No BULLSHIT guide to

MATH & PHYSICS

DIFF v5.0 (Aug 2014) $\rightarrow$ v5.1 (Jul 2016)

Ivan Savov

July 14, 2016

Concept map

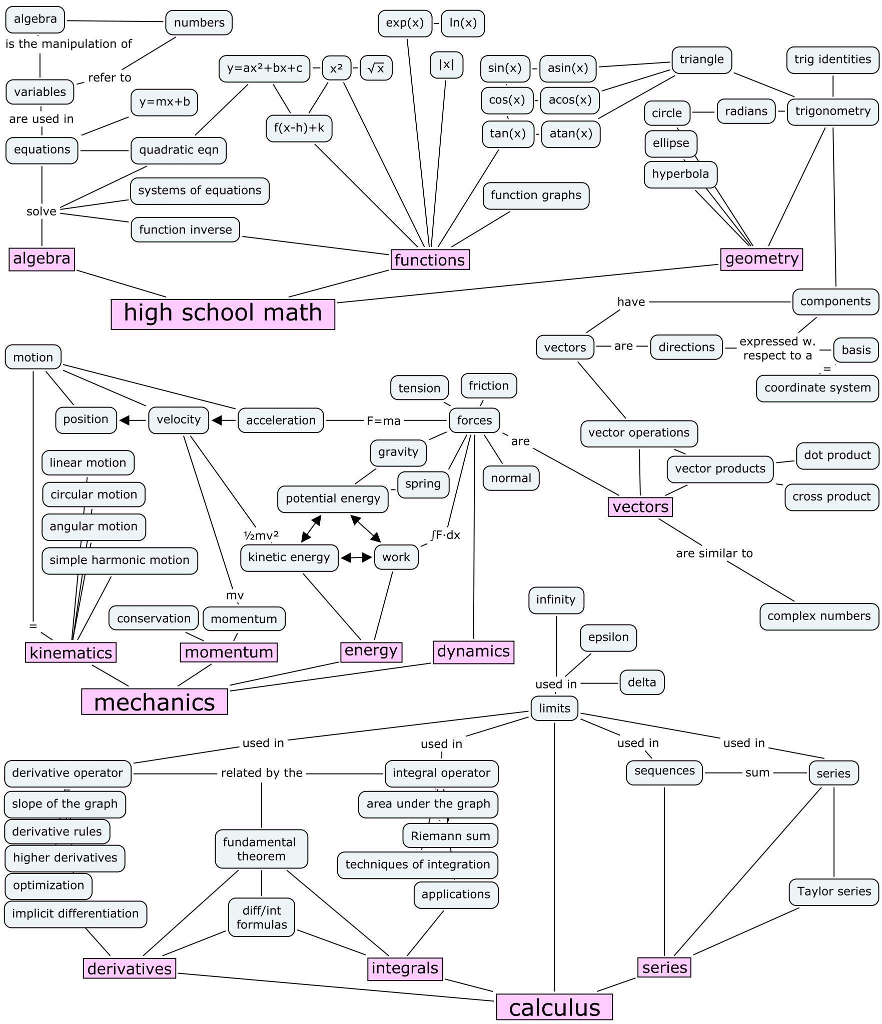


Figure 1: ~~Each concept~~ Diagram showing the concepts, topics, and subjects covered in ~~this diagram corresponds~~ the book. Consult the index on page ?? to ~~one section~~ find the pages where each concept is defined and used. Not all connections are shown in order not to overwhelm the ~~book~~diagram.

Preface

This book contains lessons on topics in math and physics, written in a style that is jargon-free and to the point. Each lesson covers one concept at the depth required for a first-year university-level course. The main focus of this book is to highlight the intricate connections between the concepts of math and physics. Seeing the similarities and parallels between the concepts is the key to understanding.

Why?

The genesis of this book dates back to my student days when I was required to purchase expensive textbooks for my courses. Not only are these textbooks expensive, they are also tedious to read. Who has the energy to go through thousands of pages of explanations? I began to wonder, "What's the deal with these thick books?" Later, I realized mainstream textbooks are long because the textbook industry wants to make more profits. You don't need to read 1000 pages to learn calculus; the numerous full-page colour pictures and the repetitive text that are used to "pad" calculus textbooks are there to make the ~~$130~~ $200 price seem reasonable.

Looking at this situation, I said to myself, "something must be done," and I sat down and wrote a modern textbook to explain math and physics clearly, concisely, and affordably. There was no way I was going to let mainstream publishers ruin the learning experience of these beautiful subjects for the next generation of students.

How?

The sections in this book are **self-contained tutorials**. Each section covers the definitions, formulas, and explanations associated with a single topic. You can therefore read the sections in any order you find logical. Along the way, you will learn about the *connections* between the concepts of calculus and mechanics. Understanding mechanics is much easier if you know the ideas of calculus. At the same time, the ideas behind calculus are best illustrated through concrete physics examples. Learning the two subjects simultaneously is the best approach.

In order to make the study of mechanics and calculus accessible for all readers, we'll begin with a review chapter on numbers, algebra, equations, functions, and other prerequisite concepts. If you feel a little rusty on those concepts, be sure to check out Chapter 1.

Each chapter ends with a section of **practice problems** designed to test your understanding of the concepts developed in that chapter. Make sure you spend plenty of time on these problems to practice what you've learned. Figuring out how to use an equation on your own in the process of solving a problem is a much more valuable experience than simply memorizing the equation.

For optimal learning efficiency, I recommend that you spend as much time working through the practice problems as you will spend reading the lessons. The problems you find difficult to solve will tell you which sections of the chapter you need to revisit. An additional benefit of testing your skills on the practice problems is that you'll be prepared in case a teacher ever tries to test you.

Throughout the book, I've included **links to ~~Internet~~ internet resources** like animations, demonstrations, and webpages with further reading material. Once you understand the basics, you'll be able to understand far more ~~Internet~~ internet resources. The links provided are a starting point for further exploration.

Is this book for you?

My aim is to make learning calculus and mechanics more accessible. Anyone should be able to open this book and become proficient in calculus and mechanics, regardless of their mathematical background.

The book's primary intended audience is students. Students taking a mechanics class can read the chapters sequentially until Chapter 4, and optionally read Chapter 5 for ~~general knowledge~~bonus points. Taking a calculus course? Skip ahead directly to the calculus chapter (Chapter 5). High school students or university students taking a precalculus class will benefit from reading Chapter 1, which is a concise but thorough review of fundamental math concepts like numbers, equations, functions, and trigonometry.

MECH CLASS	CALC CLASS	PRECALC CLASS
Ch. 1	Ch. 1	Ch. 1
Ch. 2	Ch. 2	Ch. 2†
Ch. 3		
Ch. 4		
Ch. 5†	Ch. 5	

$\dagger$ = optional reading.

Non-students, don't worry: you don't need to be taking a class in order to learn math. Independent learners interested in learning university-level material will find this book very useful. Many university graduates read this book to remember the calculus they learned back in their university days.

In general, anyone interested in rekindling their relationship with mathematics should consider this book as an opportunity to repair the broken connection. Math is good stuff; you shouldn't miss out on it. People who think they absolutely *hate* math should read Chapter 1 as therapy.

About the author

I have been teaching math and physics for more than ~~10~~ 15 years as a private tutor. My tutoring experience has taught me how to explain concepts that people find difficult to understand. I've had the chance to experiment with different approaches for explaining challenging material. Fundamentally, I've learned from teaching that understanding connections between concepts is much more important than memorizing facts. It's not about how many equations you know, but about knowing how to get from one equation to another.

I completed my undergraduate studies at McGill University in electrical engineering, then did a M.Sc. in physics, and recently completed a Ph.D. in computer science. In my career as a researcher, I've been fortunate to learn from very inspirational teachers, who had the ability to distill the essential ideas and explain things in simple language. With my writing, I want to recreate the same learning experience for you. I founded the `Minireference Co.` to revolutionize the textbook industry. We make textbooks that don't suck.

Ivan Savov
Montreal, 2014

Introduction

The last two centuries have been marked by tremendous technological advances. Every sector of the economy has been transformed by the use of computers and the advent of the ~~Internet~~internet. There is no doubt technology's importance will continue to grow in the coming years.

The best part is that you don't need to know how technology works to use it. You need not understand how ~~Internet~~ internet protocols operate to check your email and find original pirate material. You don't need to be a programmer to tell a computer to automate repetitive tasks and increase your productivity. However, when it comes to building *new* things, understanding becomes important. One particularly useful skill is the ability to create mathematical models of real-world situations. The techniques of mechanics and calculus are powerful building blocks for understanding the world around us. This is why these courses are taught in the first year of university studies: they contain keys that unlock the rest of science and engineering.

Calculus and mechanics can be difficult subjects. Understanding the material isn't hard *per se*, but it takes patience and practice. Calculus and mechanics become much easier to absorb when you break down the material into manageable chunks. It is most important you learn the *connections* between concepts.

Before we start with the equations, it's worthwhile to preview the material covered in this book. After all, you should know what kind of trouble you're getting yourself into.

Chapter 1 is a comprehensive review of math fundamentals including algebra, equation solving, and functions. The exposition of each topic is brief to make for easy reading. This chapter is highly recommended for readers who haven't looked at math recently; if you need a refresher on math, Chapter 1 is for you. It is extremely important to firmly grasp the basics. What is $\sin(0)$? What is $\sin(\pi/4)$? What does the graph of $\sin(x)$ look like? Arts students interested in enriching their cultural insight with knowledge that is 2000+ years old can read this chapter as therapy to recover from any damaging educational experiences they may have encountered in high school.

In Chapter 2, we'll look at how techniques of high school math can be used to describe and model the world. We'll learn about the basic laws that govern the motion of objects in one dimension and the mathematical equations that describe the motion. By the end of this chapter, you'll be able to predict the flight time of a ball thrown in the air.

In Chapter 3, we~~will~~'ll learn about vectors. Vectors describe directional quantities like forces and velocities. We need vectors to properly understand the laws of physics. Vectors are used in many areas of science and technology, so becoming comfortable with vector calculations will pay dividends when learning other subjects.

Chapter 4 is all about mechanics. We'll study the motion of objects, predict their future trajectories, and learn how to use abstract concepts like momentum and energy. Science students who "hate" physics can study this chapter to learn how to use the 20 main equations and laws of physics. You'll see physics is actually quite simple.

Chapter 5 covers topics from differential calculus and integral calculus. We~~will~~'ll study limits, derivatives, integrals, sequences, and series. You'll find that ~~100~~ 120 pages are enough to cover all the concepts in calculus, as well as illustrate them with examples and practice exercises.

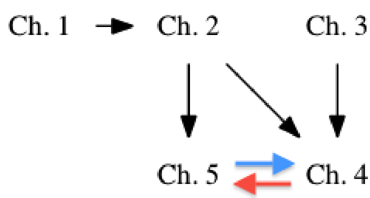


Figure 2: The prerequisite structure for the chapters in this book.

Calculus and mechanics are often taught as separate subjects. It shouldn't be like that! If you learn calculus without mechanics, it will be boring. If you learn physics without calculus, you won't truly understand. The exposition in this book covers both subjects in an integrated manner and aims to highlight the connections between them. Let's dig in.

Chapter 1

Math fundamentals

In this chapter we'll review the fundamental ideas of mathematics, including numbers, equations, and functions. To understand college-level textbooks, you need to be comfortable with mathematical calculations. Many people have trouble with math, however. Some people say they *hate* math, or could never learn it. It's not uncommon for children who score poorly on their school math exams to develop math complexes in their grown lives. If you are carrying any such emotional baggage, you can drop it right here and right now.

Do NOT worry about math! You are an adult, and you can learn math much more easily than when you were ~~in high school~~a kid. We'll review *everything* you need to know about high school math, and by the end of this chapter, you'll see that math is nothing to worry about.

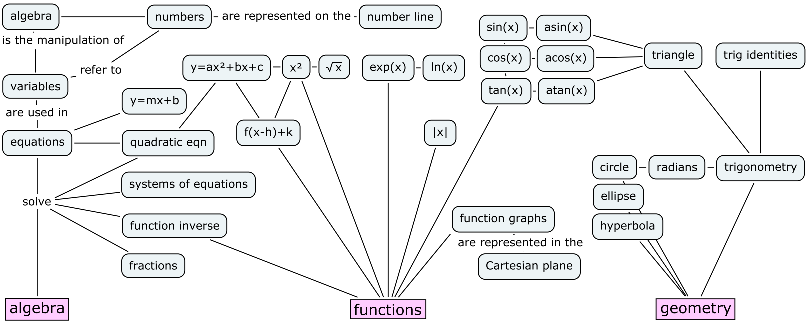


Figure 1.1: A concept map showing the mathematical topics that we will cover in this chapter. We'll learn how to solve equations using algebra, how to model the world using functions, and how to think geometrically. The material in this chapter is required for your understanding of the more advanced topics in this book.

1.1 Solving equations

Most math skills boil down to being able to manipulate and solve equations. Solving an equation means finding the value of the unknown in the equation.

Check this shit out:

$$x^2 - 4 = 45.$$

To solve the above equation is to answer the question "What is x?" More precisely, we want to find the number that can take the place of x in the equation so that the equality holds. In other words, we're asking,

"Which number times itself minus four gives 45?"

That is quite a mouthful, don't you think? To remedy this verbosity, mathematicians often use specialized ~~mathematical symbols~~ symbols to describe math operations. The problem is that these specialized symbols can be very confusing. Sometimes even the simplest math concepts are inaccessible if you don't know what the symbols mean.

What are your feelings about math, dear reader? Are you afraid of it? Do you have anxiety attacks because you think it will be too difficult for you? Chill! Relax, my brothers and sisters. There's nothing to it. Nobody can magically guess what the solution to an equation is immediately. To find the solution, you must break the problem down into simpler steps.

To find x, we can manipulate the original equation, transforming it into a different equation (as true as the first) that looks like this:

$$x \;=\; \text{only numbers}.$$

That's what it means to *solve*. The equation is solved because you can type the numbers on the right-hand side of the equation into a calculator and obtain the numerical value of x that you're seeking.

By the way, before we continue our discussion, let it be noted: the equality symbol ($=$) means that all that is to the left of $=$ is equal to all that is to the right of $=$. To keep this equality statement true, **for every change you apply to the left side of the equation, you must apply the same change to the right side of the equation**.

To find x, we need to correctly manipulate the original equation into its final form, simplifying it in each step. The only requirement is that the manipulations we make transform one true equation into another true equation. Looking at our earlier example, the first simplifying step is to add the number four to both sides of the equation:

$$x^2 - 4 + 4 = 45 + 4,$$

Multiplication

You can also multiply numbers together.

$$ab = \underbrace{a + a + \cdots + a}_{b \text{ times}} = \underbrace{b + b + \cdots + b}_{a \text{ times}}.$$

Note that multiplication can be defined in terms of repeated addition.

The visual way to think about multiplication is as an area calculation. The area of a rectangle of base a and height b is equal to ab. A rectangle with a height equal to its base is a square, and this is why we call $aa = a^2$ "a squared."

Multiplication of numbers is also commutative, $ab = ba$; and associative, $abc = (ab)c = a(bc)$. In modern notation, no special symbol is used to denote multiplication; we simply put the two factors next to each other and say the multiplication is *implicit*. Some other ways to denote multiplication are $a \cdot b$, $a \times b$, and, on computer systems, $a * b$.

Division

Division is the inverse operation of multiplication.

$$a/b = \frac{a}{b} = \text{ one } b^{\text{th}} \text{ of } a.$$

Whatever a is, you need to divide it into b equal parts and take one such part. Some texts denote division as $a \div b$.

Note that you cannot divide by 0. Try it on your calculator or computer. It will say "`error divide by zero`" because this action simply doesn't make sense. After all, what would it mean to divide something into zero equal parts?

Exponentiation

Often an equation calls for us to multiply things together many times. The act of multiplying a number by itself many times is called *exponentiation*~~, and we denote this operation as~~. We denote "a exponent n" using a superscript:

$$a^{\cancel{b}\,\cancel{=}\,n} \equiv \underbrace{aaa\cdots a}_{\cancel{b\ times}\, n \text{ times}}.$$

Note we used the symbol "$\equiv$" instead of the basic "$=$" to indicate we're defining a new quantity a^b, which is equal to the expression on the right. In this book, we'll use the symbol "$\equiv$" whenever we define new mathematical variables or expressions. Some other books use the notation "$:=$" or simply "$=$" when defining new quantities.

Negative exponents We can also encounter negative exponents. The negative in the exponent does not mean "subtract," but rather "divide by":

$$a^{-b-n} \equiv \frac{1}{a^n} = \frac{1}{a^b} = \frac{1}{\underbrace{aaa\cdots a}_{b\ times}}\,\frac{1}{\underbrace{aaa\cdots a}_{n\text{ times}}}.$$

To understand why negative exponents correspond to division, consider the following calculation:

$$a^m a^n = \underbrace{aaa\cdots a}_{m\text{ times}}\ \underbrace{aaa\cdots a}_{n\text{ times}} = \underbrace{aaaaaa\cdots\cdots aa}_{m+n\text{ times}} = a^{m+n}.$$

This is a general rule for exponential expressions "$a^m a^n = a^{m+n}$," or "add the exponents together when multiplying exponential expressions" if you prefer words. Defining $a^{-n} \equiv \frac{1}{a^n}$ ensures the rule $a^m a^n = a^{m+n}$ will continue to be valid when for negative exponents:

$$a^m a^{-n} = \underbrace{aaaaa\cdots aa}_{m\text{ times}}\ \frac{1}{\underbrace{a\cdots a}_{n\text{ times}}} = \frac{\overbrace{aaaaa\cdots aa}^{m\text{ times}}}{\underbrace{a\cdots a}_{n\text{ times}}} = \underbrace{aa\cdots a}_{m-n\text{ times}} = a^{m-n}.$$

For example the expression 2^{-3} corresponds to $\frac{1}{2^3} = \frac{1}{8}$. If we multiply together 2^5 and 2^{-3} we obtain $2^5 \cdot 2^{-3} = 2^{5-3} = 2^2 = 4$.

We defer the detailed discussion on exponents until Section 1.8, but since we already opened the subject, we should also mention the third special case of exponents.

Fractional exponents We discussed positive and negative exponents but what about exponents that are fractions? Fractional exponents describe square-root-like operations:

$$a^{\frac{1}{2}} \equiv \sqrt{a} \equiv \sqrt[2]{a}, \qquad a^{\frac{1}{3}} \equiv \sqrt[3]{a}, \qquad a^{\frac{1}{4}} \equiv \sqrt[4]{a} = a^{\frac{1}{2}\frac{1}{2}} = a^{\frac{1}{2}\frac{1}{2}} = \sqrt{\sqrt{a}}.$$

~~Square root $\sqrt{x}$ is the inverse operation of~~ Recall we saw the square-root operation $\sqrt{\ }$ in the previous section, where we used it to undo the effect of the x^2 ~~. Similarly, for any~~ operation. More generally, the "n~~we define the function $\sqrt[n]{x}$ (the n~~th root" function $\sqrt[n]{x}$ is the inverse of ~~x) to be the inverse function of~~ the function x^n.

~~It's worth clarifying what "taking the n^{th} root" means and understanding when to use this operation. The n^{th} root of a is a number which, when multiplied together n times, will give a~~If this is the first time you're seeing square roots, you might think you'll have to learn lots of new rules for manipulating square-root expressions. Or maybe you have experience with the rules of "squiggle math" already? What kind of emotions do expressions like $\sqrt[3]{27}\sqrt[3]{8}$ stir up in you? Chill! There's no new math to learn and no rules to memorize. All the "squiggle math" rules are consequences of the general rule $a^b a^c = a^{b+c}$ applied to expressions where the exponents are fractions. For example, a cube root satisfies

$$\sqrt[3]{a}\sqrt[3]{a}\sqrt[3]{a} = \sqrt[3]{a}^3 a^{\frac{1}{3}}a^{\frac{1}{3}}a^{\frac{1}{3}} = a^{\frac{1}{3}+\frac{1}{3}+\frac{1}{3}} = \sqrt[3]{a^3} a^1 = a.$$

Do you see why $\sqrt[3]{x}$ and x^3 are inverse operations? ~~The fractional exponent notation makes the meaning of roots much more explicit.~~ The number $\sqrt[3]{a}$ is the one-third of the number a with respect to multiplication (multiplying $\sqrt[3]{a}$ by itself three times produces a). We say "one third *with respect to multiplication*," because the usual meaning of "one third of a" is with respect to the addition operation (adding together three copies of $\frac{a}{3}$ produces a).

The n^{th} root of a is a number which, when multiplied together n times, will give a. The "n^{th} root of a" can be denoted in two equivalent ways:

$$\sqrt[n]{a} \equiv a^{\frac{1}{n}}.$$

~~The symbol "$\equiv$" stands for "is equivalent to" and is used when two mathematical objects are identical. Equivalence is a stronger relation than equality. Writing $\sqrt[n]{a} = a^{\frac{1}{n}}$ indicates we've found two mathematical expressions (the left-hand side and the right-hand side of the equality) that happen to be equal to each other. It is more mathematically precise to write $\sqrt[n]{a} \equiv a^{\frac{1}{n}}$, which tells us $\sqrt[n]{a}$ and $a^{\frac{1}{n}}$ are two different ways of denoting the *same* mathematical object~~Using this definition and the general rule $a^b a^c = a^{b+c}$ allows us to simplify all kinds expressions. For example, we can simplify $\sqrt[4]{a}\sqrt[4]{a}$ by rewriting it as follows $\sqrt[4]{a}\sqrt[4]{a} = a^{\frac{1}{4}}a^{\frac{1}{4}} = a^{\frac{1}{4}+\frac{1}{4}} = a^{\frac{1}{2}} = \sqrt{a}$. We can also simplify the expression $\sqrt[3]{27}\sqrt[3]{8}$ by rewriting it as $27^{\frac{1}{3}}8^{\frac{1}{3}}$ then simplifying it as follows $27^{\frac{1}{3}}8^{\frac{1}{3}} = (3\cdot 3\cdot 3)^{\frac{1}{3}}(2\cdot 2\cdot 2)^{\frac{1}{3}} = 3\cdot 2 = 6$.

~~The n^{th} root of a is equal to~~ Let's verify the claim that $\sqrt[n]{a}$ equals "one n^{th} of a with respect to multiplication. ~~To find~~" To obtain the whole number, we must multiply the number ~~$a^{\frac{1}{n}}$~~ $\sqrt[n]{a}$ times itself n times:

$$\underbrace{a^{\frac{1}{n}}a^{\frac{1}{n}}a^{\frac{1}{n}}a^{\frac{1}{n}}\cdots a^{\frac{1}{n}}a^{\frac{1}{n}}}_{n\text{ times}}(\sqrt[n]{a})^n = \left(a^{\frac{1}{n}}\right)^n = \underbrace{a^{\frac{1}{n}}a^{\frac{1}{n}}a^{\frac{1}{n}}a^{\frac{1}{n}}\cdots a^{\frac{1}{n}}a^{\frac{1}{n}}}\,{}_{n\text{ times}} = a^{\frac{n}{n}} = a^1 =$$

The n-fold product of $\frac{1}{n}$-fractional exponents of any number produces that number ~~with~~ raised to exponent one, therefore the inverse operation of $\sqrt[n]{x}$ is x^n.

Dear readers, I know I've been throwing a lot of math symbols at you in the last paragraphs, but I don't want you to worry about all the details yet. In Section 1.4 we'll define *inverse functions* more formally and in Section 1.8 we'll discuss all the rules for manipulating exponents. For now, you can think of the above formulas as "exposure therapy" that aims to give you some idea of more complicated-looking math. Continuing in the spirit of "exposure therapy," I'll give you one more paragraph about exponents.

The commutative law of multiplication $ab = ba$ implies that we can ~~see~~ write any fraction $\frac{a}{b}$ in two ~~different~~ other equivalent ways: $\frac{a}{b} = a\frac{1}{b} = \frac{1}{b}a$. We multiply by a then divide the result by b, or first we divide by b and then multiply the result by a. Similarly, when we have a fraction in the exponent, we can write the answer in two equivalent ways:

$$a^{\frac{2}{3}} = \sqrt[3]{a^2} = (\sqrt[3]{a})^2, \qquad a^{-\frac{1}{2}} = \frac{1}{a^{\frac{1}{2}}} = \frac{1}{\sqrt{a}}, \qquad a^{\frac{m}{n}} = \left(\sqrt[n]{a}\right)^m = \sqrt[n]{a^m}.$$

Make sure the above notation makes sense to you. As an exercise, try computing $5^{\frac{4}{3}}$ on your calculator and check that you obtain $8.54987973\ldots$ as the answer.

Operator precedence

There is a standard convention for the order in which mathematical operations must be performed. The basic algebra operations have the following precedence:

1. Exponents and roots
2. Products and divisions
3. Additions and subtractions

For instance, the expression $\cancel{5 \times 3^2 + 13}$ $5 \cdot 3^2 + 13$ is interpreted as "first find the square of 3, then multiply it by 5, and then add 13." Parenthesis are needed to carry out the operations in a different order: to multiply 5 times 3 first and *then* take the square, the equation should read $\cancel{(5 \times 3)^2 + 13}$$(5 \cdot 3)^2 + 13$, where parenthesis indicate that the square acts on $\cancel{(5 \times 3)}$ $(5 \cdot 3)$ as a whole and not on 3 alone.

Exercises

E1.1 Indicate which number sets the following numbers belong to.

(a) -2 (b) $\sqrt{-3}$ (c) $8^{\frac{1}{3}}$ (d) $\frac{5}{3}$ (e) $\frac{\pi}{2}$

1.1 (a) $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$; (b) $\mathbb{C}$; (c) $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$; (e) $\mathbb{Q}, \mathbb{R}, \mathbb{C}$; (d) $\mathbb{R}, \mathbb{C}$.

E1.2 Calculate the values of the following exponential expressions:

(a) $\sqrt{2}(\pi)2^{\frac{1}{2}}$ (b) $8^{\frac{2}{3}} + 8^{-\frac{2}{3}}$ (c) $\left(\frac{(\sqrt[3]{c})^3}{c}\right)^{77}$ (d) $\left(\frac{x^2\sqrt{x^4}}{x^3}\right)^2$

1.2 (a) 2π; (b) $4 + \frac{1}{4} = 4.25$; (c) 1; (d) x^2.

E1.3 Calculate the values of the following expressions:

(a) $2^3 3 - 3$ (c) $\frac{4-2}{3^3}(6 \cdot 7 - 41)$
(b) $2^3(3-3)$

1.3 (a) 21; (b) 0; (c) $\frac{2}{27}$.

Other operations

We can define all kinds of operations on numbers. The ~~above three are special~~ addition, subtraction, multiplication, division, and exponentiation operations are "special" operations since they feel simple and intuitive to apply, but we can also define arbitrary transformations on numbers. We call these transformations *functions*. Before we learn about functions, let's first cover variables.

By using the inverse function (denoted f^{-1}) we "undo" the effects of f. ~~Then we~~ We apply the inverse function f^{-1} to both sides of the equation to obtain

$$f^{-1}(f(x)) = x = f^{-1}(c)\,.$$

By definition, the inverse function f^{-1} performs the opposite action of the function f so together the two functions cancel each other out. We have $f^{-1}(f(x)) = x$ for any number x.

Provided everything is kosher (the function f^{-1} must be defined for the input c), the manipulation we made above is valid and we have obtained the answer $x = f^{-1}(c)$.

The above example introduces the notation f^{-1} for denoting the function's *inverse*. This notation is borrowed from the notion of inverse numbers: multiplication by the number a^{-1} is the inverse operation of multiplication by the number a: $a^{-1}ax = 1x = x$. In the case of functions, however, the negative-one exponent does not refer to "one over-$f(x)$" as in $\frac{1}{f(x)} = (f(x))^{-1}$; rather, it refers to the function's inverse. In other words, the number $f^{-1}(y)$ is equal to the number x such that $f(x) = y$.

Be careful: sometimes applying the inverse leads to multiple solutions. For example, the function $f(x) = x^2$ maps two input values (x and $-x$) to the same output value $x^2 = f(x) = f(-x)$. The inverse function of $f(x) = x^2$ is $f^{-1}(x) = \sqrt{x}$, but both $x = +\sqrt{c}$ and $x = -\sqrt{c}$ are solutions to the equation $x^2 = c$. In this case, this equation's solutions can be indicated in shorthand notation as $x = \pm\sqrt{c}$.

Formulas

Here is a list of common functions and their inverses:

$$\begin{array}{rcl}
\text{function } f(x) & \Leftrightarrow & \text{inverse } f^{-1}(x) \\
x+2 & \Leftrightarrow & x-2 \\
2x & \Leftrightarrow & \frac{1}{2}x \\
\cancel{-x}-1x & \Leftrightarrow & \cancel{-x}-1x \\
x^2 & \Leftrightarrow & \pm\sqrt{x} \\
2^x & \Leftrightarrow & \log_2(x) \\
3x+5 & \Leftrightarrow & \frac{1}{3}(x-5) \\
a^x & \Leftrightarrow & \log_a(x) \\
\exp(x) \equiv e^x & \Leftrightarrow & \ln(x) \equiv \log_e(x)
\end{array}$$

$$\begin{aligned} \sin(x) \quad &\Leftrightarrow \quad \sin^{-1}(x) \equiv \arcsin(x) \\ \cos(x) \quad &\Leftrightarrow \quad \cos^{-1}(x) \equiv \arccos(x) \end{aligned}$$

The function-inverse relationship is ~~*reflexive*~~*symmetric*—if you see a function on one side of the above table (pick a side, any side), you'll find its inverse on the opposite side.

Example

Let's say your teacher doesn't like you and right away, on the first day of class, he gives you a serious equation and tells you to find x:

$$\log_5\left(3 + \sqrt{6\sqrt{x} - 7}\right) = 34 + \sin(5.5) - \Psi(1).$$

See what I mean when I say the teacher doesn't like you?

First, note that it doesn't matter what Ψ (the capital Greek letter *psi*) is, since x is on the other side of the equation. You can keep copying $\Psi(1)$ from line to line, until the end, when you throw the ball back to the teacher. "My answer is in terms of *your* variables, dude. *You* go figure out what the hell Ψ is since you brought it up in the first place!" By the way, it's not actually recommended to quote me verbatim should a situation like this arise. The same goes with $\sin(5.5)$. If you don't have a calculator handy, don't worry about it. Keep the expression $\sin(5.5)$ instead of trying to find its numerical value. In general, try to work with variables as much as possible and leave the numerical computations for the last step.

Okay, enough beating about the bush. Let's just find x and get it over with! On the right-hand side of the equation, we have the sum of a bunch of terms with no x in them, so we'll leave them as they are. On the left-hand side, the outermost function is a logarithm base 5. Cool. Looking at the table of inverse functions we find the exponential function is the inverse of the logarithm: $a^x \Leftrightarrow \log_a(x)$. To get rid of $\log_5$, we must apply the exponential function base 5 to both sides:

$$5^{\log_5\left(3+\sqrt{6\sqrt{x}-7}\right)} = 5^{34+\sin(5.5)-\Psi(1)},$$

which simplifies to

$$3 + \sqrt{6\sqrt{x} - 7} = 5^{34+\sin(5.5)-\Psi(1)},$$

since 5^x cancels $\log_5 x$.

From here on, it is going to be as if Bruce Lee walked into a place with lots of bad guys. Addition of 3 is undone by subtracting 3 on both sides:

$$\sqrt{6\sqrt{x} - 7} = 5^{34+\sin(5.5)-\Psi(1)} - 3.$$

To undo a square root we take the square:

$$6\sqrt{x} - 7 = \left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2.$$

Add 7 to both sides,

$$6\sqrt{x} = \left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2 + 7,$$

divide by 6

$$\sqrt{x} = \frac{1}{6}\left(\left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2 + 7\right),$$

and square again to find the final answer:

$$x = \left[\frac{1}{6}\left(\left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2 + 7\right)\right]^2.$$

Did you see what I was doing in each step? Next time a function stands in your way, hit it with its inverse so it knows not to challenge you ever again.

Discussion

The recipe I have outlined above is not universally applicable. Sometimes x isn't alone on one side. Sometimes x appears in several places in the same equation. In these cases, you can't effortlessly work your way, Bruce Lee-style, clearing bad guys and digging toward x—you need other techniques.

The bad news is there's no general formula for solving complicated equations. The good news is the above technique of "digging toward the x" is sufficient for 80% of what you are going to be doing. You can get another 15% if you learn how to solve the quadratic equation:

$$ax^2 + bx + c = 0.$$

Solving third-degree polynomial equations like $ax^3 + bx^2 + cx + d = 0$ with pen and paper is also possible, but at this point you might as well start using a computer to solve for the unknowns.

There are all kinds of other equations you can learn how to solve: equations with multiple variables, equations with logarithms, equations with exponentials, and equations with trigonometric functions. The principle of "digging" toward the unknown by applying inverse functions is the key for solving all these types of equations, so be sure to practice using it.

Exercises

E1.4 Solve for x is the following equations:

(a) $3x = 6$ (b) $\log_5(x) = 2$ (c) $\log_{10}(\sqrt{x}) = 1$

1.4 (a) $x = 2$; (b) $x = 25$; (c) $x = 100$.

E1.5 Find the function inverse and use it to solve the given equation:

a) $f(x) = \sqrt{x}$, $f(x) = 4$. b) $g(x) = e^{-2x}$, $g(x) = 1$.

1.5 a) $f^{-1}(x) = x^2$, $x = 16$. b) $g^{-1}(x) = -\frac{1}{2}\ln(x)$, $x = 0$.

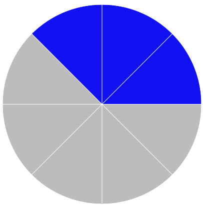
1.5 Fractions

The set of rational numbers $\mathbb{Q}$ is the set of numbers that can be written as a *fraction* of two integers:

$$\mathbb{Q} \equiv \left\{ \frac{m}{n} \,\middle|\, m \text{ and } n \text{ are in } \mathbb{Z} \text{ and } n \neq 0 \right\},$$

where $\mathbb{Z}$ denotes the set of integers $\mathbb{Z} \equiv \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.

Fractions describe what happens when a *whole* is cut into n equal parts and we are given m of those parts. For example, the fraction $\frac{3}{8}$ describes having three parts out of a whole cut into eight parts, hence the name "three eighths." We read $\frac{1}{4}$ either as *one over four* or *one quarter*, which is also equal to 0.25, but as you can see the notation $\frac{1}{4}$ is more compact and nicer. Why nicer? Check out these simple fractions:



$$\begin{aligned}
\frac{1}{1} &= 1.0 \\
\frac{1}{2} &= 0.5 \\
\frac{1}{3} &= 0.33333\ldots = 0.\overline{3} \\
\frac{1}{4} &= 0.25 \\
\frac{1}{5} &= 0.2 \\
\frac{1}{6} &= 0.166666\ldots = 0.1\overline{6} \\
\frac{1}{7} &= 0.142857142857142857\ldots = 0.\overline{142857}
\end{aligned}$$

Note that a line above some numbers means the digits underneath the line are repeated. The fractional notation on the left is preferable, because it shows the underlying *structure* of the number while avoiding the need to write infinitely long decimals.

fraction by $\frac{b}{b} = 1$ in order to make the denominator of both fractions the same:

$$\frac{a}{b} + \frac{c}{d} = \frac{a}{b}\left(\frac{d}{d}\right) + \left(\frac{b}{b}\right)\frac{c}{d} = \frac{ad}{bd} + \frac{bc}{bd}.$$

Now that we have fractions with the same denominator, we can add the numerators. The effect of multiplying the top and bottom of the fractions by the same number is the same as multiplying by 1 since the above operations did not change the fractions:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{bc}{bd} = \frac{ad + bc}{bd}.$$

More generally, to add two fractions, we can pick the *least common multiple* $\mathrm{LCM}(b, d)$ to use as the common denominator. The LCM of two numbers is obtained by multiplying the numbers together and removing their common factors:

$$\mathrm{LCM}(b, d) = \underline{\frac{b \times d}{\mathrm{GCD}(b, d)}}\,\frac{bd}{\mathrm{GCD}(b, d)},$$

where $\mathrm{GCD}(b, d)$ is the *greatest common divisor* of b and d—the largest number that divides both b and d.

For example, to add $\frac{1}{6}$ and $\frac{1}{15}$, we rewrite the fractions so they have the common denominator 6×15 (the product of the two denominators). Or, more simply, we can see $6 = 3 \times 2$, and $15 = 3 \times 5$, meaning 3 is a common divisor of both 6 and 15. The least common multiple is then $\frac{6 \times 15}{3} = 30$, so we write:

$$\frac{1}{6} + \frac{1}{15} = \frac{5 \times 1}{5 \times 6} + \frac{1 \times 2}{15 \times 2} = \frac{5}{30} + \frac{2}{30} = \frac{7}{30}.$$

Actually, all this LCM and GCD business is not *required*—but it is the most efficient way to add fractions without having to deal with excessively large numbers. If you use the common denominator ~~$b \times d$~~bd, you will arrive at the same answer as above after simplification:

$$\frac{1}{6} + \frac{1}{15} = \frac{15 \times 1}{15 \times 6} + \frac{1 \times 6}{15 \times 6} = \frac{15}{90} + \frac{6}{90} = \frac{21}{90} = \frac{7}{30}.$$

Multiplication of fractions

Fraction multiplication involves multiplying the numerators together and multiplying the denominators together:

$$\frac{a}{b} \times \frac{c}{d} = \frac{a \times c}{b \times d} = \frac{ac}{bd}.$$

Division of fractions

To divide two fractions, compute the product of the first fraction times the second fraction *flipped*:

$$\frac{a/b}{c/d} = \frac{a}{b} \div \frac{c}{d} = \frac{a}{b} \times \frac{d}{c} = \frac{a \times d}{b \times c} = \frac{ad}{bc}\,.$$

The *multiplicative inverse* of something times that something should give 1 as the answer. We obtain the multiplicative inverse of a fraction by interchanging the roles of the numerator and the denominator:

$$\left(\frac{c}{d}\right)^{-1} = \frac{d}{c}\,.$$

Any fraction times its multiplicative inverse gives

$$\frac{c}{d} \times \left(\frac{c}{d}\right)^{-1} = \frac{c}{d} \times \frac{d}{c} = \frac{cd}{cd} = 1.$$

The "flip and multiply" rule for division stems from the fact that division by a number x is the same as multiplication by $\frac{1}{x}$.

Whole and fraction notation

To indicate a fraction like $\frac{5}{3}$, which is greater than 1, we sometimes use the notation $1\frac{2}{3}$, which is read as "one and two thirds." Similarly, $\frac{22}{7} = 3\frac{1}{7}$.

There is nothing wrong with writing fractions like $\frac{5}{3}$ and $\frac{22}{7}$. However, some teachers say this way of writing fractions is *improper* and demand that fractions are written in the whole-and-fraction way, as in $1\frac{2}{3}$ and $3\frac{1}{7}$. At the end of the day, both notations are correct.

Repeating decimals

When written as decimal numbers, certain fractions have infinitely long decimal expansions. We use the overline notation to indicate the digit(s) that repeat in the expansion:

$$\frac{1}{3} = 0.\overline{3} = 0.333\ldots; \qquad \frac{1}{7} = 0.\overline{142857} = 0.14285714285714\ldots.$$

Exercsies

E1.6 Compute the value of the following expressions:

(a) $\frac{1}{2}+\frac{1}{3}$ (b) $\frac{1}{2}+\frac{1}{3}+\frac{1}{4}$ (c) $3\frac{1}{2}+2-\frac{1}{3}$

1.6 (a) $\frac{5}{6}$; (b) $\frac{13}{12}=1\frac{1}{12}$; (c) $\frac{31}{6}=5\frac{1}{6}$.

1.6 (a) To compute $\frac{1}{2}+\frac{1}{3}$, we rewrite both fractions using the common denominator 6, then compute the sum: $\frac{1}{2}+\frac{1}{3}=\frac{3}{6}+\frac{2}{6}=\frac{5}{6}$. (b) You can use the answer from part (a), or compute the triple sum directly by setting all three fractions to a common denominator: $\frac{1}{2}+\frac{1}{3}+\frac{1}{4}=\frac{6}{12}+\frac{4}{12}+\frac{3}{12}=\frac{13}{12}$. (c) Here we first rewrite $3\frac{1}{2}$ as $\frac{7}{2}$, then use the common denominator 6 for the computation: $\frac{7}{2}+2-\frac{1}{3}=\frac{21}{6}+\frac{12}{6}-\frac{2}{6}=\frac{31}{6}$.

1.6 Basic rules of algebra

It's important that you know the general rules for manipulating numbers and variables, a process otherwise known as—you guessed it—*algebra*. This little refresher will cover these concepts to make sure you're comfortable on the algebra front. We'll also review some important algebraic tricks, like *factoring* and *completing the square*, which are useful when solving equations.

When an expression contains multiple things added together, we call those things *terms*. Furthermore, terms are usually composed of many things multiplied together. When a number x is obtained as the product of other numbers like $x = abc$, we say "x factors into a, b, and c." We call a, b, and c the *factors* of x.

Given any four numbers a, b, c, and d, we can apply the following algebraic properties:

1. Associative property: $a+b+c=(a+b)+c=a+(b+c)$ and $abc=(ab)c=a(bc)$
2. Commutative property: $a+b=b+a$ and $ab=ba$
3. Distributive property: $a(b+c)=ab+ac$

We use the distributive property every time we *expand* brackets. For example $a(b+c+d) = ab+ac+ad$. The brackets, also known as parentheses, indicate the expression $(b+c+d)$ must be treated as a whole: a factor that consists of three terms. Multiplying this expression by a is the same as multiplying each term by a.

The opposite operation of expanding is called *factoring*, which consists of rewriting the expression with the common parts taken out in front of a bracket: $ab+ac=a(b+c)$. In this section, we'll discuss both of these operations and illustrate what they're capable of.

Expanding brackets

The distributive property is useful when dealing with polynomials:

$$(x+3)(x+2) = x(x+2) + 3(x+2) = x^2 + x2 + 3x + 6.$$

We can use the commutative property on the second term $x2 = 2x$, then combine the two x terms into a single term to obtain

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

Let's look at this operation in its abstract form:

$$(x+a)(x+b) = x^2 + (a+b)x + ab.$$

The product of two ~~linear terms (expressions~~ factors of the form ~~$x + ?)$~~ $(x+?)$ is equal to a quadratic expression. Observe that the middle term on the right-hand side contains the *sum* of the two constants on the left-hand side $(a+b)$, while the third term contains their product ab.

It is very common for people to confuse these terms. If you are ever confused about an algebraic expression, go back to the distributive property and expand the expression using a step-by-step approach. As a second example, consider this slightly-more-complicated algebraic expression and its expansion:

$$\begin{aligned}(x+a)(bx^2+cx+d) &= x(bx^2+cx+d) + a(bx^2+cx+d)\\ &= bx^3 + cx^2 + dx + abx^2 + acx + ad\\ &= bx^3 + (c+ab)x^2 + (d+ac)x + ad.\end{aligned}$$

Note how all terms containing x^2 are grouped into a one term, and all terms containing x are grouped into another term. We use this pattern when dealing with expressions containing different powers of x.

Example Suppose we are asked to solve for t in the equation

$$7(3+4t) = 11(6t-4).$$

Since the unknown t appears on both sides of the equation, it is not immediately obvious how to proceed.

To solve for t, we must bring all t terms to one side and all constant terms to the other side. First, expand the two brackets to obtain

$$21 + 28t = 66t - 44.$$

Then move things around to relocate all ts to the equation's right-hand side and all constants to the left-hand side:

$$21 + 44 = 66t - 28t.$$

We see t is contained in both terms on the right-hand side, so we can rewrite the equation as

$$21 + 44 = (66 - 28)t.$$

The answer is within close reach: $t = \frac{21+44}{66-28} = \frac{65}{38}$.

Factoring

Factoring involves taking out the common part(s) of a complicated expression in order to make the expression more compact. Suppose you're given the expression $6x^2y + 15x$ and must simplify it by taking out common factors. The expression has two terms and each term can be split into its constituent factors to obtain

$$6x^2y + 15x = (3)(2)(x)(x)y + (5)(3)x.$$

Since factors x and 3 appear in both terms, we can *factor them out* to the front like this:

$$6x^2y + 15x = 3x(2xy + 5).$$

The expression on the right shows $3x$ is common to both terms.

Here's another example where factoring is used:

$$2x^2y + 2x + 4x = 2x(xy + 1 + 2) = 2x(xy + 3).$$

Quadratic factoring

When dealing with a quadratic function, it is often useful to rewrite the function as a product of two factors. Suppose you're given the quadratic function $f(x) = x^2 - 5x + 6$ and asked to describe its properties. What are the *roots* of this function? In other words, for what values of x is this function equal to zero? For which values of x is the function positive, and for which ~~x values~~ values of x is the function negative?

Factoring the expression $x^2 - 5x + 6$ will help us see the properties of the function more clearly. To *factor* a quadratic expression is to express it as the product of two factors:

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

We now see at a glance the solutions (roots) are $x_1 = 2$ and $x_2 = 3$. We can also see for which x values the function will be overall positive: for $x > 3$, both factors will be positive, and for $x < 2$ both factors will be negative, and a negative times a negative gives a positive. For values of x such that $2 < x < 3$, the first factor will be positive, and the second factor negative, making the overall function negative.

For certain simple quadratics like the one above, you can simply *guess* what the factors will be. For more complicated quadratic expressions, you'll need to use the quadratic formula (page 27), which will be the subject of the next section. For now let us continue with more algebra tricks.

Completing the square

Any quadratic expression $Ax^2 + Bx + C$ can be rewritten in the form $A(x - h)^2 + k$ for some constants h and k. This process is called *completing the square* due to the reasoning we follow to find the value of k. The constants h and k can be interpreted geometrically as the horizontal and vertical shifts in the graph of the basic quadratic function. The graph of the function $f(x) = A(x - h)^2 + k$ is the same as the graph of the function $f(x) = Ax^2$ except it is shifted h units to the right and k units upward. We will discuss the geometrical meaning of h and k in more detail in Section 1.14 (page 64). For now, let's focus on the algebra steps.

Let's try to find the values of k and h in the expression $(x - h)^2 + k$ needed to complete the square in the expression $x^2 + 5x + 6$. ~~We start from the assumption that the~~ Assume the two expressions are equal, and then expand the bracket to obtain

$$\underline{x^2}+5x+6 = A(x-h)^2+k = A(x^2-2hx+h^2)+k = \underline{Ax^2}-2Ahx+Ah^2+k.$$

Observe the structure in the above equation. On both sides of the equality there is one term which contains x^2 (the quadratic term), one term that contains x^1 (the linear term), and ~~some~~ constant terms. If the expressions are equal, then the coefficient of all the term must be equal.

By focusing on the quadratic terms on both sides of the equation (they are underlined) we see $A = 1$, so we ~~can~~ rewrite the equation as

$$x^2 + \underline{5x} + 6 = x^2 \underline{-2hx} + h^2 + k.$$

Next we look at the linear terms (underlined) and infer $h = -2.5$. After rewriting, we obtain an equation in which k is the only unknown:

$$x^2 + 5x + \underline{6} = x^2 - 2(-2.5)x + \underline{(-2.5)^2 + k}.$$

We must pick a value of k that makes the constant terms equal:

$$k = 6-(-2.5)^2 = 6-(2.5)^2 = 6-\left(\frac{5}{2}\tfrac{5}{2}\right)^2 = 6\times\frac{4}{4}\tfrac{4}{4}-\frac{25}{4}\tfrac{25}{4} = \frac{24-25}{4}\tfrac{24-25}{4} = \frac{-1}{4}$$

After completing the square we obtain

$$x^2 + 5x + 6 = (x + 2.5)^2 - \frac{1}{4}\tfrac{1}{4}.$$

The right-hand side of the expression above tells us our function is equivalent to the basic function x^2, shifted 2.5 units to the left and $\frac{1}{4}$ units down. This would be very useful information if you ever had to draw the graph of this function—you could simply plot the basic graph of x^2 and then shift it appropriately.

It is important you become comfortable with this procedure for completing the square. It is not extra difficult, but it does require you to think carefully about the unknowns h and k and to choose their values appropriately. There is no general formula for finding k, but you can remember the following simple shortcut for finding h. Given an equation $Ax^2 + Bx + C = A(x - h)^2 + k$, we have $h = \frac{-B}{2A}$. Using this shortcut will save you some time, but you will still have to go through the algebra steps to find k.

Take out a pen and a piece of paper now (yes, right now!) and verify that you can correctly complete the square in these expressions: $x^2 - 6x + 13 = (x - 3)^2 + 4$ and $x^2 + 4x + 1 = (x + 2)^2 - 3$.

Exercises

E1.7 Factor the following expressions:

a) $x^2 - 8x + 7$ b) $x^2 + 4x + 4$

1.7 a) $(x - 1)(x - 7)$; b) $(x + 2)^2$.

E1.8 Expand the following expressions:

a) $(a+b)^2$ b) $(a+b)^3$ c) $(a+b)^4$ d) $(a+b)^5$

Can you spot a pattern in the coefficients of the different expressions? Do you think there is a general equation for $(a+b)^n$?

1.8 a) $a^2+2ab+b^2$; b) $a^3+3a^2b+3ab^2+b^3$; c) $a^4+4a^3b+6a^2b^2+4ab^3+b^4$; d) $a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5$.

1.8 The coefficients of the expression of $(a+b)^n$ for different values of n correspond to the rows in *Pascal's triangle*. Check out the Wikipedia page for Pascal's triangle to learn the general formula and see an interesting animation of how it can be constructed.

1.7 Solving quadratic equations

What would you do if asked to solve for x in the quadratic equation $x^2 = 45x + 23$? This is called a *quadratic equation* since it contains the unknown variable x squared. The name comes from the Latin *quadratus*, which means square. Quadratic equations appear often, so mathematicians created a general formula for solving them. In this section, we'll learn about this formula and use it to put some quadratic equations in their place.

Before we can apply the formula, we need to rewrite the equation we are trying to solve in the following form:

$$ax^2 + bx + c = 0.$$

We reach this form—called the *standard form* of the quadratic equation—by moving all the numbers and xs to one side and leaving only 0 on the other side. For example, to transform the quadratic equation $x^2 = 45x + 23$ into standard form, subtract $45x + 23$ from both sides of the equation to obtain $x^2 - 45x - 23 = 0$. What are the values of x that satisfy this formula?

Claim

The solutions to the equation $ax^2 + bx + c = 0$ are

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad \text{and} \qquad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Let's see how these formulas are used to solve $x^2 - 45x - 23 = 0$. Finding the two solutions requires the simple mechanical task of identifying $a = 1$, $b = -45$, and $c = -23$ and plugging these values into the formulas:

$$x_1 = \frac{45 + \sqrt{45^2 - 4(1)(-23)}}{2} = 45.5054\ldots,$$

$$x_2 = \frac{45 - \sqrt{45^2 - 4(1)(-23)}}{2} = -0.5054\ldots.$$

Verify using your calculator that both of the values above satisfy the original equation $x^2 = 45x + 23$.

Proof of claim

This is an important proof. I want you to see how we can *derive* the quadratic formula from first principles because this knowledge will help you understand the formula. The proof will use the completing-the-square technique from the previous section.

Starting ~~with~~ from the quadratic equation $ax^2 + bx + c = 0$, ~~first move c to the other side of the equation :~~ we divide by a to obtain the equation

$$\cancel{ax}x^2 + \cancel{bx}\frac{b}{a}x + \frac{c}{a} = \cancel{-c.}0.$$

~~Divide by a on both sides:~~

$$\cancel{x^2 + \frac{b}{a}x = -\frac{c}{a}.}$$

~~Now~~ Next we'll *complete the square* ~~on the left-hand side~~ by asking, "What are the values of h and k that satisfy the equation

$$(x-h)^2 + k \;=\; x^2 + \frac{b}{a}x + \frac{c}{a} \quad ?"$$

To find the values for h and k, ~~we'll expand the~~ let's expand the bracket on the left-hand side to obtain ~~$(x-h)^2 + k = x^2 - 2hx + h^2 + k$. We can now~~

$$x^2 - 2hx + h^2 + k \;=\; x^2 + \frac{b}{a}x + \frac{c}{a}.$$

We can identify h by looking at the coefficients in front of x on both sides of the equation. We have $-2h = \frac{b}{a}$ and hence $h = -\frac{b}{2a}$.

Let's now substitute the value $h = -\frac{b}{2a}$ into the above equation and see what we have so far:

$$\cancel{\left(x + \frac{b}{2a}\right)^2 = \left(x + \frac{b}{2a}\right)\left(x + \frac{b}{2a}\right) = x^2 + \frac{b}{2a}x + x\frac{b}{2a} + \frac{b^2}{4a^2} = x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}.}$$

~~To determine~~

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} + k \;=\; x^2 + \frac{b}{a}x + \frac{c}{a}.$$

To determine the value of k, we need to ~~move that last term to the other side:~~ ensure the constant terms on both sides of the equation are equal then isolate k: ~~We can continue with the proof where we left off:~~

$$x^2 + \frac{b}{a}x = -\frac{c}{a}.$$

~~Replace the left-hand side with the complete-the-square expression and obtain~~ Having found the values of both h and k we can write the equation $ax^2 + bx + c = 0$ in the form $(x - h) + k = 0$ as follows:

$$\left(x + \frac{b}{2a}\right)^2 + \frac{c}{a} - \frac{b^2}{4a^2} = -\frac{c}{a}.0.$$

From here on, we can use the standard procedure for ~~. Arrange all constants on~~ "digging" towards the x which we saw in Section 1.1. Move all constants to the right-hand side:

$$\left(x + \frac{b}{2a}\right)^2 = -\frac{c}{a} + \frac{b^2}{4a^2}.$$

Next, take the square root of both sides to undo the square function. Since the square function maps both positive and negative numbers to the same value, this step yields two solutions:

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

Let's take a moment to tidy up the mess under the square root:

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

We obtain

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

which is just one step from the final answer,

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

This completes the proof. □

and x_2 are called the *roots* of the function: these points are where the function $f(x)$ touches the x-axis.

You now have the ability to factor any quadratic equation. Use the quadratic formula to find the two solutions, x_1 and x_2, then rewrite the expression as $(x - x_1)(x - x_2)$.

Some quadratic expressions cannot be factored, however. These "unfactorable" expressions correspond to quadratic functions whose graphs do not touch the x-axis. They have no solutions (no roots). There is a quick test you can use to check if a quadratic function $f(x) = ax^2 + bx + c$ has roots (touches or crosses the x-axis) or doesn't have roots (never touches the x-axis). If $b^2 - 4ac > 0$ then the function f has two roots. If $b^2 - 4ac = 0$, the function has only one root, indicating the special case when the function touches the x-axis at only one point. If $b^2 - 4ac < 0$, the function has no roots. In this case the quadratic formula fails because it requires taking the square root of a negative number, which is not allowed. Think about it—how could you square a number and obtain a negative number?

