Compact Calculus

with relation to existence

Puripat Thumbanthu

1E

Prerequisites: *set theory, algebra, geometry, basic trigonometry*

Contents

Preface

Acknowledgement

Reading Guide

Big disclaimer. I am a physicist, and this is calculus from the physicist's point of view. So, there's going to be some contents from physics that pure mathematicians might not adore, I'm terribly sorry for that.

The abstract contains the guide, the overview, and the mindset of the chapter. *Reading the abstract is a necessity.*

The chapter contains the main idea of the topic. A chapter might have supplementary unnumbered chapters or sections which is optional.

Appendices either supplements or extend the chapter. Some even goes much beyond the chapter. They all vary. It's recommended to study them separately.

The interludes provide a *historical background* for the next chapter. It's meant to connect one chapters with the other. It acts as a storyline bridge. If not taken, the next chapter might seem too terse. So, I recommend the reader to skim through the fruitfulness of historical development.

This book is separated into five parts

I The fundamentals

II The applications

III The extensions

IV The foundations, reimagined: real analysis

V Beyond imagination: complex analysis

Part I (The fundamentals) focuses on the basics of calculus: from derivatives to antiderivatives and some of its applications. I've swapped around the order of contents a lot as I see fit. I've also introduced some applications that's not found of normal pedagogy.

Part II (The applications) focuses on further applications calculus to real world problems, mostly in physics.

Part III. (The extensions) explores the realm of specialized calculus, most of the aren't even taught in universities: newer branches of calculus.

Part IV. (Real analysis) and *Part V. (Complex analysis)* as the name suggests, explores the full behaviors of real and complex functions. It reconsiders all the basics of calculus and dives deep into the backbone of all symbols that are abstracted away from the physical world.

At last, a fair warning; not all contents follow accurate historical order.

Part I

The foundations

CHAPTER 1

Writing down nature: derivatives

1.1 Invitation to calculus from a physicist

Prerequisites: *graphs, functions, kinematic variables*

I might've made many mathematicians angry by writing this book. Calculus

For me, calculus is a way to describe a physical system by mathematical equations. Physics is all about how a physical quantity changes with time, whether it be how things move or disperses. We then solve these equations to determine the future of a system. In this chapter, we're going to focus on how we can mathematically describe a system, and we'll see how the attempt naturally give rise to *derivatives*: a mathematical measure of rate of change.

At the end, you should develop the intuition that

1. Derivative measures the rate of change of a function w.r.t. a variable.
2. Derivatives can be thought of the slope of a graph.
3. The universe is described in the language of differential equations

1.1.1 The notation of differential calculus

If a is a variable, then $\mathrm{d}a$ is a very small quantity of a a.k.a. an **infinitesimal**. E.g., if $\mathbf{x}$ is displacement, then $\mathrm{d}\mathbf{x}$ is a very small displacement. If t is the time, then $\mathrm{d}t$, a very short time.

We can also find ratio between infinitesimal. E.g., the ratio between some small distance and some short time, $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$, we get the speed $\mathbf{v}$.

1.2 Speed and instantaneous rate of change

To set things off, let's think about how a **speedometer** work. If we're traveling at a speed $5\,\mathrm{m\,s^{-1}}$, when *exactly* are we traveling at $5\,\mathrm{m\,s^{-1}}$? You could say $\mathbf{v} = 5\,\mathrm{m}$ *at the moment of measurement*. That's like saying, "Oh, I can find the speed of the car by just taking a picture of it." But that's illegal! To calculate speed, we have to compare two points in space through time. Or, the rate of change of distance through time:

$$\mathbf{v} = \frac{\mathbf{s}_2 - \mathbf{s}_1}{t_2 - t_1}. \tag{1.1}$$

While it may seem like cameras grab snapshots in an instant, they actually need time to take in light to construct an image, they need some *exposure time* Δt.

To get a "not blurred" image of a moving object, we reduce the exposure time. If the exposure time is too long, the object will be smeared out. This effect is known as motion blur, which is normally undesired. But motion blur actually helps out a lot with measuring velocity. In which the velocity is just

$$\mathbf{v} = \frac{\text{Distance between the smear and the main object}}{\text{Exposure time}}. \tag{1.2}$$

As illustrated in fig. 1.1, the motion blur clearly shows us the change in position of the car over the exposure time.

FIG. 1.1 | CALCULATING THE VELOCITY OF A CAR FROM MOTION BLUR.

But what will happen with shorter exposure time? Does the motion blur disappear? No! The blur is still there, but it's just smaller. Typically, 12 ms exposure time is short enough to create a "focused image". But it's just an illusion that came from the limitation of the screen's ability to reproduce such little blur. If our camera and screen is good enough, we can *always* calculate the velocity of the car from the blur. The smaller the exposure time, the more detailed the image is, and the closer you'll get to the exact $\mathbf{v}$ at that moment in time. Finally, if we let the exposure time become infinitesimally short, we can say the $\mathbf{v}$ we got is *the velocity at that exact point*. Or as we call it, the **instantaneous velocity**. By using the said calculus notation in section 1.1.1, we can just write this as

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}, \tag{1.3}$$

and here it is ladies and gentlemen, the **derivative**: the measure of rate of change.

From here on out, I shall use *derivative* and *rate of change* interchangeably. So, every time you see *derivative*, think *rate of change*.

1.3 An attempt to define derivatives

Mathematically, a **derivative** is a measure of a function's rate of change with respect to a variable. In the previous example, $\mathbf{x}$ is a function that's dependent on time, and its derivative w.r.t. time, $\mathbf{v}$, is measuring the rate of change of function $\mathbf{x}(t)$ through time $\mathrm{d}t$.

To find an explicit expression for derivative, let's say we have two points in spacetime $(\mathbf{x}_1, t_1)$ and $(\mathbf{x}_2, t_2)$. The change in $\mathbf{x}$ is $\mathbf{x}_2 - \mathbf{x}_1$, and the change in time is $t_2 - t_1$. The derivative of position w.r.t. time is then just

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \frac{\mathbf{x}_2 - \mathbf{x}_1}{t_2 - t_1}. \tag{1.4}$$

Let $t_2 = t + h$. Then, $\mathbf{x}_2 = \mathbf{x}(t+h)$ and $\mathbf{x}_1 = \mathbf{x}(t)$. The time difference used to calculate $\mathbf{v}$ must be miniscule: infinitesimally small. I have to introduce the notion of limits, which is just a fancier way of saying "very close to, but not"

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \lim_{h\to 0} \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{(t+h) - t} \tag{1.5}$$

$$= \lim_{h\to 0} \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{h}. \tag{1.6}$$

The equation above is what we generally refer to as the *definition of derivative*. Then, we just extend this relation to any function $f(x)$, which requires just a substitution of variables. And then we get:

Definition 1: Naive definition of derivative

A derivative of a function $f(x)$ w.r.t. a variable x is the rate of change of $f(x)$ w.r.t. x, and it is written as

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} \quad \text{or,} \quad \frac{\mathrm{d}}{\mathrm{d}x}f(x), \tag{1.7}$$

where

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}. \tag{1.8}$$

Equation (1.8) directly reads:

> The derivative of $f(x)$ with respect to x is $\dfrac{f(x+h)-f(x)}{h}$ where h is very close to 0, but h is strictly not 0.

On notation: So far, we've been using the Leibniz's notation for derivatives, and it has the property that derivatives behave exactly like fractions, and you can cancel terms.

$$\frac{\mathrm{d}a}{\mathrm{d}b} \times \frac{\mathrm{d}b}{\mathrm{d}c} = \frac{\mathrm{d}a}{\mathrm{d}c}. \tag{1.9}$$

But, this isn't the only accepted notation. Multiple great mathematicians have come up with their own, and some are better than others in certain cases. I'd introduce other notations later on if necessary.

1.4 The geometrical interpretation of the derivative

One way to interpret derivatives is by using slope. Notice that eq. (1.4) looks a lot like the slope equation

$$m = \frac{y_2 - y_1}{x_2 - x_1}. \tag{1.10}$$

We've also seen that eq. (1.4) is analogous to the definition of derivative eq. (1.8). So is the derivative just the slope of a line? If it is, then what is the y-axis and the x-axis of a graph?

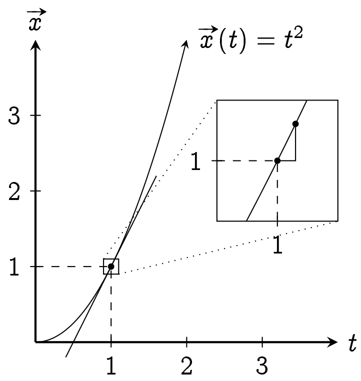


FIG. 1.2 | POSITION VS. TIME GRAPH WHERE $\mathbf{x}(t) = t^2$.

We can compare eq. (1.10) to eq. (1.4): the y-axis should be the position $\mathbf{x}$, and the x-axis, the time t. If we draw that out, we'll get the x-t graph, which can encode the exact trajectory of an object. An example of which is shown in fig. 1.2.

But what does this have to do with derivatives? Equation (1.10) only works for straight line! Well, here's the beauty of it. If you zoom into any points on a curve, eventually, it will look like a line. And thus, *the derivative zooms into the curve at some point, and chooses two very close points on the curve and calculate its slope. In which, that slope represents the rate of change of the function at that point.*

1.5 Evaluation of derivatives: method of increments

The derivative's definition can be used to directly evaluate derivatives. This is called the **method of increment**. E.g., in fig. 1.2, let's evaluate the velocity at time $t = 1\,\mathrm{s}$ where the position function, $\mathbf{x}(t) = t^2$. $t = 1\,\mathrm{s}$. We start from eq. (1.8):

$$\begin{aligned}
\mathbf{v}(1) = \frac{\mathbf{x_2} - \mathbf{x_1}}{t_2 - t_1} &= \lim_{h \to 0} \frac{\mathbf{x}(1+h) - \mathbf{x}(1)}{(1+h) - 1} \\
&= \lim_{h \to 0} \frac{(1+h)^2 - 1}{h} \\
&= \lim_{h \to 0} \frac{2h + h^2}{h} \\
&= \lim_{h \to 0} 2 + h\,.
\end{aligned}$$

Since h is very close to zero, we approximate $2 + h$ as 2. Imagine comparing 2 to 10^{-10}. The 10^{-10} wouldn't make a noticeable difference, and we can ignore it. Therefore, $\mathbf{v}(t = 1\,\mathrm{s}) = 2\,\mathrm{m\,s^{-1}}$.

Now, try evaluating $\mathbf{v}(t = 3\,\mathrm{s})$ for $\mathbf{x}(t) = t^3$. You should get $81\,\mathrm{m\,s^{-1}}$. As a hint, you can also ignore h^2 because if $h < 1$, then $h^2 < h$.

1.6 Higher order derivatives

Order	Name
1	Velocity/Speed
2	Acceleration
3	Jerk
4	Snap/Jounce
5	Crackle
6	Pop

TABLE 1.1 | HIGHER ORDER DERIVATIVES OF POSITION W.R.T. TIME

In kinematics, there are a whole set of quantities that can describe an object's trajectory, e.g., the acceleration, which is defined to be the rate of change of velocity w.r.t. time:

$$\mathbf{a}(t) = \frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t}.$$

But the $\mathbf{v}$ is also the rate of change of position w.r.t. time. Thus,

$$\mathbf{a}(t) = \frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}\mathbf{r}(t)}{\mathrm{d}t}\right) = \frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2},$$

where $\mathbf{r}$ is any position vector.

The $\mathrm{d}^2\mathbf{r}$ and $\mathrm{d}t^2$ is just a matter of symbolic manipulation and should only be interpreted as just a shorthand. $\mathbf{a}$ is called the **second order derivative** of $\mathbf{r}$ because you've differentiated $\mathbf{r}$ twice. **Higher order derivatives** of position w.r.t. time is listed in table 1.1.

1.7 Expressing nature: basic differential equations

The dynamics of a physical system is universally described by the famous Newton's second law $\mathbf{F} = m\mathbf{a}(t)$, which in derivatives form becomes:

$$\mathbf{F} = m\frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2}. \tag{1.11}$$

Let's try to describe a simple system with this. In fig. 1.3, a ball is dropped from height h. The Earth's gravity pulls the ball with force mg where m is the mass of the ball, and $\mathbf{g}$, the acceleration from Earth's gravity. Newton's second law tells us that

$$\mathbf{F} = m\frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2}$$

$$m\mathbf{g} = m\frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2} \quad \text{or,} \quad \mathbf{g} = \frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2}, \tag{1.12}$$

which is sometimes called the **equation of motion**, which packs every information about this system you'd ever want. It directly reads as:

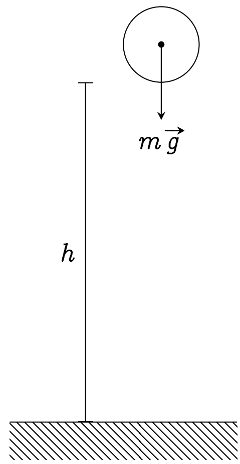


FIG. 1.3 | A BALL DROPPED FROM HEIGHT h

The acceleration of the ball is equals to g.

If you've stayed for this long, congratulations! You now have the power to describe every physical systems with mathematical equations using derivatives: the measure of the rate of change. But it might not be as useful as yet, just as a hammer may seem useless if used to paint, derivatives falls apart when you ask about the future of the system. E.g., how long the ball takes to reach the ground? Or, what's the position of the ball at a certain time? That's the job of the integral to solve, and we'll do so in the next chapter.

1.8 Conclusion for Chapter 1

1. The concept of approaching can be used to bypass dividing by zero.
2. Derivatives are rate of change of a function w.r.t. a variable which can be evaluated by the method of increments.

3. Derivatives can be thought as the slope of a graph, or the tangent to a curve.
4. Physical systems can be described by differential equations of different forms. One of them is the Newton's formulation stated in eq. (1.11)

Remarks on chapter 1.1

1. In section 1.2, we zoomed in on the graph to approximate the function as a line. Actually, this is quite literally the whole idea of derivatives. If we dig in further in calculus, sometimes the rate of change analogy doesn't even make sense. However, saying that the derivative tries to approximate every function as a line works in all scenario. Though, it's quite abstracted away from the world.

