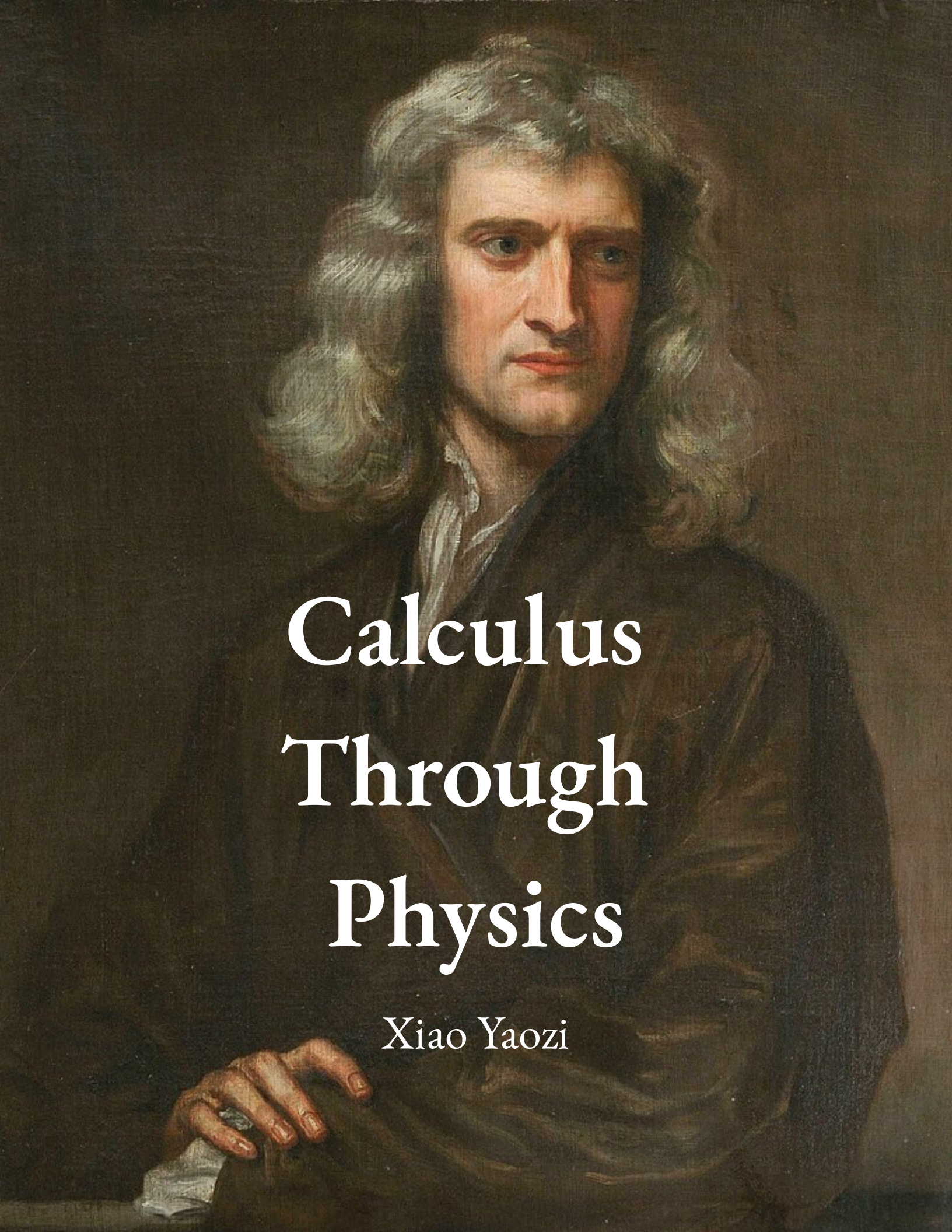

Calculus Through Physics

Xiao Yaozi

Xiao Yaozi

Calculus through Physics

Copyright notice

To my parents, who have supported me in everything I do.

Preface

This book is a fairly standard calculus textbook which covers the most of the standard Calculus I to IV material. What is *different* is that this book attacks the problem of pedagogy of calculus by bringing in *physics*, more specifically, classical mechanics. To be clear, this is not a physics textbook in any capacity, it is just that I believe that by utilizing classical mechanics, readers are able to tap into their intuition as forces and classical mechanics are quite intuitive. Therefore, I hope that calculus will be easier to learn.

What is unfortunate is that calculus has gained a "notorious" reputation among high school students as being a hard to learn, esoteric subject. Those who have learned calculus seemed to have entered a sort of cult. While it is true that calculus is quite a big departure from precalculus, it is by no means impossible to learn and if the reader puts in the work, I can almost assure you that it is worth it (that is if you are pursuing any STEM subject). Another way to emphasize the importance of calculus, is by stating the seemingly sad fact that one cannot hope to run away from calculus, be it physics, chemistry or even biology. However, this is because calculus naturally arises from the problems in these subjects, not because scientists are evil and want to gate keep.

I must make it absolutely clear, though, that the reader must have a basic knowledge of physics (forces, velocity etc.) to understand this book. In addition, the first part on precalculus does not use classical mechanics to frame the subject matter and is very brief. I expect the reader to ensure that they have a solid understanding of the subject matter. Problems in the text are sometimes accompanied by numbers which indicate the difficulty of a problem:

1. **(1)** represents an essential problem the reader should attempt, it should not be very difficult.
2. **(2)** represents a slightly difficult problem that is good for the reader to attempt, and with some careful thinking, it should be doable.
3. **(3)** represents a very difficult problem, usually requiring some extra knowledge to attempt.

No numbers mean that a problem is trivial and can be done if the reader needs more practice.

Another point to note about the book is that for mathematical programming, I shall be using SageMath, a free and open source mathematics software system. Please download it from its official website to follow along certain parts of the book.

Contents

Part I
Precalculus

Chapter 1
Functions

1.1 Function basics

Before we start on the calculus topics, this part of the book will acquaint you with the mathematics you would need to understand in order to read this book. If you are familiar with the these topics, you may skip to the Part II on Calculus I. For the rest of you, the pace of this part will be very fast, it is recommended that you read up on the concepts you do not understand very well. We will start with the idea of a *function*. By now, I am sure that many of you may have seen equations like these:

$$x^2 - 5x + 6 = 0 \tag{1.1}$$

$$\sin(x) = 0.5, 0 \leq x \leq \frac{\pi}{2} \tag{1.2}$$

$$y = mx + c \tag{1.3}$$

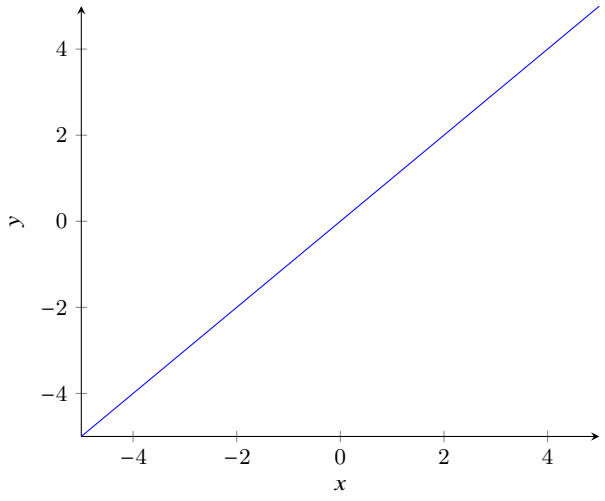

Fig. 1.1 Graph of $y = x$.

Out of the three equations, notice that Equation (1.3) is special. There is not some fixed x values that you must plug in. Instead, the y value varies with the x value. Given a x value, the equation "produces" a y value. This brings us to the idea of a function. A function is simply something that takes in a value, and "produces" another value based on the input. You may have came across functions before, such as the trigonometric functions like $\sin$ and $\cos$. The value of $\sin(x)$ varies with different values of x. Of course, it is time for me to give you the definition of a function.

Definition 1.1 A **function** f is a mathematical object that assigns its inputs, or *domain*, to its outputs, or *range*. The notation $f : A \to B$ means that the function has a domain A and range B [1]. The notation $f(a)$ where $a \in A$ denotes the element of B to which a is assigned to.

Example 1.1 Here are some examples of functions:

$$f(x) = x + 1 \tag{1.4}$$

$$g(x) = \sin(x) \tag{1.5}$$

$$h(x) = x^2 - 5x + 6 \tag{1.6}$$

To evaluate the output of a function, you simply substitute the input to the expression that defines the function. Notice that for Equation (1.4), it is simply the line $y = x + 1$. Each value of $f(x)$ for some $x \in \mathbb{R}$ is just y. Take $x = 1$ for example, $f(1) = 1 + 1 = 2$ and for the line $y = x + 1$, notice the y value here is also 2. The function assigns each x value to $x + 1$. Hence, in this case, $f : \mathbb{R} \to \mathbb{R}$. For Equation (1.5) , the function takes any real x value and maps it to $\sin(x)$. As a result, its domain would be $\mathbb{R}$ and its range would be $[-1, 1]$. Again, it is just like the graph $y = \sin(x)$. Lastly, the quadratic in Equation (1.6) is another real function(i.e. it has a real domain and range). Think about what happens when $h(x) = 0$, what value of x must you plug in to make $h(x) = 0$. Well, isn't it just the roots of the quadratic. Hence, the roots of a function are just the values x such that $h(x) = 0$. In this case, it would be 2 and 3. Thus, $h(2) = h(3) = 0$.

Actually, functions can be more than the usual $f(x) =$ something. Let's say you want to have a function that outputs "1" if the input is rational and "0" if it is not, you can define it as such: [2]

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases} \tag{1.7}$$

Here, $f(\frac{1}{2})$ would be 1 but $f(\sqrt{2})$ would be 0. (You would probably get a migraine from graphing it.) This is an example of having conditionals in a function. Lastly, functions may take more than one input and produce more than one output, examples include:

$$f(x, y) = x^2 + y^2 \tag{1.8}$$

$$g(t) = (\cos(t), \sin(t)) \tag{1.9}$$

Equation (1.8)[3] is an example of a *multivariate* function, in other words, it takes in multiple inputs. If $x = 2$ and $y = 2$, then $f(2, 2) = 8$. On the other hand, Equation (1.9) does not have a single output. Instead,

[1] Note that A and B are sets.

[2] This is actually the famous Dirichlet function that shatters certain intuitions on functions and will be discussed in later chapters.

[3] As you will see in Part IV of the book, this is the *parametric* equation of a circle.

it has a tuple [1] as its output. Evaluating it at $t = \pi$, notice that the output is $(-1, 0)$. What is interesting is that if you trace out its outputs on a plane for all $t \in \mathbb{R}$, you will get Figure 1.2.

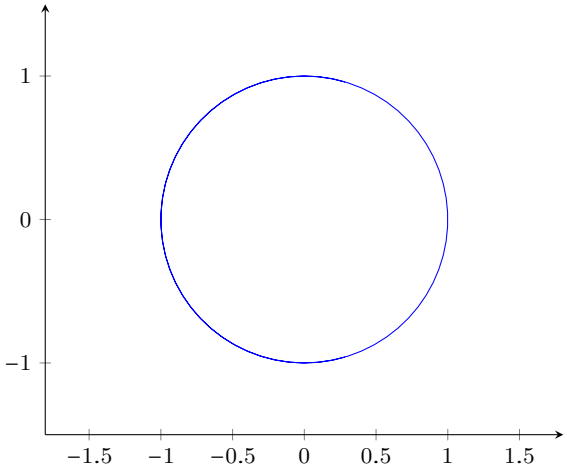

Fig. 1.2 Plot of Equation (1.9) for all $t \in \mathbb{R}$

Simple pendulum

In physics, we often deal with functions which enables us to describe many things whose behaviour is dependent on certain inputs. Take a pendulum for example. We can describe the angle of the pendulum to the normal line using the following formula[2]:

$$\theta(t) = \theta_0 \cos(\sqrt{\frac{g}{l}}t) \tag{1.10}$$

Here, $\theta(t)$ is the angle of the pendulum at time t, θ_0 is the initial angle of the pendulum, T is the period of the pendulum, l is the length of the rod. (g is the gravitational field strength.) While this function may look sightly complicated, we can deduce a lot about it without computing much. Firstly, at $t = 0$, $\theta(0) = \theta_0$, which makes sense because at the start, the pendulum is at its initial angle.

The periodic nature of the pendulum is also expressed in the cosine part of the function. Moreover, since cos can only attain values between -1 and 1, the pendulum is guaranteed to be oscillating between $-\theta_0$ and θ_0. This satisfies the conservation of energy because if this was not guaranteed and no energy is introduced into the system, energy must have been created or destroyed.

Challenge 1. *This being said, in the real world, is energy lost to the surroundings? How could energy be lost to the surroundings? How would this affect the motion of the pendulum?*

[1] A tuple is basically a set where order matters. You can treat it as if it is a coordinate.

[2] While not completely accurate, this formula is good enough for an example.

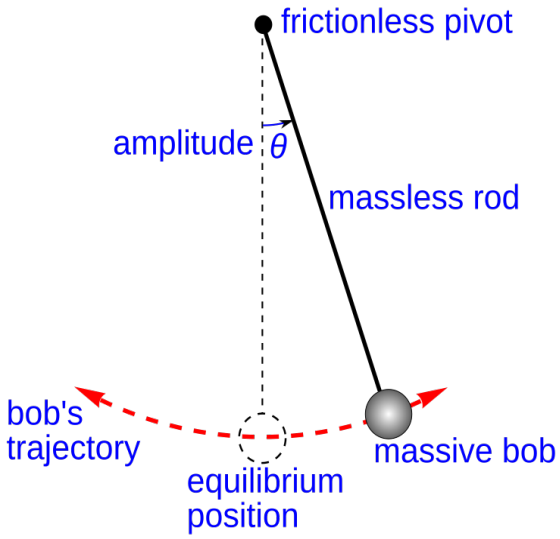

Fig. 1.3 Figure of pendulum.

Example 1.2 Let us calculate the period of the pendulum. We know the formula for the angle of the pendulum for any time t. The period of a function $\cos(ax)$ is simply $\frac{2\pi}{a}$, hence the period of the pendulum is $2\pi\sqrt{\frac{l}{g}}$. This results shows us that the period of the pendulum is **independent** of the initial angle. This property is called "isochronism" that Galileo discovered.

In the real world, we often add a *damping factor* $d(t)$ to $\theta(t)$. A damped pendulum has the function describing its angle as $\theta_d(t) = \theta(t)d(t)$. The damping factor is to reduce the amplitude of the pendulum as time passes, mainly because air resistance opposes the motion of the pendulum.

Challenge 2. *From the following damping factor $d(t)$, choose the most suitable one (you may want to use a graphing calculator to assist you):*

$$e^t \qquad e^{-t} \qquad -t \qquad a, a \in \mathbb{R}$$

Problem 1.1 Given that $x = 2$ and $y = \pi$, evaluate $f(x)$ for:

a) $f(x) = x^3 + 1$
b) $f(x) = 2^x$
c) $f(x) = \begin{cases} x^4 + 1 & \text{if } x < 2 \\ 3^x & \text{otherwise} \end{cases}$
d) $f(x, y) = \frac{\sin(y)}{x}$

Problem 1.2 Given the following functions, write down $f : A \to B$ where A is the domain and B is the range(do not consider complex inputs and outputs, we shall be ignorant to it for the time being to simplify things):

a) $f(x) = x$ [1]
b) $f(x) = \frac{1}{x}$
c) $f(x) = \sqrt{x}$

Problem 1.3 Find the root(s) of the following functions:

a) $f(x) = x^2$
b) $f(x) = 2^x - 1$
c) $f(x) = \sin(x)$

Problem 1.4 (1) A *recursive* function is one where the function uses itself in its definition. For example, the Fibonacci sequence can be defined like this for any Fibonacci number n:

$$f(n) = \begin{cases} f(n-1) + f(n-2) & n > 1 \\ 1 & n = 1 \\ 0 & n = 0 \end{cases}$$

a) Compute $f(3)$ explicitly using the definition of f.
b) Define the factorial of a positive integer in a similar way using recursion, define it as $g(a)$.
c) Sketch the graph of $g(a)$.