Exercises

E1.9 Solve for x in the quadratic equation $2x^2 - x = 3$.

1.9 $x_1 = \frac{3}{2}$ and $x_2 = -1$.

E1.10 Solve for x in the equation $x^4 - 4x^2 + 4 = 0$.
Hint: Use the substitution $y = x^2$.

1.10 $x = \pm\sqrt{2}$.

1.8 Exponents

In math we must often multiply together the same number many times, so we use the notation

$$b^n = \underbrace{bbb\cdots bb}_{n \text{ times}}$$

to denote some number b multiplied by itself n times. In this section we'll review the basic terminology associated with exponents and discuss their properties.

Definitions

The fundamental ideas of exponents are:

- b^n: the number b raised to the power n

Discussion

Even and odd exponents

The function $f(x) = x^n$ behaves differently depending on whether the exponent n is even or odd. If n is odd we have

$$\left(\sqrt[n]{b}\right)^n = \sqrt[n]{b^n} = b, \qquad \text{when } n \text{ is odd}.$$

However, if n is even, the function x^n destroys the sign of the number (see x^2, which maps both $-x$ and x to x^2). The successive application of exponentiation by n and the n^{th} root has the same effect as the absolute value function:

$$\sqrt[n]{b^n} = |b|, \qquad \text{when } n \text{ is even}.$$

Recall that the absolute value function $|x|$ discards the information about the sign of x. The expression $(\sqrt[n]{b})^n$ cannot be computed whenever b is a negative number. The reason is that we can't evaluate $\sqrt[n]{b}$ for $b < 0$ in terms of real numbers, since there is no real number which, multiplied ~~times~~ by itself an even number of times, gives a negative number.

Scientific notation

In science we often work with very large numbers like *the speed of light* ($c = 299\,792\,458$[m/s]), and very small numbers like *the permeability of free space* ($\mu_0 = 0.000001256637\ldots$[N/A^2]). It can be difficult to judge the magnitude of such numbers and to carry out calculations on them using the usual decimal notation.

Dealing with such numbers is much easier if we use *scientific notation*. For example, the speed of light can be written as $c = 2.99792458 \times 10^8$[m/s], and the permeability of free space is denoted as $\mu_0 = 1.256637 \times 10^{-6}$[N/A^2]. In both cases, we express the number as a decimal number between 1.0 and 9.9999... followed by the number 10 raised to some power. The effect of multiplying by 10^8 is to move the decimal point eight steps to the right, making the number bigger. Multiplying by 10^{-6} has the opposite effect, moving the decimal to the left by six steps and making the number smaller. Scientific notation is useful because it allows us to clearly see the *size* of numbers: 1.23×10^6 is $1\,230\,000$ whereas 1.23×10^{-10} is $0.000\,000\,000\,123$. With scientific notation you don't have to count the zeros! ~~Cool, yeah?~~

The number of decimal places we use when specifying a certain physical quantity is usually an indicator of the *precision* with which we are able to measure this quantity. Taking into account the precision of the measurements we make is an important aspect of all quantitative research. Since elaborating further would be a digression, we will not go into a full discussion about the topic of *significant digits* here. Feel free to check out the Wikipedia article on the subject ~~if you want to know~~ to learn more.

On computer systems, *floating point numbers* are represented in scientific notation: they have a decimal part and an exponent. To separate the decimal part from the exponent when entering a floating point number into the computer, use the character `e`, which stands for "exponent." The base is assumed to be 10. For example, the speed of light is written as `2.99792458e8` and the permeability of free space is `1.256637e-6`.

Links

[Further reading on exponentiation]
`http://en.wikipedia.org/wiki/Exponentiation`

[More details on scientific notation]
`http://en.wikipedia.org/wiki/Scientific_notation`

Exercises

E1.11 Simplify the following exponential expressions.

a) $2^3 ef \frac{\sqrt{ef}}{(\sqrt{ef})^3}$ b) $\frac{abc}{a^2b^3c^4}$ c) $\frac{(2\alpha)^3}{\alpha}$ d) $(a^3)^2(\frac{1}{b})^2$

1.11 a) 8; b) $a^{-1}b^{-2}c^{-3} = \frac{1}{ab^2c^3}$; c) $8\alpha^2$; d) a^6b^{-2}.

E1.12 Find all the values of x that satisfy these equations:

a) $x^2 = a$ b) $x^3 = b$ c) $x^4 = c$ d) $x^5 = d$

1.12 a) $x = \sqrt{a}$ and $x = -\sqrt{a}$; b) $x = \sqrt[3]{b}$; c) $x = \sqrt[4]{c}$ and $x = -\sqrt[4]{c}$; d) $x = \sqrt[5]{d}$. Bonus points if you can also solve $x^2 = -1$. We'll get to that in Section 3.5.

E1.13 Coulomb's constant k_e is defined by the formula $k_e = \frac{1}{4\pi\varepsilon_0}$, where ε_0 is the permittivity of free space. Use a calculator to compute the value of k_e starting from $\varepsilon_0 = 8.854 \times 10^{-12}$ and $\pi = 3.14159265$. Report your answer with an appropriate number of digits, even if the calculator gives you a number with more digits.

1.13 $k_e = 8.988 \times 10^9$.

1.13 If you're using a very basic calculator, you should first compute the expression in the denominator then invert it. Calculators that support scientific notation have an "`exp`" or "`E`" button, which allows you to enter ε_0 as `8.854e-12`. If your calculator supports expressions, you can type in the whole expression `1/(4*pi*8.854e-12)`. We report an answer with four significant digits because we started from a value of ε_0 with four significant digits of precision.

1.9 Logarithms

Some people think the word "logarithm" refers to some mythical, mathematical beast. Legend has it that logarithms are many-headed, breathe fire, and are extremely difficult to understand. Nonsense! Logarithms are simple. It will take you at most a couple of pages to get used to manipulating them, and that is a good thing because logarithms are used all over the place.

The strength of your sound system is measured in logarithmic units called decibels [dB]. This is because your ears are sensitive only to exponential differences in sound intensity. Logarithms allow us to compare very large numbers and very small numbers on the same scale. If sound were measured in linear units instead of logarithmic units, your sound system's volume control would need to range from 1 to 1 048 576. That would be weird, no? This is why we use the logarithmic scale for volume notches. Using a logarithmic scale, we can go from sound intensity level 1 to sound intensity level 1 048 576 in 20 "progressive" steps. Assume each notch doubles the sound intensity, rather than increasing the intensity by a fixed amount. If the first notch corresponds to 2, the second notch is 4—still probably inaudible, turn it up! By the time you get to the sixth notch you're at $2^6 = 64$ sound intensity, which is the level of audible music. The tenth notch corresponds to sound intensity $2^{10} = 1024$ (medium-strength sound), and finally the twentieth notch reaches a max power of $2^{20} = 1\,048\,576$, at which point the neighbours come ~~knocking~~ to complain.

Definitions

You're hopefully familiar with these following concepts from the previous section:

- b^x: the exponential function base b
- $\exp(x) = e^x$: the exponential function base e, Euler's number
- 2^x: exponential function base 2
- $f(x)$: the notion of a function $f : \mathbb{R} \to \mathbb{R}$

Taking the logarithm of both sides gives us

$$\log_b(xy) = \log_b\left(b^{m+n}\right) = m + n = \log_b(x) + \log_b(y).$$

The last step above uses the definition of the log function again, which states that

$$b^m = x \quad \Leftrightarrow \quad m = \log_b(x) \qquad \text{and} \qquad b^n = y \quad \Leftrightarrow \quad n = \log_b(y).$$

Property 2

This property helps us change from one base to another.

We can express the logarithm in any base B in terms of a ratio of logarithms in another base b. The general formula is

$$\log_B(x) = \frac{\log_b(x)}{\log_b(B)}.$$

For example, the logarithm base 10 of a number S can be expressed as a logarithm base 2 or base e as follows:

$$\log_{10}(S) = \frac{\log_{10}(S)}{1} = \frac{\log_{10}(S)}{\log_{10}(10)} = \frac{\log_2(S)}{\log_2(10)} = \frac{\ln(S)}{\ln(10)}.$$

This property will help if you ever need to compute a logarithm in a base that is not available on your calculator. Suppose you are asked to compute $\log_7(S)$, but your calculator only has a $\boxed{\log_{10}}$ button. You can simulate $\log_7(S)$ by computing $\log_{10}(S)$ and dividing by $\log_{10}(7)$.

Exercises

E1.14 Use the properties of logarithms to simplify the expressions

(1) $\log(x) + \log(2y)$ (2) $\log(z) - \log(2z)$ (3) $\log(x) + \log(y/x)$

(4) $\log_2(8)$ (5) $\log_3(\frac{1}{27})$ (6) $\log_{10}(10000)$

1.14 (1) $\log(2xy)$. (2) $-\log(z)$. (3) $\log(y)$. (4) 3. (5) -3. (6) 4.

1.10 The number line

The number line is a useful graphical representation for numbers. The integers $\mathbb{Z}$ correspond to the notches on the line while the rationals $\mathbb{Q}$ and the reals $\mathbb{R}$ densely cover the whole line.

Exercises

E1.15 Find the values of x that satisfy the following inequalities.

(1) $2x < 3$ (2) $-4x \geq 20$ (3) $|2x - 3| < 5$
(4) $3x + 3 < 5x - 5$ (5) $\frac{1}{2}x - 2 \geq \frac{1}{3}$
(6) $(x + 1)^2 \geq 9$

Express your answer as an interval with appropriate endpoints.

1.15 (1) $x \in (-\infty, \frac{3}{2})$; (2) $x \in (-\infty, -5]$; (3) $x \in (-1, 4)$; (4) $x \in (4, \infty)$; (5) $x \in [\frac{14}{3}, \infty)$; (6) $(-\infty, -4] \cup [2, \infty)$.

1.15 (1) Dividing both sides of the inequality by two gives $x < \frac{3}{2}$. (2) Divide both sides by negative four to obtain $x \leq -5$. Note the "$\geq$" changed to "$\leq$" since we divided by a negative number. (3) If the absolute value of $(2x - 3)$ is less than five, then $(2x - 3)$ must lie in the interval $(-5, 5)$. We can therefore rewrite the inequality as $-5 < 2x - 3 < 5$, then add three to both sides to obtain $-2 < 2x < 8$, and dividing by two gives us the final answer $-1 < x < 4$. (4) Let's collect all the x-terms on the right an all the constants on the left side: $8 < 2x$, which leads to $4 < x$. (5) To simplify, add two to both sides of the inequality to obtain $\frac{1}{2}x \geq \frac{1}{3} + 2$. You remember how to add fractions right? We have $\frac{1}{3} + 2 = \frac{1}{3} + \frac{6}{3} = \frac{7}{3}$, therefore $\frac{1}{2}x \geq \frac{7}{3}$. Multiply both sides by two to obtain $x \geq \frac{14}{3}$. (6) The first step is to get rid of the square, by taking the square root operation on both sides: $\sqrt{(x + 1)^2} \geq \sqrt{9}$. Recall that $\sqrt{x^2} = |x|$, so we have $|x + 1| \geq 3$. There are two ways for the absolute value of $(x + 1)$ to be greater than three. Either $x + 1 \geq 3$ or $x + 1 \leq -3$. We subtract one in each of these inequalities to find $x \geq 2$ or $x \leq -4$. The solution to this inequality is the union of these two intervals.

1.12 The Cartesian plane

Named after famous philosopher and mathematician René Descartes, the Cartesian plane is a graphical representation for *pairs* of numbers.

Generally, we call the plane's horizontal axis "the x-axis" and its vertical axis "the y-axis." We put notches at regular intervals on each axis so we can measure distances. Figure 1.5 is an example of an empty Cartesian coordinate system. Think of the coordinate system as an empty canvas. What can you draw on this canvas?

Vectors and points

A *point* $P = (P_x, P_y)$ in the Cartesian plane has an x-coordinate and a y-coordinate. To find this point, start from the origin—the point hosting the concert is $C = \$2000$, then the profit from the concert P can be modelled as

$$\begin{aligned} P(n) &= R(n) \ - \ C \\ &= 0.95 \cdot \$25 \cdot n \ - \ \$2000 \end{aligned}$$

The function $P(n) = 23.75n - 2000$ models the profit from the concert as a function of the number of tickets sold. This is a pretty good model already, and you can always update it later on as you find out more information.

The more functions you know, the more tools you have for modelling reality. To "know" a function, you must be able to understand and connect several of its aspects. First you need to know the function's mathematical **definition**, which describes exactly what the function does. Starting from the function's definition, you can use your existing math skills to find the function's domain, its ~~range~~image, and its inverse function. You must also know the **graph** of the function; what the function looks like if you plot x versus $f(x)$ in the Cartesian plane (page 43). It's also a good idea to remember the **values** of the function for some important inputs. Finally—and this is the part that takes time—you must learn about the function's **relations** to other functions.

Definitions

A *function* is a mathematical object that takes numbers as inputs and gives numbers as outputs. We use the notation

$$f\colon A \to B$$

to denote a function from the input set A to the output set B. In this book, we mostly study functions that take real numbers as inputs and give real numbers as outputs: $f\colon \mathbb{R} \to \mathbb{R}$.

We now define some fancy technical terms used to describe the input and output sets.

- The *domain* of a function is the set of allowed input values.
- The *image* or *range* of the function f is the set of all possible output values of the function.
- The *codomain* of a function describes the type of outputs the function has.

To illustrate the subtle difference between the image of a function and its codomain, consider the function $f(x) = x^2$. The quadratic function is of the form $f\colon \mathbb{R} \to \mathbb{R}$. The function's domain is $\mathbb{R}$ (it takes real numbers as inputs) and its codomain is $\mathbb{R}$ (the outputs are real numbers too), however, not all outputs are possible. The *image* of the function $f(x) = x^2$ consists only of the nonnegative real numbers $\cancel{[0, \infty \equiv \{y \in \mathbb{R} \mid y \geq 0\}.}$ $[0, \infty) \equiv \{y \in \mathbb{R} \mid y \geq 0\}.$ A function is not a number; rather, it is a *mapping* from numbers to numbers. For any input x, the output value of f for that input is denoted $f(x)$.

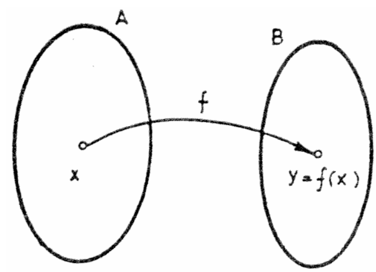


Figure 1.8: An abstract representation of a function f from the set A to the set B. The function f is the arrow which *maps* each input x in A to an output $f(x)$ in B. The output of the function $f(x)$ is also denoted y.

We say "f maps x to $f(x)$," and use the following terminology to classify the type of mapping that a function performs:

- A function is *one-to-one* or *injective* if it maps different inputs to different outputs.
- A function is *onto* or *surjective* if it covers the entire output set (in other words, if the image of the function is equal to the function's codomain).
- A function is *bijective* if it is both injective and surjective. In this case, f is a *one-to-one correspondence* between the input set and the output set: for each of the possible outputs $y \in Y$ (surjective part), there exists exactly one input $x \in X$, such that $f(x) = y$ (injective part).

The term *injective* is an allusion from the 1940s inviting us to picture the actions of injective functions as pipes through which numbers flow like fluids. Since a fluid cannot be compressed, the output space must be at least as large as the input space. A modern synonym for injective functions is to say they are *two-to-two*. If we imagine two specks of paint floating around in the "input fluid," an injective function will contain two distinct specks of paint in the "output fluid." In contrast, non-injective functions can map several different inputs to the same output. For example $f(x) = x^2$ is not injective since the inputs 2 and -2 are both mapped to the output value 4.

understand all of physics and calculus and handle *any* problem your teacher may throw at you.

~~Links~~

~~Tank game where you specify the function of the projectile~~

1.14 Function reference

Your *function vocabulary* determines how well you can express yourself mathematically in the same way that your English vocabulary determines how well you can express yourself in English. The following pages aim to embiggen your function vocabulary so you won't be caught with your pants down when the teacher tries to pull some trick on you at the final.

If you are seeing these functions for the first time, don't worry about remembering all the facts and properties on the first reading. We will use these functions throughout the rest of the book so you will have plenty of time to become familiar with them. Just remember to come back to this section if you ever get stuck on a function.

Line

The equation of a line describes an input-output relationship where the change in the output is *proportional* to the change in the input. The equation of a line is

$$f(x) = mx + b.$$

The constant m describes the slope of the line. The constant b is called the y-intercept and it corresponds to the value of the function when $x = 0$.

The equation of the line $f(x) = mx + b$ is so important that it's worth taking the time to contemplate it for a few seconds. I'll leave some blank space here to give you "pages-turned" credit for taking the time.

Square

The function x *squared*, is also called the *quadratic* function, or *parabola*. The formula for the quadratic function is

$$f(x) = x^2.$$

The name "quadratic" comes from the Latin *quadratus* for square, since the expression for the area of a square with side length x is x^2.

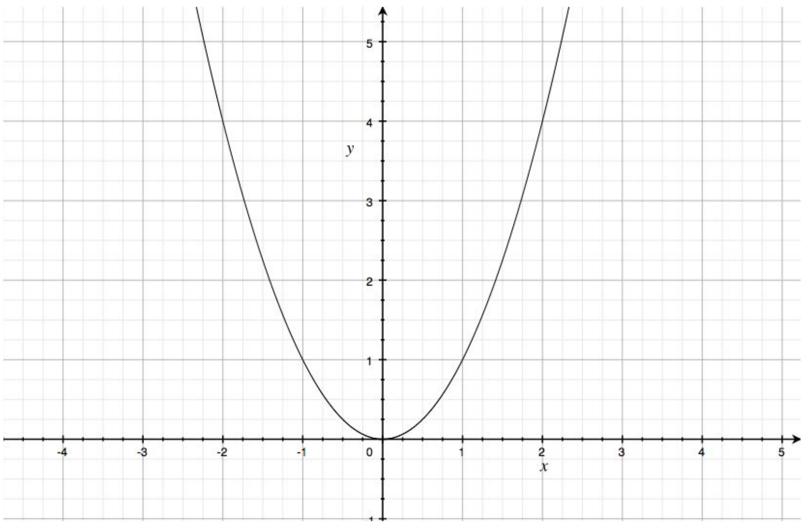


Figure 1.11: Plot of the quadratic function $f(x) = x^2$. The graph of the function passes through the following (x, y) coordinates: $(-2, 4)$, $(-1, 1)$, $(0, 0)$, $(1, 1)$, $(2, 4)$, $(3, 9)$, etc.

Properties

- Domain: $x \in \mathbb{R}$.
  The function $f(x) = x^2$ is defined for all input values $x \in \mathbb{R}$.
- Image: ~~$f(x) \in [0, \infty$~~ $f(x) \in [0, \infty)$.
  The outputs are never negative: $x^2 \geq 0$, for all $x \in \mathbb{R}$.
- The function x^2 is the inverse of the square root function $\sqrt{x}$.
- $f(x) = x^2$ is *two-to-one*: it sends both x and $-x$ to the same output value $x^2 = (-x)^2$.
- The quadratic function is *convex*, meaning it curves upward.

Square root

The square root function is denoted

$$f(x) = \sqrt{x} \equiv x^{\frac{1}{2}}.$$

The square root $\sqrt{x}$ is the inverse function of the square function x^2 for $x \geq 0$. The symbol $\sqrt{c}$ refers to the *positive* solution of $x^2 = c$. Note that $-\sqrt{c}$ is also a solution of $x^2 = c$.

Graph

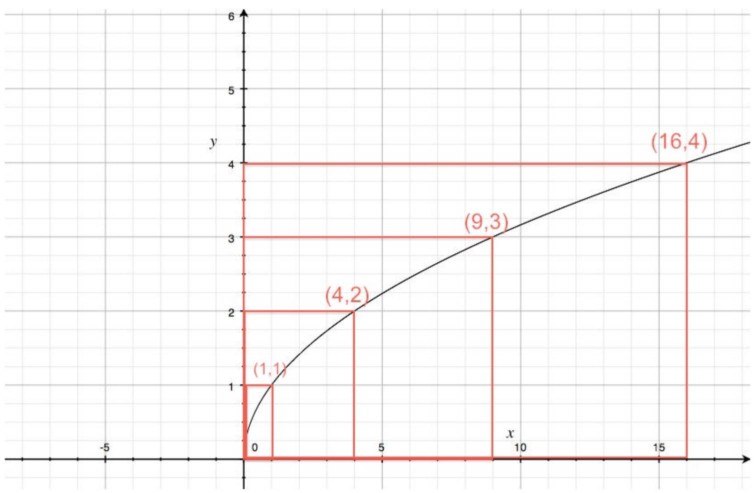


Figure 1.12: The graph of the function $f(x) = \sqrt{x}$. The domain of the function is $\cancel{x \in [0, \infty}$$x \in [0, \infty)$. You can't take the square root of a negative number.

Properties

- Domain: $\cancel{x \in [0, \infty}$$x \in [0, \infty)$.
  The function $f(x) = \sqrt{x}$ is only defined for nonnegative inputs $x \geq 0$. There is no real number y such that y^2 is negative, hence the function $f(x) = \sqrt{x}$ is not defined for negative inputs x.
- Image: $\cancel{f(x) \in [0, \infty}$$f(x) \in [0, \infty)$.
  The outputs of the function $f(x) = \sqrt{x}$ are never negative: $\sqrt{x} \geq 0$, for all $\cancel{x \in [0, \infty}$$x \in [0, \infty)$.

In addition to *square* root, there is also *cube* root $f(x) = \sqrt[3]{x} \equiv x^{\frac{1}{3}}$, which is the inverse function for the cubic function $f(x) = x^3$. We have $\sqrt[3]{8} = 2$ since $2 \times 2 \times 2 = 8$. More generally, we can define the n^{th}-root function $\sqrt[n]{x}$ as the inverse function of x^n.

Absolute value

The ~~absolute value~~ *absolute value* function tells us the ~~*size*~~ size of numbers without paying attention to whether the number is positive or negative. We can compute a number's absolute value by *ignoring the sign* of the ~~input number. Thus, a~~ number. A number's absolute value corresponds to its distance from the origin of the number line.

Another way of thinking about the absolute value function is to say it multiplies negative numbers by -1 to "cancel" their negative sign:

$$f(x) = |x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

Graph

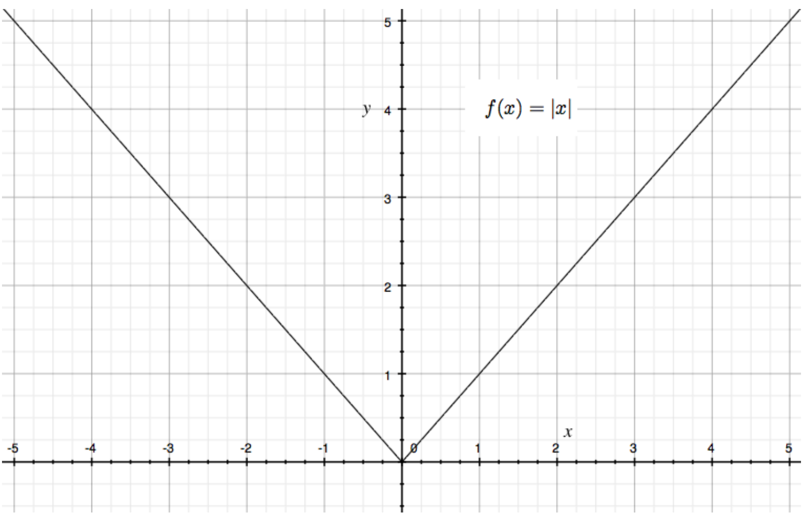


Figure 1.13: The graph of the absolute value function $f(x) = |x|$.

Properties

- Always returns a ~~non-negative number~~ nonnegative number
- The combination of squaring followed by square-root is equivalent to the absolute value function:

  $$\sqrt{x^2} \equiv |x|,$$

  since squaring destroys the sign.

Polynomial functions

The general equation for a polynomial function of degree n is written,

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n.$$

The constants a_i are known as the *coefficients* of the polynomial.

Parameters

- n: the *degree* of the polynomial
- a_0: the constant term
- a_1: the *linear* coefficient, or ~~first-order coefficient~~ *first-order* coefficient
- a_2: the *quadratic* coefficient
- a_3: the *cubic* coefficient
- a_n: the n^{th} order coefficient

A polynomial of degree n has $n+1$ coefficients: $a_0, a_1, a_2, \ldots, a_n$.

Properties

- Domain: $x \in \mathbb{R}$. Polynomials are defined for all inputs $x \in \mathbb{R}$.
- Image: depends on the coefficients
- The sum of two polynomials is also a polynomial.

Even and odd functions

The polynomials form an entire family of functions. Depending on the choice of degree n and coefficients a_0, a_1, ..., a_n, a polynomial function can take on many different shapes. We'll study polynomials and their properties in more detail in Section 1.15, but for now consider the following observations about the symmetries of polynomials:

- If a polynomial contains only even powers of x, like $f(x) = 1 + x^2 - x^4$ for example, we call this polynomial *even*. Even polynomials have the property $f(x) = f(-x)$. The sign of the input doesn't matter.
- If a polynomial contains only odd powers of x, for example $g(x) = x + x^3 - x^9$, we call this polynomial *odd*. Odd polynomials have the property $g(x) = -g(-x)$.
- If a polynomial has both even and odd terms then it is neither even nor odd.

The terminology of *odd* and *even* applies to functions in general and not just to polynomials. All functions that satisfy $f(x) = f(-x)$ are called *even functions*, and all functions that satisfy $f(x) = -f(-x)$ are called *odd functions*.

Sine

The sine function represents a fundamental unit of vibration. The graph of $\sin(x)$ *oscillates* up and down and crosses the x-axis multiple times. The shape of the graph of $\sin(x)$ corresponds to the shape of a vibrating string. See Figure 1.14.

In the remainder of this book, we'll meet the function $\sin(x)$ many times. We~~will~~'ll define the function $\sin(x)$ more formally as a trigonometric ratio in Section 1.16 (page 74). In Chapter 3 we~~will~~'ll use $\sin(x)$ and $\cos(x)$ (another trigonometric ratio) to work out the *components* of vectors. Later in Chapter 4, we~~will~~'ll learn how the sine function can be used to describe waves and periodic motion.

At this point in the book, however, we don't want to go into too much detail about all these applications. Let's hold off the discussion about vectors, triangles, angles, and ratios of lengths of sides and instead just focus on the graph of the function $f(x) = \sin(x)$.

Graph

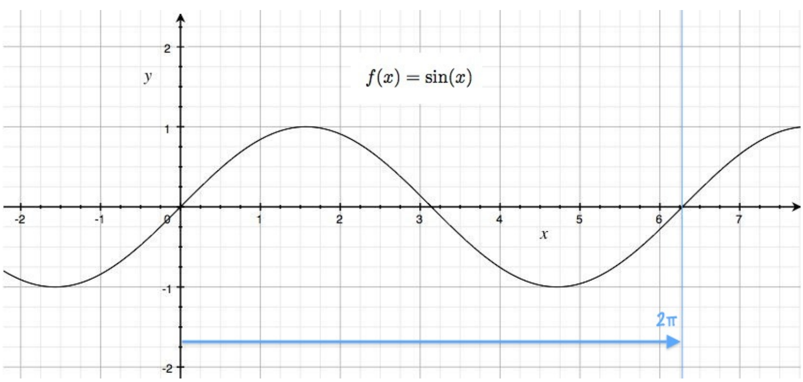
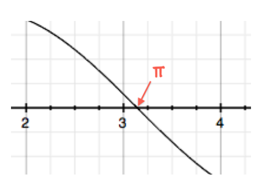


Figure 1.14: The graph of the function $y = \sin(x)$ passes through the following (x, y) coordinates: $(0, 0)$, $(\frac{\pi}{6}, \frac{1}{2})$, $(\frac{\pi}{4}, \frac{\sqrt{2}}{2})$, $(\frac{\pi}{3}, \frac{\sqrt{3}}{2})$, $(\frac{\pi}{2}, 1)$, $(\frac{2\pi}{3}, \frac{\sqrt{3}}{2})$, $(\frac{3\pi}{4}, \frac{\sqrt{2}}{2})$, $(\frac{5\pi}{6}, \frac{1}{2})$, and $(\pi, 0)$. For $x \in [\pi, 2\pi]$ the function has the same shape as for $x \in [0, \pi]$ but with negative values.



Let's start at $x = 0$ and follow the graph of the function $\sin(x)$ as it goes up and down. The graph starts from $(0, 0)$ and smoothly increases until it reaches the maximum value at $x = \frac{\pi}{2}$. Afterward, the function comes back down to cross the x-axis at $x = \pi$. After π, the function drops below the x-axis and reaches its

Tangent

The tangent function is the ratio of the sine and cosine functions:

$$f(x) = \tan(x) \equiv \frac{\sin(x)}{\cos(x)}.$$

Graph

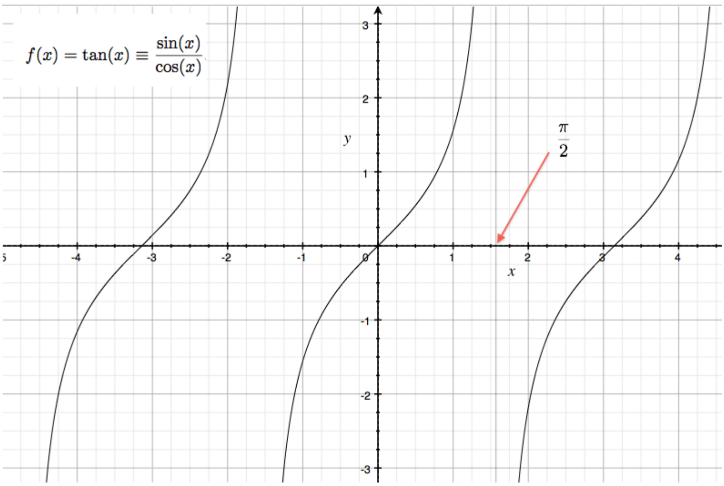


Figure 1.17: The graph of the function $f(x) = \tan(x)$.

Properties

- Domain: $\{x \in \mathbb{R} \mid x \neq \frac{(2n+1)\pi}{2}$ for any $n \in \mathbb{Z}\}$.
- ~~Range~~Image: $x \in \mathbb{R}$.
- The function tan is periodic with period π.
- The tan function "blows up" at values of x where $\cos x = 0$. These are called *asymptotes* of the function and their locations are $x = \ldots, \frac{-3\pi}{2}, \frac{-\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \ldots$.
- Value at $x = 0$: $\tan(0) = \frac{0}{1} = 0$, because $\sin(0) = 0$.
- Value at $x = \frac{\pi}{4}$: $\tan\left(\frac{\pi}{4}\right) = \frac{\sin(\frac{\pi}{4})}{\cos(\frac{\pi}{4})} = \frac{\frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}} = 1$.
- The number $\tan(\theta)$ is the length-ratio of the vertical and the horizontal sides in a right-angle triangle with angle θ.
- The inverse function of $\tan(x)$ is $\tan^{-1}(x)$.
- The inverse tangent function is used to compute the angle at the base in a right-angle triangle with horizontal side length ℓ_h and vertical side length ℓ_v: $\theta = \tan^{-1}\left(\frac{\ell_v}{\ell_h}\right)$.

Exponential

The exponential function base $e = 2.7182818\ldots$ is denoted

$$f(x) = e^x \equiv \exp(x).$$

Graph

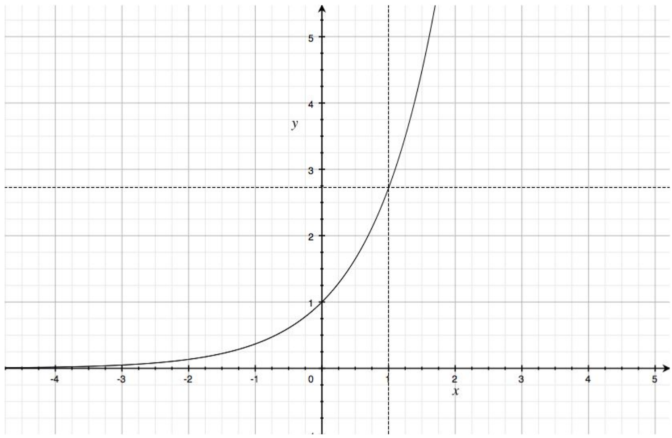


Figure 1.18: The graph of the exponential function $f(x) = e^x$ passes through the following (x, y) coordinates: $(-2, \frac{1}{e^2})$, $(-1, \frac{1}{e})$, $(0, 1)$, $(1, e)$, $(2, e^2)$, $(3, e^3 = 20.08\ldots)$, $(5, 148.41\ldots)$, and $(10, 22026.46\ldots)$.

Properties

- Domain: $x \in \mathbb{R}$
- ~~Range: $e^x \in (0, \infty$~~ Image: $e^x \in (0, \infty)$
- $f(a)f(b) = f(a + b)$ since $e^a e^b = e^{a+b}$.
- The derivative (the slope of the graph) of the exponential function is the exponential function: $f(x) = e^x \quad \Rightarrow \quad f'(x) = e^x$.

A more general exponential function would be $f(x) = Ae^{\gamma x}$, where A is the initial value, and γ (the Greek letter *gamma*) is the *rate* of the exponential. For $\gamma > 0$, the function $f(x)$ is increasing, as in Figure 1.18. For $\gamma < 0$, the function is decreasing and tends to zero for large values of x. The case $\gamma = 0$ is special since $e^0 = 1$, so $f(x)$ is a constant of $f(x) = A1^x = A$.

Links

[The exponential function 2^x evaluated]
`http://www.youtube.com/watch?v=e4MSN6IImpI`

The function $f(x) = 6.75(x^3 - 2x^2 + x)$, when stretched vertically by a factor of $A = 2$, becomes the function

$$g(x) = 2f(x) = 13.5(x^3 - 2x^2 + x).$$

The x-intercepts $f(0) = 0$ and $f(1) = 0$ do not move, and remain at $g(0) = 0$ and $g(1) = 0$. ~~The maximum at $x = \frac{1}{3}$ has doubled in value as $g(\frac{1}{3}) = 2$. Indeed, all~~ All values of $f(x)$ have been stretched upward by a factor of 2, as we can verify using the point $f(1.5) = 2.5$, which has become $g(1.5) = 5$. The maximum at $x = \frac{1}{3}$ has doubled in value becoming $g(\frac{1}{3}) = 2$.

Horizontal scaling

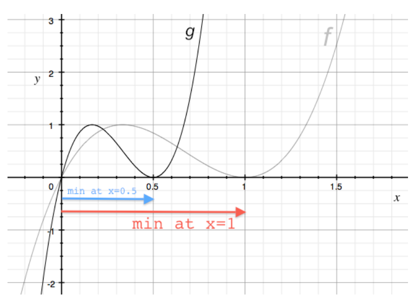
To stretch or compress a function horizontally, we can multiply the input value by some constant a to obtain:

$$g(x) = f(ax).$$

If $|a| > 1$, the function will be compressed. If $|a| < 1$, the function will be stretched. Note that the behaviour here is the opposite of vertical scaling. If a is a negative number, the function will also flip horizontally, which is a reflection through the y-axis.

The graph on the right shows $f(x) = 6.75(x^3 - 2x^2 + x)$, as well as the function $g(x)$, which is $f(x)$ compressed horizontally by a factor of $a = 2$:

$$\begin{aligned} g(x) &= f(2x) \\ &= 6.75\big[(2x)^3 - 2(2x)^2 + (2x)\big]. \end{aligned}$$



The x-intercept $f(0) = 0$ does not move since it is on the y-axis. The x-intercept $f(1) = 0$ does move, however, and we have $g(0.5) = 0$. The maximum at $x = \frac{1}{3}$ moves to $g(\frac{1}{6}) = 1$. All ~~values~~ points of $f(x)$ are compressed toward the y-axis by a factor of 2.

General quadratic function

The general quadratic function takes the form

$$f(x) = A(x - h)^2 + k,$$

where x is the input, and A, h, and k are the *parameters*.

General sine function

Introducing all possible parameters into the sine function gives us:

$$f(x) = A\sin\left(\tfrac{2\pi}{\lambda}x - \phi\right),$$

where A, λ, and ϕ are the function's parameters.

Parameters

- A: the amplitude describes the distance above and below the x-axis that the function reaches as it oscillates.
- λ: the *wavelength* of the function:

  $\lambda \equiv \{$ <u>the distance from one peak to the next</u> the horizontal distance fron

- ϕ: is a phase shift, analogous to the horizontal shift h, which we have seen. This number dictates where the oscillation starts. The default sine function has zero phase shift ($\phi = 0$), so it passes through the origin with an increasing slope.

The "bare" sin function $f(x) = \sin(x)$ has wavelength 2π and produces outputs that oscillate between -1 and $+1$. When we multiply the bare function by the constant A, the oscillations will range between $-A$ and A. When the input x is scaled by the factor $\frac{2\pi}{\lambda}$, the wavelength of the function becomes λ.

Exercises

E1.16 Given $f(x) = 2$ and $g(x) = x^2$, find $g \circ f(x)$.

1.16 $g(f(x)) = g(2) = 2^2 = 4$.

E1.17 Find the domain, the image, and the roots of $f(x) = 2\cos(x)$.

1.17 Domain: $x \in \mathbb{R}$. Image: $f(x) \in [-2, 2]$. Roots: $[\ldots, -\frac{\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots]$.

E1.18 Find the coefficients a, b, and c such that the quadratic function $f(x) = ax^2 + bx + c$ passes through the points $(0, 5)$, $(1, 4)$, and $(2, 5)$.
Hint: Find the equation $f(x) = A(x-h) + k$ first.

1.18 $f(x) = x^2 - 2x + 5$.

E1.19 Find the values α and β that will make the function $g(x) = 2\sqrt{x-\alpha} + \beta$ pass through the points $(3, -2)$, $(4, 0)$, and $(7, 2)$.

1.19 $g(x) = 2\sqrt{x-3} - 2$.

E1.20 What is the degree of the following polynomials? Are they even, odd, or neither?

a) $p(x) = x^2 - 5x^4 + 1$ b) $q(x) = x - x^3 + x^5 - x^7$

1.20 a) $p(x)$ is even and has degree 4. b) $q(x)$ is odd and has degree 7.

1.15 Polynomials

The polynomials are a simple and useful family of functions. For example, quadratic polynomials of the form $f(x) = ax^2 + bx + c$ often arise when describing physics phenomena.

Definitions

- x: the variable
- $f(x)$: the polynomial. We sometimes denote polynomials $P(x)$ to distinguish them from generic function $f(x)$.
- Degree of $f(x)$: the largest power of x that appears in the polynomial
- Roots of $f(x)$: the values of x for which $f(x) = 0$

Using a computer

When solving real-world problems, you'll often run into much more complicated equations. To find the solutions of anything more complicated than the quadratic equation, I recommend using a computer algebra system like `SymPy`: `http://live.sympy.org`.

To make the computer solve the equation $x^2 - 3x + 2 = 0$ for you, type in the following:

```
>>> solve(x**2 - 3*x + 2, x)          # usage: solve(expr, var)
[1, 2]
```

The function `solve` will find the roots of any equation of the form `expr = 0`. Indeed, we can verify that $x^2 - 3x + 2 = (x-1)(x-2)$, so $x = 1$ and $x = 2$ are the two roots.

Substitution trick

Sometimes you can solve fourth-degree polynomials by using the quadratic formula. Say you're asked to solve for x in

$$g(x) = x^4 - \cancel{3x}7x^2 \cancel{-10}+10 = 0.$$

Imagine this problem is on your exam, where you are not allowed the use of a computer. How does the teacher expect you to solve for x? The trick is to substitute $y = x^2$ and rewrite the same equation as

$$g(y) = y^2 - \cancel{3y - 10}7y + 10 = 0,$$

which you can solve by applying the quadratic formula. If you obtain the solutions $y = \alpha$ and $y = \beta$, then the solutions to the original fourth-degree polynomial are ~~$x = \sqrt{\alpha}$ and $x = \sqrt{\beta}$~~$x = \pm\sqrt{\alpha}$ and $x = \pm\sqrt{\beta}$, since $y = x^2$.

Since we're not taking an exam right now, we are allowed to use the computer to find the roots: Note how the second-degree polynomial has two roots, while the fourth-degree polynomial has four roots~~(two of which are imaginary, since we had to take the square root of anegative number to obtain them)~~. ~~The imaginary roots contain the unit imaginary number $i \equiv \sqrt{-1}$.~~

~~If you see this kind of problem on an exam, you should report the two real solutions as your answer—in this case $-\sqrt{5}$ and $\sqrt{5}$—without mentioning the imaginary solutions because you are not supposed to know about imaginary numbers yet. If you feel impatient and are ready to know about the imaginary numbers right now, feel free to skip ahead to the section on .~~

Exercises

E1.21 Solve for x in the following polynomial equations.

a) $3x + x^2 = x - 15 + 2x^2$
b) $3x^2 - 4x - 4 + x^3 = x^3 + 2x + 2$

1.21 a) $x = 5$ and $x = -3$; b) $x = 1 + \sqrt{3}$ and $x = 1 - \sqrt{3}$.

1.21 a) Rewrite the equation putting all terms on the right-hand side: $0 = x^2 - 2x - 15$. We can factor this quadratic by inspection. Are there numbers a and b such that $a + b = -2$ and $ab = -15$? Yes, $a = -5$ and $b = 3$, so $0 = (x - 5)(x + 3)$. b) Rewrite the equation so all terms are on the left-hand side: $3x^2 - 6x - 6 = 0$. Nice, the cubic terms cancel! We'll use the quadratic formula to solve this equation $x = \frac{6 \pm \sqrt{(-6)^2 - 4(3)(-6)}}{6} = \frac{6 \pm 6\sqrt{3}}{6} = 1 \pm \sqrt{3}$.

1.16 Trigonometry

We can put any three lines together to make a triangle. What's more, if one of the triangle's angles is equal to 90°, we call this triangle a *right-angle triangle*.

In this section we'll discuss right-angle triangles in great detail and get to know their properties. We'll learn some fancy new terms like *hypotenuse*, *opposite*, and *adjacent*, which are used to refer to the different sides of a triangle. We'll also use the functions *sine*, *cosine*, and *tangent* to compute the *ratios of lengths* in right triangles.

Understanding triangles and their associated trigonometric functions is of fundamental importance: you'll need this knowledge for your future understanding of mathematical subjects like vectors and complex numbers, as well as physics subjects like oscillations and waves.

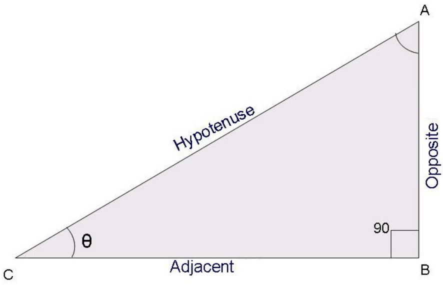


Figure 1.21: A right-angle triangle. The angle θ and the names of the sides of the triangle are indicated.

"Both I guess?"

"Well, as a function, I take angles as inputs and I give ratios as answers. More specifically, I tell you how 'wide' a triangle with that angle will be," says cos all in one breath.

"What do you mean wide?" you ask.

"Oh yeah, I forgot to say, the triangle must have a hypotenuse of length 1. What happens is there is a point P that moves around on a circle of radius 1, and we *imagine* a triangle formed by the point P, the origin, and the point on the x-axis located directly below the point P."

"I am not sure I get it," you confess.

"Let me try explaining," says sin. "Look on the next page, and you'll see a circle. This is the unit circle because it has a radius of 1. You see it, yes?"

"Yes."

"~~This circle is really cool. Imagine~~ Now imagine a point P that ~~starts from the point $P(0) = (1, 0)$ and~~ moves along the circle of radius ~~1.~~ 1, starting from the point $P(0) = (1, 0)$. The x and y coordinates of the point $P(\theta) = (P_x(\theta),\ P_y(\theta))$ as a function of θ are

$$P(\theta) = (P_x(\theta),\ P_y(\theta)) = (\cos\theta,\ \sin\theta).$$

So, *either* you can think of us in the context of triangles, or you think of us in the context of the unit circle."

"Cool. I kind of get it. Thanks so much," you say, but in reality you are weirded out. Talking functions? "Well guys. It was nice to meet you, but I have to get going, to finish the rest of the book."

"See you later," says cos.

"Peace out," says sin.

The unit circle

The ~~unit circle~~ *unit circle* consists of all points (x, y) that satisfy the equation $x^2 + y^2 = 1$. A point $P = (P_x, P_y)$ on the unit circle has coordinates $(P_x, P_y) = (\cos\theta, \sin\theta)$, where θ is the angle P makes with the x-axis.

Calculators

Make sure to set your calculator to the correct units for working with angles. What should you type into your calculator to compute the sine of 30 degrees? If your calculator is set to degrees, simply type: `30`, `sin`, `=`.

If your calculator is set to radians, you have two options:

1. Change the `mode` of the calculator so it works in degrees.
2. Convert 30° to radians

$$30\ [°] \times \frac{2\pi\ [\text{rad}]}{360\ [°]} = \frac{\pi}{6}\ [\text{rad}],$$

and type: π, `/`, `6`, `sin`, `=` on your calculator.

Exercises

E1.22 Given a circle with radius $r = 5$, find the x- and y-coordinates of the point at $\theta = 45°$. What is the circumference of the circle?

1.22 $x = 5\cos(45°) = 3.54$, $y = 5\sin(45°) = 3.54$; $C = 10\pi$.

E1.23 Convert the following angles from degrees to radians.

(a) 30° (b) 45° (c) 60° (d) 270°

1.23 (a) $\frac{\pi}{6}$[rad]; (b) $\frac{\pi}{4}$[rad]; (c) $\frac{\pi}{3}$[rad]; (d) $\frac{3\pi}{2}$[rad].

1.23 To convert an angle measure from degrees to radians we must multiply it by the conversion ratio $\frac{\pi}{180}$[rad/°].

Links

[Unit-circle walkthrough and tricks by patrickJMT on YouTube]
http://bit.ly/1mQg9Cj and http://bit.ly/1hvA702

1.17 Trigonometric identities

There are a number of important relationships between the values of the functions sin and cos. Here are three of these relationships, known as *trigonometric identities*. There about a dozen other identities that are less important, but you should memorize these three.

The three identities to remember are:

1. Unit hypotenuse

$$\sin^2(\theta) + \cos^2(\theta) = 1.$$

The unit hypotenuse identity is true by the Pythagoras theorem and the definitions of sin and cos. The sum of the squares of the sides of a triangle is equal to the square of the hypotenuse.

2. sico + sico

$$\sin(a+b) = \sin(a)\cos(b) + \sin(b)\cos(a).$$

The mnemonic for this identity is "sico + sico."

3. coco − sisi

$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b).$$

The mnemonic for this identity is "coco - sisi." The negative sign is there because it's not good to be a sissy.

Derived formulas

If you remember the above three formulas, you can derive pretty much all the other trigonometric identities.

Double angle formulas

Starting from the ~~sico-sico~~ sico+sico identity as explained above, and setting $a = b = x$, we can derive the following identity:

$$\sin(2x) = 2\sin(x)\cos(x).$$

Starting from the coco-sisi identity, we obtain

$$\begin{aligned}\cos(2x) &= \cos^2(x) - \sin^2(x) \\ &= 2\cos^2(x) - 1 \ = 2\left(1 - \sin^2(x)\right) - 1 \ = 1 - 2\sin^2(x).\end{aligned}$$

The formulas for expressing $\sin(2x)$ and $\cos(2x)$ in terms of $\sin(x)$ and $\cos(x)$ are called *double angle formulas.*

If we rewrite the double-angle formula for $\cos(2x)$ to isolate the $\sin^2$ or the $\cos^2$ term, we obtain the *power-reduction formulas*:

$$\cos^2(x) = \frac{1}{2}\left(1 + \cos(2x)\right), \qquad \sin^2(x) = \frac{1}{2}\left(1 - \cos(2x)\right).$$

Exercises

E1.24 Given $a = \pi$ and $b = \frac{\pi}{2}$, find

a) $\sin(a + b)$ b) $\cos(2a)$ c) $\cos(a + b)$

1.24 a) -1; b) 1; c) 0.

E1.25 Simplify the following expressions and compute their value without using a calculator.

a) $\cos(x) + \cos(\pi - x)$ b) $2\sin^2(x) + \cos(2x)$

c) $\sin(\frac{5\pi}{4})\sin(-\frac{\pi}{4})$ d) $2\cos(\frac{5\pi}{4})\cos(-\frac{\pi}{4})\cos(\pi)$

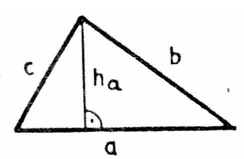
1.25 a) 0; b) 1; c) $\frac{1}{2}$; d) 1.

1.18 Geometry

Triangles

The area of a triangle is equal to $\frac{1}{2}$ times the length of its base times its height:

$$A = \frac{1}{2} a h_a.$$

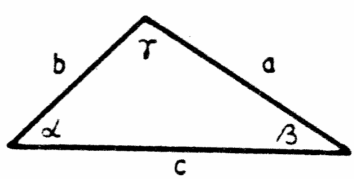


Note that h_a is the height of the triangle *relative to* the side a.

The perimeter of a triangle is

$$P = a + b + c.$$

Consider a triangle with internal angles α, β and γ. The sum of the inner angles in any triangle is equal to two right angles: $\alpha + \beta + \gamma = 180°$.



Sine rule The sine law is

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

where α is the angle opposite to a, β is the angle opposite to b, and γ is the angle opposite to c.

Cosine rule The cosine rules are

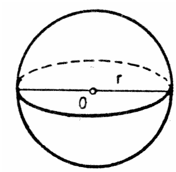
$$a^2 = b^2 + c^2 - 2bc\cos(\alpha),$$
$$b^2 = a^2 + c^2 - 2ac\cos(\beta),$$
$$c^2 = a^2 + b^2 - 2ab\cos(\gamma).$$

The cosine rule is useful when you know two sides of a triangle and the angle between them.

Sphere

A sphere is described by the equation

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



The surface area of a sphere is

$$A = 4\pi r^2,$$

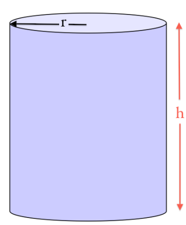
and the volume of a sphere is

$$V = \frac{4}{3}\pi r^3.$$

Cylinder

The surface area of a cylinder consists of the top and bottom circular surfaces, plus the area of the side of the cylinder:

$$A = 2\left(\pi r^2\right) + (2\pi r)h.$$



The formula for the volume of a cylinder is the product of the area of the cylinder's base times its height:

$$V = \left(\pi r^2\right) h.$$

Example You open the hood of your car and see 2.0 L written on top of the engine. The 2.0 L refers to the combined volume of the four pistons, which are cylindrical in shape. The owner's manual tells you the diameter of each piston (bore) is 87.5 mm, and the height of each piston (stroke) is 83.1 mm. Verify that the total volume of the cylinder displacement of your engine is indeed 1998789 mm^3 $\approx$ 2 L.

Exercises

E1.26 Find the length of side x in the triangle below.

E1.27 Find the volume and the surface area of a sphere with radius 2.

1.27 $V = 33.51$ and $A = 50.26$.

1.27 The volume of the sphere with radius $r = 2$ is $V = \frac{4}{3}\pi 2^3 = 33.51$. It's surface are is $A = 4\pi 2^2 = 50.26$.

1.19 Circle

The *circle* is a set of points located a constant distance from a centre point. This geometrical shape appears in many situations.

Definitions

- r: the radius of the circle
- A: the area of the circle
- C: the circumference of the circle
- (x, y): a point on the circle
- θ: the angle (measured from the x-axis) of some point on the circle

Formulas

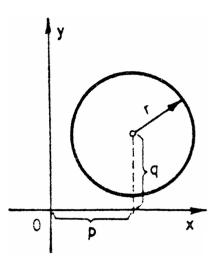
A circle with radius r centred at the origin is described by the equation

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

All points (x, y) that satisfy this equation are part of the circle.

Rather than staying centred at the origin, the circle's centre can be located at any point (p, q) on the plane:

$$(x - p)^2 + (y - q)^2 = r^2.$$



Explicit function

The equation of a circle is a *relation* or an *implicit function* involving x and y. To obtain an *explicit function* $y = f(x)$ for the circle, we can solve for y to obtain

$$y = \sqrt{r^2 - x^2}, \quad -r \leq x \leq r,$$

and

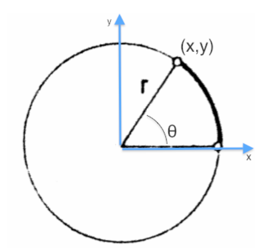
$$y = -\sqrt{r^2 - x^2}, \quad -r \leq x \leq r.$$

The explicit expression is really two functions, because a vertical line crosses the circle in two places. The first function corresponds to the top half of the circle, and the second function corresponds to the bottom half.

Polar coordinates

Circles are so common in mathematics that mathematicians developed a special "circular coordinate system" in order to describe them more easily.

It is possible to specify the coordinates (x, y) of any point on the circle in terms of the *polar coordinates* $r\angle\theta$, where r measures the distance of the point from the origin, and θ is the angle measured from the x-axis.



To convert from the polar coordinates $r\angle\theta$ to the (x, y) coordinates, use the trigonometric functions cos and sin:

$$x = r\cos\theta \qquad \text{and} \qquad y = r\sin\theta.$$

Parametric equation

We can describe *all* the points on the circle if we specify a fixed radius r and vary the angle θ over all angles: $\theta \in [0, 360°)$. A *parametric equation* specifies the coordinates $(x(\theta), y(\theta))$ for the points on a curve, for all values of the *parameter* θ. The parametric equation for a circle of radius r is given by

$$\{(x, y) \in \mathbb{R}^2 \mid x = r\cos\theta,\ y = r\sin\theta,\ \theta \in [0, 360°)\}.$$

In words, this expression describes the set of points (x, y) in the Cartesian plane whose x-coordinates are described by $r\cos\theta$ and whose y-coordinates are described by $r\sin\theta$, where the angle θ varies from $0°$ to $360°$. Try to visualize the curve traced by the point

Measuring angles in radians is equivalent to measuring arc length on a circle with radius $r = 1$.

Exercises

E1.28 Every Montrealer knows they should bring their bike home if they don't want it to get stolen. On a rainy day, Laura brings her bike home and the wet bicycle tires leave traces on the floor. What is the length of the wet trace left by the bike's rear tire (diameter 73[cm]) if the wheel makes five full turns.

