given angle θ and adjacent sides can be computed as

Theorem. *Given a triangle with a given angle θ in which $0 \leq \theta \leq 180^\circ$ and corresponding adjacent sides of length a and b , the area of the triangle is*

$$\frac{1}{2}ab \sin \theta$$

2.4.2 The Sine Law

Our first law in discussion of trigonometric function is the **sine law** - the law inheriting the charateristics of sine function.

Theorem (Sine Law). $\frac{a}{\sin \alpha} = \frac{b}{\sin \beta} = \frac{c_i}{\sin \gamma_i}$.

Proof by extended right-angle triangle. We first complete the case in T_2 . It is not hard to see that

$$\begin{aligned}
 c_2 \sin \alpha &= a \sin \gamma_2 \\
 \frac{c_2}{\sin \gamma_2} &= \frac{a}{\sin \alpha}
 \end{aligned}$$

By rotation, similar results can be proven for other equality. We are done in the case of T_2 .

We then look at the case of T_1 . As in area computation, we shall see that

$$\sin \gamma_1 = \sin(180^\circ - \gamma_1)$$

Then

$$c_1 \sin \alpha = a \sin(180^\circ - \gamma_1)$$

$$\frac{c_1}{\sin \gamma_1} = \frac{a}{\sin \alpha}$$

□

Proof by area formula. Consider the area of the triangle, we obtain the following equality:

$$\frac{1}{2}ab \sin \gamma = \frac{1}{2}bc \sin \alpha = \frac{1}{2}ac \sin \beta$$

which yields the following result when dividing each expression by $\frac{abc}{2}$:

$$\frac{\sin \gamma}{c} = \frac{\sin \alpha}{a} = \frac{\sin \beta}{b}$$

□

The law connects an angle with its opposite sides by the ratio of sine function. This is a good perspective to apply the formula to solve any triangle with given two sides and an angle.

Example. Suppose a triangle $\triangle ABC$ with $AB = 3$, $AC = 5$, $\angle B = 60^\circ$. Then it is not hard to see

$$\begin{aligned} \frac{AC}{\sin \angle B} &= \frac{AB}{\sin \angle C} = \frac{BC}{\sin \angle A} \\ \sin \angle C &= \frac{3 \sin 60^\circ}{5} \\ &= \frac{3\sqrt{3}}{10} \\ \sin \angle A &= \sin \angle B \cos \angle C + \sin \angle C \cos \angle B \\ &= \frac{\sqrt{3}}{2} \cdot \frac{\sqrt{73}}{10} + \frac{3\sqrt{3}}{10} \cdot \frac{1}{2} \\ &= \frac{\sqrt{219} + 3\sqrt{3}}{20} \\ BC &= \frac{10\sqrt{3}}{3} \cdot \frac{\sqrt{219} + 3\sqrt{3}}{20} \\ &= \frac{3 + \sqrt{73}}{2} \end{aligned}$$

Remark. It is the usual sine function when one of the angle becomes 90° .

2.4.3 The Cosine Law

The second law in trigonometry to be discussed will be the **cosine law** - the law inheriting cosine function.

Theorem (Cosine Law). *The following formulae holds:*

- $c^2 = a^2 + b^2 - 2ab \cos \gamma \iff \cos \gamma = \frac{a^2 + b^2 - c^2}{2ab}.$
- $b^2 = a^2 + c^2 - 2ac \cos \beta \iff \cos \beta = \frac{a^2 + c^2 - b^2}{2ac}.$
- $a^2 = b^2 + c^2 - 2bc \cos \alpha \iff \cos \alpha = \frac{b^2 + c^2 - a^2}{2bc}.$

Standard proof. We first see the case in T_2 . It is not hard to see if we partition b into $x + y$ according to their order, we could construct the relations

$$\begin{aligned} \begin{cases} x + y &= b \\ a^2 - x^2 &= h^2 = c_2^2 - y^2 \\ x &= a \cos \gamma_2 \end{cases} \implies \begin{cases} a^2 - x^2 &= c_2^2 - (b - x)^2 \\ x &= a \cos \gamma_2 \end{cases} \\ \implies \begin{cases} a^2 + b^2 - 2bx &= c_2^2 \\ x &= a \cos \gamma_2 \end{cases} \\ \implies c_2^2 = a^2 + b^2 - 2ab \cos \gamma_2 \end{aligned}$$

On the other hand, for T_1 , we have to tackle particularly the obtused angle case. Note that if T_1 is divided into two right-angled triangle, then γ_1 can be divided into $\theta_1 + \theta_2$. It is equivalent to say

$$\begin{aligned} \cos \gamma_1 &= \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \\ &= \sin(90^\circ - \theta_1) \sin(90^\circ - \theta_2) - \cos(90^\circ - \theta_1) \cos(90^\circ - \theta_2) \\ &= -\cos(180^\circ - \gamma_1) \end{aligned}$$

Therefore, we pick to consider $x = a \cos(180^\circ - \gamma_1)$, then

$$\begin{aligned} c_1^2 - (b + x)^2 &= a^2 - x^2 \\ c_1^2 - b^2 - 2bx &= a^2 \\ c_1^2 &= a^2 + b^2 + 2ab \cos(180^\circ - \gamma_1) \\ c_1^2 &= a^2 + b^2 - 2ab \cos \gamma_1 \end{aligned}$$

This concludes that in any case, the cosine law holds. □

The best part of cosine law is that it generalizes the Pythagoras theorem to arbitrary triangles. It is also more applicable than Sine law to different situations.

Example. Refer to the same example in sine law. We can now directly find the length of BC .

$$AC^2 = AB^2 + BC^2 - 2 \cdot AB \cdot BC \cos \angle B$$

$$5^2 = 3^2 + BC^2 - 2 \cdot 3 \cdot BC \cos 60^\circ$$

$$0 = BC^2 - 3BC - 16$$

Shall we learn to solve quadratic equations, the solution to the equation is $\frac{3 \pm \sqrt{3^2 - 4(-16)}}{2} = \frac{3 \pm \sqrt{73}}{2}$. Since $\sqrt{73} > 3$, we need $BC > 0$, so $BC = \frac{3 + \sqrt{73}}{2}$.

2.4.4 The Tangent Law

It is the third law in trigonometry, which uses less but still understandable to write down. It is the **tangent law** - a law inheriting tangent function as its argument.

Theorem (Tangent Law). *The following formulae holds:*

- $\frac{a-b}{a+b} = \frac{\tan \frac{\alpha-\beta}{2}}{\tan \frac{\alpha+\beta}{2}}$.
- $\frac{b-c}{b+c} = \frac{\tan \frac{\beta-\gamma}{2}}{\tan \frac{\beta+\gamma}{2}}$.
- $\frac{c-a}{c+a} = \frac{\tan \frac{\gamma-\alpha}{2}}{\tan \frac{\gamma+\alpha}{2}}$.

Algebraic proof. Consider $\frac{\tan \frac{\alpha-\beta}{2}}{\tan \frac{\alpha+\beta}{2}}$. Denote $t_\alpha = \tan \frac{\alpha}{2}$ and $t_\beta = \tan \frac{\beta}{2}$.

$$\begin{aligned}
\frac{\tan \frac{\alpha-\beta}{2}}{\tan \frac{\alpha+\beta}{2}} &= \frac{\tan \frac{\alpha}{2} - \tan \frac{\beta}{2}}{1 + \tan \frac{\alpha}{2} \tan \frac{\beta}{2}} \frac{1 - \tan \frac{\alpha}{2} \tan \frac{\beta}{2}}{\tan \frac{\alpha}{2} + \tan \frac{\beta}{2}} \\
&= \frac{t_\alpha - t_\beta - t_\alpha^2 t_\beta + t_\alpha t_\beta^2}{t_\alpha + t_\beta + t_\alpha^2 t_\beta + t_\alpha t_\beta^2} \\
&= \frac{t_\alpha(\sec^2 \frac{\beta}{2}) - t_\beta(\sec^2 \frac{\alpha}{2})}{t_\alpha(\sec^2 \frac{\beta}{2}) + t_\beta(\sec^2 \frac{\alpha}{2})} \\
&= \frac{t_\alpha(\cos^2 \frac{\alpha}{2}) - t_\beta(\cos^2 \frac{\beta}{2})}{t_\alpha(\cos^2 \frac{\alpha}{2}) + t_\beta(\cos^2 \frac{\beta}{2})} \\
&= \frac{\sin \alpha - \sin \beta}{\sin \alpha + \sin \beta} \\
&= \frac{ab \sin \alpha - ab \sin \beta}{ab \sin \alpha + ab \sin \beta} \\
&= \frac{a - b}{a + b}
\end{aligned}$$

□

The tangent law relates two angles with two corresponding opposite sides. The application makes an agreement with the sine law, however, without advantages to apply the tangent law because of its complexity. We just do it a favor to learn a complete view to all three trigonometric functions.

