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 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 the solution to an equation immediately. To find the solution, you must break the problem into simpler steps. Let's walk through this one together.

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* an equation: the equation is solved because the unknown is isolated on one side, while the constants are grouped on the other side. You can type the numbers on the right-hand side into a calculator and obtain the numerical value of x.

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 manipulate the original equation into its final form, simplifying it step by step until it can't be simplified any further. The only requirement is that the manipulations we make transform one true equation into another true equation. In this example, the first simplifying step is to add the number four to both sides of the equation:

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

which simplifies to

$$x^2 = 49.$$

Now the expression looks simpler, yes? How did I know to perform this operation? I wanted to "undo" the effects of the operation -4. We undo an operation by applying its *inverse*. In the case where the operation is the subtraction of some amount, the inverse operation is the addition of the same amount. We'll learn more about function inverses in Section 1.4.

We're getting closer to our goal of *isolating* x on one side of the equation, leaving only numbers on the other side. The next step is to undo the square x^2 operation. The inverse operation of squaring a number x^2 is to take its square root $\sqrt{\ }$, so that's what we'll do next. We obtain

$$\sqrt{x^2} = \sqrt{49}.$$

Notice how we applied the square root to both sides of the equation? If we don't apply the same operation to both sides, we'll break the equality!

The equation $\sqrt{x^2} = \sqrt{49}$ simplifies to

$$|x| = 7.$$

What's up with the vertical bars around x? The notation $|x|$ stands for the *absolute value* of x, which is the same as x except we ignore the sign that indicates whether x is positive or negative. For example $|5| = 5$ and $|-5| = 5$, too. The equation $|x| = 7$ indicates that both $x = 7$ and $x = -7$ satisfy the equation $x^2 = 49$. Seven squared is 49, $7^2 = 49$, and negative seven squared is also 49, $(-7)^2 = 49$, because the two negative signs cancel each other out.

The final solutions to the equation $x^2 - 4 = 45$ are

$$x = 7 \qquad \text{and} \qquad x = -7.$$

Yes, there are *two* possible answers. You can check that both of the above values of x satisfy the initial equation $x^2 - 4 = 45$.

If you are comfortable with all the notions of high school math and you feel you could have solved the equation $x^2 - 4 = 45$ on your own, then you can skim through this chapter quickly. If on the other hand you are wondering how the squiggle killed the power two, then this chapter is for you! In the following sections we will review all the essential concepts from high school math that you will need to power through the rest of this book. First, let me tell you about the different kinds of numbers.

1.2 Numbers

In the beginning, we must define the main players in the world of math: numbers.

Definitions

Numbers are the basic objects we use to count, measure, quantify, and calculate things. Mathematicians like to classify the different kinds of number-like objects into categories called *sets*:

- The natural numbers: $\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, 7, \ldots\}$
- The integers: $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
- The rational numbers: $\mathbb{Q} = \{\frac{5}{3}, \frac{22}{7}, 1.5, 0.125, -7, \ldots\}$
- The real numbers: $\mathbb{R} = \{-1, 0, 1, \sqrt{2}, e, \pi, 4.94\ldots, \ldots\}$
- The complex numbers: $\mathbb{C} = \{-1, 0, 1, i, 1+i, 2+3i, \ldots\}$

These categories of numbers should be somewhat familiar to you. Think of them as neat classification labels for everything that you would normally call a number. Each group in the above list is a *set*. A set is a collection of items of the same kind. Each collection has a name and a precise definition for which items belong in that collection. Note also that each of the sets in the list contains all the sets above it, as illustrated in Figure 1.2. For now, we don't need to go into the details of sets and set notation, but we do need to be aware of the different sets of numbers.

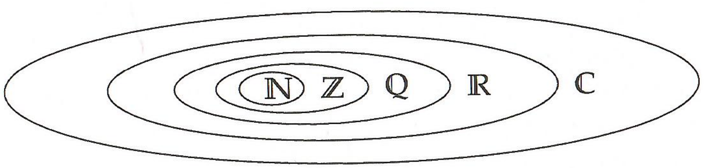


Figure 1.2: An illustration of the nested containment structure of the different number sets. The set of natural numbers is contained in the set of integers, which in turn is contained in the set of rational numbers. The set of rational numbers is contained in the set of real numbers, which is contained in the set of complex numbers.

Why do we need so many different sets of numbers? Each set of numbers is associated with more and more advanced mathematical problems.

The simplest numbers are the natural numbers $\mathbb{N}$, which are sufficient for all your math needs if all you're going to do is *count* things. How many goats? Five goats here and six goats there so the total is

11 goats. The sum of any two natural numbers is also a natural number.

As soon as you start using *subtraction* (the inverse operation of addition), you start running into negative numbers, which are numbers outside the set of natural numbers. If the only mathematical operations you will ever use are *addition* and *subtraction*, then the set of integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ will be sufficient. Think about it. Any integer plus or minus any other integer is still an integer.

You can do a lot of interesting math with integers. There is an entire field in math called *number theory* that deals with integers. However, to restrict yourself solely to integers is somewhat limiting—a rotisserie menu that offers $\frac{1}{2}$ of a chicken would be totally confusing.

If you want to use division in your mathematical calculations, you'll need the rationals $\mathbb{Q}$. The set of rational numbers corresponds to all numbers that can be expressed as *fractions* of the form $\frac{m}{n}$ where m and n are integers, and $n \neq 0$. You can add, subtract, multiply, and divide rational numbers, and the result will always be a rational number. However, even the rationals are not enough for all of math!

In geometry, we can obtain *irrational* quantities like $\sqrt{2}$ (the diagonal of a square with side 1) and π (the ratio between a circle's circumference and its diameter). There are no integers x and y such that $\sqrt{2} = \frac{x}{y}$, therefore we say that $\sqrt{2}$ is *irrational* (not in the set $\mathbb{Q}$). An irrational number has an infinitely long decimal expansion that doesn't repeat. For example, $\pi = 3.141592653589793\ldots$ where the dots indicate that the decimal expansion of π continues all the way to infinity.

Combining the irrational numbers with the rationals gives us all the useful numbers, which we call the set of real numbers $\mathbb{R}$. The set $\mathbb{R}$ contains the integers, the rational numbers $\mathbb{Q}$, as well as irrational numbers like $\sqrt{2} = 1.4142135\ldots$. By using the reals you can compute pretty much anything you want. From here on in the text, when I say *number*, I mean an element of the set of real numbers $\mathbb{R}$.

The only thing you can't do with the reals is to take the square root of a negative number—you need the complex numbers $\mathbb{C}$ for that. We defer the discussion on complex numbers until Section 1.14.

Operations on numbers

Addition

You can add numbers. I'll assume you're familiar with this stuff:

$$2 + 3 = 5, \quad 45 + 56 = 101, \quad 9\,999 + 1 = 10\,000.$$

You can visualize numbers as sticks of different length. Adding numbers is like adding sticks together: the resulting stick has a length equal to the sum of the lengths of the constituent sticks, as illustrated in Figure 1.3.