1.28 Length of trace $= 5C = 5\pi d = 11.47$[m].

E1.29 Write a formula to describe a circle of radius 3 centred at $(1, 4)$. Express you answer in Cartesian and polar coordinates.

1.29 $(x-1)^2 + (y-4)^2 = 9$ or $\{(x, y) \in \mathbb{R}^2 \mid x = 1 + 3\cos\theta, y = 4 + 3\sin\theta, \theta \in [0, 2\pi)\}$.

1.20 Ellipse

The *ellipse* is a fundamental shape that occurs in nature. The orbit of planet Earth around the Sun is an ellipse.

Parameters

- a: the half-length of the ellipse along the x-axis, also known as the semi-major axis
- b: the half-length of the ellipse along the y-axis
- ε: the *eccentricity* of the ellipse, $\varepsilon \equiv \sqrt{1 - \frac{b^2}{a^2}}$
- F_1, F_2: the two *focal points* of the ellipse
- r_1: the distance from a point on the ellipse to F_1
- r_2: the distance from a point on the ellipse to F_2

Definition

An ellipse is the curve found by tracing along all the points for which the sum of the distances to the two focal points is a constant:

$$r_1 + r_2 = \text{const.}$$

There's a neat way to draw a perfect ellipse using a piece of string and two tacks or pins. Take a piece of string and tack it to a picnic table at two points, leaving some loose slack in the middle of the string. Now take a pencil, and without touching the table, use the pencil to of mass m must be of the form $F_g = \frac{GMm}{r^2}$. We'll discuss more about the law of gravitation in Chapter 4.

For now, let's give props to Newton for connecting the dots, and props to Johannes Kepler for studying the orbital periods, and Tycho Brahe for doing all the astronomical measurements. Above all, we owe some props to the ellipse for being such an awesome shape!

By the way, the varying distance between the Earth and the Sun is not the reason we have seasons. The ellipse had nothing to do with seasons! Seasons are predominantly caused by the *axial tilt* of the Earth. The axis of rotation of the Earth is tilted by 23.4° relative to the plane of its orbit around the Sun. In the Northern hemisphere, the longest day of the year is the summer solstice, which occurs around the 21~~th~~st of June. On that day, the Earth's spin axis is tilted toward the Sun so the Northern hemisphere receives the most sunlight.

Links

[Further reading about Earth-Sun geometry]
`http://www.physicalgeography.net/fundamentals/6h.html`

1.21 Hyperbola

The *hyperbola* is another fundamental shape of nature. A horizontal hyperbola is the set of points (x, y) which satisfy the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

The numbers a and b are arbitrary constants. This hyperbola passes through the points $(-a, 0)$ and $(a, 0)$. The eccentricity of this hyperbola is defined as

$$\varepsilon = \sqrt{1 + \frac{b^2}{a^2}}\,.$$

Eccentricity is an important parameter of the hyperbola, as it determines the hyperbola's shape. Recall the ellipse is also defined by an eccentricity parameter, though the formula is slightly different. This could be a coincidence—or is there a connection? Let's see.

Graph

The graph of a hyperbola consists of two separate *branches*, as illustrated in Figure 1.27. We'll focus our discussion mostly on the right branch of the hyperbola.

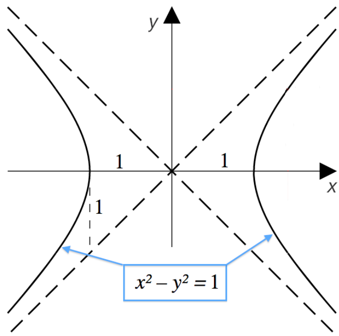
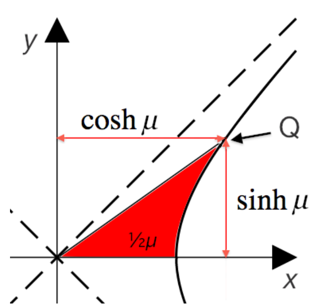


Figure 1.27: The unit hyperbola $x^2 - y^2 = 1$. The graph of the hyperbola has two branches, opening to the sides. The dashed lines are called the *asymptotes* of the hyperbola. The eccentricity determines the angle between the asymptotes. The eccentricity of $x^2 - y^2 = 1$ is $\varepsilon = \sqrt{1 + \frac{1}{1}} = \sqrt{2}$.

Hyperbolic trigonometry

The trigonometric functions sin and cos describe the geometry of the unit circle. The point $P = (\cos\theta, \sin\theta)$ traces out the unit circle as the angle θ goes from 0 to 2π. The function cos is defined as the x-coordinate of the point P, and sin is the y-coordinate. The study of the geometry of the points on the unit circle is called *circular trigonometry.*

Instead of looking at a point P on the unit circle $x^2 + y^2 = 1$, let's trace out the path of a point Q on the unit hyperbola $x^2 - y^2 = 1$. We will now define *hyperbolic* variants of the sin and cos functions to describe the coordinates of the point Q. This is called *hyperbolic trigonometry.* Doesn't that sound awesome? Next time your friends ask what you have been up to, tell them you are learning about hyperbolic trigonometry.



The coordinates of a point Q on the unit hyperbola are $Q = (\cosh\mu, \sinh\mu)$, where μ is the *hyperbolic angle*. The x-coordinate of the point Q is $x = \cosh\mu$, and its y-coordinate is $y = \sinh\mu$. The name hyperbolic angle is a bit of a misnomer, since ~~$\mu \in [0, \infty$~~ $\mu \in [0, \infty)$ actually measures an area. The area of the highlighted region in the figure corresponds to $\frac{1}{2}\mu$.

Recall the circular-trigonometric identity $\cos^2\theta + \sin^2\theta = 1$, which fol-

Percentages

We often talk about ratios between quantities, rather than mentioning the quantities themselves. For example, we can imagine average Joe, who invests \$1000 in the stock market and loses \$300 because the boys on Wall Street keep pulling dirty tricks on him. To put the number \$300 into perspective, we can say Joe lost 0.3 of his investment, or ~~alternately,~~ alternatively, we can say Joe lost 30% of his investment.

To express a ratio as a percentage, multiply it by 100. The ratio of Joe's loss to investment is

$$R = 300/1000 = 0.3.$$

The same ratio expressed as a percentage gives

$$R = 300/1000 \times 100 = 30\%.$$

To convert from a percentage to a ratio, divide the percentage by 100.

Interest rates

Say you take out a \$1000 loan with an interest rate of 6% compounded annually. How much will you owe in interest at the end of the year?

Since 6% corresponds to a ratio of 6/100, and since you borrowed \$1000, the accumulated interest at the end of the year will be

$$I_1 = \frac{6}{100} \times \$1000 = \$60.$$

At year's end, you'll owe the bank a total of

$$L_1 = \left(1 + \frac{6}{100}\right) 1000 = (1 + 0.06)1000 = 1.06 \times 1000 = \$1060.$$

The total money owed after 6 years will be

$$L_6 = (1.06)^6 \times 1000 = \$1418.52.$$

You borrowed \$1000, but in six years you will need to give back \$1418.52. This is a terrible deal! But it gets worse. The above scenario assumes that the bank compounds interest only once per year. In practice, interest is compounded each month.

Monthly compounding

An annual compounding schedule is disadvantageous to the bank, and since the bank writes the rules, compounding is usually performed every month.

The monthly interest rate can be used to find the annual rate. The bank quotes the *nominal annual percentage rate* (APR), which is equal to

$$\cancel{\text{nAPR}}\text{nominal APR} = 12 \times r,$$

where r is the monthly interest rate.

Suppose we have a nominal APR of 6%, which gives a monthly interest rate of $r = 0.5\%$. If you borrow \$1000 at that interest rate, at the end of the first year you will owe

$$L_1 = \left(1 + \frac{0.5}{100}\right)^{12} \times 1000 = \$1061.68,$$

and after 6 years you will owe

$$L_6 = \left(1 + \frac{0.5}{100}\right)^{72} \times 1000 = 1.061677^6 \times 1000 = \$1432.04.$$

Note how the bank tries to pull a fast one: the *effective* APR is actually 6.16%, not 6%. Every twelve months, the amount due will increase by the following factor:

$$\cancel{\text{eAPR}}\text{effective APR} = \left(1 + \frac{0.5}{100}\right)^{12} = 1.0616.$$

Thus the effective annual percent rate is ~~eAPR = 6.16%.~~ 6.16%, but it's legal for banks to advertise it as "6% nominal APR." Sneaky stuff.

Compounding infinitely often

~~What is~~ We saw that more frequent compounding leads to higher effective interest rates. Let's find a formula for the effective APR if the nominal APR is 6% and the bank performs the compounding n times per year~~?~~.

The annual growth ratio will be

$$\left(1 + \frac{6}{100n}\right)^{n},$$

where the interest rate per compounding period is $\frac{6}{n}\%$, and there are n periods per year.

Consider a scenario in which the compounding is performed infinitely often. This corresponds to the case when the number n in the above equation tends to infinity (denoted $n \to \infty$). This is not a practical question, but it is an interesting avenue to explore nevertheless because it leads to the definition of the natural exponential function $f(x) = e^x$.

When we set $n \to \infty$ in the above expression, the annual growth ratio will be described by the exponential function base e as follows:

$$\lim_{n\to\infty} \left(1 + \frac{6}{100n}\right)^n = \exp\!\left(\frac{6}{100}\right) = 1.0618365.$$

The expression "$\lim_{n\to\infty}$" is to be read as "in the limit when n tends to infinity." We will learn more about limits in Chapter 5.

A nominal APR of 6% with compounding that occurs infinitely often has ~~an eAPR = 6.183%~~effective APR of 6.183%. After six years you will owe

$$L_6 = \exp\!\left(\frac{6}{100}\right)^6 \times 1000 = \$1433.33.$$

The nominal APR is 6% in each case, yet, the more frequent the compounding schedule, the more money you'll owe after six years.

Exercises

E1.33 Studious Jack borrowed \$40 000 to complete his university studies and made no payments since graduation. Calculate how much money he owes after 10 years in each of the following scenarios.

(a) Nominal annual interest rate of 3% with monthly compounding.

(b) Effective annual interest rate of 4%.

(c) Nominal annual interest rate of 5% with infinite compounding.

1.33 (a) \$53 974.14; (b) \$59 209.77; (c) \$65 948.79.

1.33 (a) Since the compounding is performed monthly, we first calculate the monthly interest rate $r = \frac{3\%}{12} = 0.25\% = 0.0025$. The sum Jack owes after ten years is $\$40\,000(1.0025)^{120} = \$53\,974.14$. (b) The calculation using the effective annual interest rate is more direct $\$40\,000(1.04)^{10} = \$59\,209.77$. (c) When compounding infinitely often at a nominal annual interest rate of 5%, the amount owed will grow by $\exp(\frac{5}{100}) = 1.051271$ each year. After ten years Jack will owe $\$40\,000(1.051271)^{10} = \$65\,948.79$.

E1.34 Entrepreneurial Kate borrowed \$20 000 to start a business. Initially her loan had an effective annual percentage rate of 6%, but after five years she negotiates with the bank and obtains a lower rate of 4%. How much money does she owe after 10 years?

1.34 \$32 563.11.

1.34 Since there are two different interest rates in effect, we have to perform two separate calculations. At the end of the first five years Kate owes $\$20\,000(1.06)^5 = \$26\,764.51$. For the remaining five years the interest changed to 4%, so the sum Kate owes after 10 years is $\$26\,764.51(1.04)^5 = \$32\,563.11$.

1.24 Set notation

A *set* is the mathematically precise notion for describing a group of objects. You don't need to know about sets to perform simple math; but more advanced topics require an understanding of what sets are and how to denote set membership and subset relations between sets.

Definitions

- *set*: a collection of mathematical objects
- S, T: the usual variable names for sets
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$: some important sets of numbers: the naturals, the integers, the rationals, and the real numbers, respectively.
- $\{$ definition $\}$: the curly brackets surround the definition of a set, and the expression inside the curly brackets describes what the set contains.

Set operations:

- $S \cup T$: the *union* of two sets. The union of S and T corresponds to the elements in either S or T.
- $S \cap T$: the *intersection* of the two sets. The intersection of S and T corresponds to the elements that are in both S and T.
- $S \setminus T$: *set difference* or *set minus*. ~~The~~ The set difference $S \setminus T$ corresponds to the elements of S that are not in T.

Set relations:

- $\subset$: is a subset of
- $\subseteq$: is a subset of or equal to

Special mathematical shorthand symbols and their corresponding meanings:

- $\forall$: for all
- $\exists$: there exists
- $\nexists$: there doesn't exist
- $|$: such that
- $\in$: element of
- $\notin$: not an element of

Sets

Much of math's power comes from *abstraction*: the ability to see the bigger picture and think *meta* thoughts about the common relationships between math objects. We can think of individual numbers like 3, -5, and π, or we can talk about the *set* of *all* numbers.

It is often useful to restrict our attention to a specific *subset* of the numbers as in the following examples.

Example 1: The nonnegative real numbers

Define $\mathbb{R}_+ \subset \mathbb{R}$ (read "$\mathbb{R}_+$ a subset of $\mathbb{R}$") to be the set of ~~non-negative~~ nonnegative real numbers:

$$\mathbb{R}_+ \equiv \{\text{all } x \text{ in } \mathbb{R} \text{ such that } x \geq 0\},$$

or expressed more compactly,

$$\mathbb{R}_+ \equiv \{x \in \mathbb{R} \mid x \geq 0\}.$$

If we were to translate the above expression into plain English, it would read "the set $\mathbb{R}_+$ is defined as the set of all real numbers x such that x is greater or equal to zero."

Example 2: Odd and even integers

Define the set of even integers as

$$E \equiv \{nm \in \mathbb{Z} \mid m = 2n,\ n \in \mathbb{Z}\} = \{\ldots, -4, -2, 0, 2, 4, 6, \ldots\}$$

and the set of odd integers as

$$O \equiv \{nm \in \mathbb{Z} \mid m = 2n+1,\ n \in \mathbb{Z}\} = \{\ldots, -3, -1, 1, 3, 5, \ldots\}.$$

Indeed, every even number is divisible by two so it can be written in the form $2n$ for some integer n. Odd numbers can be obtained from the "template" $2n+1$, with n varying over all integers.

In both of the above examples, we use the mathematical notation $\{\ldots \mid \ldots\}$ to define the sets. Inside the curly brackets we first describe the general kind of objects we are talking about, followed by the symbol "$|$" (read "such that"), followed by the conditions that must be satisfied by all elements of the set.

Number sets

Recall how we defined the fundamental sets of numbers in the beginning of this chapter. It is worthwhile to review them briefly.

The *natural* numbers form the set derived when you start from 0 and add 1 any number of times:

$$\mathbb{N} \equiv \{0, 1, 2, 3, 4, 5, 6, \ldots\}.$$

The integers are the numbers derived by adding or subtracting 1 some number of times:

$$\mathbb{Z} \equiv \{x \mid x = \pm n, n \in \mathbb{N}\}.$$

When we allow for divisions between integers, we get the rational numbers:

$$\mathbb{Q} \equiv \left\{ z \;\middle|\; z = \frac{x}{y} \text{ where } x \text{ and } y \text{ are in } \mathbb{Z}, \text{ and } y \neq 0 \right\}.$$

The broader class of real numbers also includes all rationals as well as irrational numbers like $\sqrt{2}$ and π:

$$\mathbb{R} \equiv \{\pi, e, -1.53929411\ldots, \; 4.99401940129401\ldots, \; \ldots\}.$$

Finally, we have the set of complex numbers:

$$\mathbb{C} \equiv \{1, i, 1+i, 2+3i, \ldots\}.$$

Note that the definitions of $\mathbb{R}$ and $\mathbb{C}$ are not very precise. Rather than giving a precise definition of each set inside the curly brackets as we did for $\mathbb{Z}$ and $\mathbb{Q}$, we instead stated some examples of the elements in the set. Mathematicians sometimes do this and expect you to guess the general pattern for all the elements in the set.

The following inclusion relationship holds for the fundamental sets of numbers:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

This relationship means every natural number is also an integer. Every integer is a rational number. Every rational number is a real. Every real number is also a complex number.

Set relations and set operations

We'll now introduce a useful graphical representation for set relations and set operations. Sets are purely mathematical constructs so they have no "shape," but we can draw *Venn diagrams* to visualize relationships between sets and different subsets.

Figure 1.29 illustrates the notion of a set B that is strictly contained in the set A. We say $B \subset A$ if $\forall b \in B$, we also have $b \in A$.

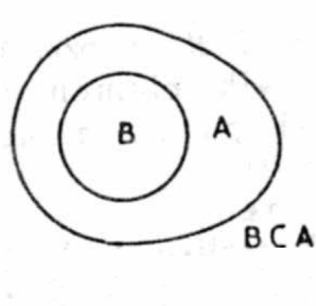


Figure 1.29: An illustration of a set B which is strictly contained in another set A, denoted $B \subset A$.

This is the picture that mathematicians think about when they say "the set B is contained in the set A." Set containment is an abstract mathematical notion, but the picture helps us visualize the situation.

Mathematicians actually have two distinct ways to denote set containment in order to specify whether they refer to a "strict" containment relation or a "subset or equal" relation. This choice of notation is analogous to the *less than* ($<$) and *less than or equal* ($\leq$) relations between numbers. The notation $\subset$ refers to strict containment: if $B \subset A$ then $\forall b \in B$, we also have $b \in A$, and $\exists a \in A$ such that $a \notin B$. In other words, we write $B \subset A$ whenever the set A contains B, but there exists at least one element in A that is not an element of B. Some mathematicians prefer the more descriptive $\subsetneq$ to describe this relation. The expression $B \subseteq A$ is a weaker statement than $B \subset A$. In writing $B \subseteq A$, a mathematician claims "if $\forall b \in B$, we also have $b \in A$," but makes no claim about the existence of elements e that are in A but not in B.

For example, the set of even numbers is a strict subset of the integers, $E \subset \mathbb{Z}$, since every even number is an integer but there exist integers that are not even (odd numbers).

Venn diagrams are also useful to visualize the subsets obtained from set operations applied on sets. The set intersection $A \cap B$ and set difference $A \setminus B$ are illustrated in Figure 1.30.

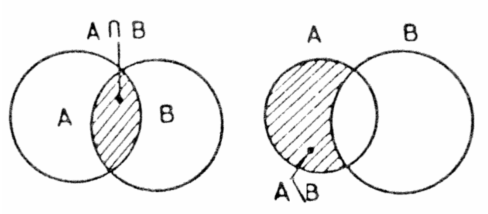


Figure 1.30: The left side of the figure illustrates $A \cap B$, the intersection of the sets A and B. The intersection of two sets contains the elements that are part of both sets. The right side of the figure shows the set difference $A \setminus B$, which consists of all the elements that are in A but not in B.

The "picture" you should have in your head when thinking about the union of two sets $A \cup B$ is shown in Figure 1.31. The shaded area in the figure indicates all the elements that are in either A or B (or both). If $e \in A \cup B$, then $e \in A$ or $e \in B$.

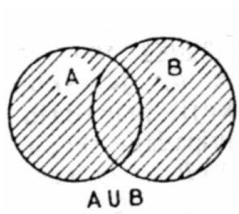


Figure 1.31: The union of two sets $A \cup B$.

Note the English expression "in A or B" is somewhat ambiguous, since it could be interpreted as an "exclusive or" meaning: "in A or B, but not both." Mathematicians use "or" in the inclusive sense when they say "in A or B," and when they write $A \cup B$. The expression $(A \cup B) \setminus (A \cap B)$ can be used to obtain the "exclusive or" of the sets.

Recall the set of even numbers $E \subset \mathbb{Z}$ and the set of odd numbers $O \subset \mathbb{Z}$ defined above. Since every integer is either an even number or an odd number, we know $\mathbb{Z} \subseteq E \cup O$. The union of two subsets is always contained within the parent set, so we also know $E \cup O \subseteq \mathbb{Z}$. Combining these facts we can establish the equality $E \cup O = \mathbb{Z}$, which states the (rather obvious) fact "the combination of all even and odd numbers is the same as all integers."

Example 3: Set operations

Consider the four sets $A = \{a, b, c\}$, $B = \{b, c, d\}$, and $C = \{c, d, e\}$. Using set operations we can define new sets like

$$A \cup B = \{a, b, c, d\}, \qquad A \cap B = \{b, c\}, \qquad A \setminus B = \{a\},$$

which correspond to elements that are in either A or B, the set of elements that are in A and B, and the set of elements that are in A but not in B. We can also construct expressions involving three sets:

$$A \cup B \cup C = \{a, b, c, d\}, \qquad A \cap B \cap C = \{c\}.$$

We can also write more elaborate set expressions like

$$(A \cup B) \setminus C = \{a, b\},$$

which is the set of elements that are in A or B but not in C.

Another example of a complicated set expression is

$$(A \cap B) \cup (B \cap C) = \{b, c, d\},$$

which describes the set of elements that are in both A and B or in both B and C. As you can see, set notation is a compact and precise language for writing complicated set expressions.

Example 4: Word problem

A startup is looking to hire student interns for the summer. Let S denote the whole set of students. Define C to be the subset of students who are good with computers, M the subset of students who know math, D the students who have design skills, and L the students who have good language skills.

Using set notation we can specify different subsets of the students that startups would be interested in hiring. A math textbook startup would like to hire students from the set $M \cap L$, which corresponds to students that are good at math and also have good language skills. A startup that builds websites needs designers for certain tasks and coders for other tasks, so it would be interested in students from the set $D \cup C$.

New vocabulary

The specialized notation used by mathematicians can be difficult to get used to. You must learn how to read symbols like $\exists$, $\subset$, $|$, and

$\in$ and translate their meaning in the sentence. Indeed, learning advanced mathematics notation is akin to learning a new language.

To help you practice the new vocabulary, we~~will look at an~~'ll now look at some mathematical arguments that make use of the new symbols.

Simple example

Claim: Given $J(n) = 3n + 2 - n$, $J(n) \in E$ for all $n \in \mathbb{Z}$. The claim is that $J(n)$ is always an even number, whenever n is an integer. This means no matter which integer number n we choose, the function $J(n) = 3n + 2 - n$ will always output an even number.

Proof: We want to show $J(n) \in E$ for all $n \in \mathbb{Z}$. Let's first review the definition of the set of even numbers $E \equiv \{m \in \mathbb{Z} \mid m = 2n, n \in \mathbb{Z}\}$. A number is even if it is equal to $2n$ for some integer n. Next let's simplify the expression for $J(n)$ as follows

$$J(n) = 3n + 2 - n = 2n + 2 = 2(n + 1).$$

Observe that the number $(n + 1)$ is always an integer whenever n is an integer. Since the output of $J(n) = 2(n + 1)$ is equal to $2m$ for some integer m, we have proved that $J(n) \in E$, for all $n \in \mathbb{Z}$.

Less simple example: Square-root of 2 is irrational

The following is an ancient mathematical proof ~~and express it~~ expressed in terms of modern ~~mathematical~~ math symbols.

~~Square-root of 2 is irrational~~

Claim: $\sqrt{2} \notin \mathbb{Q}$. ~~This means there do not~~ The claim is that $\sqrt{2}$ is not part of the set of rational numbers. Recall the definition of the set of rational numbers $\mathbb{Q} \equiv \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0 \right\}$. If $\sqrt{2} \notin \mathbb{Q}$, this means there don't exist numbers $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$ such that $m/n = \sqrt{2}$. The last sentence expressed in mathematical notation would read,

$$\nexists\, m \in \mathbb{Z}, n \in \mathbb{Z} \;\mid\; m/n = \sqrt{2}.$$

To prove the claim we'll use a technique called *proof by contradiction*. We begin by assuming the opposite of what we want to prove: that there exist numbers $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$ such that $m/n = \sqrt{2}$. We'll then carry out some simple algebra steps and in the end we'll obtain an equation that is not true—we'll arrive at a contradiction. Arriving

~~Set relations and operations~~

~~Figure **??** illustrates the notion of a set B that is strictly contained in the set A. We say $B \subset A$ if $\forall b \in B$, we also have $b \in A$, and $\exists a \in A$ such that $a \notin B$. In other words, we write $B \subset A$ whenever the set A contains B, but there exists at least one element in A that is not an element of B. Also illustrated in Figure **??** is the union of two sets $A \cup B$, which includes all the elements of both A and B. If $e \in A \cup B$, then $e \in A$ and/or $e \in B$. The left side of the figure is an illustration of a set B which is strictly contained in another set A, denoted $B \subset A$. The right side of the figure illustrates the union of two sets $A \cup B$.~~

~~The set intersection $A \cap B$ and set minus $A \setminus B$ are illustrated in Figure 1.30. The left side of the figure shows the intersection between the sets $A \cap B$. The intersection of two sets contains the elements that are part of both sets. The right side of the figure shows the set difference $A \setminus B$, which consists of all the elements that are in A but not in B.~~

Sets related to functions

A function that takes real variables as inputs and produces real numbers as outputs is denoted $f : \mathbb{R} \to \mathbb{R}$. The *domain* of a function is the set of all possible inputs to the function that produce an output:

$$\mathrm{Dom}(f) \equiv \{x \in \mathbb{R} \mid f(x) \in \mathbb{R}\}.$$

Inputs for which the function is undefined are not part of the domain. For instance the function $f(x) = \sqrt{x}$ is not defined for negative inputs, so we have $\mathrm{Dom}(f) = \mathbb{R}_+$.

The *image*~~set~~ of a function is the set of all possible outputs of the function:

$$\mathrm{Im}(f) \equiv \{y \in \mathbb{R} \mid \exists x \in \mathbb{R},\ y = f(x)\}.$$

For example, the function $f(x) = x^2$ has the image set ~~$\mathrm{Im}(f) = \mathbb{R}_+$~~ $\mathrm{Im}(f) = \mathbb{R}_+$ since the outputs it produces are always ~~non-negative.~~ nonnegative.

Discussion

Knowledge of the precise mathematical jargon introduced in this section is not crucial to understanding the rest of this book. That being said, I wanted to expose you to it here because this is the language in which mathematicians think. Most advanced math textbooks will assume you understand technical mathematical notation.

Exercises

E1.35 Given the three sets $A = \{1, 2, 3, 4, 5, 6, 7\}$, $B = \{1, 3, 5\}$, and $C = \{2, 4, 6\}$, compute the following set expressions.

a) $A \setminus B$ b) $B \cup C$ c) $A \cap B$ d) $B \cap C$

e) $A \cup B \cup C$ f) $A \setminus (B \cup C)$ g) $(A \setminus B) \cup C$ h) $(A \cap B) \cap C$

1.35 a) $\{2, 4, 6, 7\}$; b) $\{1, 2, 3, 4, 5, 6\}$; c) $\{1, 3, 5\}$; d) $\emptyset$; e) $\{1, 2, 3, 4, 5, 6, 7\}$; f) $\{7\}$; g) $\{2, 4, 6, 7\}$; h) $\emptyset$.

(1) $e^x e^{-x} e^z$ (2) $\left(\frac{xy^{-2}z^{-3}}{x^2y^3z^{-4}}\right)^{-3}$ (3) $(8x^6)^{-\frac{2}{3}}$

(4) $\log_4(\sqrt{2})$ (5) $\log_{10}(0.001)$ (6) $\ln(x^2-1)-\ln(x-1)$

1.22 (1) e^z. (2) $\frac{x^3y^{15}}{z^3}$. (3) $\frac{1}{4x^4}$. (4) $\frac{1}{4}$. (5) -3. (6) $\ln(x+1)$.

P1.23 When representing numbers on a computer, the number of digits of precision n in base b and the approximation error ϵ are related by the equation $n = -\log_b(\epsilon)$. A `float64` has 53 bits of precision (digits base 2). What is the approximation error ϵ for a `float64`? How many digits of precision does a `float64` have in decimal (base 10)?

1.23 $\epsilon = 1.110 \times 10^{-16}$; $n = 15.95$ in decimal.

1.23 See `bit.ly/float64prec` for the calculations.

P1.24 Find the values of x that satisfy the following inequalities:

(1) $2x - 5 > 3$ (2) $5 \leq 3x - 4 \leq 14$ (3) $2x^2 + x \geq 1$

1.24 (1) $\cancel{x \in (4, \infty}x \in (4, \infty)$. (2) $x \in [3, 6]$. (3) $\cancel{x \in -\infty, -1] \cup [\frac{1}{2}, \infty}x \in (-\infty, -1] \cup [\frac{1}{2}$

1.24 For (3), $1\frac{3}{4} + 1\frac{31}{32} = \frac{7}{4} + \frac{63}{32} = \frac{56}{32} + \frac{63}{32} = \frac{119}{32} = 3\frac{23}{32}$.

P1.25 Two algorithms, P and Q, can be used to solve a certain problem. The running time of Algorithm P as a function of the size of the problem n is described by the function $P(n) = 0.002n^2$. The running time of Algorithm Q is described by $Q(n) = 0.5n$. For small problems, Algorithm P runs faster. Starting from what n will Algorithm Q be faster?

1.25 For $n > 250$, Algorithm Q is faster.

1.25 The running time of Algorithm Q grows linearly with the size of the problem, whereas Algorithm P's running time grows quadratically. To find the size of the problem when the algorithms take the same time, we solve $P(n) = Q(n)$, which is $0.002n^2 = 0.5n$. The solution is $n = 250$. For $n > 250$, the linear-time algorithm (Algorithm Q) will take less time.

P1.26 Consider a right-angle triangle in which the shorter sides are 8 cm and 6 cm. What is the length of the triangle's longest side?

1.26 10 cm.

P1.27 A television screen measures 26 inches on the diagonal. The screen height is 13 inches. How wide is the screen?

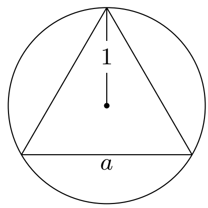
1.27 22.52 in.

P1.28 A ladder of length 3.33 m leans against a wall and its foot is 1.44 m from the wall. What is the height h where the ladder touches the wall?

Hint: Use Pythagoras' theorem twice and ~~the tan function~~ then the function tan.

1.33 $x = \tan\theta\sqrt{a^2 + b^2 + c^2}$.

P1.34 An equilateral triangle is inscribed in a circle of radius 1. Find the side length a and the area of the inscribed triangle $A_\triangle$.



Hint: Split the triangle into three sub-triangles.

1.34 $a = \sqrt{3}$, $A_\triangle = \frac{3\sqrt{3}}{4}$.

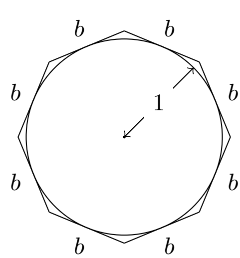
1.34 The internal angles of an equilateral triangle are all $60°$. Draw three radial lines that connect the centre of the circle to each vertex of the triangle. The equilateral triangle is split into three obtuse triangles with angle measures $30°$, $30°$, and $120°$. Split each of these obtuse sub-triangles down the middle to obtain six right-angle triangles with hypotenuse 1. The side of the equilateral triangle is equal to two times the base of the right-angle triangles $a = 2\cos(30°) = \sqrt{3}$. To find the area, we use $A_\triangle = \frac{1}{2}ah$, where $h = 1 + \sin(30°)$.

P1.35 Use the power-reduction trigonometric identities (page 81) to express $\sin^2\theta\cos^2\theta$ in terms of $\cos 4\theta$.

1.35 $\sin^2\theta\cos^2\theta = \frac{1-\cos 4\theta}{8}$.

1.35 We know $\sin^2(\theta) = \frac{1}{2}(1 - \cos(2\theta))$ and $\cos^2(\theta) = \frac{1}{2}(1 + \cos(2\theta))$, so their product is $\frac{1}{4}(1 - \cos(2\theta)\cos(2\theta))$. Note $\cos(2\theta)\cos(2\theta) = \cos^2(2\theta)$. Using the power-reduction formula on the term $\cos^2(2\theta)$ leads to the final answer $\sin^2\theta\cos^2\theta = \frac{1}{4}\left(1 - \frac{1}{2}(1 + \cos(4\theta))\right)$.

P1.36 A circle of radius 1 is inscribed inside a *regular octagon* (a polygon with eight sides of length b). Calculate the octagon's perimeter and its area.



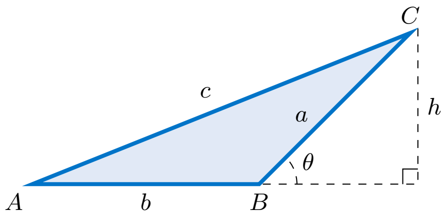
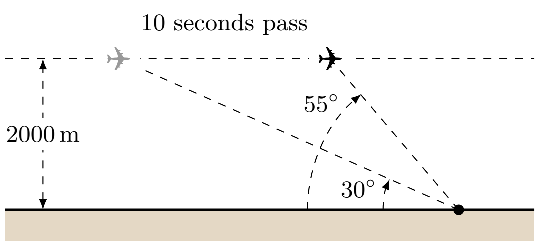


Figure 1.33: A triangle with base b and height h.

Hint: The sum of the internal angle measures of a triangle is $180°$.

1.39 $B = 44.8°$, $C = 110.2°$. $c = \frac{a \sin 110.2°}{\sin 25°} \approx 39.97$.

P1.40 An observer on the ground measures an angle of inclination of $30°$ to an approaching airplane, and 10 seconds later measures an angle of inclination of $55°$. If the airplane is flying at a constant speed at an altitude of $2000\,\text{m}$ in a straight line directly over the observer, find the speed of the airplane in kilometres per hour.



1.40 $v = 742.92\ \text{km/h}$.

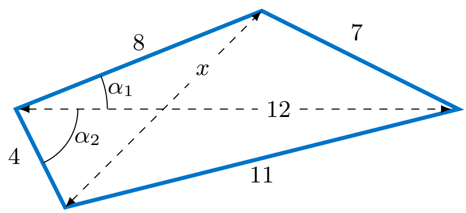
1.40 Initially the horizontal distance between the observer and the plane is $d_1 = \frac{2000}{\tan 30°}$ m. After 10 seconds, the distance is $d_2 = \frac{2000}{\tan 55°}$ m. Velocity is change in distance divided by the time $v = \frac{d_1 - d_2}{10} = 206.36\ \text{m/s}$. To convert m/s into km/h, we must multiply by the appropriate conversion factors: $206.36\,\text{m/s} \times \frac{1\,\text{km}}{1000\,\text{m}} \times \frac{3600\,\text{s}}{1\,\text{h}} = 742.92\ \text{km/h}$.

P1.41 Satoshi likes warm saké. He places 1 litre of water in a sauce pan with diameter 17 cm. How much will the height of the water level rise when Satoshi immerses a saké bottle with diameter 7.5 cm?
Hint: You'll need the volume conversion ratio $1\ \text{litre} = 1000\,\text{cm}^3$.

1.41 ~~1.33~~ 1.06 cm.

1.41 The volume of the water stays constant and is equal to $1000\,\text{cm}^3$. Initially the height of the water h_1 can be obtained from the formula for the volume of a cylinder ~~$1000\,\text{cm}^3 = h_1\pi(17\,\text{cm})^2$, so $h_1 = 5.51$~~ $1000\,\text{cm}^3 = h_1\pi(8.5\,\text{cm})^2$, so $h_1 = 4.41$ cm. After the bottle is inserted, the volume of water has the shape of a cylinder ~~from which~~ with a cylindrical part ~~is missing and $1000\,\text{cm}^3 = h_2\left(\pi(17\,\text{cm})^2 - \pi(7.5\,\text{cm})^2\right)$~~ missing and $1000\,\text{cm}^3 = h_2\left(\pi(8.5\,\text{cm})^2 - \pi(3.7\right.$ We find ~~$h_2 = 6.84$~~ $h_2 = 5.47$ cm. The change in ~~water height is $h_2 - h_1 = 1.33$ cm.~~ height is $h_2 - h_1 = 5.47 - 4.41 = 1.06$ cm.

P1.42 Find the length x of the diagonal of the quadrilateral below.

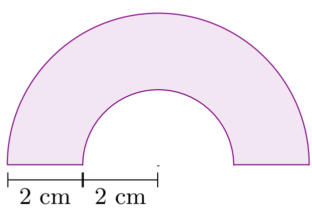


Hint: Use the law of cosines once to find α_1 and α_2, and again to find x.

1.42 $x = 9.55$.

1.42 Using the law of cosines for the angles α_1 and α_2, we obtain the equations $7^2 = 8^2 + 12^2 - 2(8)(12)\cos\alpha_1$ and $11^2 = 4^2 + 12^2 - 2(4)(12)\cos\alpha_2$ from which we find $\alpha_1 = 34.09°$ and $\alpha_2 = 66.03°$. In the last step we use the law of cosines again to obtain $x^2 = 8^2 + 4^2 - 2(8)(4)\cos(34.09° + 66.03°)$.

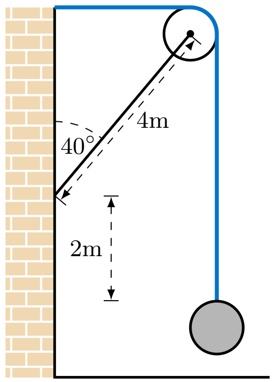
P1.43 Find the area of the shaded region.



Hint: Find the area of the outer circle, subtract the area of missing centre disk, then divide by two.

1.43 $\frac{1}{2}(\pi 4^2 - \pi 2^2) = 18.85\ \text{cm}^2$.

P1.44 In preparation for the shooting of a music video, you're asked to suspend a wrecking ball hanging from a circular pulley. The pulley has a radius of 50 cm. The other lengths are indicated in the figure. What is the total length of the rope required?



Hint: The total length of rope consists of two straight parts and the curved section that wraps around the pulley.

1.44 $\cancel{\ell_{\text{rope}} = 7.83}$ $\ell_{\text{rope}} = 8.42$ m.

1.44 The length of the horizontal part of the rope is $\ell_h = 4\sin 40$. The circular portion of the rope that hugs the pulley has length $\frac{1}{4}$ of the circumference of a circle with radius $r = 50\,\text{cm} = 0.5\,\text{m}$. Using the formula $C = 2\pi r$, we find $\cancel{\ell_c = \frac{1}{4}\pi(0.5)^2 = \frac{\pi}{16}}$ $\ell_c = \frac{1}{4}(2\pi(0.5)) = \frac{\pi}{4}$. The vertical part of the rope has length $\ell_v = 4\cos 40 + 2$. The total length of rope is $\cancel{\ell_h + \ell_c + \ell_v = 7.83}$ $\ell_h + \ell_c + \ell_v = 8.42$ m.

P1.45 The length of a rectangle is $c + 2$ and its height is 5. What is the area of the rectangle?

1.45 $A_{\text{rect}} = 5c + 10$.

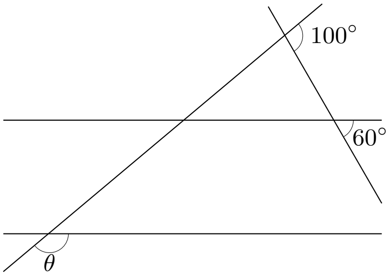
1.45 The rectangle's area is equal to its length times its height $A_{\text{rect}} = \ell h$.

P1.46 A box of facial tissues has dimensions 10.5 cm by 7 cm by 22.3 cm. What is the volume of the box in litres?
Hint: $1\,\text{L} = 1000\,\text{cm}^3$.

1.46 $V_{\text{box}} = 1.639\,\text{L}$.

1.46 The box's volume is $V = w \times h \times \ell = 10.5 \times 7 \times 22.3 = 1639\,\text{cm}^3$=$1.639\,\text{L}$.

P1.47 What is the measure of the angle θ in the figure below?

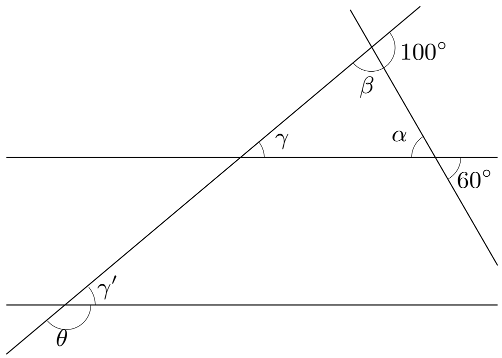


Hint: At the intersection of two lines, vertically opposite angles are equal.

1.47 $\cancel{\theta = 120°}$ $\theta = 140°$.

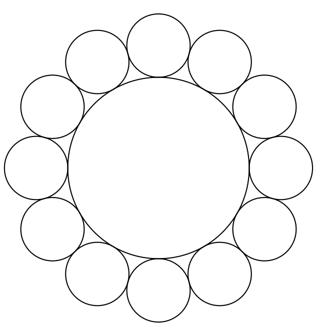
1.47 We didn't really cover these concepts in the book, but since we're on the topic let's define some vocabulary. The *complement* of an acute angle is its defect from a right angle; that is, the angle by which it falls short of a right angle. (i) Two angles are complementary when their sum is 90°. The *supplement* of an angle is its defect from two right angles, that is, the angle by which it falls short of 180°. (ii) Two angles are supplementary when their sum is 180°. Angles that are complementary or supplementary to the same angle are equal to one another.

We'll now use these facts and the diagram below to find the angle θ.



The angle α is vertically opposite to the angle 60° so $\alpha = 60°$. The angle β is supplementary to the angle 100° so $\beta = 180 - 100 = 80°$. The sum of the angles in a triangle is 180° so $\gamma = 180° - \alpha - \beta = 40°$. The two horizontal lines are parallel so the diagonally cutting line makes the same angle with them: $\gamma' = \gamma = 40°$. The angle θ is supplementary to the angle γ' so $\cancel{\theta = 180 - 40 = 120°}$ $\theta = 180 - 40 = 140°$.

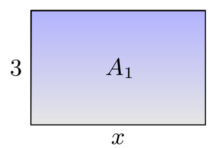
P1.48 A large circle of radius R is surrounded by 12 smaller circles of radius r. Find the ratio $\frac{R}{r}$ rounded to four decimals.

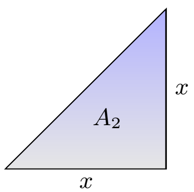


P1.69 When starting a business, one sometimes needs to find investors. Define M to be the set of investors with money, and C to be the set of investors with connections. Describe the following sets in words: (a) $M \setminus C$, (b) $C \setminus M$, and the most desirable set (c) $M \cap C$.

1.69 (a) Investors with money but without connections. (b) Investors with connections but no money. (c) Investors with both money and connections.

P1.70 Write the formulas for the functions $A_1(x)$ and $A_2(x)$ that describe the areas of the following geometrical shapes.





1.70 $A_1(x) = 3x$ and $A_2(x) = \frac{1}{2}x^2$.

Chapter 2

Introduction to physics

2.1 Introduction

One of the coolest things about understanding math is that you will automatically start to understand the laws of physics too. Indeed, most physics laws are expressed as mathematical equations. If you know how to manipulate equations and you know how to solve for the unknowns in them, then you know half of physics already.

Ever since Newton figured out the whole $F = ma$ thing, people have used mechanics to achieve great technological feats, like landing spaceships on the Moon and Mars. You can be part of this science thing too. Learning physics will give you the following superpowers:

1. The power to **predict the future motion of objects** using equations. For most types of motion, it is possible to find an equation that describes the position of an object as a function of time $x(t)$. You can use this equation to predict the position of the object at all times t, including the future. "Yo G! Where's the particle going to be at $t = 1.3$ seconds?" you are asked. "It is going to be at $x(1.3)$ metres, bro." Simple as that. The equation $x(t)$ describes the object's position for *all* times t during the motion. Knowing this, you can plug $t = 1.3$ seconds into $x(t)$ to find the object's location at that time.

2. Special **physics vision** for seeing the world. After learning physics, you will start to think in ~~term~~ terms of concepts like force, acceleration, and velocity. You can use these concepts to precisely describe all aspects of the motion of objects. Without physics vision, when you throw a ball into the air you will see it go up, reach the top, then fall down. Not very exciting. Now *with* physics vision, you will see that at $t = 0$[s], the same ball is

Concepts

The key notions for describing the motion of objects are:

- t: the time. Time is measured in seconds [s].
- $x(t)$: an object's position as a function of time—also known as the equation of motion. Position is measured in metres [m] and depends on the time t.
- $v(t)$: the object's velocity as a function of time. Velocity is measured in metres per second [m/s].
- $a(t)$: the object's acceleration as a function of time. Acceleration is measured in metres per second squared [m/s^2].
- $x_i = x(0), v_i = v(0)$: the object's initial position and velocity, as measured at $t = 0$. Together x_i and v_i are known as the *initial conditions*.

Position, velocity, and acceleration

The motion of an object is characterized by three functions: the position function $x(t)$, the velocity function $v(t)$, and the acceleration function $a(t)$. The functions $x(t)$, $v(t)$, and $a(t)$ are connected—they all describe different aspects of the same motion.

You are already familiar with these notions from your experience of riding in a car. The equation of motion $x(t)$ describes the position of the car as a function of time. The velocity describes the change in the position of the car, or mathematically,

$$v(t) \equiv \text{rate of change in } x(t).$$

If we measure x in metres [m] and time t in seconds [s], then the units of $v(t)$ will be metres per second [m/s]. For example, an object moving with a constant velocity of +30[m/s] will increase its position by 30[m] each second. Note that the velocity $v(t)$ could be positive or negative. The *speed* of an object is defined as the absolute value of its velocity $|v(t)|$. The speed is always a nonnegative quantity, whereas velocity is positive or negative depending on the direction of motion.

The rate of change of an object's velocity is called *acceleration*:

$$a(t) \equiv \text{rate of change in } v(t).$$

Acceleration is measured in metres per second squared [m/s^2]. A constant positive acceleration means the velocity of the motion is steadily increasing, similar to pressing the gas pedal. A constant negative acceleration means the velocity is steadily decreasing, similar to pressing the brake pedal.

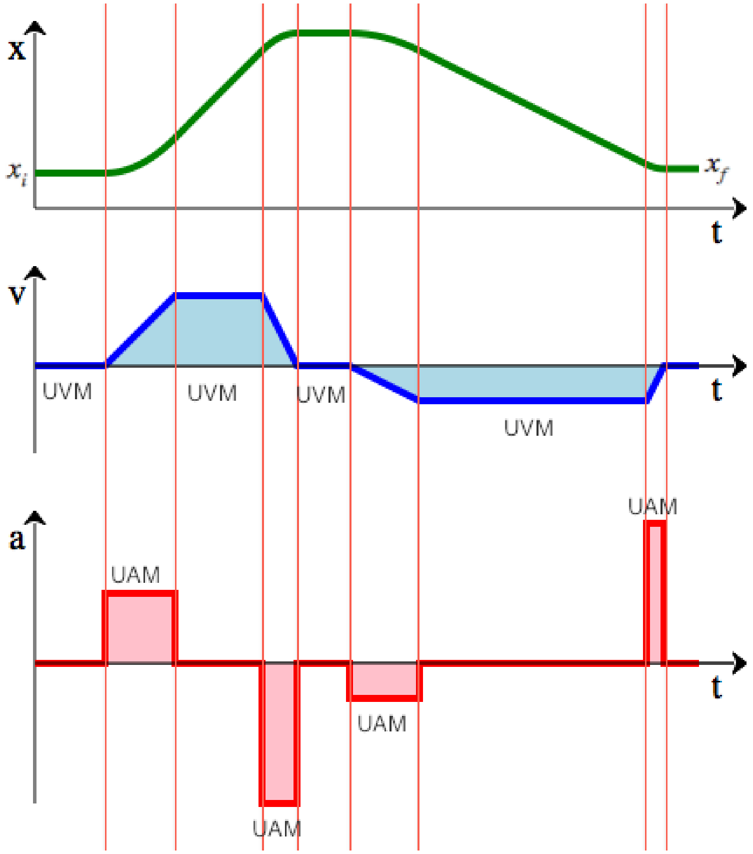


Figure 2.2: The illustration shows the simultaneous graphs of the position, velocity, and acceleration of a car during some time interval. The car starts from an initial position x_i where it sits still for some time. The driver then floors the pedal to produce a maximum acceleration for some time, and the car picks up speed. The driver then eases off the accelerator, keeping it pressed enough to maintain a constant speed. Suddenly the driver sees a police vehicle in the distance and slams on the brakes (negative acceleration) and shortly afterward brings the car to a stop. The driver waits for a few seconds to make sure the cops have passed. Next, the driver switches into reverse gear and adds gas. The car accelerates backward for a bit, then maintains a constant backward speed for an extended period of time. Note how “moving backward” corresponds to negative velocity. In the end the driver slams on the brakes again to stop the car. Notice that braking corresponds to positive acceleration when the motion is in the negative direction. The car’s final position is x_f.

In a couple of paragraphs, we'll discuss the exact mathematical equations for $x(t)$, $v(t)$, and $a(t)$, but before we dig into the math, let's look at the example of the motion of a car illustrated in Figure 2.2. We can observe two distinct types of motion~~in the situation described in Figure 2.2~~. During some times, the car undergoes motion at a constant velocity (uniform velocity motion, UVM). During other times, the car undergoes motion with constant acceleration (uniform acceleration motion, UAM). There exist many other types of motion, but for the purpose of this section we'll focus on these two types of motion.

- UVM: During times when there is no acceleration, the car maintains a uniform velocity and therefore $v(t)$ is a constant function. For motion with constant velocity, the position function is a line with a constant slope because, by definition, $v(t) =$ slope of $x(t)$.
- UAM: During times where the car experiences a constant acceleration $a(t) = a$, the velocity of the function changes at a constant rate. The rate of change of the velocity is constant $a =$ slope of $v(t)$, so the velocity function looks like a line with slope a. The position function $x(t)$ has a curved shape (quadratic) during moments of constant acceleration.

Formulas

There are basically four equations you need to know for this entire section. Together, these four equations fully describe all aspects of motion with constant acceleration.

Uniformly accelerated motion (UAM)

If the object undergoes a *constant* acceleration $a(t) = a$—like a car when you floor the *accelerator*—then its motion can be described by the following equations:

$$a(t) = a, \tag{2.1}$$
$$v(t) = at + v_i, \tag{2.2}$$
$$x(t) = \tfrac{1}{2}at^2 + v_i t + x_i, \tag{2.3}$$

where v_i is the initial velocity of the object and x_i is its initial position.

Here is another useful equation to remember:

$$[v(t)]^2 = v_i^2 + 2a[x(t) - x_i],$$

which is usually written

$$v_f^2 = v_i^2 + 2a\Delta x, \tag{2.4}$$

We start by writing the general UAM equation:

$$y(t) = \tfrac{1}{2}a_y t^2 + v_i t + y_i.$$

To find the time when the ball will hit the ground, we ~~must~~ can solve for t in the equation $y(t) = 0$~~. Plug all the known~~ , then plug the values into the UAM equation,

$$y(t) = 0 = \tfrac{1}{2}(-9.81)t^2 - 10t + 44.145,$$

and solve for t using the quadratic formula. First, rewrite the quadratic equation in standard form:

$$0 = \underbrace{4.905}_{a}\,t^2 + \underbrace{10.0}_{b}\,t \quad \underbrace{-44.145}_{c}.$$

$0 = \underbrace{4.905}_{a}\,t^2 + \underbrace{10.0}_{b}\,t \quad \underbrace{-44.145}_{c}.$ Then solve using the quadratic equation:

$$t_{\text{fall}} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-10 \pm \sqrt{100 + 866.12}}{9.81} = 2.15 \qquad [\text{s}].$$

We ignore the negative-time solution because it corresponds to a time in the past. Compared to the first Moroccan example, we see that throwing the ball downward makes it fall to the ground ~~faster.~~ in less time.

Discussion

Most kinematics problems you'll be solving follow the same pattern as the examples above. Given some initial values, you'll be asked to solve for some unknown quantity.

It's important to keep in mind the *signs* of the numbers you plug into the equations. You should always draw the coordinate system and indicate clearly (to yourself) the x-axis, which measures the object's displacement. A velocity or acceleration quantity that points in the same direction as the x-axis is a positive number, while quantities pointing in the opposite direction are negative numbers.

By the way, all this talk about $v(t)$ being the "rate of change of $x(t)$" is starting to get on my nerves. The expression "rate of change of" is an indirect way of saying the calculus term *derivative*. In order to use this more precise terminology throughout the remainder of the book, we'll now take a short excursion into the land of calculus to define two fundamental concepts: derivatives and integrals.

Exercises

E2.1 Calculate the time it will take a rocket launched with initial velocity v_i[m/s] and constant acceleration of a[m/s^2] to reach the velocity of 100[m/s]. What is the total distance travelled?

a) $v_i = 20$[m/s], $a = 5$[m/s^2] b) $v_i = 30$[m/s], $a = 10$[m/s^2]

2.1 a) $t_f = 20$[s], $d = 1400$[m]; b) $t_f = 10$[s], $d = 800$[m].

2.1 We can find t_f from the UAM velocity equation $v_f = 100 = 5(t_f) + v_i$. We can then compute the total distance travelled using the UAM position equation: $d = x(t_f) = 0 + v_i t_f + \frac{1}{2}at_f^2$.

2.3 Introduction to calculus

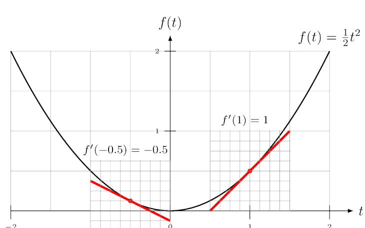
Calculus is the study of functions and their properties. The two operations in the study of calculus are derivatives—which describe how quantities *change over time*—and integrals, which are used to calculate the total amount of a quantity *accumulated* over a time period.

Derivatives

The derivative function $f'(t)$ describes how the function $f(t)$ changes over time. The derivative encodes the information about the *instantaneous rate of change* of the function $f(t)$, which is the same as the *slope* of the graph of the function at that point:

$$f'(t) \equiv \text{slope}_f(t) = \frac{\text{change in } f(t)}{\text{change in } t} = \frac{f(t+\Delta t) - f(t)}{\Delta t}.$$

If the derivative $f'(t)$ is equal to 5 units per second, this means that $f(t)$ changes by 5 units each second. The derivative of the constant function is zero because it has zero rise-over-run everywhere. The derivative of the function $f(t) = mt + b$ (a line) is the constant function $f'(t) = m$. More generally, the instantaneous slope of a function is different for different values of t, as illustrated in the figure.

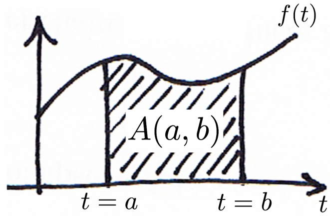


The derivative operation is denoted by several names and symbols: $Df(t) = f'(t) = \frac{df}{dt} = \frac{d}{dt}\{f(t)\} = \dot{f}$~~and all~~. All these symbols carry the same meaning. Think of $f'(t)$ not as a separate entity from $f(t)$, but as a *property* of the function $f(t)$. It's best to think of the derivative as an *operator* $\frac{d}{dt}$ ~~that you can apply to any function to obtain its~~ you apply to functions to obtain their slope information.