CHAPTER 2

Integrals and antiderivative

Prerequisites: *chapter 1, sigma summation notation*

Terminologies: Every line, including straight, is a **curve**. You might see other textbooks use the term "*area under the graph*" to refer to integrals. But, a **graph** is *a diagram consisting of a line or lines, showing how two or more sets of numbers are related to each other* [1], not the curve itself. Therefore, I'll refer to a curve as any line that connects two points, whether straight or not. Now let's start.

2.1 Invitation: mission impossible

2.1.1 The mindset of integral calculus

In the previous chapter, we've learned how to describe the universe using derivatives. But derivatives falls short when we want to predict the future of a system. But, what do we mean when we say "predict the system?"

Literally everything that involves a quantity changing is written in the language of derivatives. When derivatives are used in equations to describe a system, it becomes *a differential equation*. Its solution contains the future of the entire system in question. However, it's quite hard to solve; some even impossible.

In classical physics, a **state** represents the configuration that the system *at one point in time*. To predict the system, we need to know the initial state of a system. Classical physics says that if you know the rules that the system plays by (In this case, $\mathbf{F} = m\dfrac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}$), and the initial condition, we can always determine the state of the system at any point in time. That is, classical physics guarantees that there is always a function, which takes in time as input, and outputs the state of a system. And in order to predict the system, we must know that function. But what may that function be?

Consider the example from chapter 1. The equation of motion of the ball dropped from a height h is a second-order differential equation

$$g = \frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2}, \tag{2.1}$$

with initial condition being $\mathbf{r} = h$. The function that describes the state is $\mathbf{r}(t)$, which outputs the position of the ball at a certain time t. You can see that this function is the derivative sign. To solve the differential equation for this function, we have to isolate it out, and undo the derivative sign. But how?

It seems impossible at first. We only know that $\mathbf{r}(t)$ must satisfy eq. (2.1), i.e., the second derivative of $\mathbf{r}(t)$ is g. It's like we have to search through a gigantic pool of functions that mathematics have to offer to find a single function $\mathbf{r}(t)$ that satisfies eq. (2.1). It'd be like finding a needle in the haystack!

Of course, this is a textbook, there must be a solution. If there is a will, there is a way. You might have to reverse engineer derivatives, which might look tedious at first. But well, you might find something interesting along the way.

2.1.2 Brief notation of integral calculus

The $\int$, a.k.a. the **integral**[1], means to sum. This integral symbol is basically sigma summation symbol, but for infinitesimals. Therefore, we have bounds called **integral bounds**. E.g., if we sum a lot of little time step $\mathrm{d}t$ together from t_A to t_B we get $t_B - t_A$, the total time step. Thus,

$$\int_{t_A}^{t_B} \mathrm{d}t = t_B - t_A. \tag{2.2}$$

2.2 Finding a function in the haystack

Equation (2.1) has a second order derivative, let's go slowly and undo one derivative at a time. We'll undo the derivative of the R.H.S. to get the velocity first. Then, we'll undo the derivative again to get the position.

2.2.1 Step one: the velocity function from acceleration

One good strategy for solving any kind of equation is **separation of variables**. We isolate the variable we're interested in solving, which is $\mathbf{r}$, and move everything else to the other side. Here, write $\frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2}$ as $\frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t}$. Then, isolate $\mathbf{v}$ on one side and move $\mathbf{t}$ to the other.

$$\mathbf{g} = \frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t} \tag{2.3}$$

$$\mathrm{d}\mathbf{v}(t) = \mathbf{g}\,\mathrm{d}t. \tag{2.4}$$

[1] Which also looks like a beansprout

This equation reads

> *A small change in velocity* $\mathrm{d}\mathbf{v}$ *is product of* $\mathbf{g}$ *and a small time interval* $\mathrm{d}t$.

To find the total change in velocity, we sum up a lot of small changes in velocity $\mathrm{d}\mathbf{v}(t)$, which is equal to $\mathbf{g}\,\mathrm{d}t$. Because $\mathrm{d}v$ is directly proportional to $\mathrm{d}t$, the total change in $\mathbf{v}$ is simply $\mathbf{g}t$. However, *changes* doesn't say anything about the initial condition, so we add a term C to compensate. Therefore,

$$\mathbf{v}(t) = \mathbf{g}t + C. \tag{2.5}$$

To find what C is, just plug in the initial condition. If $\mathbf{v}(t = 0) = \mathbf{v}_0$, then

$$\mathbf{v}(t = 0) = v_0 = \mathbf{g} \times 0 + C \tag{2.6}$$

$$v_0 = C. \tag{2.7}$$

Thus,

$$\mathbf{v}(t) = \mathbf{g}t + \mathbf{v}_0. \tag{2.8}$$

We can also express these ideas symbolically using the integral symbol (section 2.1.2) as

$$\mathbf{g} = \frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t} \tag{2.9}$$

$$\mathrm{d}\mathbf{v} = \mathbf{g}\,\mathrm{d}t \tag{2.10}$$

$$\int \mathrm{d}\mathbf{v} = \int \mathbf{g}\,\mathrm{d}t \tag{2.11}$$

$$\mathbf{v} = \mathbf{g}t + \mathbf{v}_0. \tag{2.12}$$

2.2.2 Step two: the position function from velocity

Rewrite $\mathbf{v}$ as $\dfrac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}$, then do separation of variables.

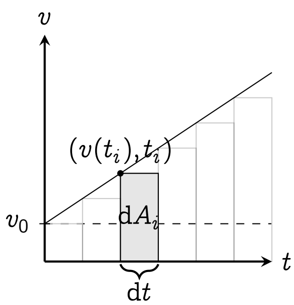
$$\mathbf{v} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{g}t + \mathbf{v}_0 \tag{2.13}$$

$$\mathrm{d}\mathbf{r} = \mathrm{d}t(\mathbf{g}t + \mathbf{v}_0), \tag{2.14}$$

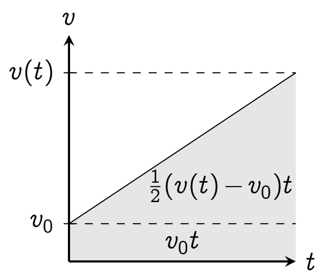
which reads,

A small change in position $\mathrm{d}\mathbf{r}$ *is the product of* $\mathbf{g}t + \mathbf{v}_0$ *and a small time interval* $\mathrm{d}t$.

If we want to find the total change in position $\mathbf{r}$, we just have to sum all $\mathrm{d}\mathbf{r}$'s. This time, it's not as obvious, because there's $\mathbf{g}t + \mathbf{v}_0$, which is also changing with time. So at each point in time, the rate of change of position is different.



(A) A SUBDIVIDED GRAPH



(B) A GRAPH SHOWING THE AREA OF THE GRAY TRAPEZOID BROKEN INTO TWO PARTS

FIG. 2.1 | A v-t GRAPH OF A BALL DROPPED FROM A BUILDING

A good strategy in math if you don't know what to do is to just graph the function. The graph of eq. (2.13) is shown in fig. 2.1a. A small time interval represents a little step in the t axis. The curve shown in the graph represents $\mathbf{g}t + \mathbf{v}_0$. Therefore, $\mathrm{d}t(\mathbf{g}t + \mathbf{v}_0)$ would just represent an area of a little rectangle $\mathrm{d}A$ as shown in the figure.

The total change in position is the sum all those rectangles. When $\mathrm{d}t \to 0$, the sum of all $\mathrm{d}t(\mathbf{g}t + \mathbf{v}_0)$ approaches the area under the graph, which can be calculated geometrically as shown in fig. 2.1b. Thus,

$$\mathbf{r} = \frac{1}{2}(\mathbf{v}(t) - \mathbf{t}_0)t + \mathbf{v}_0 t + C \tag{2.15}$$

$$= \frac{1}{2}(\mathbf{g}t - \mathbf{t}_0)t + \mathbf{v}_0 t + C \tag{2.16}$$

$$= \frac{1}{2}\mathbf{g} \times t^2 + \mathbf{v}_0 t + C \tag{2.17}$$

When $t = 0, \mathbf{r} = \mathbf{r}_0$. Therefore,

$$\mathbf{r}_0 = \frac{1}{2}\mathbf{g} \times 0^2 + \mathbf{v}_0 \times 0 + C$$
$$\mathbf{r}_0 = C.$$

Therefore,

$$\mathbf{r}(t) = \frac{1}{2}\mathbf{g}t^2 + \mathbf{v}_0 t + \mathbf{r}_0. \tag{2.18}$$

To model the trajectory of the ball, we set

1. $\mathbf{v}_0 = 0$ (object is dropped and starts at zero speed)
2. $\mathbf{r}_0 = 0$ (convenient initial condition placement)

thus we get,

$$\mathbf{r} = \frac{1}{2}\mathbf{g}t^2 \implies t = \sqrt{\frac{2\mathbf{r}}{\mathbf{g}}} \tag{2.19}$$

Since the ground is at $\mathbf{r} = h$, the time that the ball hits the ground is then $\sqrt{2h/\mathbf{g}}$.

2.2.3 Conclusion: area under the curve and antiderivative

The form of the differential equation that we've solved is in the form $\frac{\mathrm{d}x}{\mathrm{d}t} = f(x)$. We have to undo the derivative, and the simplest way is to do separation of variables and turn the equation into

$$\mathrm{d}x = f(x)\,\mathrm{d}t, \tag{2.20}$$

which reads,

> A small change in x, i.e., $\mathrm{d}x$ is represented by the area of a rectangle width $\mathrm{d}t$ and height $f(x)$.

And, the total change x is represented by the sum all those little rectangles, which is the area under the curve $f(x)$. Then, we add a constant C to compensate for the initial condition. In symbolic form introduced in section 2.1.2, it's just

$$\int \mathrm{d}x = \int g(x)\,\mathrm{d}t \tag{2.21}$$

$$x = \int g(x)\,\mathrm{d}t. \tag{2.22}$$

For now, we could say that integration is the reverse of derivatives. But to clearly see how this is linked for every function, we must study the fundamental theorem of calculus, which is the bridge between integration and differentiation.

Before we go there, let me clarify some terminologies. An **integral** refers to the area under the curve evaluated between two points. We say that an integral must have an **integral bound**. If we want to find the area under a function $f(x)$ from $x = a$ to $x = b$, we write it as

$$A = \int_a^b f(x)\,\mathrm{d}x. \tag{2.23}$$

This just reads

> The area A under the curve $f(x)$ from $x = a$ to $x = b$ is equal to the sum of the area of many thin stripes width $\mathrm{d}x$ height $f(x)$ that lies between $x = a$ and $x = b$.

The **antiderivative** however, refers to the function which takes in a value a, and output the integral of $f(x)$, evaluated from 0 to a. Therefore, if $A(a)$ is the antiderivative of $f(x)$, then

$$A(a) = \int_0^a f(x)\,\mathrm{d}x. \tag{2.24}$$

It also implies that if the area between a and b can be evaluated by

$$\int_a^b f(x)\,\mathrm{d}x = A(b) - A(a). \tag{2.25}$$

2.3 The fundamental theorem of calculus

The intuition of fundamental theorem of calculus states that derivatives and integrals are essentially inverse of each other. In this section, I'll clarify this fact and make it more rigorous.

Theorem 1: The (first) fundamental theorem of calculus If a function $f(x)$ has an antiderivative $A(x)$, then

$$\frac{\mathrm{d}A(x)}{\mathrm{d}x} = f(x). \tag{2.26}$$

If $A(x)$ is the antiderivative of $f(x)$, then

$$\int_0^x f(x)\,\mathrm{d}x = A(x). \tag{2.27}$$

The *actual* area of one of the stripes (not rectangles) width $\mathrm{d}x$ shown in fig. 2.2, it's obviously $A(x+\mathrm{d}x) - A(x)$. The riemann sum approximation approximates the area by a small rectangle area $f(x)\,\mathrm{d}x$. We can write the

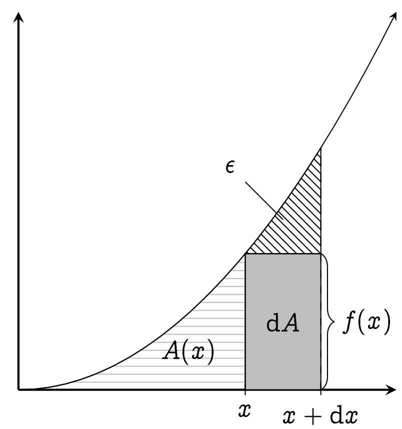


FIG. 2.2 | THE GEOMETRICAL INTERPRETATION OF PART ONE OF THE FUNDAMENTAL THEOREM OF CALCULUS (THEOREM 1).

relation between the actual and the approximated area as

$$A(x + \mathrm{d}x) - A(x) \approx f(x)\,\mathrm{d}x. \tag{2.28}$$

To turn this into an equality, we add a correction term ϵ

$$A(x + \mathrm{d}x) - A(x) = f(x)\,\mathrm{d}x + \epsilon. \tag{2.29}$$

If we let $\mathrm{d}x \to 0$, ϵ is negligible compared to $f(x)\,\mathrm{d}x$; therefore,

$$\lim_{\mathrm{d}x \to 0} A(x + \mathrm{d}x) - A(x) = \lim_{\mathrm{d}x \to 0} f(x)\,\mathrm{d}x. \tag{2.30}$$

Since $f(x)$ is not a variable that's controlled by the limit sign,

$$\begin{aligned}
\lim_{\mathrm{d}x \to 0} A(x + \mathrm{d}x) - A(x) &= f(x) \lim_{\mathrm{d}x \to 0} \mathrm{d}x \\
\lim_{\mathrm{d}x \to 0} \frac{A(x + \mathrm{d}x) - A(x)}{\mathrm{d}x} &= f(x).
\end{aligned}$$

And here, we see that the L.H.S. is just the derivative of $A(x)$ w.r.t. x, thus

$$\frac{\mathrm{d}A(x)}{\mathrm{d}x} = f(x), \tag{2.31}$$

or *"the rate of change in area is the function itself"*. But the area function is given by the integral. This means

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int f(x)\,\mathrm{d}x\right) = f(x): \tag{2.32}$$

integrals and derivatives are inverses of each other. If we rephrase theorem 1, we see that *"the integral is the cumulative effect of the function."*

The second fundamental theorem of calculus,

> **Theorem 2: The second fundamental theorem of calculus (Newton-Leibniz rule)** If a function $f(x)$ has an antideriva-

tive $A(x)$, then its indefinite integral from a to b is

$$\int_a^b f(x)\,\mathrm{d}x = \int_0^b f(x)\,\mathrm{d}x - \int_0^a f(x)\,\mathrm{d}x = A(b) - A(a). \tag{2.33}$$

follows as a direct consequence of the geometrical interpretation of integrals: area under the curve.