Figure 1.3: The addition of numbers corresponds to adding lengths.

Addition is *commutative*, which means that $a + b = b + a$. In other words, the order of the numbers in a summation doesn't matter. It is also *associative*, which means that if you have a long summation like $a + b + c$ you can compute it in any order $(a + b) + c$ or $a + (b + c)$, and you'll get the same answer.

Subtraction

Subtraction is the inverse operation of addition:

$$2 - 3 = -1, \quad 45 - 56 = -11, \quad 999 - 1 = 998.$$

Unlike addition, subtraction is not a commutative operation. The expression $a - b$ is not equal to the expression $b - a$, or written mathematically:

$$a - b \neq b - a.$$

Instead we have $b - a = -(a - b)$, which shows that changing the order of a and b in the expression changes its sign.

Subtraction is not associative either:

$$(a - b) - c \neq a - (b - c).$$

For example $(7 - 2) - 3 = 2$ while $7 - (2 - 3) = 8$.

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 width a and height b is equal to ab. A rectangle with a height equal to its width is a square, and this is why we call $aa = a^2$ "a squared."

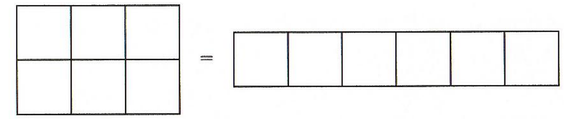


Figure 1.4: The area of a rectangle with width 3 m and height 2 m is equal to 6 m^2, which is equivalent to six squares with area 1 m^2 each.

Multiplication of numbers is also commutative, $ab = ba$, and associative, $abc = (ab)c = a(bc)$. In modern math notation, no special symbol is required 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.

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

Whatever a is, you need to divide it into b equal parts and take one such part.

Division is not a commutative operation since a/b is not equal to b/a. Division is not associative either: $(a \div b) \div c \neq a \div (b \div c)$. For example, when $a = 6$, $b = 3$, and $c = 2$, we get $(6/3)/2 = 1$ while $6/(3/2) = 4$.

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

The act of multiplying a number by itself many times is called *exponentiation*. We denote "a exponent n" using a superscript, where n is the number of times the base a is multiplied by itself:

$$a^n = \underbrace{aa \cdots a}_{n \text{ times}}.$$

In words, we say "a raised to the power of n."

To visualize how exponents work, we can draw a connection between the value of exponents and the dimensions of geometric objects. Figure 1.5 illustrates how the same length 2 corresponds to different geometric objects when raised to different exponents. The number 2 corresponds to a line segment of length two, which is a geometric object in a one-dimensional space. If we add a line segment of length two in a second dimension, we obtain a square with area 2^2 in a two-dimensional space. Adding a third dimension, we obtain a cube with volume 2^3 in a three-dimensional space. Indeed, raising a base a to the exponent 2 is commonly called "a squared," and raising a to the power of 3 is called "a cubed."

The geometrical analogy about one-dimensional quantities as lengths, two-dimensional quantities as areas, and three-dimensional quantities as volumes is good to keep in mind.

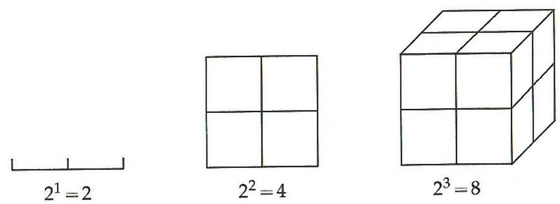


Figure 1.5: Geometric interpretation for exponents 1, 2, and 3. A length raised to exponent 2 corresponds to the area of a square. The same length raised to exponent 3 corresponds to the volume of a cube.

Our visual intuition works very well up to three dimensions, but we can use other means of visualizing higher exponents, as demonstrated in Figure 1.6.

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. Parentheses
2. Exponents
3. Multiplication and Division
4. Addition and Subtraction

If you're seeing this list for the first time, the acronym PEMDAS and the associated mnemonic "Please Excuse My Dear Aunt Sally," might help you remember the order of operations.

For instance, the expression $5 \cdot 3^2 + 13$ is interpreted as "First find the square of 3, then multiply it by 5, and then add 13." Parentheses 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

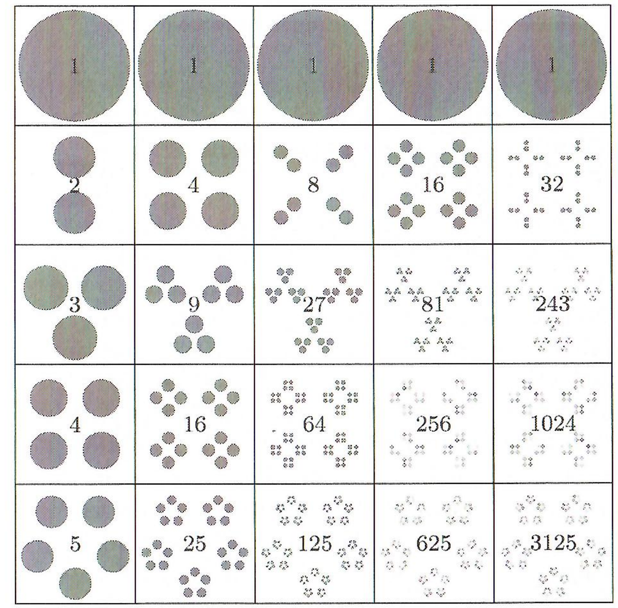


Figure 1.6: Visualization of numbers raised to different exponents. Each box in this grid contains a^n dots, where the base a varies from one through five, and the exponent n varies from one through five. In the first row we see that the number $a = 1$ raised to any exponent is equal to itself. The second row corresponds to the base $a = 2$ so the number of dots doubles each time we increase the exponent by one. Starting from $2^1 = 2$ in the first column, we end up with $2^5 = 32$ in the last column. The rest of the rows show how exponentiation works for different bases.

$(5 \cdot 3)^2 + 13$, where parentheses indicate that the square acts on $(5 \cdot 3)$ as a whole and not on 3 alone.

Exercises

E1.1 Solve for the unknown x in the following equations:

a) $3x + 2 - 5 = 4 + 2$ **b)** $\frac{1}{2}x - 3 = \sqrt{3} + 12 - \sqrt{3}$

c) $\frac{7x-4}{2} + 1 = 8 - 2$ **d)** $5x - 2 + 3 = 3x - 5$

E1.2 Indicate all the number sets the following numbers belong to.

a) -2 **b)** $\sqrt{-3}$ **c)** $8 \div 4$ **d)** $\frac{5}{3}$ **e)** $\frac{\pi}{2}$

E1.3 Calculate the values of the following expressions:

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

1.3 Variables

In math we use a lot of *variables* and *constants*, which are placeholder names for *any* number or unknown. Variables allow us to perform calculations without knowing all the details.