Problem 1.5 (2) Find $f : \mathbb{R} \setminus \{0, 1\} \to \mathbb{R}$ such that $f(x) + f(\frac{x-1}{x}) = 1 + x$ for any $x \in \mathbb{R} \setminus \{0, 1\}$.

[1] This function has a special name, the **identity** function. Notice how its input and output are the same.

1.2 Function composition and inverses

In this chapter, we will look at composition of functions and inverse functions. You can use the normal arithmetic operations on functions like how you would to numbers:

Example 1.3 If $x = 2$ and $f(x) = x^2 + 4$:

$$f(x) + f(x) = 2^2 + 4 + 2^2 + 4 = 16$$

$$\frac{f(x)}{2} = \frac{2^2 + 4}{2} = 4$$

What is more interesting, perhaps, is taking the *output* of a function and use it as the *input* to another function. You may also think of it as "piping" one function into another. This is called *function composition*.

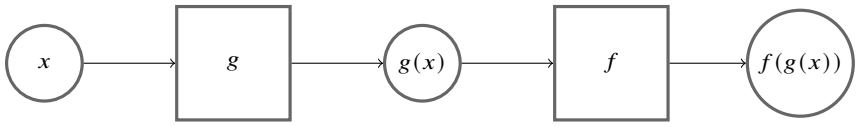

Fig. 1.4 Visual representation of function composition.[1]

Another notation for composing functions together (which we will not really use except for this chapter) is as such:

$$f(g(x)) = (f \circ g)(x)$$

When you compose the same function n-times, the notation for it is[2]:

$$f^n(x) = \underbrace{f(f(\ldots f(x)))}_{n-times}$$

Example 1.4 Here are some examples of composing functions together: If $x = 3$ and $f(x) = \pi x$, $g(x) = \sin(x) + 1$:

$$g(f(x)) = \sin(\pi \times 3) + 1 = 1 \tag{1.11}$$

$$f^3(x) = \pi^3 x \tag{1.12}$$

There is nothing much here, really, just evaluate the function from the innermost function to the outermost.

Theorem 1.1 *Given functions $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$, prove that the composition of the three functions is associative, in other words:*

$$h \circ (g \circ f(x)) = (h \circ g) \circ f(x)$$

Proof. Let $x \in X$, we have:

[1] Note that the rectangles are functions and circles are values.

[2] However, this does not apply to trigonometric functions. $\sin^2(x)$ is $\sin(x) \times \sin(x)$ instead of $\sin(\sin(x))$. Annoying mathematical notation strikes again. What is even better, is that $f^{-1}(x)$ denotes the inverse of f.

$$h \circ (g \circ f(x)) = h(g \circ f(x)) = h(g(f(x))) = h \circ g(f(x)) = (h \circ g)f(x)$$

This proves the associativity of function composition. □

Remark 1.1 This is just a fun little exercise for the reader and ultimately is not that important to the entire book, just a good identity to keep in the back of your mind.

Next, we will delve into inverse functions. Let us start with one I believe you may have seen before, look at the following figure:

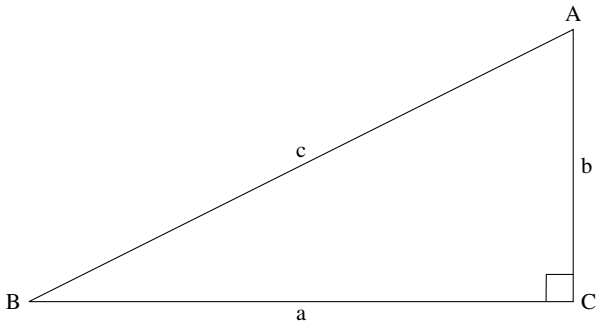

Fig. 1.5 Figure of right triangle ABC.

If I ask you what is $\arcsin(B)$, you may say it is $\frac{b}{c}$. But, what is $\arcsin(\sin(B))$, well, that is just B, isn't it. Hence, the $\arcsin$ sort of "cancels" the $\sin$. We call $\arcsin$ the *inverse function* or just inverse of $\sin$.[1] Generally, given a function $f : X \to Y$, f^{-1} denotes the inverse function that satisfies:

$$f \circ f^{-1}(y) = y, y \in Y \text{ and } f^{-1} \circ f(x) = x, x \in X \tag{1.13}$$

To find the inverse function of a function f, we have to find the function that maps the outputs of the f back to its inputs. To solve algebraically, we set $f(x)$ for some arbitrary x to y, and make x the subject of the equation.

Example 1.5 Given that $f(x) = \frac{x+7}{3}$:

$$\begin{aligned} f(x) &= y \\ \frac{x+7}{3} &= y \\ x + 7 &= 3y \\ x &= 3y - 7 \end{aligned}$$

Hence, the inverse function is $f^{-1}(y) = 3y - 7$.

Example 1.6 Some common examples of inverse functions:

[1] We will use $\arcsin$ as the inverse of $\sin$ instead of $\sin^{-1}$ as it is clearer.

1. Trigonometry:

$$\arcsin \qquad \arccos \qquad \arctan \qquad \dots$$

2. Logarithms (Inverses of exponential functions):

$$\ln \qquad \log_b$$

Problem 1.6 Given $f(x) = \frac{x^2+3}{2}$ and $g(x) = \tan(x) + 1$, evaluate:

a) $f(g(\frac{\pi}{4}))$
b) $f^3(1)$
c) $f(g(\arcsin(0)))$

Problem 1.7 Given that $h(x) = (x-7)^2$, $x \geq 7$, find the inverse of h. Hence, evaluate:

a) $h^{-1}(1)$
b) $h(h^{-1}(69))$ *As Peter Lorre would have said it, "Do it ze kveek vay, Johnny".*

Problem 1.8 (1) The *dirac delta function* is defined as

$$\delta(x) = \begin{cases} \infty & x = 0 \\ 0 & \textit{otherwise} \end{cases} \tag{1.14}$$

It also has an area of 1 "underneath" it.

a) Show that $f(x)\delta(x-a) = f(a)\delta(x-a)$.
b) Hence, simplify $(3x^2 - 2x - 1)\delta(x-3)$.

Problem 1.9 (1) Given that the function $f(x)$ is defined as $f(x) > f(x-1) + f(x-2)$, and when $x < 3$, $f(x) = x$, show that $f(20) > 1000$. *(Adapted from the 2024 Chinese College Entrance Examination)*

1.3 Injection, surjection and bijection

You may have noticed in the previous exercise that question 2 is particularly interesting. There is a restriction on the domain of $h(x)$. Why is that? This is what this chapter is all about. However, before we explain that, I will go over some definitions of properties of functions. Firstly, we have *injection*.

Definition 1.2 An **injective** function, f, or a *one-to-one* function, is a function such that every distinct output of the function corresponds to only *one* distinct input, that is [1]:

$$f(x_1) = f(x_2) \text{ implies } x_1 = x_2$$

Of course, this explanation might not have been very clear. Here is a visualization:

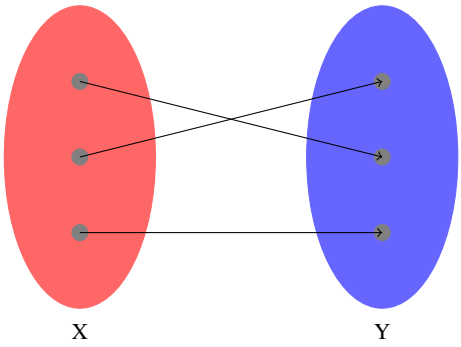

Fig. 1.6 Illustration of an injective function with domain X and range Y.

As you can see in Figure 1.6, every output in Y of the function is associated with only one input in X, hence the function is *injective*. (Kind of like how marriages *should* be.) To prove that a function f is injective, you can either show that if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$, or if $f(x_1) = f(x_2)$, then $x_1 = x_2$.

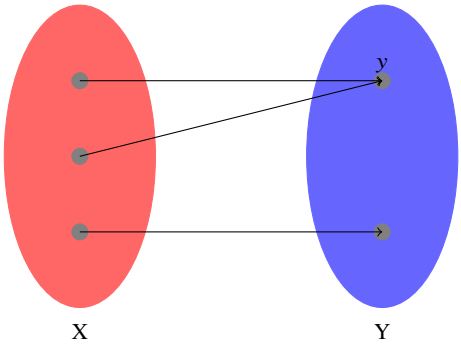

Fig. 1.7 Illustration of a non-injective function with domain X and range Y.

[1] By contraposition, $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$. The contraposition of a statement "p then q" is "not q then p".

Here, the function is *not* injective. As you can see in Figure 1.7, the output y is associated with more than one input in X. With this in mind, it is time to talk about *surjection*.

Definition 1.3 A **surjective** function f is a function such that for every output y, there exists at least one input x such that $f(x) = y$.

From Figure 1.6 and Figure 1.7, both functions are surjective as every output is associated with at least one input. Do take note that it does not matter how many inputs associate to one output, as long every output is associated with at least one input, the function is surjective. To prove that a function is surjective, you must show that there exists an element in the domain of the function for any arbitrary output in the range.

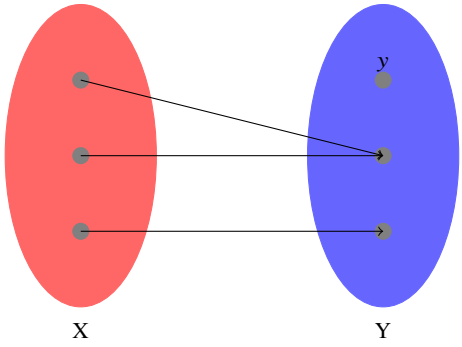

Fig. 1.8 Illustration of non-surjective function with domain X and range Y.

From Figure 1.8, we can see that y has no input associated with it. As a result, the function here is not surjective. Lastly, *bijective* functions can't be any easier.

Definition 1.4 A **bijective** function is a function that is both injective and surjective.

However, bijective functions have a superpower. A function is only bijective *if and only if* it has an inverse.[1]

Theorem 1.2 *A function, $f : A \to B$, has an inverse if and only if it is bijective.*

Proof. To begin such proofs where there is a "if and only if" clause, we will prove both cases separately. First, we claim that if the function has an inverse, then it is bijective. To prove this, we simply prove that f is injective and surjective. Then, we claim that if a function is bijective, then it has an inverse. With both proofs below, we show that f has an inverse if and only if it is a bijection. □

Lemma 1.1 *If a function, $f : A \to B$, has an inverse, then it is bijective.*

Proof. Suppose f^{-1} is the inverse of f. Notice that $f^{-1}(f(x))$ is injective as it is simply $f(x) = x$. Suppose $a, a' \in A$ and $f(a) = f(a')$, then $f^{-1}(f(a)) = f^{-1}(f(a')))$. Notice that by injectivity of $f^{-1}(f(x))$, $a = a'$. Hence, f is injective. Then, suppose $b \in B$, we wish to show that the function is surjective. Also, $f(f^{-1}(x))$ is surjective(it is the identity function). Let $f^{-1}(b) = x$ and $f(f^{-1}(b)) = c$ for some $c \in B$. By a simple substitution, $f(f^{-1}(b)) = f(x) = c$, thereby showing there exists a $x \in A$ for every $c \in B$. The two cases proves that f is bijective if it has an inverse. □

[1] P iff Q, or "P if and only if Q", means that "if P then Q" and "if Q then P".

Lemma 1.2 *If a function, $f : A \to B$, is a bijection, then it has an inverse.*

Proof. Let us define f^{-1} as $f^{-1} : B \to A$. Let $b \in B$. Since f is surjective, there exists $a \in A$ such that $f(a) = b$. Let $f^{-1}(b) = a$. Since f is injective, f^{-1} is well-defined, or a function that outputs the same value given the same input. Next, we will show that f^{-1} is indeed the inverse of f by showing the composition of the two is an identity function. Again, let $a \in A$ and $b = f(a)$ which implies, by definition, $f^{-1}(b) = a$. Composing the two functions, $f(f^{-1}(b)) = f(a) = b$, which shows that f^{-1} is indeed the inverse function. The other case of $f^{-1} \circ f$ is similar. Hence, if f is a bijection, then it has an inverse. □

Remark 1.2 This serves as a very good example showing how the reader should prove the injectivity, surjectivity, and bijectivity of functions. Also, this explains why in the previous exercise, the function $h(x) = (x-7)^2$ is bounded, only taking x values larger than 7. The function is not bijective in $\mathbb{R}$, as evident in the graph as it is not injective. Hence, the bounds are needed for h to have an inverse.

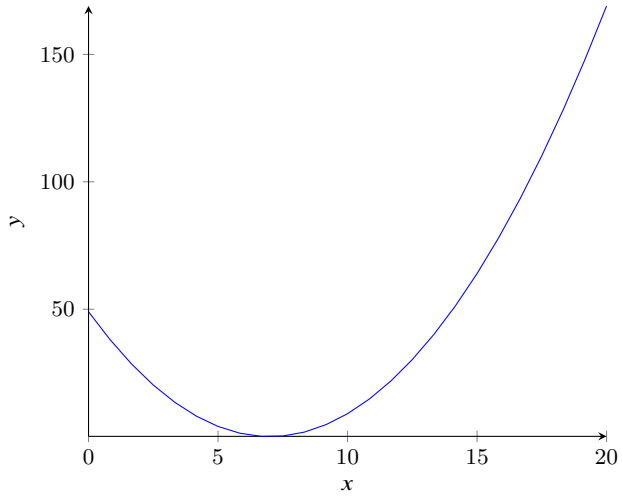

Fig. 1.9 Plot of $(x-7)^2$.

Problem 1.10 Determine if the following functions are injective or surjective or bijective in its domain:

a) $f(x) = \sqrt{x}, f : \mathbb{R}^+ \to \mathbb{R}^+$
b) $f(x) = \frac{1}{x}, x > 0$
c) $f(x) = (x-7)^2, x \geq 5$

Problem 1.11 (1) Prove that:

a) The composition of two injective functions is *injective*.
b) The inverse of a function is *injective*.
c) The inverse of a function is *surjective*.

Problem 1.12 (1) Given a quadratic function $q(x) = ax^2 + bx + c$, and $a \neq 0$, show that q cannot be bijective.

Further problems

Problem 1.13 (1) *(Polynomial remainder theorem)* Let $f(x)$ be a polynomial. If $f(a) = 0$, prove that $x - a$ is a factor of $f(x)$.

Problem 1.14 (1) We shall define an area function, $A(f, x_1, x_2)$ which gives us the area under a given function $f(x)$ [1] bound by $x_1 \leq x \leq x_2$.

a) Using an example, show that $A(f, x_1, x_2)$ cannot be bijective.
b) Also using an example, disprove that $A(fg, x_1, x_2) = A(f, x_1, x_2)A(g, x_1, x_2)$.
c) From the examples, verify geometrically that $A(f + g, x_1, x_2) = A(f, x_1, x_2) + A(g, x_1, x_2)$:
    a. $f(x) = x + 1$ and $g(x) = x + 2$
    b. $f(x) = x$ and $g(x) = 0$
    c. $f(x) = 1$ and $g(x) = 2$
d) If δ is the dirac delta function found in Problem 1.8, show that $A((x^2 + x + 1)\delta(x - 1), -\infty, \infty) = 3$. It is given that $A(cf(x), \ldots) = cA(f(x), \ldots)$ for some constant c.