The fundamental theorem of calculus allows us to evaluate integrals using derivatives. For example, if the derivative of x^2 is $2x$, then we also know that the integral of $2x\,\mathrm{d}x$ is just x^2, which we'll discuss how to do that in the next chapter.

2.4 How to calculate an integral? Riemann sum

In section 2.2, we used geometry to find the area under a curve. However, that is not always possible, e.g., try integrating fig. 2.3 geometrically[2]. It'd be impossible. But we can still approximate its area by slicing the area under the curve into thin rectangular stripes, then summing them. The approximated area is called the **Riemann sum**. Due to its com-

[2]This is what you'd get if you solve the simple harmonic oscillator

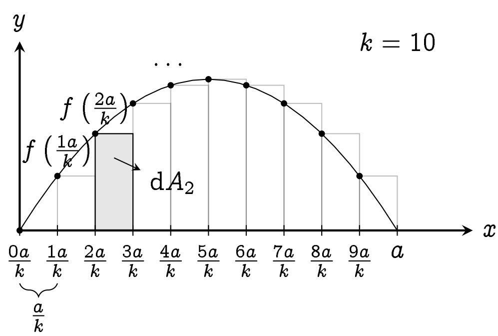


FIG. 2.3 | ILLUSTRATION OF RIEMANN SUM OF A FUNCTION $f(x)$ FROM 0 TO a BY SETTING $k = 10$

putational cost, you don't really want to use this method. However, to develop a good intuition at the integral, we should still know its symbolic form.

Let there be a function $A(s)$ that represents the actual area of a function $f(x)$ from 0 to s. The approximated area is then

$$\sum_{i=0}^{k} \mathrm{d}A_i = \sum_{i=0}^{k} \text{width} \times \text{height} \tag{2.34}$$

where k is the amount of subdivisions. From fig. 2.3, the width of each stripe is a / k, and the height of the i'th stripe is $f(ia / k)$. Therefore,

$$\sum_{i=0}^{k} \frac{a}{k} \times f\left(\frac{ia}{k}\right).$$

In fig. 2.3, we $k = 10$. The area of the second rectangle $\mathrm{d}A_2$ is

$$\mathrm{d}A_2 = \frac{a}{k} f\left(\frac{2a}{k}\right).$$

For an arbitrarily finite k, the Riemann sum is just an approximation. If you want to find the *actual* area under the curve, let $k \to \infty$. The limit as $k \to \infty$ is what we actually call the **integral**. Thus, we say

Definition 2: Naive definition of integrals

The (definite) integral, or the area under the curve of $f(x) = y$ from 0 to a, is defined as

$$\int_0^a \mathrm{d}A = \int_0^a f(x)\,\mathrm{d}x = \lim_{k\to\infty} \sum_{i=0}^{k} \frac{a}{k} \times f\left(\frac{ia}{k}\right) = A(a). \tag{2.35}$$

where $\int$ is the integral sign[a]. Here, 0 is the lower bound of integration, and a, the higher bound. The function $A(x)$ is called the **antiderivative**, or the **indefinite integral** of $f(x)$.

[a] Famously known for looking like a beansprout

I shall put these definitions into perspective in the next two examples. It might use a bit of series knowledge. If you don't know, you can simply search up the summation identities that I'll use in *Wikipedia* [2].

Example 2.4.1: Riemann sum and antiderivative of x^2

Let $f(x) = x^2$, and let $A(a)$ be the antiderivative of $f(x)$, i.e., $A(a)$ is the area under the curve of $f(x)$ from 0 to a. Note that

$$\sum_{i=0}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}, \tag{2.36}$$

The Riemann sum is then

$$\sum_{i=0}^{k} \frac{a}{k} \times f\left(\frac{ia}{k}\right) = \sum_{i=0}^{k} \frac{a}{k} \times \left(\frac{ia}{k}\right)^2 \tag{2.37}$$

$$= \sum_{i=0}^{k} \frac{a^3}{k^3} \times i^2 \tag{2.38}$$

$$= \frac{a^3}{k^3} \times \frac{k(k+1)(2k+1)}{6}. \tag{2.39}$$

To find the antiderivative, let $k \to \infty$. Notice that the $+1$ in the parenthesis are negligible when $k \to \infty$. Therefore, we can write the antiderivative as

$$A(a) = \int_0^a f(x)\,\mathrm{d}x \tag{2.40}$$

$$= \lim_{k\to\infty} \frac{a^3}{k^3} \times \frac{k(k+1)(2k+1)}{6} \tag{2.41}$$

$$= \lim_{k\to\infty} \frac{a^3}{k^3} \times \frac{2k^3}{6} \tag{2.42}$$

$$= \lim_{k\to\infty} \frac{x^3}{3} = \frac{x^3}{3}. \tag{2.43}$$

Example 2.4.2: Riemann sum and antiderivative of x^3

Let $f(x) = 4x^3$, and let $A(a)$ be the antiderivative of $f(x)$, i.e., $A(a)$ is the area under the curve of $f(x)$ from 0 to a. Note that

$$\sum_{i=0}^{k} i^3 = \left(\frac{k(k+1)}{2}\right)^2, \tag{2.44}$$

The Riemann sum is then

$$\begin{aligned} \sum_{i=0}^{k} \frac{a}{k} \times f\left(\frac{ia}{k}\right) &= \sum_{i=0}^{k} \frac{a}{k} \times 4\left(\frac{ia}{k}\right)^3 \\ &= 4\sum_{i=0}^{k}\left(\frac{x}{k}\right)^4 i^3 \\ &= 4\left(\frac{x}{k}\right)^4\left(\frac{k(k+1)}{2}\right)^2. \end{aligned}$$

To find the antiderivative, let $k \to \infty$. The $+1$ in $(k+1)$ can be ignored as $k \to \infty$. Therefore,

$$\begin{aligned} A(a) &= \int_0^a f(x)\,\mathrm{d}x \\ &= \lim_{k\to\infty} 4\left(\frac{x}{k}\right)^4\left(\frac{k^2}{2}\right)^2 \\ &= \lim_{k\to\infty} x^4 = x^4. \end{aligned}$$

2.5 Basic applications of the integrals

2.5.1 Five equations of linear motion

So far, we have developed an intuition for derivatives, integrals, and their relationship. Let's apply those concepts to derive the famous equations of linear motion:

$$v(t) = v_0 + at, \tag{2.45}$$

$$s = \frac{1}{2}(v_0 + v(t))t, \tag{2.46}$$

$$s = v_0(t)t + \frac{1}{2}at^2, \tag{2.47}$$

$$s = v(t)t - \frac{1}{2}at^2, \tag{2.48}$$

$$v(t)^2 = v_0^2 + 2as. \tag{2.49}$$

Here, v_0 represents the initial velocity, $v(t)$, the velocity at any time t, a, the acceleration, and s, the displacement. These equations are derived from the Newton's second law on *constant/uniformed acceleration* motion in one dimension, which we can evaluate using the geometrical interpretation of integrals developed earlier.

Because we've constrained the object to accelerate at a, the force exerted must be ma. Newton's second law $F = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}$ then simplifies to $a = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}$. The same idea applies, we rewrite the equation in terms of v.

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} \tag{2.50}$$

Which reads, "*the rate of change of v w.r.t. t is a*" or "*the slope of the v-t curve is always equals to a*". Because a is constant, the v-t curve must be a

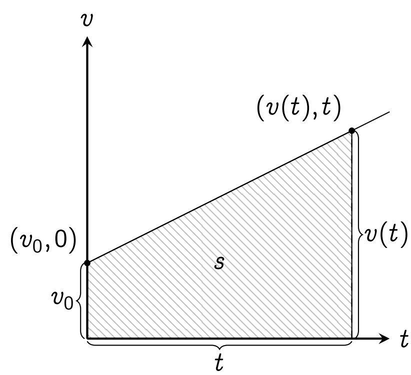


FIG. 2.4 | v-t GRAPH OF AN OBJECT UNDER CONSTANT OR UNIFORMED ACCELERATION

straight line with slope a, shown in fig. 2.4.

Derivation of eq. (2.45). We take advantage of the linearness of the curve. Just pick two points on fig. 2.4, as already shown, then

$$
\begin{aligned}
m = a &= \frac{\Delta v}{\Delta t} = \frac{v(t) - v_0}{t - 0} \\
a &= \frac{v(t) - v_0}{t} \\
v(t) &= v_0 + at.
\end{aligned}
$$
□

Derivation of eq. (2.46). Use the reverse of theorem 1. We know that[3] $v = \frac{\mathrm{d}}{\mathrm{d}t}s$; therefore,

$$\int v \, \mathrm{d}t = s,$$

which reads "*The displacement s is the area under the curve of a v-t graph*". From fig. 2.4, the area under the curve is a trapezoid with side length v_0, $v(t)$, and width t. Thus,

$$s = \frac{1}{2}(v_0 + v(t))t,$$

which is just eq. (2.46): the area of a trapezoid. □

Derivation of eqs. (2.47) *to* (2.49). We can arrange eq. (2.45) into three different ways, then plug in eq. (2.46). First, $v(t) = v_0 + at$

$$
\begin{aligned}
s &= \frac{1}{2}(v_0 + v_0 + at)t \\
&= v_0 t + \frac{1}{2}at^2.
\end{aligned}
$$

Second, $v_0 = v(t) - at$

$$
\begin{aligned}
s &= \frac{1}{2}(v(t) - at + v(t))t \\
&= v(t)t - \frac{1}{2}at^2.
\end{aligned}
$$

Third, $t = \dfrac{v(t) - v_0}{a}$

$$
\begin{aligned}
s &= \frac{1}{2}(v(t) + v_0)\frac{(v(t) - v_0)}{a} \\
2as &= (v(t) + v_0)(v(t) - v_0) \\
2as &= v(t)^2 - v_0^2 \\
v(t)^2 &= 2as + v_0^2.
\end{aligned}
$$
□

2.5.2 The area of a circle

The perimeter of the circle is $2\pi r$. For now, we only know how to find areas of polygons, but we want to know the circle's area. So, is there

[3] Here, x is replaced with s to represent displacement in one dimension.

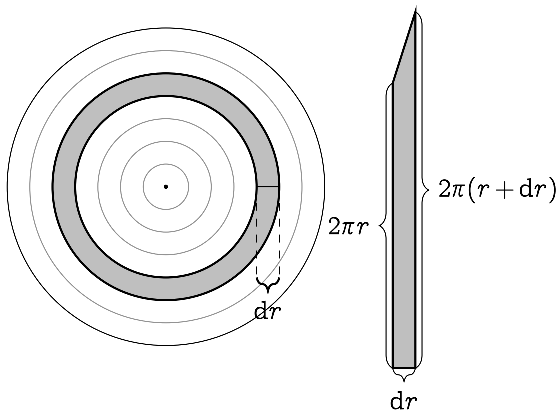


FIG. 2.5 | (LEFT) DISSECTING A CIRCLE RADIUS R RADIALLY INTO RINGS, EACH RING $\mathrm{d}r$ THIN. (RIGHT) STRETCHING A RING INTO A TRAPEZOID (NOT TO SCALE).

any way to turn a circle into a polygon? From section 2.4, that the riemann sum can approximate areas under the curve using rectangular stripes. It'd be great if this circle can be turned into multiple rectangular stripes on a graph right?

So, we dissect a circle radially into small rings $\mathrm{d}r$ thin, as shown in fig. 2.5. Then, stretch all the rings into a very thin trapezoid-like shape. It might seem impossible at first, but considering that our ring is very thin, it's quite easy to stretch it without breaking. I encourage you to grab a piece of paper, cut a really thin ring and try it out. If you actually do it, it'll look a bit warped. However, the warpedness will go away the thinner you go.

Since we sliced our trapezoid from a circle, for a stripe positioned at r, the inner side will be $2\pi r$ long and the outer side, $2\pi(r + dr)$. The area of the little trapezoid $\mathrm{d}A$ then becomes

$$\mathrm{d}A = \frac{1}{2}\,\mathrm{d}r(2\pi r + 2\pi(r + dr))$$

$$= 2\pi r\,\mathrm{d}r + 2\pi\,\mathrm{d}r^2.$$

As $\mathrm{d}r \to 0$, $\mathrm{d}r^2$ becomes negligible. Therefore,

$$\mathrm{d}A = 2\pi r\,\mathrm{d}r. \tag{2.51}$$

This equation says that for $\mathrm{d}r \to 0$, the trapezoid becomes a rectangle side length $\mathrm{d}r$ and $2\pi r$. Now, we have to sum it together. If we can put all these rectangles onto a graph, we can easily use the Riemann's sum to evaluate it. A natural way to do this is to put all the rectangles that we got from stretching the rings of the circle onto a graph one by one. The result would look something like fig. 2.6[4].

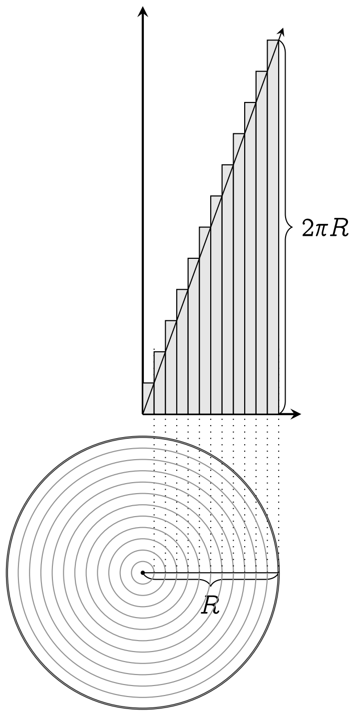


Fig. 2.6 | Rearranging all the approximated rectangles onto a graph (Not to scale).