Example You're having tacos for lunch today and wondering how many you can eat without going over your caloric budget. Your goal is to eat 800 calories for lunch and you want to do the calculation before getting to the restaurant because you fear your math abilities might be affected in the presence of tacos. You're not sure how many calories each taco contains, so you invent the variable c to denote this unknown. You also define the variable x to represent the number of tacos you will eat, and come up with the equation $800 = cx$ to represent the total number of calories of your lunch. Solving for x, you find the total number of tacos you should order is $x = \frac{800}{c}$. If the restaurant serves tacos that contain $c = 200$ calories each, then you should order $x = \frac{800}{200} = 4$ of them. If the restaurant serves only giant tacos worth $c = 400$ calories each, then you can only eat $x = \frac{800}{400} = 2$ of them. Observe we were able to solve for x even before knowing the value of c.

Variable names

There are common naming patterns for variables:

- x: name used for the unknown in equations. We also use x to denote function inputs and the position of objects in physics.
- i, j, k, m, n: common names for integer variables
- a, b, c, d: letters near the beginning of the alphabet are often used to denote constants (fixed quantities that do not change).
- θ, ϕ: the Greek letters *theta* and *phi* are used to denote angles
- C: costs in business, along with P for profit, and R for revenue
- X: capital letters are used to denote random variables in probability theory

Variable substitution

We can often *change variables* and replace one unknown variable with another to simplify an equation. For example, say you don't feel comfortable around square roots. Every time you see a square root, you freak out until one day you find yourself taking an exam trying

to solve for x in the following equation:

$$\frac{6}{5-\sqrt{x}} = \sqrt{x}.$$

Don't freak out! In crucial moments like this, substitution can help with your root phobia. Just write, "Let $u = \sqrt{x}$" on your exam, and voila, you can rewrite the equation in terms of the variable u:

$$\frac{6}{5-u} = u,$$

which contains no square roots.

The next step to solve for u is to undo the division operation. Multiply both sides of the equation by $(5-u)$ to obtain

$$\frac{6}{5-u}(5-u) = u(5-u),$$

which simplifies to

$$6 = 5u - u^2.$$

This can be rewritten as the equation $u^2 - 5u + 6 = 0$, which in turn can be rewritten as $(u-2)(u-3) = 0$ using the techniques we'll learn in Section 1.6.

We now see that the solutions are $u_1 = 2$ and $u_2 = 3$. The last step is to convert our u-answers into x-answers by using $u = \sqrt{x}$, which is equivalent to $x = u^2$. The final answers are $x_1 = 2^2 = 4$ and $x_2 = 3^2 = 9$. Try plugging these x values into the original square root equation to verify that they satisfy it.

Compact notation

Symbolic manipulation is a powerful tool because it allows us to manage complexity. Say you're solving a physics problem in which you're told the mass of an object is $m = 140$ kg. If there are many steps in the calculation, would you rather use the number 140 kg in each step, or the shorter symbol m? It's much easier to use m throughout your calculation, and wait until the last step to substitute the value 140 kg when computing the final numerical answer.

1.4 Functions and their inverses

As we saw in the section on solving equations, the ability to "undo" functions is a key skill for solving equations.

Example Suppose we're solving for x in the equation

$$f(x) = c,$$

where f is some function and c is some constant. We're looking for the unknown x such that $f(x)$ equals c. Our goal is to isolate x on one side of the equation, but the function f stands in our way.

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

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

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 inverse function. This notation is inspired by the notation for reciprocals. Recall that multiplication by the reciprocal 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 inverse function. In other words, the number $f^{-1}(y)$ is equal to the number x such that $f(x) = y$.

Be careful: sometimes an equation can have 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 $f^{-1}(y) = \sqrt{y}$, 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:

$$\begin{aligned}
\text{function } f(x) &\Leftrightarrow \text{inverse } f^{-1}(x) \\
x+2 &\Leftrightarrow x-2 \\
2x &\Leftrightarrow \tfrac{1}{2}x \\
-1x &\Leftrightarrow -1x \\
x^2 &\Leftrightarrow \pm\sqrt{x} \\
2^x &\Leftrightarrow \log_2(x)
\end{aligned}$$

$$\begin{aligned}
3x+5 &\Leftrightarrow \tfrac{1}{3}(x-5) \\
a^x &\Leftrightarrow \log_a(x) \\
\exp(x) = e^x &\Leftrightarrow \ln(x) = \log_e(x) \\
\sin(x) &\Leftrightarrow \sin^{-1}(x) = \arcsin(x) \\
\cos(x) &\Leftrightarrow \cos^{-1}(x) = \arccos(x)
\end{aligned}$$

The function-inverse relationship is *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.

Don't be surprised to see $-1x \Leftrightarrow -1x$ in the list of function inverses. Indeed, the opposite operation of multiplying by -1 is to multiply by -1 once more: $(-(-x) = x)$.

Example 1

If you want to solve the equation $x - 4 = 5$, you can apply the inverse function of $x - 4$, which is $x + 4$. After adding four to both sides of the equation, $x - 4 + 4 = 5 + 4$, we obtain the answer $x = 9$.

Example 2

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:

$$\log_5\left(3 + \sqrt{6\sqrt{x} - 7}\right) = 34 + \sin(8) - \Psi(1).$$

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

First, note that it doesn't matter what Ψ (the 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 $\sin(8)$. If you don't have a calculator handy, don't worry about it. Keep the expression $\sin(8)$ 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:

$$5^{\log_5\left(3+\sqrt{6\sqrt{x}-7}\right)} = 5^{34+\sin(8)-\Psi(1)},$$

which simplifies to

$$3+\sqrt{6\sqrt{x}-7} = 5^{34+\sin(8)-\Psi(1)},$$

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:

$$\sqrt{6\sqrt{x}-7} = 5^{34+\sin(8)-\Psi(1)} - 3.$$

To undo a square root we take the square:

$$6\sqrt{x}-7 = \left(5^{34+\sin(8)-\Psi(1)} - 3\right)^2.$$

Add 7 to both sides,

$$6\sqrt{x} = \left(5^{34+\sin(8)-\Psi(1)} - 3\right)^2 + 7,$$

divide by 6

$$\sqrt{x} = \frac{1}{6}\left(\left(5^{34+\sin(8)-\Psi(1)} - 3\right)^2 + 7\right),$$

and square again to find the final answer:

$$x = \left[\frac{1}{6}\left(\left(5^{34+\sin(8)-\Psi(1)} - 3\right)^2 + 7\right)\right]^2.$$

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:

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

We'll show a formula for solving quadratic equations in Section 1.6. Solving cubic equations like $ax^3 + bx^2 + cx + d = 0$ using a formula 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 in the following equations:

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

E1.5 Find the function inverse and use it to solve the problems.

a) Solve the equation $f(x) = 4$, where $f(x) = \sqrt{x}$.

b) Solve for x in the equation $g(x) = 1$, given $g(x) = e^{-2x}$.

1.5 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.