Integrals

An integral corresponds to the computation of the *area* enclosed between the curve $f(t)$ and the $\cancel{x}t$-axis over some interval:

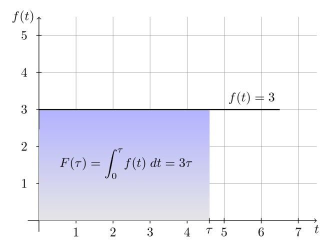
$$A(a,b) \equiv \int_{t=a}^{t=b} f(t)\;dt.$$



The symbol $\int$ is shorthand for *sum*.
Indeed, the area under the curve corresponds to the sum of the values of the function $f(t)$ between $t=a$ and $t=b$. The integral is the total of f between a and b.

Example 1

We can easily find the area under the graph of the constant function $f(t) = 3$ between any two points because the region under the curve is rectangular. Choosing $t = 0$ as the starting point, we obtain the integral function $F(\tau)$, which corresponds to the area under $f(t)$ between $t = 0$ and $t = \tau$:



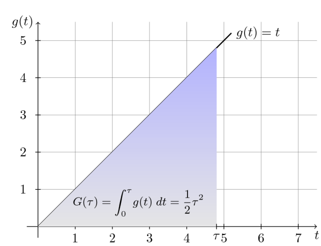
$$F(\tau) \equiv A(0,\tau) = \int_0^\tau f(t)\;dt = 3\tau.$$

The area is equal to the rectangle's height times its width.

Example 2

Consider now the area under the graph of the line $g(t) = t$, starting from $t = 0$. Since the region under the curve is triangular, we can compute its area. Recall the area of a triangle is given by the length of its base times its height divided by 2.

The general formula for the area under $g(t)$ from $t = 0$ until $t = \tau$ is described by the following integral calculation:



$$G(\tau) \equiv A(0,\tau) = \int_0^\tau g(t)\;dt = \frac{\tau \times \tau}{2} = \tfrac{1}{2}\tau^2.$$

We're able to compute the above integrals thanks to the simple geometries of the areas under the graphs. Later in this book (Chapter 5), we'll develop techniques for finding integrals (areas under the curve) of more complicated functions. In fact, there is an entire course called integral calculus, which is dedicated to the task of finding integrals.

But don't worry, you don't need to know *everything* about integrals to learn physics. What is important right now is that you understand the concept of integration. The integral of a function is the area under the graph of the function, which is in some sense the total amount of the function accumulated during some interval of time. For the most part of the first-year physics, the only integral formulas you'll need to know are

$$\int_0^\tau a\,dt = a\tau \qquad \text{and} \qquad \int_0^\tau at\,dt = \tfrac{1}{2}a\tau^2.$$

The first integral describes the general calculation of the area under a constant function, like in Example 1. The second formula is a generalization of the formula we derived in Example 2. Using these formulas in combination, you can compute the integral of an arbitrary line $h(t) = mt + b$ as follows:

$$H(\tau) = \int_0^\tau h(t)\,dt = \int_0^\tau (mt+b)\,dt = \int_0^\tau mt\,dt + \int_0^\tau b\,dt = \tfrac{1}{2}m\tau^2 + b\tau.$$

Regroup

At this point you might be on the fence about the new calculus concepts. On one hand, calculating slopes (derivatives) and areas under the curve (integrals) seem like trivial tasks. On the other hand, seeing five different notations for the derivative and the weird integral sign has probably put some fear in you. You might be wondering whether you *really* need to learn about derivatives and integrals. How often do you have to compute the area under the graph of a function in the real world? It turns out that "calculating the area under a curve" is very useful since it is the "undo operation" for the derivative.

Inverse operations

The integral is the inverse operation of the derivative. Many equations in math and physics involve the *derivative* of some unknown function. Understanding the inverse relationship between integrals and derivatives will allow you to solve for the unknown function in these equations.

You should already be familiar with the inverse relationship between functions. When solving equations, we use inverse functions

Concepts

Recall the kinematics concepts related to the motion of objects:

- t: time
- $x(t)$: position as a function of time
- $v(t)$: velocity as a function of time
- $a(t)$: acceleration as a function of time
- $\cancel{x_i = x(0), v_i = v(0)}$$x_i \equiv x(0), v_i \equiv v(0)$: the initial conditions

Position, velocity, and acceleration revisited

The equations of kinematics are used to predict the motion of objects. Suppose you know the acceleration of the object $a(t)$ at all times t. Can you find $x(t)$ starting from $a(t)$?

The equations of motion $x(t)$, $v(t)$, and $a(t)$ are related:

$$a(t) \xleftarrow{\frac{d}{dt}} v(t) \xleftarrow{\frac{d}{dt}} x(t).$$

The velocity function is the derivative of the position function and the acceleration function is the derivative of the velocity function.

General procedure

If you know the acceleration of an object as a function of time $a(t)$, and you know its initial velocity $v_i = v(0)$, you can find its velocity function $v(t)$ for all later times using integration. This is because the acceleration function $a(t)$ describes the change in the object's velocity. ~~If you know the~~

Let's see how this works. The object started with an initial velocity of $v_i \equiv v(0)$ ~~, the~~ at $t = 0$. The velocity at a later time $t = \tau$ is equal to v_i plus the total acceleration of the object between $t = 0$ and $t = \tau$:

$$v(\tau) = v_i + \int_0^\tau a(t)\, dt.$$

If you know the initial position x_i and the velocity function $v(t)$, you can find the position function $x(t)$ by using integration. We find the position at time $t = \tau$ by adding all the velocities (the changes in the object's position) that occurred between $t = 0$ and $t = \tau$:

$$x(\tau) = x_i + \int_0^\tau v(t)\, dt.$$

The procedure for finding $x(t)$ starting from $a(t)$ can be summarized as follows:

$$a(t) \quad \xrightarrow{v_i + \int dt} \quad v(t) \quad \xrightarrow{x_i + \int dt} \quad x(t).$$

Applications of derivatives

Recall that the velocity and the acceleration functions are obtained by taking derivatives of the position function:

$$x(t) \xrightarrow{\frac{d}{dt}} v(t) \xrightarrow{\frac{d}{dt}} a(t).$$

We just saw how to use integration to follow this chain of operations in reverse to obtain $x(t)$ for the special case of constant acceleration:

$$\begin{aligned} a(t) &\equiv a, \\ v(t) &\equiv v_i + \int_0^t a(\tau)\, d\tau = v_i + at, \\ x(t) &\equiv x_i + \int_0^t v(\tau)\, d\tau = x_i + v_i t + \tfrac{1}{2}at^2. \end{aligned}$$

Note that, in addition to the integral calculations, the formulas for $v(t)$ and $x(t)$ require some additional information—the initial value of the function.

Earlier we defined the derivative operator $\frac{d}{dt}$ that computes the derivative function $f'(t)$, which tells us the slope of the function $f(t)$. There are several derivative formulas that you need to learn to be proficient at calculus. We'll get to them in Chapter 5. For now, the only derivative formula you'll need is the *power rule* for derivatives:

$$\text{if} \quad f(t) = At^n \quad \text{then} \quad f'(t) = nAt^{n-1}.$$

Using this formula on each term in the function $f(t) = A + Bt + Ct^2$ we find its derivative is $\frac{df}{dt} \equiv f'(t) = 0 + B + 2Ct$.

Let's now use the derivative to verify that the equations of motion we obtained above satisfy ~~$x'(t) = v(t)$ and~~ $v'(t) = a(t)$ and $x'(t) = v(t)$. Applying the derivative operation to both sides of the equations we obtain

$$\begin{aligned} a'(t) &\equiv 0, \\ v'(t) &\equiv \tfrac{d}{dt}\{v_i + at\} = \cancel{\tfrac{d}{dt}\{v_i\}} + \tfrac{d}{dt}\{at\} = 0 + a &&= a(t), \\ x'(t) &\equiv \cancel{\tfrac{d}{dt}\{x_i\}} + \tfrac{d}{dt}\{v_i t\} + \tfrac{d}{dt}\{\tfrac{1}{2}at^2\} = 0 + v_i + at &&= v(t). \end{aligned}$$

Note that computing the derivative of a function kills the information about its initial value; the derivative contains only information about the changes in $f(t)$.

Let's summarize what we've learned so far about derivatives and integrals. Integrals are useful because they allow us to compute $v(t)$

2. Calculate the velocity of the car at $t = 4$[s].
3. What is the car's velocity when it is 9[m] away from point A?

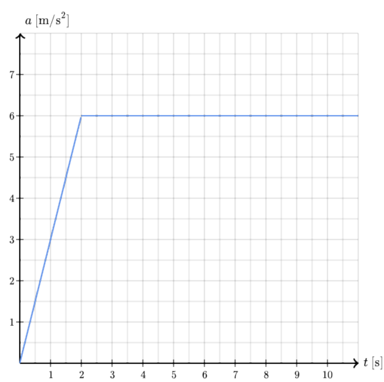
Hint: Use Newton's 2$^{\text{nd}}$ law and integration.

2.9 (1) $a(t) = 2[\text{m/s}^2]$, $v(t) = 2t[\text{m/s}]$, and $x(t) = t^2[\text{m}]$. (2) $v(4) = 8[\text{m/s}]$. (3) $\cancel{v = 10}$$v = 6$[m/s] when $x = 9$[m].

2.9 (1) Calculate $a(t)$ from Newton's 2$^{\text{nd}}$ law to obtain $a(t) = 2[\text{m/s}^2]$. Integrate with respect to time to obtain the velocity and then integrate again to find the position. (2) Plug $t = 4$[s] into $v(t)$. (3) Use the fourth equation to find the speed after the car has travelled 9[m].

P2.10 Below is an acceleration-vs-time graph of a particle. At $t = 0$[s], the particle starts moving from rest at $x = 0$[m]. The particle's acceleration from $t = 0$[s] to $t = 3$[s] is given by $a(t) = 3t[\text{m/s}^2]$. After $t = 2$[s], the acceleration is constant $a = 6[\text{m/s}^2]$.

1. Find the velocity $v(2)$ and position $x(2)$ of the particle at $t = 2$[s].
2. Construct the functions of time that describe the acceleration, the velocity, and the position of the particle after $t = 2$[s].
3. How much time is needed for the particle to reach $x = 49$[m]?
4. At what distance from the origin will the particle's velocity reach 12[m/s]?



Hint: Use integration to find the velocity and the position. The integral of $f(t) = t^2$ is $F(t) = \frac{1}{3}t^3$. Make sure that when $t = 2$[s], the functions $v(t)$ and $x(t)$ in Part 2 match your answer from Part 1.

2.10 (1) $v(2) = 6$[m/s], $x(2) = 4$[m]. (2) For $t > 2$[s]: $v(t) = 6 + 6(t - 2)$[m/s], $x(t) = 4 + 6(t - 2) + 3(t - 2)^2$[m]. (3) When $x = 49$[m], $t = 5$[s]. (4) $v = 12$[m/s] when $\cancel{x = 14}$$x = 13$[m].

Okay, back to vectors. In this case, the *directions* can be also written as a vector $\vec{d}$, expressed as:

$$\vec{d} = 2\text{km}\,\hat{N} + 3\text{km}\,\hat{E}.$$

This is the mathematical expression that corresponds to the directions "walk 2 km north then 3 km east." Here, $\hat{N}$ is a *direction* and the number in front of the direction tells you the distance to walk in that direction.

Act 2: Equivalent directions

Later during your vacation, you decide to return to the location X. You arrive at the bus stop to find there is a slight problem. From your position, you can see a kilometre to the north, where a group of armed and threatening-looking men stand, waiting to ambush anyone who tries to cross what has now become a trail through the crops. Clearly the word has spread about X and constant visitors have drawn too much attention to the location.

Well, technically speaking, there is no problem at X. The problem lies on the route that starts north and travels through the ambush squad. Can you find an alternate route that leads to X?

`"Use math, Luke! Use math!"`

Recall the commutative property of ~~addition for numbers~~number addition: $a + b = b + a$. Maybe an analogous property holds for vectors? Indeed, this ~~is the case~~it does:

$$\vec{d} = 2\text{km}\,\hat{N} + 3\text{km}\,\hat{E} = 3\text{km}\,\hat{E} + 2\text{km}\,\hat{N}.$$

The $\hat{N}$ directions and the $\hat{E}$ directions obey the commutative property. Since the directions can be followed in any order, you can first walk the 3 km east, then walk 2 km north and arrive at X again.

Act 3: Efficiency

It takes 5 km of walking to travel from the bus stop to X, and another 5 km to travel back to the bus stop. Thus, it takes a total of 10 km walking every time you want to go to X. Can you find a quicker route? What is the fastest way from the bus stop to the destination?

Instead of walking in the east and north directions, it would be quicker if you take the diagonal to the destination. Using Pythagoras' theorem you can calculate the length of the diagonal. When the side lengths are 3 and 2, the diagonal has length $\sqrt{3^2 + 2^2} = \sqrt{9 + 4} =$

$\sqrt{13} = 3.60555\ldots$. The length of the diagonal route is just 3.6 km, which means the diagonal route saves you a whole 1.4 km of walking in each direction.

But perhaps seeking efficiency is not always necessary! You could take a longer path on the way back and give yourself time to enjoy the great outdoors.

Discussion

Vectors are directions for getting from one point to another point. To indicate directions on maps, we use the four cardinal directions: $\hat{N}$, $\hat{S}$, $\hat{E}$, $\hat{W}$. In math, however, we will use only two of the cardinals—$\hat{E} \equiv \hat{x}$ and $\hat{N} \equiv \hat{y}$—since they fit nicely with the usual way of drawing the Cartesian plane. We don't need an $\hat{S}$ direction because we can represent downward distances as negative distances in the $\hat{N}$ direction. Similarly, $\hat{W}$ is the same as negative $\hat{E}$.

From now on, when we talk about vectors we will always represent them with respect to the standard coordinate system $\hat{x}$ and $\hat{y}$, and use *bracket notation*,

$$(v_x, v_y) \equiv v_x\,\hat{x} \ + \ v_y\,\hat{y}.$$

Bracket notation is nice because it's compact, which is good since we will be doing a lot of calculations with vectors. Instead of explicitly writing out all the directions, we will automatically assume that the first number in the bracket is the $\hat{x}$ distance and the second number is the $\hat{y}$ distance.

3.2 Vectors

Vectors are extremely useful in all areas of life. In physics, for example, we use a vector to describe the velocity of an object. It is not sufficient to say that the speed of a tennis ball is 20[m/s]: we must also specify the direction in which the ball is moving. Both of the two velocities

$$\vec{v}_1 = (20, 0) \qquad \text{and} \qquad \vec{v}_2 = (0, 20)$$

describe motion at the speed of 20[m/s]; but since one velocity points along the x-axis, and the other points along the y-axis, they are *completely* different velocities. The velocity vector contains information about the object's speed *and* direction. The direction makes a big difference. If it turns out that the tennis ball is ~~coming your way, you need to~~ hurtling toward you, you better get out of the way!

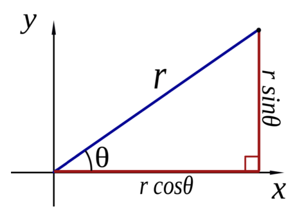
This section's main idea is that **vectors are not the same as numbers**. A vector is a special kind of mathematical object that is of $\alpha = 0.5$. The scaled-down vector is $\vec{w} = 0.5\vec{v} = (1.5, 1)$.

Conversely, we can think of the vector $\vec{v}$ as being twice as long as the vector $\vec{w}$.

Length-and-direction representation

So far, we've seen how to represent a vector in terms of its components. There is also another way of representing vectors: we can specify a vector in terms of its length $||\vec{v}||$ and its direction—the angle it makes with the x-axis. For example, the vector $(1, 1)$ can also be written as $\sqrt{2}\angle 45°$. This magnitude-and-direction notation is useful because it makes it easy to see the "size" of vectors. On the other hand, vector arithmetic operations are much easier to carry out in the component notation. We will use the following formulas for converting between the two notations.

To convert the length-and-direction vector $\|\vec{r}\|\angle\theta$ into an x-component and a y-component (r_x, r_y), use the formulas



$$r_x = \|\vec{r}\|\cos\theta \quad \text{and} \quad r_y = \|\vec{r}\|\sin\theta.$$

To convert from component notation (r_x, r_y) to length-and-direction $\|\vec{r}\|\angle\theta$, use

$$r = \|\vec{r}\| = \sqrt{r_x^2 + r_y^2} \qquad \text{and} \qquad \theta = \tan^{-1}\left(\frac{r_y}{r_x}\right).$$

Note that the second part of the equation involves the inverse tangent function. By convention, the function $\tan^{-1}$ returns values between $\pi/2$ (90°) and $-\pi/2$ (−90°). You must be careful when finding the θ of vectors with an angle outside of this range. Specifically, for vectors with $v_x < 0$, you must add π (180°) to $\tan^{-1}(r_y/r_x)$ to obtain the correct θ.

Unit vector notation

~~As discussed above~~In three dimensions, we can think of a vector $\vec{v} = (v_x, v_y, v_z)$ as a command to "go a distance v_x in the x-direction, a distance v_y in the y-direction, and v_z in the z-direction."

To write this set of commands more explicitly, we can use multiples of the vectors $\hat{\imath}, \hat{\jmath}$, and $\hat{k}$. These are the unit vectors pointing in the x, y, and z directions, respectively:

$$\hat{\imath} = (1, 0, 0), \qquad \hat{\jmath} = (0, 1, 0), \quad \text{and} \quad \hat{k} = (0, 0, 1).$$

If the boat travels directly upstream at full throttle $12\hat{\imath}$, then the speed of the boat relative to the shore will be

$$12\hat{\imath} - 5\hat{\imath} = 7\hat{\imath},$$

since we have to "deduct" the speed of the current from the speed of the boat relative to the water.

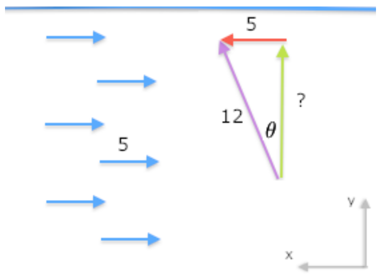


Figure 3.2: Part of the boat's thrust must be used to cancel the current.

If the crew wants to cross the river perpendicular to the current flow, they can use some of the boat's thrust to counterbalance the current, and the remaining thrust to push across. The situation is illustrated in Figure 3.2. In what direction should the boat sail to cross the river? We are looking for the direction of $\vec{v}$ the boat should take such that, after adding in the velocity of the current, the boat moves in a straight line between the two banks (in the $\hat{\jmath}$ direction).

~~A picture is necessary: draw a river, then draw a triangle in the river with its long leg perpendicular~~ Let's analyze the vector diagram. The opposite side of the triangle is parallel to the current flow ~~. Make the short leg of~~ and has length 5. We ~~will~~ take the up-the-river component of the speed $\vec{v}$ to be equal to $5\hat{\imath}$, so that it cancels exactly the $-5\hat{\imath}$ flow of the river. ~~Finally, label the hypotenuse with~~ The hypotenuse has length 12 ~~,~~ since this is the speed of the boat relative to the surface of the water.

From all this we can answer the question like professionals. You want the angle? Well, we have that $\cancel{12\sin(\theta) = 5}\, \frac{\text{opp}}{\text{hyp}} = \frac{5}{12} = \sin(\theta)$, where θ is the angle of the boat's course relative to the straight line between the two banks. We can use the inverse-sin function to solve for the angle:

$$\theta = \sin^{-1}\left(\frac{5}{12}\right) = 24.62°.$$

The across-the-river component of the velocity can be calculated using $v_y = 12\cos(\theta) = 10.91$, or from Pythagoras' theorem if you prefer $v_y = \sqrt{\|\vec{v}\|^2 - v_x^2} = \sqrt{12^2 - 5^2} = 10.91$.

Vector dimensions

The most common types of vectors are two-dimensional vectors (like the ones in the Cartesian plane), and three-dimensional vectors (directions in 3D space). 2D and 3D vectors are easier to work with because we can visualize them and draw them in diagrams. In general, vectors can exist in any number of dimensions. An example of a n-dimensional vector is

$$\vec{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n.$$

The rules of vector algebra apply in higher dimensions, but our ability to visualize stops at three dimensions.

Coordinate system

~~The geometrical interpretation of vectors depends~~ Vector components depend on the coordinate system in which the vectors are represented. Throughout this section we ~~have~~ used the x, y, and z axes ~~, and we've~~ as the coordinate system, and we described vectors as components along each of these ~~directions~~axes. This is a very convenient coordinate system; we have a set of three *perpendicular* axes, and a set of three unit vectors $\{\hat{\imath}, \hat{\jmath}, \hat{k}\}$ that point along each of the three axis directions. Every vector is implicitly defined in terms of this coordinate system. When you and I talk about the vector $\vec{v} = 3\hat{\imath} + 4\hat{\jmath} + 2\hat{k}$, we are really saying, "start from the origin $(0, 0, 0)$, move 3 units in the x-direction, then move 4 units in the y-direction, and finally move 2 units in the z-direction." It is simpler to express these directions as $\vec{v} = (3, 4, 2)$, while remembering that the numbers in the bracket measure distances *relative* to the xyz-coordinate system.

It turns out, using the xyz-coordinate system and the vectors $\{\hat{\imath}, \hat{\jmath}, \hat{k}\}$ is just one of many possible ways we can represent vectors. We can represent a vector $\vec{v}$ as coefficients (v_1, v_2, v_3) with respect to any *basis* $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ as follows: $\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2 + v_3\hat{e}_3$. What is a basis, you ask? I'm glad you asked, because this is the subject of the next section.

3.3 Basis

One of the most important concepts in the study of vectors is the concept of a ~~basis~~*basis*. Consider the ~~space of~~ three-dimensional ~~vectors~~ vector space $\mathbb{R}^3$. A *basis* for $\mathbb{R}^3$ is a set of vectors $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ ~~which~~ that can be used as a coordinate system for $\mathbb{R}^3$. If the set of vectors $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ is a basis, then you can *represent* any vector $\vec{v} \in \mathbb{R}^3$ as coefficients (v_1, v_2, v_3) *with respect to* that basis:

$$\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2 + v_3\hat{e}_3.$$

The vector $\vec{v}$ is obtained by measuring out a distance v_1 in the $\hat{e}_1$ direction, a distance v_2 in the $\hat{e}_2$ direction, and a distance v_3 in the $\hat{e}_3$ direction.

You are already familiar with the *standard* basis $\{\hat{\imath}, \hat{\jmath}, \hat{k}\}$, which is associated with the xyz-coordinate system. You know that any vector $\vec{v} \in \mathbb{R}^3$ can be expressed as a ~~triplet~~ triple (v_x, v_y, v_z) with respect to the basis $\{\hat{\imath}, \hat{\jmath}, \hat{k}\}$ through the formula $\vec{v} = v_x\hat{\imath} + v_y\hat{\jmath} + v_z\hat{k}$. In this section, we'll discuss how to represent vectors with respect to other bases.

An analogy

Let's start with a simple example of a basis. If you look at the HTML source code behind any web page, you're sure to find at least one mention of the colour stylesheet directive such as `background-color:#336699;`. The numbers should be interpreted as a ~~triplet~~ triple of values $(33, 66, 99)$, each value describing the amount of red, green, and blue needed to create a given colour. Let us call the colour described by the ~~triplet~~ triple $(33, 66, 99)$ CoolBlue. This convention for colour representation is called the RGB colour model and we can think of it as the *RGB basis*. A basis is a set of elements that can be combined together to express something more complicated. In our case, the **R**, **G**, and **B** elements are pure colours that can create any colour when mixed appropriately. Schematically, we can write this mixing idea as

$$\text{CoolBlue} = (33, 66, 99)_{RGB} = 33\mathbf{R} + 66\mathbf{G} + 99\mathbf{B},$$

where the *coefficients* determine the strength of each colour component. To create the colour, we combine its components as symbolized by the $+$ operation.

The cyan, magenta, and yellow (CMY) colour model is another basis for representing colours. To express the "cool blue" colour in the CMY basis, you will need the following coefficients:

$$(33, 66, 99)_{RGB} = \text{CoolBlue} = (222, 189, 156)_{CMY} = 222\mathbf{C} + 189\mathbf{M} + 156\mathbf{Y}.$$

The *same* colour CoolBlue is represented by a *different* set of coefficients when the CMY colour basis is used.

Note that a ~~triplet~~ triple of coefficients by itself does not mean anything unless we know the basis being used. For example, if we were to interpret the ~~triplet~~ triple of coordinates $(33, 66, 99)$ with respect to the CMY basis, will would obtain a completely different colour, which would not be cool at all.

A basis is required to convert mathematical objects like the ~~triplet~~ triple (a, b, c) into real-world ideas like colours. As exemplified above, to avoid any ambiguity we can use a subscript after the bracket to indicate the basis associated with each ~~triplet~~ triple of coefficients.

Discussion

It's hard to over-emphasize the importance of the basis—the coordinate system you will use to describe vectors. The choice of coordinate system is the bridge between real-world vector quantities and their mathematical representation in terms of components. Every time you solve a problem with vectors, **the first thing you should do is draw a coordinate system**, and think of vector components as measuring out a distance along this coordinate system.

3.4 Vector products

If addition of two vectors $\vec{v}$ and $\vec{w}$ corresponds to the addition of their components $(v_x + w_x, v_y + w_y, v_z + w_z)$, you might logically think that the product of two vectors will correspond to the product of their components $(v_x w_x, v_y w_y, v_z w_z)$, however, this way of multiplying vectors is not used in practice. Instead, we use the dot product and the cross product.

The *dot product* tells you how similar two vectors are to each other:

$$\vec{v} \cdot \vec{w} \equiv v_x w_x + v_y w_y + v_z w_z \equiv \|\vec{v}\| \|\vec{w}\| \cos(\varphi) \quad \in \mathbb{R},$$

where φ is the angle between the two vectors. The factor $\cos(\varphi)$ is largest when the two vectors point in the same direction because the angle between them will be $\varphi = 0$ and $\cos(0) = 1$.

The exact formula for the *cross product* is more complicated so I will not show it to you just yet. What is important to know is that the cross product of two vectors is another vector:

$$\vec{v} \times \vec{w} = \{ \text{ a vector perpendicular to both } \vec{v} \text{ and } \vec{w} \ \} \quad \in \mathbb{R}^3.$$

If you take the cross product of one vector pointing in the x-direction with another vector pointing in the y-direction, the result will be a vector in the z-direction.

The polar representation of complex numbers:

- $z = |z|\angle\varphi_z = |z|\cos\varphi_z + i|z|\sin\varphi_z$
- $|z| = \sqrt{\bar{z}z} = \sqrt{a^2+b^2}$: the *magnitude* of $z = a+bi$
- $\varphi_z = \tan^{-1}(b/a)$: the *phase* or *argument* of $z = a+bi$
- $\text{Re}\{z\} = |z|\cos\varphi_z$
- $\text{Im}\{z\} = |z|\sin\varphi_z$

Formulas

Addition and subtraction

Just as we performed the addition of vectors component by component, we perform addition on complex numbers by adding the real parts together and adding the imaginary parts together:

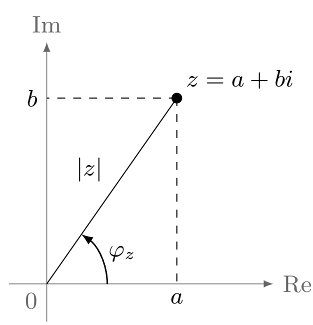
$$(a+bi)+(c+di) = (a+c)+(b+d)i.$$

Polar representation

We can give a geometrical interpretation of the complex numbers by extending the real number line into a two-dimensional plane called the *complex plane*. The horizontal axis in the complex plane measures the *real* part of the number. The vertical axis measures the *imaginary* part. Complex numbers are vectors in the complex plane.

It is possible to represent any complex number $z = a+bi$ in terms of its *magnitude* and its *phase*:

$$z = |z|\angle\varphi_z = \underbrace{|z|\cos\varphi_z}_{a} + \underbrace{|z|\sin\varphi_z}_{b}\, i.$$



The ~~magnitude~~ *magnitude* (or *absolute value*) of a complex number $z = a+bi$ is

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

This corresponds to the *length* of the vector that represents the complex number in the complex plane. The formula is obtained by using Pythagoras' theorem.

The *phase*, also known as the *argument* of the complex number $z = a+bi$ is

$$\varphi_z \equiv \arg z \;=\; \texttt{atan2}(b,a) \;\overset{\dagger}{=}\; \tan^{-1}(b/a).$$

The phase corresponds to the angle z forms with the real axis. Note the equality labelled † is true only when $a > 0$, because the function $\tan^{-1}$ always returns numbers in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Manual corrections of the output of $\tan^{-1}(b/a)$ are required for complex numbers with $a < 0$.

Some programming languages provide the 2-input math function `atan2(y,x)` that correctly computes the angle the vector (x, y) makes with the x-axis in all four quadrants. Complex numbers behave like 2-dimensional vectors so you can use `atan2` to compute their phase.

Complex numbers have vector-like properties like magnitude and phase, but we can also do other operations with them that are not defined for vectors. The set of complex numbers $\mathbb{C}$ is a *field*. This means, in addition to the addition and subtraction operations, we can also perform multiplication and division with complex numbers.

Multiplication

The product of two complex numbers is computed using the usual rules of algebra:

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

In the polar representation, the product formula is

$$(p\angle\phi)(q\angle\psi) = pq\angle(\phi + \psi).$$

To multiply two complex numbers, multiply their magnitudes and add their phases.

Division

Let's look at the procedure for dividing complex numbers:

$$\frac{(a + bi)}{(c + di)} = \frac{(a + bi)}{(c + di)}\frac{(c - di)}{(c - di)} = (a + bi)\frac{(c - di)}{(c^2 + d^2)} = (a + bi)\frac{\overline{c + di}}{|c + di|^2}.$$

In other words, to divide the number z by the complex number s, compute $\bar{s}$ and $|s|^2 = s\bar{s}$ and then use

$$z/s = z\frac{\bar{s}}{|s|^2}.$$

You can think of $\frac{\bar{s}}{|s|^2}$ as being equivalent to s^{-1}.

De Moivre's ~~theorem~~formula

By replacing θ in Euler's formula with $n\theta$, we obtain de Moivre's ~~theorem~~formula:

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta.$$

De Moivre's ~~Theorem~~ formula makes sense if you think of the complex number $z = e^{i\theta} = \cos\theta + i\sin\theta$, raised to the n^{th} power:

$$(\cos\theta + i\sin\theta)^n = z^n = (e^{i\theta})^n = e^{in\theta} = \cos n\theta + i\sin n\theta.$$

Setting $n = 2$ in de Moivre's formula, we can derive the double angle formulas (page 81) as the real and imaginary parts of the following equation:

$$(\cos^2\theta - \sin^2\theta) + (2\sin\theta\cos\theta)i = \cos(2\theta) + \sin(2\theta)i.$$

Links

[Mini tutorial on the complex numbers]
`http://paste.lisp.org/display/133628`

P3.7 Given the vectors $\vec{p} = (1, 1, 0, 3, 3)$ and $\vec{q} = (1, 2, 3, 4, 5)$, calculate the following expressions:

(a) $\vec{p} + \vec{q}$ (b) $\vec{p} - \vec{q}$ (c) $\vec{p} \cdot \vec{q}$

3.7 (a) $(2, 3, 3, 7, 8)$. (b) $(0, -1, -3, -1, -2)$. (c) 30.

P3.8 Find a unit-length vector that is perpendicular to both $\vec{u} = (1, 0, 1)$ and $\vec{v} = (1, 2, 0)$.
Hint: Use the cross product.

3.8 $(-\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ or $(\frac{2}{3}, -\frac{1}{3}, -\frac{2}{3})$.

3.8 See `bit.ly/1cOa8yo` for calculations.

P3.9 Find a vector that is orthogonal to both $\vec{u}_1 = (1, 0, 1)$ and $\vec{u}_2 = (1, 3, 0)$ and whose dot product with the vector $\vec{v} = (1, 1, 0)$ is equal to 8.

3.9 $(12, -4, -12)$.

3.9 Any multiple of the vector $\vec{u}_1 \times \vec{u}_2 = (-3, 1, 3)$ is perpendicular to both $\vec{u}_1$ and $\vec{u}_2$. We must find a multiplier $t \in \mathbb{R}$ such that $t(-3, 1, 3) \cdot (1, 1, 0) = 8$. Computing the dot product we find $-3t + t = 8$, so $t = -4$. The vector we're looking for is $(12, -4, -12)$. See `bit.ly/1nmYH8T` for calculations.

P3.10 Compute the following expressions:

(a) $\sqrt{-4}$ (b) $\dfrac{2 + 3i}{2 + 2i}$ (c) $e^{3i}(2 + i)e^{-3i}$

3.10 (a) $2i$. (b) $\frac{1}{4}(5 + i)$. (c) $2 + i$.

P3.11 Solve for $x \in \mathbb{C}$ in the following equations:

(a) $x^2 = -4$ (b) $\sqrt{x} = 4i$

(c) $x^2 + 2x + 2 = 0$ (d) $x^4 + 4x^2 + 3 = 0$

Hint: To solve (d), use the substitution $u = x^2$.

3.11 (a) $x = \pm 2i$. (b) $x = -16$. (c) $x = -1 - i$ and $x = -1 + i$. (d) $x = i$, $x = -i$, $x = \sqrt{3}i$, and $x = -\sqrt{3}i$.

P3.12 Given the numbers $z_1 = 2 + i$, $z_2 = 2 - i$, and $z_3 = -1 - i$, compute

(a) $|z_1|$ (b) $\dfrac{z_1}{z_3}$ (c) $z_1 z_2 z_3$

3.12 (a) $\sqrt{5}$. (b) $\frac{1}{2}(-3 + i)$. (c) $-5 - 5i$.

P3.13 A real business is a business that is profitable. An imaginary business is an idea that is just turning around in your head. We can model the real-imaginary nature of a business project by representing the *project state* as a complex number $p \in \mathbb{C}$. For example, a business idea is described by the state $p_o = 100i$. In other words, it is 100% imaginary.

To bring an idea from the imaginary into the real, you must work on it. We'll model the work done on the project as a multiplication by the complex number $e^{-i\alpha h}$, where h is the number of hours of work and α is a constant that depends on the project. After h hours of work, the initial state of the project is transformed as follows: ~~$p_f = e^{-i\alpha h} p_i$~~ $p_f = e^{-i\alpha h} p_o$. Working on the project for one hour "rotates" its state by $-\alpha$[rad], making it less imaginary and more real.

If you start from an idea $p_o = 100i$ and the cumulative number of hours invested after t weeks of working on the project is $h(t) = 0.2t^2$, how long will it take for the project to become 100% real? Assume $\alpha = 2.904 \times 10^{-3}$. Hint: A project is 100% real if $\text{Re}\{p\} = p$.

3.13 $t = 52$ weeks.

3.13 We want the final state of the project to be 100% real: $p_f = 100$. ~~Given that~~ Since we start from ~~$p_i = 100i$~~ $p_o = 100i$, the rotation required is $e^{-i\alpha h(t)} = e^{-i\frac{\pi}{2}}$, which means $\alpha h(t) = \frac{\pi}{2}$. We can rewrite this equation as $h(t) = 0.2t^2 = \frac{\pi}{2\alpha}$ and solving for t we find $t = \sqrt{\frac{\pi}{2(0.002904)(0.2)}} = 52$ weeks.

P3.14 A farmer with a side-interest in robotics has come up with a prototype of a robotic tractor. The tractor is programmed to move with a speed of 0.524[km/h] and follow the direction of the hour-hand on a conventional watch. Assume the tractor starts at 12PM (noon) and is left to roam about in a field until 6PM. What is the shape of the trajectory that the tractor will follow? What is the total distance travelled by the tractor after six hours?

3.14 The tractor's trajectory is a half-circle. The total distance travelled is 3.14[km].

3.14 The direction of the tractor changes constantly throughout the day, and the overall trajectory has the shape of a half-circle. The total distance travelled by the tractor is equal to half of the circumference of a circle of radius R. Since we know it took the tractor 6 hours of movement at $v = 0.522$[km/h] to complete this half-circle, we have $\pi R = v(t_f - t_i) = 0.524(6)$, from which we find $R = 1$[km] and the total distance travelled by the tractor is $\pi R = 3.14$[km].

To predict the motion of objects after a collision, we can use a *momentum* calculation. An object of mass m moving with velocity $\vec{v}$ has momentum $\vec{p} \equiv m\vec{v}$. The principle of conservation of momentum states that **the total amount of momentum in a system before and after a collision is conserved**. Thus, if two objects with initial momenta $\vec{p}_{i1}$ and $\vec{p}_{i2}$ collide, the total momentum before the collision must be equal to the total momentum after the collision:

$$\sum \vec{p}_i = \sum \vec{p}_f \qquad \Rightarrow \qquad \vec{p}_{i1} + \vec{p}_{i2} = \vec{p}_{f1} + \vec{p}_{f2}.$$

We use this equation to calculate the final momenta $\vec{p}_{f1}$, $\vec{p}_{f2}$ of the objects after the collision.

There is another way to solve physics problems by applying the concept of *energy*. Instead of trying to describe the entire motion of the object, we can focus only on the initial parameters and the final parameters of an object's movement. The law of conservation of energy states that **the total energy of the system is conserved**:

$$\sum E_i = \sum E_f.$$

By knowing the total initial energy of a system, we can find the final energy in the system, and from the final energy we can calculate the final motion parameters.

Units

In math we work with numbers—we solve questions where the answers are numbers without dimensions like 3, 5 or 12.34. The universal power of math comes precisely from this abstraction of things into numbers. We could be solving for the number of sheep in a pen, the surface area of a sphere, or the annual revenue of your startup; we can apply the same mathematical techniques to each example, even though the numbers we use will represent very different kinds of quantities.

~~In physics we use numbers too, and because physics~~ Because physics deals with real-world concepts and quantities, each number in physics always comes with a *measurement unit*. We must pay attention to the units of physical quantities and—most ~~importantly—to distinguish~~ importantly—distinguish between the different *dimensions* of numerical quantities. An answer in physics is a number that represents a length, a time, a velocity, an acceleration, or some other physical quantity. It doesn't make sense to add a *time* and a *mass*, because the two numbers measure different kinds of quantities.

Gravitation

The force of gravity exists between any two ~~massive objects .~~ objects with mass. The magnitude of the gravitational force between two objects of mass M[kg] and m[kg] separated by a distance r[m] is given by the formula

$$F_g = \frac{GMm}{r^2}, \tag{4.6}$$

where $G = 6.67 \times 10^{-11}[\frac{\text{Nm}^2}{\text{kg}^2}]$ is the *gravitational constant*. This is one of Newton's biggest discoveries—the famous one-over-r-squared law ~~.~~ of gravitation.

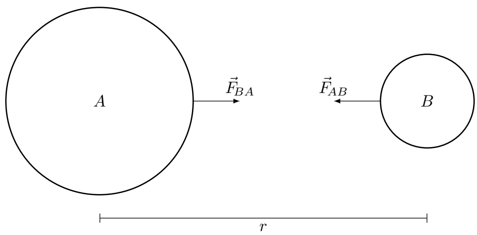


Figure 4.3: The gravitational force between two planets acts to pull them together. Assume Planet A has mass m_A and planet B has mass m_B. The force vector $\vec{F}_{AB}$ describes "Planet A's pull on Planet B." The force vector $\vec{F}_{BA}$ describes "Planet B's pull on Planet A." The magnitude of the gravitational pull is $F_g = \frac{Gm_Am_B}{r^2} = \|\vec{F}_{AB}\| = \|\vec{F}_{BA}\|$.

On the surface of the Earth, which has mass $M = 5.972 \times 10^{24}$[kg] and radius $r = 6.367 \times 10^6$[m], the force of gravity on an object of mass m is given by

$$F_g = \frac{GMm}{r^2} = \underbrace{\frac{GM}{r^2}}_{g} m = 9.81m = W.$$

We call this force the *weight* of the object, and to be precise we ~~should~~ write $\vec{W} = -mg\hat{\jmath}$ to indicate that the ~~force~~ weight acts *downward*, in the negative y-direction. Verify using ~~your~~ a calculator that $\frac{GM}{r^2} = 9.81 \equiv g$.

Force of a spring

A spring is a piece of metal twisted into a coil that has a certain natural length. The spring will resist any attempts to stretch or compress it. The force exerted by a spring is given by

$$\vec{F}_s = \cancel{-k}-kx,$$

where x is the amount by which the spring is displaced from its natural length, and the constant k[N/m] is a measure of the spring's *strength*. Note the negative sign indicates that the spring always acts to oppose the displacement.

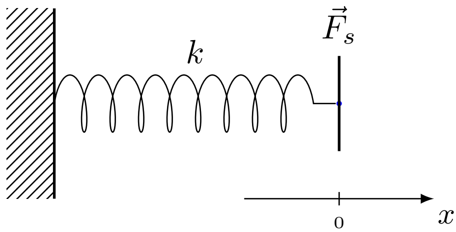


Figure 4.4: The force exerted by a spring is proportional to its displacement from its natural length, denoted x. When $x > 0$ the spring is stretched. When $x < 0$ the spring is compressed. When one end of the spring is fixed, the force exerted by the spring at the other end is $\vec{F}_s = -kx$.

If you try to stretch the spring, displacing it in the positive x-direction, then the force of the spring will pull against you (the spring will pull in the negative x-direction). Similarly, if you try to compress the spring (a displacement in the negative x-direction), the spring will push back against you, in the positive x-direction.

Normal force

The *normal* force is the force between two surfaces in contact. In this context, the word *normal* means "perpendicular to the surface of." The reason my coffee mug is not falling to the floor right now is that the table exerts a normal force $\vec{N}$ on the mug, keeping it in place.

Force of friction

In addition to the normal force between surfaces, there is also the force of friction $\vec{F}_f$, which acts to impede any sliding motion between the surfaces. There are two kinds of friction forces, and both are proportional to the amount of normal force between the surfaces:

$$\max\{\vec{F}_{fs}\} = \mu_s\|\vec{N}\| \quad \text{(static)}, \quad \text{and} \quad \vec{F}_{fk} = \mu_k\|\vec{N}\| \quad \text{(kinetic)},$$

where μ_s and μ_k are the static and kinetic *friction coefficients*. It makes sense that the force of friction should be proportional to the magnitude of the normal force $\|\vec{N}\|$, since the harder the two surfaces push against each other, the more difficult it becomes to make

The vector $v_x\hat{\imath}$ corresponds to the part of $\vec{v}$ that points in the x-direction.

Shortly, I'll be asking you over and over again to

find the component of $\vec{F}$ in the ? direction,

which is another way of asking you to find the number $v_?$.

The answer is usually equal to the length $\|\vec{F}\|$ multiplied by either cos or sin or sometimes -1, **depending on way the coordinate system is chosen**. So don't guess. Look at the coordinate system. If the vector points in the direction where x increases, then v_x should be a positive number. If $\vec{v}$ points in the opposite direction, then v_x should be negative.

To add forces $\vec{F}_1$ and $\vec{F}_2$, you need to add their components:

$$\vec{F}_1 + \vec{F}_2 = (F_{1x}, F_{1y}) + (F_{1x2x}, F_{2y}) = (F_{1x} + F_{2x}, F_{1y} + F_{2y}) = \vec{F}_{\text{net}}.$$

However, instead of dealing with vectors in the bracket notation, when solving force diagrams it is easier to simply write the x equation on one line, and the y equation on a separate line below it:

$$F_{\text{net},x} = F_{1x} + F_{2x},$$

$$F_{\text{net},y} = F_{1y} + F_{2y}.$$

It's a good idea to always write those two equations together as a block, so it's clear the first row represents the x dimension and the second row represents the y dimension for the same problem.

Force check

It is important to account for *all* the forces acting on an object. First of all, any object with mass on the surface of the Earth will feel a downward gravitational pull of magnitude $F_g = W = m\vec{g}$. Then you must consider whether any of the other forces are present: $\vec{T}$, $\vec{N}$, $\vec{F}_f$, and $\vec{F}_s$. Any time you see a rope tugging on the object, you can know there must be some tension $\vec{T}$, which is a force vector pulling on the object. Any time you have an object sitting on a surface, the surface pushes back with a *normal* force $\vec{N}$. If the object slides on the surface, then a force of friction acts against the direction of motion:

$$F_{fk} = \mu_k \|\vec{N}\|.$$

If the object is not moving, you must use the equation for the static force of friction. The maximum static friction force that the contact equal to zero, which means the forces acting on the object must be counter-balancing each other.

Step 5: Suppose the teacher now asks, "What is the magnitude of the normal force?" By looking at the second equation, you can answer, "$N = mg$ bro!"

Moving the fridge

You are trying to push your fridge across the kitchen floor. It weighs quite a lot, and is strongly "gripping" the floor when you try to push it. The static coefficient of friction between the bottom of your fridge and the tiles on the floor is μ_s. How much force $\vec{F}_{\text{ext}}$ will it take to cause the fridge to start moving?

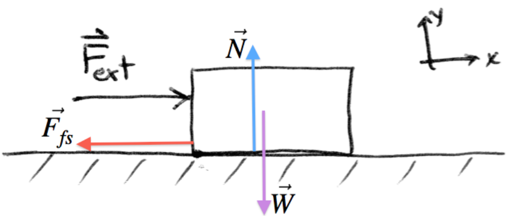


Figure 4.6: How ~~strongly~~ hard do you need to push before the fridge starts to slip?

$$\sum F_x = F_{\text{ext}} - F_{fs} = 0,$$
$$\sum F_y = N - mg = 0.$$

If you push with a force of $F_{\text{ext}} = 30$[N], the fridge will push back via its connection to the floor with a force $F_{fs} = 30$[N]. If you push harder, the fridge will push back harder and it still will not move. Only when you reach the slipping threshold will the fridge move. This means you'll need to push with a force equal to the *maximum* static friction force $F_{fs} = \mu_s N$, so we have

$$\sum F_x = F_{\text{ext}} - \mu_s N = 0,$$
$$\sum F_y = N - mg = 0.$$

To solve for F_{ext}, ~~first~~ look at the bottom equation~~and~~, isolate $N = mg$, then substitute the value of N in the top equation to find $F_{\text{ext}} = \mu_s mg$.

Friction slowing you down

Okay, now you're moving the fridge at a steady pace across the room~~:~~. The forces acting on the fridge are illustrated in Figure 4.7.

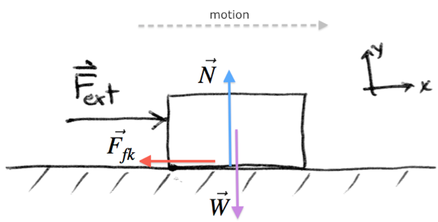


Figure 4.7: The fridge is moving. What is the magnitude of the external force $\vec{F}_{\text{ext}}$ required to counterbalance the kinetic force of friction?

Your equation of motion is expressed as

$$\sum F_x = F_{\text{ext}} - F_{fk} = ma_x,$$
$$\sum F_y = N - mg = 0.$$

In particular, if you want to keep a steady speed ($v = \text{const}$) as you move the fridge across the room, you'll need to push the fridge with a force that exactly balances the friction force and keeps $a_x = 0$.

To find the value of F_{ext} that allows you to keep a constant speed, solve

$$\sum F_x = F_{\text{ext}} - \mu_k N = 0,$$
$$\sum F_y = N - mg = 0.$$

The above set of equations are similar to the equations we obtained for the fridge that was not moving. The only difference is the kinetic coefficient of friction μ_k replaces the static coefficient of friction μ_s. Keeping the fridge moving with a constant velocity requires an external force $F_{\text{ext}} = \mu_k mg$. Generally, $\mu_k < \mu_s$, so less force is needed to keep the fridge moving than is needed to start the fridge moving.

Let's approach this whole friction thing from a different slant.

Incline

At this point, my dear readers, we're delving into the crucial question that you will—without a doubt—be asked to solve in your homework or at the final exam.

A block is sliding down an incline. What is its acceleration?

Since we know the values of m_1, m_2, μ_k, α, and β, we can calculate all the quantities on the left-hand side and solve for a. We thus obtain $a = \frac{m_1 g(\sin\alpha - \mu_k \cos\alpha) + m_2 g(\sin\beta - \mu_k \cos\beta)}{m_1 + m_2}$. Observe the final answer has the form $a = \frac{F_T}{m_T}$, where F_T is sum of forces acting on the system divided by the total mass of the system.

Other types of problems

Each of the previous examples asked you to find the acceleration, but sometimes a problem might give you the acceleration and ask you to solve for a different unknown. Regardless of what you must solve for, you should always start with a diagram and a sum-of-the-forces template. Once these equations are in front of you, you'll be able to reason through the problem more easily.

Experiment

You remove the spring from a retractable pen, and from the spring you suspend an object of known mass—say a 100[g] chocolate bar. With a ruler, you measure how much the spring stretches in the process. What is the spring constant k?

Discussion

In previous sections we discussed the *kinematics* problem of finding an object's position $x(t)$ given its acceleration function $a(t)$, and given the initial conditions x_i and v_i. In this section we studied the *dynamics* problem, which involves drawing force diagrams and calculating the net force acting on an object. Understanding these topics means you fully understand Newton's equation $F = ma$, which is perhaps the most important equation in this book.

We can summarize the entire procedure for predicting the position of an object $x(t)$ from first principles in the following equation:

$$\frac{1}{m}\underbrace{\left(\sum \vec{F} = \vec{F}_{\text{net}}\right)}_{\text{dynamics}} = \underbrace{\vec{a}(t) \overset{\vec{v}_i + \int dt}{\longrightarrow} \vec{v}(t) \overset{\vec{r}_i + \int dt}{\longrightarrow} \vec{r}(t)}_{\text{kinematics}}.$$

The left-hand side calculates the net force acting on an object, which is the *cause* of acceleration. The right-hand side indicates how we can calculate the position vector $\vec{r}(t)$ starting from the acceleration and the initial conditions. If you know the forces acting on any object (rocks, projectiles, cars, stars, planets, etc.) then you can predict the object's motion using this equation, which is pretty cool.

fluids, fields, and even collisions involving atomic particles described by quantum mechanics. The quantity of motion (a.k.a. momentum) cannot be created or destroyed—it can only be exchanged between systems.

Examples

Example 1 It's a rainy day, and from your balcony you throw—horizontally, at a speed of 10[m/s]—a piece of rolled-up carton with a mass of 0.4[g]. Shortly after it leaves your hand, the piece collides with a rain drop that weighs 2[g] and is falling straight down at a speed of 30[m/s]. What will the resulting velocity be if the two objects stick together after the collision?

The conservation of momentum equation says,

$$\vec{p}_{\text{in},1} + \vec{p}_{\text{in},2} = \vec{p}_{\text{out}}.$$

Plugging in the values, we obtain the equation

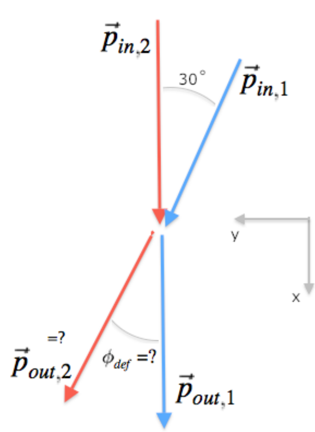
$$\begin{array}{ccccc} m_1\vec{v}_1 & + & m_2\vec{v}_2 & = & (m_1+m_2)\vec{v}_{\text{out}}, \\ 0.4\times(10,0) & + & 2\times(0,-30) & = & 2.4\times\vec{v}_{\text{out}}. \end{array}$$

Solving for $\vec{v}_{\text{out}}$ we find

$$\vec{v}_{\text{out}} = \frac{0.4(10,0)+2(0,-30)}{2.4} = (1.666,-25.0) = 1.666\hat{\imath} - 25.0\hat{\jmath}.$$

Example 2: Hipsters on bikes Two hipsters on fixed-gear bikes are headed toward the same intersection. Both hipsters have a speed of 50[km/h]. The first hipster crosses the street at a diagonal of 30 degrees when the two bikers collide. Did anyone else see this coming? Apparently, the second hipster didn't, because the thick frames of his glasses were blocking his peripheral vision.

Let's first look at the ~~hipster moving in the straight line, and let's assume the combined weight of the hipster and his bike is 100kg. As for the~~ street-crossing-at-30-degrees hipster ~~, his weight combined with the weight of~~ and assume his weight combined with the weight of his bike ~~frame (which is lighter and more expensive) totals~~ totals 90[kg]. For the hipster moving in the straight line, we assume the combined weight



of the hipster and his bike is 100[kg].

The story will continue in a moment, but first let's review the information I've given you so far:

$$\begin{aligned}\vec{p}_{\text{in},1} &= 90 \times 50\angle 30 \\ &= 90(50\cos 30, 50\sin 30),\end{aligned}$$

$$\begin{aligned}\vec{p}_{\text{in},2} &= 100 \times 50\angle 0 \\ &= 100(50, 0),\end{aligned}$$

where the x-coordinate points down the street, and the y-coordinate is perpendicular to the street.

Surprisingly, nobody gets hurt in this collision. The bikers bump shoulder-to-shoulder and bounce off each other. The hipster who was trying to cross the street is redirected down the street, while the hipster travelling down the street is deflected to the side and rerouted onto a bike path. I know what you are thinking: couldn't they get hurt at least a little bit? Okay, let's say the whiplash from the shoulder-to-shoulder collision sends the hipsters' heads flying toward each other and smashes their glasses. There you have it.

Suppose the velocity of the first hipster after the collision is 60 [km/h]. What is the velocity and the deflected direction of the second hipster? As given above, the outgoing momentum of the first hipster is $\vec{p}_{\text{out},1} = 90(60, 0)$, and we're looking to find $\vec{p}_{\text{out},2}$.

We can solve this problem with the conservation of momentum formula, which tells us that

$$\vec{p}_{\text{in},1} + \vec{p}_{\text{in},2} \ = \ \vec{p}_{\text{out},1} + \vec{p}_{\text{out},2}.$$

We know three of the above quantities, so we can solve for the remaining unknown vector by isolating it on one side of the equation:

$$\vec{p}_{\text{out},2} = \vec{p}_{\text{in},1} + \vec{p}_{\text{in},2} - \vec{p}_{\text{out},1},$$

$$\vec{p}_{\text{out},2} = 90(50\cos 30, 50\sin 30) + 100(50, 0) - 90(60, 0).$$

The x-component of the momentum $\vec{p}_{\text{out},2}$ is

$$p_{\text{out},2,x} = 90 \times 50\cos 30 + 5000 - 90 \times 60 = 3497.11,$$

and the y-component is $p_{\text{out},2,y} = 90 \times 50\sin 30 = 2250$.