For every stripe at r, its height is $2\pi r$. If you were to plot the height of all rectangles when $\mathrm{d}r \to 0$, it'll eventually look like a curve that's given by $f(r) = 2\pi r$. We're interested in the area of the circle from 0 to R. Now, it's transformed into the area under the curve of $f(r) = 2\pi r$: a triangle with base R and height $f(R) = 2\pi R$. Thus, the area of a circle becomes

$$A = \frac{1}{2}R(2\pi R) = \pi R^2.$$

And there you go, you've turned essentially turned circle into a triangle and evaluate its area from there. I'd like to end off this chapter

[4] Not to scale

by mentioning the spirit of mathematics. Sometimes, you can't solve the problem directly. Most of the time, you have to re-frame the problem into another more-solvable problem. Problems like these often have the most sublime connections to the foundations of mathematics. This is a common theme in most of mathematics, especailly calculus. So, be sure to keep this in mind while reading through.

2.6 Conclusion for Chapter 2

1. Riemann sum are used to approximate areas under the curve of a function by using little rectangles then summing it.

2. Integrals or anti-derivatives are functions that output the area under the graph of other functions.

3. The limit where the width of the rectangles in the Riemann sum approaches zero, the Riemann sum becomes an integral.

4. Integrals are the cumulative effect of a function.

5. Integrals and derivatives are inverses of each other, and they're related by the fundamental theorem of calculus

6. Integrals can be used in various ways by reframing questions into another simpler question.

CHAPTER 3

Basic derivatives and antiderivatives

This chapter covers the derivatives and integrals of common functions: polynomials, exponential, and logarithms; focusing on their geometrical interpretation.

Prerequisites: *binomial theorem (appendix B), basic trigonometry, derivatives, and integrals*

Terminologies: The integral refers to the area under the curve. The antiderivative refers to the function $A(x')$ that outputs the area under the curve from 0 to x'.

3.1 Trivial rules

The chain rule The Leibniz' notation treats derivative as fractions. You can cancel terms as seen earlier in eq. (1.9). This property can be used to take derivatives of composite functions. E.g., finding the derivative of $f(g(x))$ but you only know $f(x)$ and $g(x)$. First, substitute $g(x)$ as u.

$$\frac{\mathrm{d}f(g(x))}{\mathrm{d}x} = \frac{\mathrm{d}f(u)}{\mathrm{d}x}.$$

We know that *one* is the multiplicative identity, and *one* is any number divided by itself[1]. Let $1 = \frac{\mathrm{d}u}{\mathrm{d}u}$, then

$$\frac{\mathrm{d}f(g(x))}{\mathrm{d}x} = \frac{\mathrm{d}f(u)}{\mathrm{d}x} \times \frac{\mathrm{d}u}{\mathrm{d}u}.$$

Performing change of denominator, then substitute $u = g(x)$:

$$\frac{\mathrm{d}f(g(x))}{\mathrm{d}k} = \frac{\mathrm{d}f(u)}{\mathrm{d}u} \times \frac{\mathrm{d}g(x)}{\mathrm{d}x}.$$

This is what we call the chain rule, or more generally

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}u} \times \frac{\mathrm{d}u}{\mathrm{d}x}. \tag{3.1}$$

where u is a function of y.

The chain rule holds the intuition of how rate of changes relate to each other. E.g., the cheetah's speed is 10 times the bicycle's speed, which is 4 times the walking speed. The ratio between the cheetah's speed compared to walking speed would obviously be $10 \cdot 4 = 40$:

$$\frac{\mathrm{dCheetah}}{\mathrm{dWalking}} = \frac{\mathrm{dCheetah}}{\mathrm{dBicycle}} \times \frac{\mathrm{dBicycle}}{\mathrm{dWalking}}. \tag{3.2}$$

Integral constant In section 2.2, each time we evaluate the antiderivative, we add an initial condition term, i.e., v_0 and r_0. So for any function $f(x)$,

$$\int f(x)\,\mathrm{d}x = A(x) + C \tag{3.3}$$

where C is any constant. However, theorem 2 still holds for integrals with bounds.

[1]Except zero, of course.

Integral of the infinitesimal The antiderivative of the small rectangles $\mathrm{d}x$ is just the total rectangle x plus the integral constant C.

$$\int \mathrm{d}x = x + C. \tag{3.4}$$

Rules of equality If two arguments are equal, their derivatives and antiderivatives w.r.t. the same variable must also be equal.

$$\text{If } f = g, \text{ then } \frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}x}, \text{ and } \int f \,\mathrm{d}x = \int g \,\mathrm{d}x + C$$

Derivative of a constant A constant doesn't change; thus, the derivative of a constant is zero.

$$\frac{\mathrm{d}}{\mathrm{d}x}(c) = 0. \tag{3.5}$$

3.2 Linearity of differentiation and integration

A **linear** mathematical entity are defined as anything that is compatible with addition and scaling. E.g., for a function $f(x)$ to be linear, it must obey

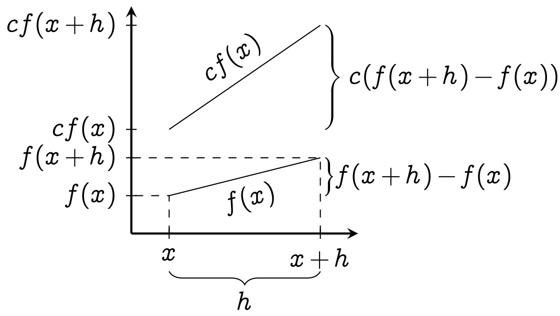
1. Additivity: $f(x + y) = f(x) + f(y)$
2. Homogeneity of degree one: $f(ax) = af(x)$ for all constant a

The simplest linear function there is, is a line that passes through the origin: $f(x) = ax$, in which

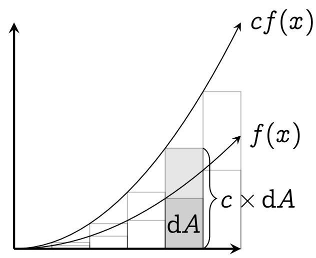
$$f(x + y) = a(x + y) = ax + ay = f(x) + f(y), \tag{3.6}$$

$$f(cx) = a(cx) = c(ax) = cf(x). \tag{3.7}$$

Because this property, we associate a line with being linear. However, linearity doesn't have to always refer to lines.



(A) FOR DERIVATIVES



(B) FOR INTEGRALS

FIG. 3.1 | CONSTANT MULTIPLE RULES

Both derivatives and integrals are linear; for any function $f(x)$, $g(x)$, and constants a, b, the following rules follow.

The constant multiple rule:

$$\frac{\mathrm{d}}{\mathrm{d}x}(af(x)) = a\frac{\mathrm{d}f(x)}{\mathrm{d}x}, \qquad \text{Illustrated in fig. 3.1a} \tag{3.8}$$

$$\int af(x)\,\mathrm{d}x = a\int f(x)\,\mathrm{d}x. \qquad \text{Illustrated in fig. 3.1b} \tag{3.9}$$

The geometrical interpretation of these two rules are very simple. When a function is multiplied by a constant c, its value is increased by a factor of c everywhere. The slope must also be increased by c, and thus the function's area also.

The sum rule:

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(x)+g(x)) = \frac{\mathrm{d}}{\mathrm{d}x}f(x) + \frac{\mathrm{d}}{\mathrm{d}x}g(x), \tag{3.10}$$

$$\int f(x) + g(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x. \tag{3.11}$$

I.e., adding two functions increases its slope, thus also increasing its area under the graph.

3.3 Derivatives and antiderivatives of polynomials

Now that we've discussed the "trivial rules", we're ready to tackle the easiest family of functions: the **polynomials**. They're in the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots . \tag{3.12}$$

Let's focus on the antiderivative first; we'll see later how the antiderivative of polynomials can be found with just a simple substitution trick.

The form mentioned in eq. (3.12) is quite useless if we want to make progress. We can break it down by using the linearity of derivatives:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}(a_0 + a_1 x^1 + a_2 x^2 + a_3 x^3 + \dots) \tag{3.13}$$

$$= a_1 \frac{\mathrm{d}x^1}{\mathrm{d}x} + a_2 \frac{\mathrm{d}x^2}{\mathrm{d}x} + a_3 \frac{\mathrm{d}x^3}{\mathrm{d}x} + \dots . \tag{3.14}$$

Now we're left with derivatives of *monomials* in the form of x^n. So let's do that instead.

3.3.1 Derivatives of monomials: the power rule

The method of increments allow us to quickly evaluate the derivative of x^n.

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^n) = \lim_{h \to 0} \frac{(x+h)^n - x^n}{h}. \tag{3.15}$$

By using the binomial expansion (appendix B),

$$= \lim_{h \to 0} \frac{\left(\sum_{k=0}^{n} \binom{n}{k} x^{n-k} \cdot h^n\right) - x^n}{h} \tag{3.16}$$

$$
\begin{aligned}
&= \lim_{h \to 0} \frac{\binom{n}{0} x^n h^0 + \binom{n}{1} x^{n-1} h^1 + \binom{n}{2} x^{n-2} h^2 + \ldots \binom{n}{n} + x^0 h^n - x^n}{h} && (3.17)\\
&= \lim_{h \to 0} \frac{x^n + n x^{n-1} h^1 + \binom{n}{2} x^{n-2} h^2 + \cdots + h^n - x^n}{h} && (3.18)\\
&= \lim_{h \to 0} n x^{n-1} + \binom{n}{2} x^{n-2} h + \binom{n}{3} x^{n-3} h^2 + h^{n-1} && (3.19)
\end{aligned}
$$

The terms with h vanishes when $h \to 0$; therefore,

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^n) = n x^{n-1}, \tag{3.20}$$

which is what we call the power rule. Because the binomial theorem work for all real numbers, this is true for any powers of n. But what about the geometrical interpretation given?

Let's now focus on the geometrical interpretation of the derivative of x^2: a function that represents the area of a square sidelength x. Its derivative then represents the ratio between the change in x, and the change of area when the sidelength is increased by $\mathrm{d}x$.

Illustrated in fig. 3.2,

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^2) = \lim_{\mathrm{d}x \to 0} \frac{x\,\mathrm{d}x + x\,\mathrm{d}x + \mathrm{d}x^2}{\mathrm{d}x}. \tag{3.21}$$

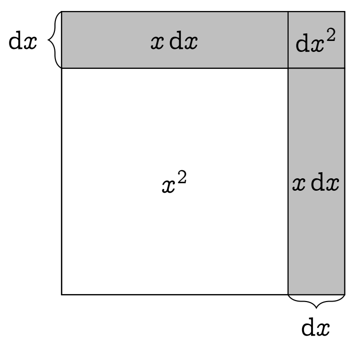


Fig. 3.2 | Interpretation of $\frac{\mathrm{d}x^2}{\mathrm{d}x}$ ($\mathrm{d}x$ exaggerated)

When $\mathrm{d}x \to 0$, the $\mathrm{d}x^2$ square would just become a single point. Compared to the big $x\,\mathrm{d}x$ on the side, it's negligible; therefore, we can safely ignore it. The equation then becomes

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^2) = \lim_{\mathrm{d}x \to 0} \frac{2x\,\mathrm{d}x}{\mathrm{d}x} = 2x,$$

which is equivalent to the result from the power rule.

Deriving $\frac{\mathrm{d}}{\mathrm{d}x}(x^3)$ geometrically shouldn't be too hard either. Take a cube sidelength x; increase its side length by $\mathrm{d}x$, and compute the ratio between the change in volume and $\mathrm{d}x$. The final answer should be $3x^2$. You could try with more higher order of n, but it'd be very hard to visualize.

Here I leave some exercises which shouldn't be too hard to do

1. $\frac{\mathrm{d}}{\mathrm{d}x}(x^2 - 2x + 16)$ $\qquad 2x - 2$
2. $\frac{\mathrm{d}}{\mathrm{d}x}(x^3 + x^2 + x + 1)$ $\qquad 3x^2 + 2x + 1$
3. $\frac{\mathrm{d}}{\mathrm{d}x}(3x^4 + 24x^3 - 2x^2 - 32x + 88)$ $\qquad 12x^3 + 72x^2 - 4x - 32$

3.3.2 Antiderivatives of polynomials: the reversed power rule

For the antiderivative of x^n, we use the fundamental theorem of calculus (theorem 1) on the power rule (eq. (3.20)),

$$x^n = \int nx^{n-1}\,\mathrm{d}x.$$

Substitute n with $n+1$

$$\int (n+1)x^n\,\mathrm{d}x = x^{n+1},$$

and by the linearity of integrations (eq. (3.9)), we get the **reversed power rule**:

$$\int x^n\,\mathrm{d}x = \frac{x^{n+1}}{n+1}. \tag{3.22}$$

Notice, this rule does not work for $n = 1$ because we can't divide by zero...*or can we*??? (Discussed further in section 3.8)

3.4 Extending the equations of linear motion

Let's now apply our power rule to some problems. From section 2.5.1, we discussed the equation of motion of objects with uniformed acceleration. What if now, the acceleration is changing over time, but is change uniformly? Mathematically, this means that the derivative of acceleration w.r.t. time, aka the jerk, is constant.

$$j = \frac{\mathrm{d}a}{\mathrm{d}t}.$$

We can then move $\mathrm{d}t$ around and integrate both sides by using the reversed power rule on it.

$$\begin{aligned} \int j \,\mathrm{d}t &= \int \mathrm{d}a \\ j \int \mathrm{d}t &= \int \mathrm{d}a && \text{Linearity of integrals} \\ jt + a_0 &= a && \text{Integral constant } a_0 \end{aligned}$$

Because a is the derivative of v,

$$\begin{aligned} \frac{\mathrm{d}v}{\mathrm{d}t} &= jt + a_0 \\ \int \mathrm{d}v &= \int jt + a_0 \,\mathrm{d}t \\ v &= j \int t \,\mathrm{d}t + \int a_0 \,\mathrm{d}t && \text{Linearity of integrals} \\ v &= \frac{1}{2} jt^2 + a_0 t + v_0. && \text{Reversed power rule} \end{aligned}$$