Let's define some terminology for referring to different parts of math expressions. 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 three numbers a, b, and c, we can apply the following algebraic properties:

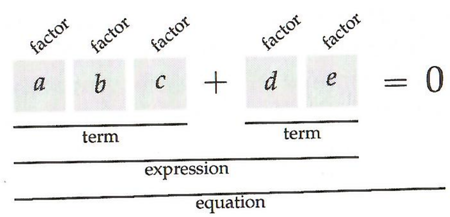


Figure 1.7: Diagram showing the names used to describe the different parts of the equation $abc + de = 0$.

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; as a factor consisting 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 all algebra operations and illustrate what they're capable of.

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 can 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 "factor it out" by rewriting the equation as

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

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

Expanding brackets

To *expand* a bracket is to multiply each term inside the bracket by the factor outside the bracket. The key thing to remember when expanding brackets is to apply the *distributive* property: $a(x + y) = ax + ay$. For longer expressions, we may need to apply the distributive property several times, until there are no more brackets left:

$$\begin{aligned}(a + b)(x + y + z) &= a(x + y + z) + b(x + y + z)\\ &= ax + ay + az + bx + by + bz.\end{aligned}$$

After expanding the brackets in this expression, we end up with six terms—one term for each of the six possible combinations of products between the terms in $(a + b)$ and the terms in $(x + y + z)$.

The distributive property is often used to manipulate expressions containing different powers of the variable x. For instance,

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

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

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

The bracket-expanding and simplification techniques demonstrated above are very common in math, and I recommend you solve some algebra practice problems to get the hang of them. Most math textbooks skip simplification steps and jump straight to the answer, since they assume readers are capable of doing simplifications on their own. It would be too long (and annoying) to show the simplifications for each expression. For example, the sentence "We can rewrite $(x + 3)(x + 2)$ as $x^2 + 5x + 6$," is the short version of the longer sentence, "We can apply the distributive property twice on $(x + 3)(x + 2)$ then combine the terms with the same power of x to get $x^2 + 5x + 6$."

It's not unusual for people to make math mistakes when expanding brackets and manipulating long algebra expressions. To avoid mistakes, use a step-by-step approach and apply only one operation in each step. Write legibly and keep the equations "organized" so it's easy to check the calculations performed in each step. 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 we sorted the terms in the final expression according to the different powers of x, with the terms containing x^2 grouped together, and the terms containing x grouped together. This approach helps keep things organized when dealing with expressions containing many terms.

Factoring

Factoring involves "taking out" the common parts of a complicated expression in order to make the expression more compact. Suppose we're given the expression $6x^2y + 15x$. We can simplify this expression by taking out the common factors and moving them in front of a bracket. Let's see how to do this, step by step.

The expression $6x^2y + 15x$ has two terms. Let's split each term into its constituent factors:

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

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

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

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

Here's another example of factoring—notice the common factors are taken out and moved in front of the bracket:

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

Factoring quadratic expressions

A *quadratic expression* is an expression of the form $ax^2 + bx + c$. The expression contains a *quadratic term* ax^2, a *linear term* bx, and a constant term c. The numbers a, b, and c are called *coefficients*: the quadratic coefficient is a, the linear coefficient is b, and the constant coefficient is c.

To *factor* the quadratic expression $ax^2 + bx + c$ is to rewrite it as the product of a constant and two factors like $(x + p)$ and $(x + q)$:

$$ax^2 + bx + c = a(x + p)(x + q).$$

Rewriting quadratic expressions in factored form helps us better understand and describe their properties.

Example Suppose we're asked to describe the properties of the function $f(x) = x^2 - 5x + 6$. Specifically, we're asked to find the function's *roots*, which are the values of x for which the function equals zero.

Factoring the expression $x^2 - 5x + 6$ helps us see its properties more clearly, and makes our job of finding the roots of $f(x)$ easier. The factored form of this quadratic expression is

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

Now we can see at a glance that the values of x for which $f(x) = 0$ are $x = 2$ and $x = 3$. When $x = 2$, the factor $(x - 2)$ is zero and hence $f(x) = 0$. Similarly, when $x = 3$, the factor $(x - 3)$ is zero so $f(x) = 0$.

How did we know that the factors of $x^2 - 5x + 6$ are $(x - 2)$ and $(x - 3)$ in the above example? For simple quadratics like the one above, we can simply *guess* the values of p and q in the equation $x^2 - 5x + 6 = (x + p)(x + q)$. Before we start guessing, let's look at the expanded version of the product between $(x + p)$ and $(x + q)$:

$$(x + p)(x + q) = x^2 + (p + q)x + pq.$$

Note the linear term on the right-hand side contains the sum of the unknowns $(p + q)$, while the constant term contains their product pq. If we want the equation $x^2 - 5x + 6 = x^2 + (p + q)x + pq$ to hold, we must find two numbers p and q whose sum equals -5 and whose product equals 6. After a couple of attempts we find $p = -2$ and $q = -3$. This guessing approach is an effective strategy for many of the factoring problems we will likely be asked to solve, since math teachers often choose simple numbers like ± 1, ± 2, ± 3, or ± 4 for the constants p and q. For more complicated quadratic expressions, we'll need to use the quadratic formula, which we'll talk about in Section 1.6.

Common quadratic forms

Let's look at some common variations of quadratic expressions you might encounter when doing algebra calculations.

The quadratic expression $x^2 - p^2$ is called a *difference of squares*, and it can be obtained by multiplying the factors $(x + p)$ and $(x - p)$:

$$(x + p)(x - p) = x^2 \cancel{-xp} \cancel{+px} - p^2 = x^2 - p^2.$$

There's no linear term because the $-xp$ term cancels the $+px$ term. Any time you see an expression like $x^2 - p^2$, you can know it comes from a product of the form $(x + p)(x - p)$.

A *perfect square* is a quadratic expression that can be written as the product of repeated factors $(x+p)$:

$$x^2 + 2px + p^2 = (x+p)(x+p) = (x+p)^2.$$

Note $x^2 - 2qx + q^2 = (x-q)^2$ is also a perfect square.

Completing the square

In this section we'll learn about an ancient algebra technique called *completing the square*, which allows us to rewrite *any* quadratic expression of the form $x^2 + Bx + C$ as a perfect square plus some constant correction factor $(x+p)^2 + k$. This algebra technique was described in one of the first books on *al-jabr* (algebra), written by Al-Khwarizmi around the year 800 CE. The name "completing the square" comes from the ingenious geometric construction used by this procedure. Yes, we can use geometry to solve algebra problems!

We assume the starting point for the procedure is a quadratic expression whose quadratic coefficient is one, $1x^2 + Bx + C$, and use capital letters B and C to denote the linear and constant coefficients. The capital letters are to avoid any confusion with the quadratic expression $ax^2 + bx + c$, for which $a \neq 1$. Note we can always write $ax^2 + bx + c$ as $a(x^2 + \frac{b}{a}x + \frac{c}{a})$ and apply the procedure to the expression inside the brackets, identifying $\frac{b}{a}$ with B and $\frac{c}{a}$ with C.