In advance, let's introduce one more result related to the tangent law to close this extra section.

Theorem (Mollweide's Equations). *The following holds:*

$$\frac{a - b}{c} = \frac{\sin \frac{\alpha - \beta}{2}}{\cos \frac{\gamma}{2}} \quad (1)$$

$$\frac{a + b}{c} = \frac{\cos \frac{\alpha - \beta}{2}}{\sin \frac{\gamma}{2}} \quad (2)$$

Proof. As α, β, γ represents 3 angles in a triangle, their sum agrees with $\alpha + \beta + \gamma = 180^\circ$, therefore $\frac{\gamma}{2} = 90^\circ - \frac{\alpha + \beta}{2}$. We have the equality

$$\tan \frac{\alpha + \beta}{2} = \frac{1}{\tan \frac{\gamma}{2}}$$

By proved result on the Tangent Law:

$$\begin{aligned}
 \frac{a-b}{a+b} &= \frac{\tan \frac{\alpha-\beta}{2}}{\tan \frac{\alpha+\beta}{2}} \\
 &= \frac{\sin \frac{\alpha-\beta}{2} \cos \frac{\alpha+\beta}{2}}{\sin \frac{\alpha+\beta}{2} \cos \frac{\alpha-\beta}{2}} \\
 &= \frac{\sin \frac{\alpha-\beta}{2} \sin \frac{\gamma}{2}}{\cos \frac{\gamma}{2} \cos \frac{\alpha-\beta}{2}} \\
 &= \frac{\sin \frac{\alpha-\beta}{2}}{\cos \frac{\gamma}{2}} \frac{2 \sin \frac{\gamma}{2} \cos \frac{\gamma}{2}}{2 \cos \frac{\alpha-\beta}{2} \cos \frac{\gamma}{2}} \\
 &= \frac{\sin \frac{\alpha-\beta}{2}}{\cos \frac{\gamma}{2}} \frac{\sin \gamma}{\sin \alpha + \sin \beta} \\
 &= \frac{\sin \frac{\alpha-\beta}{2}}{\cos \frac{\gamma}{2}} \frac{c}{a+b} \\
 \frac{a-b}{c} &= \frac{\sin \frac{\alpha-\beta}{2}}{\cos \frac{\gamma}{2}}
 \end{aligned}$$

This proves the first equation.

For the second equation, consider

$$\begin{aligned}
 \frac{a+b}{c} &= \frac{a+b}{a-b} \cdot \frac{a-b}{c} \\
 &= \frac{\tan \frac{\alpha+\beta}{2} \sin \frac{\alpha-\beta}{2}}{\tan \frac{\alpha-\beta}{2} \cos \frac{\gamma}{2}} \\
 &= \frac{\cos \frac{\alpha-\beta}{2}}{\sin \frac{\gamma}{2}}
 \end{aligned}$$

Then the Mollweide's equations are proved. □

2.4.5 Heron's formula

In order to compute formula from all perspective, we now look at a well-known formula for computing area of an arbitrary triangle using side length only. This is the Heron's formula.

Theorem (Heron's formula). *Let a, b, c be the side length of 3 sides of a triangle. Then the area of the triangle is computed by*

$$\sqrt{s(s-a)(s-b)(s-c)}$$

where $s = \frac{a+b+c}{2}$.

Geometric proof of Heron's formula. Let r be the radius of the inscribe circle. Then the area of the triangle can be computed by

$$\frac{ra + rb + rc}{2} = rs$$

where $s = \frac{a+b+c}{2}$.

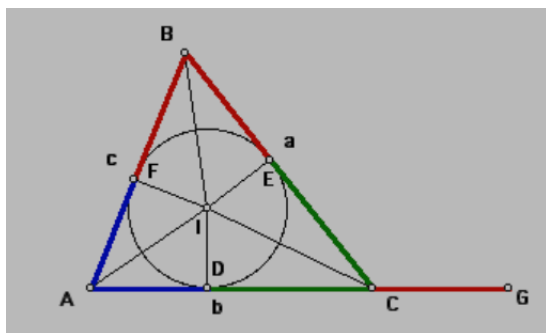


Figure 3: Reference: <http://jwilson.coe.uga.edu/EMT668/EMAT6680.2000/Umberger/MATH7200/HeronFormulaProject/GeometricProof/geoproof.html>

The following steps will try to resolve rs from other perspective. Here we orientate the side AC to be the base of the triangle (parallel to horizon), and extend the segment to G as in the figure. The area becomes $r \cdot AG$.

Next we adjoin the figure with a line from I so that $\angle AIJ = 90^\circ$ as follows.

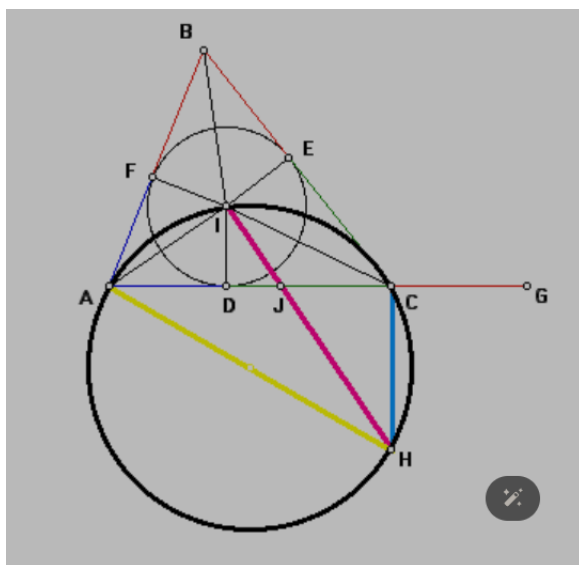


Figure 4: Reference: <http://jwilson.coe.uga.edu/EMT668/EMAT6680.2000/Umberger/MATH7200/HeronFormulaProject/GeometricProof/geoproof.html>

We try to figure out some similar triangles.

$$\angle AHC + \angle AIC = 180^\circ \quad (\text{concyelic, } \angle \text{ supp.})$$

$$\angle BIE + \angle AIC = 180^\circ \quad (\angle s \text{ at a pt.})$$

$$\therefore \angle AHC = \angle BIE$$

$$\angle BEI = \angle ACH = 90^\circ \quad (\text{By construction})$$

$$\therefore \triangle ACH \sim \triangle BEI$$

On the other hand, it is not hard to see that

$$\triangle IDJ \sim \triangle HCJ$$

Then the following relation holds:

$$\frac{AC}{CH} = \frac{BE}{IE} \quad (\triangle ACH \sim \triangle BEI)$$

$$\frac{AC}{CG} = \frac{CH}{ID} \quad (BE = CG, IE = ID)$$

$$= \frac{CJ}{DJ} \quad (\triangle IDJ \sim \triangle HCJ)$$

$$\frac{AG}{CG} = \frac{CD}{DJ} \quad (+1 \text{ on both sides})$$

$$AG \cdot ID = \frac{CD}{DJ} CG \cdot ID$$

$$= CD \cdot CG \cdot \frac{AD}{ID} \quad \left(\frac{ID}{DJ} = \frac{AD}{ID}, \triangle ADI \sim \triangle IDJ \right)$$

$$= CD \cdot CG \cdot \frac{AD \cdot AG}{ID \cdot AG}$$

$$AG \cdot ID = \sqrt{AD \cdot DC \cdot CG \cdot AG} \quad (\text{rearrangement})$$

Which, the area of $\triangle ABC$ is exactly

$$AG \cdot ID = \sqrt{AD \cdot DC \cdot CG \cdot AG}$$

and the proof is done. □

We have also the algebraic proof of heron's formula which is much sensible.