Problem 1.15 (2) The *wavefunction* Ψ of a particle can be thought of as the quantum state of a particle. In this problem, you will solve for the wavefunction of an one-dimensional particle (that lies on the x-axis) in an infinite square well. It is given that $|\psi(x)|^2$ gives the probability of the particle existing at x and $|\Psi(x, t)|^2$ is the probability of a particle existing at x at time t after a measurement. (Hence, note that $\Psi(x) \neq 0$ and $\psi(x) \neq 0$. That would imply that the particle cannot be found anywhere.) The potential function (potential energy) of the infinite square well is

$$V(x) = \begin{cases} 0 & 0 \leq x \leq a \\ \infty & \textit{otherwise} \end{cases} \tag{1.15}$$

a) Can the particle exist outside of $0 \leq x \leq a$, or more specifically, what is $|\psi(x)|^2$ (and therefore $\psi(x)$) for $x < 0$ or $x > a$?
b) To ensure that ψ is continuous, $\psi(0)$ and $\psi(a)$ are equal to your answer of $\psi(x)$ in (a). Also, it is known that the form of general solution inside the well (to wit: $0 \leq x \leq a$) to ψ is $\psi(x) = A\sin(kx) + B\cos(kx)$. Show that $\psi(x) = A\sin(kx)$.
c) Using your answer in (b), solve for k. (Hint: k has multiple solutions, generalize all the solutions by writing down k_n for $n \in \mathbb{Z} \setminus \{0\}$.)
d) Show that we can discard all negative solutions of k_n, hence $n \in \mathbb{N}$.
e) If $E(n) = \frac{\hbar^2 k_n^2}{2m}$, find $E(1)$ or the energy of the *ground state* of the particle. $\hbar$ is the reduced plank's constant. What happens as n increases?
f) The relation between Ψ and ψ is $\Psi(x, t) = \psi(x)e^{-\frac{iE(n)t}{\hbar}}$. Find $|\Psi|^2$ for $t = 1$ and $t = 2$, what happens to the graph of $|\Psi|^2$ as time passes? Take A, m, a as positive real numbers and choose some natural n.

Problem 1.16 (3) If $f(x) = 2^x - x$, find x_1 such that $f(x_1) = 5$.

[1] If you are skeptical about the area function provides the area for any function, you are absolutely right. This property that a function can be integrated is called integrability. If you are interested, there are many types, such as the physicist favourite "square-integrability". However, the point of the problem remains.

1.4 An aside: A more rigorous definition of a function

In the previous section, we defined a function as a rule that assigns to each element in a set (the domain) exactly one element in another set (the range). While this intuitive definition is often sufficient, we can provide a more rigorous definition using the concept of ordered pairs and Cartesian products. But first, what even is an ordered pair?

Definition 1.5 An **ordered pair** is a collection of two objects where the order of the objects matters. We denote an ordered pair by (a, b), where a is the first element and b is the second element. Two ordered pairs (a, b) and (c, d) are considered equal if and only if $a = c$ and $b = d$.

Definition 1.6 Given two sets A and B, the **cartesian product** of A and B, denoted by $A \times B$, is the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$. Formally,

$$A \times B = \{(a, b) \mid a \in A, b \in B\}. \tag{1.16}$$

Example 1.7 Consider the sets $A = \{1, 2\}$ and $B = \{a, b\}$. The Cartesian product $A \times B$ is given by:

$$A \times B = \{(1, a), (1, b), (2, a), (2, b)\}. \tag{1.17}$$

Here, each ordered pair consists of an element from A and an element from B. Notice that the order matters: $(1, a)$ is different from $(a, 1)$. This is different from the usual sets where the order of elements does not matter. Next, consider something more complex, $\mathbb{Z} \times \mathbb{Z}$:

$$\mathbb{Z} \times \mathbb{Z} = \{(m, n) \mid m, n \in \mathbb{Z}\}. \tag{1.18}$$

This set contains all possible ordered pairs of integers. For example, $(1, 2), (3, 4), (-1, 0), (0, -3)$ are all elements of $\mathbb{Z} \times \mathbb{Z}$. This Cartesian product is often visualized as a grid or lattice in the coordinate plane, where each point on the grid corresponds to an ordered pair of integers.

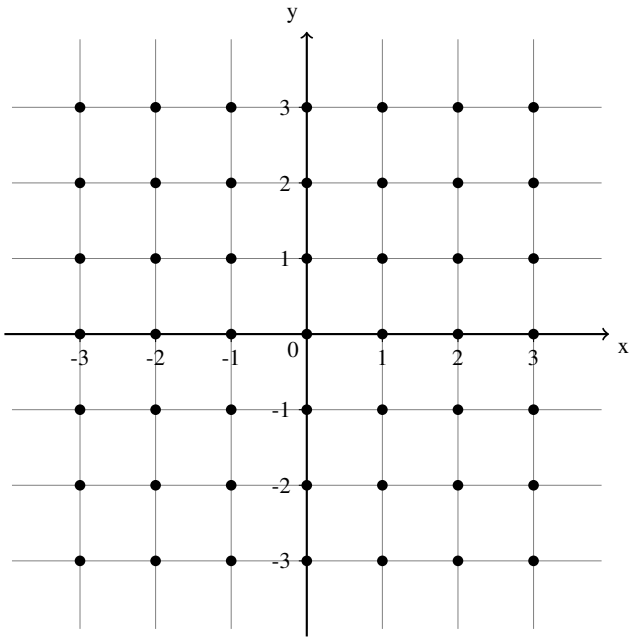

Fig. 1.10 A visualization of the Cartesian product $\mathbb{Z} \times \mathbb{Z}$. Each dot represents an ordered pair of integers.

But why are ordered pairs and Cartesian products relevant to functions? Well, we can represent a function as a set of ordered pairs, where each ordered pair consists of an element from the domain and its corresponding element in the range. This leads us to a more formal definition of a function.

Definition 1.7 A function f from a set A (the domain) to a set B (the range) is a subset of the Cartesian product $A \times B$ such that for every element $a \in A$, there exists exactly one element $b \in B$ such that the ordered pair (a, b) is in the subset. We denote this function as $f : A \to B$ and write $f(a) = b$.

Example 1.8 Consider the function $f : \{1, 2, 3\} \to \{a, b, c\}$ defined by the rule $f(1) = a$, $f(2) = b$, and $f(3) = c$. We can represent this function as the set of ordered pairs:

$$f = \{(1, a), (2, b), (3, c)\}. \tag{1.19}$$

Here, the domain is $\{1, 2, 3\}$ and the range is $\{a, b, c\}$. Each element in the domain is paired with exactly one element in the range, satisfying the definition of a function. Then, consider the Dirichlet function defined back in Section 1.1. This function can be represented as the set of ordered pairs:

$$D = \{(x, 1) \mid x \in \mathbb{Q}\} \cup \{(x, 0) \mid x \in \mathbb{R} \setminus \mathbb{Q}\}. \tag{1.20}$$

You would notice that the set is infinite and not very practical to write out in full, but it still adheres to the definition of a function since each real number x is paired with exactly one value (either 0 or 1) based on whether x is rational or irrational. Also, through our definitions, we can write:

$$D \subseteq \mathbb{R} \times \{0, 1\}$$

$$f \subseteq \{1, 2, 3\} \times \{a, b, c\}$$

I wish this is a pretty fresh way to look at functions. It is a bit more abstract, but it is also more precise. This definition allows us to work with functions in a more formal way, especially in advanced mathematics where precision is crucial. It also helps us understand the properties of functions better, such as injectivity, surjectivity, and bijectivity. The notation of functions is also slightly demystified through this lens. For instance, when we write $f(a) = b$, we are essentially "retrieving" the second element of the ordered pair (a, b) from the set that defines the function f. In calculus, we usually deal with real functions, or functions where both the domain and range are subsets of the real numbers $\mathbb{R}$. However, this set-theoretic definition of functions is general and can be applied to any sets, not just numbers. This abstraction is powerful because it allows us to study functions in a wide variety of contexts, from simple numerical functions to more complex structures in higher mathematics. For example, we can have functions between sets of matrices, or even functions between sets of functions (operators). This generality is one of the strengths of the set-theoretic approach to defining functions.

If you have read Chapter 2 or know linear transformations, you might be delighted to know that linear transformations are basically functions that map vectors to vectors. For instance, consider a function $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{r}) = \begin{bmatrix} 2\mathbf{r}_x \\ 3\mathbf{r}_y \end{bmatrix}$. But why is it a linear transformation? Because it satisfies the properties of additivity and homogeneity.[1]

Definition 1.8 A linear transformation (or actually any function) T is a function such that

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$
$$T(c\mathbf{u}) = cT(\mathbf{u})$$

I hope you see the connection between functions and linear transformations now. Linear transformations are just a special type of function that preserves vector space operations. It also explains why linear transformations are *linear*. This is a beautiful example of how abstract mathematical concepts can be applied in various contexts, from simple functions to complex transformations in vector spaces. I also hope you see this definition of functions more than just a novelty, but a powerful tool that can enhance your understanding of mathematics as you progress in your studies.

Problem 1.17 (1) Let $A = \{1, 2, 7, 9\}$ and $B = \{x, y, z\}$. List all the elements of the Cartesian product $A \times B$. Then, define a function $f : A \to B$ by specifying its set of ordered pairs such that $f(1) = x$, $f(2) = y$, $f(7) = z$, and $f(9) = x$. Write out the set that represents this function.

Problem 1.18 (1) Are the following linear transformations? Justify your answer.

a) $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{r}) = \begin{bmatrix} 2\mathbf{r}_x + 1 \\ 3\mathbf{r}_y \end{bmatrix}$
b) $S : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $S(\mathbf{r}) = \begin{bmatrix} -\mathbf{r}_y \\ \mathbf{r}_x \end{bmatrix}$
c) $U : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $U(\mathbf{r}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Problem 1.19 (2) Prove that any linear transformation must map the zero vector to the zero vector. That is, show that if $T : \mathbb{R}^n \to \mathbb{R}^n$ [2] is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$. (Hint: $\mathbf{0} + \mathbf{0} = \mathbf{0}$.)

---

[1] Please do verify these properties for yourself.

[2] Technically, linear transformations can map vectors to spaces of different dimensions as long as the properties of linearity are satisfied.

Chapter 2
Linear Algebra

2.1 Vector Basics

Stepping into this chapter, you may feel overwhelmed by the name, *Linear Algebra*. What is that? Please take note though, the knowledge in this chapter will not be used until Calculus III to IV in part IV of the book. If you are just willing to learn Calc I and II, you do not need the knowledge here. With that said, let's delve into Vectors. A concept that is very ubiquitous in this world is a coordinate, like the GPS coordinate in the maps app. A coordinate simply describes the position of a point in space. Generally, it is denoted (x, y) in 2 dimensions and (x, y, z) in 3 dimensions. Likewise, vectors are also behaves like coordinates and are written down in a *similar* fashion. Think of a vector as a coordinate.

Definition 2.1 A **vector** is a mathematical object with a magnitude, or length, and a direction. A vector, $\mathbf{a}$ is a point (x, y) in space drawn with an arrow. It is denoted as [1]:

$$\mathbf{a} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Of course, vectors by themselves are not very useful. Hence, we shall define the vector space. For the purposes of this text, you may think of vector spaces as the 2 and 3-dimensional Euclidean spaces you are very familiar with. In a vector space, we can finally meaningfully manipulate them. Take a look at the vector $\mathbf{a} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ and the vector $\mathbf{b} = \left[\begin{smallmatrix} -2 \\ 1 \end{smallmatrix}\right]$ in the two-dimensional Euclidean space:

You may have noticed that the coordinate axes in Figure 2.1 are the same as the ones you are familiar with. However, in linear algebra, instead of the usual xy axes to describe the components of a vector, you have $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ to describe the components of a vector. (So instead of saying the "x" component of a vector, you would say the "i hat" component of a vector.) [2] This is a convention in linear algebra. $\hat{\mathbf{i}}$ translates to the x axis and $\hat{\mathbf{j}}$ translates to the y axis. Likewise, in the third dimension, $\hat{\mathbf{k}}$ represents the z axis. Another thing to note are the *basis vectors*:

Definition 2.2 The set of vectors in a vector space are called **basis** if every vector in the vector space can be written uniquely written as a finite *linear combination* [3] of the basis vectors. An example would be these

[1] Likewise, a vector at (x, y, z) is denoted $\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$

[2] Some authors use $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ instead of the usual $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$. We shall use the latter.

[3] A linear combination of x and y is $ax + by$ where a, b are constants.

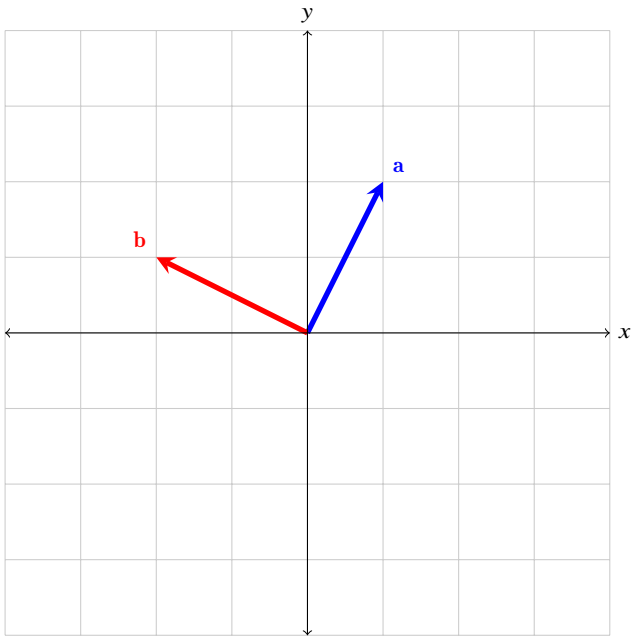

Fig. 2.1 Figure of vectors **a** and **b**.

vectors $\hat{\mathbf{i}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\hat{\mathbf{j}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ in the 2-dimensional Euclidean space which can be thought of one unit of distance in the x and y direction respectively.

With this in mind, another way of writing a vector $\mathbf{a} = \begin{bmatrix} x \\ y \end{bmatrix}$ is $\mathbf{a} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}}$.[1] Of course $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are not the only basis vectors in 2d, if you have learned polar coordinates before, there is the equivalent $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$, but this is a topic for a later chapter. Next, we shall look at the *addition* of vectors. A vector $\mathbf{a} = \begin{bmatrix} x_a \\ y_a \end{bmatrix}$ added to a vector $\mathbf{b} = \begin{bmatrix} x_b \\ y_b \end{bmatrix}$ is the vector $\begin{bmatrix} x_a + x_b \\ y_a + y_b \end{bmatrix}$. As you can see, the addition of two vectors results in another vector. Here is a illustration to help clear up things: [2]

Let us analyze Figure 2.2. Firstly, notice that $\mathbf{a} = 1\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$ and $\mathbf{b} = 1\hat{\mathbf{i}} + 1\hat{\mathbf{j}}$. Thus, $\mathbf{a} + \mathbf{b} = 2\hat{\mathbf{i}} + 3\hat{\mathbf{j}}$, exactly like in our figure. One way to understand this is to "attach" the vector **a**'s tail to **b**'s head by translating **a**. [3] You may have seen this in your physics class before. Similarly, the subtraction of vectors $\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} c \\ d \end{bmatrix}$ is $\mathbf{x} - \mathbf{y} = \begin{bmatrix} a-c \\ b-d \end{bmatrix}$. Of course, we can use the same "head to tail" model to explain vector subtraction. Subtracting **y** from **x** is simply finding the vector whose tail is on the head of **y** and whose head is on the head of **x**. However, some of you may be scratching your heads right now, what if a vector does not start from the *origin* (or the coordinate $(0, 0)$) but its tail lands on, say $(1, 1)$. How do we express such a vector? The answer may be surprising, both vectors have the same representation. However, unless specifically stated, *all* vectors start from the origin. This is also why the "head to tail" model works, while the "tail" of the vector is not at the origin, translating it to the origin results in an *identical* vector.