First let's rewrite the quadratic expression $x^2 + Bx + C$ by splitting the linear term into two equal parts:

$$x^2 + \tfrac{B}{2}x + \tfrac{B}{2}x + C.$$

We can interpret the first three terms geometrically as follows: the x^2 term corresponds to a square with side length x, while the two $\frac{B}{2}x$ terms correspond to rectangles with sides $\frac{B}{2}$ and x. See the left side of Figure 1.8 for an illustration.

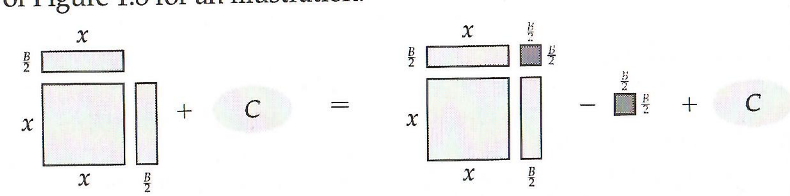


Figure 1.8: To complete the square in the expression $x^2 + Bx + C$, we need to add the quantity $(\frac{B}{2})^2$, which corresponds to a square (shown in darker colour) with sides equal to half the coefficient of the linear term. We also subtract $(\frac{B}{2})^2$ so the overall value of the expression remains unchanged.

The square with area x^2 and the two rectangles can be positioned to form a larger square with side length $(x + \frac{B}{2})$. Note there's a small piece of sides $\frac{B}{2}$ by $\frac{B}{2}$ missing from the corner. To *complete the square*, we can add a term $(\frac{B}{2})^2$ to this expression. To preserve the equality, we also subtract $(\frac{B}{2})^2$ from the expression to obtain:

$$\begin{aligned} x^2 + \tfrac{B}{2}x + \tfrac{B}{2}x + C &= \underbrace{x^2 + \tfrac{B}{2}x + \tfrac{B}{2}x + \left(\tfrac{B}{2}\right)^2} - \left(\tfrac{B}{2}\right)^2 + C \\ &= \qquad \left(x + \tfrac{B}{2}\right)^2 \qquad - \left(\tfrac{B}{2}\right)^2 + C. \end{aligned}$$

The right-hand side of this equation describes the area of the square with side length $(x + \frac{B}{2})$, minus the area of the small square $(\frac{B}{2})^2$, plus the constant C, as illustrated on the right side of Figure 1.8.

We can summarize the entire procedure in one equation:

$$x^2 + Bx + C = (x \underbrace{+\tfrac{B}{2}}_{(1)})^2 + C \underbrace{-\left(\tfrac{B}{2}\right)^2}_{(2)}.$$

There are two things to remember when you want to apply the complete-the-square trick: (1) choose the constant inside the bracket to be $\frac{B}{2}$ (half of the linear coefficient), and (2) subtract $(\frac{B}{2})^2$ outside the bracket in order to keep the equation balanced.

Solving quadratic equations

Suppose we want to solve the quadratic equation $x^2 + Bx + C = 0$. It's not possible to solve this equation with the digging-toward-the-x approach from Section 1.1 (since x appears in both the quadratic term x^2 and the linear term Bx). Enter the completing-the-square trick!

Example Let's find the solutions of the equation $x^2 + 5x + 6 = 0$. The coefficient of the linear term is $B = 5$, so we choose $\frac{B}{2} = \frac{5}{2}$ for the constant inside the bracket, and subtract $(\frac{B}{2})^2 = (\frac{5}{2})^2$ outside the bracket to keep the equation balanced. Completing the square gives

$$x^2 + 5x + 6 = \left(x + \tfrac{5}{2}\right)^2 + 6 - \left(\tfrac{5}{2}\right)^2 = 0.$$

Next we use fraction arithmetic to simplify the constant terms in the expression: $6 - (\frac{5}{2})^2 = 6 \cdot \frac{4}{4} - \frac{25}{4} = \frac{24-25}{4} = \frac{-1}{4} = -0.25$.

We're left with the equation

$$(x + 2.5)^2 - 0.25 = 0,$$

which we can now solve by digging toward x. First move 0.25 to the right-hand side to get $(x+2.5)^2 = 0.25$. Then take the square root on both sides to obtain $(x + 2.5) = \pm 0.5$, which simplifies to

$x = -2.5 \pm 0.5$. The two solutions are $x = -2.5 + 0.5 = -2$ and $x = -2.5 - 0.5 = -3$. You can verify these solutions by substituting the values in the original equation $(-2)^2 + 5(-2) + 6 = 0$ and similarly $(-3)^2 + 5(-3) + 6 = 0$. Congratulations, you just solved a quadratic equation using a 1200-year-old algebra technique!

In the next section, we'll learn how to leverage the complete-the-square trick to obtain a general-purpose formula for quickly solving quadratic equations.

Exercises

E1.6 Factor the following quadratic expressions:

a) $x^2 - 8x + 7$ **b)** $x^2 + 4x + 4$ **c)** $x^2 - 9$

Hint: Guess the values p and q in the expression $(x+p)(x+q)$.

E1.7 Solve the equations by completing the square.

a) $x^2 + 2x - 15 = 0$ **b)** $x^2 + 4x + 1 = 0$

1.6 Solving quadratic equations

What would you do if asked to solve for x in the quadratic equation $2x^2 = 4x + 6$? 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:

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

This is called the *standard form* of the quadratic equation. We obtain this form 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 $2x^2 = 4x + 6$ into standard form, we subtract $4x + 6$ from both sides of the equation to obtain $2x^2 - 4x - 6 = 0$. What are the values of x that satisfy this equation?

Quadratic formula

The solutions to the equation $ax^2 + bx + c = 0$ for $a \neq 0$ are

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

The quadratic formula is usually abbreviated $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$, where the sign "$\pm$" stands for both "$+$" and "$-$." The notation "$\pm$" allows us to express both solutions x_1 and x_2 in one equation, but you should keep in mind there are really two solutions.

Let's see how the quadratic formula is used to solve the equation $2x^2 - 4x - 6 = 0$. Finding the two solutions requires the simple mechanical task of identifying $a = 2$, $b = -4$, and $c = -6$, then plugging these values into the two parts of the formula:

$$x_1 = \frac{4 + \sqrt{4^2 - 4(2)(-6)}}{4} = \frac{4 + \sqrt{16 + 48}}{4} = \frac{4 + \sqrt{64}}{4} = 3,$$

$$x_2 = \frac{4 - \sqrt{4^2 - 4(2)(-6)}}{4} = \frac{4 - \sqrt{16 + 48}}{4} = \frac{4 - \sqrt{64}}{4} = -1.$$

We can easily verify that value $x_1 = 3$ and $x_2 = -1$ both satisfy the original equation $2x^2 = 4x + 6$.

Proof of the quadratic formula

Every claim made by a mathematician comes with a *proof*, which is a step-by-step argument that shows why the claim is true. It's easy to see where a proof starts and where a proof ends in mathematical texts. Each proof begins with the heading *Proof* (usually in italics) and has the symbol "□" at its end. The purpose of these demarcations is to give readers the option to skip the proof. It's not necessary to read and understand the proofs of all math statements, but reading proofs can often lead you to a more solid understanding of the material.