Algebraic proof of Heron's formula. Consider the sine law and cosine law where we have proven to be true in any case. We have the area formula

$$A = \frac{1}{2}ab \sin \gamma$$

and thus if we do a little trick, square it, we have

$$\begin{aligned}
 A^2 &= \frac{1}{4}a^2b^2 \sin^2 \gamma \\
 &= \frac{1}{4}a^2b^2(1 - \cos^2 \gamma) \\
 &= \frac{1}{4}a^2b^2\left(1 - \left(\frac{a^2 + b^2 - c^2}{2ab}\right)^2\right) \\
 &= \frac{1}{4}a^2b^2\left(1 + \frac{a^2 + b^2 - c^2}{2ab}\right)\left(1 - \frac{a^2 + b^2 - c^2}{2ab}\right) \\
 &= \frac{1}{16}(a^2 + 2ab + b^2 - c^2)(c^2 - a^2 + 2ab - b^2) \\
 &= \frac{1}{16}[(a + b)^2 - c^2][c^2 - (a - b)^2] \\
 &= \frac{1}{16}(a + b + c)(a + b - c)(-a + b + c)(a - b + c) \\
 &= s(s - a)(s - b)(s - c)
 \end{aligned}$$

Thus, by the positivity of area, we have

$$A = \sqrt{s(s - a)(s - b)(s - c)}$$

where $s = \frac{a + b + c}{2}$

□

Example. Let $\triangle ABC$ be with sides 7, 10, 13. Define $s = \frac{7 + 10 + 13}{2} = 15$. Then the area of the triangle is

$$S_{\triangle ABC} = \sqrt{15(15 - 7)(15 - 10)(15 - 13)} = \sqrt{1200} = 20\sqrt{3}$$

Essential Practice 2.4.5.1. Compute the area of $\triangle ABC$ if the side lengths are:

1. 1, 2, 3.
2. 3, 4, 5.
3. 2, 6, 7.
4. 13, 15, 17.
5. 9, 9, 6.

2.5 Extending trigonometric functions to describing circles

Trigonometric functions are not bounded by a triangle. In fact, if we want to, and trust to, extend trigonometric functions to any degree, we could still give a definition.

2.5.1 The relation between rectangular coordinates and polar coordinates

2.5.2 Defining trigonometry as coordinate function

2.5.3 Reduction formula for trigonometric functions

2.5.4 The C-A-S-T graph and reduction strategy

2.6 Challenging Questions

3 Linear functions

In this section, we focus on a specific form of function, which is called **Linear functions**, a type of functions that generate straight lines in the Euclidean space, or we may call it the xy-plane, denoted by \mathbb{R}^2 . It is the simplest form of function, and has many properties we are interested in.

3.1 Fundamental concepts of points

Revisiting the xy-plane, we may define some useful tools for measurement and ratio properties, since we shall always be equipped with the sense of measuring and performing comparisons.

3.1.1 Distance between two points

One important practical measurement in the xy-plane is the so-called **distance between two points**. We recall that given a right-angled triangle $\triangle ABC$ with $\angle ABC = 90^\circ$, the following relation on the side lengths holds:

$$|\overline{AB}|^2 + |\overline{BC}|^2 = |\overline{AC}|^2$$

where the stroke sign $|\cdot|$ denotes the length of a certain line and each line is mentioned explicitly using the over-lined notation \overline{XY} to indicate the line connecting point X and point Y . As we learned the relation is mentioned by *Pythagoras Theorem* or *Pythagorean Theorem*.

Following from the Pythagoras Theorem, we shall examine the distance between any two points on the xy-plane by the length of the straight line connecting them. With the help of the following figure, let us carry out our definition of distance between two points.

Definition 3.1 (Distance between two points). *Let $A(x_1, y_1)$ and $B(x_2, y_2)$ be two points on \mathbb{R}^2 -plane. The distance between A and B is computed by the formula*

$$\text{dist}(A, B) := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

We may think of the following construction for distance between two points:

Taking the right-angled triangle as distance reference, the vertical displacement becomes $|y_1 - y_2|$ and the horizontal displacement becomes $|x_1 - x_2|$. Putting the oblique displacement $\text{dist}(A, B)$ gives the Pythagorean relation, and we could get the desired formula.

Example. *Let $A(1, 2)$ and $B(4, -2)$ be two points on the rectangular coordinate plane, i.e. the xy-plane. Then,*

$$\text{dist}(A, B) = \sqrt{(1 - 4)^2 + (2 - (-2))^2} = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$$

Example. Let $A(1, 2)$ and $B(k, k - 1)$ be two points on the rectangular coordinate plane such that $\text{dist}(A, B) = 2$. Then,

$$\begin{aligned}\sqrt{(k-1)^2 + (k-3)^2} &= 2 \\ (k-1)^2 + (k-3)^2 &= 4 \\ k^2 - 2k + 1 + k^2 - 6k + 9 &= 4 \\ 2k^2 - 8k + 6 &= 0 \\ k^2 - 4k + 3 &= 0 \\ (k-1)(k-3) &= 0\end{aligned}$$

Then $k = 1$ or $k = 3$, i.e. the possible coordinates of B are $(1, 0)$ or $(3, 2)$.

Example. Given two trajectories $T_1 : (x_1(t), y_1(t))$ and $T_2 : (x_2(t), y_2(t))$, we say the distance between T_1 and T_2 is

$$\text{dist}(T_1, T_2) := \min_{t \geq 0} \sqrt{(x_1(t) - x_2(t))^2 + (y_1(t) - y_2(t))^2}$$

It is quite abstract to talk about the distance between two trajectories in this stage. Just keep in mind that this is a truth, or say you may believe in it.

Essential Practice 3.1.1.1. Compute the distance between following pair of points:

1. $(1, 0)$ and $(0, 1)$;
2. $(1, -1)$ and $(-1, 1)$;
3. $(3, 4)$ and $(12, -13)$;
4. $(7, -8)$ and $(-3, 0)$;

Essential Practice 3.1.1.2. Given the distance between two points, find the possible value(s) of the unknowns:

1. $A(0, 0), B(3, k)$; $\text{dist}(A, B) = 3$.
2. $A(2, 0), B(3, k)$; $\text{dist}(A, B) = 1$.
3. $A(k, 0), B(0, k)$; $\text{dist}(A, B) = 5$.
4. $A(3 + k, 1 - k), B(3, -3)$; $\text{dist}(A, B) = 8$.

3.1.2 Mid-point and Division points

Definition 3.2 (Mid-way of a trajectory). *Given a trajectory defined by $(x(t), y(t))$ where $t \in [0, r]$ with $r > 0$ a real number, the **midway** (or **half-way**) of the trajectory is given by the point M such that*

$$\text{arclength}((x(0), y(0)), M) = \text{arclength}(M, (x(r), y(r)))$$

In particular, if we are talking about a mid-point, we usually pick the shortest trajectory between the two points. This is due to the uniqueness of the shortest trajectory if our distance is well-defined. In the Euclidean space, the shortest trajectory is usually the straight line between the two points.