Example 2.1 Here are some examples of vectors:

[1] You can think of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ as one unit of direction in the x and y axis respectively.

[2] This can be extrapolated into any dimensions: $\begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} d \\ e \\ f \end{bmatrix} = \begin{bmatrix} a+d \\ b+e \\ c+f \end{bmatrix}$.

[3] Translation means moving something without rotation or any scaling.

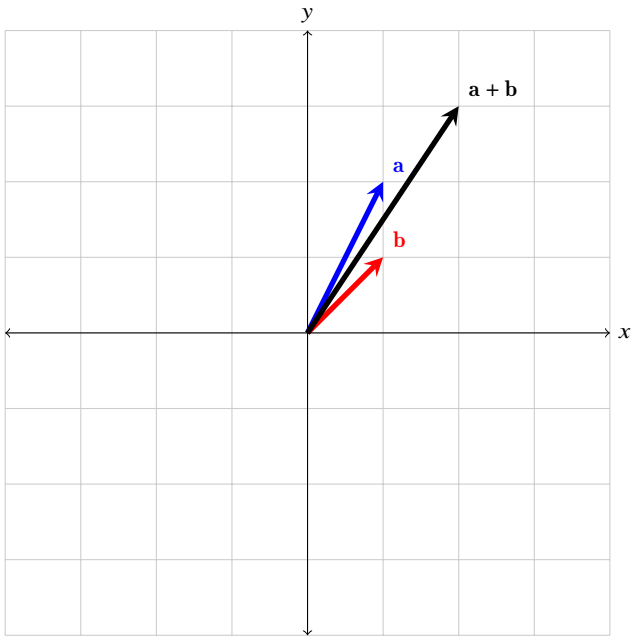

Fig. 2.2 Illustration of vector addition.

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \mathbf{b} = \begin{bmatrix} -3 \\ 4 \end{bmatrix} \mathbf{c} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

Here are some examples of adding and subtracting vectors. Note that addition are *commutative*, in other words the order does not matter:

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} 1 - 3 \\ 2 + 4 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$$

$$\mathbf{a} - \mathbf{c} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

$$\mathbf{a} + \mathbf{b} - \mathbf{c} = \begin{bmatrix} -2 \\ 6 \end{bmatrix} - \mathbf{c} = \begin{bmatrix} -4 \\ 8 \end{bmatrix}$$

Given a vector $\mathbf{a}$, $-\mathbf{a}$ can be thought of the same vector with the same magnitude but pointing in the *opposite* direction. To find it, simply flip the signs of all the components in $\mathbf{a}$. If you add $\mathbf{a}$ to $-\mathbf{a}$, you get what is called the *zero vector*, or the vector $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. This is a vector without direction has zero magnitude. We shall denote it $\mathbf{0}$. Lastly, I want to briefly touch on the idea of *magnitude* which is basically the length of a vector. The derivation of the following formula is simply the Pythagorean theorem.

Definition 2.3 The **magnitude** of a vector $\mathbf{v}$ in n-dimensional space is as such:

$$||\mathbf{v}|| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

A **unit** vector is a vector with magnitude 1. A **zero** vector has magnitude 0 and is denoted $\mathbf{0}$.

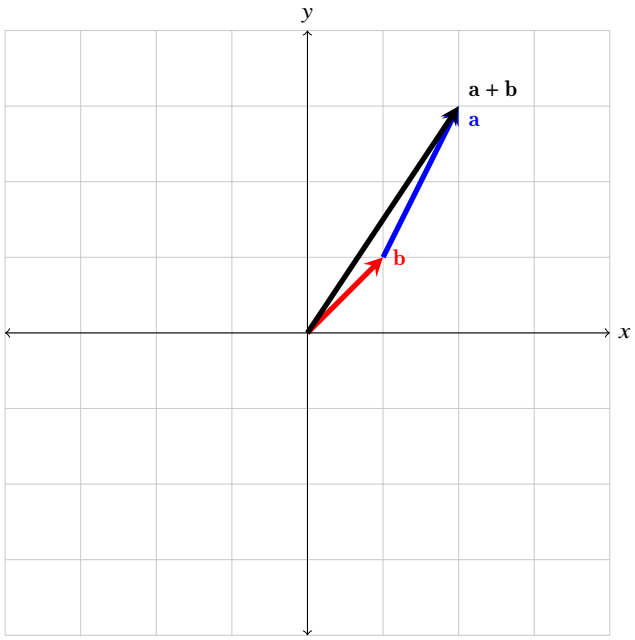

Fig. 2.3 Illustration of vector "head to tail" addition.

Problem 2.1 Given that $\mathbf{a} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$, calculate:

a) $\mathbf{a} + \mathbf{b}$
b) $\mathbf{a} - \mathbf{c}$
c) $\mathbf{a} - \mathbf{b} - \mathbf{c}$

Problem 2.2 Are the following statements true or false:

a) The subtraction of two vectors *always* results in a vector.
b) When any vector is added to the zero vector, the result *always* is not a zero vector.

Problem 2.3 Find the magnitude of $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$.

Problem 2.4 (1) Prove that the addition of vectors is *associative* given that the addition in $\mathbb{R}$ is associative. [1]

Problem 2.5 (2) Consider the 3-dimensional vector space where all vectors are described by $a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}$.

a) Consider the vector space which is the subset of vectors where $a_z = 0$. What is the dimension of this vector space?
b) Prove that the components of a vector with respect to this basis $(\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}})$ are unique.

[1] An operation $\circ$ is associative if order in which it is applied does not matter, in this case $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.

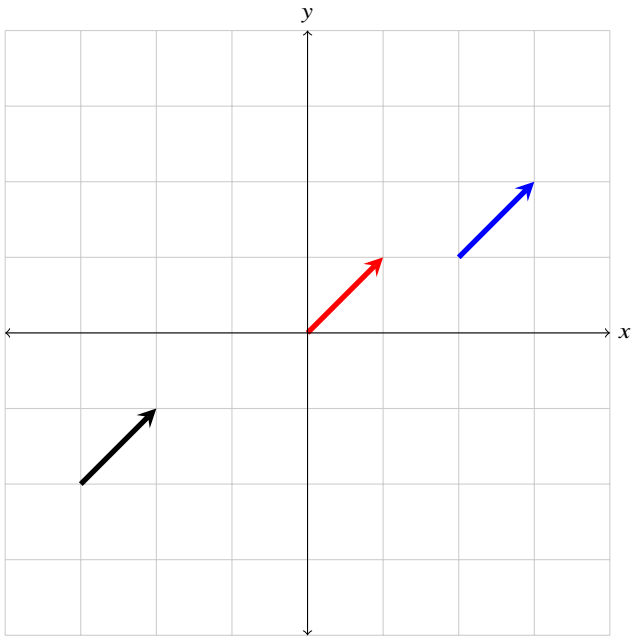

Fig. 2.4 Three identical vectors.

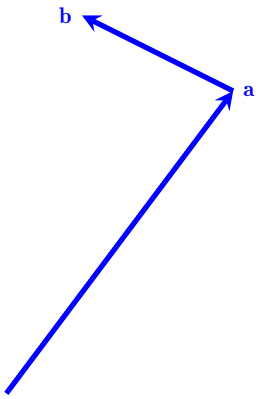

Fig. 2.5 Visualization of "head to tail" addition of **a** and **b**.

2.2 Scaling and dot product

In this chapter we will deal two kinds of multiplication with vectors. By far the easiest would be scaling a vector. As we have previously discussed, the magnitude of a vector is the length of a vector. But what if I want to "elongate" the vector by some amount to make it longer or shorter. This is called *scaling*.

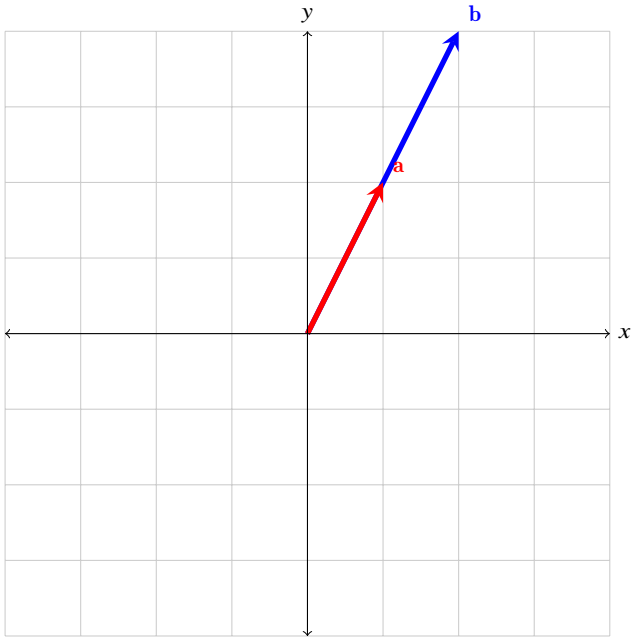

Fig. 2.6 Two vectors, one twice as long as the other.

Take a look at Figure 2.6, $\mathbf{b}$ is twice as long as $\mathbf{a}$. What should we multiply to $\mathbf{a}$ to make as long as $\mathbf{b}$? If you guessed 2, then you are right. (Not sure what else you could have guessed.) To multiply a vector by a constant, we multiply each individual component of the vector by the constant. From the figure, we can identify that $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ so $2\mathbf{a} = \begin{bmatrix} 2\times1 \\ 2\times2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ which is equal to $\mathbf{b}$. You may extrapolate this to higher dimensions.

Definition 2.4 Scaling an n-dimensional vector $\mathbf{v}$ by a constant n is defined as such:

$$n\mathbf{v} = n\begin{bmatrix} v_1 \\ \cdots \\ v_n \end{bmatrix} = \begin{bmatrix} n\times v_1 \\ \cdots \\ n\times v_n \end{bmatrix}$$

Let us try something fun, what if we multiplied the vector by a negative number? Well, the computation remains the same. As you can see from Figure 2.7, $\mathbf{a} = (-0.5)\mathbf{b}$. You may wish to compute this yourself to verify. One thing to note here is that the vector gets flipped across the origin when the constant the vector is multiplied to is negative.

Challenge 3. *Also, when scaling a vector by some constant, the reader may want to verify that the magnitude of the vector is also scaled by that constant. This should be a trivial proof.*

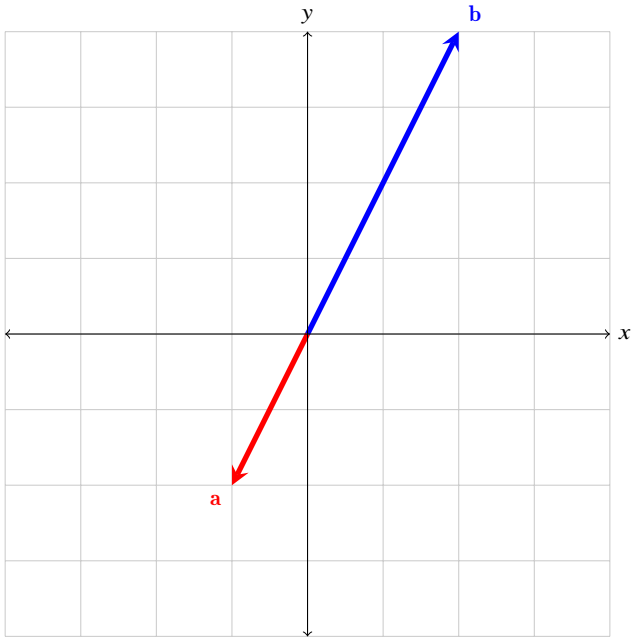

Fig. 2.7 Two vectors, one is scaled by -0.5.

You may have noticed that I have abstained from using either the $\cdot$ or $\times$ operators. The reason is bad notation which is sprinkled throughout mathematics. These two operators have special meanings in the realm of linear algebra. We shall first explore the easier one, the *dot product*.

Dot products use the $\cdot$ operator and only acts on vectors, do not abuse notation by using the $\cdot$ operator in the context of scaling a vector. Dot products are actually fairly straightforward to compute:

Definition 2.5 The **dot product** of two vectors $\mathbf{a}$ and $\mathbf{b}$ in n-dimensional space is:

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = (||\mathbf{a}||)(||\mathbf{b}||) \cos\theta \tag{2.1}$$

where θ is the angle between the two vectors. We shall refer to the formula between the first and second equal signs the "first definition of dot products" and the formula after the second equal sign the "second definition of dot products".

From Equation (2.1), we can tell that the dot product tells us something about the relationship between the two vectors. In essence, if the dot product between two vectors is large and positive, then the vectors are pointing in roughly the same direction. Else, if the product is very negative, then the dot product is very negative.

Firstly, note down the components of each vector in Figure 2.8. Then, let us compute $\mathbf{a}\cdot\mathbf{b}$. By Equation (2.1), it is $1(-2) + 2(1) = 0$. Also, it is important to take note that these two vectors are orthogonal(at right angles with each other). This brings us to an important property of dot products, if two vectors are orthogonal, then their dot product is 0. This can be seen from the fact that $\cos(\frac{\pi}{2}) = 0$ and from the second definition of dot products in Equation (2.1). In a sense, the dot product tells you how aligned are the directions of two vectors. Another example would be taking the dot product of two identical vectors, such as $\mathbf{a} \cdot \mathbf{a}$. The dot

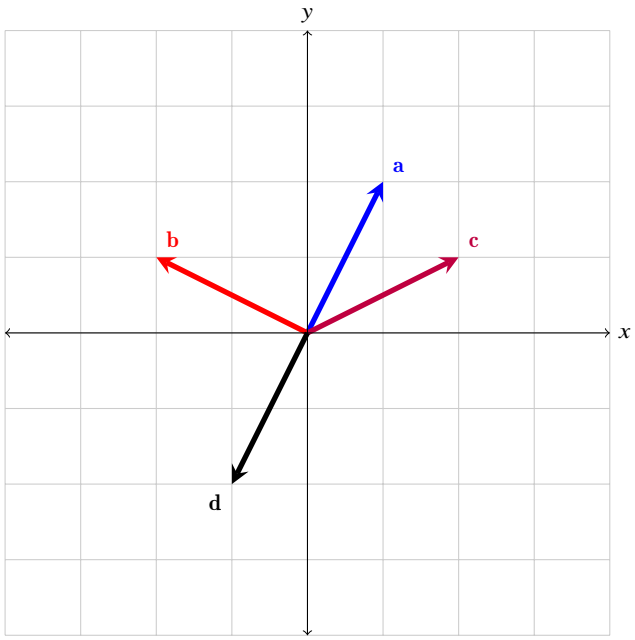

Fig. 2.8 Four vectors pointing in different directions.

product there is $1(1) + 2(2) = 5$. In addition, we know $\cos 0 = 1$ so according to the second definition, $\mathbf{a} \cdot \mathbf{a} = (||\mathbf{a}||)(||\mathbf{a}||) = (||\mathbf{a}||)^2$. If we take the square root on both sides, $||\mathbf{a}|| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$ which can be rewritten as $||\mathbf{a}|| = \sqrt{a_1^2 + \cdots + a_n^n}$. Drum roll please. This is exactly the formula we got for the magnitude of a vector. This is a pretty nice result.

Example 2.2 From Figure 2.8, we can see that:

$$\mathbf{a} \cdot \mathbf{c} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 1(2) + 2(1) = 4$$

$$\mathbf{a} \cdot \mathbf{d} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -2 \end{bmatrix} = 1(-1) + 2(-2) = -5$$

Remark 2.1 Generally, when finding the dot product, we use the first form in Equation (2.1) as it is typically easier to compute. It is also important to know that the dot product of two vectors results in a number, not a vector like in scaling.

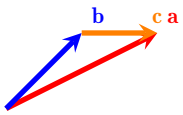

Fig. 2.9 A triangle formed by three vectors.