I want you to see the proof of the quadratic formula because it's an important result that you'll use very often to solve math problems. Reading the proof will help you understand where the quadratic formula comes from. The proof relies on the completing-the-square technique from the previous section, and some basic algebra operations. You can totally handle this!

Proof. We're starting from the quadratic equation $ax^2 + bx + c = 0$, and we're making the additional assumption that $a \neq 0$. We want to find the value or values of x that satisfy this equation.

The first thing we want to do is divide by a to obtain the equivalent equation

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

We are allowed to divide by a since we assumed that $a \neq 0$.

Next we apply the *complete the square* trick to the quadratic expression, to obtain an equivalent expression of the form $(x+?)^2+?$. Recall that the trick for completing the square is to choose the number inside the bracket to be half the coefficient of the linear term of the quadratic expression, which is $\frac{b}{2a}$ in this case. We must also subtract the square of this term outside the bracket in order to maintain the equality. After completing the square, we're left with the following equation:

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

From here, we use the standard digging-toward-the-x procedure. Move all constants to the right-hand side,

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

and take the square root of both sides to undo the square function:

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

Since any number and its opposite have the same square, taking the square root gives us two possible solutions, which we denote using the "$\pm$" symbol.

Next we subtract $\frac{b}{2a}$ from both sides of the equation to isolate x and obtain $x = -\frac{b}{2a} \pm \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}}$. We tidy up the mess under the square root, $\sqrt{\frac{b^2}{4a^2} - \frac{c}{a}} = \sqrt{\frac{b^2}{4a^2} - \frac{4a \cdot c}{4a \cdot a}} = \sqrt{\frac{b^2-4ac}{4a^2}} = \frac{\sqrt{b^2-4ac}}{2a}$, and add the fractions on the right-hand side to obtain $x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$. The solutions to the quadratic equation $ax^2 + bx + c = 0$ are

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

This completes the proof of the quadratic formula. □

The expression $b^2 - 4ac$ is called the *discriminant* of the equation. The discriminant tells us important information about the solutions of the equation $ax^2 + bx + c = 0$. The solutions x_1 and x_2 correspond to real numbers if the discriminant is positive or zero: $b^2 - 4ac \geqslant 0$. When the discriminant is zero ($b^2 - 4ac = 0$), the equation has only one solution since $x_1 = x_2 = \frac{-b}{2a}$. If the discriminant is negative, $b^2 - 4ac < 0$, the quadratic formula requires computing the square root of a negative number, which is not allowed for real numbers.

Alternative proof

To prove the quadratic formula, we don't necessarily need to show the algebra steps we followed to obtain the formula as outlined above. The quadratic formula states that x_1 and x_2 are solutions. To prove the formula is correct we can simply plug x_1 and x_2 into the equation $ax^2 + bx + c = 0$ to verify that x_1 and x_2 are solutions. Verify this on your own.

Applications

The golden ratio

The *golden ratio* is an essential proportion in geometry, art, aesthetics, biology, and mysticism, and is usually denoted as $\varphi = \frac{1+\sqrt{5}}{2} = 1.6180339\ldots$. This ratio is determined as the positive solution to the quadratic equation

$$x^2 - x - 1 = 0.$$

Applying the quadratic formula to this equation yields two solutions,

$$x_1 = \frac{1+\sqrt{5}}{2} = \varphi \quad \text{and} \quad x_2 = \frac{1-\sqrt{5}}{2} = -\frac{1}{\varphi}.$$

You can learn more about the various contexts in which the golden ratio appears from the Wikipedia article on the subject.

Explanations

Multiple solutions

Often, we are interested in only one of the two solutions to the quadratic equation. It will usually be obvious from the context of the problem which of the two solutions should be kept and which should be discarded. For example, the *time of flight* of a ball thrown in the air from a height of 3 metres with an initial velocity of 12 metres per second is obtained by solving the equation $(-4.9)t^2 + 12t + 3 = 0$. The two solutions of the quadratic equation are $t_1 = -0.229$ and $t_2 = 2.678$. The first answer t_1 corresponds to a time in the past so we reject it as invalid. The correct answer is t_2. The ball will hit the ground after $t = 2.678$ seconds.

Relation to factoring

In the previous section we discussed the *quadratic factoring* operation by which we could rewrite a quadratic function as the product of a constant and two factors:

$$f(x) = ax^2 + bx + c = a(x - x_1)(x - x_2).$$

The two numbers x_1 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 $a(x - x_1)(x - x_2)$.

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 real 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 (for now). We'll come back to the idea of taking square roots of negative numbers in Section 1.14 (see page 96).

Links

[Algebra explanation of the quadratic formula]
`https://www.youtube.com/watch?v=r3SEkdtpobo`

[Visual explanation of the quadratic formula derivation]
`https://www.youtube.com/watch?v=EBbtoFMJvFc`

Exercises

E1.8 Solve for x in the quadratic equation $2x^2 - x = 3$.

E1.9 Solve for x in the equation $x^4 - 4x^2 + 4 = 0$.
Hint: Use the substitution $y = x^2$.

1.7 The Cartesian plane

The Cartesian plane, named after famous philosopher and mathematician René Descartes, is used to visualize pairs of numbers (x, y).

Consider first the *number line* representation for numbers.

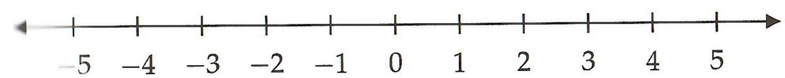


Figure 1.9: Every real number x corresponds to a point on the number line. The number line extends indefinitely to the left (toward negative infinity) and to the right (toward positive infinity).

The Cartesian plane is the two-dimensional generalization of the number line. 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.

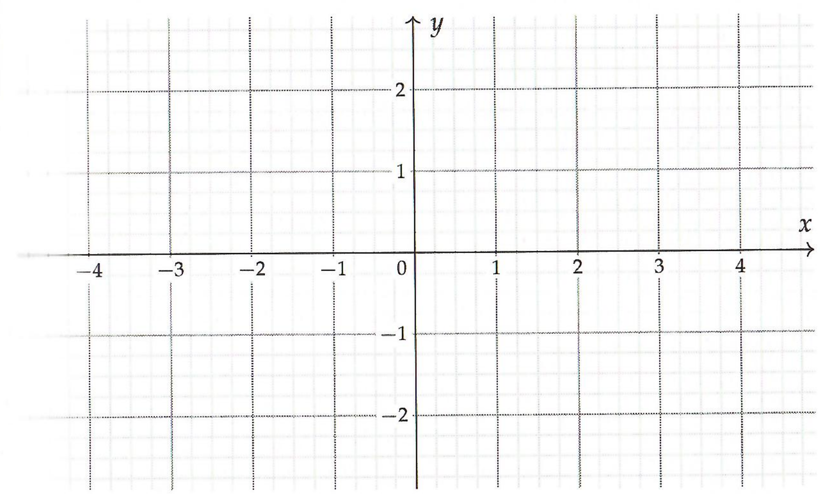


Figure 1.10: Every point in the Cartesian plane corresponds to a pair of real numbers (x, y). Points $P = (P_x, P_y)$, vectors $\vec{v} = (v_x, v_y)$, and graphs of functions $(x, f(x))$ live here.