Theorem (Mid-point of two points). *Given $A(x_1, y_1)$ and $B(x_2, y_2)$ be two points on the Euclidean space, the **mid-point** M of A and B is defined by*

$$\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right)$$

Proof. Let L be the straight line connecting A and B . Assume that $M := \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right)$, then

$$\begin{aligned} \text{dist}(A, M) &= \sqrt{\left(x_1 - \frac{x_1 + x_2}{2}\right)^2 + \left(y_1 - \frac{y_1 + y_2}{2}\right)^2} \\ &= \frac{1}{2} \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\ &= \frac{1}{2} \text{dist}(A, B) \\ \text{dist}(M, B) &= \sqrt{\left(\frac{x_1 + x_2}{2} - x_2\right)^2 + \left(\frac{y_1 + y_2}{2} - y_2\right)^2} \\ &= \frac{1}{2} \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\ &= \frac{1}{2} \text{dist}(A, B) \end{aligned}$$

In particular,

$$\text{dist}(A, M) + \text{dist}(M, B) = \text{dist}(A, B)$$

Thus M preserve a point on the straight line connecting A and B and in fact the mid-point of A and B . □

Example. *Given $A(100, 20)$ and $B(30, 70)$, their mid-point is the point $M(65, 45)$.*

Essential Practice 3.1.2.1. *For the following given pairs of points, find their mid-points:*

1. $(2, 2)$ and $(0, 0)$;
2. $(2, 1)$ and $(-2, -1)$.

Other than the mid-point, we also consider the general version of line separation, which is the point of division.

Definition 3.3 (Point of division). *Given two points A and B , any point P lies in between A and B is called **the point of division with ratio** $\text{dist}(A, P) : \text{dist}(P, B)$.*

Theorem (Formula for point of division). *Let $A(x_1, y_1)$ and $B(x_2, y_2)$ be a point on the Euclidean space. Let P be the point of division of A and B with ratio $r : s$. Then the coordinates of P is given by*

$$\frac{sA + rB}{r + s} = \left(\frac{sx_1 + rx_2}{r + s}, \frac{sy_1 + ry_2}{r + s} \right)$$

In particular, the mid-point of A and B is given by setting $s = r = 1$.

Proof. Similar to the proof of mid-point. Let $P := \left(\frac{sx_1 + rx_2}{r + s}, \frac{sy_1 + ry_2}{r + s} \right)$, then

$$\begin{aligned} \text{dist}(A, P) &= \sqrt{\left(x_1 - \frac{sx_1 + rx_2}{r + s}\right)^2 + \left(y_1 - \frac{sy_1 + ry_2}{r + s}\right)^2} \\ &= \frac{1}{r + s} \sqrt{(rx_1 - rx_2)^2 + (ry_1 - ry_2)^2} \\ &= \frac{r}{r + s} \text{dist}(A, B) \\ \text{dist}(P, B) &= \sqrt{\left(\frac{sx_1 + rx_2}{r + s} - x_2\right)^2 + \left(\frac{sy_1 + ry_2}{r + s} - y_2\right)^2} \\ &= \frac{1}{r + s} \sqrt{(sx_1 - sx_2)^2 + (sy_1 - sy_2)^2} \\ &= \frac{s}{r + s} \text{dist}(A, B) \end{aligned}$$

In particular,

$$\text{dist}(A, P) + \text{dist}(P, B) = \text{dist}(A, B)$$

Thus P preserve a point on the straight line connecting A and B and satisfies $\text{dist}(A, P) : \text{dist}(P, B) = \frac{r}{r + s} : \frac{s}{r + s} = r : s$. □

Example. *Let $A(2, 4)$ and $B(-4, 7)$. To trisect the line connecting A and B , we need two points on the line where they divide the straight line in the ratio of $1 : 1 : 1$. In particular, let the point*

closer to A be X and the other be Y , we have X the division point of ratio $1 : 2$ and Y the division point of ratio $2 : 1$. Therefore,

$$X := \left(\frac{2(2) + 1(-4)}{1 + 2}, \frac{2(4) + 1(7)}{1 + 2} \right) = (0, 5)$$

$$Y := \left(\frac{1(2) + 2(-4)}{1 + 2}, \frac{1(4) + 2(7)}{1 + 2} \right) = (-2, 6)$$

Essential Practice 3.1.2.2. Given $A(-100, -10)$ and $B(100, 100)$. Find the point of division for the following given ratio $r : s$:

1. $r : s = 1 : 1$;
2. $r : s = 1 : 3$;
3. $r : s = 3 : 7$;
4. $r : s = 19 : 91$.

3.1.3 Slope of two points

For two points, we want to know not only how far away they are, but also how steep they see each other, i.e. the direction of proceeding. We call it the slope of the two points.

To be professional, we say a slope of the two points is how line goes from the left point to the right point; that is, whether y increase with x or y decrease with x .

Definition 3.4 (Slope). Let $A(x_1, y_1)$ and $B(x_2, y_2)$ be two points on the Euclidean space. Then the **slope** of A and B , denoted by \mathfrak{M}_{AB} , is given by

$$\mathfrak{M}_{AB} := \frac{y_2 - y_1}{x_2 - x_1}$$

Example. The slope of $A(1, 2)$ and $B(4, 1)$ is $\mathfrak{M}_{AB} = \frac{1 - 2}{4 - 1} = -\frac{1}{3}$.

We could observe the following fact. For if two points are aligned horizontally on the plane, then their slope is computed equal to 0; if the right point is higher than the left, then the slope is positive, meaning the corresponding line is increasing with x ; otherwise the slope is negative, meaning the corresponding line is decreasing with x . For if two points are aligned vertically on the plane, then the slope is undefined, since we could not explicitly tell whether the corresponding line is increasing or decreasing - it could be both. Due to the uncertainty, we identify it as undefined slope. However, we can still define the slope of vertical lines for consistent usage, but that's another page of discussion.

Essential Practice 3.1.3.1. *Compute the slope of the following pairs of points:*

1. $(1, 0)$ and $(0, 1)$;
2. $(1, -1)$ and $(-1, 1)$;
3. $(3, 4)$ and $(12, -13)$;
4. $(7, -8)$ and $(-3, 0)$;

Definition 3.5 (Collinear). *3 points are called **collinear** if there is a straight line passes through 3 points at the same time.*

In fact, collinearity is somehow the converse of computing points of division, by the definition of point of division. We will make use of the equivalence in proving the following theorem:

Theorem. *3 points on the plane are collinear if and only if the slope of any two of the three points are the same.*

Proof. For the direction from collinear to slope equality, we have if 3 points are collinear, then they can be written as

$$A(x_1, x_2), B\left(\frac{sx_1 + rx_2}{r + s}, \frac{sy_1 + ry_2}{r + s}\right), C(x_2, y_2)$$

We then have

$$\begin{aligned} \mathfrak{M}_{AB} &= \frac{\frac{sy_1 + ry_2}{r + s} - y_1}{\frac{sx_1 + rx_2}{r + s} - x_1} \\ &= \frac{sy_1 + ry_2 - (r + s)y_1}{sx_1 + rx_2 - (r + s)x_1} \\ &= \frac{ry_2 - ry_1}{rx_2 - rx_1} \\ &= \frac{y_2 - y_1}{x_2 - x_1} \\ \mathfrak{M}_{BC} &= \frac{y_2 - \frac{sy_1 + ry_2}{r + s}}{x_2 - \frac{sx_1 + rx_2}{r + s}} \\ &= \frac{(r + s)y_2 - sy_1 + ry_2}{(r + s)x_2 - sx_1 + rx_2} \\ &= \frac{sy_2 - sy_1}{sx_2 - sx_1} \\ &= \frac{y_2 - y_1}{x_2 - x_1} \\ \mathfrak{M}_{AC} &= \frac{y_2 - y_1}{x_2 - x_1} \end{aligned}$$

Then collinear implies the equality of the slopes.