Problem 2.6 If $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, compute the following:

a) $2\mathbf{x}$
b) $\mathbf{x} \cdot \mathbf{y}$
c) $\mathbf{y} \cdot \mathbf{x}$

Problem 2.7 (1) Prove or disprove that the dot product *commutes*. [1]

Problem 2.8 (1) Newton's second law states that $\mathbf{F} = m\mathbf{a}$. Consider an object with a mass 10kg in the 2d plane.

a) If the object is accelerating in the x-axis with a magnitude of 3m/s^2 and accelerating in the y-axis with a magnitude of 4m/s^2, what is the x component of the force?
b) What is the y-component of the force in (a). Hence, what is the magnitude of the force.

[1] *An operation $\circ$ is commutative if $x \circ y = y \circ x$.*

2.3 Matrices

In this chapter, we will introduce a new mathematical object that you have probably heard before but not sure what it is, *linear transformation*. A great place to start is the *basis vectors*. All vectors can be written as a unique linear combination of the basis vectors. But $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are not the only basis vectors out there, what if I wanted to describe vectors using another set of basis vectors? Perhaps that would be more convenient for my special use case. Well, it is actually pretty easy to do so. For the sake of this argument, let us try to rotate all of the vectors in a normal 2d vector field anti-clockwise by $90°$. Notice how we could not describe $\mathbf{b}$ as some

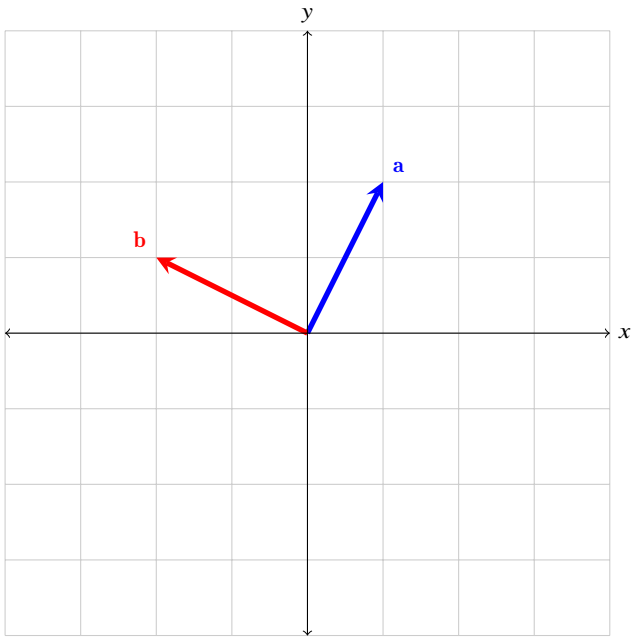

Fig. 2.10 Two vectors, one rotated anti-clockwise by $90°$.

constant multiplied to $\mathbf{a}$ in Figure 2.10. We need some other tool to help us. If we change the basis vectors of $\mathbf{a}$ to another pair of basis vectors, can we somehow get $\mathbf{b}$? Here is a thought: replace the $\hat{\mathbf{i}}$ with $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\hat{\mathbf{j}}$ with $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$. So, $\mathbf{a} = 1\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$ will become $\mathbf{a} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 2\begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$. This is pretty intriguing, by changing the basis vectors, we have basically "morphed" $\mathbf{a}$ into $\mathbf{b}$. This "morphing" is called a *linear transformation* or *linear map*. In some of the visualisations of the linear transformations I might keep the original grid in the background in a lighter color.

Example 2.3 Here are some examples of linear transformations, the blue grid is the vector space after the transformation. Take note where the basis vectors are after the transformations.

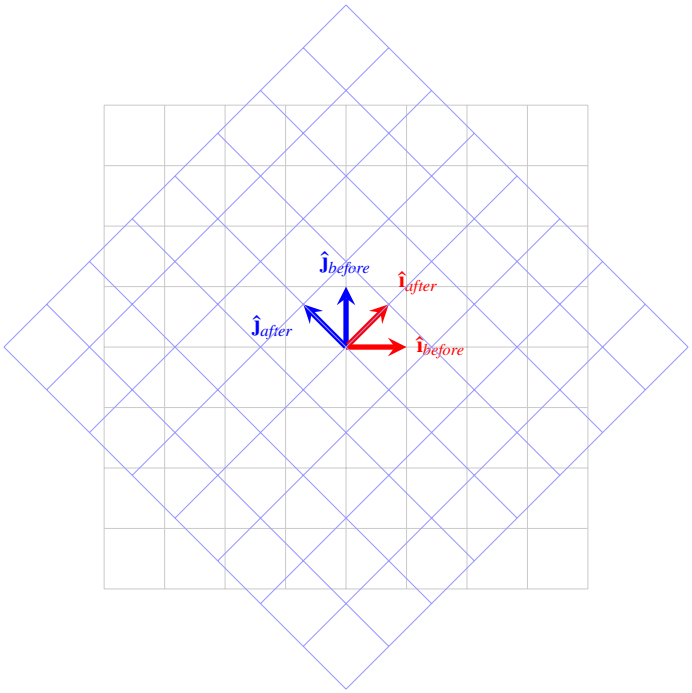

Fig. 2.11 A transformation in the form of an anti-clockwise *rotation*.

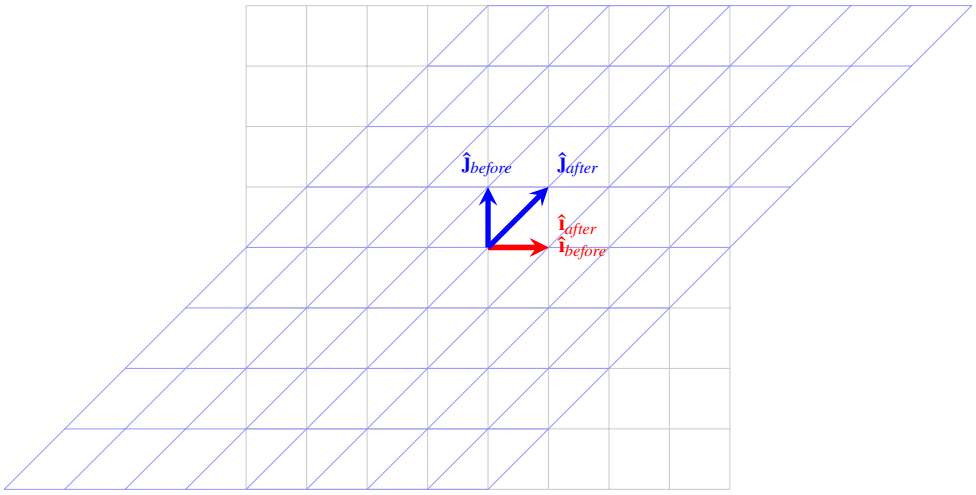

Fig. 2.12 A transformation in the form of a *shear*.[1]

[1] In Figure 2.12, the $\hat{\mathbf{i}}$ vector remains at the same location.

To describe these transformations, we actually only need to describe where the basis vectors land after the transformation. No other information required. We present the location of the new basis vectors in the form of a *matrix*:

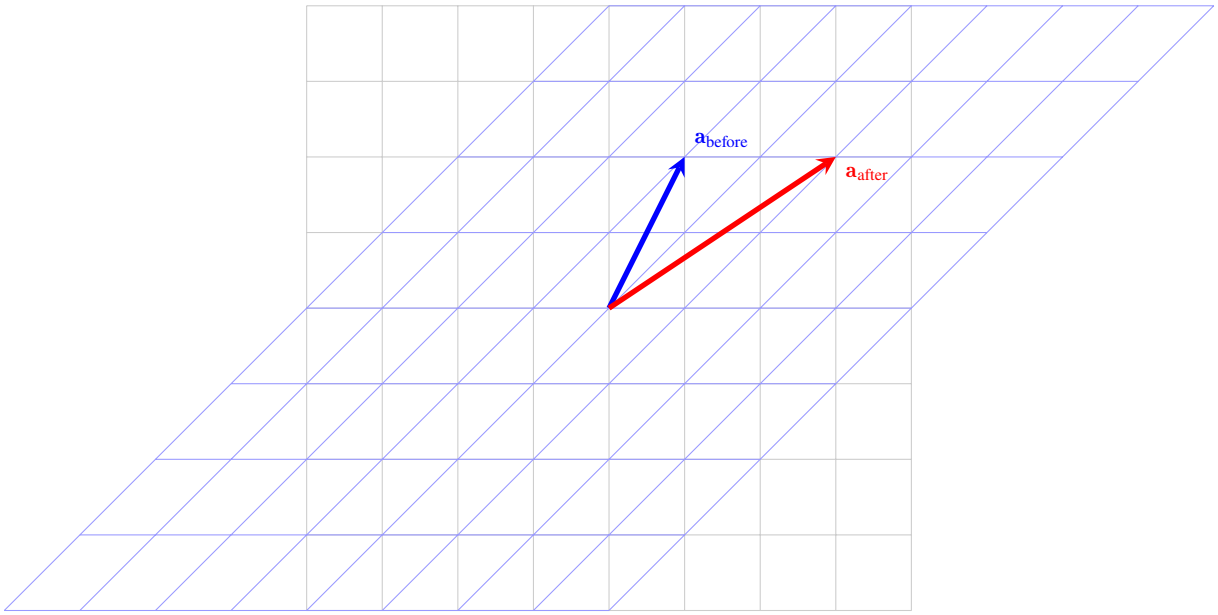

Fig. 2.13 Figure of a vector before and after a shear transformation like in Figure 2.12.

Definition 2.6 In linear algebra, **matrices** generally represent linear transformations. An $n \times m$ matrix is one with n rows and m columns. The nth column of the matrix denotes the representation of the corresponding basis vector after the linear transformation.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

represents a matrix that moves the $\hat{\mathbf{i}}$ basis vector to $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ and moves the $\hat{\mathbf{j}}$ basis vector to $\left[\begin{smallmatrix} b \\ d \end{smallmatrix}\right]$.

Remark 2.2 Matrices can also represent a variety of things apart from linear transformations. It is just a mathematical object representing a bunch of things arranged in rows and columns.

It should go without saying that any vector in the original vector space will get moved after a transformation. To calculate where a vector ends up after a transformation, we multiply the matrix by the vector. This is where many textbooks make linear algebra feel laborious. In order to make the process more intuitive, recall that a vector has a $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ component. Now, we swap the basis vectors for the new basis vectors described in our matrix. This is just multiplying a vector by a constant, which you have done before.

Example 2.4 Take the transformation in Figure 2.12 which is represented by the matrix $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$. Let us say in the original vector space, there is a vector $\mathbf{a} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$. Then, the vector's representation after the transformation:

$$\begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix} = 1\begin{bmatrix}1\\0\end{bmatrix} + 2\begin{bmatrix}1\\1\end{bmatrix}$$
$$= \begin{bmatrix}3\\2\end{bmatrix}$$

If you have come from a linear algebra course, this may feel very refreshing. Instead of memorizing what needs to be multiplied, you just swap the basis vectors for some other basis vectors. While we get the representation of **a** from Figure 2.13, it is also important to take note that in the new vector space after the transformation, **a** has one unit of direction in the horizontal direction and two units of direction in the "vertical" direction. This is just like the original representation of **a**, just that the vector space has been sheared. This change of basis vector is important enough to have its own name, *change of basis*.

With this knowledge, let us proceed to one of the most hated parts of linear algebra, *matrix multiplication*. These few words can bring out so much frustration out of so many students. It is quite a laborious and boring process if you do not grasp the meaning and motivation behind it. In actuality, the idea is simple. When you multiply two matrices together, you are effectively applying two linear transformations, one after another. The result is also a matrix, which is a linear transformation, hence the operation is *closed*. [1]

Example 2.5 Say you have two matrices, a 90°anti-clockwise rotation $x = \left[\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right]$ and a horizontal shear $y = \left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]$. xy is simply applying the shear first before the rotation. (The right to left reading is similar to that of function composition)

[1] This is only true if the matrix is $n \times n$, which are the ones we are dealing with.

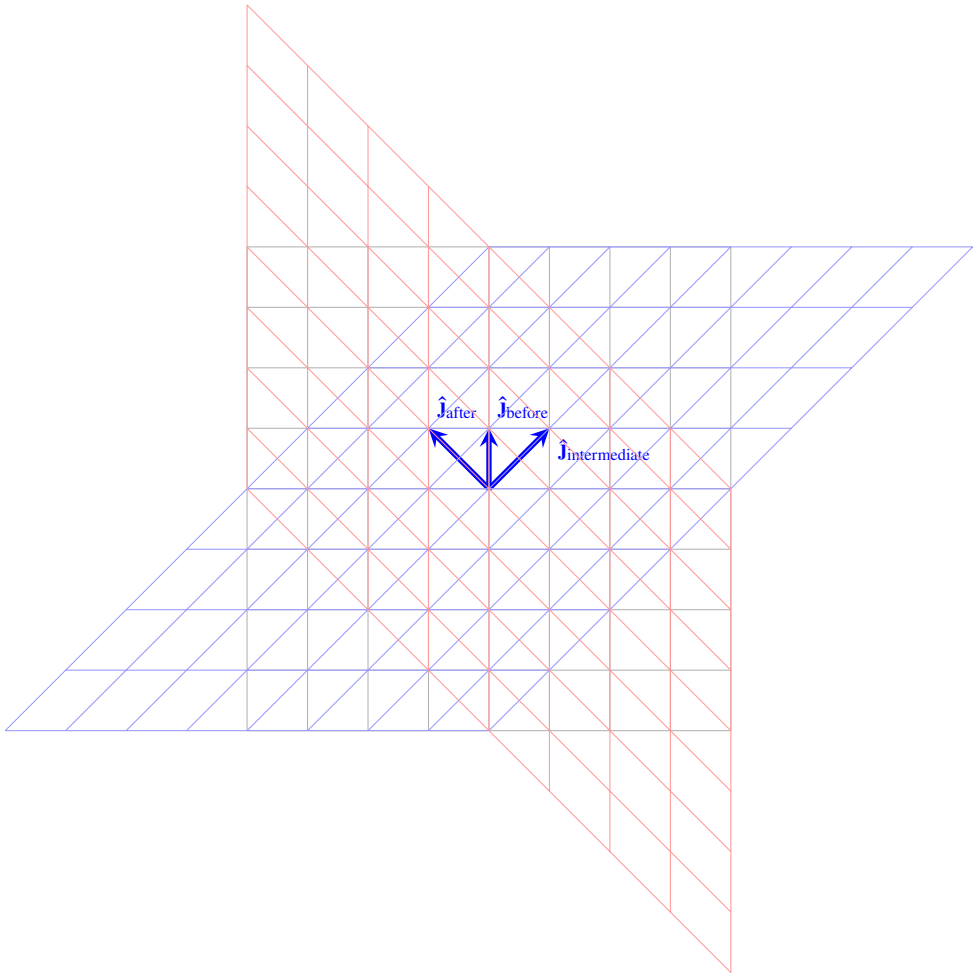

Fig. 2.14 Two linear transformations.