We add v_0 as an integral constant in a similar manner. Then because v is the derivative of r,

$$\begin{aligned}
\frac{\mathrm{d}r}{\mathrm{d}t} &= \frac{1}{2}jt^2 + a_0 t + v_0 \\
\int \mathrm{d}r &= \int \frac{1}{2}jt^2 + a_0 t + v_0 \, \mathrm{d}t \\
r &= \frac{1}{2}j \int t^2 \, \mathrm{d}t + a_0 \int t \, \mathrm{d}t + v_0 \int \mathrm{d}t \\
r &= \frac{1}{6}jt^3 + \frac{1}{2}a_0 t^2 + v_0 t + r_0.
\end{aligned}$$

And there you have it! Technically we can extend this to whatever we want. The jerk also doesn't have to be uniformed, it could be a function of time itself. But, that's probably enough. If you'd like to try, extend the equation of motion to uniformed snap s. The final equation should be

$$r = \frac{1}{24}st^4 + \frac{1}{6}jt^3 + \frac{1}{2}a_0 t^2 + v_0 t + r_0.$$

3.5 Exponentials and growth

The next function that we're going to discuss is exponentials: the mathematical representation of growth and decay. As an example, let's say there's a magical drop of water that doubles its volume V every hour. I.e., for any time t,

$$V(t+1) = 2V(t). \tag{3.23}$$

If the drop starts at one unit of volume, $V(0) = 1$; thus,

$$\begin{aligned}
V(1) &= 2V(0) = 2, \\
V(2) &= 2V(1) = 2(2) = 2^2, \\
V(3) &= 2V(2) = 2(2^2) = 2^3, \\
V(4) &= 2V(3) = 2(2^3) = 2^4, \\
&\vdots
\end{aligned}$$

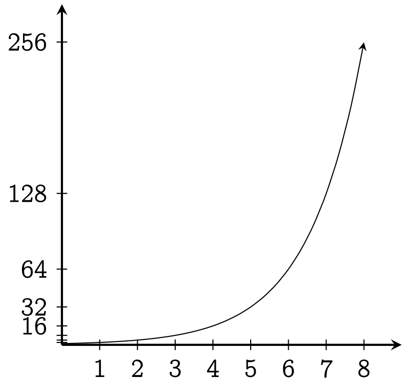


Fig. 3.3 | Exponential function 2^x plotted from 0 to 8

t	$V(t)$	$V(t) - V(t-1)$
0	1	
1	2	$2 - 1 = 1$
2	4	$4 - 2 = 2$
3	8	$8 - 4 = 4$
4	16	$16 - 8 = 8$
5	32	$32 - 16 = 16$
6	64	$64 - 32 = 32$
7	128	$128 - 64 = 64$
8	256	$256 - 128 = 128$
9	512	$512 - 256 = 256$
10	1024	$1024 - 512 = 12$

Table 3.1 | Tables of 2^x plotted at interval 1 from 0 to 10

It's clear that the pattern is $V(t) = 2^t$: an exponential function. A natural question to ask is "*what is its rate of change?*". Let's start by plotting the function over time (fig. 3.3). But, exponentials grow too quick to plot! By $V(7)$, we're already in the hundreds. So It'd probably be better to list the values of each point on a table. Notice that on the right most column of table 3.1, the difference between $V(t)$ and $V(t-1)$ is exactly $V(t-1)$: the function changes as much as its past-self. So does that mean that $\frac{\mathrm{d}}{\mathrm{d}x}(2^x) = 2^x$?

Well sadly not, but close. See, table 3.1 only shows a *discrete* step. You can write it out as

$$\frac{2^{x+1} - 2^x}{1} = 2^x \left(\frac{2-1}{1}\right) = 2^x, \tag{3.24}$$

that is why $V(t) - V(t-1) = V(t-1)$. But we still need the method of increments to calculate the derivative of 2^x:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(2^x) &= \lim_{h\to 0}\frac{2^{x+h}-2^x}{\mathrm{d}x}\\ &= 2^x \lim_{h\to 0}\left(\frac{2^h-1}{h}\right).\end{aligned}$$

Now you could try plugging in a really small value of h, say 0.000001. The term $\dfrac{2^h-1}{h}$ will approach 0.69314 ... . If you try other bases of exponents, say 3, you might see a pattern emerging.

$$\frac{\mathrm{d}}{\mathrm{d}x}(3^x) = 3^x \lim_{h\to 0}\frac{3^h-1}{h}. \tag{3.25}$$

The rate of change of an exponential function is always itself times a proportionality constant. For 3^x, it's about 1.09851 ... . If we could find a number n where $\dfrac{n^h-1}{h} = 0$, we'd have a very pretty function which it is its own derivative. So let's find that!

3.5.1 A function that is its own derivative

Let's set a goal: find the function that is its own derivative. I shall introduce a substantial concept in calculus: the expansion of functions. Every function has a polynomial expansion[2] called the **power series**. For every $f(x)$,

$$f(x) = a_0 + a_1x^1 + a_2x^2 + a_3x^3 + \dots . \tag{3.26}$$

E.g., $\sin(x)$ can be written as

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \dots . \tag{3.27}$$

we will derive this expression later in section 5.4.1. For now, we can just use eq. (3.26) to find the expression for the function that is its own derivative.

[2]Although the convergence of the series derived is quite questionable; thankfully, the power series of n^x converges everywhere.

We've seen that the exponential is a possible candidate for a function that is its own derivative. Now, assume that for some real number n,

$$\frac{\mathrm{d}}{\mathrm{d}x}(n^x) = n^x. \tag{3.28}$$

Then, we use the polynomial expansion and the power rule,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(a_0 + a_1x^1 + a_2x^2 + a_3x^3 + \dots) &= a_1 + 2a_2x^1 + 3a_3x^2 + 4a_4x^3 + \dots \\ &= n^x\end{aligned}$$

If the function is its own derivative, the polynomial expansion of the function and its derivative must be the same.

$$n^x = a_0 + a_1x^1 + a_2x^2 + a_3x^3 + \dots, \tag{3.29}$$

$$n^x = a_1 + 2a_2x^1 + 3a_3x^2 + 4a_4x^3 + \dots. \tag{3.30}$$

Since both are polynomials, we can match the coefficient here:

$$\begin{aligned} a_0 &= a_1 \\ a_1 &= 2a_2 \\ a_2 &= 3a_3 \\ a_3 &= 4a_4 \\ &\vdots \end{aligned} \tag{3.31}$$

a_0 and a_1 is relatively easy to find. As we've seen, n must be between 2 and 3. By the properties of exponentials, $x = 0 \implies n^x = 1$. We can then plug $x = 0$ and set $n^x = 1$ into eq. (3.30):

$$\begin{aligned} 1 &= a_0 + a_1(0)^1 + a_2(0)^2 + a_3(0)^3 + \dots \\ a_0 &= 1. \end{aligned}$$

Since $a_0 = a_1$, a_1 must also be 1. We can then go back to eq. (3.31) and get

$$n^x = 1 + 1 + \frac{1}{2!}x + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots$$

The pattern here is clearly $a_n = n!$. If we want to find n, we just let $x = 1$.

$$n = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \dots,$$

and there we have an expression for n which is an irrational number. If you work this out, it's around $2.71828\dots$. Because $(2.71828\dots)^x$ is its own derivative, it's very useful in mathematics and appears everywhere, even at the seams of mathematics that doesn't even seems related to growths: the patterns of prime number, this constant $2.71828\dots$ has a name and symbol: the Euler's number[3], written as e where

$$\mathrm{e} = \sum_{i=0}^{\infty} \frac{1}{i!} \tag{3.32}$$

3.5.2 Another interpretation of e: infinite bank interests

There are two types of bank interests: simple and compound. Simple interests is the thing that you don't really want: the interest is always the same and doesn't grow with your account. You can calculate it by using

$$n(t) = n_0 + tr \tag{3.33}$$

where $n(t)$ is the total money at time t, n_0 the initial money in your bank account, and r, the interest rate.

Compound interest in the other hand calculates your interest based on how much money you have at that moment:

$$n(t+1) = n(t)r + n(t). \tag{3.34}$$

We can find the expression for $n(t)$ in a similar fashion to what we've done in eq. (3.23). You'll get

$$n(t) = (1+r)^t n_0 \tag{3.35}$$

[3]Not to be confused with the "Euler's constant" which is another constant written γ, and is around $0.57721\dots$

which is an exponential function.

Let's say you deposit 100$ into a bank and the bank is offering you two options on **compound interests** rate. 1) Take 100% interest in 1 year, 2) Take 100 / 2% twice a year, or 3) Take 100 / 356% daily. If you take option one, you'd end up with 200$. Option two takes you to 225$, and option three takes you to around 271.447 ... $. You might see a theme here. If you get 100 / n% interest, n times a year, the result keeps getting higher. Is there an upper limit to this?

If we write it in terms of limits as $n \to \infty$ and use eq. (3.35), the compound interests formula,

$$x = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n n_0.$$

Where x is the total money after a year. We're interested in the upper limit, so we'll just let n_0 for now. The expression will become

$$x = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n \tag{3.36}$$

Technically, we could go in and substitute a very high n, such as 1000000. But I believe you could already see that it would be a nightmare to calculate: exponentiation is not at all an easy task. However, notice that from option three earlier, the total money is 271.447 ... $ which is suspiciously similar to e at 2.71828 ... . If eq. (3.36) equals eq. (3.31), we'd find the upper limit for this problem and solve the mystery.

We can use the binomial theorem on eq. (3.36) and get

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} \frac{1}{n^k} \\
&= \lim_{n\to\infty} \binom{n}{0}\frac{1}{n^0} + \binom{n}{1}\frac{1}{n^1} + \binom{n}{2}\frac{1}{n^2} + \binom{n}{3}\frac{1}{n^3} + \ldots .
\end{aligned}$$

Then, we use the definition of n choose k,

$$= 1 + \lim_{n\to\infty} \frac{n!}{1!(n-1)!}\frac{1}{n^1} + \frac{n!}{2!(n-2)!}\frac{1}{n^2} + \frac{n!}{3!(n-3)!}\frac{1}{n^3} + \ldots .$$

Now, we can cancel the $n!$ on the numerator to the denominator and isolate the factorials.

$$= 1 + \lim_{n\to\infty} \frac{n(n-1)!}{(n-1)!}\frac{1}{1!n^1} + \frac{n(n-1)(n-2)!}{(n-2)!}\frac{1}{2!n^2} + \ldots$$
$$= 1 + \frac{1}{1!} + \lim_{n\to\infty} \frac{n(n-1)}{2!n^2} + \frac{n(n-1)(n-2)}{3!n^3} + \frac{n(n-1)(n-2)(n-3)}{4!n^4} + \ldots$$

Notice, as $n \to \infty$, the ratio between $n+R$ and n where R is any real numbers would be literally negligible. For every terms in our series, both the numerator and the denominator has the same polynomic degrees. Therefore, all the n's in the series cancel out and we get

$$x = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots \tag{3.37}$$

which is literally eq. (3.31). That means, the upper limit that the bank can give you is e. Geometrically, it should make sense. Because we're gradually turning a discrete interest into a continuous one, e should appear in the limit of the continuous bank interests.

3.5.3 The antiderivative of exponential functions

It should be trivial that if e^x is the derivative of itself, so is its antiderivative

$$\int \mathrm{e}^x \,\mathrm{d}x = \mathrm{e}^x + C. \tag{3.38}$$

For other bases, we could use theorem 1, the fundamental theorem of calculus, to find its antiderivative. That is, if

$$\frac{\mathrm{d}(n^x)}{\mathrm{d}x} = n^x \lim_{h\to 0} \frac{n^h - 1}{h},$$

then

$$\frac{\mathrm{d}}{\mathrm{d}x}(n^x) = n^x \lim_{h\to 0} \frac{n^h - 1}{h}$$

$$n^x = \lim_{h\to 0} \frac{n^h - 1}{h} \int n^x \, \mathrm{d}x$$
$$\int n^x \, \mathrm{d}x = n^x \lim_{h\to 0} \left(\frac{n^h - 1}{h} \right)^{-1}.$$

The term in the limit sign still appears here. If we want to uncover how this term comes to be, we must discuss the logarithms.

3.6 Logarithms

Monomials have their inverse functions: the roots, exponentials also has an inverse functions: the logarithms. Here's a simple example to illustrate what I mean.

$$\sqrt[a]{x^a} = x, \text{ but with logarithms, } \log_a(a^x) = x.$$

Logarithms are inverses functions of exponentials: it cancels exponentials. With it comes the following properties:

$$\log_a(x) + \log_a(y) = \log_a(xy) \tag{3.39}$$
$$\log_a(x) - \log_a(y) = \log_a\left(\frac{x}{y}\right) \tag{3.40}$$
$$\log_a(x) = \frac{\log_b(x)}{\log_b(a)} \tag{3.41}$$

Since e^x is shown to be a very important function in modelling continuous growth, and is its own derivative, we give its inverse function its own name: the **natural logarithm**, written as $\ln(x)$.

So what's the growth of $\ln(x)$? At first sight, since the exponential grows so fast, the inverse of exponentials must grow very slowly. It might just be the reciprocal of x: $1 \,/\, x$. However, we have to derive it somehow. You could try the method of increments and get

$$\frac{\mathrm{d}\ln(x)}{\mathrm{d}x} = \lim_{h\to 0} \frac{\ln(x+h) - \ln(x)}{h}. \tag{3.42}$$

As seen, there are no properties of logarithm that can manipulate this. However, I shall introduce quite a sneaky concept here called **implicit differentiation**.

Notice that we've only concerned the function that's clearly written in the form $f(x) = y$ and its rate of change. We call these neat functions **explicit functions**. However, not all functions are written in this form. E.g., $\sqrt{x^2 + y^2} = 2$. These functions are called **implicit functions**, and we can actually differentiate it.

If I let $\ln(x) = y$, I can raise e to the power of both sides and get

$$\mathrm{e}^{\ln(x)} = \mathrm{e}^{y}. \tag{3.43}$$

Because logarithms are inverses of exponential,

$$x = \mathrm{e}^{y}. \tag{3.44}$$

From the rule of equality, (section 3.1), we take the derivative of both sides w.r.t. y instead of x:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}y} &= \frac{\mathrm{d}\mathrm{e}^{y}}{\mathrm{d}y} \\ \frac{\mathrm{d}x}{\mathrm{d}y} &= \mathrm{e}^{y}. \end{aligned}$$

But we're looking for the derivative of y (which is just $\ln(x)$) w.r.t. x, not the derivative of x w.r.t. y. Here's where Leibniz's notation comes into clutch: we can swap the numerator with the denominator for both sides then substitute in the y:

$$\begin{aligned} \frac{\mathrm{d}\ln(x)}{\mathrm{d}x} &= \frac{1}{\mathrm{e}^{\ln(x)}} \\ \therefore \frac{\mathrm{d}\ln(x)}{\mathrm{d}x} &= \frac{1}{x}, \end{aligned}$$

and there: the derivative of the natural logarithm is the reciprocal.