Conversely, suppose 3 points A, B, C such that $\mathfrak{M}_{AB} = \mathfrak{M}_{BC} = \mathfrak{M}_{AC} = m$. We may compute that

$$\begin{aligned}\text{dist}(A, B) &= \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \\ &= |x_A - x_B| \sqrt{1 + \mathfrak{M}_{AB}^2}\end{aligned}$$

Similarly, we have

$$\begin{aligned}\text{dist}(B, C) &= |x_B - x_C| \sqrt{1 + \mathfrak{M}_{BC}^2} \\ \text{dist}(A, C) &= |x_A - x_C| \sqrt{1 + \mathfrak{M}_{AC}^2}\end{aligned}$$

Without loss of generality, we may assume their x-coordinates are in fixed order, where I will set $x_A < x_B < x_C$. Then

$$\begin{aligned}\text{dist}(A, B) + \text{dist}(B, C) &= (|x_A - x_B| + |x_B - x_C|) \sqrt{1 + m^2} \\ &= (x_B - x_A + x_C - x_B) \sqrt{1 + m^2} \\ &= |x_A - x_C| \sqrt{1 + \mathfrak{M}_{AC}^2} \\ &= \text{dist}(A, C)\end{aligned}$$

This shows that B is a point of division of A and C , and hence A, B, C are collinear. \square

Corollary. *3 points on the plane are not collinear if and only if the slope of any two of the three points are not the same.*

By the theorem, we could easily show the collinearity of any 3 points, and of course we may push a little forward from it.

Example. *Determine the collinearity of following sets of points:*

1. $(0, 0), (1, 2), (2, 4)$.

Sol. Denote $(0, 0), (1, 2), (2, 4)$ by A, B, C respectively. Then

$$\begin{aligned}\mathfrak{M}_{AB} &= \frac{2 - 0}{1 - 0} = 2 \\ \mathfrak{M}_{BC} &= \frac{4 - 2}{2 - 1} = 2\end{aligned}$$

Then, by the previous theorem, A, B, C are collinear.

2. $(0, 1), (1, 2), (2, 2)$.

Sol. Denote $(0, 0), (1, 2), (2, 4)$ by A, B, C respectively. Then

$$\mathfrak{M}_{AB} = \frac{2 - 1}{1 - 0} = 2$$

$$\mathfrak{M}_{BC} = \frac{2 - 2}{2 - 1} = 0$$

Then, by the previous theorem, A, B, C are not collinear.

Essential Practice 3.1.3.2. Determine the collinearity of the following sets of points:

1. $(0, 0), (1, 3), (2, 6)$.
2. $(1, 0), (0, -1), (-1, -2)$.
3. $(0, 2), (1, -3), (-2, 6)$.
4. $(k, 4), (4 + k, 4 + k), (4, k)$.

This measurement provides another way of line construction. In fact, by one of its powerful property, there is a generalized version of slope with various application in higher mathematics and other fields including computer science.

Theorem (Invariant properties of slope of two points). *Let A and B be two points on the Euclidean space, and \mathfrak{M}_{AB} be the slope of A and B . Under any simultaneous translation to A and B , \mathfrak{M}_{AB} remains unchanged.*

Proof. Let $A(x_1, y_1)$ and $B(x_2, y_2)$ be two points. Suppose A' and B' be the resulting coordinates after simultaneously translating A and B , we may write $A'(x_1 + \Delta x, y_1 + \Delta y)$ and $B'(x_2 + \Delta x, y_2 + \Delta y)$. Then the slope after simultaneous translation is

$$\mathfrak{M}_{A'B'} = \frac{(y_2 + \Delta y) - (y_1 + \Delta y)}{(x_2 + \Delta x) - (x_1 + \Delta x)} = \frac{y_2 - y_1}{x_2 - x_1} = \mathfrak{M}_{AB}$$

□

3.2 Different forms of a linear function

Following the concepts of distance, point of division and slope, we can now discuss the form of a linear function.

3.2.1 Segment

We may first consider a segment in the plane, which is the straight line connecting any two points in the Euclidean plane, ends at that two points.

Definition 3.6 (Segment). *A **segment** with two endpoints A and B is the collection of all division points between A and B . In other words, it is a straight line with endpoints A and B .*

By the definition of division point, we may let $A(x_1, y_1)$ and $B(x_2, y_2)$ be two points, and two numbers r and s such that every division point of ratio $r : s$ could be represented in the following form

$$\left(\frac{sx_1 + rx_2}{r + s}, \frac{sy_1 + ry_2}{r + s} \right)$$

By simple algebra, we can see that $\frac{s}{r + s} = 1 - \frac{r}{r + s}$, so we may let $t := \frac{r}{r + s}$ be a number in between 0 and 1 such that every division point can be written as

$$((1 - t)x_1 + tx_2, (1 - t)y_1 + ty_2) = (x_1 + t(x_2 - x_1), y_1 + t(y_2 - y_1))$$

with $0 \leq t \leq 1$. Now, we could see the relationship between x - and y - coordinates in the following derivation:

$$\begin{aligned} y &:= (1 - t)y_1 + ty_2 \\ &= y_1 + t(y_2 - y_1) \\ &= y_1 + t \frac{y_2 - y_1}{x_2 - x_1} (x_2 - x_1) \\ &= y_1 + t\mathfrak{M}_{AB}(x_2 - x_1) \\ &= y_1 + \mathfrak{M}_{AB}(x - x_1) \end{aligned}$$

3.2.2 Line as extension of a segment

We have established that a segment can be represented by the equation

$$y = y_1 + \mathfrak{M}_{AB}(x - x_1)$$

where it is interesting to see the equation involves no variable t . In fact, it is independent of t and could be extended to out of the segment. We call this extension a **straight line** passes through A and B .

The above equation looks simple, but let us confirm that the derived equation is indeed satisfying the extension of a segment. We shall return to the fundamental form of a segment but revising the domain of t :

$$((1 - t)x_1 + tx_2, (1 - t)y_1 + ty_2)$$

where $-\infty < t < \infty$. In fact, a line should be proven to be this form rigorously. Let us dive into it.

Theorem. *A straight line passes through (x_1, y_1) and (x_2, y_2) is in the form of*

$$((1-t)x_1 + tx_2, (1-t)y_1 + ty_2)$$

where $-\infty < t < \infty$.

Proof. The major discussion is the domain of t diverges to infinity. Let us prove first that for every real value t , the point $((1-t)x_1 + tx_2, (1-t)y_1 + ty_2)$ is collinear with $A(x_1, y_1)$ and $B(x_2, y_2)$. In fact,

$$\mathfrak{M} = \frac{(1-t)y_1 + ty_2 - y_1}{(1-t)x_1 + tx_2 - x_1} = \frac{y_2 - y_1}{x_2 - x_1}$$

is true for all $t \neq 0$. At the same time, the case $t = 0$ is just considering (x_1, y_1) , which is immediately true. Hence, every point in that form is collinear with A and B .

Next, we shall show that t diverges at both ends. Suppose on contrary that t is bounded above by M , then the longest segment ends at $V_M((1-M)x_1 + Mx_2, (1-M)y_1 + My_2)$, but letting $V_{M+1}((-M)x_1 + (M+1)x_2, (-M)y_1 + (M+1)y_2)$

$$\text{dist}(A, V_{M+1}) = (M+1)\text{dist}(A, B) > M\text{dist}(A, B) = \text{dist}(A, V_M)$$

This contradicts that AV_M is the longest segment we have. Thus t is unbounded. The argument is similar for t 's lower bound and hence $-\infty < t < \infty$. \square

3.2.3 Point-slope form

Observe that

$$y = y_1 + \mathfrak{M}_{AB}(x - x_1) \iff y - y_1 = \mathfrak{M}_{AB}(x - x_1) \iff \frac{y - y_1}{x - x_1} = \mathfrak{M}_{AB}$$

for given points $A(x_1, y_1)$ and slope \mathfrak{M}_{AB} , we may give the following result.