The magnitude of the momentum of hipster 2 is given by

$$\|\vec{p}_{\text{out},2}\| = \sqrt{p_{\text{out},2,x}^2 + p_{\text{out},2,y}^2} = 4158.39 \quad [\text{kg km/h}].$$

initial potential energy: $K_f = U_i$. Since the formula for kinetic energy is $K = \frac{1}{2}mv^2$[J], we have $\frac{1}{2}mv_f^2 = mgh$. We cancel the mass on both sides of the equation and solve for v_f to obtain $v_f = \sqrt{2gh}$[m/s].

Both methods of solving the example problem lead us to the same ~~conclusion~~answer, but the energy reasoning is arguably more intuitive than blindly plugging values into a formula. In science, it is really important to know different ways of arriving at the same answer. Knowing about these alternate routes will allow you to check your answers and better understand concepts.

Concepts

Energy is measured in Joules [J] and it arises in several contexts:

- K = **kinetic energy**: the type of energy objects have by virtue of their motion
- W = **work**: the amount of energy an external force adds or subtracts from a system. Positive work corresponds to energy added to the system while negative work corresponds to energy withdrawn from the system.
- U_g = **gravitational potential energy**: the energy an object has by virtue of its position above the ground. We say this energy is *potential* because it is a form of *stored work*. Potential energy corresponds to the amount of work the force of gravity will add to an object when the object falls to the ground.
- U_s = **spring potential energy**: the energy stored in a spring when it is displaced (stretched or compressed) from its relaxed position.

There are many other kinds of energy—electrical energy, magnetic energy, sound energy, thermal energy, and so on. However, we'll limit our focus in this section to include only the *mechanical* energy concepts described above.

Formulas

Kinetic energy

An object of mass m moving at velocity $\vec{v}$ has a *kinetic energy* of

$$K = \frac{1}{2}m\|\vec{v}\|^2 \qquad \text{[J]}.$$

Note ~~,~~ the kinetic energy depends ~~only~~ on the speed $\|\vec{v}\|$ of the object and ~~is not affected by~~ not on the direction of motion.

The expression $\sum E_i$ corresponds to the sum of all the different types of energy the system contains in its initial state. Similarly, $\sum E_f$ corresponds to the sum of the final energies of the system. In mechanics, we consider three types of energy: kinetic energy, gravitational potential energy, and spring potential energy. Thus the conservation of energy equation in mechanics is

$$K_i + U_{gi} + U_{si} \quad + W_{\text{in}} \quad = \quad K_f + U_{gf} + U_{sf} \quad + W_{\text{out}}.$$

Usually, we're able to drop some of the terms in this lengthy expression. For example, we do not need to consider the spring potential energy U_s in physics problems that do not involve springs.

Explanations

Work and energy are measured in Joules [J]. Joules can be expressed in terms of other fundamental units:

$$[\text{J}] = [\text{N m}] = [\text{kg m}^2/\text{s}^2].$$

The first equality follows from the definition of work as force times displacement. The second equality comes from the definition of the Newton: $[\text{N}] = [\text{kg m/s}^2]$, which comes from $F = ma$.

Kinetic energy

A moving object has energy $K = \frac{1}{2}m\|\vec{v}\|^2[\text{J}]$, called *kinetic* energy from the Greek word for motion, *kinema*.

Note that velocity $\vec{v}$ and speed $\|\vec{v}\|$ are not the same as energy. Suppose you have two objects of the same mass, and one is moving two times faster than the other. The faster object will have twice the velocity of the slower object, and four times more kinetic energy.

Work

When hiring movers to help you move, you must pay them for the *work* they do. Work is the product of the amount of force needed for the move and the distance of the move. When the move requires a lot of force, more work will be done. And the bigger the displacement (think moving to ~~the South Shore~~ a different city versus moving next door), the more money the movers will ask for.

The amount of *work* done by a force $\vec{F}$ on an object that moves along some path p is given by

$$W = \int_p \vec{F}(x) \cdot d\vec{x}.$$

Gravitational potential energy

The force of gravity is given by $\vec{F}_g = -mg\hat{\jmath}$. The direction of the gravitational force ~~pulls~~ is downward, toward the centre of the Earth.

The gravitational potential energy of lifting an object from a height of $y = 0$ to a height of $y = h$ is given by

$$\begin{aligned} U_g(h) \equiv -W_{\text{done}} &= -\int_0^h \vec{F}_g \cdot d\vec{y} = -\int_0^h (-mg\hat{\jmath}) \cdot \hat{\jmath}\, dy \\ &= mg\int_0^h 1\, dy = mgy\Big|_{y=0}^{y=h} \\ &= mgh \qquad [\text{J}]. \end{aligned}$$

Spring potential energy

The force of a spring ~~when stretched a distance~~ when displaced by $\vec{x}$[m] from its natural position is given by $\vec{F}_s(\vec{x}) = -k\vec{x}$. The potential energy stored in a spring as it is ~~compressed~~ stretched from $y = 0$ to $y = x$[m] is given by

$$\begin{aligned} U_s(x) \equiv -W_{\text{done}} &= -\int_0^x \vec{F}_s(y) \cdot d\vec{y} \\ &= -\int_0^x (-ky)\, dy = k\int_0^x y\, dy = k\frac{1}{2}y^2\Big|_{y=0}^{y=x} \\ &= \frac{1}{2}kx^2 \qquad [\text{J}]. \end{aligned}$$

Note the formula applies when $x > 0$ (a spring stretched by $|x|$) and when $x < 0$ (a spring compressed by length $|x|$).

Conservation of energy

Energy cannot be created or destroyed. It can only be transformed from one form to another. If no external forces act on the system, then the system obeys the *conservation of energy* equation:

$$\sum E_i \;=\; \sum E_f.$$

~~If any~~ If external forces like friction do work on the system, we must account for their energy contributions:

$$\sum E_i \,+\, W_{\text{in}} = \sum E_f, \quad \text{or} \quad \sum E_i = \sum E_f \,+\, W_{\text{out}}.$$

The conservation of energy is one of the most important equations you will find in this book. It allows you to solve complicated problems by simply accounting for all the different kinds of energy involved in a system.

Thus, the block's velocity just after the bullet's impact is $v_{\text{out}} = \frac{mv}{M+m}$.

Next, we go to the conservation of energy principle to relate the initial kinetic energy of the block-plus-bullet to the height h by which the block rises:

$$K_i + U_i = K_f + U_f,$$

$$\frac{1}{2}(M+m)v_{\text{out}}^2 + 0 = 0 + (m+M)gh.$$

Isolate v_{out} in the above equation and set it equal to the v_{out} we obtained from the momentum calculation:

$$v_{\text{out}} = \frac{mv}{M+m} = \sqrt{2gh} = v_{\text{out}}.$$

We can use this equation to find the speed of the incoming bullet:

$$v = \frac{M+m}{m}\sqrt{2gh}.$$

Incline and spring

A block of mass m is released from rest at point (A), located on top of an incline at coordinate $y = y_i$. The block slides down the frictionless incline to the point (B) $y = 0$. The coordinate $y = 0$ corresponds to the relaxed length of a spring with spring constant k. The block then compresses the spring all the way to point (C), corresponding to $y = y_f$, where the block momentarily comes to rest~~again~~. The angle of the incline's slope is θ.

What is the speed of the block at $y = 0$? Find the value of y_f, the compression of the spring when the block stops. Bonus points if you can express your answer for y_f in terms of Δh, the difference in height between points (A) and (C).

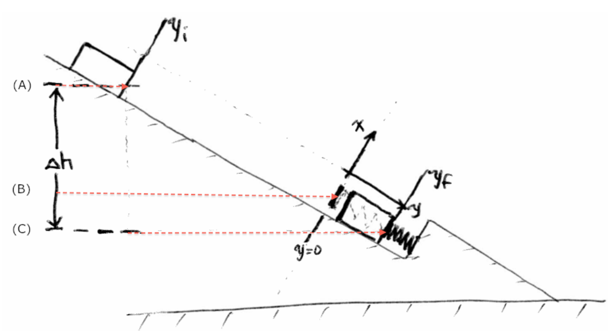


Figure 4.12: A block is released from the point (A) and slides down a frictionless incline to the point (B). The motion of the block is then slowed by a spring at the bottom of the incline. The block comes to rest at the point (C), after the spring is compressed by a length y_f.

Essentially, we have two problems: the block's motion from (A) to (B) in which its gravitational potential energy is converted into kinetic energy; and the block's motion from (B) to (C), in which all its energy is converted into spring potential energy.

There is no friction in either movement, so we can use the conservation of energy formula:

$$\sum E_i = \sum E_f.$$

For the block's motion from (A) to (B), we have

$$K_i + U_i = K_f + U_f.$$

The block starts from rest, so $K_i = 0$. The difference in potential energy is equal to mgh, and in this case the block is $|y_i| \sin\theta$ [m] higher at (A) than it is at (B), so we write

$$0 + mg|y_i| \sin\theta = \frac{1}{2}mv_B^2 + 0.$$

In the formula above, we assume the block has zero gravitational potential energy at point (B). The potential energy at point (A) is $U_i = mgh = mg|y_i - 0| \sin\theta$ *relative* to point (B), since point (A) is $h = |y_i - 0| \sin\theta$ metres higher than point (B).

Solving for v_B in this equation answers the first part of our question:

$$v_B = \sqrt{2g|y_i| \sin\theta}.$$

Now for the second part of the block's motion. The law of conservation of energy dictates that

$$K_i + U_{gi} + U_{si} = K_f + U_{gf} + U_{sf},$$

where i now refers to the moment (B), and f refers to the moment (C). Initially the spring is uncompressed, so $U_{si} = 0$. By the end of the motion, the spring is compressed by a total of $\Delta y = |y_f - 0|$[m], so its spring potential energy is $U_{sf} = \frac{1}{2}k|y_f|^2$. We choose the height of (C) as the reference potential energy; thus $U_{gf} = 0$. Since the difference in gravitational potential energy is $U_{gi} - U_{gf} = mgh = |y_f - 0| \sin\theta$, we can complete the entire energy equation:

$$\frac{1}{2}mv_B^2 + mg|y_f| \sin\theta + 0 = 0 + 0 + \frac{1}{2}k|y_f|^2.$$

Assuming the values of k and m are given, and knowing v_B from the first part of the question, we can solve for $|y_f|$ in the above equation.

To obtain the answer $|y_f|$ in terms of Δh, we'll use $\sum E_i = \sum E_f$ again, but this time i will refer to moment (A) and f to moment (C). The conservation of energy equation tells us $\cancel{mg\Delta h = \frac{1}{2}k|y_f|}$ $mg\Delta h = \frac{1}{2}k|y_f|^2$,

from which we obtain $\cancel{|y_f| = \frac{2mg\Delta h}{k}}$ $|y_f| = \sqrt{\frac{2mg\Delta h}{k}}$.

- $\hat{z}$: the usual $\hat{z}$-direction

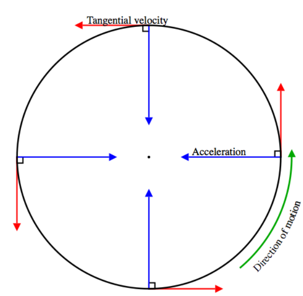
From a static observer's point of view, the tangential and radial directions constantly change their orientation as the object rotates around in a circle. From the rotating object's point of view, the tangential and radial directions are fixed. The tangential direction is always "forward" and the radial direction is always "to the side."

We can use the new coordinate system to describe the position, velocity, and acceleration of an object undergoing circular motion:

- $\vec{v} = (v_r, v_t)_{\hat{r}\hat{t}}$: the object's *velocity* expressed with respect to the $\hat{r}\hat{t}$-coordinate system
- $\vec{a} = (a_r, a_t)_{\hat{r}\hat{t}}$: the object's *acceleration*

The most important ~~parameters of motion~~ motion parameters are the tangential velocity v_t, the radial acceleration a_r, and the radius of the circle of motion R. We have $v_r = 0$ since the motion is entirely in the $\hat{t}$-direction, and $a_t = 0$ because in this case we assume the tangential velocity v_t remains constant (*uniform* circular motion).

Radial acceleration



The defining feature of circular motion is the presence of an acceleration that acts perpendicularly to direction of motion. At each instant, the object wants to continue moving along the tangential direction, but the radial acceleration causes the object's velocity to change direction. This constant inward acceleration causes the object to follow a circular path.

The *radial acceleration* a_r of an object moving in a circle of radius R with a tangential velocity v_t is given by

$$a_r = \frac{v_t^2}{R}. \tag{4.12}$$

This important equation relates the three key parameters of circular motion.

According to Newton's second law $\vec{F} = m\vec{a}$, an object's radial acceleration must be caused by a *radial force*. We can calculate the magnitude of this radial force F_r as follows:

$$F_r = ma_r = m\frac{v_t^2}{R}.$$

This formula connects the observable aspects of a circular motion v_t and R with the motion's cause: the force F_r, which always acts toward the centre of rotation.

To phrase it another way, we can say circular motion *requires* a radial force. From now on, when you see an object in circular motion, you can try to visualize the radial force that is causing the circular motion.

In the rock-on-a-rope example introduced in the beginning of the section (page 227), circular motion is caused by the tension ~~of a~~ in the rope that always acts in the radial direction (toward the centre of rotation). We're now in a position to calculate the value of the tension T in the rope using the equation

$$F_r = T = ma_r \qquad \Rightarrow \qquad T = m\frac{v_t^2}{R}\,.$$

The heavier the ~~object~~ rock and the faster it goes, the higher the tension in the rope. Inversely, the bigger the circle's radius, the less tension is required for the same v_t.

Example

During a student protest, a young activist named David is stationed on the rooftop of a building of height 12[m]. A mob of blood-thirsty neoconservatives is slowly approaching his position, determined to lynch him because of his leftist views. David has assembled a makeshift weapon by attaching a 0.3[kg] rock to the end of a shoelace of length 1.5[m]. The maximum tension the shoelace can support is 500[N]. What is the maximum tangential velocity $\max\{v_t\}$ the shoelace can support? What is the projectile's maximum *range* when it is launched from the roof?

The first part of the question ~~is answered with the $T = m\frac{v_t^2}{R}$ formula : $\max\{v_t\} = \sqrt{\frac{RT}{m}} = \sqrt{\frac{1.5\times 500}{0.3}} = 50$m/s.~~ can be answered using the formula $T = m\frac{v_t^2}{R}$:

$$\max\{v_t\} = \sqrt{\frac{RT}{m}} = \sqrt{\frac{1.5 \times 500}{0.3}} = 50[\text{m/s}].$$

To answer the second question, we must solve for the distance travelled by a projectile with initial velocity $\vec{v}_i = (v_{ix}, v_{iy}) = (50, 0)[\text{m/s}]$, launched from $\vec{r}_i = (x_i, y_i) = (0, 12)[\text{m}]$. First, solve for the total time of flight $t_f = \sqrt{2 \times 12/9.81} = 1.56[\text{s}]$. Then find the range of the rock by multiplying the projectile's horizontal speed by the time of flight $x(t_f) = 0 + v_{ix}t_f = 50 \times 1.56 = 78.20[\text{m}]$.

The angular velocity ω is useful because it describes the speed of a circular motion without any reference to the radius. If we know the angular velocity of an object is ω, we can obtain the tangential velocity by multiplying angular velocity times the radius: $v_t = R\omega[\text{m/s}]$.

You'll be asked to compute some angular velocities in the upcoming examples.

Bicycle odometer

Imagine you place a small speed detector gadget on one of the spokes of your bicycle's front wheel. Your bike's wheels have a radius $R = 14[\text{in}]$, and the gadget is attached at a distance $\frac{3}{4}R[\text{m}]$ from the wheel's centre. Find the wheel's angular velocity ω, period T, and frequency f of rotation when the bicycle's speed relative to the ground is $40[\text{km/h}]$. What is the tangential velocity v_t of the detector gadget?

The bicycle's velocity relative to the ground $v_{\text{bike}} = 40[\text{km/h}]$ is equal to the tangential velocity of the rim of the wheel:

$$v_{\text{bike}} = v_{\text{rim}} = 40[\cancel{\text{km}}/\cancel{\text{h}}] \times \frac{1000[\text{m}]}{1[\cancel{\text{km}}]} \times \frac{1[\cancel{\text{h}}]}{3600[\text{s}]} = 11.11[\text{m/s}].$$

We can find the wheel's angular velocity using $\omega = \frac{v_{\text{rim}}}{R}$ and the radius of the wheel $R = 14[\text{in}] = 0.355[\text{m}]$. We obtain $\omega = \frac{11.11}{0.355} = 31.24[\text{rad/s}]$. From here we calculate $T = \frac{2\pi}{\omega} = 0.20[\text{s}]$ and $f = \frac{1}{0.20} = 5[\text{Hz}]$. Finally, to compute the gadget's tangential velocity, multiply the wheel's angular velocity ω by its radius of rotation: $v_{\text{det}} = \omega \times \frac{3}{4}R = 8.333[\text{m/s}]$.

Rotation of the Earth

It takes exactly 23 hours, 56 minutes and 4.09 seconds for the Earth to compete one full turn (2π radians) around its axis of rotation. What is the Earth's angular velocity? What is the tangential speed of a person standing in Montreal, at a latitude of $45°$?

We can find ω by carrying out a simple conversion:

$$\frac{2\pi[\text{rad}]}{1[\cancel{\text{day}}]} \cdot \frac{1[\cancel{\text{day}}]}{23.93447[\cancel{\text{h}}]} \cdot \frac{1[\cancel{\text{h}}]}{3600[\text{s}]} = 7.2921 \times 10^{-5}[\text{rad/s}].$$

The radius of the trajectory traced by someone located at a latitude of $45°$ is given by $r = R\cos(45°) = 4.5025 \times 10^6[\text{m}]$, where $R = 6.3675 \times 10^6[\text{m}]$ is the radius of the Earth. Though it may not feel like you're moving, you are actually hurtling through space at a speed of

$$v_t = r\omega = 4.5025 \times 10^6 \times 7.2921 \times 10^{-5} = 464.32328.32[\text{m/s}],$$

which is equal to ~~1671.56~~1181.95[km/h]. Imagine that! You can attempt to present this fact if you are ever stopped by the cops for a speeding infraction: “Yes officer, I was doing 130[km/h], but this is really a negligible speed relative to the ~~1671~~1200[km/h] the Earth is doing around its axis of rotation.”

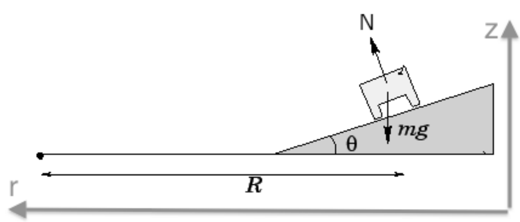
Three dimensions

For some problems involving circular motion, we'll need to consider the z-direction in the force diagram. In these cases, the best approach is to draw the force diagram as a cross section that is perpendicular to the tangential direction. Your diagram should show the $\hat{r}$ and $\hat{z}$ axes.

Using the force diagram, you can find all forces in the radial and vertical directions, as well as solve for accelerations a_r, a_z. Remember, you can always use the relation $a_r = \frac{v_t^2}{R}$, which connects the value of a_r with the tangential velocity v_t and the radius of rotation R.

Example Japanese people of the future design a giant racetrack for retired superconducting speed trains. The shape of the race track is a big circle with radius $R = 3$[km]. Because the trains are magnetically levitated, there is no friction between the track and the train $\mu_s = 0, \mu_k = 0$. What is the bank angle required for the racetrack so trains moving at a speed of exactly 400[km/h] will stay on the track without moving laterally?

We begin by drawing a force diagram which shows a cross section of the train in the $\hat{r}$ and $\hat{z}$ directions. The bank angle of the racetrack is θ. This is the unknown we're looking for. Because of the frictionless-ness of levitated superconducting suspension, there cannot be any force of friction F_f. Therefore, the only forces acting on the train are its weight $\vec{W}$ and the normal force $\vec{N}$.



The next step is to write two force equations that represent the $\hat{r}$ and $\hat{z}$ directions:

$$\sum F_r = N\sin\theta = ma_r = m\frac{v_t^2}{R} \quad \Rightarrow \quad N\sin\theta = m\frac{v_t^2}{R},$$
$$\sum F_z = N\cos\theta - mg = 0 \quad \Rightarrow \quad N\cos\theta = mg.$$

Note how the normal force $\vec{N}$ is split into two parts~~: the~~. The vertical component counterbalances the train's weight so it doesn't slide down the track. The component of $\vec{N}$ in the $\hat{r}$-direction is the force that causes the train's rotational motion.

We want to solve for θ in the above equations. It's a common trick to solve equations containing multiple trigonometric functions by dividing one equation by the other. Doing this, we obtain

$$\frac{N\sin\theta}{N\cos\theta} = \cancel{\frac{m\frac{v_t^2}{R}}{mg}}\,\frac{m\frac{v_t^2}{R}}{mg} \qquad \Rightarrow \qquad \tan\theta = \frac{v_t^2}{Rg}.$$

The final answer is $\theta = \tan^{-1}\!\left(\frac{v_t^2}{gR}\right) = \tan^{-1}\!\left(\frac{(400\times\frac{1000}{3600})^2}{9.81\times3000}\right) = 22.76°$. If the angle were any steeper, the trains would fall toward the track's centre. If the bank angle were any shallower, the trains would fly off to the side. The angle 22.76° is just right.

Discussion

Radial acceleration

In the kinematics section (page 134) we studied problems involving *linear acceleration*, in which an acceleration $\vec{a}$ acted in the direction of the velocity, causing a change in the magnitude of the velocity $\vec{v}$.

Circular motion deals with a different situation in which the object's speed $\|\vec{v}\|$ remains constant while its velocity $\vec{v}$ changes direction. At each point along the circle, the object's velocity points along the tangential direction; simultaneously, the radial acceleration pulls the object inwards, causing it to rotate.

Another term for radial acceleration is *centripetal* acceleration, which literally means "tending toward the centre."

Nonexistence of the centrifugal force

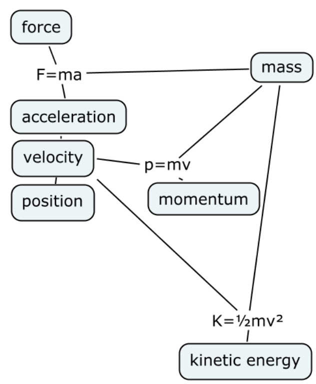
When a car makes a left turn, the passenger riding shotgun will feel pushed toward the right, into the passenger door. It would be erroneous to attribute this effect to a *centrifugal force* that acts away from the centre of rotation. In fact, no force is directly responsible for the feeling of being flung out of a car during a sharp turn.

The passenger is pushed into the door because of Newton's first law, which says that in the absence of external forces, an object will continue moving in a straight line. Since the initial motion occurs in the forward direction, the passenger's body naturally wants to continue moving in that direction. The force that the passenger feels from the car's door is necessary to cause the circular trajectory. If it

As you'll see shortly, the basic concepts we'll use to describe angular motion are directly analogous to the concepts of linear motion: position, velocity, acceleration, force, momentum, and energy.

Review of linear motion

It will be helpful to begin with a quick review of the concepts and formulas used to describe the linear motion of objects.



The linear motion of an object is described by its position $x(t)$, velocity $v(t)$, and acceleration $a(t)$ as functions of time. The position function tells you where the object is, the velocity tells you how fast it is moving, and the acceleration measures the change in the object's velocity.

The motion of objects is governed by Newton's first and second laws. In the absence of external forces, objects will maintain a uniform velocity (UVM), which corresponds to the equations of motion $x(t) = x_i + v_i t$ and $v(t) = v_i$. If a net force $\vec{F}$ acts on the object, the force will cause the object to accelerate. We obtain the magnitude of this acceleration with the formula $F = ma$. A constant force acting on an object will produce a constant acceleration (UAM), which corresponds to the equations of motion $x(t) = x_i + v_i t + \frac{1}{2}at^2$ and $v(t) = v_i + at$.

We also learned how to quantify the *momentum* $\vec{p} = m\vec{v}$ and the *kinetic energy* $K = \frac{1}{2}mv^2$ of moving objects. The momentum vector is the natural measure of the "quantity of motion," which plays a key role in collisions. The kinetic energy measures ~~how much energy the object has~~ the energy stored in the object by virtue of its motion.

An object's mass m is also an important factor in many physics equations. In the equation $F = ma$, the mass m measures the object's *inertia*—the object's resistance to being moved. The object's mass also appears in the formulas for momentum and kinetic energy; the heavier the object, the larger its momentum and its kinetic energy will be.

Concepts

We're ready to introduce the new concepts for describing the angular motion of objects.

- The kinematics of rotating objects is described in terms of angular quantities:
  - ▷ $\theta(t)$[rad]: the object's angular position
  - ▷ $\omega(t)$[rad/s]: the object's angular velocity
  - ▷ $\alpha(t)$[rad/s^2]: the object's angular acceleration
- I[kg m^2]: the *moment of inertia* tells you how difficult it is to make the object turn. The quantity I plays the same role in angular motion as the mass m plays in linear motion.
- $\mathcal{T}$[N m]: *torque* measures angular force. Torque is the cause of angular acceleration. The angular equivalent of Newton's second law $\sum F = ma$ is given by the equation $\sum \mathcal{T} = I\alpha$. This law states that applying an angular force (torque) $\mathcal{T}$ will produce an amount of angular acceleration α that is inversely proportional to the object's moment of inertia I.
- $L = I\omega$[kg m^2/s]: the *angular momentum* of a rotating object describes the "quantity of rotational motion."
- $K_r = \frac{1}{2}I\omega^2$[J]: the ~~*angular* or~~ *rotational* kinetic energy quantifies the amount of energy an object has by virtue of its rotational motion.

Formulas

Angular kinematics

Instead of talking about position x, velocity v, and acceleration a, for angular motion we ~~will~~ use the angular position θ, angular velocity ω, and angular acceleration α. Except for this change of ingredients, the "recipe" for finding the equations of motion remains the same:

$$\alpha(t) \quad \overset{\omega_i + \int dt}{\longrightarrow} \quad \omega(t) \quad \overset{\theta_i + \int dt}{\longrightarrow} \quad \theta(t).$$

Given the knowledge of an object's angular acceleration $\alpha(t)$, its initial angular velocity ω_i, and its initial angular position θ_i, we can use integration to find the equation of motion $\theta(t)$ that describes the angular position of the rotating object at all times.

Though this recipe can be applied to any form of angular acceleration function, you are only *required* to know the equations of motion for two special cases: the case of constant angular acceleration $\alpha(t) = \alpha$, and the case of zero angular acceleration $\alpha(t) = 0$. These are the angular analogues of *uniform acceleration motion* and *uniform velocity motion* we studied in the kinematics section.

The contribution of each piece of the object's mass dm to the total moment of inertia is multiplied by the squared distance of that piece from the object's centre, hence the units [kg m^2].

We rarely use the integral formula to calculate objects' moments of inertia. Most physics problems you'll be asked to solve will involve geometrical shapes for which the moment of inertia is given by simple formulas:

$$I_{\text{disk}} = \frac{1}{2}mR^2,\ I_{\text{ring}} = mR^2,\ I_{\text{sphere}} = \frac{2}{5}mR^2,\ I_{\text{sph.shell}} = \frac{2}{3}mR^2.$$

When you learn more about calculus (Chapter 5), you will be able to derive each of the above formulas on your own. For now, just try to remember the formulas for the ~~inertia of the disk and the~~ moment of inertia of a disk and a ring, as they are likely to show up in problems.

The quantity I plays the same role in the equations of angular motion as the mass m plays in the equations of linear motion.

Torques cause angular acceleration

Recall Newton's second law $F = ma$, which describes the amount of acceleration produced by a given force acting on an object. The angular analogue of Newton's second law is expressed as

$$\mathcal{T} = I\alpha. \tag{4.13}$$

This equation indicates that the angular acceleration produced by the torque $\mathcal{T}$ is inversely proportional to the object's moment of inertia. Torque is the cause of angular acceleration.

Angular momentum

The angular momentum of a spinning object measures the "amount of rotational motion." The formula for the angular momentum of an object with moment of inertia I rotating at an angular velocity ω is

$$L = I\omega \qquad [\text{kg m}^2/\text{s}].$$

In the absence of external torques, an object's angular momentum is a conserved quantity:

$$L_{\text{in}} = L_{\text{out}}. \tag{4.14}$$

This property is similar to the way momentum $\vec{p}$ is a conserved quantity in the absence of external forces.

Conservation of angular momentum

A spinning figure skater starts from an initial angular velocity of $\omega_i = 12[\text{rad/s}]$ with her arms extending away from her body. In this position, her body's moment of inertia is $I_i = 3[\text{kg m}^2]$. The skater then brings her arms close to her body, and in the process her moment of inertia changes to $I_f = 0.5[\text{kg m}^2]$. What is her new angular velocity?

This ~~looks like~~ is a job for the law of conservation of angular momentum:

$$L_i = L_f \qquad \Rightarrow \qquad I_i\omega_i = I_f\omega_f.$$

We know I_i, ω_i, and I_f, so we can solve for the final angular velocity ω_f. The answer is $\omega_f = I_i\omega_i/I_f = 3 \times 12/0.5 = 72[\text{rad/s}]$, which corresponds to 11.46 turns per second.

Conservation of energy

You have a 14[in] bicycle wheel with mass $m = 4[\text{kg}]$, with nearly all of its mass concentrated near the outside rim. The wheel is set in rolling motion up an incline at a velocity of 20[m/s]. How far up the incline will the wheel reach before it stops?

We can solve this problem with the principle of conservation of energy $\sum E_i = \sum E_f$. We must account for both the linear and rotational kinetic energies of the wheel:

$$\begin{aligned} K_i \quad + \quad K_{ri} \ + U_i &= K_f + K_{rf} + U_f\,, \\ \frac{1}{2}mv^2 + \frac{1}{2}I\omega^2 + 0 \ &= \ 0 \ + \ 0 \ + mgh. \end{aligned}$$

First, calculate I_{wheel} using the formula $I_{\text{ring}} = mR^2 = 4 \times (0.355)^2 = 0.5[\text{kg m}^2]$. If the wheel's linear velocity is 20[m/s], then its angular velocity is $\omega = v_t/R = 20/0.355 = 56.34[\text{rad/s}]$. We can now use these values in the conservation of energy equation:

$$\frac{1}{2}(4)(20)^2 + \frac{1}{2}(0.5)(56.34)^2 + 0 = 800.0 + 793.55 = (4)(9.81)h.$$

The wheel will reach a maximum height of $h = 40.61[\text{m}]$.

Note that roughly half the wheel's kinetic energy is stored in its rotational motion. This demonstrates the importance of accounting for K_r when solving energy problems involving rotating objects.

Static equilibrium

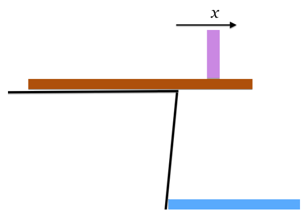
We say a system is in *equilibrium* when all forces and torques acting on the system "balance each other out." If an object is not moving, we say the object is in *static equilibrium*. Basically, zero ~~net force implies zero motion. We can also use this reasoning in reverse.~~ motion implies zero net force. If you see an object that is completely still, then the forces and torques acting on it must be in equilibrium:

$$\sum F_x = 0, \quad \sum F_y = 0, \quad \sum \mathcal{T} = 0. \tag{4.15}$$

Equilibrium means there is zero net force in the x-direction, zero net force in the y-direction, and zero net torque acting on the object.

Example: Walking the plank

A heavy wooden plank is placed so one-third of its length protrudes from the side of a pirate ship. The plank has a total length of 12[m] and a total weight of 120[kg]~~; this means 40kgof its weight is suspended above the ocean, while 80kgrests on the ship's deck~~. How far onto the plank can a person weighing 100[kg] walk before the plank tips into the ocean?

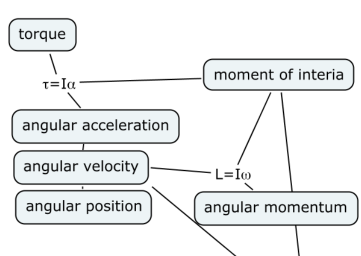


We'll use the torque equilibrium equation $\sum \mathcal{T}_E = 0$, where we calculate the torques relative to the edge of the ship, the point around which the plank will pivot. There are two torques involved: the torque produced by the plank's weight and the torque produced by the person's weight. The plank's weight acts in its centre of mass, which is located 2[m] from the edge of the ship. The torque produced by the weight of the plank is therefore given by $\mathcal{T}_2 = 120g \times 2 = 240g$[N m]. The torque produced by the person when he reaches a distance of x[m] from the edge of the ship is $\mathcal{T}_1 = -100gx$[N m]. Thus, the maximum distance the person can walk before the plank tips is $x = \frac{240g}{100g} =$ 2.4[m]. ~~After that it's all for the sharks.~~

Discussion

In this section, we applied the techniques and ideas from linear motion in order to describe the rotational motion of objects. Our coverage of rotational motion has been relatively brief because there were no new notions of physics to be learned.

Calling upon our prior knowledge of physics, we explored the parallels between the new rotational concepts and their linear counterparts. It is important you understand these parallels. To help



spring is maximally stretched or compressed—these are the locations where the pull of the spring is the strongest.

You'll definitely be asked to solve for the quantities v_{max} and a_{max} in exercises and exams. This is an easy task if you remember the above formulas and you know the values of the amplitude A and the angular frequency ω. Note the term *amplitude* applies more generally to the constant in front of *any* sin or cos function. Thus, we say that A is the amplitude of the position, $v_{\text{max}} = A\omega$ is the amplitude of the velocity, and $a_{\text{max}} = A\omega^2$ is the amplitude of the acceleration.

Energy

The potential energy stored in a spring that is stretched or compressed by a length x is given by the formula $U_s = \frac{1}{2}kx^2$. Since we know $x(t)$, we can obtain the potential energy of the mass-spring system as a function of time:

$$U_s(t) = \frac{1}{2}kxk[x(t)]^2 = \frac{1}{2}kA^2\cos^2(\omega t + \phi).$$

The potential energy reaches its maximum value $U_{s,\text{max}} = \frac{1}{2}kA^2$ when the spring is fully stretched or fully compressed.

The kinetic energy of the mass as a function of time is given by

$$K(t) = \frac{1}{2}mvm[v(t)]^2 = \frac{1}{2}m\omega^2A^2\sin^2(\omega t + \phi).$$

The kinetic energy is maximum when the mass passes through the centre position. The maximum kinetic energy is given by $K_{\text{max}} = \frac{1}{2}mv_{\text{max}}^2 = \frac{1}{2}mA^2\omega^2$.

Conservation of energy

Since the mass-spring system does not experience any dissipative forces (friction), the total energy of the system $E_T(t) = U_s(t) + K(t)$ must be *conserved*. The conservation of energy principle says that the sum of the system's potential energy and its kinetic energy must be the same at any two instants t_1 and t_2:

$$E_T(t_1) = U_s(t_1) + K(t_1) = U_s(t_2) + K(t_2) = E_T(t_2).$$

According to this equation, even if $U_s(t)$ and $K(t)$ change over time, the system's total energy $E_T(t)$ always remains constant.

Let us convince ourselves the system's total energy is indeed a constant. We can use the identity $\cos^2\theta + \sin^2\theta = 1$ to find the value of this constant:

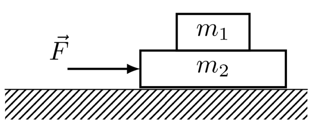
$$E_T(t) = U_s(t) + K(t)$$

P4.22 A spring with stiffness $k = 115$[N/m] is compressed by $\Delta x = 40$[cm] then released to push a ball of radius $R = 10$[cm] placed in front of it. When the ball leaves the spring, it rolls without skidding at an angular velocity of $\omega = 30$[rad/s]. What is the mass of the ball?
Hint: Use the angular velocity and the radius to find the linear velocity of the ball. Recall that $I_{\text{ball}} = \frac{2}{5}mR^2$.

4.22 $m = 1.46$[kg].

4.22 The spring's potential energy is equal to the sum of the linear and rotational kinetic energies of the ball: $\frac{1}{2}k(\Delta x)^2 = \frac{1}{2}mv^2 + \frac{1}{2}I_{\text{ball}}\omega^2$. Because the ball is rolling without skidding, we can calculate the linear velocity from the angular velocity and the radius $v = \omega R = 3$[m/s]. Substituting all knowns into the energy equation, we can solve for the mass m.

P4.23 Two blocks are stacked on top of each other ~~. The mass of the upper block is~~ and lying on a frictionless surface. The upper block has mass $m_1 = 0.25$[kg]~~and~~ , the lower block ~~'s mass is~~ has mass $m_2 = 1.00$[kg], and the static coefficient of friction between the two blocks is $\mu_s = 0.24$. You push the lower block with a constant force ~~$F = 3.0$~~$\vec{F}$[N] as shown in the figure. Find the ~~minimum coefficient of static friction between the blocks~~ maximum force you can apply for which the upper block will not slip.



Hint: To remain connected, both blocks must have the same acceleration.

4.23 ~~$\mu_s = 0.24$~~$\|\vec{F}\| = 3$[N].

4.23 ~~Calculate the acceleration of the two blocks if they were to move together , then calculate the friction force required to provide that acceleration to the~~ In the regime where the static friction between the blocks is strong enough for them to remain stuck together without slipping, the two blocks will act as a single block of mass $m_1 + m_2 = 1.25$[kg] pushed by an external force $\vec{F}$[N]. We can also analyze the contact forces (normal force and friction force) between the two blocks. From the point of view of the upper block, ~~and finally solve for μ_s in $F_{fs} = \mu_s N$ where $N = m_1 g$.~~ it feels a normal force of $N = 2.45$[N] and the maximum static friction that can exist between it and the bottom block is $\max F_{fs} = \mu_s N = 0.6$[N]. The upper block feels the static friction force F_{fs} in the positive x-direction, this is the force that is making the block accelerate in tandem with the block below it. Using $F = m_1 a_1$ for the upper block, we find the maximum acceleration the friction force can cause is $a_1 = \frac{\max F_{fs}}{m_1} = 2.4$[m/s^2]. We can now find the amount of the external force F required to accelerate both blocks at 2.4[m/s^2] using $F = (m_1 + m_2)a = (0.25 + 1)2.4 = 3$[N].

P4.39 The disk brake pads on your new bicycle squeeze the brake disks with a force of 5000[N] from each side ~~. There is one brake pad on each~~ of the tire. The coefficient of friction between the brake pads and ~~brake disks~~ the disk is $\mu_k = 0.3$. The brake disks have radius $r = 6$[cm] and the bike's tires ~~have radius~~ radius is $R = 20$[cm].
1. What is the total force of friction exerted by each brake?
2. What is the torque exerted by each brake?
3. Suppose you're moving at 10[m/s] when you apply broth brakes. The combined mass of you and your bicycle is 100[kg]. How many times will the wheels turn before the bike stops?
4. What will the braking distance be?

4.39 (1) $F_f = 3000$[N] per wheel. (2) $\mathcal{T} = 180$[N m]. (3) 2.21 turns. (4) $2.\overline{7}$[m].

4.39 (1) The friction force is proportional to the normal force. The friction on each side of each disk is $F_f = 0.3 \times 5000 = 1500$[N] for a total friction force of $F_f = 3000$[N] per wheel. (2) The friction force of the brakes acts with a leverage of 0.06[m], so the torque produced by each brake is $\mathcal{T} = 0.06 \times 3000 = 180$[N m]. (3) The kinetic energy of a 100[kg] object moving at 10[m/s] is equal to $K_i = \frac{1}{2}100(10)^2 = 5000$[J]. We'll use $K_i - W = 0$, where W is the work done by the brakes. Let θ_{stop} be the angle of rotation of the tires when the bike stops. The work done by each brake is $180\theta_{\text{stop}}$. It will take a total of $\theta_{\text{stop}} = \frac{5000}{360} = 13.\overline{8}$[rad] to stop the bike. This angle corresponds to 2.21 turns of the wheels. (4) Your stopping distance will be $13.\overline{8} \times 0.20 = 2.\overline{7}$[m]. Yay for disk brakes!

P4.40 **Tarzan** A half-naked dude swings from a long rope attached ~~to the ceiling~~six metres above the ground. The rope has length 6[m]. The dude swings from an initial angle of $-50°$ (50° to the left of the rope's vertical line), almost grazes the ground, then swings all the way to the angle $+10°$, at which point he lets go of the rope. See Figure ??. How far will Tarzan fall, as measured from the centre position of the swing motion? Find $x_f = 6\sin(10) + d$ where d is the distance travelled by Tarzan after he lets go.
Hint: This is an energy problem followed by a projectile motion problem.

4.40 $\cancel{x_f = 2.08}$$x_f = 2.09$[m].

4.40 The energy equation $\sum E_i = \sum E_f$ in this case is $U_i = U_f + K_f$, or $mg(6 - 6\cos 50°) = mg(6 - 6\cos 10°) + \frac{1}{2}mv^2$, which can be simplified to $v^2 = 12g(\cos 10° - \cos 50°)$. Solving for v we find $v = 4.48$[m/s]. Now for the projectile motion part. The initial velocity is 4.48[m/s] at an angle of 10° with respect to the horizontal, so $\vec{v}_i = (4.42, 0.778)$[m/s]. Tarzan's initial position is $(x_i, y_i) = (6\sin(10), 6[1-\cos(10)]) = (1.04, 0.0911)$[m]. To find the total time of flight, we solve for t in $0 = -4.9t^2 + 0.778t + 0.0911$ and find $t = 0.237$[s]. Tarzan will land at $\cancel{x_f = 6\sin(10) + 4.42t = 2.08}$$x_f = 6\sin(10) + 4.42t = 2.09$[m].

P4.41 A disgruntled airport employee decides to vandalize a moving walkway by suspending a leaking-paint-bucket pendulum above it. The pendulum is composed of a long cable (considered massless) and a paint bucket with a hole in the bottom. The pendulum's oscillations are small, and ~~transverse~~ perpendicular to the direction of the walkway's motion. Find the equation $y(x)$ of the pattern of paint that forms on the moving walkway in terms of the pendulum's maximum angular displacement $\theta_{\max}$, its length ℓ, and the speed of the walkway v. Assume x measures distance along the walkway and y denotes the ~~transversal displacement~~ bucket's transversal displacement, measured from the centre of the walkway.
Hint: This is a simple harmonic motion question involving a pendulum.

4.41 $y(x) = \ell \sin(\theta_{\max}) \cos((\omega/v)x)$.

4.41 We begin by writing the general equation of motion for a pendulum: $\theta(t) = \theta_{\max} \cos(\omega t)$, where $\omega = \sqrt{g/\ell}$. Enter the walkway, which is moving to the left at velocity v. If we choose the $x = 0$ coordinate at a time when $\theta(t) = \theta_{\max}$, the pattern on the walkway can be described by the equation $y(x) = \ell \sin(\theta_{\max}) \cos(kx)$, where $k = 2\pi/\lambda$, and λ tells us how long (measured as a distance in the x-direction) it takes for the pendulum to complete one cycle. One full swing of the bucket takes $T = 2\pi/\omega$[s]. In that time, the moving walkway will have moved a distance of vT metres. So one cycle in space (one wavelength) is $\lambda = vT = v2\pi/\omega$. We conclude that the equation of the paint on the moving sidewalk is $y(x) = \ell \sin(\theta_{\max}) \cos((\omega/v)x)$.

Links

~~A wikibook of physics exercises with solutions en.wikibooks.org/wiki/Physics~~[

Lots of interesting worked examples]
http://farside.ph.utexas.edu/teaching/301/lectures/lectures.html
~~Interactive exercises from the MIT mechanics course on edX www.edx.org/course/m~~

Recall Newton's second law $F_{\text{net}}(t) = ma(t)$, which can also be written as

$$\frac{F_{\text{net}}(t)}{m} = a(t) = x''(t) = \frac{d}{dx}\left(\frac{d}{dx}x(t)\right).$$

In Chapter 2 we learned how to use integration to solve for $x(t)$ in special cases where the net force is constant $F_{\text{net}}(t) = F_{\text{net}}$. In this chapter, we'll revisit the procedure for finding $x(t)$, and learn how to calculate the motion of an object affected by an external force that varies over time $F_{\text{net}}(t)$.

Limits

The main new tool we'll use in ~~our study of~~ calculus is the notion of a *limit*. ~~In calculus, we often~~ In calculus we use limits to describe what happens to mathematical expressions when one variable becomes very large, or alternately becomes very small.

For example, to describe a situation where a number n becomes bigger and bigger, we can say,

$$\lim_{n\to\infty} (\text{expression involving } n).$$

This expression is read, "in the limit as n goes to infinity, expression involving n."

Another type of limit occurs when a small, positive number—for example $\delta > 0$, the Greek letter *delta*—becomes progressively smaller and smaller. The precise mathematical statement that describes what happens when the number δ tends to 0 is

$$\lim_{\delta\to 0} (\text{expression involving } \delta),$$

which is read as, "the limit as δ goes to zero, expression involving δ."

Derivative and integral operations are both defined in terms of limits, so understanding limits is essential for calculus. We'll explore limits in more detail and discuss their properties in Section 5.4.

Sequences

So far, we've studied functions defined for real-valued inputs $x \in \mathbb{R}$. We can also study functions defined for natural number inputs $n \in \mathbb{N}$. These functions are called *sequences*.

A sequence is a function of the form $a : \mathbb{N} \to \mathbb{R}$. The sequence's input variable is usually denoted n or k, and it corresponds to the *index* or number in the sequence. We describe sequences either by

~~Techniques~~Series techniques

The main mathematical question we'll study with series is the question of their convergence. We say a series $\sum a_n$ *converges* if the infinite sum $S_\infty \equiv \sum_{n\in\mathbb{N}} a_n$ equals some finite number $L \in \mathbb{R}$.

$$S_\infty = \sum_{n=0}^{\infty} a_n = L \quad \Rightarrow \quad \text{the series } \sum a_n \text{ converges.}$$

We call L the *limit* of the series $\sum a_n$.

If the infinite sum $S_\infty \equiv \sum_{n\in\mathbb{N}} a_n$ grows to infinity, we say the series $\sum a_n$ *diverges*.

$$S_\infty = \sum_{n=0}^{\infty} a_n = \pm\infty \quad \Rightarrow \quad \text{the series } \sum a_n \text{ diverges.}$$

The main series technique you need to learn is how to spot the differences between series that converge and series that diverge. You'll learn how to perform different *convergence tests* on the terms in the series, which will indicate whether the infinite sum converges or diverges.

Applications

Series are a powerful computational tool. We can use series to compute approximations to numbers and functions.

For example, the number e can be computed as the following series:

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + 1 + \frac{1}{2\cdot 1} + \frac{1}{3\cdot 2} + \frac{1}{4\cdot 3\cdot 2} + \frac{1}{5\cdot 4\cdot 3\cdot 2} + \cdots.$$

The factorial operation $n!$ is the product of n times all integers smaller than n: $n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$. As we compute more terms from the series, our estimate of the number e becomes more accurate. The partial sum of the first six terms (as shown above) gives us an approximation of e that is accurate to three decimals. The partial sum of the first 12 terms gives us e to an accuracy of nine decimals.

Another useful thing you can do with series is approximate functions by infinitely long polynomials. The ~~*Taylor*~~ *power series* approximation for a function $f(x)$ is defined as the series

$$f(x) = \sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \cdots.$$

Each term in the series is of the form $a_n = c_n x^n$, where c_n is a constant that depends on the function $f(x)$.

For example, the power series of $\sin(x)$ is

$$\sin(x) = \underbrace{\overbrace{x}^{T_1(x)} - \frac{x^3}{3!} + \frac{x^5}{5!}}_{T_5(x)} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \cdots.$$

We can truncate the infinite series anywhere to obtain an approximation to the function. The function $T_5(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!}$ is the best approximation to the function $\sin(x)$ by a polynomial of degree 5. The equation of the tangent line $T_1(x)$ at $x = 0$ is a special case of the ~~Taylor~~ power series approximation procedure, which approximates the function as a first-degree polynomial. We will continue the discussion on series, their properties, and their applications in Section 5.19.

If you haven't noticed yet from glancing at the examples so far, the common theme underpinning all the topics of calculus is the notion of *infinity*. We now turn our attention to the infinite.

5.3 Infinity

Working with infinitely small quantities and infinitely large quantities can be tricky business. It~~is important that you develop an intuitive understanding of~~ 's important that we develop the appropriate language for describing these concepts as soon as possible. Like, now.

Infinitely large

The number ∞ is *really* large. How large? Larger than any number you can think of. Think of any number n. It is true that $n < \infty$. Now think of a bigger number N. It will still hold true that $N < \infty$. In fact, any finite number you can think of, no matter how large, will always be less than ∞.

Technically speaking, ∞ is not a number; infinity is a *process*. You can think of ∞ as the answer you obtain by starting from 0 and continuously adding 1 *forever*.

To see why $N < \infty$ for any finite number N, consider the following reasoning. When we add 1 to a number, we obtain a larger number. The operation $+1$ is equivalent to taking one unit step to the right on the number line. For any n, it is true that $n < n + 1$. To get to infinity we start from $n = 0$ and keep adding 1. After N steps, we'll arrive at $n = N$. But then we must continue adding 1 and obtain $N+1$, $N+2$, $N+3$, and so on. Since adding 1 always creates a larger number, the following chain of inequalities is true:

$$N \;<\; N+1 \;<\; N+2 \;<\; N+3 < \;\cdots\; < \;\infty.$$

Therefore $N < \infty$ for any finite N.

When we say a number n "goes to" infinity, we~~are~~'re saying n becomes increasingly larger and larger. No number ever actually arrives at infinity since infinity is obtained by adding 1 forever. There is no number $n \in \mathbb{R}$ such that $n = \infty$. Nevertheless, sometimes we can write $N = \infty$, which is an informal way of saying $N = \lim_{n\to\infty} n$.

Infinitely small

The opposite of ~~infinitely large is infinitely small~~an infinitely large number is an infinitely small number. As a mathematical convention, infinitely small numbers are denoted by the Greek letters ϵ (*epsilon*) and δ (*delta*). The infinitely small number $\epsilon > 0$ is a nonzero number smaller than any number you can think of. The number 0.00001 is pretty small, but it's true that $\epsilon < 0.00001$. The number 10^{-16} extends for 15 zeros after the decimal point, but still ϵ is smaller than it: $\epsilon < 10^{-16}$. Most often, the variable ϵ appears in limit expressions as a quantity that tends toward 0. The expression $\lim_{\epsilon\to 0}$ describes the process of ϵ becoming smaller and smaller, but never actually reaching zero, since by definition $\epsilon > 0$.

Infinitely many

The interval $[0, 1]$ of the number line contains infinitely many numbers. Think of the sequence $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$, and so forth. There is an infinite number of such fractions, and they all lie in the interval $[0, 1]$.

The ancient Greek philosopher Zeno was confused by this fact. He reasoned as follows. Suppose an archer shoots an arrow and sends it flying toward a target. After some time, the arrow will have travelled half the distance to the target. At some later time, the arrow will have travelled half of the remaining distance and so on, always getting closer and closer to the target. Zeno observed that no matter how little distance remains between the arrow and the target, there will always remain some distance to travel. To reach the target, the arrow would need to pass through an infinite number of points, which is impossible. "How could an infinite number of points fit inside a finite interval?" he figured.

Zeno's argument is not quite right. It is true that the arrow must pass through infinitely many points before it hits the target, but these points "fit" fine in the interval $[0, 1]$. These are mathematical points—they don't take up any space at all. We can commend Zeno for thinking about limits centuries before calculus was invented, but we shouldn't repeat his mistake. You must learn how to make limit arguments, because limits are important. Imagine if Zeno had tried to verify his theory experimentally by placing himself in front of an arrow. A wrong argument about limits could get you killed!

Interlude

If the concept of infinity were a person, it would have several problematic character traits. Let's see what we know about infinity so far. The bit about the infinitely large shows signs of megalomania. There is enough of this whole "more, more, more" mentality in the world already, so the last thing you want is someone like this as a friend. Conversely, the obsession with the infinitely small ϵ could be a sign of abnormal altruism: the willingness to give up all and leave less and less for oneself. You don't want someone *that* altruistic in your group. And that last part about how infinitely many numbers can fit in a finite interval of the number line sounds infinitely theoretical—definitely not someone to invite to a party.

Let's learn about one redeeming, practical quality of the concept of infinity. Who knows, you might become friends after all.

Infinitely precise

A computer science (CS) student and a math student are chatting over lunch. The CS student recently learned how to write code that computes mathematical functions as infinitely long series:

$$f(x) = e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2 \cdot 1} + \frac{x^3}{3 \cdot 2} + \frac{x^4}{4 \cdot 3 \cdot 2} + \cdots$$

She wants to tell her friend about her newly acquired powers.

The math student is also learning cool stuff about ~~transcendental~~ numbers. For example the number e can be defined as $e \equiv \lim_{n\to\infty} (1 + \frac{1}{n})^n$, but can never be computed exactly—it can only be approximated.

"You know, math is *soooo* much better than CS," says the math student, baiting her friend into an argument about the relative merits of their fields of study.

"What? No way. I can do *anything* on a computer," replies the incredulous scholar of code.

"But can you find exact answers?" the mathematician asks. "Can you compute the number e *exactly*?"

"Sure," says the computer scientist, opening her laptop and typing in a few commands. "The answer is $e = 2.718281828459045$."

"That is not exact," the mathematician points out, "it is just an approximation."

"Whatever—I gave you an approximation to fifteen digits after the decimal. If you're not satisfied with this, then I don't know what your problem is."

"Well, I asked for the *exact* value of e and you only gave me an approximation. Can you find e to 25 digits of precision?" asks the mathematician.

The computer scientist goes back to her laptop.

"Okay, $e = 2.7182818284590452353602875$," she says.

"What about computing e to 50 digits of precision?"

"$e = 2.71828182845904523536028747135266249775724709369995$," says the computer scientist a few seconds later.

"What about—"

"Listen,—" says her friend, "I have this code here that computes e in terms of its power series. The more terms I add in this series, the better my approximation will become. **I can achieve any precision you could possibly ask for**," she explains.

"Then you really know e!" exclaims the mathematician, convinced.

The computer scientist and the mathematician are discussing how to compute approximations to the number e. The mathematician thinks of the number e as the limit $e \equiv \lim_{n\to\infty} (1+\frac{1}{n})^n$. The computer scientist thinks of the number e as the infinite series

$$e = e^1 = \lim_{N\to\infty} \sum_{n=0}^{N} \frac{1}{n!} = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots .$$

Both formulas for e are correct. Observe that we can never compute the value of e exactly, since the formulas for e involve limits to infinity. Because no number ever arrives at infinity, we can never arrive at e either. The number e is a limit. We can only compute numbers that *approach* e.