It should not be difficult to see that the end result of applying the two linear transformations in the previous examples resulted in another linear transformation (the red grid). We wish to find the representation of this linear transformation as a matrix. Let us go through this step by step. First of all, the basis vectors of the *identity matrix* [1], gets moved to some new location. In other words, given a matrix $a = \left[\begin{smallmatrix} a_1 & a_2 \\ a_3 & a_4 \end{smallmatrix}\right]$, we know that the basis vectors gets moved to $\left[\begin{smallmatrix} a_1 \\ a_3 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} a_2 \\ a_4 \end{smallmatrix}\right]$ respectively for the $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ vectors. Now, we have another transformation $b = \left[\begin{smallmatrix} b_1 & b_2 \\ b_3 & b_4 \end{smallmatrix}\right]$. To know where the $\hat{\mathbf{i}}$ vector lands after both transformations, notice that it is simply multiplying the intermediate position of it by b, which is $\left[\begin{smallmatrix} b_1 & b_2 \\ b_3 & b_4 \end{smallmatrix}\right]\left[\begin{smallmatrix} a_1 \\ a_3 \end{smallmatrix}\right] = \left[\begin{smallmatrix} a_1b_1+a_3b_2 \\ a_1b_3+a_3b_4 \end{smallmatrix}\right]$. A similar line of reasoning may be used for $\hat{\mathbf{j}}$. Hence, the result is:

[1] The identity matrix is simply the transformation that leaves the vector space unchanged.

$$ba = \begin{bmatrix} a_1b_1 + a_3b_2 & a_2b_1 + a_4b_2 \\ a_1b_3 + a_3b_4 & a_2b_3 + a_4b_4 \end{bmatrix}$$

This is the reason many students find linear algebra tedious. Most courses make students memorize the formula which is extremely rote. I wish that you understand what is going on behind the scenes on this formula. With that said, it is best to get some experience with these multiplications:

Example 2.6 Given that $x = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ and $y = \left[\begin{smallmatrix} 1 & 3 \\ 2 & 4 \end{smallmatrix}\right]$ and $z = \left[\begin{smallmatrix} 5 & 6 \\ 7 & 8 \end{smallmatrix}\right]$,

$$xy = \begin{bmatrix} 5 & 11 \\ 11 & 25 \end{bmatrix}$$
$$yx = \begin{bmatrix} 10 & 14 \\ 14 & 20 \end{bmatrix}$$
$$x(yz) = \begin{bmatrix} 102 & 118 \\ 230 & 266 \end{bmatrix}$$
$$(xy)z = \begin{bmatrix} 102 & 118 \\ 230 & 266 \end{bmatrix}$$

This examples shows that matrix multiplication does not commute but it is associative.

Problem 2.9 If $\mathbf{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, $a = \left[\begin{smallmatrix} 2 & 3 \\ 4 & 5 \end{smallmatrix}\right]$ and $b = \left[\begin{smallmatrix} 6 & 7 \\ 8 & 9 \end{smallmatrix}\right]$, compute:

a) $a\mathbf{x}$
b) ab
c) $ab\mathbf{x}$

Problem 2.10 (1) Prove that matrix multiplication is *associative* for all 2×2 matrices.

Problem 2.11 (1) The *trace* of an $n \times n$ matrix A is defined as

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} \tag{2.2}$$

where a_{ii} denotes the entry for the ith column and ith row of A.

a) Let $A = \left[\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right]$, what is the trace of A.
b) Prove that, for square matrices A and B, $\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$.
c) Hence, if $B = \left[\begin{smallmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{smallmatrix}\right]$, verify that $\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$.

2.4 Determinants, inverses and Cramer's rule

At this juncture, you should be familiar with the concept that a matrix represents a linear transformation and it morphs the vector space. Also, if you imagine the basis vectors of any vector space as lengths of a parallelogram, then you might want to find the area of that parallelogram. For example, the identity matrix's basis vectors forms a square with area 1. Likewise, given a matrix $x = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, we can display it as:

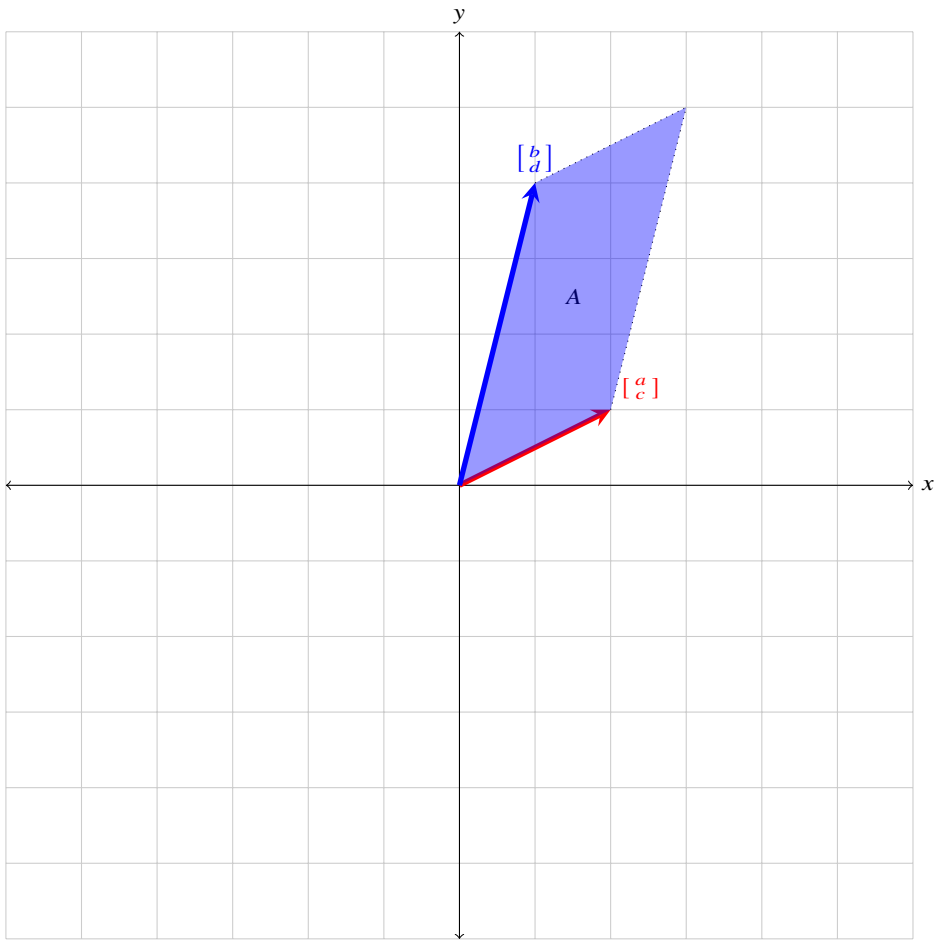

Fig. 2.15 Figure of area A formed by a pair of basis vectors.

To find the area of A in Figure 2.15, we may use the sine rule (let $\mathbf{u} = \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ and $\mathbf{v} = \left[\begin{smallmatrix} b \\ d \end{smallmatrix}\right]$):

$$\begin{aligned}
A &= (||\mathbf{u}||)(||\mathbf{v}||)\sin(\theta) \quad \text{where } \theta \text{ is the angle between the vectors} \\
&= (||\mathbf{u'}||)(||\mathbf{v}||)\cos(\theta') \quad \text{where } \mathbf{u'} \text{ is } \left[\begin{smallmatrix} a \\ -c \end{smallmatrix}\right], \text{ perpendicular to } \mathbf{v} \text{ and } \theta' \text{ is } 90° \\
&= \begin{bmatrix} -c \\ a \end{bmatrix} \begin{bmatrix} b \\ d \end{bmatrix} \quad \text{by Equation (2.1), definition of dot product} \\
&= ad - bc
\end{aligned} \tag{2.3}$$

The resultant area is also known as the *determinant* of a matrix.

Definition 2.7 The **determinant** of a 2×2 matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is represented as:

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

and it represents the signed area of a parallelogram with sides of the basis vectors described in A. Meanwhile, a 3×3 matrix $B = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ has a determinant of:

$$\begin{aligned}
\det(B) &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \\
&= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}
\end{aligned}$$

and it represents the volume of the parallelepiped formed by the three basis vectors in B.

Example 2.7 Given that [1] $A = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$,$B = \left[\begin{smallmatrix} 2 & 5 \\ 2 & 5 \end{smallmatrix}\right]$ and $C = \left[\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right]$:

$$\begin{aligned}
\det(A) &= 1 \times 4 - 2 \times 3 \\
&= -2
\end{aligned}$$

$$\det(B) = 0 \tag{2.4}$$

$$\begin{aligned}
\det(C) &= 1 \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\
&= (5 \times 9 - 8 \times 6) - 2(4 \times 9 - 6 \times 7) + 3(4 \times 8 - 7 \times 5) \\
&= 0
\end{aligned}$$

Next, we would move onto the *inverse* of a matrix. Think of a number x, $x^{-1} \times x = 1$, 1 is also called the multiplicative identity. Likewise, given a matrix Y, $YY^{-1} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ which is the identity matrix. If you think of Y as a transformation, then Y^{-1} "undo" the transformation of Y.

Definition 2.8 The (multiplicative) **inverse** of a matrix X is denoted X^{-1} where:

$$XX^{-1} = I$$

[1] Equation (2.4) shows a more general property, given that $X = \left[\begin{smallmatrix} a & b \\ a & b \end{smallmatrix}\right]$, $\det(X) = 0$. The proof is trivial and shall be left as a verification to the reader.

where I is the identity matrix. If X has an inverse, it is called an *invertible* matrix. X has an inverse if and only if $\det(X) \neq 0$. To find the inverse of a matrix $Y = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, we use the following formula [1]:

$$Y^{-1} = \frac{1}{\det(Y)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{2.5}$$

Example 2.8 Given $A = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 2 & 5 \\ 2 & 5 \end{smallmatrix}\right]$:

$$\begin{aligned} A^{-1} &= \frac{1}{-2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix} \end{aligned}$$

$\det(B) = 0$, hence B is not invertible.

Until now, this chapter has been focusing mainly on computation. The reader may feel that these computations are excessively dull and boring. That being said, here is a application of determinants, *Cramer's rule*. Cramer's rule may be used to solve a system of linear equations. Firstly, we will represent a simultaneous equation as a product of vectors and matrices. Given a simultaneous equation:

$$\begin{aligned} x + y &= 10 \\ 2x - y &= 14 \end{aligned} \tag{2.6}$$

We write it as:

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 10 \\ 14 \end{bmatrix} \tag{2.7}$$

This is also called the "matrix form" of a simultaneous equation. It also important to recognize that Equation (2.7) is in the form of $A\mathbf{v} = B$, it is easy to see that $\mathbf{v} = BA^{-1}$. From Equation (2.5), we can find BA^{-1}

$$\begin{aligned} BA^{-1} &= \begin{bmatrix} 10 \\ 14 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \\ &= \begin{bmatrix} 8 \\ 2 \end{bmatrix} \end{aligned}$$

This is also the solution for Equation (2.7), x is indeed 8 and y is indeed 2. However, this method might not be the best. If there were three unknowns and three equations, how could you solve the equations? We may apply Cramer's rule. Consider a system of n linear equations with n unknowns $(x_1, x_2, \ldots, x_n)$, it can be represented as $Ax = B$ as we have seen before. The solutions for the equations are:

$$x_i = \frac{\det(A_i)}{\det(A)} \quad i = 1, \ldots, n$$

Where x_i is the i-th solution and A_i is the matrix formed by replacing the i-th column of A by B.

Example 2.9 Given the following system of linear equations of three unknowns:

[1] When multiplying a matrix by a constant, you multiply each entry in matrix by the constant.

$$\begin{aligned} x_1 + x_2 + x_3 &= 12 \\ x_1 - x_2 + x_3 &= 4 \\ x_1 - x_2 - x_3 &= -6 \end{aligned}$$

and its matrix form is:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 4 \\ -6 \end{bmatrix}$$

The solutions are:

$$\begin{aligned} x_1 &= \frac{\det(A_1)}{\det(A)} \\ &= \frac{\det(\begin{bmatrix} 12 & 1 & 1 \\ 4 & -1 & 1 \\ -6 & -1 & -1 \end{bmatrix})}{\det(\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix})} \\ &= \frac{12}{4} \\ &= 3 \end{aligned}$$

$$\begin{aligned} x_2 &= \frac{\det(A_2)}{\det(A)} \\ &= \frac{\det(\begin{bmatrix} 1 & 12 & 1 \\ 1 & 4 & 1 \\ 1 & -6 & -1 \end{bmatrix})}{\det(\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix})} \\ &= \frac{16}{4} \\ &= 4 \end{aligned}$$

$$
\begin{aligned}
x_3 &= \frac{\det(A_3)}{\det(A)} \\
&= \frac{\det(\begin{bmatrix} 1 & 1 & 12 \\ 1 & -1 & 4 \\ 1 & -1 & -6 \end{bmatrix})}{\det(\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix})} \\
&= \frac{20}{4} \\
&= 5
\end{aligned}
$$

As you can tell, Cramer's rule is very powerful. In the real world, however, *gaussian elimination* is a much faster alternative than Cramer's rule. It is also unlikely for the reader to be computing large matrices determinants by hand, we have computers for that. This also applies to matrix multiplication. This book and more generally calculus courses only need knowledge on computing determinants of up to 3×3 matrices.

Problem 2.12 Find the determinants and inverses, if any, for the following matrices:

a) $\begin{bmatrix} 1 & 9 \\ 2 & 10 \end{bmatrix}$

b) $\begin{bmatrix} 3 & 2 \\ 6 & 4 \end{bmatrix}$

c) $\begin{bmatrix} 3 & 4 \\ 7 & 9 \end{bmatrix}$

Problem 2.13 (1) Find the determinant of the following matrix: $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 2 & 3 & 6 \end{bmatrix}$

Problem 2.14 (1) Solve the following simultaneous equations using matrices:

$$
\begin{aligned}
x_1 + x_2 + x_3 &= 3 \\
x_1 + 2x_2 + x_3 &= 4 \\
x_1 + 2x_2 + 2x_3 &= 6
\end{aligned}
$$

2.5 Cross products and triple products

The previous section perfectly segues into this section. *Cross products* is an operation on vectors that happens solely in the third dimension. Do not confuse it with the dot product, both uses different operators and should not be used interchangeably. Firstly, we must recall that one of the basis vectors is $\hat{\mathbf{k}}$, and it is one unit in the z direction.

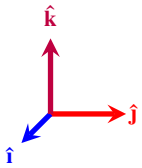

Fig. 2.16 Figure of the three basis vectors.

The cross product is a vector, whose magnitude is the area of a parallelogram formed by the two operands (which are vectors). [1] More formally, the cross product is defined as:

Definition 2.9 The **cross product** is an operation on two vectors and the result is another vector which is perpendicular to both vectors. It is defined by the formula:

$$\mathbf{a} \times \mathbf{b} = (||\mathbf{a}||)(||\mathbf{b}||)\sin(\theta)\mathbf{n}$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$ and $\mathbf{n}$ is the unit vector perpendicular to both of the operands such that all three vectors obey the "right-hand rule"[2]. The common way to compute the cross product is using the following formula:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \mathbf{a}_x & \mathbf{a}_y & \mathbf{a}_z \\ \mathbf{b}_x & \mathbf{b}_y & \mathbf{b}_z \end{vmatrix} \tag{2.8}$$

Lastly, do not confuse $\mathbf{n}$ with $\hat{\mathbf{k}}$ as both vectors are sometimes not the same.

Example 2.10 Given that $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$, then $\mathbf{a} \times \mathbf{b}$ would be:

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 1 & 2 & 0 \\ -2 & 1 & 0 \end{vmatrix} \\ &= 0\hat{\mathbf{i}} - 0\hat{\mathbf{j}} + 5\hat{\mathbf{k}} \\ &= 5\hat{\mathbf{k}} \end{aligned}$$

[1] The operands of $\mathbf{a} \times \mathbf{b}$ are $\mathbf{a}$ and $\mathbf{b}$.

[2] The right hand rule states that if you use your right hand and point the index finger in the direction of $\mathbf{a}$, your middle finger in the direction of $\mathbf{b}$, your thumb should be pointing in the direction of $\mathbf{n}$.

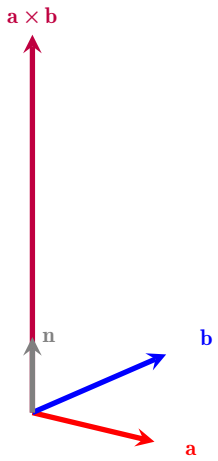

Fig. 2.17 Figure of two vectors and their cross product.