Theorem. *If a line L passes through point $A(x_1, y_1)$ and having slope \mathfrak{M} , then L can be constructed from writing*

$$L : \frac{y - y_1}{x - x_1} = \mathfrak{M}$$

The meaning of this form can be thought of as: Given any point (x, y) on the line L , its slope with A is the same as the given slope \mathfrak{M} .

Example. *Suppose L is a straight line passes through $(0, 0)$ with slope being equal to 3. Then L could be modified as*

$$L : \frac{y - 0}{x - 0} = 3 \iff y = 3x$$

Example. Suppose L is a straight line passes through $(1, 2)$ with slope being equal to 4. Then L could be modified as

$$L : \frac{y - 2}{x - 1} = 4 \iff y = 4x - 2$$

3.2.4 Two-point form

With any two given points on the Euclidean plane, we could rewrite the slope \mathfrak{M} into $\frac{y_2 - y_1}{x_2 - x_1}$. This means the line passes through any two given points could be modelled by

Theorem. If a line L passes through point $A(x_1, y_1)$ and $B(x_2, y_2)$, then L can be constructed from writing

$$L : \frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}$$

Example. Suppose L is a straight line passes through $(0, 0)$ and $(1, 2)$. Then L could be modified as

$$L : \frac{y - 0}{x - 0} = \frac{2 - 0}{1 - 0} \iff y = 2x$$

Example. Suppose L is a straight line passes through $(1, 2)$ and $(10, 3)$. Then L could be modified as

$$L : \frac{y - 2}{x - 1} = \frac{3 - 2}{10 - 1} \iff y = \frac{1}{9}x + \frac{17}{9}$$

3.2.5 Slope-intercept form

Given the slope \mathfrak{M} and the y-intercept c , we could take the y-intercept as one particular coordinate $(0, c)$. This means the line with given information is

Theorem. If a line L has slope \mathfrak{M} and y-intercept c , then L can be constructed from writing

$$L : \frac{y - c}{x - 0} = \mathfrak{M}$$

Example. Suppose L is a straight line with slope 5 and y-intercept 0. Then L could be modified as

$$L : \frac{y - 0}{x - 0} = 5 \iff y = 5x$$

Example. Suppose L is a straight line with slope $\frac{1}{10}$ and y-intercept 5. Then L could be modified as

$$L : \frac{y - 5}{x - 0} = \frac{1}{10} \iff y = \frac{1}{10}x + 5$$

3.2.6 Intercepts form

Except the previous forms, an intercepts form is a form purely taken from the two intercepts. Let a be the x -intercept and b be the y -intercept of a line L . We see from the view of division point, the line shall be formed by writing

$$((1-t)a, tb)$$

for all the points on L , where $-\infty < t < \infty$. We could derive that

$$\frac{x}{a} + \frac{y}{b} = 1$$

for any value of t .

Theorem. Let L be having x -intercept a and y -intercept b . Then L could be modelled by

$$L : \frac{x}{a} + \frac{y}{b} = 1$$

Example. Suppose L has x -intercept 5 and y -intercept 5. Then L could be modified as

$$L : \frac{x}{5} + \frac{y}{5} = 1 \iff y = 5 - x$$

Example. Suppose L is a straight line with x -intercept k and no y -intercept. Then L could be modified as

$$L : \frac{x}{k} = 1 \iff x = k$$

Example. Suppose L is a straight line with y -intercept k and no x -intercept. Then L could be modified as

$$L : \frac{y}{k} = 1 \iff y = k$$

Example. Suppose L is a straight line completely lies on x -axis. This means x -intercept is undefined. Then L could be modified as

$$L : y = 0$$

Example. Suppose L is a straight line completely lies on y -axis. This means y -intercept is undefined. Then L could be modified as

$$L : x = 0$$

3.2.7 General form

In convenience to solving linear equations, we aim to set a form that is always constructive. We put our attention to what a general form is.

Recall that a straight line could be either a vertical line, an oblique line, or a horizontal line. We shall introduce a form that is generally true to write a straight line. This thought should eliminate the presence of slope. Perhaps the intercepts form provides a good start to write a general form. Indeed,

$$\frac{x}{a} + \frac{y}{b} = 1 \iff bx + ay = ab$$

gives us the form purely by multiplication and addition. This can so be rearranged such that

$$bx + ay = ab \iff Ax + By + C = 0$$

with $A = b$, $B = a$ and $C = -ab$. Of its elimination of slope, this is the desired form of a straight line.

Theorem. *Any straight line L can be written in the form of*

$$Ax + By + C = 0$$

where A, B, C are constants. Furthermore, the slope of L is determined $-\frac{A}{B}$, the x -intercept $-\frac{C}{A}$ and y -intercept $-\frac{C}{B}$.

Proof. The first part of the proof is to determine the connection between different forms of equation of straight line. Indeed, the point-slope form connects with two-point form, while the point-slope form also connects with slope intercept form. The general form comes from intercepts form, so it remains to show the connection between slope intercept form and general form. In fact,

$$\begin{aligned} y = \mathfrak{M}x + c &\iff \mathfrak{M}x - y + c = 0 \\ &\iff Ax + By + C = 0 \end{aligned}$$

shows the connection between slope-intercept form and general form. This completes the first part.

The second part of the proof should show the "furthermore" implication. Consider the definitions of x -intercept and y -intercept, we will see if x_0 denotes the x -intercept, then $Ax_0 + B(0) + C = 0$ implies $x_0 = -\frac{C}{A}$; Similarly, denoting y_0 the y -intercept, we have $y_0 = -\frac{C}{B}$. For the slope we take the provided intercepts and write

$$\mathfrak{M} = \frac{y_0 - 0}{0 - x_0} = -\frac{A}{B}$$

concludes the proof. □

It is the best to write the general form with integers instead of fractions or decimal numbers, since it is more convenient and intuitive to solve equations using integers. In general, if we have to handle irrational numbers, we prefer to write everything without quotient and shall be simplest.

Example. *It is better to write $x + 2y + 3 = 0$ instead of $\frac{1}{2}x + y + \frac{3}{2} = 0$.*

Example. *If we need to handle irrationals, say $\frac{\sqrt{3}}{2} + \sqrt{5}y + \frac{3}{\sqrt{7}} = 0$, it is more desired to write $\sqrt{21}x + 2\sqrt{35}y + 6 = 0$.*

Essential Practice 3.2.7.1. *Find the straight line in general form with given conditions.*

1. *Of slope 2 and passes through $(1, 2)$.*
2. *Of slope $-\frac{2}{3}$ and passes through $(1, -1)$.*
3. *Passes through $(1, 3)$ and $(-2, 9)$.*
4. *Passes through $(1, 0)$ and $(-2, 0)$.*
5. *Of slope 1 and y-intercept 3.*
6. *Of slope $\frac{5}{7}$ and y-intercept $\frac{2}{3}$.*
7. *x-intercept 1 and y-intercept -2 .*
8. *x-intercept $-\frac{1}{2}$ and y-intercept $\frac{1}{3}$.*

Essential Practice 3.2.7.2. *Determine the slope, the x-intercept and the y-intercept for the following straight lines.*

1. $x + y + 1 = 0$.
2. $3x + 4y + 9 = 0$.
3. $2x = 4$.
4. $7y = 1$.

3.3 Rotational geometry

So far, we have determined the slope of a straight line which relates to the direction of procession very much, and we aim to discuss more about the real-world application of slope of straight line. We will take inspiration from polar coordinate and circular trigonometry.