The computer scientist can obtain approximations to e by computing the partial sum of the first N terms in the series:

$$e_N = \sum_{n=0}^{N} \frac{1}{n!} = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots + \frac{1}{N!} .$$

Let us denote as ϵ the required precision of the approximation. The more terms she adds, the more accurate the approximation e_N will become. She can always choose a value for N such that the approximation e_N satisfies $|e_N - e| < \epsilon$.

The computer scientist's first answer has a precision of $\epsilon = 10^{-15}$. To obtain an approximation to e with this precision, it is sufficient to

5.4 Limits

Limits are the mathematically precise way to talk about infinity. You must understand the language of limits to truly understand the infinitely small, the infinitely large, and the infinitely precise. Once you become comfortable with limits, you'll be able to understand the formal definitions of the derivative and integral operations.

Example

Let's begin with a simple example. Say you have a string of length ℓ and you want to divide it into infinitely many, infinitely short segments. There are infinitely many segments, and they are infinitely short, so together the segments add to the string's total length ℓ.

It's easy enough to describe this process in words. Now let's describe the same process using the notion of a limit. If we divide the length of the string ℓ into N equal pieces then each piece will have a length of

$$\delta \equiv \frac{\ell}{N}.$$

Let's make sure that N pieces of length δ added together equal the string's total length:

$$N\delta = N\frac{\ell}{N} = \ell.$$

Now imagine what happens when the variable N becomes larger and larger. The larger N becomes, the shorter the pieces of string will become. In fact, if N goes to infinity (written $N \to \infty$), then the pieces of string will have zero length:

$$\lim_{N\to\infty} \delta = \lim_{N\to\infty} \frac{\ell}{N} = 0.$$

In the limit as $N \to \infty$, the pieces of string are *infinitely small*.

Note we can still add the pieces of string together to obtain the whole length:

$$\lim_{N\to\infty} (N\delta) = \lim_{N\to\infty} \left(N\frac{\ell}{N} \right) = \ell.$$

Even if the pieces of string are *infinitely small*, because there are *infinitely many* of them, they still add to ℓ.

The ~~notion of a limit is one of the central ideas in this course. As~~ take-home message is that, as long as you clearly define your limits, you can use infinitely small numbers in your calculations. The notion of a limit is one of the central ideas in this course.

Limits of trigonometric functions:

$$\lim_{x\to 0}\frac{\sin(x)}{x}=1,\quad \lim_{x\to 0}\cos(x)=1,\quad \lim_{x\to 0}\frac{1-\cos x}{x}=0,\quad \lim_{x\to 0}\frac{\tan(x)}{x}=1.$$

A polynomial of degree n and the exponential function ~~base a with $a>1$~~ e^x both go to infinity as x goes to infinity:

$$\lim_{x\to\infty} x^n=\infty, \qquad\qquad \lim_{x\to\infty} ae^x=\infty.$$

Though both functions grow to infinity, the exponential function grows much faster. The limit of the ratio of the exponential function divided by any polynomial function is

$$\lim_{x\to\infty}\frac{a^x}{x^n}\frac{e^x}{x^n}=\infty, \qquad \text{for all } n\in\mathbb{N}, |a|>1..$$

In computer science, this distinction is a big deal when comparing the running times of algorithms. Imagine x represents the *size* of the problem we want to solve. A *polynomial-time algorithm* will take fewer than Cx^n steps to compute the answer, for some constants C and n. An *exponential-time algorithm* takes an exponential number of steps to compute the answer—the number of steps is described by the expression $\cancel{Da^x}$ De^x, for some ~~constants~~ constant D~~and a~~. Exponential-time algorithms are kind of useless because their running time becomes prohibitively long for large problems. With a large enough input x, an exponential-time algorithm with running time $\cancel{Da^x}$ De^x will take longer than the age of the universe to finish! The above results hold not only for the exponential function base e, but for all exponential functions with base a, so long as $a>1$:

$$\lim_{x\to\infty}\frac{a^x}{x^n}=\infty, \qquad \text{for all } n\in\mathbb{N},\ \text{for all } a>1.$$

The exponential function a^x with base $a>1$ grows faster than any polynomial function.

We'll now look at some limit formulas involving the logarithms. The logarithmic function is "weaker" than any polynomials function:

$$\lim_{x\to\infty}\frac{\ln(x)}{x^n}=0, \qquad \forall n\in\mathbb{N}_+.$$

Both $\ln(x)$ and x^n go to infinity as x becomes very large, but logarithmic function grow more slowly, so the ratio goes to zero. In fact, the logarithmic function is weaker than a polynomial even when raised to power p:

$$\lim_{x\to\infty}\frac{\ln^p(x)}{x^n}=0, \qquad \forall n\in\mathbb{N}_+ \text{ and } p<\infty.$$

Also of interest is the behaviour of the logarithmic function as x approaches zero from the right:

$$\lim_{x\to 0^+} x^n\ln(x)=0, \qquad \forall n\in\mathbb{N}_+.$$

The two factors in this limit expression pull in different directions. The logarithmic function goes to $-\infty$ as x approaches zero, but the polynomial factor x^n goes to zero as x goes to zero. Since the limit is equal to zero, we know the polynomial factor wins. Near $x=0^+$, the polynomial function x^n goes to zero faster than the logarithmic function $\ln(x)$ goes to negative infinity.

A third point of interest for the logarithmic function is near the value $x=1$, where the following limit holds:

$$\lim_{x\to1}\frac{\ln(x)}{x-1}=1.$$

In other words, the shape of $\ln(x)$ near $x=1$ resembles the function $x-1$, which is a line with slope one passing through the point $(1,0)$.

Euler's number

The number e is defined as the following limit:

$$e\equiv\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \qquad \text{or alternately as} \qquad e\cancel{=}\equiv\lim_{\epsilon\to0}(1+\epsilon)^{1/\epsilon}.$$

The first expression corresponds to a compound interest calculation with an annual interest rate of 100% where compounding is performed infinitely often.

The exponential function e^x can be obtained through similar limit expressions:

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n=e^x \qquad \text{or} \qquad \lim_{\epsilon\to0}(1+\epsilon x)^{1/\epsilon}=e^x.$$

For future reference, here ~~are some other limits~~ is another limit formula involving the exponential function:

$$\lim_{x\to0}\cancel{\frac{e^x-1}{x}}\frac{e^x-1}{x}=1,\ \cancel{\lim_{n\to\infty}1+\frac{x}{n}^n=e^x.}$$

~~And here are some limits involving logarithms:~~

$$\lim_{x\to 0^+} x^a \ln(x) = 0, \qquad \lim_{x\to\infty} \frac{\ln^p(x)}{x^a} = 0, \ \forall p < \infty$$

$$\lim_{x\to 0} \frac{\ln(x+a)}{x} = a, \qquad \lim_{x\to 0} \left(a^{1/x} - 1\right) = \ln(a).$$

which tells us of the $e^x - 1$ is similar to the function x near $x = 0$.

Properties

The calculation of the limit of the sum, difference, product, and quotient of two functions is computed as follows, respectively:

$$\begin{aligned}
\lim_{x\to a}(f(x)+g(x)) &= \lim_{x\to a} f(x) + \lim_{x\to a} g(x),\\
\lim_{x\to a}(f(x)-g(x)) &= \lim_{x\to a} f(x) - \lim_{x\to a} g(x),\\
\lim_{x\to a} f(x)g(x) &= \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x),\\
\lim_{x\to a} \frac{f(x)}{g(x)} &= \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}.
\end{aligned}$$

The above formulas indicate we are allowed to *take the limit inside* of the basic arithmetic operations.

L'Hopital's rule

If you are taking the limit of a fraction of two functions $\frac{f(x)}{g(x)}$ that obey $\lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to\infty} g(x) = \infty$, the limit of their ratio is

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \frac{\lim_{x\to\infty} f(x)}{\lim_{x\to\infty} g(x)} = \frac{0}{\infty} = 0.$$

Both the numerator and the denominator help drive the ratio to zero. Alternately, if you ever obtain a fraction of the form $\frac{\infty}{0}$ as a limit, where both the large numerator and the small denominator make the fraction grow to infinity, you can write $\frac{\infty}{0} = \infty$.

Sometimes, when evaluating limits of fractions $\frac{f(x)}{g(x)}$, you might end up with a fraction like

$$\frac{0}{0} \qquad \text{or} \qquad \frac{\infty}{\infty}.$$

These are called *undecidable* conditions. They are undecidable because we cannot tell whether the function in the numerator or the

Congratulations, you have just calculated your first derivative! The calculation wasn't that complicated, but the process was pretty long and tedious. The good news is you only need to calculate the derivative from first principles once. Once you obtain a *derivative formula* for a particular function, you can use the formula every time you see a function of that form.

The power rule

The derivative formula we obtained in the last example is a special case of the general formula for computing derivatives of powers of x. The *power rule* formula states:

$$\text{if} \quad f(x) = x^n \quad \text{then} \quad f'(x) = nx^{n-1}.$$

The proof of this formula proceeds by steps analogous to the steps used in the example above.

Example 2 Use the power rule to compute the derivatives of the following functions:

$$f(x) = x^{10}, \qquad g(x) = \sqrt{x^3}, \qquad h(x) = \frac{1}{x^3}.$$

In the first case, we apply the formula directly to find the derivative $f'(x) = 10x^9$. In the second case, we begin with the fact that square root is equivalent to an exponent of $\frac{1}{2}$, thus we rewrite the function as $g(x) = x^{\frac{3}{2}}$. After rewriting, we find $\cancel{g'(x) = \frac{3}{2}x^{\frac{1}{2}} = \frac{3}{2}\sqrt{x}}$$g'(x) = \frac{3}{2}x^{\frac{1}{2}} = \frac{3}{2}\sqrt{x}$. We can rewrite the third function as $h(x) = x^{-3}$, then use the power rule to compute the derivative $h'(x) = -3x^{-4} = -\frac{3}{x^4}$.

Applications of derivatives

Optimization

Consider some real-world problem in which a quantity is described by the function $f(x)$. The derivative function $f'(x)$ describes how the quantity $f(x)$ *changes* ~~over time~~as x changes. Often, we don't actually care about the value of $f'(x)$ and only need to find the sign of the derivative. If the derivative is positive $f'(x) > 0$, the function is *increasing*. If $f'(x) < 0$, the function is *decreasing*. If the function is horizontal at a certain point $x = x_0$, then $f'(x_0) = 0$. The points where $f'(x) = 0$ are important for finding the maximum and minimum values of $f(x)$.

and for any constant α, we have

$$[\alpha f(x)]' = \alpha f'(x).$$

The derivative of a linear combination of functions $\alpha f(x) + \beta g(x)$ is equal to the linear combination of the derivatives $\alpha f'(x) + \beta g'(x)$.

Product rule

The derivative of a product of two functions is obtained as follows:

$$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x).$$

Quotient rule

The *quotient rule* tells us how to obtain the derivative of a fraction of two functions:

$$\left[\frac{f(x)}{g(x)}\right]' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

Chain rule

If you encounter a situation that includes an inner function and an outer function, like $f(g(x))$, you can obtain the derivative by a two-step process:

$$[f(g(x))]' = f'(g(x))g'(x).$$

In the first step, leave the inner function $g(x)$ alone~~. Focus~~ , and focus on taking the derivative of the outer function $f(x)$~~, and leave the expression $g(x)$ inside the~~. This step gives us $f'(g(x))$, the value of f' ~~expression~~evaluated at $g(x)$. As the second step, we multiply the resulting expression by the derivative of the *inner* function $g'(x)$.

The chain rule tells us the derivative of a composite function is calculated as the product of the derivative of the outer function and the derivative of the inner function.

Simple example Consider the following derivative calculation:

$$\left[\sin(x^2))\right]' = \cos(x^2)\left[x^2\right]' = \cos(x^2)2x.$$

Proving the correctness of the chain rule for derivatives is a little bit more complicated. Actually, it is *a lot* more complicated. The argument presented in the next section is the most technical part of this book, and it's totally fine if you're not able to follow all the details. It's my duty as your calculus teacher to prove to you that the formula $[f(g(x))]' = f'(g(x))g'(x)$ is correct, but the proof is included only for readers who insist on seeing the full, excruciating details. Other readers should feel free to skip the next section and continue reading on page 319.

THE FOLLOWING PROOF IS COMPLETELY NEW, PLEASE CHECK IN FULL EVEN IF IT IS NOT BLUE — THOUGHT IT MIGHT BE EASIER TO READ A LONG PASSAGE IF IT IS PRINTED IN REGULAR BLACK FONT

Derivation of the chain rule

Assume $f(x)$ and $g(x)$ are differentiable functions. We want to show that the derivative of $f(g(x))$ equals $f'(g(x))g'(x)$, which is the chain rule for derivatives:

$$[f(g(x))]' = f'(g(x))g'(x).$$

Before we begin, I'd like to remark on the notation used to define derivatives. I happen to like the Greek letter δ (lowercase *delta*), so I defined the derivative of $f(x)$ as

$$f'(x) = \lim_{\delta \to 0} \frac{f(x+\delta) - f(x)}{\delta}.$$

Instead, we could use the variable Δ (uppercase *delta*) and write

$$f'(x) \equiv \lim_{\Delta \to 0} \frac{f(x+\Delta) - f(x)}{\Delta}.$$

In fact, we can use *any* variable for the limit expression. All that matters is that we divide by the *same* non-zero quantity as the quantity added to x inside the function, and that this quantity goes to zero. If we're not careful with our choice of limit variable we could run into trouble. Specifically, the definition of a limit depends on a "small nonzero number Δ," which is then used in the limit $\Delta \to 0$. The condition $\Delta \neq 0$ is essential because the expression $\frac{f(x+\Delta)-f(x)}{\Delta}$ is not well defined when $\Delta = 0$, since it leads to a divide-by-zero error.

In order to avoid any possibility of such errors, we define the following piecewise function

$$R(y,b) \equiv \begin{cases} \frac{f(y)-f(b)}{y-b} & \text{if } y \neq b, \\ f'(b) & \text{if } y = b. \end{cases}$$

Observe the function $R(y,b)$ is continuous in y, when we treat b as a constant. This follows from the definition of the derivative formula and the assumption that $f(x)$ is differentiable. Using the function $R(y,b)$, we can write the formula for the derivative of $f(x)$ as $f'(x) = \lim_{\Delta \to 0} R(x+\Delta, x)$. Note this formula is valid even in the case $\Delta = 0$.

To prove the chain rule, we'll need the function $R\big(g(x+\delta), g(x)\big)$ which is defined as follows:

$$R\big(g(x+\delta), g(x)\big) \equiv \begin{cases} \frac{f(g(x+\delta))-f(g(x))}{g(x+\delta)-g(x)} & \text{if } g(x+\delta) \neq g(x), \\ f'(g(x)) & \text{if } g(x+\delta) = g(x). \end{cases}$$

Okay, we're done with the preliminaries, so we can get back to proving the chain rule $[f(g(x))]' = f'(g(x))g'(x)$. We start with the limit expression for the left hand side of the equation:

$$[f(g(x))]' = \lim_{\delta \to 0} \frac{f(g(x+\delta)) - f(g(x))}{\delta}.$$

Observe that the fraction inside the limit can be written as

$$\frac{f(g(x+\delta)) - f(g(x))}{\delta} = R\big(g(x+\delta), g(x)\big)\frac{g(x+\delta) - g(x)}{\delta}.$$

This is the most tricky part of the proof so let's analyze carefully why this equation holds. We must check the equation holds in the two special cases in the definition of $R\big(g(x+\delta), g(x)\big)$.

Case A Whenever $g(x+\delta) \neq g(x)$ we have

$$\begin{aligned}
\frac{f(g(x+\delta)) - f(g(x))}{\delta} &= \frac{f(g(x+\delta)) - f(g(x))}{\delta}\frac{g(x+\delta) - g(x)}{g(x+\delta) - g(x)} \\
&= \frac{f(g(x+\delta)) - f(g(x))}{g(x+\delta) - g(x)}\frac{g(x+\delta) - g(x)}{\delta} \\
&= R\big(g(x+\delta), g(x)\big)\frac{g(x+\delta) - g(x)}{\delta}
\end{aligned}$$

Case B For points where $g(x+\delta) = g(x)$ we have

$$\frac{f(g(x+\delta)) - f(g(x))}{\delta} = \frac{0}{\delta} = 0,$$

and

$$R\big(g(x+\delta), g(x)\big)\frac{g(x+\delta) - g(x)}{\delta} = f'(g(x))\frac{0}{\delta} = 0.$$

Thus, the equation $\frac{f(g(x+\delta)) - f(g(x))}{\delta} = R\big(g(x+\delta), g(x)\big)\frac{g(x+\delta) - g(x)}{\delta}$ holds in both cases.

We can now rewrite the limit expression for $[f(g(x))]'$ using the equation established above:

$$\begin{aligned}
[f(g(x))]' &= \lim_{\delta \to 0} \frac{f(g(x+\delta)) - f(g(x))}{\delta} \\
&= \lim_{\delta \to 0}\left(\underbrace{R\big(g(x+\delta), g(x)\big)}_{F_1} \underbrace{\frac{g(x+\delta) - g(x)}{\delta}}_{F_2} \right)
\end{aligned}$$

We're trying to evaluate a limit expression that is the product of two factors $\lim_{\delta \to 0} F_1 F_2$. The limit of a product exists if the limits of both factors $\lim_{\delta\to 0} F_1$ and $\lim_{\delta\to 0} F_2$ exist. Before we proceed, we must evaluate the limit $\delta \to 0$ for both factors to ensure they exist.

To obtain the limit of the first factor, we'll rely on the continuity of the functions $g(x)$ and $R(y,b)$:

$$\lim_{\delta\to 0} g(x+\delta) = g(x) \quad \text{and} \quad \lim_{\Delta\to 0} R\big(b+\Delta, b\big) = R\big(b,b\big) = f'(b).$$

We define the quantity $\Delta \equiv g(x+\delta) - g(x)$ and using the continuity of $g(x)$, we can establish $\Delta \to 0$ as $\delta \to 0$. We are therefore allowed to change the limit variable from δ to Δ, and evaluate the limit of the first factor as follows:

$$\begin{aligned}
\lim_{\delta\to 0} F_1 &= \lim_{\delta\to 0} R\big(g(x+\delta), g(x)\big) \\
&= \lim_{\Delta\to 0} R\big(g(x)+\Delta, g(x)\big) \\
&= R\big(g(x), g(x)\big) = f'(g(x)).
\end{aligned}$$

We also know the limit of the second factor exists because it corresponds to the derivative of $g(x)$:

$$\lim_{\delta\to 0} F_2 = \lim_{\delta\to 0} \frac{g(x+\delta) - g(x)}{\delta} = g'(x),$$

and we assumed $g(x)$ is differentiable so its derivative must exists.

Since the limits of both factors $\lim_{\delta\to 0} F_1$ and $\lim_{\delta\to 0} F_2$ exist and are well defined, we can now complete the proof:

$$\begin{aligned}
[f(g(x))]' &= \lim_{\delta\to 0} \left(R\big(g(x+\delta), g(x)\big) \frac{g(x+\delta) - g(x)}{\delta} \right) \\
&= \left(\lim_{\delta\to 0} R\big(g(x+\delta), g(x)\big) \right) \left(\lim_{\delta\to 0} \frac{g(x+\delta) - g(x)}{\delta} \right) \\
&= f'(g(x)) g'(x).
\end{aligned}$$

This establishes the validity of the chain rule $[f(g(x))]' = f'(g(x))g'(x)$.

END OF NEW PROOF

Alternate notation

The presence of so many primes and brackets can make ~~the expressions above~~ derivative formulas difficult to read. As an alternative, we sometimes use ~~another notation~~ the *Leibniz notation* for derivatives. The three rules of derivatives in ~~the alternate~~ Leibniz notation are written as follows:

- Linearity: $\frac{d}{dx}(\alpha f(x) + \beta g(x)) = \alpha \frac{df}{dx} + \beta \frac{dg}{dx}$
- Product rule: $\frac{d}{dx}(f(x)g(x)) = \frac{df}{dx}g(x) + f(x)\frac{dg}{dx}$
- Chain rule: $\frac{d}{dx}\left(f(g(x))\right) = \frac{df}{dg}\frac{dg}{dx}$

Some authors prefer the notation $\frac{df}{dx}$ for the derivative of ~~the function~~ $f(x)$ ~~,~~ because it is more evocative of a rise-over-run calculation.

5.9 Higher derivatives

In the previous section we learned how to calculate the derivative $f'(x)$ of any function $f(x)$. The second derivative of $f(x)$ is the derivative of the derivative of $f(x)$, and is denoted

$$f''(x) \equiv [f'(x)]' \equiv \frac{d}{dx}f'(x) \equiv \frac{d^2}{dx^2}f(x).$$

This process can be continued to calculate higher derivatives of $f(x)$.

In practice, the first and second derivatives are most important because they have a geometrical interpretation. The first derivative of $f(x)$ describes the *slope* of $f(x)$ while the second derivative describes the *curvature* of $f(x)$.

Definitions

- $f(x)$: the original function
- $f'(x)$: the first derivative of the function $f(x)$. The first derivative contains information about the *slope* of the function $f(x)$.
- $f''(x)$: the second derivative of the function $f(x)$. The second derivative contains information about the *curvature* of the function $f(x)$.
  - ▷ If $f''(x) > 0$ for all x, the function $f(x)$ is *convex*. Convex functions open upward, like $f(x) = x^2$.
  - ▷ If $f''(x) < 0$ for all x, the function $f(x)$ is *concave*. Concave functions open downward, like $f(x) = -x^2$.

Later in this chapter, we will learn how to compute the Taylor series of a function, which is a procedure used to find polynomial approximations to any function $f(x)$:

$$f(x) \;\approx\; c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + \cdots + c_nx^n.$$

The values of the coefficients c_0, c_1, ..., c_n in the approximation require us to compute higher derivatives of $f(x)$. The coefficient c_n tells us whether $f(x)$ is more similar to $+x^n$ ($c_n > 0$), or to $-x^n$ ($c_n < 0$), or to neither of the two ($c_n = 0$).

Example Compute the third derivative of $f(x) = \sin(x)$.

The first derivative is $f'(x) = \cos(x)$. The second derivative will be $f''(x) = -\sin(x)$ so the third derivative must be $f'''(x) = -\cos(x)$. Note that $f^{(4)}(x) = f(x)$.

Optimization: the killer app of calculus

Knowing your derivatives will allow you to *optimize* any function—a crucial calculus skill. Suppose you can choose the input of $f(x)$ and you want to pick the *best* value of x. The best value usually means the *maximum* value (if the function measures something desirable like profits) or the *minimum* value (if the function describes something undesirable like costs). We'll discuss the *optimization algorithm* in more detail in the next section, but first let us look at an example.

~~Crime TV~~Example

~~A calculus teacher turned screenwriter is working on the pilot episode for a new TV series. Here is the story he has written so far.~~

~~The local drug boss~~ The boss of a large drug organization has recently been running into problems ~~as police are intercepting his dealers on the street.~~ with the authorities. The more drugs he sells, the more money he makes; but if he sells too much, ~~police arrests will increase~~ the authorities will start to crack down on his organization and he'll lose money. Fed up with this situation, he decides to find the *optimal* amount of drugs to ~~release on the streets~~push: as much as possible, but not enough to ~~trigger the police raids~~run into trouble with the law. One day he tells ~~his brothers and sisters in crime~~ all his advisors and underbosses to leave the room and picks up a pencil and a piece of paper to do some calculus.

If x is the amount of drugs ~~he releases on the street~~ his organization sells every day, then the amount of money he makes is given by the function

$$f(x) = 3000xe^{-0.25x},$$

where the linear ~~part~~ factor $3000x$ represents his profits ~~with no police involvement and the~~ and the factor $e^{-0.25x}$ represents the effects of the ~~police~~ authorities stepping up their actions as more drugs are released.

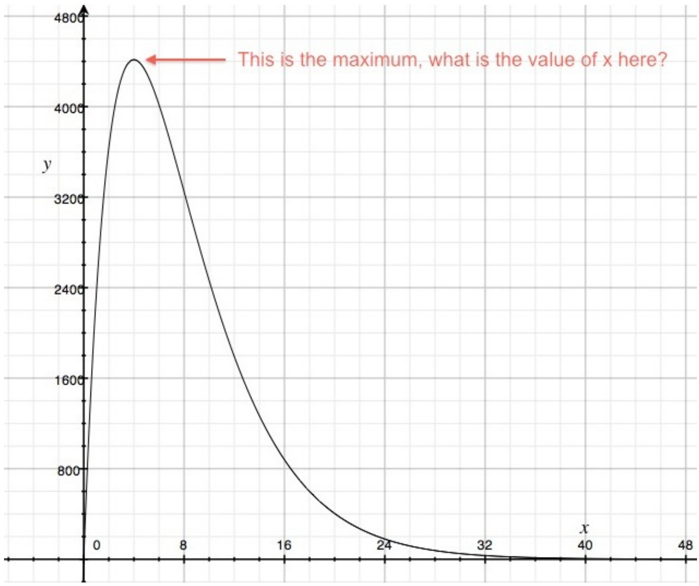


Figure 5.7: The graph of profit as a function of quantity sold.

Looking at the function, the drug boss asks, “What is the value of x which will give me the most profit~~from my criminal dealings?”~~ ?” Stated mathematically, he is asking,

$$\underset{x}{\operatorname{argmax}}\ 3000xe^{-0.25x} = ?$$

which means “find the value of the argument x that gives the *maximum* value of $f(x)$.”

Remembering a conversation with a crooked financial analyst he met in prison, the drug boss recalls the steps required to find the maximum of a function. First he must take the function’s derivative. Because the function is a product of two functions, he applies the product rule $[g(x)h(x)]' = g'(x)h(x) + g(x)h'(x)$. Taking the derivative of $f(x)$, he obtains

$$f'(x) = 3000e^{-0.25x} + 3000x(-0.25)e^{-0.25x}.$$

Whenever $f'(x) = 0$, the function $f(x)$ has zero slope. A maximum is exactly the kind of place where you’ll find zero slope—think of a mountain peak with steep slopes on all sides; the mountain is momentarily horizontal at its peak.

So when is the derivative zero? We set up the equation,

$$f'(x) = 3000e^{-0.25x} + 3000x(-0.25)e^{-0.25x} = 0.$$

We factor out the 3000 and the exponential function to obtain

$$3000e^{-0.25x}(1 - 0.25x) = 0.$$

Since $3000 \neq 0$ and $e^{-0.25x} \neq 0$, the term in the bracket must be equal to zero:

$$(1 - 0.25x) = 0,$$

or $x = 4$. The slope of $f(x)$ is equal to zero when $x = 4$. This x value corresponds to the peak of the curve.

Then and there, the ~~crime~~ boss calls his posse back into the room and proudly announces that from then on, his organization will release exactly four kilograms of drugs per day.

"Boss, how much money will we make per day if we sell four kilograms?" asks one of the ~~gangsters wearing sports pants.~~ underbosses, dressed in a slick suit.

"We'll make the *maximum* possible!" replies the boss.

"Yes I know Boss, but how much money is the maximum?"

The dude in ~~sports pants~~ suit is asking a good question. It is one thing to know *where* the maximum occurs, and it's another to know the ~~value~~ *value* of the function at this point. The dude is asking the following mathematical question:

$$\max_x \; 3000xe^{-0.25x} \; = ?$$

Since we already know the maximum occurs at $x^* = 4$, we can plug this ~~value~~ number into the function $f(x)$ to find

$$\max_x f(x) = f(4) = 3000(4)e^{-0.25(4)} = \frac{12000}{e} \approx 4414.55.$$

After ~~this conversation is complete, everyone, including the boss, begins to question their choice of occupation in life. When you crunch the numbers, is crime really worth it?~~ the meeting the boss feels exceptionally good about himself. He feels he's doing his job really well. As the CEO of a pharmaceutical corporation, his job is to blindly maximize the corporate profits, irrespective of possible negative outcomes and side effects.

A word of caution

~~It may seem funny to imagine calculus in the hands of the "bad guys," but in reality this is often the case.~~ The System is obsessed with this whole optimization thing. Optimize to make more profits, optimize to minimize costs, optimize stealing of natural resources from Third World countries, optimize anything that moves, basically. Therefore, the System wants *you*—the young and powerful generation of the future—to learn this important skill and become faithful employees of corporations. The corporates want you to learn calculus so you can help them optimize things, ensuring the smooth continuation of the whole enterprise.

Mathematical knowledge does not come with an ethics manual to help you decide what should and should not be optimized; this responsibility falls on you. If, like me, you don't want to become a corporate sellout, you can always choose to use calculus for science. It doesn't matter whether it will be physics, medicine, or running your own company, it is all good. Just stay away from the System. Please do this for yourself and our future, will you?

Having said these words of warning, let's now proceed so I can show you the powerful optimization algorithm.

5.10 Optimization algorithm

This section shows and explains the details of the algorithm for finding the maximum of a function. This is called *optimization*, as in finding the optimal solution to a problem.

Say you have the function $f(x)$, which represents a real-world phenomenon. For example, $f(x)$ could represent how much *fun* you have as a function of alcohol consumed during one evening. We all know that with too much x, the fun stops and you find yourself, as the Irish say, "talking to God on the big white phone." Too little x and you might not have enough Dutch courage to chat up that girl/guy from the table across the room. To have as much fun as possible, you want to find the alcohol consumption x^* where f takes on its maximum value.

This is one of the prominent applications of calculus (I'm talking about optimization, not alcohol consumption). This is why you've been learning about all those limits, derivative formulas, and differentiation rules in previous sections.

Definitions

- x: the *variable* we can control

- $[x_i, x_f]$: the interval of values from which x can be chosen. The values of x must obey $x_i \leq x \leq x_f$. These are the *constraints* of the optimization problem. For the drinking optimization problem, $x \geq 0$ since you can't drink negative alcohol, and probably $x < 2$ (in litres of hard booze) because roughly at this point a person will die from alcohol poisoning. So we are searching for the optimal amount of alcohol x in the interval $[0, 2]$.
- $f(x)$: the *function* we want to optimize. This function must be *differentiable*, meaning we can take its derivative.
- $f'(x)$: the *derivative* of $f(x)$. The derivative contains information about the slope of $f(x)$.
- *maximum*: a place where the function reaches a peak. When there are multiple peaks, we call the highest peak the *global maximum*, while all other peaks are *local maxima*.
- *minimum* : a place where the function reaches a low point at the bottom of a valley. The *global minimum* is the lowest point overall, whereas a *local minimum* is the minimum in some neighbourhood.
- *extremum*: a general term to describe both maximum and minimum points.
- *saddle point* : a place where $f'(x) = 0$ at a point that is neither a max nor a min. For example, the function $f(x) = x^3$ has a saddle point at $x = 0$.

Suppose some function $f(x)$ has a global maximum at x^*, and the value of that maximum is $f(x^*) = M$. The following mathematical notations apply:

- $\max_x f(x) = M$: the maximum value of $f(x)$
- $\text{argmax}_x\ f(x) = x^*$: the argmax operator tells you the location (the *argument* of the function) where the maximum occurs ~~$\max_x f(x) = M$: the maximum value~~

Algorithm for finding extrema

Input: a function $f(x)$ and a constraint region $C = [x_i, x_f]$
Output: the ~~location and value~~ locations and values of all maxima and minima of $f(x)$

Follow this algorithm step-by-step to find the extrema of a function:

1. First, *look* at $f(x)$. If you can plot it, plot it. If not, try to imagine what the function looks like.

2. Find the derivative $f'(x)$.

3. Solve the equation $f'(x) = 0$. Usually, there will be multiple solutions. Make a list of them. We'll call this the list of *candidates*.

4. For each candidate x^* ~~on~~ in the list, check to see whether it is a ~~max, a min~~maximum, a minimum, or a saddle point:

   - If $f'(x^* - 0.1)$ is positive and $f'(x^* + 0.1)$ is negative, then the point x^* is a ~~max~~maximum. The function goes up, flattens at x^*, then goes down after x^*. Therefore, x^* must be a peak.
   - If $f'(x^* - 0.1)$ is negative and $f'(x^* + 0.1)$ is positive, the point x^* is a ~~min~~minimum. The function goes down, flattens, then goes up, so the point must be a minimum.
   - If $f'(x^* - 0.1)$ and $f'(x^* + 0.1)$ have the same sign, the point x^* is a saddle point. Remove it from the list of candidates.

5. Now go through the list one more time and reject all candidates x^* that do not satisfy the constraints ~~C~~$\mathcal{C}$. In other words, if $x \in [x_i, x_f]$, the candidate stays; but if $x \notin [x_i, x_f]$, we remove it since this solution is not *feasible*. Returning to the alcohol consumption example, if you have a candidate solution that says you should drink 5[L] of booze, you must reject it because otherwise you would die.

6. Add x_i and x_f to the list of candidates. These are the boundaries of the constraint region and should also be considered. If no constraint was specified, use the *default* constraint region $-\infty < x < \infty$ and add $-\infty$ and ∞ to the list of candidates.

7. For each candidate x^*, calculate the function value $f(x^*)$.

The resulting list is a collection of *local* extrema: maxima, minima, and endpoints. The *global maximum* is the largest value from the list of local maxima. The *global minimum* is the smallest of the local minima.

~~Note that in~~ When dealing with points at infinity ~~,~~ such as $x^* = \infty$, we don't actually calculate a value; rather, we calculate the limit $\lim_{x\to\infty} f(x)$. ~~Usually, the~~ The function either blows up ~~$f(\infty) = \infty$~~ $\lim_{x\to\infty} f(x) = \infty$ (like x, x^2, e^x~~, ...~~), drops down indefinitely ~~$f(\infty) = -\infty$~~ $\lim_{x\to\infty} f(x) = -\infty$ (like $-x$, $-x^2$, $-e^x$~~, ...~~), or ~~reaches some~~ tends to some finite value (like $\lim_{x\to\infty} \frac{1}{x} = 0$, $\lim_{x\to\infty} e^{-x} = 0$). If a function goes to positive ∞ it doesn't have a global maximum and instead continues growing indefinitely. Similarly, functions that go toward negative ∞ don't have a global minimum.

Example 1 Find all the maxima and minima of the function

$$f(x) = x^4 - 8x^2 + 356.$$

Since no interval is specified, we'll use the default interval $x \in \mathbb{R}$. Let's go through the steps of the algorithm.

1. We don't know what the ~~x^4 function~~ function x^4 looks like, but it~~is~~'s probably similar to the x^2—it goes up to infinity on the far left and the far right.

2. Using the ~~formula for derivative of polynomials~~ power rule for derivatives we find

$$f'(x) = 4x^3 - 16x.$$

3. Now we must solve

$$4x^3 - 16x = 0,$$

which is the same as

$$4x(x^2 - 4) = 0,$$

which is the same as

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

The list of candidate points is ~~$\{x = -2, x = 0, x = 2\}$.~~ $\{x = 0, x = 2, x = -2$

4. For each of these points, we'll check to see if it is a max, a min, or a saddle point.

   (a) For $x = -2$, we check $f'(-2.1) = 4(-2.1)(-2.1-2)(-2.1+2) < 0$ and $f'(-1.9) = 4(-1.9)(-1.9-2)(-1.9+2) > 0$ to conclude $x = -2$ must be a minimum.

   (b) For $x = 0$ we try $f'(-0.1) = 4(-0.1)(-0.1-2)(-0.1+2) > 0$ and $f'(0.1) = 4(0.1)(0.1-2)(0.1+2) < 0$, which reveals we have a maximum at $x = 0$.

   (c) For $x = 2$, we check $f'(1.9) = 4(1.9)(1.9-2)(1.9+2) < 0$ and $f'(2.1) = 4(2.1)(2.1-2)(2.1+2) > 0$, so $x = 2$ must be a minimum.

5. We don't have any ~~constraints~~constraint region, so all of the above candidates make the cut.

6. We add the ~~two constraint~~ default boundaries $-\infty$ and ∞ to the list of candidates. At this point, our final shortlist of candidates contains $\{x = -\infty, x = -2, x = 0, x = 2, x = \infty\}$.

7. We now evaluate the function $f(x)$ for each of the values to obtain location-value pairs $(x, f(x))$, like so: $\{(-\infty, \infty), (-2, 340), (0, 356), (2, 340), (\infty, \infty)\}$. Note that $f(\infty) = \lim_{x\to\infty} f(x) = \infty^4 - 8\infty^2 + 356 = \infty$ and the same is true for $f(-\infty) = \infty$.

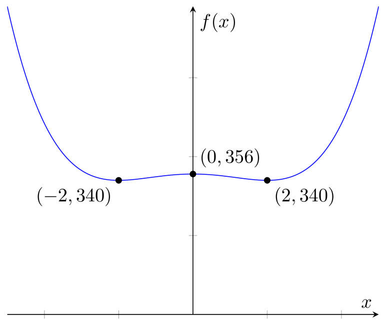


Figure 5.8: Graph of the function $f(x) = x^4 - 8x^2 + 356$ indicating the three critical points of the function located at $x = -2$, $x = 0$, and $x = 2$.

We are done. The function has no global maximum since it increases to infinity. It has a local maximum at $x = 0$ with value 356. It also has two global minima at $x = -2$ and $x = 2$, both of which have value 340. Thank you, come again.

Alternate algorithm

Instead of checking nearby points to the left and right of each critical point, we can modify the algorithm with an alternate Step 4 known as the *second derivative test*. Recall the second derivative tells us the function's *curvature*. If the second derivative is positive at a critical point x^*, then the point x^* must be a minimum. If, on the other hand, the second derivative at a critical point is negative, the function must be maximum at x^*. If the second derivative is zero, the test is inconclusive.

Alternate Step 4

- For each candidate x^*, see if it~~is~~'s a max, a min, or a saddle point:
  - ▷ If $f''(x^*) < 0$ then x^* is a max.
  - ▷ If $f''(x^*) > 0$ then x^* is a min.
  - ▷ If $f''(x^*) = 0$, the second derivative test fails. We must revert back to checking nearby values $f'(x^*-\delta)$ and $f'(x^*+\delta)$ to determine if x^* is a max, a min, or a saddle point.

Limitations

The optimization algorithm described above applies to *differentiable* functions of a single variable. Not all functions are differentiable. Functions with sharp corners in their graphs like the absolute value function $|x|$ are not differentiable everywhere, and therefore cannot be analyzed using the optimization algorithm. Functions with jumps in them, like the ~~Heaviside step function~~*Heaviside step function*, are not continuous and therefore not differentiable—the algorithm cannot be used on them either.

We can generalize the optimization procedure to optimize multivariable functions like $f(x, y)$. You'll learn how to do this in a *multivariable calculus* course. Multivariable optimization techniques are similar to the steps above, but with more variables and with more intricate constraint regions.

At last, I want to comment on the fact that ~~you can only maximize *one* function at a time~~**we can only maximize *one* function at a time**. Say the ~~drug boss from the TV series~~ CEO of the pharmaceutical corporation wanted to maximize ~~his funds~~ the company's profits $f(x)$ *and* ~~his gangster street cred~~ and the company's growth $g(x)$. This is not a well-posed problem; either you maximize $f(x)$ or you maximize $g(x)$, but you can't do both at the same time. There is no reason why a single x would give the highest value for both $f(x)$ and $g(x)$. If both functions are important~~to you, you~~, we can make a new function that combines the original two $F(x) = f(x) + g(x)$ and maximize $F(x)$. If ~~gangster street cred~~ company growth is three times more important to ~~you than funds, you could optimize~~ than profits, then the function to optimize is $F(x) = f(x)+3g(x)$, but it is mathematically and logically impossible to maximize both $f(x)$ and $g(x)$ at the same time.

Exercises

E5.3 The function $f(x) = x^3 - 2x^2 + x$ has a local maximum on the interval $x \in [0, 1]$. Find where this maximum occurs, and find the value of f at that point.

5.3 Max at $x = \frac{1}{3}$; $f\left(\frac{1}{3}\right) = \frac{4}{27}$.

5.11 Implicit differentiation

Thus far, we discussed how to compute derivatives of functions $f(x)$. We identified the function's output with the variable y and wrote

$y = f(x)$ to show the output y depends on the input x through the function $f(x)$. The slope of this function is the rise in the y-direction divided by the run in the x-direction and is equal to the value of the derivative function $\frac{dy}{dx} \equiv f'(x)$ at that point.

We can also use the derivative operation to compute the slope of graphs describing *mathematical relations* that are not expressed in the form $f(x) =$ something. For example, consider the equation that describes a circle of radius R:

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

The equation of a circle describes a *relation* between the variables x and y without specifying one variable as a function of the other. Nevertheless, we can still treat y as a function of x. We say the function $y(x)$ is *implicit*.

If we want to make the functional relationship between y and x *explicit*, we can rewrite the equation as $x^2 + (y(x))^2 = R^2$ and then solve for $y(x)$ to obtain

$$y(x) = \pm\sqrt{R^2 - x^2},$$

which shows ~~*explicitly*~~ explicitly how y depends on the variable x. Note there are actually two ys for each x: the top half and the bottom half of the circle.

~~**Problem**~~

Example

Consider the point $P = (x_P, y_P)$ that lies on the circle ~~$x^2 + y^2 = R$~~ $x^2 + y^2 = R^2$. Find the *slope* of the tangent line to the circle at P.

This problem is asking us to find $y'(x_P)$. Using the explicit function $y(x)$, we would first compute the derivative function $y'(x) = \pm\frac{1}{2}\frac{1}{\sqrt{R^2-x^2}}(-2x)$ and then substitute the value x_P into $y'(x)$. The slope of the tangent line to the circle at the point $P = (x_P, y_P)$ is

$$y'(x_P) = \frac{-x_P}{\sqrt{R^2 - x_P^2}} = -\frac{x_P}{y_P}.$$

But do we really need to go through the explicit equation? Let me show you a faster way to solve the problem, without the need to compute the function $y(x)$ explicitly. Apply the derivative operator $\frac{d}{dx}$

Example In the corporate world, ~~a man~~an executive officer's ego E is related to ~~his~~ their salary S by the following equation:

$$E^2 = S^3.$$

Suppose both E and S are functions of time. What is the rate of change of the ego of ~~Corporate Joe, the insurance analyst, when his~~ an executive, when their salary is 60k and ~~his~~ their salary increases at a rate of 5k per year?

This is called a *related rates* problem. We're given the relation $E^2 = S^3$ and the rate $\frac{dS}{dt} = 5000$ and we're asked to find $\frac{dE}{dt}$ when $S = 60000$. First, take the implicit derivative of the salary-to-ego relation:

$$\frac{d}{dt}\left[\ E^2\ \right] = \frac{d}{dt}\left[\ S^3\ \right],$$

$$2E\frac{dE}{dt} = 3S^2\frac{dS}{dt}.$$

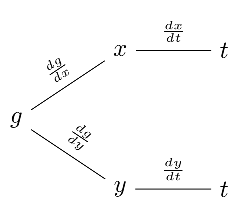
We're interested in the point where $S = 60000$. To find ~~Joe's ego at this point~~the ego points, solve for E in the relation $E^2 = S^3$; $E = \sqrt{60000^3} = 14696938.46$ ego points when $S = 60000$. Substituting all known values into the derivative of the relation, we find

$$2(14696938.46)\frac{dE}{dt} = 3(60000)^2(5000).$$

~~Joe~~The executive's ego is growing at $\frac{dE}{dt} = \frac{3(60000)^2(5000)}{2(14696938.46)} = 1837117.31$ ego points per year. Yay, ego points! I wonder what you can redeem these for.

Total derivative

Consider again a relation $g(x, y) = 0$, but this time assume that both x and y are implicit functions of a third variable t. To compute the derivative of the expression $g(x, y)$ with respect to t we must trace g's dependence on t through both x and y:



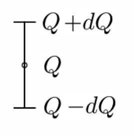
$$\frac{dg}{dt} = \frac{dg}{dx}\frac{dx}{dt} + \frac{dg}{dy}\frac{dy}{dt}.$$

We call this the *total derivative* of g because it represents the total dependence between g and t through both functions $x(t)$ and $y(t)$.

The reasoning behind the total derivative is similar to the reasoning behind the product rule. The derivative of $g(t) = f(t)h(t)$

Application of differentials to computing error bars

In science, when we report the results of an experimental measurement of some quantity Q, we write $Q \pm dQ$, where dQ is an estimate of the error of the measurement. The measurement error dQ is represented graphically as an "error bar" as shown on the right. The *precision* of a measurement is defined as the *ratio* of the error of the measurement divided by the size of the quantity being measured $\frac{dQ}{Q}$, or as a percentage.



Suppose the quantity Q depends on the variables x and y. We can express the dependence between the error in the measurement of Q and the error in the measurement of x and y as the formula:

$$dQ = \frac{dQ}{dx}dx + \frac{dQ}{dy}dy.$$

This is the *total differential* of Q. Note the similarity of the total differential formula to the total derivative formula.

Example ~~Suppose you~~ You want to calculate the kinetic energy of a particle ~~. Recall the formula for kinetic energy is~~ using the formula $K = \frac{1}{2}mv^2$. You measure the particle's mass m with precision 3%, and the particle's velocity with precision 2%. What is the precision of your kinetic energy calculation?

We want to find $\frac{dK}{K}$ and we're told $\frac{dm}{m} = 0.03$ and $\frac{dv}{v} = 0.02$. The first step is to calculate the *total differential* of the kinetic energy:

$$dK = d\left(\frac{1}{2}mv^2\right) = \frac{dK}{dm}dm + \frac{dK}{dv}dv = \frac{1}{2}v^2(dm) + mv(dv),$$

in which we used the product rule and the chain rule for derivatives. To obtain the relative error, divide both sides by K to obtain

$$\frac{dK}{K} = \frac{\frac{1}{2}v^2\,dm \;+\; mv\,dv}{\frac{1}{2}mv^2} = \frac{dm}{m} + 2\frac{dv}{v}.$$

The precision of the kinetic energy calculation in your experiment is $\frac{dK}{K} = 0.03 + 2(0.02) = 0.07$ or 7%. Note the error in the velocity measurement dv contributes twice as much as the error in the mass measurement dm~~. This is because the velocity~~ , since it appears with exponent two in the formula ~~$K = \frac{1}{2}mv^2$.~~ .

Discussion

We have reached the half-point of the calculus chapter. We learned about derivatives and described applications of derivatives to optimization problems, finding tangent lines, related rates, etc.

Before you continue reading about integrals in the second half of the chapter, I highly recommend you attempt to solve some of the derivative problems starting on page 408. ~~Understanding the theory is important, but it is by solving exercises that you will become a calculus expert.~~

5.12 Integrals

We now begin our discussion of integrals, the second topic in calculus. An integral is a fancy way of computing the area under the graph of a function. Integral calculus is usually taught as a separate course after differential calculus, but this separation can be counter-productive. The easiest way to understand integration is to think of it as the inverse of the derivative operation. Integrals are antiderivatives. Once you realize this fundamental fact, you'll be able to apply all your differential calculus knowledge to the domain of integral calculus. In differential calculus, we learned how to take a function $f(x)$ and find its derivative $f'(x)$. In integral calculus, we'll be given a function $f(x)$ and we'll be asked to find its *antiderivative* function $F(x)$. The antiderivative of $f(x)$ is a function $F(x)$ whose derivative equals $f(x)$.

In this section, we'll learn about two tasks: how to compute antiderivatives, and how to compute the area under the graph of $f(x)$. Confusingly, both of these ~~tasks~~ operations are called *integration*. To avoid any possibility of confusion, we'll define the two concepts right away:

- The *indefinite integral* of $f(x)$ is denoted $\int f(x)dx = F(x) + C$. To compute the indefinite integral of $f(x)$, you must find a function $F : \mathbb{R} \to \mathbb{R}$, such that $F'(x) = f(x)$. The indefinite integral is the antiderivative function.

- The *definite integral* of $f(x)$ between $x = a$ and $x = b$ is denoted $\int_a^b f(x)dx = A(a,b)$. Definite integrals correspond to the computation of the area under the function $f(x)$ between $x = a$ and $x = b$. The definite integral is a number $A(a,b) \in \mathbb{R}$.

The two integration ~~tasks~~ operations are related. The area under the curve $A(a,b)$ can be computed as the *change* in the antiderivative function, using to the formula $A(a,b) = \big[F(x) + C\big]_a^b = F(b) - F(a)$.

Definitions

You should already be familiar with these concepts:

- $\mathbb{R}$: the set of real numbers
- $f(x)$: a function of the form $f : \mathbb{R} \to \mathbb{R}$, which means f takes real numbers as inputs and produces real numbers as outputs
- $\lim_{\delta\to 0}$: a limit expression in which the number δ tends to zero
- $f'(x)$: the derivative of $f(x)$ is the rate of change of f at x. The derivative is a function of the form $f' : \mathbb{R} \to \mathbb{R}$.

These are the new concepts, which we will learn about in integral calculus:

- $A(a,b)$: the value of the *area* under the curve $f(x)$ from $x=a$ until $x=b$. The area $A(a,b)$ is computed as the following integral
$$A(a,b) = \int_a^b f(x)\,dx.$$
The $\int$ sign stands for *sum*. Indeed, the integral is the "sum" of $f(x)$ for all values of x between a and b.

- $A_0(x)$: the *integral function* of $f(x)$. The integral function corresponds to the computation of the area under $f(x)$ as a function of the upper limit of integration:
$$A_0(x) \equiv A(0,x) = \int_0^x f(u)\,du.$$
The choice of $x=0$ as the lower limit of integration is arbitrary.

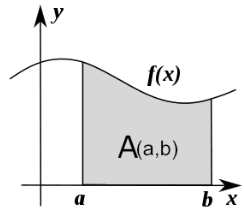
- $F(x)+C$: The *antiderivative* function of the function $f(x)$. An antiderivative function is defined as a function whose derivative equals to $f(x)$. The antiderivative function always includes an additive constant C. If the function $F(x)$ is an antiderivative (obeys $F'(x) = f(x)$) then the function $F(x)+C$ is also an antiderivative since
$$\frac{d}{dx}[F(x)+C] = f(x),$$
for any constant C.

- The fundamental theorem of calculus (FTC) states that the integral function $A_0(x)$ is equal to the antiderivative function $F(x)$ up to an additive constant C:
$$A(0,x) \equiv A_0(x) \overset{\text{FTC}}{=} F(x) + C.$$
The fundamental theorem leads us to the following formula for computing the area $A(a,b)$:
$$A(a,b) = A(0,b) - A(0,a) = A_0(b) - A_0(a) = F(b) - F(a).$$
The area under the curve, $A(a,b)$, is equal to the change in the antiderivative function $F(x)$ between $x=a$ and $x=b$.

The area under the curve

An integral describes the computation of the area under the curve $f(x)$ between $x = a$ and $x = b$:

$$A(a,b) \equiv \int_a^b f(x)\, dx.$$



We refer to the numbers a and b as the *limits of integration*. The location where the integral starts, $x = a$, is called the *lower limit* of integration. The location where the integral stops, $x = b$, is called the *upper limit* of integration.

The integral as a function

The *integral function* of $f(x)$ describes the "running total" of the area under the curve $f(x)$ as a function of the upper limit of integration:

$$A_0(x) \equiv A(0,x) \equiv \int_0^x f(u)\, du.$$

The variable x represents the upper limit of integration. The variable u inside the integral is called the *integration variable* and its value varies between $u = 0$ and $u = x$. The name of the integration variable u is not important; we can write $\int_0^x f(y)dy$ or $\int_0^x f(z)dz$ or even $\int_0^x f(\xi)d\xi$ and all of these represent the same function $A_0(x)$.

The choice of the lower limit of integration is also not important. For the sake of concreteness, we define the integral function to start at $x = 0$. A different choice for the lower limit of integration would lead to a different integral function. For example, the integral function that describes the area under $f(x)$ starting from $x = a$ is defined as $A_a(x) \equiv A(a,x) \equiv \int_a^x f(u)du$. The ~~function~~ functions $A_0(x)$ ~~can be obtained from the function~~ and $A_a(x)$ ~~by adding the missing area $A(0,a)$~~ differ only by the constant factor $A(0,a) = A_0(a)$:

$$\begin{aligned} \int_0^x f(u)\, du &= \int_0^a f(v)\, dv + \int_a^x f(w)\, dw \\ A(0,x) &= A(0,a) \qquad + \quad A(a,x) \\ A_0(x) &= A_0(a) \qquad + \quad A_a(x). \end{aligned}$$

The area $A(a,b) \equiv A_a(b)$ can be computed as the *change* in the value of $A_0(x)$ between $x = a$ and $x = b$:

$$A(a,b) \equiv \int_a^b f(x)\, dx = A_0(b) - A_0(a).$$

Note the formula $A(a,b) = A_c(b) - A_c(a)$ applies for all $c \in \mathbb{R}$.

The second step is to compute the area $A(a,b)$ as the change in the antiderivative function between $x = a$ and $x = b$:

$$A(a,b) = \left[F(x) + C\right]\Big|_{x=a}^{x=b} = [F(b) + C] - [F(a) + C] = F(b) - F(a).$$

Note the new "vertical bar" notation: $\cancel{g(x)\big|_{\alpha}^{\beta} = g(\beta) - g(\alpha)}$ $g(x)\big|_{\alpha}^{\beta} \equiv g(\beta) - g(\alpha)$. This is a useful shorthand for denoting the change in the function $g(x)$ between two points. Figure 5.9 illustrates the meaning of this procedure.

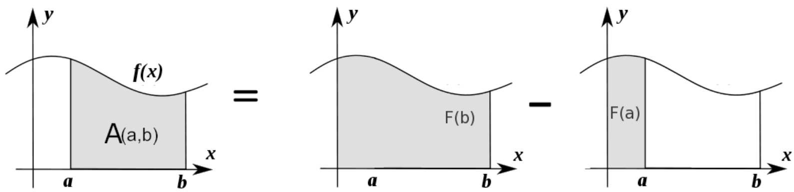


Figure 5.9: The function $F(x)$ measures the area under the curve $f(x)$. The area under $f(x)$ ~~between~~ from $x = a$ ~~and~~ to $x = b$ is $\cancel{A(a,b) = F(b) - F(a)}$ $A(a,b) = F(x)\big|_a^b \equiv F(b) - F(a)$.

You can also try rearranging the plots in Figure 5.9 to visualize the equation $A(0,b) = A(0,a) + A(a,b)$. The "running total" of the area under $f(x)$ until $x = b$ is equal to the "running total" of the area under $f(x)$ until $x = a$, plus the area $A(a,b)$. The formula $A(a,b) = F(b) - F(a)$ follows from combining the equation $A(0,b) = A(0,a) + A(a,b)$ with the result of the fundamental theorem of calculus: $A(0,x) \equiv A_0(x) \overset{\text{FTC}}{=} F(x)$.

Example 4 The antiderivative of $f(x) = x^2$ is $F(x) = \frac{1}{3}x^3 + C$. Use this fact to find the value of the definite integral $\int_a^b x^2\, dx$.