With the power of natural logarithms, we can actually go back at the derivative of n^x and finally uncover the mystery behind the proportionality term that's lingering around. Start with the manipulation of n^x.

$$n^x = \left(\mathrm{e}^{\ln(n)}\right)^x = \mathrm{e}^{x\ln(x)}. \tag{3.45}$$

With the chain rule (section 3.1), let $u = x\ln(n)$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(n^x) &= \frac{\mathrm{d}\mathrm{e}^{x\ln(n)}}{\mathrm{d}x}\\ &= \frac{\mathrm{d}\mathrm{e}^u}{\mathrm{d}x}\times\frac{\mathrm{d}u}{\mathrm{d}u}\\ &= \frac{\mathrm{d}\mathrm{e}^u}{\mathrm{d}u}\times\frac{\mathrm{d}u}{\mathrm{d}x}\\ &= \mathrm{e}^{x\ln(n)}\times\frac{\mathrm{d}x\ln(n)}{\mathrm{d}x}\\ &= n^x\ln(n).\end{aligned}$$

And here it is. The mystery proportionality constant is just a consequence of the natural logarithm. Thus, one way to define the natural log would be

$$\ln(n) = \lim_{h\to 0}\frac{n^h - 1}{h}. \tag{3.46}$$

The antiderivative of other bases exponents are then given by

$$\int n^x = \frac{1}{\ln(n)}n^x + C. \tag{3.47}$$

3.6.1 The product rule and the quotient rule

Sometimes, we have to multiply the two functions together before taking the derivatives. There are two ways to do this. To keep the spirit of visualization, I shall first introduce the geometrical way, then the analytical way.

The derivative of $f(x)g(x)$ w.r.t. x can be thought of a rectangle with side length that's governed by $f(x)$ and $g(x)$. As shown in fig. 3.4, the

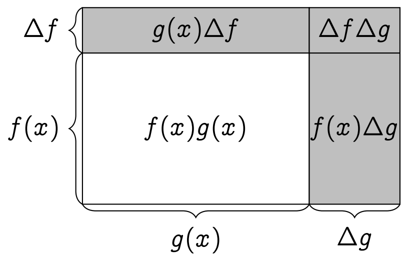


FIG. 3.4 | THE GEOMETRICAL INTERPRETATION OF THE PRODUCT RULE.

area increase on side $f(x)$ is $g(x)\Delta f$ and on $g(x)$, $f(x)\Delta g$. The $\Delta f \Delta g$ part is basically negligible. Therefore,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(f(x)g(x)) &= \lim_{h\to 0} \frac{g(x)\Delta f + f(x)\Delta g}{h} \\
&= \lim_{h\to 0} \frac{f(x)(g(x+h) - g(x)) + g(x)(f(x+h) - f(x))}{h} \\
&= f(x) \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} + g(x) \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= f(x)\frac{\mathrm{d}g(x)}{\mathrm{d}x} + g(x)\frac{\mathrm{d}f(x)}{\mathrm{d}x}.
\end{aligned}$$

Which is what we call the product rule. Notice the alternation between f and g in the two terms. The derivative of $f(x)$ is multiplied by $g(x)$, and the derivative of $g(x)$ is multiplied by $f(x)$. This is a direct consequence of the diagram: the change in $f(x)$ is multiplied by $g(x)$ to give the area and also the other way around. You could check this with the method of increments, and it would still be true. I encourage you to do it.

To take derivatives of quotients of functions, just plug in $1 \,/\, g(x)$ instead of $g(x)$. The final form should be

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)}{g(x)}\right) = \frac{f(x)\dfrac{\mathrm{d}g(x)}{\mathrm{d}x} - g(x)\dfrac{\mathrm{d}f(x)}{\mathrm{d}x}}{f(x)^2}. \tag{3.48}$$

But how's about the product of multiple functions? We can't really geometrically interpret those anymore. So we have to turn ourselves to the analytical method.

Let's think of this through. We don't know the derivative of products, but we know the derivative of *sums*: the linearity property. So if we can turn products into sum, this problem would be so easy! Gladly, the logarithms can do exactly that. To make our life easier, we shall use the natural logarithm. Because I want to save space, let's write $a(x)b(x)c(x)$ as just abc. Do know that these functions are all dependent on x. Start off with a manipulation of products.

$$\begin{aligned}\frac{\mathrm{d}(abc\dots)}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^{\ln(abc\dots)}) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^{\ln(a)+\ln(b)+\ln(c)+\dots}).\end{aligned}$$

Now, let $\ln(a)+\ln(b)+\ln(c)+\cdots = u$ and use the chain rule,

$$\begin{aligned}&= \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^{u})\cdot\frac{\mathrm{d}u}{\mathrm{d}u} \\ &= \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^{u})\cdot\frac{\mathrm{d}}{\mathrm{d}x}\left(\ln(a)+\ln(b)+\ln(c)+\dots\right) \\ &= \mathrm{e}^{u}\left(\frac{\mathrm{d}\ln(a)}{\mathrm{d}x}+\frac{\mathrm{d}\ln(b)}{\mathrm{d}x}+\frac{\mathrm{d}\ln(c)}{\mathrm{d}x}+\dots\right).\end{aligned}$$

Then use the chain rule again on the terms in the parenthesis

$$\begin{aligned}&= \mathrm{e}^{u}\left(\frac{\mathrm{d}\ln(a)}{\mathrm{d}x}\frac{\mathrm{d}a}{\mathrm{d}a}+\frac{\mathrm{d}\ln(b)}{\mathrm{d}x}\frac{\mathrm{d}b}{\mathrm{d}b}+\frac{\mathrm{d}\ln(c)}{\mathrm{d}x}\frac{\mathrm{d}c}{\mathrm{d}c}+\dots\right) \\ &= (abc\dots)\left(\frac{\mathrm{d}\ln(a)}{\mathrm{d}a}\frac{\mathrm{d}a}{\mathrm{d}x}+\frac{\mathrm{d}\ln(b)}{\mathrm{d}b}\frac{\mathrm{d}b}{\mathrm{d}x}+\frac{\mathrm{d}\ln(c)}{\mathrm{d}c}\frac{\mathrm{d}c}{\mathrm{d}x}+\dots\right) \\ &= (abc\dots)\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}x}+\frac{1}{b}\frac{\mathrm{d}b}{\mathrm{d}x}+\frac{1}{c}\frac{\mathrm{d}c}{\mathrm{d}x}+\dots\right).\end{aligned}$$

And there we have it: the generalized product rule.

3.6.2 Alternate derivations for the power rule

The power rule can also be derived using the same technique we just used. However, we use a different property of logarithm: $\ln(x^n) =$

$n \ln(x)$.

$$\frac{\mathrm{d}x^n}{\mathrm{d}x} = \frac{\mathrm{d}e^{n\ln(x)}}{\mathrm{d}x}.$$

Let $n\ln(x) = u$ then use the chain rule

$$\begin{aligned}
&= \frac{\mathrm{d}e^u}{\mathrm{d}x} \cdot \frac{\mathrm{d}u}{\mathrm{d}u} \\
&= \frac{\mathrm{d}e^u}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} \\
&= e^u \cdot \frac{\mathrm{d}}{\mathrm{d}x}\big(n\ln(x)\big) \\
&= x^n \cdot n\frac{1}{x} = nx^{n-1}.
\end{aligned}$$

3.7 Implicit differentiation

We've discussed that we can differentiate implicit functions. Normally, it wouldn't be quite useful, but this **implicit differentiation** shows up when two rates of change are related to each other: the **related rates** problem.

Take the example of a sliding ladder, sketched in fig. 3.5. The ladder is 5m long. If at any one moment, what is the rate of sliding along the x-axis w.r.t. the y-axis?

The problem is asking us to find $\frac{\mathrm{d}x}{\mathrm{d}y}$. Here, the rates of sliding along the y-axis $\frac{\mathrm{d}y}{\mathrm{d}t}$ and along the x-axis $\frac{\mathrm{d}x}{\mathrm{d}t}$ is clearly related because the ladder length still stays the same over time. If y decreases, x must increase. Both variables are related by the Pythagorean theorem

$$x^2 + y^2 = 5^2. \tag{3.49}$$

Then, we can differentiate this w.r.t. t, and use the chain rule

$$\frac{\mathrm{d}x^2}{\mathrm{d}t} + \frac{\mathrm{d}y^2}{\mathrm{d}t} = \frac{\mathrm{d}(5^2)}{\mathrm{d}t}$$

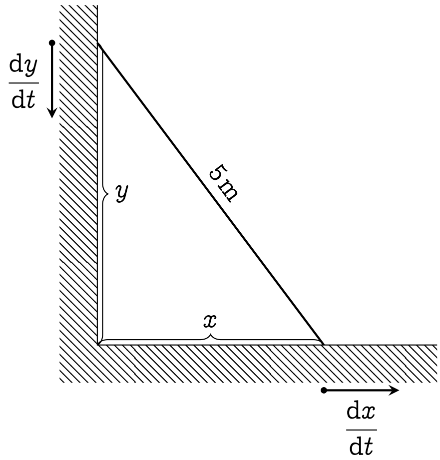


Fig. 3.5 | A ladder length 5m sliding down a corner.

$$
\begin{aligned}
\frac{\mathrm{d}x^2}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\mathrm{d}y^2}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}t} &= 0 \\
2x\frac{\mathrm{d}x}{\mathrm{d}t} + 2y\frac{\mathrm{d}y}{\mathrm{d}t} &= 0.
\end{aligned}
$$

We can take advantage of the Leibniz's notation and multiply by $\mathrm{d}t$ on both sides giving

$$2x\,\mathrm{d}x + 2y\,\mathrm{d}y = 0.$$

To find $\frac{\mathrm{d}x}{\mathrm{d}y}$, we just have to isolate the variables,

$$\frac{\mathrm{d}x}{\mathrm{d}y} = -\frac{y}{x}.$$

And this should actually make sense. Because if y is increasing by a bit, x must decrease by some amount, and that amount is $-y\,/\,x$: the higher the y, the larger the rates of sliding.

3.8 But why is the integral of the reciprocal the natural logarithm?

As we've seen, the antiderivative of $\frac{1}{x}$ is $\ln(x)$ by the fundamental theorem of calculus (theorem 1). I, however, find it disturbing and unresolved. It's a hole in the reversed power rule. From this dissatisfaction, I spent a night coming up with a way to derive this using just the reversed power rule. Enjoy the transformation!

Function	Antiderivative
x^{-3}	$-x^{-2}\,/\,2$
x^{-2}	x^{-1}
x^{-1} or $1\,/\,x$	$\ln(x)$
x^0 or 1	x
x^1	$x^2\,/\,2$

FIG. 3.6 | TABLES OF REVERSED POWER RULE FROM x^{-3} TO x^1

$$\begin{aligned}
\int \frac{1}{x}\,\mathrm{d}x &= \int \lim_{h\to 0}\left(\frac{1}{2}x^{-1+h} + \frac{1}{2}x^{-1-h}\right) \\
&= \lim_{h\to 0}\int\left(\frac{1}{2}x^{-1+h} + \frac{1}{2}x^{-1-h}\right) \\
&= \lim_{h\to 0}\int\left(\frac{1}{2}\frac{x^{-1+h+1}}{(-1+h+1)} + \frac{1}{2}\frac{x^{-1-h+1}}{(-1-h+1)}\right) \\
&= \lim_{h\to 0}\left(\frac{1}{2}\frac{x^h}{h} - \frac{1}{2}\frac{x^{-h}}{-h}\right) = \lim_{h\to 0}\left(\frac{1}{2}\frac{x^h}{h}\frac{x^h}{h} - \frac{1}{2}\frac{1}{hx^h}\right) \\
&= \lim_{h\to 0}\left(\frac{x^{2h}-1}{2hx^h}\right) = \lim_{h\to 0}\left(\frac{\mathrm{e}^{2h\ln(x)}-1}{2h\ln(x)}\cdot\frac{\ln(x)}{x^h}\right) \\
&= \lim_{h\to 0}\left(\frac{e^{2h\ln(x)}}{2h\ln(x)}\right)\cdot\lim_{h\to 0}\left(\frac{\ln(x)}{x^h}\right) \\
&= \lim_{h\to 0}\left(\frac{e^{2h\ln(x)}}{2h\ln(x)}\right)\cdot\ln(x).
\end{aligned} \tag{3.50}$$

Then, we evaluate the limit at the front by letting $u = 2h\ln(x)$. When $h \to 0$, $u \to 0$ as well. Then, use the definition of e from eq. (3.36).

$$\begin{aligned}
\lim_{h\to 0}\left(\frac{\mathrm{e}^{2h\ln(x)}}{2h\ln(x)}\right) &= \lim_{u\to 0}\left(\frac{\mathrm{e}^u - 1}{u}\right) \\
&= \lim_{u\to 0}\left(\frac{\left(\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n\right)^u}{u}\right).
\end{aligned}$$

Change the limits from $n \to \infty$ into $n \to 0$. Notice, $\lim_{n\to 0}\left(1+\frac{1}{n}\right)^n = \lim_{n\to 0}(1+n)^{1/n}$. If $n \to 0$ and $u \to 0$, that means $n = u$. Substitute $n = u$ into the limit,

$$= \lim_{u\to 0}\left(\frac{\left(\lim_{u\to 0}(1+u)^{1/u}\right)^u - 1}{u}\right)$$

$$= \lim_{u\to 0}\left(\frac{1+u-1}{u}\right) = 1.$$

Then, substitute this limit back into eq. (3.50), you'll see that

$$\int \frac{1}{x}\,\mathrm{d}x = \lim_{h\to 0}\left(\frac{\mathrm{e}^{2h\ln(x)}}{2h\ln(x)}\right)\cdot\ln(x) = \ln(x) + C.$$