Figure 1.10 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 (0,0)—and move a distance P_x on the x-axis, then move a distance P_y on the y-axis.

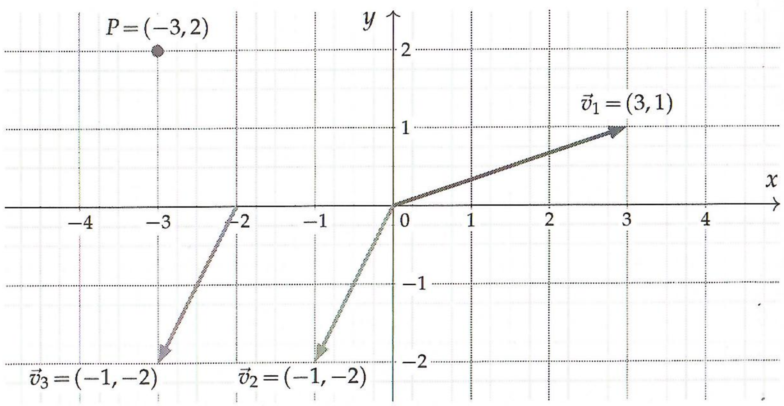


Figure 1.11: A Cartesian plane which shows the point $P = (-3, 2)$ and the vectors $\vec{v}_1 = (3,1)$ and $\vec{v}_2 = \vec{v}_3 = (-1,-2)$.

Similar to a point, a vector $\vec{v} = (v_x, v_y)$ is a pair of coordinates. Unlike points, we don't necessarily start from the plane's origin when mapping vectors. We draw vectors as arrows that explicitly mark where the vector starts and where it ends. Note that vectors $\vec{v}_2$ and $\vec{v}_3$ illustrated in Figure 1.11 are actually the *same* vector—the "displace left by 1 and down by 2" vector. It doesn't matter where you draw this vector, it will always be the same whether it begins at the plane's origin or elsewhere.

Graphs of functions

The Cartesian plane is great for visualizing functions. You can think of a function as a set of input-output pairs $(x, f(x))$. You can draw the *graph* of a function by letting the y-coordinate represent the function's output value:

$$(x, y) = (x, f(x)).$$

For example, with the function $f(x) = x^2$, we can pass a line through the set of points

$$(x, y) = (x, x^2),$$

and obtain the graph shown in Figure 1.12.

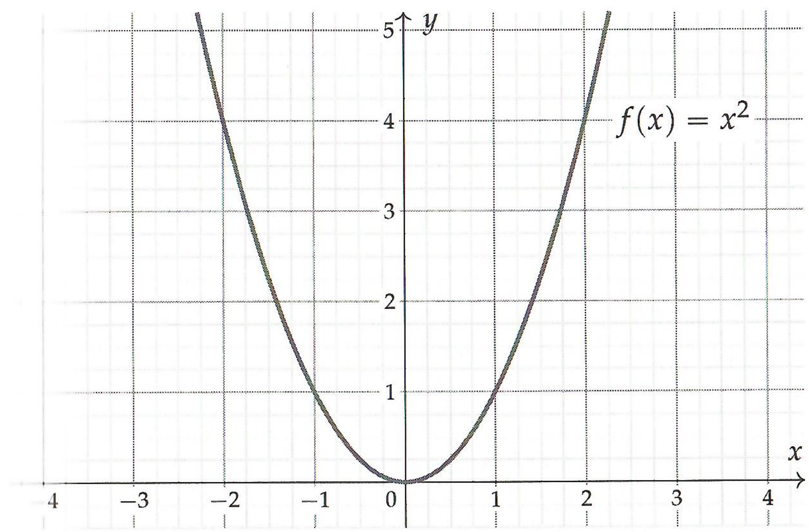


Figure 1.12: The graph of the function $f(x) = x^2$ consists of all pairs of points (x, y) in the Cartesian plane that satisfy $y = x^2$.

When plotting functions by setting $y = f(x)$, we use a special terminology for the two axes. The x-axis represents the *independent* variable (the one that varies freely), and the y-axis represents the *dependent* variable $f(x)$, since $f(x)$ depends on x.

To draw the graph of any function $f(x)$, use the following procedure. Imagine making a sweep over all of the possible input values for the function. For each input x, put a point at the coordinates $(x, y) = (x, f(x))$ in the Cartesian plane. Using the graph of a function, you can literally *see* what the function does: the "height" y of the graph at a given x-coordinate tells you the value of the function $f(x)$.

Dimensions

The number line is one-dimensional. Every number x can be visualized as a point on the number line. The Cartesian plane has two dimensions: the x dimension and the y dimension. If we need to visualize math concepts in 3D, we can use a three-dimensional coordinate system with x, y, and z axes (see Figure 1.55 on page 92).

1.8 Functions

We need to have a relationship talk. We need to talk about functions. We use functions to describe the relationships between variables. In particular, functions describe how one variable *depends* on another.

For example, the revenue R from a music concert depends on the number of tickets sold n. If each ticket costs \$25, the revenue from the concert can be written *as a function of* n as follows: $R(n) = 25n$. Solving for n in the equation $R(n) = 7000$ tells us the number of ticket sales needed to generate \$7000 in revenue. This is a simple model of a function; as your knowledge of functions builds, you'll learn how to build more detailed models of reality. For instance, if you need to include a 5% processing charge for issuing the tickets, you can update the revenue model to $R(n) = 0.95 \cdot 25 \cdot n$. If the estimated cost of 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 as you learn 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 **properties**. 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. 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 produces 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: $f\colon \mathbb{R} \to \mathbb{R}$.

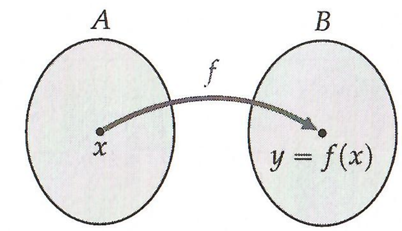


Figure 1.13: An abstract representation of a function f from the set A to the set B. The function f is the arrow which *maps* each input x in A to an output $f(x)$ in B. The output of the function $f(x)$ is also denoted y.

A function is not a number; rather, it is a *mapping* from numbers to numbers. We say "f maps x to $f(x)$." For any input x, the output value of f for that input is denoted $f(x)$, which is read as "f of x."

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

- A: the *source set* of the function describes the types of numbers that the function takes as inputs.
- $\text{Dom}(f)$: the *domain* of a function is the set of allowed input values for the function.
- B: the *target set* of a function describes the type of outputs the function has. The target set is sometimes called the *codomain*.
- $\text{Im}(f)$: the *image* of the function is the set of all possible output values of the function. The image is sometimes called the *range*.

See Figure 1.14 for an illustration of these concepts. The purpose of introducing all this math terminology is so we'll have words to distinguish the general types of inputs and outputs of the function (real numbers, complex numbers, vectors) from the specific properties of the function like its domain and image.