3.3.1 Angle of elevation from horizon**3.3.2 Angle of depression from horizon****3.3.3 Angle between two arbitrary lines****3.3.4 Rotation of points and lines****3.4 Relation between lines****3.4.1 Point of intersection****3.4.2 Parallel lines****3.4.3 Perpendicular lines****3.4.4 Number of intersections****3.4.5 Angle bisector****3.5 Additional content: Point-line distance**

By definition, the distance between two points on a \mathbb{R}^2 plane is

Definition 3.7 (Distance between two points). *Let $A(x_1, y_1)$ and $B(x_2, y_2)$ be two points on \mathbb{R}^2 -plane. The distance between A and B is computed by the formula*

$$\text{dist}(A, B) := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Now let $L : ax + by + c = 0$ be a straight line and $P(x_0, y_0)$ be a point on \mathbb{R}^2 -plane. Note that with the following axiom

Axiom. *The distance between a point P and a line L is defined by the shortest distance between P and a point on L*

$$\text{dist}(P, L) := \inf\{\text{dist}(P, Q) : Q \in L\}$$

we can choose the perpendicular displacement of P from L to define the distance. Using the line Γ perpendicular to L passing through P , we can find such Q by following computation:

$$\begin{cases} L : ax + by + c = 0 \\ \Gamma : bx - ay + (ay_0 - bx_0) = 0 \end{cases} \implies Q\left(\frac{b^2x_0 - aby_0 - ac}{a^2 + b^2}, \frac{a^2y_0 - abx_0 - bc}{a^2 + b^2}\right)$$

Therefore, the distance between P and Q is

$$\begin{aligned}
 \text{dist}(P, Q) &= \frac{1}{a^2 + b^2} \sqrt{[(a^2 + b^2)x_0 - (b^2x_0 - aby_0 - ac)]^2 + [(a^2 + b^2)y_0 - (a^2y_0 - abx_0 - bc)]^2} \\
 &= \frac{1}{a^2 + b^2} \sqrt{(a^2x_0 + aby_0 + ac)^2 + (b^2y_0 + abx_0 + bc)^2} \\
 &= \frac{\sqrt{a^2 + b^2} \sqrt{(ax_0 + by_0 + c)^2}}{a^2 + b^2} \\
 &= \frac{\sqrt{(ax_0 + by_0 + c)^2}}{\sqrt{a^2 + b^2}} \\
 &= \left| \frac{ax_0 + by_0 + c}{\sqrt{a^2 + b^2}} \right|
 \end{aligned}$$

3.6 Linear inequalities

3.6.1 One variable inequality

3.6.2 Two variable inequality

3.6.3 Linear programming

3.7 Challenging questions

4 Quadratic functions

4.1 Solving Quadratic Equations

4.1.1 The square-root method

4.1.2 General form and Completing the square method

4.1.3 The quadratic formula

4.2 Solvability of Quadratic Equations

4.2.1 Discriminant: the factor affecting solvability

4.2.2 Application of solvability

4.3 Relation between coefficients and roots

4.3.1 What does it mean by a root?

4.3.2 The Vieta formula

4.3.3 Forming quadratic equations with given roots

4.4 Vertex form of a quadratic function

4.4.1 Obtaining vertex form using Transformation

4.4.2 Relation between vertex form and general form

4.4.3 Opening direction

4.4.4 The axis of symmetry

4.4.5 The extremum of quadratic functions

4.5 Quadratic inequalities

4.5.1 Boolean algebra

4.5.2 Solving quadratic inequalities

4.6 Challenging questions

5 Polynomial functions

5.1 Arithmetic rules for polynomials

5.1.1 Addition, Subtraction and Multiplication

5.1.2 Division

5.2 Divisibility of polynomials

5.2.1 Division algorithm

5.2.2 Remainder Theorem

5.2.3 Factor Theorem and General Vieta formula

5.3 G.C.D. and L.C.M.

5.3.1 Greatest Common Divisor

5.3.2 Least Common Multiple

5.4 Rational Function

5.4.1 Simplification of Rational Functions

5.4.2 Extended: Partial Fractions

5.5 Positional notation

5.5.1 Binary, Octal, Decimal and Hexadecimal

5.5.2 Conversion between different bases

5.6 Challenging questions

6 Exponential and Logarithmic functions

The evolution of computational concepts comes from shortening the notation of recursive action. One has the '+' notation for shortening process of counting, and '×' notation for shortening the process of adding. Here we have a new notation, call it **exponentiation** for shortening the process of multiplying.

6.1 Index notation

We first revisit the concept of self-multiplication.

6.1.1 From multiplication to power

6.1.2 negative integral power

6.1.3 Law of integral index

6.2 Rational power

6.2.1 The n-th root notation

6.2.2 Fractional power

6.2.3 Law of fractional index

6.3 Exponential functions

6.3.1 Law of Exponential algebra

6.3.2 Seeing surds as fractional exponents

6.3.3 Solving equations using exponentiation method

6.4 Logarithmic functions

6.4.1 What is Logarithm?

6.4.2 Law of Logarithmic algebra

6.4.3 Solving equations using Logarithmic method

6.5 Challenging questions

7 Sequence as a function of natural numbers

7.1 What is a sequence?

7.2 Arithmetic sequence

7.2.1 General form

7.2.2 Summation of arithmetic sequence

7.3 Geometric sequence

7.3.1 General form

7.3.2 Summation on Geometric sequence

7.3.3 Infinite sum of Geometric sequence

7.4 Additional content: Arithmetic-Geometric sequence

7.5 Challenging questions

8 Probability functions

8.1 Counting Principle

Let's introduce one notation for sum and one notation for product. We will see their convenience during conceptualisation.

Definition 8.1 (Summation notation). *We use the symbol sigma Σ to emphasis the process of addition. Usually we define the addition process using indexing format. The notation*

$$\sum_{i=1}^n x_i$$

means adding up from x_1 to x_n . In other words, it is saying

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n$$

Example. $\sum_{i=1}^3 1 = 1 + 1 + 1 = 3.$

Example. $\sum_{i=1}^4 i = 1 + 2 + 3 + 4 = 10.$

Example. $\sum_{i=2}^5 \frac{i^2}{2} = 27.$

Example. $\sum_{i=2}^5 \frac{i^2}{2} + 2 = 29.$

Example. $\sum_{i=2}^5 (\frac{i^2}{2} + 2) = 35.$

Remark. *The summation notation can usually be a misconception to many students, but bear in mind that the sum is computed by adding up terms.*

Theorem (Properties of summation notation). *The following properties holds for summation notation:*

1. $\sum_{i=1}^n a_i + \sum_{i=1}^n b_i = \sum_{i=1}^n (a_i + b_i);$

2. $c \sum_{i=1}^n a_i = \sum_{i=1}^n ca_i;$

$$3. \left(\sum_{i=1}^n a_i \right) \left(\sum_{j=1}^m b_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j;$$

$$4. \sum_{i=1}^n \sum_{j=1}^m a_i b_j = \sum_{j=1}^m \sum_{i=1}^n a_i b_j.$$

Definition 8.2 (Product notation). *We use the symbol \prod to emphasis the process of multiplication. Usually we define the multiplication process using indexing format. The notation*

$$\prod_{i=1}^n x_i$$

means multiplying up from x_1 to x_n . In other words, it is saying

$$\prod_{i=1}^n x_i = x_1 x_2 \cdots x_n$$

Example. $\prod_{i=1}^3 1 = 1 \cdot 1 \cdot 1 = 1.$

Example. $\prod_{i=1}^4 2 = 2 \cdot 2 \cdot 2 \cdot 2 = 16.$

Example. $\prod_{i=1}^4 i = 1 \cdot 2 \cdot 3 \cdot 4 = 24.$

8.1.1 Addition rule

The addition rule starts by finding the total number of things. We count by doing adding values.

Definition 8.3 (Addition rule). *Let A and B be two collections with m and n objects respectively. Then the total number of objects in A and in B is $m + n$.*

Example. *Let Alice has 3 apples and Bob has 4 apples. Then they have a total of 7 apples.*

Definition 8.4 (General version of addition rule). *Let A_1, A_2, \dots, A_n be n collections with x_1, x_2, \dots, x_n objects respectively. Then the total number of objects among all these collections is $\sum_{i=1}^n x_i$.*

Example. *Let Alice has 2 apples, Bob has 5 apples and Chris has 7 apples. Then they have a total of 14 apples.*

8.1.2 Multiplication rule

The multiplication rule comes from finding the total number of combinations. We count by spanning a table.