The definite integral is computed by evaluating the value of the antiderivative function at the upper limit and subtracting the value of the antiderivative function at the lower limit:

$$\int_a^b x^2\, dx = \left[\tfrac{1}{3}x^3 + C\right]_{x=a}^{x=b} = \left[\tfrac{1}{3}b^3 + C\right] - \left[\tfrac{1}{3}a^3 + C\right] = \frac{1}{3}(b^3 - a^3).$$

Example 5 What is the area under the curve $f(x) = \sin(x)$, between $x = 0$ and $x = \pi$? First we take the antiderivative

$$F(x) = \int \sin(x)\, dx = -\cos(x) + C.$$

Now we calculate the difference between $F(x)$ at the upper limit minus $F(x)$ at the lower limit:

$$\begin{aligned} A(0,\pi) &= \int_0^{\pi} \sin(x)\ dx \\ &= \underbrace{[-\cos(x) + C]}_{F(x)}\Big|_0^{\pi} \\ &= [-\cos\pi + C] - [-\cos(0) + C] \\ &= \cos(0) - \cos\pi \quad = \quad 1 - (-1) = 2. \end{aligned}$$

The final answer does not depend on the constant C because we evaluate the *change* in $F(x) + C$~~and so C cancels out.~~.

In case you are wondering what the "area under the curve" calculation is used for in practice, you should recall how we derived the kinematics equations in Chapter 2. The velocity $v(t)$ measures change in position $x(t)$ over time. The total change in position between $t = 0$ and $t = \tau$ is obtained by calculating the integral of $v(t)$ as follows:

$$x(\tau) - x(0) = \int_0^{\tau} v(t)\ dt.$$

Note how the dimensions work in this equation. Time is measured in seconds [s], and $v(t)$ is measured in [m/s], so the area under $v(t)$ has dimensions of [m/s]×[s] = [m].

Properties of integrals

Signed area

The value of a definite integral can be either positive or negative. If the limits of integration a and b satisfy $a < b$ (b is to the right of a on the number line), and if $f(x) > 0$ (meaning $f(x)$ is a positive function), then the area under the curve will be positive:

$$A(a,b) = \int_a^b f(x)\ dx \quad > \quad 0.$$

For a function $g(x) < 0$, the integral from a to b corresponds to a negative area. In general, if $f(x)$ is above the x-axis in some places, these zones ~~will~~ contribute positively to the total area under the curve; places where $f(x)$ is below the x-axis ~~will~~ contribute negatively to the total area $A(a,b)$.

We can also obtain a negative area if we swap the limits of integration. Suppose we have $f(x) > 0$, and limits of integration a and b

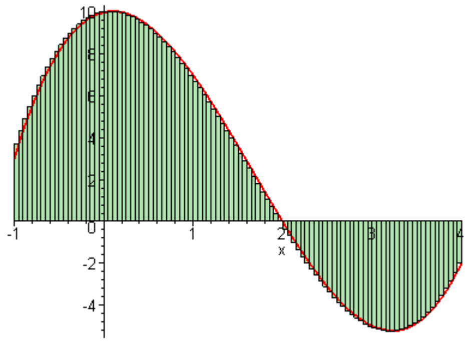


Figure 5.12: An approximation of the area under the function $f(x) = x^3 - 5x^2 + x + 10$ between $x = -1$ and $x = 4$ using $n = 100$ rectangles.

Using $n = 1000$ rectangles, we obtain an approximation to the area $S_{1000}(-1, 4) = 12.9041562$, which is accurate to the first decimal.

In the long run, when n grows really large, the Riemann sum approximations will get better and better and approach the true value of the area under the curve. Imagine cutting the region into $n = 10000$ rectangles; isn't $S_{10000}(-1, 4)$ a pretty accurate approximation of the actual area $A(-1, 4)$?

The integral as a limit

In the limit as the number of rectangles n approaches ∞, the Riemann sum approximation to the area under the curve becomes *arbitrarily close* to the true area:

$$\lim_{n\to\infty} \sum_{k=1}^{n} f(a + k\Delta x)\Delta x = A(a, b).$$

The definite integral between $x = a$ and $x = b$ is *defined* as the limit of a Riemann sum as n goes to infinity:

$$\int_a^b f(x)\, dx \equiv \lim_{n\to\infty} \sum_{k=1}^{n} f(a + k\Delta x)\Delta x \equiv A(a, b).$$

Perhaps now the weird notation we use for integrals will start to make more sense to you. An integral is, literally, the sum of the function at the different sample points! In the limit as $n \to \infty$, the summation sign $\sum$ becomes an integral sign $\int$, and the step size Δx becomes an ~~infinitesimally~~ infinitely small step dx.

It is not computationally practical to make $n \to \infty$; we can simply stop at some finite n which produces the desired accuracy of approximation. The approximation using 1 million rectangles is accurate to the fourth decimal place, which you can verify by entering the following commands on `live.sympy.org`:

Formal definition of the integral

We rarely compute integrals using Riemann sums. The Riemann sum is a *theoretical construct* like the rise-over-run calculation that we use to define the derivative operation:

$$f'(x) = \lim_{\delta \to 0} \frac{f(x+\delta) - f(x)}{\delta} .$$

The integral is defined as the approximation of the area under the curve with infinitely many rectangles:

$$\int_a^b f(x)\, dx \equiv \lim_{n \to \infty} \sum_{k=1}^{n} f(a + k\Delta x)\Delta x, \qquad \Delta x = \frac{b-a}{n} .$$

It is usually much easier to refer to a table of derivative formulas (see page 310) rather than compute a derivative starting from the formal definition and taking the limit $\delta \to 0$. Similarly, it is easier to refer to a table of integral formulas (also see page 310), rather than computing the integral by taking the limit as $n \to \infty$ of a Riemann sum.

Now that we have established a formal definition of the integral, we'll be able to understand why integral formulas are equivalent to derivative formulas applied in the opposite direction. In the next section we'll give a formal proof of the inverse relationship between the derivative operation and the integral operation.

Links

[A Riemann sum demonstration]
`http://www.geogebratube.org/student/m68523`

[Riemann sum wizard]
`http://mathworld.wolfram.com/RiemannSum.html`

5.14 The fundamental theorem of calculus

In Section 5.12 we defined the integral function $A_0(x)$ that corresponds to the calculation of the area under $f(x)$ starting from $x = 0$:

$$A_0(x) \equiv \int_0^x f(\underline{u}t)\, \underline{du}dt.$$

Plenty of integrals have no *closed-form solution*, meaning the function has no antiderivative. There is no simple procedure to follow such that you input a function and "turn the crank" until the integral comes out. Integration is a bit of an art.

Which functions *can* we integrate, and how? Back in the day, scientists collected big tables with integral formulas for various complicated functions. We can use these tables to *look up* a specific integral formula. Such table is given on page 481 in the back of the book.

We can also learn some *integration techniques* to help make complicated integrals simpler. Think of the techniques presented in this section as *adapters*. You can reach for these adapters when the function you need to integrate doesn't appear in your table of integrals, but a similar one is found in the table.

A note to all our students in the audience who are taking an integral calculus course. These integration techniques are exactly the skills you'll be expected to demonstrate on the final. Instead of using the table of integrals to look up complicated integrals, you'll need to know how to fill in the table.

For people interested in learning physics, I'll honestly tell you that if you skip this next section you won't miss much. You should read the important section on *substitution*, but there's no need to read the details of all the recipes for integrating things. For most intents and purposes, once you understand what an integral is, you can use a computer to calculate it. A good tool for calculating integrals is the computer algebra system at `live.sympy.org`.

```
>>> integrate( sin(x) )
-cos(x)
>>> integrate( x**2*exp(x) )
x**2*exp(x) - 2*x*exp(x) + 2*exp(x)
```

You can use `SymPy` for all your integration needs.

A comment to those of you reading this book for ~~*general culture*~~fun, without the added stress of homework and exams. Consider the next dozen pages as an ethnographic snapshot of the daily life of the undergraduate experience in science. Try to visualize the life of first-year science students, busy integrating things they don't want to integrate for many, long hours. Picture some unlucky science student locked in her room, crunching calculus while hundreds of dangling integrals scream for attention, keeping her from hanging with friends.

Actually, it is not that bad. There are, like, four tricks to learn. If you **practice**, you can learn all of them in a week or so. Mastering these four tricks is essentially the purpose of the entire integral calculus ~~class~~course. If you understand the material in this section, you'll be done with integral calculus and you'll have two months to chill.

Substitution

Say you're integrating some complicated function that contains a square root $\sqrt{x}$. You wonder how to compute this integral:

$$\int \frac{1}{x-\sqrt{x}}\, dx \;=\; ?$$

Sometimes you can simplify an integral by *substituting* a new variable into the expression. Let $u = \sqrt{x}$. Substitution is like search-and-replace in a word processor. Every time you see the expression $\sqrt{x}$, replace it with u:

$$\int \frac{1}{x-\sqrt{x}}\, dx = \int \frac{1}{u^2-u}\, dx.$$

Note we also replaced $x = (\sqrt{x})^2$ with u^2.

We're not done yet. To change from the x variable to the u variable, we must also change dx to du. Can we simply replace dx with du? Unfortunately no, otherwise it would be like saying the "short step" du is equal in length to the "short step" dx, which is only true for the trivial substitution $u = x$.

To find the relation between the ~~infinitesimals~~small step du and the small step dx, we take the derivative:

$$u(x) = \sqrt{x} \quad \Rightarrow \quad u'(x) = \frac{du}{dx} = \frac{1}{2\sqrt{x}}.$$

For the next step, I need you to stop thinking about the expression $\frac{du}{dx}$ as a whole, and instead think about it as a rise-over-run fraction that can be split. Let's move the *run* dx to the other side of the equation:

$$du = \frac{1}{2\sqrt{x}}\, dx.$$

Next, to isolate dx, multiply both sides by $2\sqrt{x}$:

$$dx = 2\sqrt{x}\, du = 2u\, du,$$

where we use the fact that $u = \sqrt{x}$ in the last step.

We now have an expression for dx expressed entirely in terms of the variable u. After the substitution, the integral looks like

$$\int \frac{1}{x-\sqrt{x}}\, dx = \int \frac{1}{u^2-u} 2u\ 2u\, du = \int \frac{2}{u-1}\, du.$$

We can recognize the general form of the function inside the integral, $f(u) = \frac{2}{u-1}$, to be similar to the function $f(u) = \frac{1}{u}$. Recall that the integral of $\frac{1}{u}$ is $\ln(u)$. Accounting for the -1 horizontal shift and the factor of 2 in the numerator, we obtain the answer:

$$\int \frac{1}{x - \sqrt{x}}\, dx = \int \frac{2}{u-1}\, du = 2\ln(u-1) = 2\ln(\sqrt{x}-1).$$

Note in the last step, we changed back to the x variable to give the final answer. The variable u exists only in our calculation. We invented it out of thin air when we said, "Let $u = \sqrt{x}$" in the beginning.

Thanks to the substitution, the integral ~~becomes simplersince we~~ became simpler: we were able to eliminate the square roots. The extra u that ~~comes~~ came from the expression $dx = 2u\, du$ ~~cancels~~ canceled with one of the us in the denominator, ~~making things~~ thus making the expression even simpler. In practice, substituting x with u inside f is the easy part. The hard part is making sure our choice of substitution leads to a replacement for dx that helps to simplify the integral.

For definite integrals—that is, integrals with limits of integration—there is an extra step we need to take when changing variables: we must change the x-limits of integration to u-limits. In our expression, when changing to the u variable, we write

$$\int_a^b \frac{1}{x - \sqrt{x}}\, dx = \int_{u(a)}^{u(b)} \frac{2}{u-1}\, du.$$

Say we are asked to compute the definite integral between $x = 4$ and $x = 9$ for the same expression. In this case, the new limits are $u = \sqrt{4} = 2$ and $u = \sqrt{9} = 3$, and we have

$$\int_4^9 \frac{1}{x - \sqrt{x}} dx = \int_2^3 \frac{2}{u-1} du = 2\ln(u-1)\Big|_2^3 = 2(\ln(2) - \ln(1)) = 2\ln(2).$$

Let's recap. Substitution involves three steps:

1. Replace all occurrences of $u(x)$ with u.
2. Replace dx with $\frac{1}{u'(x)} du$.
3. If there are limits, replace the x-limits with u-limits.

If the resulting integral is simpler to solve, then good for you!

Example Find $\int \tan(x)\,dx$. We know $\tan(x) = \frac{\sin(x)}{\cos(x)}$, so we can use the substitution $u = \cos(x)$, $du = -\sin(x)dx$ as follows:

$$\begin{aligned}
\int \tan(x)dx &= \int \frac{\sin(x)}{\cos(x)}dx \\
&= \int \frac{-1}{u}du \\
&= -\ln|u| + C \\
&= -\ln|\cos(x)| + C.
\end{aligned}$$

Integrals of trig functions

Because sin, cos, tan, and the other trig functions are related, we can often express one function in terms of another in order to simplify integrals.

Recall the trigonometric identity,

$$\cos^2(x) + \sin^2(x) = 1,$$

which is the statement of Pythagoras' theorem.

If we choose to make the substitution $u = \sin(x)$, we can replace all kinds of trigonometric terms with the new variable u:

$$\begin{aligned}
\sin^2(x) &= u^2, \\
\cos^2(x) &= 1 - \sin^2(x) = 1 - u^2, \\
\tan^2(x) &= \frac{\sin^2(x)}{\cos^2(x)} = \frac{u^2}{1-u^2}.
\end{aligned}$$

Of course the change of variable $u = \sin(x)$ means you must also change the ~~$du = u'(x)dx = \cos(x)dx$~~ integration step $du = \cos(x)dx \;\Rightarrow\; dx = \frac{1}{\cos(x)}$ so there better be ~~something to cancel this~~ an extra $\cos(x)$ ~~term~~ factor in the integral to cancel this $\frac{1}{\cos(x)}$.

Let me show you one example where things work perfectly. Suppose m is some arbitrary number and you need to integrate:

$$\int (\sin(x))^m \cos^3(x)\,dx \equiv \int \sin^m(x)\cos^3(x)\,dx.$$

This integral contains m powers of the sin function and three powers of the cos function. Let us split the cos term into two parts:

$$\int \sin^m(x)\cos^3(x)\,dx = \int \sin^m(x)\cos^2(x)\cos(x)\,dx.$$

Making the change of variables ($u = \sin(x)$ and $du = \cos(x)dx$) means we can replace $\sin^m(x)$ by u^m, and $\cos^2(x) = 1 - u^2$ in the above expression to obtain

$$\int \sin^m(x)\cos^2(x)\cos(x)\;dx = \int u^m\left(1-u^2\right)\cos(x)\;dx.$$

Conveniently, we happen to have $dx = \frac{1}{\cos(x)}\;du$, so the complete change-of-variable step is

$$\int \sin^m(x)\cos^2(x)\cos(x)\;dx = \int u^m\left(1-u^2\right)\;du.$$

This is what I was talking about earlier when I mentioned "having an extra $\cos(x)$" to cancel the one that appears as a result of the $dx \to du$ change.

What is the answer then? It is a simple integral of a polynomial:

$$\begin{aligned}\int u^m\left(1-u^2\right)\;du &= \int \left(u^m - u^{m+2}\right)\;du\\ &= \frac{1}{m+1}u^{m+1} - \frac{1}{m+3}u^{m+3}\\ &= \frac{1}{m+1}\sin^{m+1}(x) - \frac{1}{m+3}\sin^{m+3}(x).\end{aligned}$$

You might be wondering how useful this substitution technique actually is. I mean, how often will you need to integrate such particular combinations of sin and cos powers, where substitution works perfectly? You might be surprised! Sin and cos functions are used often in this thing called the *Fourier transform*, which is a way of expressing a sound wave $f(t)$ in terms of the frequencies it contains.

Also, integrals with trig functions are known favourites of teachers to ask on exams. A trig substitution question will test if you can perform substitutions, *and* teachers use them to check whether you remember all the trigonometric identities (page 80), which you are supposed to have learned in high school.

Are there other trig substitution tricks you should know about? On an exam, you should try any possible substitution you can think of, combined with any trigonometric identity that seems to simplify things. Some common substitutions are described below.

Cos

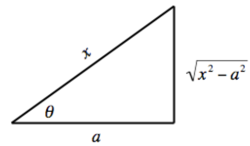
Just as we can substitute sin, we can also substitute $u = \cos(x)$ and use $\sin^2(x) = 1 - u^2$. Again, this substitution only makes sense when there is a leftover sin somewhere in the integral that can cancel with the sin in $dx = \frac{-1}{\sin x}\;du$.

Recall the trigonometric identity $1 + \tan^2\theta = \sec^2\theta$ or, rewritten differently,

$$\sec^2\theta - 1 = \tan^2\theta.$$

The appropriate substitution for terms like $\sqrt{x^2 - a^2}$ is

$$x = a\sec\theta, \qquad dx = a\tan\theta\sec\theta\, d\theta.$$



The substitution procedure is the same as in previous cases of the sin substitution and the tan substitution we discussed, so we won't elaborate here in detail. We can label the sides of the triangle accordingly, as

$$\sec\theta = \frac{x}{a} = \frac{\text{hyp}}{\text{adj}}.$$

We'll use this triangle when converting back from θ to x in the final steps of the integral calculation.

Interlude

By now, things are starting to get pretty tight for your calculus teacher. You are beginning to understand how to "handle" any kind of integral he can throw at you: polynomials, fractions with x^2, plus or minus a^2, and square roots. He can't even fool you with dirty trigonometric tricks involving sin, cos, and tan, since you know about these, too. Are there any integrals left that he can drop on the exam to trick you up?

Substitution is the most important integration technique. Recall the steps involved: (1) the choice of substitution $u = \ldots$, (2) the associated dx to du change, and (3) the change in the limits of integration required for definite integrals. With medium to advanced substitution skills, you'll score at least an 80% on your integral calculus final.

What will the remaining 20% of the exam depend on? How many more techniques could there possibly be? I know all these ~~integrations~~ integration techniques that I've been throwing at you during the last ten pages may seem arduous and difficult to understand, but this is what you got yourself into when you signed up for the course "integral calculus." In this course, there are lots of *integrals* and you *calculate* them.

The good news is that we are almost done. Only one more "trick" remains, and afterward, I'll finally tell you about the *integration by parts* procedure, which is very useful.

Don't bother memorizing the steps in each ~~problem. The~~ of the examples discussed: the correct substitution of $u = \ldots$ will be different in each problem. Think of integration techniques as general recipe guidelines you must adapt based on the ingredients available to you at the moment of cooking. ~~You can always return to this section when~~ When faced with a complicated integral ~~; check to see which of this section 's examples looks the most similar and follow~~ problem, you can always return to this section and find the example that is most similar to your problem and use the same approach.

Partial fractions

Suppose you need to integrate a rational function $\frac{P(x)}{Q(x)}$, where P and Q are polynomials.

For example, you could be asked to integrate

$$\frac{P(x)}{Q(x)} = \frac{ax+b}{cx^2+dx+e},$$

where a, b, c, d, and e are arbitrary constants. To get even more specific, let's say you are asked to calculate

$$\int \frac{3x+1}{x^2+x}\,dx.$$

By magical powers, I can transform the function in this integral into two *simple fractions*:

$$\int \frac{3x+1}{x^2+x}\,dx \qquad = \qquad \int \frac{1}{x}\,dx \quad + \quad \int \frac{2}{x+1}\,dx.$$

We split the complicated-looking rational expression into two partial fractions.

Now that the hard part is done, all that remains is to compute the two integrals. Recall that $\frac{d}{dx}\ln(x) = \frac{1}{x}$, so the integrals will give ln-like terms. The final answer is

$$\int \frac{3x+1}{x^2+x}\,dx = \ln|x| + 2\ln|x+1| + C.$$

How did I split the problem into partial fractions? Is it really magic or is there a method? The answer is, a little bit of both. My method was to *assume* the existence of constants A and B such that

$$\frac{3x+1}{x^2+x} = \frac{3x+1}{x(x+1)} = \frac{A}{x} + \frac{B}{x+1}.$$

Then I solved the above equation for A and B by computing the sum of the two fractions:

$$\frac{3x+1}{x(x+1)} = \frac{A(x+1)+Bx}{x(x+1)}.$$

- Repeated factors, like $(x-\beta)^n$, for which we assume the existence of n different terms on the right-hand side:
$$\frac{B}{x-\beta}+\frac{C}{(x-\beta)^2}+\cdots+\frac{F}{(x-\beta)^n}.$$
- If one of the factors is a polynomial ax^2+bx+c that cannot be factored, such as x^2+1, we must preserve this portion as a whole and assume that a term of the form
$$\frac{Gx+H}{ax^2+bx+c}$$
exists on the right-hand side.

2. Add all the parts on the equation's right-hand side by first cross-multiplying each part in order to bring all fractions to a common denominator, and then adding the fractions together. If you followed the steps correctly in Part 1, the ~~least common denominator (LCD)~~ common denominator will turn out to be $Q(x)$, and both sides will have the same denominator. Solve for the unknown coefficients $A, B, C, \ldots$ in the numerators. Find the coefficients of each power of x on the right-hand side and set them equal to the corresponding coefficient in the numerator $P(x)$ of the left-hand side.
3. Use the appropriate integral formula for each kind of term:
   - For simple factors, use
$$\int \frac{1}{x-\alpha}\,dx = A\ln|x-\alpha|+C.$$
   - For higher powers in the denominator, use
$$\int \frac{1}{(x-\beta)^m}\,dx = \frac{1-m}{(x-\beta)^{m-1}}+C.$$
   - For the quadratic denominator terms with "matching" numerator terms, use
$$\int \frac{2ax+b}{ax^2+bx+c}\,dx = \ln|ax^2+bx+c|+C.$$
For quadratic terms with only a constant in the numerator, use a two-step substitution process. First, change x to the complete-the-square variable $y=x-h$:
$$\int \frac{1}{ax^2+bx+c}\,dx = \int \frac{1/a}{(x-h)^2+k}\,dx = \frac{1}{a}\int \frac{1}{y^2+k}\,dy.$$
Then apply a trig substitution $y=\sqrt{k}\tan\theta$ to obtain
$$\frac{1}{a}\int \frac{1}{y^2+k}\,dy = \frac{\sqrt{k}}{a}\tan^{-1}\!\left(\frac{y}{\sqrt{k}}\right) = \frac{\sqrt{k}}{a}\tan^{-1}\!\left(\frac{x-h}{\sqrt{k}}\right).$$

Example Find $\int \frac{1}{(x+1)(x+2)^2}dx$.

Here, $P(x) = 1$ and $Q(x) = (x+1)(x+2)^2$. If I wanted to be sneaky, I could have asked for $\int \frac{1}{x^3+5x^2+8x+4}dx$ instead—which is the same question, but you'd need to do the factoring yourself.

According to the recipe outlined above, we must look for a split fraction of the form

$$\frac{1}{(x+1)(x+2)^2} = \frac{A}{x+1} + \frac{B}{x+2} + \frac{C}{(x+2)^2}.$$

To make the equation more explicit, let's add the fractions on the right. Set all of them to the least common denominator and add:

$$\begin{aligned}
\frac{1}{(x+1)(x+2)^2} &= \frac{A}{x+1} + \frac{B}{x+2} + \frac{C}{(x+2)^2} \\
&= \frac{A(x+2)^2}{(x+1)(x+2)^2} + \frac{B(x+1)(x+2)}{(x+1)(x+2)^2} + \frac{C(x+1)}{(x+1)(x+2)^2} \\
&= \frac{A(x+2)^2 + B(x+1)(x+2) + C(x+1)}{(x+1)(x+2)^2}.
\end{aligned}$$

The denominators are the same on both sides of the above equation, so we can focus our attention on the numerator:

$$A(x+2)^2 + B(x+1)(x+2) + C(x+1) = 1.$$

We can evaluate this equation for three different values of x to find the values of A, B, and C:

$$\begin{array}{cc}
x = 0 & 1 = 2^2A + 2B + C \\
x = -1 & 1 = A \\
x = -2 & 1 = -C
\end{array}$$

so $A = 1$, $B = -1$, and $C = -1$. Thus,

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

We can now calculate the integral by integrating each of the terms:

$$\int \frac{1}{(x+1)(x+2)^2}dx = \ln(x+1) - \ln(x+2) + \frac{1}{x+2} + C.$$

The partial fractions technique for integrating rational functions is best understood using using a hands-on approach. Try solving the following exercises to see if you can apply the techniques.

The gravitational potential energy of two planets ~~a~~ separated by distance R ~~apart~~ is negative since work is *required* to pull the planets apart once they have come together.

There is an important physics lesson to learn here. For each conservative force $\vec{F}_?(x)$, there is an associated potential energy function $U_?(x)$ that is defined as the negative of the work done when moving an object against the force $\vec{F}_?(x)$:

$$\text{Given } \vec{F}_?(x) \qquad \Rightarrow \qquad U_?(x) \equiv -\int_0^x \vec{F}_?(u) \cdot d\vec{u}.$$

We can use this relationship in the other direction too. Given a potential energy function $U_?(x)$, we can find the force $F_?(x)$ associated with that potential energy function by taking the derivative:

$$\text{Given } U_?(x) \qquad \Rightarrow \qquad F_?(x) \equiv -\frac{d}{dx}\Big[U_?(x)\Big].$$

The negative of the derivative of the gravitational potential energy $U_g(y) = mgy$ gives $F_g(y) = -mg$. The negative of the derivative of the spring potential energy $U_s(x) = \frac{1}{2}kx^2$ gives $F_s(x) = -kx$.

Integrals over circular objects

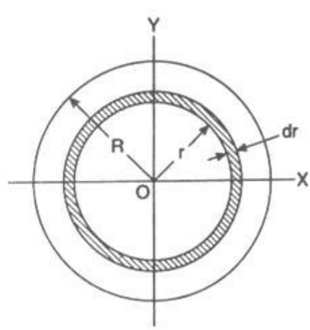
We can use integration to ~~calculate the~~ obtain area and volume formulas of objects with circular symmetries.

Consider the disk-shaped region described by the equation $D = \{x, y \in \mathbb{R} \mid x^2 + y^2 \leq R^2\}$. In polar coordinates we describe this region as $r \leq R$, where it is implicit that the angle θ varies between 0 and 2π. Because this region is two-dimensional, computing an integral over this region requires a *double integral*, which is the subject of multivariable calculus. Even before you learn about double integrals, you still know enough to integrate over a circular region by breaking the region into thin circle-shaped slices of disk dD.

Similar to the way a horizontal slice through an onion consists of many thin onion rings, we can break the disk into a number of thin circular strips of width dr. The circular strip with radius r has an area of

$$dD = 2\pi r \, dr,$$

since $2\pi r$ is the circumference of a circle with radius r, and since the width of the strip is dr.



We can perform the integral over the whole disk by adding the contributions of all the strips:

$$I_{\text{disk}} = \int_0^R r^2\, dm = \int_0^R r^2 \sigma 2\pi r\, dr = \int_0^R r^2 \frac{m}{\pi R^2} 2\pi r\, dr =$$

$$= \frac{2m}{R^2}\int_0^R r^3\, dr = \frac{2m}{R^2}\left[\frac{r^4}{4}\right]_0^R = \frac{2m}{R^2}\frac{R^4}{4} = \frac{1}{2}mR^2.$$

Arc length of a curve

Given a function $y = f(x)$, how can you calculate the total ~~*length* ℓ~~ *arc length* of the graph of $f(x)$ between $x = a$ and $x = b$?

If $f(x)$ is the equation of a line, the length of its graph can be calculated as the hypotenuse of the change-in-x and the change-in-y triangle: $\ell = \sqrt{\text{run}^2 + \text{rise}^2} = \sqrt{(b-a)^2 + (f(b) - f(a))^2}$.

However, if the function is *not* a straight line, we need to apply this hypotenuse calculation to each piece of the curve $d\ell = \sqrt{dx^2 + dy^2}$, then ~~add~~ sum all the contributions ~~as an integral~~using integration $\ell = \int d\ell$.

The arc length ℓ of ~~a graph $y = f(x)$ on the interval $x \in [a, b]$~~ the function $f(x)$ between $x = a$ and $x = b$ is

$$\ell = \int d\ell = \int \sqrt{dx^2 + dy^2} = \int \sqrt{\left(1 + \frac{dy^2}{dx^2}\right) dx^2} = \int_a^b \sqrt{1 + \left(f'(x)\right)^2}\, dx.$$

Example Use the arc length formula to compute the length of the curve $f(x) = \frac{2}{3}\sqrt{x^3}$ between the points $(1, \frac{2}{3})$ and $(6, 4\sqrt{6})$.

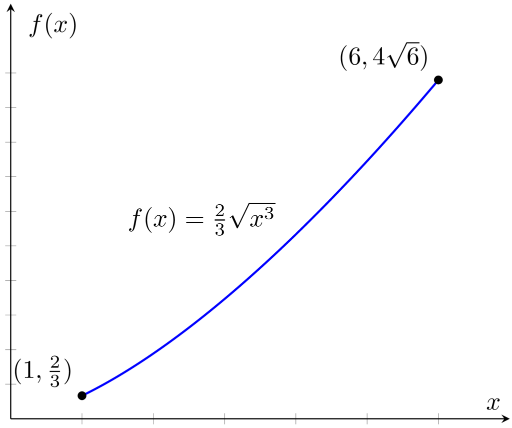


Figure 5.14: Graph of the function $f(x) = \frac{2}{3}x^{\frac{3}{2}}$ between $x = 1$ and $x = 6$.

The formula for length of a curve is $\ell = \int d\ell = \int_1^6 \sqrt{1 + (f'(x))^2}\, dx$. Applying this formula to the function $f(x)$ we obtain $f'(x) = \sqrt{x}$ then find $\sqrt{1 + (f'(x))^2} = \sqrt{1 + x}$. The length of the curve between $x = 1$ and $x = 6$ is

$$\ell = \int_1^6 \sqrt{1 + x}\, dx = \frac{2}{3}(x + 1)^{\frac{3}{2}}\Big|_1^6 = \frac{14}{3}\sqrt{7} - \frac{4}{3}\sqrt{2}.$$

Area of a surface of revolution

We ~~modify the arc-length~~ can modify the arc length formula to calculate the surface area A of a ~~solid of revolution. An object~~ *surface of revolution*. A surface with circular symmetry can be generated by a revolution of ~~some~~ a curve $f(x)$ around the x-axis.

~~Each piece of length $d\ell$ must be multiplied by $2\pi f(x)$, since each piece rotates around~~ Imagine a potter's wheel whose axis of rotation corresponds to the x-axis~~in a circle of radius~~. If the potter's fingers trace out the curve $f(x)$ in space, the result will be a circular vase whose sides have the shape $f(x)$.

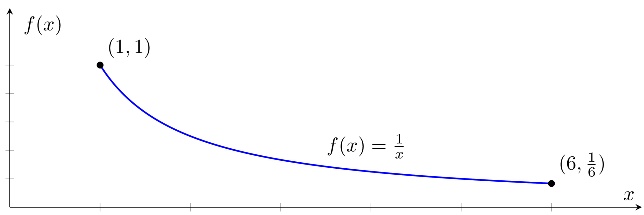


Figure 5.15: Graph of $f(x) = \frac{1}{x}$ between $x = 1$ and $x = 6$. What surface will be produced if we rotate this curve around the x-axis?

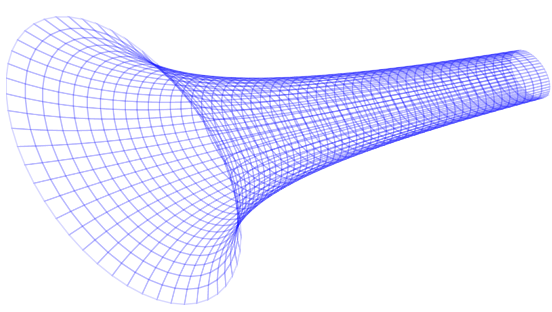


Figure 5.16: The surface of revolution generated by the section of the curve $f(x) = \frac{1}{x}$ between $x = 1$ and $x = 6$ when rotated around the x-axis.

We can split up the surface area into circular strips by cutting the surface at regular intervals dx along the x-axis. The radius of each strip varies according to $f(x)$ and the width of each strip is given by the arc length $d\ell = \sqrt{dx^2 + dy^2} = \sqrt{1 + (f'(x))^2}dx$. We can approximate the area of each circular strip as $2\pi f(x)d\ell$.

The area of the surface of revolution traced by the graph of $f(x)$ between $x = a$ and $x = b$ as it rotates around the x-axis is given by the ~~following integral:~~ formula

$$A = \int_a^b 2\pi f(x) d\ell = \int_{\cancel{x_i}\,a}^{\cancel{x_f}\,b} 2\pi f(x)\ \sqrt{1 + (f'(x))^2}\ dx.$$

Volumes of revolution

Let's move on to three-dimensional integrals, or integrals over volumes. Again, we'll use the circular symmetry of the object's volume to split the object into little "pieces of volume" and then compute an integral to find the total volume.

Disk method

We can describe the volume of an object with circular symmetry as the sum of a number of disks. Each disk will have thickness dx and a radius proportional to the function $f(x)$. In other words, the function $f(x)$ describes the object's outer boundary. The area of each disk is $\pi(f(x))^2$ and its thickness is dx.

The volume of a solid of revolution with boundary $f(x)$ rotated around the x-axis is given by the formula

$$V = \int A_{\text{disk}}\, dx = \int \pi(f(x))^2\, dx.$$

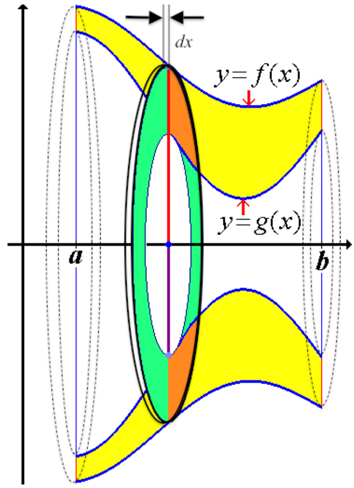
$$\begin{aligned} V &= \int A_{\text{disk}}(x)\, dx \\ &= \int \pi(f(x))^2\, dx. \end{aligned}$$

Example Use the disk method to calculate the volume of a sphere with radius R. The volume we want to calculate is bounded by the curve $f(x) = \sqrt{R^2 - x^2}$ and the horizontal line $y = 0$. Our limits of integration are the x-values where the curve intersects the line $y = 0$, namely, $x = \pm R$. We have

$$\begin{aligned} V_{\text{sphere}} &= \int_{-R}^{R} \pi(R^2 - x^2) dx \\ &= \pi\left(\int_{-R}^{R} R^2\, dx - \int_{-R}^{R} x^2\, dx\right) \\ &= \pi\left(R^2 x\Big|_{-R}^{R} - \frac{x^3}{3}\Big|_{-R}^{R}\right) \end{aligned}$$

Washer method

The washer method is a generalization of the disk method for computing volumes. Consider a volume of revolution with an inner radius described by the function $g(x)$ and an outer radius described by the function $f(x)$. The diagram on the right shows such a volume of revolution. Instead of using thin disks, we can represent this volume as the sum of thin washers, which are disks of radius $f(x)$ with a middle section of radius $g(x)$ removed.



The volume dV of each washer is equal to the outer area $\pi(f(x))^2$ minus the removed inner area $\pi(g(x))^2$, times the thickness dx. The total volume is given by

$$\begin{aligned} V &= \int dV \\ &= \int A_{\text{washer}}(x)\, dx \\ &= \int \Big(A_{\text{outer}}(x) - A_{\text{inner}}(x) \Big)\, dx \\ &= \int \pi \left[(f(x))^2 - (g(x))^2 \right] dx. \end{aligned}$$

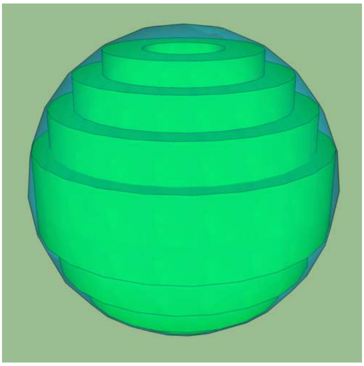
Cylindrical shell method

We can split any circularly symmetric volume into thin, cylindrical shells of thickness dr. If the volume has a circular symmetry and is bounded from above by $F(r)$ and from below by $G(r)$, then the integral over the volume will be

$$\begin{aligned} V &= \int C_{\text{shell}}(r)\, h_{\text{shell}}(r)\, dr \\ &= \int_a^b 2\pi r |F(r) - G(r)|\, dr. \end{aligned}$$

The cylindrical shell with radius r has circumference $2\pi r$, thickness dr, and its height is described by the expression $|F(r) - G(r)|$.

Example Calculate the volume of a sphere of radius R using the cylindrical shell method. We are talking about the region enclosed by the surface $x^2 + y^2 + z^2 = R^2$.



At radius $r = \sqrt{x^2+y^2}$, the cylindrical shell will be bounded from above by $z = F(r) = \sqrt{R^2 - r^2}$, and bounded from below by $z = G(r) = -\sqrt{R^2-r^2}$. The circumference of the shell is $2\pi r$ and its width is dr. The integral proceeds like this:

$$\begin{aligned} V &= \int_0^R 2\pi r|F(r) - G(r)|\, dr \\ &= \int_0^R 2\pi r 2\sqrt{R^2-r^2}\, dr \\ &= -2\pi \int_{R^2}^0 \sqrt{u}\, du \qquad \text{(using } u = R^2 - r^2, du = -2rdr) \\ &= -2\pi \left[\tfrac{2}{3}u^{3/2}\right]_{R^2}^0 \\ &= -2\pi[0 - \tfrac{2}{3}R^3] \\ &= \frac{4}{3}\pi R^3. \end{aligned}$$

Discussion

The formulas for calculating the surface areas and volumes of revolutions can be adapted for computing other quantities that have circular symmetry. For example, the formula for the surface area of a revolution can be used to compute the moment of inertia of a thin spherical shell. The formula for the surface are of a spherical shell of radius R can be obtained as the area of revolution for the curve $f(x) = \sqrt{R^2 - x^2}$ around the x-axis:

$$A_{\text{sph.shell}} = \int_{-R}^{R} dA = \int_{-R}^{R} 2\pi f(x)\sqrt{1 + (f'(x))^2}\, dx = 4\pi R^2.$$

We can adapt this formula to calculate the moment of inertia of a spherical shell of mass m. The general formula for the moment of inertia is $I = \int r^2\, dm$, which tells us the contribution of each "piece of mass" dm must be multiplied by its squared distance from the axis of rotation. For a spherical shell, each "piece of mass" can be written as $dm = \sigma dA$, where $\sigma = \frac{m}{4\pi R^2}$ is the mass density. The distance of the pieces from the rotation axis varies according to $f(x)$. The moment of inertia of a thin spherical shell is therefore given by the integral

$$I_{\text{sph.shell}} = \int_{-R}^{R} (f(x))^2 \sigma dA = \int_{-R}^{R} (f(x))^2\, \sigma 2\pi f(x) \sqrt{1 + (f'(x))^2}\, dx.$$

In problem **P5.126** you'll be asked to complete this calculation and verify the formula $I_{\text{sph.shell}} = \frac{2}{3}mR^2$, which was stated in Section 4.8 without proof.

A similar approach can be used to calculate the moment of inertia of a solid sphere $I_{\text{sphere}} = \frac{2}{5}mR^2$. We can use either the disk method or the cylindrical shell method, adapting the volume of revolution formula to transform each piece of volume dV into a piece of mass dm, multiplied by its squared distance from the axis of rotation. See **P5.127** and **P5.128** for the calculations.

Exercises

E5.5 Calculate the volume of a cone with radius R and height h that is generated by the revolution around the x-axis of the region bounded by the curve $y = R - \frac{R}{h}x$ and the lines $y = 0$ and $x = 0$.

5.5 $\frac{\pi R^2 h}{3}$.

E5.6 Find the volume of a vertical cone with radius R and height h formed by the revolution of the region bounded by the curves $y = 0$, $y = h - \frac{h}{R}x$, and $x = 0$ around the y-axis. Use cylindrical shells.

5.6 $\frac{\pi R^2 h}{3}$.

Links

[An animation showing how a volume of revolution is constructed]
`http://mathforum.org/mathimages/index.php/Volume_of_Revolution`

5.17 Improper integrals

Imagine you want to find the area under the function $f(x) = \frac{1}{x^2}$ from $x = 1$ all the way to $x = \infty$. This kind of calculation is known as an *improper integral* since one of the endpoints of the integration is not a regular number, but infinity.

Geometric sequence

By using the index as the exponent of a fixed number r, we obtain the geometric series

$$a_n = r^n, \quad n \in \mathbb{N},$$

which is a sequence of the form

$$\left(1, r, r^2, r^3, r^4, r^5, r^6, \ldots\right).$$

If we choose $r = \frac{1}{2}$, then the geometric series with this ratio will be

$$\left(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \ldots\right).$$

Fibonacci

The Fibonacci numbers are constructed according to the following pattern. The first Fibonacci number is 0, the second Fibonacci number is 1, and each subsequent number is the sum of the two preceding it:

$$a_0 = 0, a_1 = 1, \qquad a_n = a_{n-1} + a_{n-2}, \quad n \geq \gtrsim 2.$$

$$(0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, \ldots).$$

Convergence

We say a sequence a_n *converges* to a limit L, written mathematically as

$$\lim_{n\to\infty} a_n \;=\; L,$$

if for large values of n the terms in the sequence become arbitrarily close to the value L.

More precisely, the limit expression $\lim\limits_{n\to\infty} a_n \;=\; L$ means that for *any* precisions $\epsilon > 0$, we can pick a number N_ϵ such that

$$|a_n - L| < \epsilon, \qquad \forall n \geq N_\epsilon.$$

The notion of a limit of a sequence is the same as the notion of a limit of a function. Just as we learned how to calculate which number the function $f(x)$ tends to for large x values (see page 295), we can study which number the sequence a_n tends to for large n values.

We'll now state without proof a number of other formulas where the sum of a series can be obtained as a closed-form expression.

The formulas for the sum of the first N positive integers and the sum of the squares of the first N positive integers are

$$\sum_{n=1}^{N} n = \frac{N(N+1)}{2}, \qquad \sum_{n=1}^{N} n^2 = \frac{N(N+1)(2N+1)}{6}.$$

See problem **P5.119** for the derivations of these formulas. The sum of the first ~~$N+1$~~ N terms in an arithmetic sequence is

$$\sum_{\cancel{n=0}\,n=1}^{N}(a_0+nd) = a_0\cancel{(}N+\cancel{1)+}\frac{N(N+1)}{2}d.$$

It will be important to remember these formulas because they can come up in calculus problems. For example, computing the integral of the function $f(x) = ax^2 + bx + c$ using an infinite Riemann sum requires these formulas.

~~Another group of series with exact formulas for their sums are the~~ There are many other series whose infinite sum is described by an exact formula. Below we state some known formulas for the sums of certain infinite series. The p-series involving even values of p ~~:~~ can be computed:

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}, \qquad \sum_{n=1}^{\infty}\frac{1}{n^4} = \frac{\pi^4}{90}, \qquad \sum_{n=1}^{\infty}\frac{1}{n^6} = \frac{\pi^6}{945}.$$

Note it's not required for you to memorize these formulas. They are given here as examples of what can be done. Other closed-form ~~sums include :~~ expressions for infinite series include

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12}, \qquad \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n} = \ln(2), \qquad \sum_{n=1}^{\infty}\frac{1}{4n^2-1} = \frac{1}{2},$$

$$\sum_{n=0}^{\infty}\frac{(-1)^{n}}{2n+1} = \frac{\pi}{4}, \qquad \sum_{n=0}^{\infty}\frac{1}{(2n+1)^2} = \frac{\pi^2}{8}, \qquad \sum_{n=0}^{\infty}\frac{(-1)^{n}}{(2n+1)^3} = \frac{\pi^3}{32}.$$

Wow, that's a lot of formulas! Don't worry about memorizing all these formulas; just think of them as a "trophy case" that illustrates some mathematical success stories. Mathematicians are really proud when they manage to make sense of some complicated, infinite sum expression, by finding a simple formula to describe its value. In general most infinite series do not have such closed-form expressions, so you can understand mathematicians' excitement and their desire to build a trophy case of known formulas. The series formulas shown above are analogous to the "trophy case" of integral formulas on page 481.

The number N_ϵ corresponds to how many terms of the series you need for the partial sum S_N to become ϵ-close to the limit L.

Convergence tests

The main thing you need to know about series are the different *tests* you can perform to check whether a series converges or diverges. ~~We'll now discuss a number of these tests.~~

Divergence test

The only way the infinite sum $\sum_{n=0}^{\infty} a_n$ will converge is if the elements of the sequence a_n tend to zero for large n. This observation gives us a simple series *divergence test*. If $\lim_{n\to\infty} a_n \neq 0$ then $\sum_{n=0}^{\infty} a_n$ diverges. How could an infinite sum of non-zero quantities add to a finite number?

Absolute convergence

If $\sum_n |a_n|$ converges, $\sum_n a_n$ also converges. The opposite is not necessarily true, since the convergence of a_n might be due to negative terms *cancelling* with positive terms.

A sequence a_n for which $\sum_n |a_n|$ converges is called *absolutely convergent*. A sequence b_n for which $\sum_n b_n$ converges but $\sum_n |b_n|$ diverges is called *conditionally convergent*.

Decreasing alternating sequences

An alternating series a_n in which the absolute values of the terms is decreasing ($|a_n| > |a_{n+1}|$), and tend to zero ($\lim a_n = 0$) converges. For example, we know the series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots$ converges because it is a decreasing alternating series and $\lim_{n\to\infty} \frac{1}{n} = 0$.

Integral test

If the integral $\int_a^{\infty} f(x)\, dx$ is finite, then the series $\sum_n f(n)$ converges. If the integral $\int_a^{\infty} f(x)\, dx$ diverges, then the series $\sum_n f(n)$ also diverges.

The improper integral is defined as ~~the~~ a limit expression:

$$\int_a^{\infty} f(x)\, dx \equiv \lim_{b\to\infty} \int_a^b f(x)\, dx.$$

Discussion

You can think of the Taylor series as containing the "similarity coefficients" between $f(x)$ and the different powers of x. We choose the terms in the Taylor series of $f(x)$ to ensure the series approximation has the same n^{th} derivative as the function $f(x)$. For a Maclaurin series, the similarity between $f(x)$ and its power series representation is measured at $x = 0$, so the coefficients are chosen as $c_n = \frac{f^{(n)}(0)}{n!}$. The more general Taylor series allows us to build an approximation to $f(x)$ at any point $x = a$, ~~so the~~ and its similarity coefficients are calculated to match the derivatives at that point: $c_n = \frac{f^{(n)}(a)}{n!}$.

Another way of looking at the Maclaurin series is to imagine it is a kind of X-ray picture for each function $f(x)$. The zero$^{\text{th}}$ coefficient c_0 in the Maclaurin series tells you how much of the constant function is in $f(x)$. The first coefficient, c_1, tells you how much of the linear function x is in f; the coefficient c_2 tells you about the x^2 contents of f, and so on.

Now get ready for some crazy shit. I want you to go back to page 400 and take a careful look at the Maclaurin series of e^x, $\sin(x)$, and $\cos(x)$. As you will observe, it's as if e^x *contains* both $\sin(x)$ and $\cos(x)$, the only difference being the presence of the alternating negative signs. How about that? Do you remember Euler's formula $e^{ix} = \cos x + i \sin x$? Verify Euler's formula (page 177) by substituting ix into the power series for e^x.

Another interesting equation to think about in terms of series is $e^x = \cosh x + \sinh x$.

Links

[Animation showing Taylor series approximations to $\sin(x)$]
http://mathforum.org/mathimages/index.php/Taylor_Series

[Good summary with many interesting examples]
http://en.wikipedia.org/wiki/Series_(mathematics)

[A comprehensive list of important math series]
http://en.wikipedia.org/wiki/List_of_mathematical_series

5.20 Conclusion

Now you know how to take derivatives, calculate integrals, and find sums of infinite series. These practical skills will come in handy in the future, especially if you choose to pursue a career in science.

The exposure you had to formal math definitions prepared you for more advanced math classes. In particular, you learned how to deal with limits involving infinitely small quantities like ϵ and δ. Recall that both the derivative and the integral are defined as limit expressions. The derivative is defined as a rise-over-run calculation for an infinitely short run. The integral is defined as a Riemann sum with infinitely narrow rectangles.

~~The~~ We also learned important facts about sequences and series. Series teach us how to think about computations with an infinite number of steps. In particular the notion of a ~~limit is essentially the main new idea we learn in calculus . Limits allow us to talk about infinity.~~ Taylor series is a foundational idea for understanding functions.

Let's close the chapter by comparing the new procedures we learned in calculus with the type of math procedures used in high school math. Most of high school algebra tricks involve a finite number of steps: we start from a description of a problem, model it using an equation, and then solve this equation to obtain the answer using a couple of algebra steps. For example, we can solve the equation $x^2 - 2 = 0$ by rewriting it as $x^2 = 2$ and computing the square root of both sides of the equation to obtain $x = \pm\sqrt{2}$. Note it took us two steps of algebra to find the answer. A more complicated equation might require more steps, but in general, a finite number of steps are sufficient to solve most high school math problems.

In calculus we learn to solve a much broader class of problems. Using limits allows us to obtain answers computed by mathematical procedures with an infinite number of steps! For example, the answer to the question "what is the effective interest rate for a loan with a nominal interest rate of 100%, compounded infinitely often," is equal to the number $e = \lim_{n\to\infty}(1 + \frac{1}{n})^n$. If you borrow N dollars today, you~~will~~'ll owe Ne dollars ~~at the end of~~ after one year. Note the number e cannot be computed by any mathematical procedure with a finite number of steps. A limit is required to model the "infinitely frequent" compounding.

~~We also learned some facts about sequences and series. Series teach us how to think about computations with an infinite number of steps. In particular the notion of a Taylor series is a foundational idea for understanding functions.~~

Above all, the purpose of calculus is to solve problems. Speaking

5.21 Calculus problems

In this chapter we learned about derivatives and integrals, which are mathematical operations relating to the slope of a function and the area under the graph of a function. We also learned about limits, sequences, and series. It's now time to see how much you've really learned by trying to solve some calculus problems.

Calculus hasn't changed much in the last hundred years. It is testament to this fact that many of the problems presented here were adapted from the book "Calculus Made Easy" by Silvanus Thompson, originally published[1] in 1910. These problems remain as pertinent and interesting today as they were 100 years ago.

As much as calculus is about understanding things conceptually and seeing the big picture (abstraction), calculus is also about practice. There are more than ~~100~~ 120 problems to solve in this section. The goal is to turn differentiation and integration into routine operations that you can carry out without stressing out. You should vanquish as many problems as you need to feel comfortable with the procedures of calculus.

Okay, enough prep talk. Let's get to the problems!

Limits problems

P5.1 Use the graph of the function $f(x)$ shown in Figure 5.18 to calculate the following limit expressions:

(1) $\lim\limits_{x\to -5^-} f(x)$ (2) $\lim\limits_{x\to -5^+} f(x)$ (3) $\lim\limits_{x\to -5} f(x)$

(4) $\lim\limits_{x\to 2^-} f(x)$ (5) $\lim\limits_{x\to 2^+} f(x)$ (6) $\lim\limits_{x\to 2} f(x)$

(7) $\lim\limits_{x\to 5^-} f(x)$ (8) $\lim\limits_{x\to 5^+} f(x)$ (9) $\lim\limits_{x\to 5} f(x)$

(10) Is the function $f(x)$ continuous at $x = 5$?

(11) What are the intervals where the function $f(x)$ is continuous?

5.1 (1) -6. (2) 2. (3) Doesn't exist. (4) 8.6 (eyeballing it). (5) -5. (6) Doesn't exist. (7) -2. (8) -2. (9) -2. (10) No. (11)$[-10,-5)$, $[-5,2)$, $[2,5)$, $(5,10]$.

5.1 (10) For $x = 5$, we have $\lim_{x\to 5} f(x) \neq f(5)$ so the $f(x)$ is discontinuous at $x = 5$. This is called a removable discontinuity. (11) The function is continuous everywhere except for the discontinuities. The function $f(x)$ is continuous from the right at $x = -5$ and $x = 2$, so these endpoints are included in the intervals to the right.

P5.2 Find the value of the following limit expressions:

[1] Full text is available at `http://gutenberg.org/ebooks/33283` (public domain).

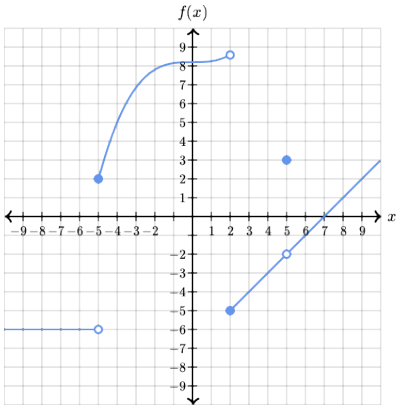


Figure 5.18: The graph of a piecewise-continuous function $f(x)$. The function $f(x)$ has two jump discontinuities at $x = -5$ and $x = 2$ and one removable discontinuity at $x = 5$.

P5.4 Calculate the limit if it exists, or explain why it doesn't.

(1) $\lim\limits_{x\to-\infty} \sin(x)$ (2) $\lim\limits_{x\to 0^+} \sin(x)$ (3) $\lim\limits_{x\to\infty} \sin(x)$

(4) $\lim\limits_{x\to-\infty} \sin(\frac{1}{x})$ (5) $\lim\limits_{x\to 0^+} \sin(\frac{1}{x})$ (6) $\lim\limits_{x\to\infty} \sin(\frac{1}{x})$

(7) $\lim\limits_{x\to-\infty} x\sin(\frac{1}{x})$ (8) $\lim\limits_{x\to 0^+} x\sin(\frac{1}{x})$ (9) $\lim\limits_{x\to\infty} x\sin(\frac{1}{x})$

Hint: Use the substitution $y = \frac{1}{x}$ to rewrite $\lim\limits_{x\to 0^+} \sin(\frac{1}{x})$ as $\lim\limits_{y\to\infty} \sin(y)$.

5.4 (1) Doesn't exist. (2) 0. (3) Doesn't exist. (4) 0. (5) Doesn't exist. (6) 0. (7) 1. (8) 0. (9) 1.

5.4 (8) We can prove $\lim_{x\to 0^+} x\sin(\frac{1}{x}) = 0$ using the squeezing principle. Observe that $-1 \leq \sin(\frac{1}{x}) \leq 1$, then $-x \leq x\sin(\frac{1}{x}) \leq x$. Since the limits of both the lower bound $\ell(x) = -x$ and the upper bound $u(x) = x$ are 0 as $x \to 0^+$, so is the limit of $x\sin(\frac{1}{x})$, which is squeezed between them.