An important fact to note is that the operation is not commutative. In fact, the operation is called *anti-commutative*. This means that $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$. In addition, if the angle between the operands is any integer multiple of π, then the result is $\mathbf{0}$. This should be trivially seen from the formula for the cross product. The derivation for the cross product is also easy, given two vectors $\mathbf{a}$ and $\mathbf{b}$ and an angle θ between them, simply use the sine rule and the result should be obvious. The cross product has many real world applications.

Toque and the moment arm formula

Torque is a force that is common in our daily lives, present in the turning of a doorknob or the turning of a wrench. The formula for torque is $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$, where $\boldsymbol{\tau}$ is the torque, $\mathbf{r}$ is the distance from the pivot point to the point where the force is applied and $\mathbf{F}$ is the force applied. The direction of the torque can be determined by the right-hand rule. Torque is also known as the moment of force, and we actually derive the infamous moment arm formula for moment $\textit{moment} = (\textit{moment arm})(\textit{force})$. We shall restrict ourselves to only two dimensions and use a Cartesian coordinate system. Set the origin such that the pivot point is at the origin. Then, we can represent $\mathbf{r}$ as the distance from the origin to where the force is applied. Let $\mathbf{F}$ be the force vector. Since there is no more z component, all the z components of the vectors are 0. We can write down the cross product of $\mathbf{r}$ and $\mathbf{F}$ (we resolve each into its x and y components):

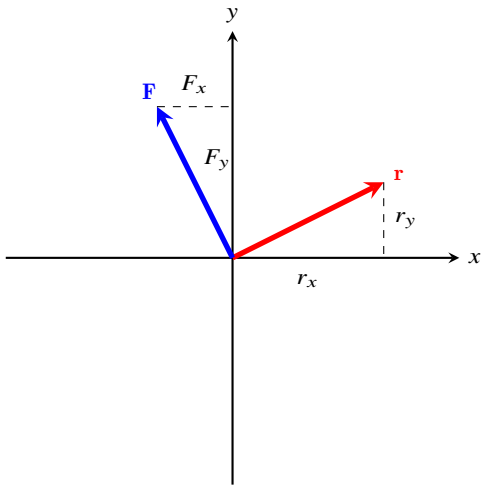

Fig. 2.18 Figure of torque in two dimensions.

$$\begin{aligned}\boldsymbol{\tau} &= \mathbf{r} \times \mathbf{F} \\ &= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ r_x & r_y & 0 \\ F_x & F_y & 0 \end{vmatrix} \\ &= (r_x F_y - r_y F_x)\hat{\mathbf{k}}\end{aligned}$$

We can consider the magnitude of the torque:

$$||\boldsymbol{\tau}|| = r_x F_y - r_y F_x \tag{2.9}$$

Next, we turn to the moment arm formula.

$$\begin{aligned}moment &= (moment\ arm)(force) \\ &= \sqrt{r_x^2 + r_y^2}\sqrt{F_x^2 + F_y^2}\end{aligned}$$

We are able to write this as I have assumed that the force and distance are perpendicular to each other (See Problem 2.18). Then, because the force and radius are perpendicular, notice that $\mathbf{r} \cdot \mathbf{F} = r_x F_x + r_y F_y = 0$. We can utilize this fact by writing down:

$$(r_x F_y - r_y F_x)^2 + (r_x F_x + r_y F_y)^2 = (r_x^2 + r_y^2)(F_x^2 + F_y^2) \tag{2.10}$$

Notice that the left-hand side of Equation (2.10) is simply $||\boldsymbol{\tau}||^2$ and the right-hand side is simply $(moment\ arm)^2(force)^2$. Hence, we have shown that $||\boldsymbol{\tau}|| = (moment\ arm)(force)$ which is the moment arm formula. An immediate difference between torque and moment is that moment is a scalar while torque is a

vector. Another fact about torque that it has a $\hat{\mathbf{k}}$ component, which means that it is perpendicular to the plane formed by $\mathbf{r}$ and $\mathbf{F}$.

Next, we will move on to triple products which happens to also be a three-dimensional topic. Firstly, take a look at the following three vectors:

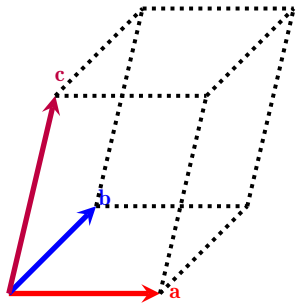

Fig. 2.19 Figure of triple vector multiplication.

Figure 2.19 is the example of triple vector multiplication. You would represent it as $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. The result here is simply the volume of the parallelepiped formed in Figure 2.19.

Definition 2.10 The **triple vector product** is a product of three 3-dimensional vectors. The most common of which, is the scalar triple product which is Figure 2.19. The formula for which is $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$, which yields the volume of the parallelepiped with the sides being the three vectors. Another type of triple vector multiplication is vector triple product. It is simply $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$. We can use **Lagrange's formula** to compute it easily: $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$.

Example 2.11 We shall derive the formula for the volume of a parallelepiped given its height, breadth and length, denote them h, b and l respectively. Notice that we define the sides of the parallelepiped like this:

$$\mathbf{a} = \begin{bmatrix} l \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix}$$

$$\mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix}$$

Now, we use the triple vector multiplication formula. However, take note of the brackets and the order of operations:

$$\begin{aligned}\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) &= \mathbf{a}\cdot\begin{vmatrix}\hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}}\\ 0 & b & 0\\ 0 & 0 & h\end{vmatrix}\\ &= \begin{bmatrix}l\\0\\0\end{bmatrix}\cdot\begin{bmatrix}bh\\0\\0\end{bmatrix}\\ &= lbh\end{aligned} \tag{2.11}$$

The result in Equation (2.11) is trivial and should come as no surprise to the reader.

As for vector triple product, the process of computation is very similar. We first evaluate the cross product of the two vectors in parentheses, then take the cross product of the first and resultant vector from the previous computation. We shall give a simple proof for Lagrange's formula.

Theorem 2.1 *(Lagrange's formula) Given* $\mathbf{a},\mathbf{b},\mathbf{c}$*, show that* $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{b}(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}(\mathbf{a}\cdot\mathbf{b})$. [1]

Proof. We will start by evaluating $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$:

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) &= \mathbf{a}\times\begin{vmatrix}\hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}}\\ b_x & b_y & b_z\\ c_x & c_y & c_z\end{vmatrix}\\ &= \mathbf{a}\times\begin{bmatrix}b_yc_z - b_zc_y\\ b_zc_x - b_xc_z\\ b_xc_y - b_yc_x\end{bmatrix}\\ &= \begin{vmatrix}\hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}}\\ a_x & a_y & a_z\\ b_yc_z - b_zc_y & b_zc_x - b_xc_z & b_xc_y - b_yc_x\end{vmatrix}\\ &= \begin{bmatrix}a_y(b_xc_y - b_yc_x) - a_z(b_zc_x - b_xc_z)\\ \cdots\end{bmatrix}\\ &= \begin{bmatrix}a_yb_xc_y + a_zb_xc_z - a_yb_yc_x - a_zb_zc_x\\ \cdots\end{bmatrix}\end{aligned} \tag{2.12}$$

We will examine the $\hat{\mathbf{i}}$ component of 2.12, the other two components are similar. Then, notice that:

$$\begin{aligned}\mathbf{b}(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}(\mathbf{a}\cdot\mathbf{b}) &= \mathbf{b}(a_xc_x + a_yc_y + a_zc_z) - \mathbf{c}(a_xb_x + a_yb_y + a_zb_z)\\ &= \begin{bmatrix}b_x(a_xc_x + a_yc_y + a_zc_z)\\ b_y(a_xc_x + a_yc_y + a_zc_z)\\ b_z(a_xc_x + a_yc_y + a_zc_z)\end{bmatrix} - \begin{bmatrix}c_x(a_xb_x + a_yb_y + a_zb_z)\\ c_y(a_xb_x + a_yb_y + a_zb_z)\\ c_z(a_xb_x + a_yb_y + a_zb_z)\end{bmatrix}\\ &= \begin{bmatrix}a_yb_xc_y + a_zb_xc_z - a_yb_yc_x - a_zb_zc_x\\ \cdots\end{bmatrix}\end{aligned} \tag{2.13}$$

As you can see, the $\hat{\mathbf{i}}$ components of 2.12 and 2.13 are identical. Hence, we have shown Lagrange's formula. □

[1] A sidenote here, when it is written $\mathbf{ab}$, this is not the cross or dot product, instead it is equal to $\begin{bmatrix}a_xb_x\\ a_yb_y\\ \cdots\end{bmatrix}$.

Remark 2.3 The above proof is a tedious pedagogical exercise mainly created to drill in the concepts of the cross products.

Problem 2.15 Find the cross products of the following vectors:

a) $\begin{bmatrix}1\\2\\5\end{bmatrix} \times \begin{bmatrix}1\\2\\7\end{bmatrix}$

b) $\begin{bmatrix}0\\2\\4\end{bmatrix} \times \begin{bmatrix}1\\3\\5\end{bmatrix}$

Problem 2.16 (1) Prove or disprove:

a) $\mathbf{a} \times \mathbf{b} = \mathbf{b} \times \mathbf{a}$
b) $\mathbf{a} \times \mathbf{a} = 0$

Problem 2.17 (1) Prove Lagrange's formula for the other two components not shown in the previous proof.

Further problems

Problem 2.18 (1) In this problem, you will derive the formula for the moment arm when the force and distance are not perpendicular. Refer to Figure 2.18 for when the force is perpendicular to the moment arm. Let the angle between $\mathbf{r}$ and $\mathbf{F}$ be θ.

a) Find the expression for the force that is perpendicular to $\mathbf{r}$ in terms of F and θ.
b) Thus, find the expression for the moment arm when the force and distance are not perpendicular.

Problem 2.19 Do not compute the cross products but just use the right-hand rule to find:

a) $\hat{\mathbf{i}} \times \hat{\mathbf{j}}$
b) $\hat{\mathbf{j}} \times \hat{\mathbf{k}}$
c) $\hat{\mathbf{k}} \times \hat{\mathbf{i}}$
d) $\hat{\mathbf{j}} \times \hat{\mathbf{i}}$
e) $\hat{\mathbf{k}} \times \hat{\mathbf{j}}$
f) $\hat{\mathbf{i}} \times \hat{\mathbf{k}}$

Problem 2.20 (2) Prove that the following determinant gives the area of a triangle with vertices at (x_1, y_1), (x_2, y_2) and (x_3, y_3):

$$Area = \left| \frac{1}{2} \det \left(\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \right) \right| \tag{2.14}$$

Therefore, if the determinant is 0, what does this imply about the three points?

Problem 2.21 (1) The vector projection of $\mathbf{a}$ onto $\mathbf{b}$ is defined as:

$$\text{proj}_{\mathbf{b}}\mathbf{a} = \left(\frac{\mathbf{a} \cdot \mathbf{b}}{||\mathbf{b}||^2} \right) \mathbf{b} \tag{2.15}$$

Hence, find the projection of $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ onto $\mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$.

Problem 2.22 (1) The *Gram-Schmidt process* is a way of finding two or more vectors that are orthogonal to each other. Given two vectors $\mathbf{a}$ and $\mathbf{b}$, we can find a vector $\mathbf{v}$ that is orthogonal to $\mathbf{a}$ by using the formula (the proj function is defined in Problem 2.21):

$$\mathbf{v} = \mathbf{b} - \text{proj}_{\mathbf{a}}\mathbf{b} \tag{2.16}$$

Hence, by using the Gram-Schmidt process, find vector $\mathbf{v_1}$ that is orthogonal to $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Repeat this process to find a vector $\mathbf{v_2}$ that is orthogonal to both $\mathbf{a}$ and $\mathbf{v_1}$. Do it once more to find a vector $\mathbf{v_3}$ that is orthogonal to $\mathbf{a}$, $\mathbf{v_1}$ and $\mathbf{v_2}$. Verify that all the vectors are orthogonal to each other using their dot products.

Remark 2.4 While the linear algebra chapter in this book is ending, this is a very small subset of a standard linear algebra course. Many concepts, such as eigenvalues, are not covered due to their not so important nature. Readers are still highly encouraged to read up some of the other concepts.

2.6 An aside: Bra-ket and functions as vectors

Bra-ket notation, also known as Dirac notation, is a standard mathematical notation used in quantum mechanics to describe quantum states. It was introduced by physicist Paul Dirac in the 1930s and has since become a fundamental tool in the field. The notation consists of two main components: bras and kets. It is indispensable in quantum mechanics, as it provides a concise and elegant way to represent quantum states, operators, and inner products. It is particularly useful in the context of *Hilbert spaces*, which are complete inner product spaces that serve as the mathematical framework for quantum mechanics. We will build up some intuition for bra-ket notation here because it is a very integral part of (quantum) physics, but a more rigorous treatment can be found in advanced quantum mechanics or linear algebra texts. To start off, the vector $\boldsymbol{\psi}$ in the traditional sense is represented as a *ket*, denoted by $|\psi\rangle$.

Example 2.12 For example, a vector in a two-dimensional complex space can be represented as:

$$|\psi\rangle = \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}$$

where ψ_1 and ψ_2 are complex numbers. The ket represents a column vector in a complex vector space, which is often used to describe the state of a quantum system. Notice that I refrained from using ψ_x and ψ_y because in quantum mechanics, the components of a (state) vector do not necessarily correspond to spatial dimensions. We might use different bases in different scenarios (see next section). I want you to divorce the idea of vectors from spatial dimensions because in quantum mechanics, vectors can represent states that are not directly tied to physical space. Adding vectors is done in the same way as traditional vector addition:

$$|\psi\rangle + |\phi\rangle = \begin{bmatrix} \psi_1+\phi_1 \\ \psi_2+\phi_2 \end{bmatrix}$$

We actually call the addition of two kets a *superposition* (of states), which is a fundamental concept in quantum mechanics. A superposition represents a vector or state that is a combination of multiple possible states, and it is a key feature of quantum systems.

Dot products here are generalized to *inner products*, which are represented using both bras and kets. A *bra* is simply denoted by $\langle\psi|$. But first, what is an inner product?