Definition 8.5 (Multiplication rule). *Let A and B be two collections with m and n objects respectively. Then the total number of combining objects in A and in B is mn .*

	1	2	\dots	m
1	(1,1)	(2,1)	\dots	(m ,1)
2	(1,2)	(2,2)	\dots	(m ,2)
\vdots	\vdots	\vdots	\ddots	\vdots
n	(1, n)	(2, n)	\dots	(m , n)

Example. *Let Alice has 3 different T-shirts and Bob has 4 different pants. Then the number of costume combination is 12.*

Example. *Suppose there are 15 boys and 18 girls in a class. Then the number of possible couples in this class is 270.*

Definition 8.6 (General version of multiplication rule). *Let A_1, A_2, \dots, A_n be n collections with x_1, x_2, \dots, x_n objects respectively. Then the total number of n -tuples among all these collections is $\prod_{i=1}^n x_i$.*

Example. *Let class A has 2 class members, class A has 5 class members and class C has 7 class members. To form a committee by selecting one students from each class, 70 different committee can be formed.*

8.2 Pigeonhole Principle

One should identify the constraints to manage how distributions are formed, and so does probability. We will see one question to facilitate the mind on distribution.

Let us see we have for instance 4 pigeonholes. If we have 1 pigeon to choose where to live in, it will fill one of the 4 holes; if we have 2 pigeons, 2 out of 4 will be filled; if 3, then 3 out of 4 will be filled; and 4 out of 4 for the case of 4. However, what happened when 5 pigeons are there?

Assume they do not fight to each other so that no dead bodies are there, and they agreed to live together in some pigeonhole. Alright, no extra space for building new holes either. Then we could see when 4 holes are all filled with 1 pigeon, 1 is left behind to choose which pigeonhole to

live in. In this stage, we observe no matter which hole it chose, **at least one of the 4 holes have to be with 2 or more pigeons**. We then extend this concept to a general case of $n + 1$ pigeons with n pigeonholes. This is called the **Pigeonhole principle**.

8.2.1 Fundamental version

Theorem (Baby pigeon). *If there are n containers with $n + 1$ objects, then at least one of the containers must have 2 or more objects in it.*

Example. *Consider 31 books are going to be distributed to 30 students, then at least one of the students has 2 or more books.*

Essential Practice 8.2.1.1. *To ensure 10 children are all having at least 1 slice of pizza, at least how many slices should the pizza be sliced into?*

8.2.2 General version

A general version of pigeonhole principle can be introduced from the perspective of division.

Theorem (Adult pigeon). *If there are n containers with $mn + 1$ objects, then at least one of the containers must have $m + 1$ or more objects in it.*

Proof. We may consider $mn + 1$ as division algorithm and decide the worst situation. Indeed, it is the worst when every container is containing the least amount of objects in it. This is equivalent to saying the distribution is uniform, i.e. averagely distributed. We then conclude after the worst case, every container is containing m objects with 1 left behind. The principle is deduced. \square

Corollary (Pigeonhole principle). *Suppose t objects are going to be distributed to n containers, where $mn + 1 \leq t \leq (m + 1)n$. Then at least one of the containers must be having $m + 1$ objects.*

Example. *Suppose 99 balls are stored in the storage room. If there are 5 basket ball players, then at least one of the players could have 20 balls.*

This is why we usually use division to measure the worst case of distribution.

8.3 Counting functions

Based on the rules of counting, we establish the following method of counting measure. Our perspective of counting starts from permutation.

8.3.1 Factorial

Suppose we are arranging some people in a queue. Let say there are 5 people to be arranged. Then how many different queues could be formed?

Assume 5 placement should be filled by 5 people, then we have no duplicated choice for 2 different placement. We call the selection be without replacement.

For the first placement, we have 5 choice; for the second placement, as we have already select 1 person out of 5, we can only choose from the remaining 4 person, so the second placement has 4 choices; for the third placement, we have a remaining of 3 person, so 3 choices for the third placement; so on and so forth, the fourth and the fifth placement has 2 and 1 choice respectively. As long as a queue can be seen as an ordered combination, we shall apply multiplication rule to compute the number of permutation:

$$5 \times 4 \times 3 \times 2 \times 1 = 120$$

To shorten the multiplication (Yes, we are lazy), we define the following notation:

Definition 8.7 (Factorial). *Let n be a non-negative integer, then the **factorial** notation is defined as recursive definition:*

$$n! = n(n-1)!, 0! = 1$$

In some sense, if we see the factorial notation as some function, we have

$$n! := \prod_{i=1}^n i$$

However, how was $0!$ be defined?

Let's consider the reverse view of factorial. We note that factorial is built from multiplication, and so can be extended using division. Consider the definition

$$n! = n(n-1)! \implies (n-1)! = \frac{n!}{n}$$

which means $0!$ can be computed as

$$0! = \frac{1!}{1} = 1$$

Example. *For a queue of 7 people, we have $7! = 5040$ different permutations.*

Example. *For a queue of 0 people, we have $0! = 1$ permutation only. That is, arranging nothing. It is so defined because this is also a valid situation to be counted!*

Essential Practice 8.3.1.1. *Compute the number of distinct queues can be formed if there are (a) 6 people; (b) 9 people; (c) 12 people.*

8.3.2 Permutation

Somehow, we have another problem of arranging not all people but part of them. We shall now consider if **not all people are selected in the arrangement**.

Suppose now we have 6 people in the selection zone, and we need to arrange any 3 of them in order. It would be easier to think of counting championships, but never mind. Similar to the part of factorial, we just need to reduce the outstanding arrangement. Therefore, we have in total

$$6 \times 5 \times 4 = 120$$

numbers of arrangement for 3 out of 6 people. It is interesting to do an algebra trick: we complete the factorials by filling up the gaps so that

$$6 \times 5 \times 4 = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1} = \frac{6!}{3!} = \frac{6!}{(6-3)!}$$

The reason to write the last equality is we shall observe and fulfill the generality of the writings. Consider we need to arrange 4 out of 6, the computation becomes

$$6 \times 5 \times 4 \times 3 = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = \frac{6!}{2!} = \frac{6!}{(6-4)!}$$

Therefore, in general, we find the permutation of r out of n objects is equal to $\frac{n!}{(n-r)!}$, and one denote the concept by taking P as the notation standing for permutation. The formal definition says:

Definition 8.8 (Permutation). *The amount of permutation for arranging r objects out of n objects is defined as*

$$P_r^n := \frac{n!}{(n-r)!}$$

Example. *There are $9 \times P_4^9$ 5-digit numbers without repeated digits.*

Essential Practice 8.3.2.1. *How many strings (alphabet permutations) of length 6 can be formed if no alphabet is used repeatedly?*

8.3.3 Combination

8.4 Fundamental Probability

8.4.1 Fair dice intuition

8.4.2 Outcomes/Events as sets

8.4.3 Probability in terms of outcomes

8.5 Probability of independent events

8.5.1 Independent events

8.5.2 Venn-diagram

8.5.3 Supplementary event

8.5.4 Rules of coherence

8.6 Probability of dependent events

8.6.1 Independence decomposition

8.6.2 Conditional probability

8.7 Discrete random variables

8.7.1 Finite domain and infinite domain

8.7.2 Mean and expected value

8.7.3 Standard deviation and variance

8.8 Distributions

8.8.1 Binomial distribution

8.8.2 Geometric distribution

8.8.3 Poisson distribution

8.8.4 Normal distribution and Standard distribution

8.9 Sampling

8.9.1 Point sampling

8.9.2 Central limit theorem

8.9.3 Confidence interval

8.10 Challenging questions