3.9 Formula for Chapter 3

3.9.1 Formula for derivatives of functions

1. $f(x) = g(x) \implies \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{\mathrm{d}g(x)}{\mathrm{d}x}$ (Rules of Equality)
2. For $c \in \mathbb{R}$, $\frac{\mathrm{d}(c)}{\mathrm{d}x} = 0$ (Derivative of a constant)
3. $\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{\mathrm{d}x}{\mathrm{d}u} \times \frac{\mathrm{d}u}{\mathrm{d}y}$ (Chain rule)
4. $\frac{\mathrm{d}}{\mathrm{d}x}(af(x) + bg(x)) = a\frac{\mathrm{d}f(x)}{\mathrm{d}x} + b\frac{\mathrm{d}g(x)}{\mathrm{d}x}$ (Linearity of differentiation)
5. $\frac{\mathrm{d}}{\mathrm{d}x}(ax^n) = anx^{n-1}$ (Power rule)
6. $\frac{\mathrm{d}}{\mathrm{d}x}(n^x) = n^x\ln(n)$, $\frac{\mathrm{d}(\mathrm{e}^x)}{\mathrm{d}x} = \mathrm{e}^x$ (Derivative of exponentials)
7. $\frac{\mathrm{d}}{\mathrm{d}x}(\ln(x)) = \frac{1}{x}$ (Derivative of natural logarithms)
8. $\frac{\mathrm{d}}{\mathrm{d}x}(f_1(x)f_2(x)) = f_1(x)\frac{\mathrm{d}f_2}{\mathrm{d}x} + f_2(x)\frac{\mathrm{d}f_1}{\mathrm{d}x}$ (Product rule for two functions)
9. $\frac{\mathrm{d}}{\mathrm{d}x}(f_1f_2 \dots f_n) = f_1f_2\dots f_n\left(\frac{1}{f_1}\frac{\mathrm{d}f_1}{\mathrm{d}x} + \frac{1}{f_2}\frac{\mathrm{d}f_2}{\mathrm{d}x} + \cdots + \frac{1}{f_n}\frac{\mathrm{d}f_n}{\mathrm{d}x}\right)$ (Generalized product rule)

3.9.2 Formula for antiderivatives of functions

1. $f(x) = g(x) \implies \int f(x)\,\mathrm{d}x = \int g(x)\,\mathrm{d}x$ (Rules of Equality)
2. $\int af(x) + bg(x)\,\mathrm{d}x = a\int f(x)\,\mathrm{d}x + b\int g(x)\,\mathrm{d}x$ (Linearity of integration)
3. $n \neq -1$, $\int ax^n\,\mathrm{d}x = a\dfrac{x^{n+1}}{n+1} + C$ (Reversed power rule)
4. $\int n^x\,\mathrm{d}x = \dfrac{1}{\ln(n)}n^x + C$, $\int \mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x + C$ (Antiderative of exponentials)
5. $\int \dfrac{1}{x}\,\mathrm{d}x = \ln(x) + C$ (Antiderivative of natural logarithms)

3.9.3 Definition for various functions and constants

1. $\mathrm{e}^x = \lim\limits_{k\to 0}\sum\limits_{i\to 0}^{k}\dfrac{1}{i!} = 1 + x + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \ldots$
2. $\mathrm{e} = \lim\limits_{n\to\infty}\left(1 + \dfrac{1}{n}\right)^n$, and $\mathrm{e} = \sum\limits_{i=0}^{\infty}\dfrac{1}{i!} = 1 + \dfrac{1}{1!} + \dfrac{1}{2!} + \dfrac{1}{3!} + \ldots$
3. $\ln(x) = \lim\limits_{n\to 0}\left(\dfrac{x^h - 1}{h}\right)$

CHAPTER 4

A study of graphs with calculus

4.1 Invitation: graph of functions

If I told you to *sketch* out x^2, it'd be easy right? Just a simple parabola would do. But if I say: plot $x^4 - 2x^3 + x^2 - x + 1$? Isn't so easy now is it. One way would be to substitute in various value and plot it into the graph directly. But I'm asking for a rough sketch, not a plot. I just want the general shape and structure of the function: where are the highest and lowest points, the x-intercept, and the y-intercept. We're going to learn how to do exactly that in this chapter.

4.2 Minima and maxima: optimization problem

4.3 Tangent and normal of a graph

4.4 Newton-Raphson root finding algorithm

CHAPTER 5

Basic calculus and trigonometry

Prerequisites: *basic trigonometry.*

5.1 Invitation: oscillations, measurement and trigonometry

Trigonometry is all about measuring triangles and circles that contains them. Trigonometric functions can be used to describe oscillation, vibrations, and anything that's going in circles; therefore, they show up as a very important tool in physics. We'll also see that trigonometric functions can be used as a tool for integrating very complex integrals. So, we shall study them in this chapter.

As an overview, our trigonometric journey starts from analyzing the behavior of sine and cosine: two of the most important trigonometric functions. Then, we'll derive a polynomial expression for these functions.

After which we'll see that these expressions come up when we're dealing with oscillating system; the simplest one being the simple harmonic oscillator, aka the spring-mass system. We'll also talk about how trigonometry can be used to describe things that go around in circles; thus, the Newton's law in radial coordinate systems. From there, we'd be able to derive some pretty well-known equations in circular motion. Then, we'll finish off with the wave equation and complex numbers.

5.2 The behavior of sine and cosine

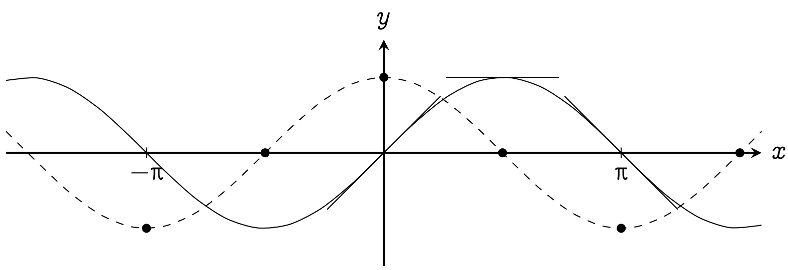


FIG. 5.1 | THE RELATION BETWEEN SINE (IN BLACK) AND COSINE (IN DASHED)

Two of the most important trigonometric functions are sine and cosine. In calculus, we study the rate of change of these functions. Let's start with the sine function. The derivative (slope) of the sine function also oscillates with the function. When $x = 0$, $\sin(x) = 0$, but the slope is rising up. When $x = \pi / 2$, $\sin(x)$ reaches its peak at 1, but it flattens down and the slope is zero. Later, it dips down at $x = \pi$, $\sin(x) = 0$, but now the slope is going down. If we plot the slope of sine at each point (shown as dots in fig. 5.1), you'll see that it resembles a cosine curve. Maybe it as well be, but we'd have to mathematically prove that.

Let's start with the definition of the derivative given in eq. (1.8),

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \lim_{h\to 0}\frac{\sin(x+h)-\sin(x)}{h}. \tag{5.1}$$

The sum of sines reads $\sin(x+h) = \sin(x)\cos(h)+\cos(x)\sin(h)$; therefore,

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \lim_{h\to 0}\frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h} \tag{5.2}$$

When $h \to 0$, $\cos(h) \to 1$

$$= \lim_{h\to 0}\frac{\sin(x)+\cos(x)\sin(h)-\sin(x)}{h} \tag{5.3}$$

$$= \cos(x)\lim_{h\to 0}\frac{\sin(h)}{h}. \tag{5.4}$$

For small angles, $\sin(h) \approx h$, therefore

$$= \cos(x)\lim_{h\to 0}\frac{h}{h} = \cos(x). \tag{5.5}$$

In conclusion, the derivative of sine is just the cosine just as we predicted with the graph.

Similarly, the derivative of cosine can also be found using the method of increments and using the sum of cosines formula $\cos(x+h) = \cos(x)\cos(h) - \sin(x)\sin(h)$

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(\cos(x)\big) = \lim_{h\to 0}\frac{\cos(x+h)-\cos(x)}{h} \tag{5.6}$$

$$= \lim_{h\to 0}\frac{\cos(x)\cos(h)-\sin(x)\sin(h)-\cos(x)}{h} \tag{5.7}$$

When $h \to 0$, $\cos(h) = 1$

$$= \lim_{h\to 0}\frac{\cos(x)-\cos(x)-\sin(x)\sin(h)}{h} \tag{5.8}$$

$$= \lim_{h\to 0}\frac{-\sin(x)\sin(h)}{h} \tag{5.9}$$

For small angles of h, $\sin(h) \approx h$

$$= \lim_{h\to 0}\big(-\sin(x)\big) = -\sin(x) \tag{5.10}$$

I.e., the derivative of cosine is the negative of sine. From those two alone, we can form a set of formulas for differentiating sines and cosines:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(\sin(x)) &= \cos(x)\\ \frac{\mathrm{d}^2}{\mathrm{d}x^2}(\sin(x)) &= \frac{\mathrm{d}}{\mathrm{d}x}(\cos(x)) = -\sin(x)\\ \frac{\mathrm{d}^3}{\mathrm{d}x^3}(\sin(x)) &= \frac{\mathrm{d}}{\mathrm{d}x}(-\sin(x)) = -\cos(x)\\ \frac{\mathrm{d}^4}{\mathrm{d}x^4}(\sin(x)) &= \frac{\mathrm{d}}{\mathrm{d}x}(-\cos(x)) = \sin(x)\end{aligned} \tag{5.11}$$

We can see that the derivative of both sine and cosine cycles in pattern of four: $\sin(x), \cos(x), -\sin(x), -\cos(x)$. From these, we can use the rules developed in chapter 3, e.g., the chain rule, power rule, etc., in order to find the derivative of other trigonometric functions.

5.3 Derivative of other trigonometric functions

Note that I wouldn't be referring back to this section often. You can continue reading the next section straightaway. But for those of you who are still interested, I'd be glad to have you here. Because this is one of the places that we need the chain rule, product rule, and power rule to help us find the derivative. So it's a good exercise. Let's start with the reciprocals of trigonometric functions first: the cosecant ($\csc(x) = \frac{1}{\sin(x)}$),

$$\frac{\mathrm{d}}{\mathrm{d}x}\csc(x) = \frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sin(x)} \tag{5.12}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{u}\cdot\frac{\mathrm{d}u}{\mathrm{d}u} \qquad \text{Chain rule: } \sin(x) = u \tag{5.13}$$

$$= \frac{\mathrm{d}}{\mathrm{d}u}u^{-1}\cdot\frac{\mathrm{d}u}{\mathrm{d}x} \tag{5.14}$$

$$= -1u^{-2}\cdot\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) \qquad \text{Power rule, } u = \sin(x) \tag{5.15}$$

$$= -\frac{1}{\sin^2(x)}\cos(x) \qquad \frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x) \tag{5.16}$$

$$= -\csc(x)\cot(x) \qquad \cot(x) = \frac{\cos(x)}{\sin(x)}. \tag{5.17}$$

And the secant function ($\sec(x) = \frac{1}{\cos(x)}$),

$$\frac{d}{dx}\sec(x) = \frac{d}{dx}\frac{1}{\cos(x)} \tag{5.18}$$

$$= \frac{d}{dx}\frac{1}{u}\cdot\frac{du}{du} \qquad \text{Chain rule: } \cos(x) = u \tag{5.19}$$

$$= \frac{d}{du}u^{-1}\cdot\frac{du}{dx} \tag{5.20}$$

$$= -1u^{-2}\cdot\frac{d}{dx}\cos(x) \qquad \text{Power rule, } u = \cos(x) \tag{5.21}$$

$$= -\frac{1}{\cos^2(x)}(-\sin(x)) \qquad \frac{d}{dx}\cos(x) = -\sin(x) \tag{5.22}$$

$$= \sec(x)\tan(x). \tag{5.23}$$

For the tangent function ($\tan(x) = \frac{\sin(x)}{\cos(x)}$), we use the product rule and the results from earlier

$$\frac{d}{dx}\tan(x) = \frac{d}{dx}\frac{\sin(x)}{\cos(x)} \tag{5.24}$$

$$= \frac{d}{dx}\sin(x)\sec(x) \tag{5.25}$$

$$= \sin(x)\frac{d}{dx}\sec(x) + \sec(x)\frac{d}{dx}\sin(x) \tag{5.26}$$

$$= \sin(x)\sec(x)\tan(x) + \sec(x)\cos(x) \tag{5.27}$$

$$= \sin(x)\frac{1}{\cos(x)}\frac{\sin(x)}{\cos(x)} + \frac{1}{\cos(x)}\cos(x) \tag{5.28}$$

$$= \left(\frac{\sin(x)}{\cos(x)}\right)^2 + 1 = \tan^2(x) + 1 = \sec^2(x) \tag{5.29}$$

The Pythagorean identity is used in the last line to convert $1 + \tan^2(x)$ to $\sec^2(x)$. Of course, we cannot miss its reciprocal function: $\cot(x) = \frac{1}{\tan(x)}$.

$$\frac{d}{dx}\cot(x) = \frac{d}{dx}\frac{1}{\tan(x)} \tag{5.30}$$

$$= \frac{d}{dx}\frac{\cos(x)}{\sin(x)} \tag{5.31}$$

$$= \frac{d}{dx}\cos(x)\csc(x) \tag{5.32}$$

$$= \cos(x)\frac{\mathrm{d}}{\mathrm{d}x}\csc(x) + \csc(x)\frac{\mathrm{d}}{\mathrm{d}x}\cos(x) \tag{5.33}$$

$$= \cos(x)\big(-\csc(x)\cot(x)\big) + \csc(x)\big(-\sin(x)\big) \tag{5.34}$$

$$= -\cos(x)\frac{1}{\sin(x)}\frac{\cos(x)}{\sin(x)} - \frac{1}{\sin(x)}\sin(x) \tag{5.35}$$

$$= -\big(\cot^2(x) + 1\big) = -\csc^2(x), \tag{5.36}$$

where we've again used the Pythagorean identity $\cot^2(x) + 1 = -\csc^2(x)$.

Altogether, I've summarized all the derivatives of trigonometric functions into table 5.1.