Definition 2.11 An **inner product** is a mathematical operation that takes two vectors and produces a scalar (a single number). In the context of an n-dimensional complex vector spaces, the inner product of two vectors $|\psi\rangle$ and $|\phi\rangle$ is denoted as $\langle\psi|\phi\rangle$, where $\langle\psi|$ is the bra corresponding to the ket $|\phi\rangle$.

$$\langle\psi|\phi\rangle = \sum_{i=1}^{n} \psi_i^* \phi_i$$

Example 2.13 Basic braket operations Let's say we have two kets:

$$|\psi\rangle = \begin{bmatrix} 1+i \\ 2 \end{bmatrix}, \quad |\phi\rangle = \begin{bmatrix} 3 \\ 4-i \end{bmatrix}$$

Now, we can compute the inner product (braket) between these two states:

$$\langle\psi|\phi\rangle = (1-i)\cdot 3 + 2\cdot(4-i) = 3 - 3i + 8 - 2i = 11 - 5i$$

Lastly, we have three conditions imposed on inner products [1]:

$$\langle\psi|\psi\rangle \geq 0 \tag{2.17}$$

$$\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^* \tag{2.18}$$

$$\langle\psi|(a|\phi\rangle + b|\phi\rangle) = a\langle\psi|\phi\rangle + b\langle\psi|\phi\rangle \tag{2.19}$$

Till now, I have not really introduced a lot of new concepts, but I have just changed the notation. The real power of bra-ket notation comes when we start dealing with operators and transformations in quantum mechanics. Operators are analogous to matrices in linear algebra, and they act on kets to produce new kets. For example, if we have an operator $\hat{A}$, we can represent its action on a ket $|\psi\rangle$ as:

$$\hat{A}|\psi\rangle = |\hat{A}\psi\rangle = |\phi\rangle \tag{2.20}$$

where $|\phi\rangle$ is the resulting ket after the operator A has acted on $|\psi\rangle$. Here is the punchline: I claim that functions are just vectors. Not exactly, because some functions are not easy to work with. For the sake of this argument, consider the polynomials [2] up to degree n. The set of all polynomials up to degree n forms a vector space, or does it? Let's see. Consider two polynomials $p(x)$ and $q(x)$:

$$p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

$$q(x) = b_0 + b_1x + b_2x^2 + \ldots + b_nx^n$$

We can add these two polynomials together:

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \ldots + (a_n + b_n)x^n$$

The resultant polynomial is still a polynomial of degree n or less. We can also multiply a polynomial by a scalar c:

$$c \cdot p(x) = ca_0 + ca_1x + ca_2x^2 + \ldots + ca_nx^n$$

Again, functions behave just as if they were vectors. The zero polynomial (where all coefficients are zero) acts as the zero vector. The additive inverse of a polynomial $p(x)$ is simply $-p(x)$, which is also a polynomial of degree n or less. Thus, the set of all polynomials up to degree n satisfies all the properties of a vector space. However, the astute among you may have noticed that I have not defined an inner product for these polynomials yet. We can define an inner product for polynomials as follows (though I have to start using calculus, so feel free to skip this part until you are down with Calculus II):

Definition 2.12 The **inner product** of two polynomials $p(x)$ and $q(x)$ over the interval $[a, b]$ is defined as:

$$\langle p|q\rangle = \int_a^b p(x)^* q(x)dx$$

where $p(x)^*$ denotes the complex conjugate of $p(x)$.

[1] You should try to prove these properties for yourself.

[2] I bet that now the part of your brain that does geometry must be yelling: this is moonshine, how can polynomials be vectors! Polynomials do not have "direction". The truth is that the geometric interpretation of vectors can be generalized for other mathematical objects. As you will see for yourself, many of the common operations for geometric vectors can be generalized for functions.

Two key properties that physicists exploit are *orthogonality* and *normalization*. Two functions $f(x)$ and $g(x)$ are said to be orthogonal over an interval $[a, b]$ if their inner product is zero:

$$\langle f|g\rangle = \int_a^b f(x)^* g(x)dx = 0 \tag{2.21}$$

A function $f(x)$ is said to be normalized over an interval $[a, b]$ if its inner product with itself is equal to one:[1]

$$\langle f|f\rangle = \int_a^b |f(x)|^2 dx = 1 \tag{2.22}$$

With these two properties, a set of functions are *orthonormal* if they are normalized and orthogonal to each other. [2] For a set of orthonormal functions $\{f_i(x)\}$, we have:

$$\langle f_i|f_j\rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{2.23}$$

Armed with these concepts, we can revisit an old problem in Chapter 1, Problem 1.15. The wave function $\psi_n(x)$ in this case is

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right) \tag{2.24}$$

These wave functions are actually orthogonal (over the interval $[0, a]$) wherever $p \neq q$ in the sense that

$$\begin{aligned}
\int_0^a \psi_p \psi_q dx &= \frac{2}{a} \int_0^a \sin\left(\frac{p\pi}{a}x\right) \sin\left(\frac{q\pi}{a}\right) dx \\
&= \frac{1}{a} \int_0^a \cos\left(\frac{p-q}{a}\pi x\right) - \cos\left(\frac{p+q}{a}\pi x\right) dx \\
&= \left(\frac{1}{(p-q)\pi} \sin\left(\frac{p-q}{a}\pi x\right) - \frac{1}{(p+q)\pi} \sin\left(\frac{p+q}{a}\pi x\right)\right)_0^a \\
&= \frac{1}{\pi}\left(\frac{\sin((p-q)\pi)}{p-q} - \frac{\sin((p+q)\pi)}{p+q}\right) \\
&= 0
\end{aligned}$$

Because of normalization [3], we also have

$$\int_0^a |\psi_n|^2 dx = 1 \tag{2.25}$$

Hence, the set of ψ_n are actually orthonormal. Why do we care so much about orthonormality in the first place? Well, because of something known as *Fourier's trick*. You see, many solutions to the infinite square well is not a simple ψ_n, but rather a linear combination or superposition of many ψ_n:

[1] Do take note that $|f(x)| = f(x)f(x)^*$.

[2] The δ is the Kronecker delta, which is 1 if the indices are equal and 0 otherwise.

[3] Recall that $|\psi|^2$ gives us the probability at x, hence the probability must add up to 1, therefore we have the following integral.

$$f(x) = \sum_{n=1}^{\infty} c_n \psi_n(x) \tag{2.26}$$

The question is, how do we find the coefficients c_n? This is where orthonormality comes in. We can multiply both sides by $\psi_m(x)$ and integrate over the interval $[0, a]$:

$$\int_0^a f(x)\psi_m(x)dx = \int_0^a \left(\sum_{n=1}^{\infty} c_n \psi_n(x)\right) \psi_m(x)dx$$

We can interchange the sum and the integral:

$$\int_0^a f(x)\psi_m(x)dx = \sum_{n=1}^{\infty} c_n \int_0^a \psi_n(x)\psi_m(x)dx$$

Due to orthonormality, all terms in the sum vanish except for the term where $n = m$:

$$\int_0^a f(x)\psi_m(x)dx = c_m \int_0^a |\psi_m(x)|^2 dx$$

Since $\psi_m(x)$ is normalized, the integral on the right side equals 1:

$$\int_0^a f(x)\psi_m(x)dx = c_m$$

Thus, we have derived a formula for the coefficients c_n:

$$c_n = \int_0^a f(x)\psi_n(x)dx \tag{2.27}$$

Just by the simple properties of orthonormality, we have derived a powerful method to find the coefficients of a function expressed as a linear combination of orthonormal basis functions.

Problem 2.23 (1) Given that $|\psi\rangle = 1\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$ and $|\phi\rangle = 3\hat{\mathbf{i}} + 4\hat{\mathbf{j}}$, find

a) $\langle\psi|\phi\rangle$
b) $|\psi\rangle + |\phi\rangle$
c) Hence, find $\langle\phi|\hat{Q}(|\psi\rangle + |\phi\rangle)$ if $\hat{Q} = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right]$

2.7 An aside: Eigenvalues and Eigenfunctions

Part II
Calculus I

Chapter 3
Derivatives

3.1 Essence of calculus

Perhaps framing an established pedagogical subject like calculus through the unconventional lens of physics is not optimal, but I will try my best. One of the most infamous problems in physics is thermal conduction, which I will frame as such for now

Problem 3.1 Bring two identical one-dimensional rods with different temperatures into contact such that they line up in a straight line, what will be the temperature distribution at some time t?

The topic of heat conduction proved to be very much non-trivial, even Newton's solution to the problem only holds for small temperature differences. However, the breakthrough would only come more than a hundred years later, when Joseph Fourier, equipped with many advancements in the study of calculus, tackled the problem with the infamous **Fourier series**. Tackling the heat conduction problem is multi-layered, requiring tools from each level of calculus. Hence, this is a rich problem no matter which stage of calculus one is at.

The core of deriving what will be known as the **heat equation** is by analyzing the small sections of the whole system, tiny sections, or *infinitesimal* sections. Infinitesimal refers to a quantity that is smaller than any positive real number, but not zero, which sounds weird at first. Remember this sentence, this implies that infinitesimals are *not* real numbers. Calculus is the study of infinitesimal changes, and through analyzing small changes, we can understand a much larger picture. It turns out, a large class of problems in physics and even other sciences can be solved by the methods of calculus. It is surprisingly powerful how we can essentially utilize tiny changes to understand the whole. But enough of philosophical musings, let's get to the math.

Definition 3.1 The **limit** of a function $f(x)$ as x approaches a is the value that $f(x)$ approaches as x gets arbitrarily close to a. We write

$$\lim_{x \to a} f(x) = L \tag{3.1}$$

where L is the limit of $f(x)$ as x approaches a. [1]

[1] Actually, $f(a)$ does not even have to exist, it can be undefined.

Instead of plunging into evaluating limits immediately, I want to motivate the reader to understand the origin of the need for limits. Take the following function as an example [1]

$$f(x) = \frac{\sin(x)}{x} \tag{3.2}$$

I am about to pose a completely nonsensical question, what is $f(0)$? Easy, inasmuch as there is a 0 in the denominator, $f(0)$ is undefined. That is not wrong, but perhaps when we plot the graph of f, we shall notice something interesting

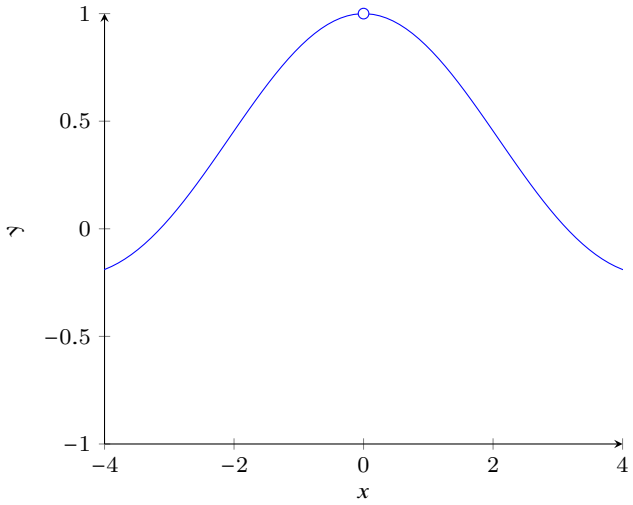

Fig. 3.1 Plot of $f(x)$.

Supposing you are the first one to notice that f seems to approach 1 as x goes to 0. You would be absolutely correct. Turns out, the limit as x approaches 0 is 1, which we write as

$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1 \tag{3.3}$$

Using limits, we are able to "eliminate" this discontinuity in the function. Limits enable us to study behavior of functions that otherwise would not be possible by conventional means. Of course, I have not shown why this limit is 1 or that it even *exists*. While the exact derivation will be left to the later chapters, we can use SageMath to help us evaluate this limit. To put it simply, Sage is a mathematics software system like Mathematica, but unlike many of its counterparts, it is free and open source. To evaluate this limit in Sage, we use the following syntax (enter the following code line by line):

```
sage: f(x)=sin(x)/x
sage: f.limit(x=0)
x |--> 1
```

[1] This is actually the sinc function, used commonly in signal processing.

Line 1 defines the function and line 2 evaluates the limit of f as x approaches 0. Expectedly, Sage yields the value 1 for the limit. We shall be using Sage for the rest of the book so do read up on the syntax of Sage. A quick Google search should suffice. Coming back to calculus, treat limits as an integral part (pun intended) of calculus, but the nitty-gritty shall be brushed aside for now. All in all, remembering that infinitesimals like x in Equation (3.3) are not real numbers, but are special quantities that enable us to reason, with rigor in due time, a lot of problems.

3.2 Basic heat flow

Coming back from limits, we return to the two rods. At some point in time, the heat distribution may look something like this[1]

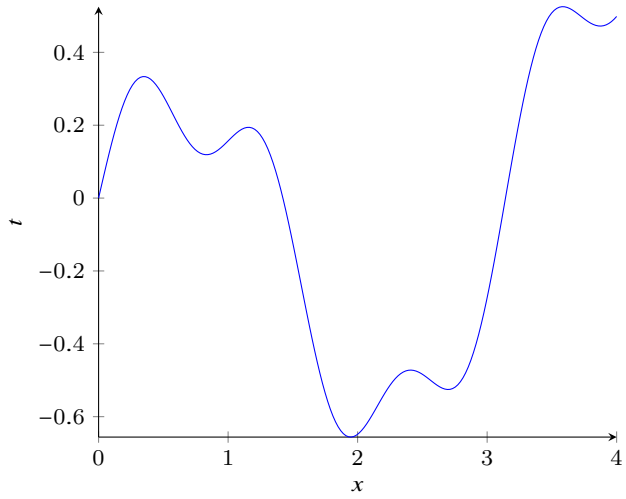

Fig. 3.2 Possible distribution of heat.

As you may see, the heat distribution may look awfully complex, even in one-dimensions. Heck, we did not even include time into this graph. Suppose I take a "leap of faith", and treat this continuous distribution with discrete points (Figure 3.3).

Zooming in to an individual point on Figure 3.3, say the third point from $x = 0$, we notice it is surrounded by two points of lower temperature. From this, we can probably infer that if we let time pass, the temperature at this point will probably decrease. How about another point, for example the point right above the $x = 3$ mark. It is surrounded by a lower point and a higher point. However, since the difference in temperature between the higher point and it is lower than the difference between it and the lower point, it will probably increase in temperature as time passes. Hence, we can generalize that

Lemma 3.1 *Given equally spaced points x_1, x_2, x_3 on the heat distribution graph where $x_1 < x_2 < x_3$, define the function T such that it yields the temperature at the point x. If the average temperature of x_1 and x_3, denoted by $\frac{T(x_1)+T(x_3)}{2}$, is larger than the temperature at x_2, then x_2 will heat up. Likewise, if the average temperature is lower, x_2 will tend to cool down.*[2]

Bravo, this at least provides some insight, albeit very qualitatively, to how the distribution of heat may change. But, we can definitely do better than this. Imagine a cup of hot tea cooling down in a room, the hotter the tea, the faster the tea cools down at that point in time. Conversely, if the tea is cooler, then the rate of heat transfer is also going to be lower. Using this analogy, I may write

[1] Of course, it might not look like this, but I am just using this graph as an example to demonstrate the complexity of the heat distribution.

[2] For the sake of completeness, the heat distribution function $T(x,t)$ also varies with time, which is explained later.

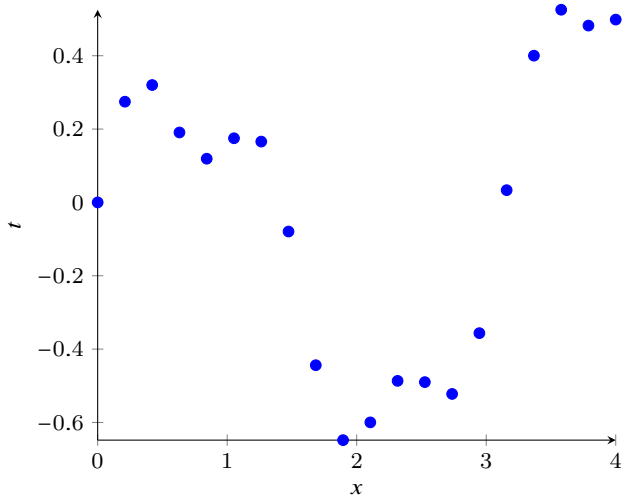

Fig. 3.3 Discrete distribution of heat.

Lemma 3.2 *The rate in which x_2 heats up is proportional to the average of x_1 and x_3 minus the temperature of x_2 itself, or* [1]

$$\frac{\Delta T(x_2,t)}{\Delta t} = \alpha\left(\frac{T(x_1,t)+T(x_3,t)}{2} - T(x_2,t)\right) \tag{3.4}$$

Equation (3.4) gives us some insight into why heat transfer is so complicated, even in the one-dimensional case. Notice that $T(x,t)$ varies with two variables, space and time. If Equation (3.4) only varies with time or space, the problem would be dumb down by a lot. We can, however, do this by shifting our focus to only one point on the curve, throwing spatial variation out of the window. This does not mean that the heat distribution will simplify into one point, but we shall look at one point at any given time.

Problem 3.2 (2) Solve for two possible $T(x,t)$ such that they satisfy Equation (3.4). (Hint: Try very trivial functions.) What do such functions imply physically?

[1] Essentially what the fraction is saying is the change of T over a certain time period Δt, also known as rate of change.