P5.5 Calculate the following limit expressions:

(1) $\lim\limits_{x\to 1} \dfrac{x^2+5x+6}{x-1}$ (2) $\lim\limits_{x\to 1} \dfrac{x^2+x-2}{x-1}$ (3) $\lim\limits_{x\to a} \dfrac{x^2-a^2}{x-a}$

Hint: L'Hopital's rule might come in handy.

5.5 (1) Doesn't exist. (2) 3. (3) $2a$.

P5.6 Use a calculator to verify numerically the limits (1) through (6):

Derivatives problems

P5.7 Find the derivative with respect to x of the functions:

(1) $y = x^{13}$	(2) $y = x^{-\frac{3}{2}}$	(3) $y = x^{2a}$
(4) $u = t^{2.4}$	(5) $z = \sqrt[3]{u}$	(6) $y = \sqrt[3]{x^{-5}}$
(7) $u = \sqrt[5]{\frac{1}{x^8}}$	(8) $y = 2x^a$	(9) $y = \sqrt[q]{x^3}$

5.7 (1) $\frac{dy}{dx} = 13x^{12}$. (2) $\frac{dy}{dx} = -\frac{3}{2}x^{-\frac{5}{2}}$. (3) $\frac{dy}{dx} = 2ax^{(2a-1)}$. (4) $\frac{du}{dt} = 2.4t^{1.4}$. (5) $\frac{dz}{du} = \frac{1}{3}u^{-\frac{2}{3}}$. (6) $\frac{dy}{dx} = -\frac{5}{3}x^{-\frac{8}{3}}$. (7) $\frac{du}{dx} = -\frac{8}{5}x^{-\frac{13}{5}}$. (8) $\frac{dy}{dx} = 2ax^{a-1}$. (9) $\frac{dy}{dx} = \frac{3}{q}x^{\frac{3-q}{q}}$.

P5.8 Differentiate the following:

(1) $y = ax^3 + 6$	(2) $y = 13x^{\frac{3}{2}} - c$	(3) $y = 12x^{\frac{1}{2}} + c^{\frac{1}{2}}$
(4) $y = c^{\frac{1}{2}}x^{\frac{1}{2}}$	(5) $u = \frac{az^n - 1}{c}$	(6) $y = 1.18t^2 + 22.4$

5.8 (1) $\frac{dy}{dx} = 3ax^2$. (2) $\frac{dy}{dx} = 13 \times \frac{3}{2}x^{\frac{1}{2}}$. (3) $\frac{dy}{dx} = 6x^{-\frac{1}{2}}$. (4) $\frac{dy}{dx} = \frac{1}{2}c^{\frac{1}{2}}x^{-\frac{1}{2}}$. (5) $\frac{du}{dz} = \frac{an}{c}z^{n-1}$. (6) $\frac{dy}{dt} = 2.36t$.

P5.9 ~~Differentiate the~~ Take the derivative with respect to x of the following expressions:

(a) $u = 1 + x + \dfrac{x^2}{1 \cdot 2} + \dfrac{x^3}{1 \cdot 2 \cdot 3} + \dfrac{x^4}{1 \cdot 2 \cdot 3 \cdot 4} + \dfrac{x^5}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} + \cdots$

(b) $y = ax^2 + bx + c$ (c) $y = (x + a)^3$

5.9 (a) $1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \ldots$; (b) $2ax + b$; (c) $3x^2 + 6ax + 3a^2$.

P5.10 Use the product rule to find the following derivatives:

(1) If $w = t(a - \frac{1}{2}bt)$, find $\frac{dw}{dt}$.
(2) Find the derivative of $y = (x + \sqrt{-1})(x - \sqrt{-1})$.
(3) Differentiate $y = (197x - 34x^2)(7 + 22x - 83x^3)$.
(4) If $x = (y + 3)(y + 5)$, what is $\frac{dx}{dy}$?
(5) Differentiate $y = 1.3709x(112.6 + 45.202x^2)$.

5.10 (1) $\frac{dw}{dt} = a - bt$. (2) $\frac{dy}{dx} = 2x$. (3) $14110x^4 - 65404x^3 - 2244x^2 + 8192x + 1379$. (4) $\frac{dx}{dy} = 2y + 8$. (5) $185.9022654x^2 + 154.36334$.

P5.11 Find the derivative of the following rational functions:

(1) $p(x) = \dfrac{2x + 3}{3x + 2}$ (2) $q(x) = \dfrac{1 + x + 2x^2 + 3x^3}{1 + x + 2x^2}$

(3) $r(x) = \dfrac{ax + b}{cx + d}$ (4) $s(x) = \dfrac{x^n + a}{x^{-n} + b}$

Hint: Use the quotient rule.

5.11 (1) $p'(x) = \frac{-5}{(3x+2)^2}$. (2) $q'(x) = \frac{6x^4+6x^3+9x^2}{(1+x+2x^2)^2}$. (3) $r'(x) = \frac{ad-bc}{(cx+d)^2}$. (4) $s'(x) = \frac{anx^{-n-1}+bnx^{n-1}+2nx^{-1}}{(x^{-n}+b)^2}$.

P5.12 Differentiate the following functions:

(a) $y = x^x$ (b) $y = (e^x)^x$ (c) $y = e^{x^x}$

Hint: Recall that $\ln(a^b) = b\ln(a)$ and $a = e^{\ln a}$, for $a > 0$.

5.19 (a) $x^x(1 + \ln x)$; (b) $2x(e^x)^x$; (c) $e^{x^x}x^x(1 + \ln x)$.

P5.20 Differentiate the following functions with respect to θ

(1) $y = A\sin(\theta - \frac{\pi}{2})$ (2) $y = \sin^2\theta$ (3) $y = \sin 5\theta$

(4) $y = \sin^3\theta$ (5) $y = 18\cos(\theta + 6)$ (6) $y = \ln\cos\theta$

5.20 (1) $A\cos\left(\theta - \frac{\pi}{2}\right)$. (2) $2\sin\theta\cos\theta = \sin 2\theta$. (3) $5\cos 5\theta$. (4) $3\sin^2\theta\cos\theta$. (5) $-18\sin(\theta + 6)$. (6) $\frac{-\sin\theta}{\cos\theta} = -\tan\theta$.

P5.21 Differentiate $y = \frac{1}{2\pi}\cos(2\pi nt)$.

5.21 $\frac{dy}{dt} = -n\sin(2\pi nt)$.

P5.22 ~~Differentiate the~~ Take the derivative with respect to x of the following functions:

(i) ~~$y = \sin a^x$~~ $y = \sin(a^x)$ (ii) $y = \sec x$ (iii) $y = \cos^{-1}(x)$

(iv) $y = \tan^{-1}(x)$ (v) $y = \sec^{-1}(x)$ (vi) $y = \tan(x)\sqrt{3\sec x}$

5.22 (i) $a^x\ln a\cos a^x$. (ii) $\frac{dy}{dx} = \tan x\,\sec x$. (iii) $\frac{dy}{dx} = -\frac{1}{\sqrt{1-x^2}}$. (iv) $\frac{dy}{dx} = \frac{1}{1+x^2}$. (v) $\frac{dy}{dx} = \frac{1}{x\sqrt{x^2-1}}$. (vi) $\frac{dy}{dx} = \frac{\sqrt{3\sec x}\left(3\sec^2 x - 1\right)}{2}$.

P5.23 Find the derivatives of the following functions:

(1) $y = \sin\theta\sin(2\theta)$ (2) $y = a\tan^m(\theta^n)$ (3) $y = e^x\sin^2 x$

(4) $y = \sin\left((2\theta + 3)^{2.3}\right)$ (5) $y = \theta^3 + 3\sin(\theta + 3) - 3^{\sin\theta} - 3^\theta$

5.23 (1) $2\sin\theta\left(3\cos^2\theta - 1\right)$. (2) $amn\theta^{n-1}\times\tan^{m-1}(\theta^n)\sec^2\theta^n$. (3) $e^x\left(\sin^2 x + \sin 2x\right.$ (4) $4.6\,(2\theta + 3)^{1.3}\cos\left((2\theta + 3)^{2.3}\right)$. (5) $3\theta^2 + 3\cos(\theta + 3) - \ln 3\left(\cos\theta\times 3^{\sin\theta} + 3\theta\right)$.

P5.24 The length of an iron rod varies with temperature. Let $\ell(t)$ denote the length of the iron rod (in metres) at temperature t[°C]. We measure the length of the rod at different temperatures and determine its length is described by the function $\ell(t) = \ell_o(1 + 0.000012t)$[m], where ℓ_o is the length of the rod at 0[°C]. Find the change of length of the rod per degree Celsius.

5.24 $\frac{d\ell}{dt} = 0.000012\ell_o$[m/°C].

P5.25 The power P[W] consumed by an incandescent light bulb when connected is given by the equation $P = aV^b$, where a and b are constants, and V is the voltage drop across the bulb's terminals.

Find the rate of change of the power with respect to the voltage. Calculate the change in power per volt at the operating voltages $V_1 = 100$[V], $V_2 = 110$[V], and $V_3 = 120$[V], in the case of a light bulb for which $a = 0.008264$ and $b = 2$.

P5.30 The voltage V of a certain type of standard cell varies with temperature $t[°\mathrm{C}]$ according to the relation

$$V(t) = 1.4340\big[1 - 0.000814(t-15) + 0.000007(t-15)^2\big]. \quad [\mathrm{V}]$$

Find the change of voltage per degree at 15°C, 20°C, and 25°C.

5.30 $V'(t) = 1.4340(0.000014t - 0.001024)$, $V'(15) = -0.00117$, $V'(20) = -0.00107$, $V'(25) = -0.00097$.

P5.31 The voltage necessary to maintain an electric arc of length ℓ with a current of intensity I was found by Mrs. Ayrton to be

$$V = a + b\ell + \frac{c + k\ell}{I},$$

where a, b, c, k are constants. Find an expression for the variation of the voltage (a) with regard to the length of the arc ℓ; and (b) with regard to the strength of the current I.

5.31 (a) $\frac{dV}{d\ell} = b + \frac{k}{I}$; (b) $\frac{dE}{dI} = -\frac{c+k\ell}{I^2}$.

P5.32 Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ for the following functions:

(1) $y = 17x + 12x^2$ (2) $y = a + bx^2 + cx^4$ (3) $y = \frac{x^2+a}{x+a}$

5.32 (1) $17 + 24x$; 24. (2) $2bx + 4cx^3$; $2b + 12cx^2$. (3) $\frac{x^2+2ax-a}{(x+a)^2}$; $\frac{2a(a+1)}{(x+a)^3}$.

P5.33 Calculate $\frac{d^2y}{dx^2}$ and $\frac{d^3y}{dx^3}$ for the functions in **P5.9**.

5.33 (a) $\frac{d^2y}{dx^2} = \frac{d^3y}{dx^3} = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \ldots$. (b) $\frac{d^2y}{dx^2} = 2a$; $\frac{d^3y}{dx^3} = 0$. (c) $\frac{d^2y}{dx^2} = 6x + 6a$; $\frac{d^3y}{dx^3} = 6$.

P5.34 Calculate second and third derivatives of the functions in **P5.10**.

5.34 (1) $-b$; 0. (2) 2; 0. (3) $56440x^3 - 196212x^2 - 4488x + 8192$; $169320x^2 - 392424x - 4488$. (4) 2; 0. (5) $371.80453x$; 371.80453.

P5.35 The distance travelled by a body falling freely in space is described by the equation $d = 16t^2$, where d is in feet, and t is in seconds. Draw a curve showing the relation between d and t. Determine the velocity of the body at the following times: $t = 2$[s], $t = 4.6$[s], and $t = 0.01$[s].

5.35 $v(2) = 64$[ft/s], $v(4.6) = 147.2$[ft/s], $v(0.01) = 0.32$[ft/s].

P5.36 ~~If $x = v_i t - \frac{1}{2}gt^2$~~Given $x(t) = v_i t - \frac{1}{2}gt^2$, find $\dot{x}$ and $\ddot{x}$.
Hint: Physicists sometimes use $\dot{x}$ and $\ddot{x}$ to denote $x'(t)$ and $x''(t)$.

5.36 $\dot{x} = v_i - gt$, $\ddot{x} = -g$.

P5.37 If a body moves according to the law $x(t) = 12 - 4.5t + 6.2t^2$[m], find its velocity and its acceleration when $t = 4$[s].

5.37 $v(4) = 45.1$[m/s], $a(4) = 12.4$[m/s^2].

5.54 $2xyz + y^2z + z^2y + 2xy^2z^2$; $2xyz + x^2z + xz^2 + 2x^2yz^2$; $2xyz + x^2y + xy^2 + 2x^2y^2z$.

P5.55 Let $r^2 = (x-a)^2+(y-b)^2+(z-c)^2$. Find the value of $\frac{dr}{dx}+\frac{dr}{dy}+\frac{dr}{dz}$. Also find $\frac{d^2r}{dx^2}+\frac{d^2r}{dy^2}+\frac{d^2r}{dz^2}$.

5.55 $\frac{1}{r}\{(x-a)+(y-b)+(z-c)\} = \frac{(x+y+z)-(a+b+c)}{r}$; $\frac{3}{r}$.

P5.56 Find the total differential of $y = u^v$.

5.56 $dy = vu^{v-1}\,du + u^v \ln(u)\,dv$.

P5.57 Find the total differential of the following expressions:

(1) $y = u^3 \sin v$ (2) $y = (\sin x)^u$ (3) $y = \frac{\ln u}{v}$

5.57 (1) $dy = 3u^2 \sin v\,du + u^3 \cos v\,dv$. (2) $dy = u(\sin x)^{u-1}\cos x\,dx + (\sin x)^u \ln(\sin x)du$. (3) $dy = \frac{1}{v}\frac{1}{u}\,du - \ln u \frac{1}{v^2}\,dv$.

P5.58 The equation of a tangent to the curve $y = 5 - 2x + 0.5x^3$ is of the form $y = mx + b$, where m and b are constants. Find the value of m and b for the tangent to the curve at $x = 2$.

5.58 $m = 4$, $b = -3$.

P5.59 Consider the line ℓ_1 defined by the equation $y = x$ and the line ℓ_2 defined by $y = 3x$. Use the equation $\theta = \tan^{-1}\left(\frac{\Delta y}{\Delta x}\right)$ to find the angle each line makes with the x-axis. Find the angle of intersection between the lines.

5.59 $\angle\ell_1$ is $45°$, $\angle\ell_2$ is $71.56°$. Lines ℓ_1 and ℓ_2 intersect at an angle $26.56°$.

P5.60 ~~At what angle do the two curves~~

$$\cancel{y = 3.5x^2 + 2 \qquad \text{and} \qquad y = x^2 - 5x + 9.5}$$

~~intersect each other~~What are the *intersection angle* between the curve $f(x) = 3.5x^2 + 2$ and the curve $g(x) = x^2 - 5x + 9.5$ at the point(s) where they meet?

5.60 Intersections at $x = 1$ and $x = -3$. Angles $153°26'$ and $2°28'$.

P5.61 Two tangent lines to the curve $y = \sqrt{25 - x^2}$ are drawn at points $x = 3$ and $x = 4$. Find the coordinates of the point where the tangent lines intersect and their angle of intersection.

5.61 The tangents intersect at $(x, y) = (3.57, 3.50)$ at an angle of $16°16'$.

P5.62 ~~A straight~~ The line $y = 2x - b$ touches the curve $y = 3x^2 + 2$ at one point. ~~What are the coordinates of the~~ Where is this point of contact and what is the value of b?

5.62 $x = \frac{1}{3}$, $y = 2\frac{1}{3}$, $b = -\frac{5}{3}$.

P5.63 Find the value(s) of x that make y maximum or minimum:

5.97 For a second-order differential equation (a differential equation involving second derivatives) there will be two independent solutions. The hint tells us these solutions are $e^{-\lambda x}$ and $xe^{-\lambda x}$ and we choose $\lambda = 1$ to satisfy the differential equation. The general solution is any linear combination of these solutions $f(x) = C_1 e^{-x} + C_2 x e^{-x}$, where C_1 and C_2 are arbitrary constants. By computing $f(0)$ and $f'(0)$, we find that the choice $C_1 = 1$ and $C_2 = 2$ satisfies the initial conditions $f(0) = 1$ and $f'(0) = 1$. See `bit.ly/1kNxhvo`.

P5.98 Calculate the length of the curve $f(x) = \frac{1}{2}x^2$ between $x = 0$ and $x = 1$.
Hint: This problem requires a long integral calculation. Start by rewriting $\sqrt{1+x^2} = \frac{1+x^2}{\sqrt{1+x^2}}$. Lookup the derivative formula for $\sinh^{-1}(x)$. Use integration by parts and the self-referential trick from page 375.

5.98 $\ell = \frac{1}{2}\sinh^{-1}(1) + \frac{1}{2}\sqrt{2}$.

5.98 The arc length of a curve $f(x)$ is $\cancel{\ell = \int_a^b d\ell = \int_a^b \sqrt{1 + (f'(x))^2}\, dx}$ $\ell = \int_a^b d\ell = \int_a^b \sqrt{1 +$
In the current problem $f'(x) = x$ so the integral we want to find is $I = \int \sqrt{1+x^2}\, dx$. Note $\sqrt{1+x^2} = \frac{1+x^2}{\sqrt{1+x^2}} = \frac{1}{\sqrt{1+x^2}} + \frac{x^2}{\sqrt{1+x^2}}$. The first term is the derivative of $\sinh^{-1}(x)$. For the second term, use integration by parts with $u = x$, $dv = \frac{x}{\sqrt{1+x^2}}\, dx$. You'll obtain the equation $I = \sinh^{-1}(x) + x\sqrt{1+x^2} - I$, where I is the integral we want to find.

Evaluating I at the endpoints, we find $\ell = I\big|_0^1 = \frac{1}{2}\sinh^{-1}(1) + \frac{1}{2}\sqrt{2}$. ~~If you obtained the~~ The answer $\frac{1}{2}\ln(1+\sqrt{2}) + \frac{1}{2}\sqrt{2}$ ~~, it's~~ is also correct because $\sinh^{-1}(x) \equiv \ln(x + \sqrt{1+x^2})$.

P5.99 The voltage coming out of a North American electric wall outlet is described by the equation $V(t) = 155.57\cos(\omega t)$[V]. The average squared voltage is calculated using the integral $V_{\text{avg}}^2 = \frac{1}{T}\int_0^T V(t)^2\, dt$. Calculate the root-mean-squared voltage $V_{\text{rms}} \equiv \sqrt{V_{\text{avg}}^2}$.
Hint: Make the substitution $\tau = \omega t$ and recall $\frac{1}{\omega} = \frac{T}{2\pi}$.

5.99 $V_{\text{rms}} = 110$[V].

5.99 We want to calculate the square of the voltage during one period T: $V_{\text{avg}}^2 = \frac{1}{T}\int_0^T 155.57^2\cos^2(\omega t)\, dt$. Making the substitution $\tau = \omega t$, we obtain an equivalent expression $V_{\text{avg}}^2 = \frac{1}{2\pi}\int_0^{2\pi} 155.57^2\cos^2(\tau)\, d\tau$, which is $\frac{155.57^2}{2\pi}\int_0^{2\pi}\cos^2(\tau)\, d\tau = \frac{155.57^2}{2\pi}\left[\frac{t}{2} + \frac{\sin(t)\cos(t)}{2}\right]_0^{2\pi} = \frac{155.57^2}{2\pi}\left[\frac{2\pi}{2}\right] = \frac{155.57^2}{2}$.
The root-mean-squared voltage is $V_{\text{rms}} = \sqrt{\frac{155.57^2}{2}} = \frac{155.57}{\sqrt{2}} = 110$[V].

P5.100 Find the volume generated by the curve $y = \sqrt{1+x^2}$ between $x = 0$ and $x = 4$ as it revolves about the x-axis.

5.100 79.4.

$$(1)\ \sum_{n=1}^{\infty} \frac{1}{2n^2 - 3n - 5} \qquad (2)\ \sum_{n=1}^{\infty} \frac{\ln n}{n^3} \qquad (3)\ \sum_{n=1}^{\infty} \frac{2^n}{3^n + 4^n}$$

Hint: Use the alternating series test.

5.108 (1) Converges. (2) Converges. (3) Converges.

P5.109 Calculate the values of the following infinite series:

$$(1)\ \sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} \qquad (2)\ \sum_{n=0}^{\infty} \left(\frac{2}{3^n} + \frac{4}{5^n}\right) \qquad (3)\ \sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^n$$

5.109 (1) 1. (2) 8. (3) 2.

5.109 In each case we use the formula for the geometric series $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$. See `bit.ly/1e9F52v` for the calculations.

P5.110 State whether the series converge absolutely, converge conditionally, or diverge.

$$(1)\ \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n^2 + 2} \qquad (2)\ \sum_{n=1}^{\infty} \frac{(-1)^n \ln n}{n} \qquad (3)\ \sum_{n=2}^{\infty} \frac{(-1)^n}{\ln n}$$

5.110 (1) Converges absolutely. (2) Converges conditionally. (3) Converges conditionally.

5.110 Use the geometric series formula $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$.

P5.111 Use the n^{th} root test or the ratio test to see whether the following series converge:

$$(a)\ \sum_{n=1}^{\infty} \frac{n!}{n^n} \qquad (b)\ \sum_{n=1}^{\infty} \frac{n^2}{n^n} \qquad (c)\ \sum_{n=1}^{\infty} \frac{(n!)^2}{n^n}$$

5.111 (a) Converges. (b) Converges. (c) Diverges.

P5.112 Find the sum of $\frac{2}{3} + \frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \frac{1}{24} + \cdots$.

5.112 $\frac{4}{3}$.

5.112 This is a geometric series with $r = \frac{1}{2}$ and $a = \frac{2}{3}$, so ~~the infinite sum is~~ $\frac{a}{1-r} = \frac{\frac{2}{3}}{1-\frac{1}{2}} = \frac{4}{3}$.

P5.113 Calculate the Maclaurin series of the function $f(x) = \frac{1}{1-x}$.

5.113 $f(x) = \sum_{n=0}^{\infty} x^n$.

P5.114 Find the Maclaurin series for the following functions:

P5.119 Prove the formulas for the infinite sum of the first N positive integers, and the sum of the squares of the first N positive integers:

$$\sum_{n=1}^{N} n = \frac{N(N+1)}{2}, \qquad \sum_{n=1}^{N} n^2 = \frac{N(N+1)(2N+1)}{6}.$$

Hint: For the first part, try writing out the terms of the summation in increasing order, and a copy of it in decreasing order below. For the second part, you might be interested in computing the sum $\sum_{n=1}^{N}(n+1)^3 - n^3$ using two different approaches, and equating the results.

5.119 Taking the approach suggested in the hint, we can write twice the summation we want to find as

$$\begin{aligned} 2\sum_{n=1}^{N} n &= 1 + 2 + \cdots + (N-1) + N + \\ &\quad N + (N-1) + \cdots + 2 + 1 \\ &= \underbrace{(N+1) + (N+1) + \cdots + (N+1) + (N+1)}_{N \text{ times}}. \end{aligned}$$

From this we conclude that $2\sum_{n=1}^{N} n = N(N+1)$.

For the second part, we can use the "telescopic" nature of this series. Except for the first and last terms in this series, the negative part of each term $(n+1)^3 - n^3$ cancels with positive part of the next term. Thus only the negative part of the first term and the positive part of the last term remain: $\sum_{n=1}^{N}(n+1)^3 - n^3 = (N+1)^3 - 1$. Using basic algebra operations we find:

$$(n+1)^3 - n^3 = n^3 + 3n^2 + 3n + 1 - n^3 = 3n^2 + 3n + 1.$$

Since we know $\sum_{n=1}^{N}(n+1)^3 - n^3 = (N+1)^3 - 1$ from the telescopic approach, we obtain the equation

$$(N+1)^3 - 1 = \sum_{n=1}^{N} 3n^2 + 3n + 1 = 3\sum_{n=1}^{N} n^2 + 3\sum_{n=1}^{N} n + \sum_{n=1}^{N} 1.$$

Using the formulas $\sum_{n=1}^{N} n = \frac{N(N+1)}{2}$ and $\sum_{n=1}^{N} 1 = N$, we rewrite as

$$(N+1)^3 - 1 = 3\sum_{n=1}^{N} n^2 + 3\frac{N(N+1)}{2} + N.$$

Isolating $\sum_{n=1}^{N} n^2$ and simplifying leads us to the desired result.

P5.120 Compute the definite integral $\int_\alpha^\beta mx + b$ using the Riemann sum approach, taking the limit $n \to \infty$ for the number of vertical rectangles.

5.120 $\int_\alpha^\beta (mx + b)\,dx = \frac{m}{2}(\beta^2 - \alpha^2) + b(\beta - \alpha)$.

5.120 Using the definition of the integral as a the limit of a Riemann sum from page 350, we can write the following formula for the integral of $f(x)$ between $x = \alpha$ and $x = \beta$:

$$\int_\alpha^\beta f(x)\,dx \equiv \lim_{n\to\infty} \sum_{k=1}^{n} f(a + k\Delta x)\Delta x,$$

where $\Delta x = \frac{\beta-\alpha}{n}$ denotes the width of the rectangles used to approximate the area under the curve. For the integral of the function $f(x) = mx + b$ the formula becomes

$$\begin{aligned}
\int_\alpha^\beta mx + b\,dx &= \lim_{n\to\infty} \sum_{k=1}^{n} \left(m(\alpha + k\Delta x) + b\right)\Delta x \\
&= \lim_{n\to\infty} \Delta x \left((m\alpha + b) \sum_{k=1}^{n} 1 + m(\Delta x) \sum_{k=1}^{n} k \right) \\
&= \lim_{n\to\infty} \Delta x \left((m\alpha + b)n + m(\Delta x)\frac{n(n+1)}{2} \right).
\end{aligned}$$

After taking the limit $n \to \infty$ and several steps of simplification we obtain the final answer $\int_\alpha^\beta (mx + b)\,dx = b(\beta - \alpha) + \frac{m}{2}(\beta^2 - \alpha^2)$.

P5.121 Find the area under the curve for the function $f(x) = ax^2 + bx + c$ between $x = 0$ and $x = d$ using the infinite Riemann sum approach.

5.121 $\int_0^d (ax^2 + bx + c)\,dx = \frac{1}{3}ad^3 + \frac{1}{2}bd^2 + cd$.

5.121 Using the definition of the integral as an infinite Riemann sum and taking $x_k = 0 + \Delta x k$ and $\Delta x = \frac{d-0}{n}$, we obtain the following formula for area under the curve of $f(x) = ax^2 + bx + c$ between $x = 0$ and $x = d$:

$$\begin{aligned}
&\int_0^d ax^2 + bx + c\,dx \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \left(a(\Delta x)^2k^2 + b(\Delta x)k + c\right)\Delta x \\
&= \lim_{n\to\infty} \Delta x \left(a(\Delta x)^2 \sum_{k=1}^{n} k^2 + b(\Delta x)\sum_{k=1}^{n} k + c\sum_{k=1}^{n} 1 \right) \\
&= \lim_{n\to\infty} \Delta x \left(a(\Delta x)^2 \sum_{k=1}^{n} \frac{n(n+1)(2n+1)}{6} + b(\Delta x)\frac{n(n+1)}{2} + cn \right),
\end{aligned}$$

which becomes $\frac{1}{3}ad^3 + \frac{1}{2}bd^2 + cd$ after using $\Delta x = \frac{d}{n}$ and taking the limit.

P5.122 Calculate the surface area of a sphere of radius R using the surface of revolution formula given on page 384.
Hint: Split the surface area into narrow strips of width $d\ell$. Imagine you're peeling an orange by starting from the top and going around in circles. Each turn, produces a ring of orange peel of area $dA = 2\pi r d\ell$. Since the surface area of an orange is equal to the area of its peels, we have $A = \int dA$.

5.122 $A_{\text{sphere}} = 4\pi R^2$.

5.122 The surface area of a sphere of radius R can be computed by splitting it into narrow circular strips of area dA, with radius varying according to $f(x) = \sqrt{R^2 - x^2}$. The width of each strip is given by the arc length formula $d\ell = \sqrt{1 + (f'(x))^2}\, dx$. To compute $A_{\text{sphere}} = \int_{-R}^{R} 2\pi f(x) \sqrt{1 + (f'(x))^2}\, dx$, we first find $f'(x) = \frac{-x}{\sqrt{R^2-x^2}}$ then obtain $\sqrt{1 + (f'(x))^2} = \frac{R}{\sqrt{R^2-x^2}}$. Substituting into the formula we find $A_{\text{sphere}} = \int_{-R}^{R} 2\pi\sqrt{R^2 - x^2} \frac{R}{\sqrt{R^2-x^2}}\, dx$, which simplifies to $2\pi R \int_{-R}^{R} 1\, dx$, which gives $4\pi R^2$ as expected.

P5.123 Calculate the volume of the solid of revolution generated by revolving the region bounded by the curve $y = x^2$ and the lines $x = 0$, $x = 1$, and $y = 0$ around the x-axis.

5.123 $\frac{\pi}{5}$.

P5.124 Calculate the volume of the solid of revolution generated by revolving the region bounded by the curves $y = x^2$ and $y = x^3$ and the lines $x = 0$ and $x = 1$ around the x-axis.

5.124 $\frac{2\pi}{35}$.

P5.125 *Torricelli's trumpet* is the surface of revolution generated by rotating the curve $f(x) = \frac{1}{x}$ around the x-axis. Figure 5.16 on page 384 shows part of the surface. Calculate the total volume enclosed by Torricelli's trumpet between $x = 1$ and $x = \infty$.
Hint: Use the disk method to compute the volume of revolution between $x = 1$ and $x = a$ then take the limit as $a \to \infty$.

5.125 $V = \pi$.

5.125 Using the disk method, we find the volume enclosed by Torricelli's trumpet between $x = 1$ and $x = a$ is $\int_1^a \pi (f(x))^2\, dx = \pi \int_1^a \frac{1}{x^2}\, dx = \pi \left(1 - \frac{1}{a}\right)$. To find the total volume we compute the limit $\lim_{a\to\infty} \pi \left(1 - \frac{1}{a}\right) = \pi$.

P5.126 Calculate the moment of inertia of a thin spherical shell with mass m and radius R by adapting the formula for the surface area of a revolution from page 384.
Hint: See page 388 if you need a reminder how to setup the integral.

5.126 $I_{\text{sph.shell}} = \frac{2}{3}mR^2$.

5.126 To evaluate $I_{\text{sph.shell}} = \int_{-R}^{R}(f(x))^2\,\sigma 2\pi f(x)\sqrt{1+(f'(x))^2}\,dx$, we first compute $f'(x) = \frac{-x}{\sqrt{R^2-x^2}}$ then obtain $\sqrt{1+(f'(x))^2} = \frac{R}{\sqrt{R^2-x^2}}$. Substituting these calculations into the integral, we find $I_{\text{sph.shell}} = 2\pi\sigma R\int_{-R}^{R}(R^2-x^2)$ which is straightforward integral of a polynomial.

P5.127 Calculate the moment of inertia of a solid sphere with mass m and radius R using the cylindrical shell method.
Hint: Adapt the formula for computing the volume of a sphere to take into account the total mass and the squared distance of each spherical shell.

5.127 $I_{\text{sphere}} = \frac{2}{5}mR^2$.

5.127 The formula for computing the volume of a sphere using the spherical shell method is $V_{\text{sphere}} = \int_0^R 2\pi r h(r) dr$, where $h(r) = 2\sqrt{R^2-r^2}$ is the height of the thin cylindrical shell at radius r, and $2\pi r$ is its circumference. To evaluate the integral $I_{\text{sphere}} = \int r^2 dm$, we define $dm = \rho A(r) dr = \rho 2\pi r h(r)\,dr$ and scale the contributions of each dm by an extra factor r^2. We thus obtain $I_{\text{sphere}} = \int_0^R r^2 \rho 2\pi r 2\sqrt{R^2-r^2}\,dr$. Computing the integral $\int_0^R r^3\sqrt{R^2-r^2}\,dr$ requires several steps, and I encourage you to perform the calculation on your own. The substitutions $r = R\sin\theta$ and later $u = \cos\theta$ might come in handy.

P5.128 Calculate the moment of inertia of a solid sphere with mass m and radius R by splitting it into thin disks. The moment of inertia of a disk of thickness dx, radius $r(x)$, and mass $m(x)$ is given by $I_{\text{disk}}(x) = \frac{1}{2}m(x)[r(x)]^2$, where the mass of the disk is proportional to the mass density $\rho = \frac{m}{\frac{4}{3}\pi R^3}$ and its area $A(x)$: $m(x) = \rho A(x) = \rho\pi[r(x)]^2$.
Hint: The radius of each disk as a function of x is described by $r(x) = \sqrt{R^2-x^2}$. The integral you want to setup looks like $\int_{-R}^{R} I_{\text{disk}}(x)\,dx$.

5.128 $I_{\text{sphere}} = \frac{2}{5}mR^2$.

5.128 The integral to calculate, $\int_{-R}^{R} I_{\text{disk}}(x)\,dx = \int_{-R}^{R}\frac{1}{2}m(x)[r(x)]^2\,dx$, becomes $\frac{1}{2}\pi\rho\int_{-R}^{R}(f(x))^4\,dx$ after substituting all the known values. After expanding $(f(x))^4$ we find $I_{\text{sphere}} = \frac{1}{2}\pi\rho\int_{-R}^{R}(R^4 - 2R^2x^2 + x^4)\,dx$, which is straightforward to compute.

Links

Here are some links to more calculus problems with solutions:

[Lots of solved calculus examples by Larry Perez]
http://saddleback.edu/faculty/lperez/algebra2go/calculus/calc3A.pdf

[Try the odd-numbered problems in Gilbert Strang's calculus ~~text~~ book]
http://bit.ly/1mnheSD

ing during my engineering days, and Guy Moore and Zaven Altounian for teaching me advanced physics topics. Among all my teachers, I owe the most to Patrick Hayden whose teaching methods have always inspired me. From him, I learned that by defining things clearly, you can *trick* students into learning advanced topics, and even make it seem that the results are obvious! Thanks go out to all my research collaborators and friends: David Avis, Arlo Breault, Juan Pablo Di Lelle, Omar Fawzi, Adriano Ferrari, Igor Khavkine, Felix Kwok, Doina Precup, Andie Sigler, and Mark M. Wilde. Thank you all for teaching me a great many things!

Preparing this book took many years and the combined efforts of many people. I want to thank Afton Lewis, Oleg Zhoglo, and Alexandra Foty for helping me proofread v2 of the book, and all the ~~v3 and v4~~ readers who reported typos and suggested clarifications. Thank you all for your comments and feedback! Georger Araujo ~~deserves a~~ and Tomasz Święcicki deserve particular mention as ~~his~~ their meticulous reading of the text led to the correction of many technical mistakes. I also want to thank Mohamad Nizar Kezzo for helping me prepare the problems and exercises for the book. Above all, I want to thank my editor Sandy Gordon, who helped me substantially improve the writing in the book. Her expertise with the English language and her advice on style and content have been absolutely invaluable.

Last but not least, I want to thank all my students for their endless questions and demands for explanations. If I have developed any skill for explaining things, I owe it to them.

Further reading

You have reached the end of this book, but you're only at the beginning of the journey of scientific discovery. There are a lot of cool things left for you to learn about. Below are some recommendation of subjects you might find interesting.

Electricity and Magnetism

Electrostatics is the study of the electric force $\vec{F}_e$ and the associated electric potential U_e. Here, you will also learn about the electric field $\vec{E}$ and electric potential V.

Magnetism is the study of the magnetic force $\vec{F}_b$ and the magnetic field $\vec{B}$, which are caused by electric currents flowing through wires. The current I is the total number of electrons passing through a cross-section of the wire in one second. By virtue of its motion through space, each electron contributes to the strength of the magnetic field surrounding the wire.

The beauty of electromagnetism is that the entire theory can be described in just four equations:

$\nabla \cdot \vec{E} = \dfrac{\rho}{\varepsilon_0}$	Gauss's law
$\nabla \cdot \vec{B} = 0$	Gauss's law for magnetism
$\nabla \times \vec{E} = -\dfrac{\partial \vec{B}}{\partial t}$	Faraday's law of induction
$\nabla \times \vec{B} = \mu_0 \vec{J} + \mu_0 \varepsilon_0 \dfrac{\partial \vec{E}}{\partial t}$	Ampère's circuital law

Together, these are known as Maxwell's equations.

Vector calculus

You may be wondering what the triangle thing ~~∇~~ is. The symbol ∇ (~~nabla~~called *nabla*) is the vector derivative operation: $\nabla \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$. Guess what—you can also do calculus with vectors.

If you take a vector calculus course, you'll learn about path integrals, surface integrals, and volume integrals of vector functions. You will also learn about vector-derivatives, as well as two vector equivalents of the fundamental theorem of calculus:

- Stokes' Theorem:

$$\iint_{\Sigma} \nabla \times \vec{F} \cdot d\vec{S} = \int_{\partial\Sigma} \vec{F} \cdot d\vec{r},$$

  which states the integral of the ~~curl$\vec{F} \equiv \nabla \times \vec{F}$~~ curl $\mathrm{curl}(\vec{F}) \equiv \nabla \times \vec{F}$ of the field $\vec{F}$ over the surface Σ is equal to the circulation of $\vec{F}$ along the boundary of the surface $\partial\Sigma$.

- Gauss' Divergence Theorem:

$$\iiint_{\mathrm{V}} \nabla \cdot \vec{F}\ \mathrm{dV}\, dV = \iint_{\partial \mathrm{V}} \vec{F} \cdot d\vec{S},$$

  which states the integral of the divergence ~~div$\vec{F} \equiv \nabla \cdot \vec{F}$~~ $\mathrm{div}(\vec{F}) \equiv \nabla \cdot \vec{F}$ of the field $\vec{F}$ over the volume V is equal to the flux of $\vec{F}$ through the volume's boundary ∂V.

Both theorems relate the total of the derivative of a quantity over some region R to the value of that quantity on the boundary of the region, which we denote ∂R. The fundamental theorem of calculus can also be interpreted in this manner:

$$\int_I F'(x)\,dx = \int_a^b F'(x)\,dx = F_{\partial I} = F(b) - F(a),$$

$z = f(x, y)$ as the height of a mountain at particular (x, y) coordinates on a map, then the gradient vector $\nabla f(x, y)$ always points uphill.

~~The~~ In the second part of multivariable calculus you'll learn how to do double integrals, which are integrals over two variables:

$$\int_{y=0}^{y=1}\int_{x=0}^{x=1} yx\, dxdy = \int_{y=0}^{y=1}\left[y\frac{x^2}{2}\right]_{x=0}^{x=1} dy = \int_{y=0}^{y=1}\frac{y}{2}\, dy = \left[\frac{1}{4}y^2\right]_0^1 = \frac{1}{4}.$$

Once you get over the initial shock of seeing two integral signs, you should be able to understand the above integral calculation. Proceeding from inside-out, a double integral is nothing more than two integral operators applied in succession. Instead of an interval of integration split into tiny integration steps dx, we have a *region of integration* split into tiny integration boxes with area $dxdy$. In the above example, the region of integration is a unit square in the xy-plane.

If you understand derivatives and integrals, you will find multivariable calculus easy: it's just he multivariable upgrade of the concepts from Chapter 5.

Linear algebra

Linear algebra is the study of vectors $\vec{v} \in \mathbb{R}^n$ and linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$. Linear transformations are *vector functions* that obey the linear property $T(\alpha\vec{v}_1 + \beta\vec{v}_2) = \alpha T(\vec{v}_1) + \beta T(\vec{v}_2)$. Using the standard notation for functions we write $T(\vec{v}) = \vec{w}$, to show the linear transformation acting on an input vector $\vec{v} \in \mathbb{R}^n$ to produce the output vector $\vec{w} \in \mathbb{R}^m$. Every linear transformation T can be *represented* as a matrix $A_T \in \mathbb{R}^{m\times n}$, which is an array of numbers with m rows and n columns. Because of the equivalence between linear transformations and matrices, we can also say that linear algebra is the study of vectors and matrices.

Learning linear algebra will open many doors for you. You need linear algebra to understand computer graphics, machine learning, quantum mechanics, error correcting codes, and many other areas of science and business. Vectors and matrices are used all over the place! If your knowledge of high-school math gives you modelling superpowers, then linear algebra is the vector-upgrade that teaches you how to model multivariable quantities.

[VIDEOS] Gilbert Strang. *Linear Algebra*, MIT OpenCourseWare, 2010, online: `http://bit.ly/StrangLAlectures`.

[BOOK] Ivan Savov. *No bullshit guide to linear algebra*, Minireference Publishing, ~~2014,~~ 2nd edition, 2016, ISBN 978-0-9920010-2-5.

Set notation

You don't need a lot of fancy notation to understand mathematics. It really helps, though, if you know a little bit of set notation.

Symbol	Read as	Denotes
$\{\ \dots\ \}$	the set ...	define a sets
$\mid$	such that	describe or restrict the elements of a set
$\mathbb{N}$	the naturals	the set $\mathbb{N} \equiv \{0, 1, 2, \dots\}$. Note $\mathbb{N}_{+} \equiv \mathbb{N} \backslash \{0\}$.
$\mathbb{Z}$	the integers	the set $\mathbb{Z} \equiv \{\dots, -2, -1, 0, 1, 2, 3, \dots\}$
$\mathbb{Q}$	the rationals	the set of fractions of integers
$\mathbb{A}$		the set of algebraic numbers
$\mathbb{R}$	the reals	the set of real numbers
$\mathbb{C}$		the set of complex numbers
$\subset$	subset	one set strictly contained in another
$\subseteq$	subset or equal	containment or equality
$\cup$	union	the combined elements from two sets
$\cap$	intersection	the elements two sets have in common
$S \setminus T$	S set minus T	the elements of S that are not in T
$a \in S$	a in S	a is an element of set S
$a \notin S$	a not in S	a is not an element of set S
$\forall x$	for all x	a statement that holds for all x
$\exists x$	there exists x	an existence statement
$\nexists x$	there doesn't exist x	a non-existence statement

An example of a condensed math statement that uses set notation is "$\nexists m, n \in \mathbb{Z}$ such that $\frac{m}{n} = \sqrt{2}$," which reads "there don't exist integers m and n whose fraction equals $\sqrt{2}$." Since we identify the set of ~~fraction~~ fractions of integers with the rationals, this statement is equivalent to the shorter "$\sqrt{2} \notin \mathbb{Q}$," which reads "$\sqrt{2}$ is irrational."

Complex numbers notation

Expression	Denotes
$\mathbb{C}$	the set of complex numbers $\mathbb{C} \equiv \{a + bi \mid a, b \in \mathbb{R}\}$
i	the unit imaginary number $i \equiv \sqrt{-1}$ or $i^2 = -1$
$\mathrm{Re}\{z\} = a$	real part of $z = a + bi$
$\mathrm{Im}\{z\} = b$	imaginary part of $z = a + bi$
$\lvert z \rvert \angle \varphi_z$	polar representation of $z = \lvert z \rvert \cos \varphi_z + i \lvert z \rvert \sin \varphi_z$
$\lvert z \rvert = \sqrt{a^2 + b^2}$	magnitude of $z = a + bi$
$\varphi_z = \tan^{-1}(b/a)$	phase or argument of $z = a + bi$
$\bar{z} = a - bi$	complex conjugate of $z = a + bi$

Units

The International System of Units (*Système International*) defines seven base units for measuring physical quantities.

Name	Sym.	Measures	Definition
metre	m	length	The distance travelled by light in vacuum during $\frac{1}{299792458^{\text{th}}}$ of a second.
kilogram	kg	mass	The mass of the *international prototype kilogram* (a cylinder of platinum-iridium kept at Sèvres near Paris).
second	s	time	The time for 9192631770 transitions in the ground state of the caesium-133 atom.
Ampere	A	electric current	One ampere is the current that has to flow in two infinitely long wires placed a distance 1[m] apart, to produce a force between them of 2×10^{-7}[N/m].
Kelvin	K	temperature	The Kelvin is $\frac{1}{273.16}$ of the thermodynamic temperature of the triple point of water.
mole	mol	# of atoms	One mole is how many carbon atoms are in 0.012[kg] of carbon-12.
candela	cd	light intensity	One candela is defined as the luminous intensity of a monochromatic source with a particular frequency and radiant intensity.

Derived units

The base SI units cover most of the fundamental quantities. Other physical units are defined as combinations of the basic units.

Name	Sym.	Measures	Definition	SI equivalent
Hertz	Hz	frequency		s^{-1}
Newton	N	force		$kg\,m\,s^{-2}$
Pascal	Pa	pressure	N/m^2	$kg\,m^{-1}\,s^{-2}$
Joule	J	energy, work, heat	N m	$kg\,m^2\,s^{-2}$
Watt	W	power	J/s	$kg\,m^2\,s^{-3}$
Coulomb	C	electric charge		s A
Volt	V	voltage, electric potential		$kg\,m^2\,s^{-3}\,A^{-1}$
Ohm	Ω	resistance, reactance	V/A	$kg\,m^2\,s^{-3}\,A^{-2}$
Siemens	S	electrical conductance	A/V	$kg^{-1}\,m^{-2}\,s^3\,A^2$
Farad	F	capacitance	C/V	$kg^{-1}\,m^{-2}\,s^4\,A^2$
Tesla	T	magnetic field strength		$kg\,s^{-2}\,A^{-1}$
Henry	H	inductance	Ω s	$kg\,m^2\,s^{-2}\,A^{-2}$
Weber	Wb	magnetic flux	$T\,m^2$	$kg\,m^2\,s^{-2}\,A^{-1}$

Other units and conversions

We often measure physical quantities like length, weight, and velocity in nonstandard units like feet, pounds, and miles per hour. The following table lists the conversion ratios which are required to covert these nonstandard measurement units to SI units.

Dimension	Symb.	Name	Conversion
length	Å	Angstrom	$1[\text{Å}] = 10^{-10}[\mathrm{m}] = 0.1[\mathrm{nm}]$
	in, "	inch	$1[\cancel{\mathrm{T}}\mathrm{in}] = \cancel{1000}2.54[\cancel{\mathrm{kg}}\mathrm{cm}]$
	ft, '	foot	$1[\mathrm{ft}] = 12[\mathrm{in}] = 0.3048[\mathrm{m}]$
	yd	yard	$1[\mathrm{yd}]= 3[\mathrm{ft}] = 0.9144[\mathrm{m}]$
	mi	mile	$1[\mathrm{mi}] = 5280[\mathrm{ft}] = 1609.344[\mathrm{m}]$
	nmi	nautical mile	$1[\mathrm{nmi}] = 1852[\mathrm{m}]$
	ly	light-year	$1[\mathrm{ly}] = 9.460\,730\,472 \times 10^{15}[\mathrm{m}]$
area	in^2	square inch	$1[\mathrm{in}^2] = 6.452 \times 10^{-4}[\mathrm{m}^2]$
	ft^2	square foot	$1[\mathrm{ft}^2] = 9.290 \times 10^{-2}[\mathrm{m}^2]$
			$1[\mathrm{m}^2] = 110.764[\mathrm{ft}^2]$
	ac	acre	$1[\mathrm{ac}] = 4840[\mathrm{yd}^2] = 4046.856[\mathrm{m}^2]$
	ha	hectare	$1[\mathrm{ha}] = 10\,000[\mathrm{m}^2]$
	mi^2	square mile	$1[\mathrm{mi}^2] = 2.589\,988 \times 10^6[\mathrm{m}^2]$
volume	L	litre	$1[\mathrm{L}] = 1[\mathrm{dm}^3] = \frac{1}{1000}[\mathrm{m}^3]$
	gal(US)	gallon (fluid)	$1[\mathrm{gal}] = 3.785[\mathrm{L}]$
weight	lb	pound	$1[\mathrm{lb}] = 0.454[\mathrm{kg}] = 453.592[\mathrm{g}]$
	t	tonne	$1[\mathrm{t}] = 1000[\mathrm{kg}]$
angle	rad	radian	$1[\mathrm{turn}] = 2\pi[\mathrm{rad}]$
	°	degree	$360[^\circ] = 2\pi[\mathrm{rad}]$
	rev	revolution	$1[\mathrm{rev}] = 360[^\circ] = 2\pi[\mathrm{rad}]$
	grad	gradian	$1[\mathrm{grad}]=\frac{1}{400}[\mathrm{rev}] = 0.9[^\circ]$
time	min	minute	$1[\mathrm{min}] = 60[\mathrm{s}]$
	h	hour	$1[\mathrm{h}] = 60[\mathrm{min}] = 3600[\mathrm{s}]$
velocity	km/h	km per hour	$1[\mathrm{km/h}] = \frac{1}{3.6}[\mathrm{m/s}] = 0.2\overline{7}[\mathrm{m/s}]$
	mph	mile per hour	$1[\mathrm{mph}] = 0.447[\mathrm{m/s}] = 1.61[\mathrm{km/h}]$
temperature	°C	Celsius	$x[^\circ\mathrm{C}] = (x + 273.15)[^\circ\mathrm{K}]$
	°F	Fahrenheit	$x[^\circ\mathrm{F}] = \frac{5}{9}(x + 459.67)[^\circ\mathrm{K}]$
			$x[^\circ\mathrm{F}] = \frac{5}{9}(x - 32)[^\circ\mathrm{C}]$
pressure	atm	atmosphere	$1[\mathrm{atm}] = 101\,325[\mathrm{Pa}]$
	bar	bar	$1[\mathrm{bar}] =10^5[\mathrm{Pa}]$

This tutorial presents many explanations as blocks of code. Be sure to try the code examples on your own by typing the commands into SymPy. It's always important to verify for yourself!

Using SymPy

The easiest way to use SymPy, provided you're connected to the ~~Internet~~internet, is to visit http://live.sympy.org. You'll be presented with an interactive prompt into which you can enter your commands—right in your browser.

If you want to use SymPy on your own computer, you must install Python and the python package sympy. You can then open a command prompt and start a SymPy session using:

```
you@host$ python
Python X.Y.Z
[GCC a.b.c (Build Info)] on platform
Type "help", "copyright", or "license" for more information.
>>> from sympy import *
>>>
```

The >>> prompt indicates you're in the Python shell which accepts Python commands. The command from sympy import * imports all the SymPy functions into the current namespace. All SymPy functions are now available to you. To exit the python shell press CTRL+D.

I highly recommend you also install ipython, which is an improved interactive python shell. If you have ipython and SymPy installed, you can start an ipython shell with SymPy pre-imported using the command isympy. For an even better experience, you can try ~~ipython~~ jupyter notebook, which is a web frontend for the ipython shell.

Each section in this appendix begins with a python import statement for the functions used in that section. If you use the statement from sympy import * in the beginning of your code, you don't need to run these individual import statements, but I've included them so you'll know which SymPy vocabulary is covered in each section.

Fundamentals of mathematics

Let's begin by learning about the basic SymPy objects and the operations we can carry out on them. We'll learn the SymPy equivalents of the math verbs we used in Chapter 1: "to solve" (an equation), "to expand" (an expression), "to factor" (a polynomial).

The function `solve` is like a Swiss Army knife you can use to solve all kind of problems. Suppose you want to *complete the square* in the expression $x^2 - 4x + 7$, that is, you want to find constants h and k such that $x^2 - 4x + 7 = (x - h)^2 + k$. There is no special "complete the square" function in `SymPy`, but you can call `solve` on the equation $(x - h)^2 + k \; - \; (x^2 - 4x + 7) = 0$ to find the unknowns h and k:

```
>>> h, k = symbols('h k')
>>> solve( (x-h)**2 + k  - (x**2-4*x+7), [h,k] )
[(2, 3)]                                # so h = 2 and k = 3
>>> ((x-2)**2+3).expand()               # verify...
x**2 - 4*x + 7
```

Learn the basic `SymPy` commands and you'll never need to suffer another tedious arithmetic calculation painstakingly performed by hand again!

Rational functions

```
>>> from sympy import together, apart
```

By default, `SymPy` will not combine or split rational expressions. You need to use `together` to symbolically calculate the addition of fractions:

```
>>> a, b, c, d = symbols('a b c d')
>>> a/b + c/d
a/b + c/d
>>> together(a/b + c/d)
(a*d + b*c)/(b*d)
```

Alternately, if you have a rational expression and want to divide the numerator by the denominator, use the `apart` function:

```
>>> apart( (x**2+x+4)/(x+2)  )
x - 1  +  6/(x + 2)
```

Exponentials and logarithms

Euler's ~~constant~~ number $e = 2.71828\ldots$ is defined one of several ways,

$$e \equiv \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \equiv \lim_{\epsilon\to 0}(1+\epsilon)^{1/\epsilon} \equiv \sum_{n=0}^{\infty}\frac{1}{n!},$$

and is denoted `E` in `SymPy`. Using `exp(x)` is equivalent to `E**x`.

The functions `log` and `ln` both compute the logarithm base e:

Uniform acceleration motion (UAM)

Let's analyze the case where the net force on the object is constant. A constant force causes a constant acceleration $a = \frac{F}{m} = \text{constant}$. If the acceleration function is constant over time $a(t) = a$. We find $v(t)$ and $x(t)$ as follows:

```
>>> t, a, v_i, x_i = symbols('t \DIFaddbegin \DIFadd{a }\DIFaddend v_i x_i
>>> v = v_i + integrate(a, (t, 0,t) )
>>> v
a*t + v_i
>>> x = x_i + integrate(v, (t, 0,t) )
>>> x
a*t**2/2 + v_i*t + x_i
```

You may remember these equations from Section 2.4 (page 145). They are the *uniform accelerated motion* (UAM) equations:

$$\begin{aligned} a(t) &= a, \\ v(t) &= v_i + at, \\ x(t) &= x_i + v_i t + \frac{1}{2}at^2. \end{aligned}$$

In high school, you probably had to memorize these equations. Now you know how to derive them yourself starting from first principles.

For the sake of completeness, we'll now derive the fourth UAM equation, which relates the object's final velocity to the initial velocity, the displacement, and the acceleration, without reference to time:

```
>>> (v*v).expand()
a**2*t**2 + 2*a*t*v_i + v_i**2
>>> ((v*v).expand() - 2*a*x).simplify()
-2*a*x_i + v_i**2
```

The above calculation shows $v_f^2 - 2ax_f = -2ax_i + v_i^2$. After moving the term $2ax_f$ to the other side of the equation, we obtain

$$(v(t))^2 \;=\; v_f^2 = v_i^2 + 2a\Delta x \;=\; v_i^2 + 2a(x_f - x_i).$$

The fourth equation is important for practical purposes because it allows us to solve physics problems ~~in a time-less manner.~~ without using the time variable.

Example

Find the position function of an object at time $t = 3$[s], if it starts from $x_i = 20$[m] with $v_i = 10$[m/s] and undergoes a constant acceleration of $a = 5$[m/s^2]. What is the object's velocity at $t = 3$[s]?