$f(x)$	$\frac{\mathrm{d}}{\mathrm{d}x}f(x)$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x)$	$\sec^2(x)$
$\csc(x)$	$-\csc(x)\cot(x)$
$\sec(x)$	$\sec(x)\tan(x)$
$\cot(x)$	$-\csc^2(x)$

TABLE 5.1 | THE DERIVATIVES OF TRIGONOMETRIC FUNCTIONS

5.4 Power series of various trigonometric functions

The approximation method that I'd be discussing are commonly taught to high schoolers under the name of **small angle approximation**, i.e., for $\theta \to 0$,

$$\sin(\theta) \approx \tan(\theta) \approx \theta \quad \text{and,} \quad \cos(\theta) \approx 1 - \frac{\theta^2}{2} \tag{5.37}$$

These approximations are very useful for calculating limits, e.g.,[1]

$$\lim_{\theta\to 0}\frac{\sin(n\theta)}{\theta} = \lim_{\theta\to 0}\frac{n\theta}{\theta} = n. \tag{5.38}$$

[1]Normally these limits are evaluated using the squeeze theorem. However, I don't want the mathematical intricacies to disrupt the flow of our journey right now. The full theorem will be discussed later at the end of the book.

Here, I want to focus on where these approximations really comes from: the power series development of trigonometric functions.

5.4.1 Power series development for sine and cosines

As we've seen in section 3.5.1, every function has a power series expansion

$$a_0 + a_1x^1 + a_2x^2 + a_3x^3 + \ldots\,. \tag{5.39}$$

In this section, we'll develop the power series expansion for sine and cosine. Let's start with the sine function,

$$\sin(x) = s_0 + s_1x^1 + s_2x^2 + s_3x^3 + \ldots \tag{5.40}$$

Because $\sin(0) = 0$,

$$\sin(0) = s_0 + s_1(0)^1 + s_2(0)^2 + s_3(0)^3 + \ldots \tag{5.41}$$

$$0 = s_0; \tag{5.42}$$

thus,

$$\sin(x) = s_1x^1 + s_2x^2 + s_3x^3. \tag{5.43}$$

To proceed further, we must infuse more information about the behavior of sines into the power series. Namely, that sine is an odd function: $\sin(-x) = -\sin(x)$.

$$\sin(-x) = -\sin(x) \tag{5.44}$$

$$\begin{aligned} s_1(-x)^1 + s_2(-x)^2 \\ + s_3(-x)^3 + s_4(-x)^4 + \ldots = {} & -s_1x^1 - s_2x^2 \\ & - s_3x^3 - s_4x^4 + \ldots \end{aligned} \tag{5.45}$$

$$\begin{aligned} -s_1x^1 + s_2x^2 \\ - s_3x^3 + s_4x^4 - \ldots = {} & -s_1x^1 - s_2x^2 - \\ & s_3x^3 - s_4x^4 - \ldots \end{aligned} \tag{5.46}$$

In the L.H.S., the sign oscillates between negative and positive. In order for the L.H.S. to match the R.H.S., the positive terms must vanish; therefore,

$$\sin(x) = s_1x^1 + s_3x^3 + s_5x^5 + \dots . \tag{5.47}$$

This is what we mean by "sine is an odd function": there are only odd terms in its polynomial expansion.

To find s_1, we can get rid of the multiplier x by taking the derivative of sine, which is cosine.

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x) = s_1 + 3s_3x^2 + 5s_5x^4 + \dots . \tag{5.48}$$

Substitute $x = 0$,

$$\cos(0) = 1 = s_1 + 3s_3(0)^2 + 5s_5(0)^4 + \dots \tag{5.49}$$

$$1 = s_1 \tag{5.50}$$

Now that we know s_1, we can take the derivative of sine again to get a recurrence relation.

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\sin(x) = 3\cdot 2s_3x + 5\cdot 4s_5x^3 + 7\cdot 6s_7x^5 + \dots \tag{5.51}$$

$$-\sin(x) = \frac{3!}{1!}s_3x + \frac{5!}{3!}s_5x^3 + \frac{7!}{5!}s_7x^5 + \dots \tag{5.52}$$

$$-s_1x - s_3x^3 - s_5x_5 - \dots = \frac{3!}{1!}s_3x + \frac{5!}{3!}s_5x^3 + \frac{7!}{5!}s_7x^5 + \dots \tag{5.53}$$

By matching terms with the same order,

$$\begin{aligned}
-s_1 &= \frac{3!}{1!}s_3x, \\
-s_3 &= \frac{5!}{3!}s_5x, \\
-s_5 &= \frac{7!}{5!}s_7x. \\
&\vdots
\end{aligned} \tag{5.54}$$

From $s_1 = 1$, we get that $s_3 = -\frac{1}{3!}$, $s_5 = -\frac{1}{5!}$, $s_7 = -\frac{1}{7!}$ etc. The negative sign alternates every term. We can then summarize the whole thing as

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n \tag{5.55}$$

$$= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \dots, \tag{5.56}$$

which is the polynomial expansion of the sine function.

To develop the power series expansion for cosine, we can use the same method. But we can just take the derivative of the sine power series once to get cosine:

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7\right) \tag{5.57}$$

$$\cos(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \dots \tag{5.58}$$

5.4.2 Approximation of trigonometric functions

One way a function can be approximated is by truncating its power series. To see why this is viable, consider the limit

$$\lim_{x\to 0}\sin(x) = \lim_{x\to 0}\left(x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \dots\right) \tag{5.59}$$

When $x \to 0$, all the higher order degrees term vanishes; thus,

$$\lim_{x\to 0}\sin(x) = x. \tag{5.60}$$

It tells us that when $x \approx 0$, the sine function behaves like a linear function. Illustrated in fig. 5.2,

When x gets larger, the higher order terms in the power series contributes more and more, so we need to include them. An excellent approximation for the sine function when $x \in [-\pi, \pi]$ is a truncation at the fifth term:

$$\sin(x) \approx x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \frac{1}{9!}x^9, \tag{5.61}$$

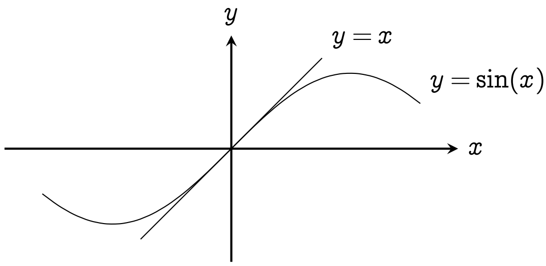


FIG. 5.2 | COMPARISON OF $y = \sin(x)$ AND $y = x$ NEAR $x = 0$

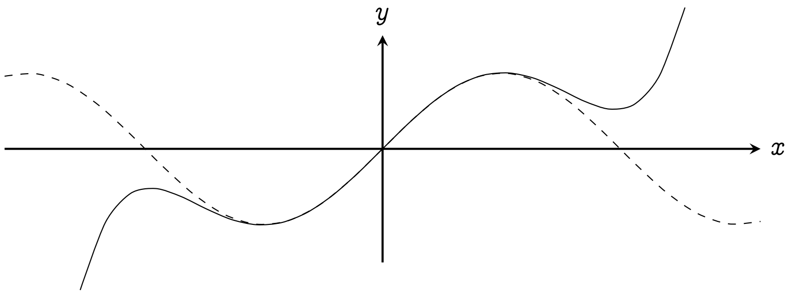


FIG. 5.3 | AN APPROXIMATION OF SINE BY TRUNCATING ITS POWER SERIES AT THE FIFTH TERM.

plotted in fig. 5.3. It is extremely accurate in that range.

For cosine,

$$\lim_{x\to 0} \cos(x) = \lim_{x\to 0}\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots\right) \approx 1 - \frac{x^2}{2!} \approx 1. \tag{5.62}$$

An acceptable approximation around zero is 1 or $1 - \frac{x^2}{2}$. Similarly, an excellent approximation when $x \in [-\pi, \pi]$ is also a truncation at the fourth term:

$$\cos(x) \approx 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 \tag{5.63}$$

plotted in fig. 5.4

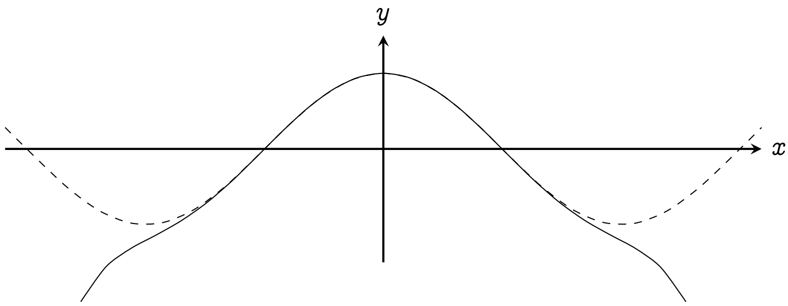


FIG. 5.4 | AN APPROXIMATION OF COSINE BY TRUNCATING ITS POWER SERIES AT THE FOURTH TERM

5.5 The harmonic oscillator

Now that we have the mathematical foundations built, let's start tackling our first oscillatory system: the harmonic oscillator. A physical harmonic oscillator can be constructed using a spring attached to a mass, illustrated in fig. 5.5. The stiffness of the spring can be described by a stiffness constant k. The higher the k, the stiffer the spring. The point $x = 0$ is set at the spring's equilibrium point, i.e., it's neither stretched or compressed. The further it is away from this equilibrium point, the more force the spring acts on the object to restore itself into the equilibrium position. By experiment, the force acted on the object for small x is a kx; the negative sign suggests that the force acts on the opposite direction from the direction of the object. Now that we know the force, we can solve for the position function. The Newton's equation, $F = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}$ then becomes

$$-kx = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} \tag{5.64}$$

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\frac{k}{m}x \tag{5.65}$$

For simplicity, let's set $c^2 = \frac{k}{m}$ where $c \in \mathbb{R}$.

I'd like to focus on the qualitative part of this system before we

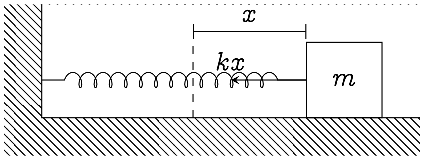


FIG. 5.5 | SETUP OF A SIMPLE HARMONIC OSCILLATOR

start to solve it analytically. The spring system is an oscillatory system. At $t = 0$, the mass is at $x = x_0$. For simplicity's sake, let $v(t = 0) = 0$. The spring's restoration force pulls on the object, accelerating it. As the object gets closer to the equilibrium, the restoration force gets smaller. The continuously accelerates until it overshoots the equilibrium position; after which, the spring is compressed and tries to push back on the object. This continues until the object finally stops at the other end. Then, the spring pulls back, the object overshoots the equilibrium position again and stops at the original position, completing one oscillation cycle.

From that description alone, we can already extract some key features to the spring system.

1. The object reaches zero velocity when the spring is fully stretched or compressed.
2. The acceleration is highest when the spring is fully stretched or compressed.
3. The velocity is highest when the object passes through the equilibrium position.

As we'll see, these features will all be reflected in the result.

There are actually two ways of solving eq. (5.65). The first one is to just guess the solution straightaway! It's called the **ansatz method**, which is commonly taught in Massachusets Institute of Technology (MIT)

5.6 Polar coordinate systems

Part II

The extensions

Part III

The fundamentals, reimagined: real analysis

CHAPTER 6

Constructing the real numbers

Abstract

Prerequisites: *intuitions of set theory, basic discrete math*

As far as the "calculus" part of this book goes, it doesn't really delve deep into the proof: the backbones of how structures work together. For example, how do you know that in the definition of derivatives, if $h \to 0$, the value actually converges to something. To be frank, real analysis is quite abstracted away from the physical reality therefore, it's a bit dry. This is quite the double edged sword of math. With real analysis, we have the power to tell clearly if something is true or not. However, most times it's mistakenly used to the bones: too abstract that the learner does not have any clear concepts leftover. All the rest is just some meaningless mathematical notation that's floating in the air. And I don't want that.

The goal for the real analysis part of this book is to provide an enjoyable experience delving in to the proofs behind the backbones of calculus. Therefore, I shall try to illustrate everything with diagrams so it's simple to visualize and not too abstracted away from reality. Now that you know my intentions, let's start.

6.1 The mindset of real analysis

Before we study the reals, we must know the mindset of real analysis first. Analysis is used to generalize and study the *exact* behaviors of mathematical entities. In real analysis, we study the *reals*. Most of the stuffs in mathematics were built way before real analysis. However, it's not rigorous and it's prone to error. Here, real analysis comes to play.

We *abstract* properties of mathematical identities away from the numbers, and we generalize it. But we can't just choose everything, we must be very wise. The properties that we select to be true are called **axioms** After all the decision has been done, we must find the most general mathematical entity that satisfies it. And thus, we shall begin with the most basics of analysis: set theories.

6.2 The Zermelo–Fraenkel set theory

In here, we shall explore what's the backbones of sets that will lead to the mechanics of numbers. And here arises the set theory. Firstly, a **set** is a group of things, whether it be mathematical entities or real world objects. If two sets contains the same elements, then it's the same set. That means, set does not care about permutation. A wiser way to state this is

Axiom 1: Axiom of Extensionality

Two sets are the same if they have the same elements.

$$\forall X \forall Y [\forall z (z \in X \iff z \in X) \implies X = Y]. \tag{6.1}$$

Translation: Set X and Y will be equal iff for all elements z, z is in both X and Y.

which just means that "A *set is uniquely determined by its members*".

Then, we also have to define that a set cannot have the same elements that is,

Axiom 2: Axiom of foundation

Every non-empty set x contains a member y such that x and y are *disjoint*.

$$\forall x [x \neq \emptyset \implies \exists y ((y \in x) \wedge (y \cup x) = \emptyset)] \tag{6.2}$$

Translation: For all non-empty set x, there exists y where both y

Part IV

Beyond imagination: complex analysis

APPENDIX A

Fundamental of physics

APPENDIX B

The binomial theorem

Index

Bibliography

[1] J. Turnbull, D. Lea, D. Parkinson, P. Phillips, B. Francis, S. Webb, V. Bull, and M. Ashby, *Oxford advanced learner's dictionary, 8th edition: paperback*, Oxford Advanced Learner's Dictionary, 8th Edition (OUP Oxford, 2010).

[2] Wikipedia contributors, *Summation: powers and logarithm of arithmetic progressions*, (Feb. 4, 2024) https://en.wikipedia.org/wiki/Summation#Powers_and_logarithm_of_arithmetic_progressions.