Let's look at an example to illustrate the difference between the source set and the domain of a function. Consider the square root function $f\colon \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \sqrt{x}$, which is shown in Figure 1.15. The source set of f is the set of real numbers—yet only nonnegative real numbers are allowed as inputs, since $\sqrt{x}$ is not defined for negative numbers. Therefore, the domain of the square root function is only the nonnegative real numbers: $\text{Dom}(f) = \mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$. Knowing the domain of a function is essential to using the function correctly. In this case, whenever you use the square root function, you need to make sure that the inputs to the function are nonnegative numbers.

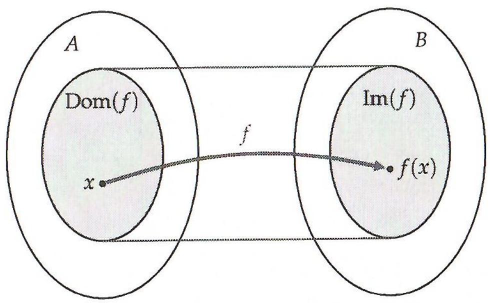


Figure 1.14: Illustration of the input and output sets of a function $f\colon A \to B$. The *source set* is denoted A and the *domain* is denoted $\text{Dom}(f)$. Note that the function's domain is a subset of its source set. The *target set* is denoted B and the *image* is denoted $\text{Im}(f)$. The image is a subset of the target set.

The complicated-looking expression between the curly brackets uses *set notation* to define the set of nonnegative numbers $\mathbb{R}_+$. In words, the expression $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$ states that "$\mathbb{R}_+$ is defined as the set of all real numbers x such that x is greater than or equal to zero." We'll discuss set notation in more detail in Section 1.16. For now, you can just remember that $\mathbb{R}_+$ represents the set of nonnegative real numbers.

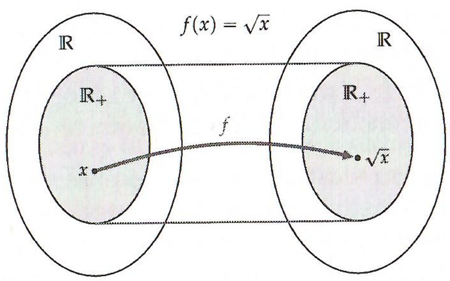


Figure 1.15: The input and output sets of the function $f(x) = \sqrt{x}$. The domain of f is the set of nonnegative real numbers $\mathbb{R}_+$ and its image is $\mathbb{R}_+$.

To illustrate the difference between the image of a function and its target set, let's look at the function $f(x) = x^2$ shown in Figure 1.16. The quadratic function is of the form $f\colon \mathbb{R} \to \mathbb{R}$. The function's source set is $\mathbb{R}$ (it takes real numbers as inputs) and its target set is $\mathbb{R}$ (the outputs are real numbers too); however, not all real numbers are possible outputs. The *image* of the function $f(x) = x^2$ consists only of the nonnegative real numbers $\mathbb{R}_+ = \{y \in \mathbb{R} \mid y \geqslant 0\}$, since $f(x) \geqslant 0$ for all x.

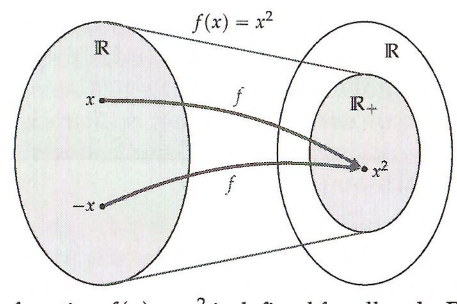


Figure 1.16: The function $f(x) = x^2$ is defined for all reals: $\text{Dom}(f) = \mathbb{R}$. The image of the function is the set of nonnegative real numbers: $\text{Im}(f) = \mathbb{R}_+$.

Function properties

We'll now introduce some additional terminology for describing three important function properties. Every function is a mapping from a source set to a target set, but what kind of mapping is it?

- A function is *injective* if it maps two different inputs to two different outputs. If x_1 and x_2 are two input values that are not equal $x_1 \neq x_2$, then the output values of an injective function will also not be equal $f(x_1) \neq f(x_2)$.
- A function is *surjective* if its image is equal to its target set. For every output y in the target set of a surjective function, there is at least one input x in its domain such that $f(x) = y$.
- A function is *bijective* if it is both injective and surjective.

I know this seems like a lot of terminology to get acquainted with, but it's important to have names for these function properties. We'll need this terminology to give a precise definition of the *inverse function* in the next section.

Injective property We can think of *injective* functions as pipes that transport fluids between containers. Since fluids cannot be compressed, the "output container" must be at least as large as the "input container." If there are two distinct points x_1 and x_2 in the input container of an injective function, then there will be two distinct points $f(x_1)$ and $f(x_2)$ in the output container of the function as well. In other words, injective functions don't smoosh things together.

In contrast, a function that doesn't have the injective property can map several different inputs to the same output value. The function $f(x) = x^2$ is not injective since it sends inputs x and $-x$ to the same output value $f(x) = f(-x) = x^2$, as illustrated in Figure 1.16.

The maps-distinct-inputs-to-distinct-outputs property of injective functions has an important consequence: given the output of an injective function y, there is only one input x such that $f(x) = y$. If a second input x' existed that also leads to the same output $f(x) = f(x') = y$, then the function f wouldn't be injective. For each of the outputs y of an injective function f, there is a *unique* input x such that $f(x) = y$. In other words, injective functions have a unique-input-for-each-output property.

Surjective property A function is *surjective* if its outputs cover the entire target set: every number in the target set is a possible output of the function for some input. For example, the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is surjective: for every number y in the target set $\mathbb{R}$, there is an input x, namely $x = \sqrt[3]{y}$, such that $f(x) = y$. The function $f(x) = x^3$ is surjective since its image is equal to its target set, $\mathrm{Im}(f) = \mathbb{R}$, as shown in Figure 1.17.

On the other hand, the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by the equation $f(x) = x^2$ is not surjective since its image is only the nonnegative numbers $\mathbb{R}_+$ and not the whole set of real numbers (see Figure 1.16). The outputs of this function do not include the negative numbers of the target set, because there is no real number x that can be used as an input to obtain a negative output value.

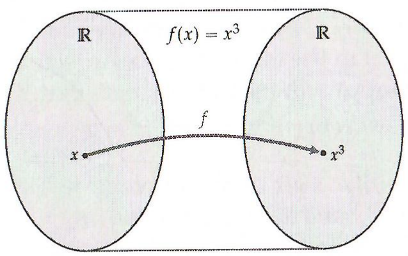


Figure 1.17: For the function $f(x) = x^3$ the image is equal to the target set of the function, $\mathrm{Im}(f) = \mathbb{R}$, therefore the function f is surjective. The function f maps two different inputs $x_1 \neq x_2$ to two different outputs $f(x_1) \neq f(x_2)$, so f is injective. Since f is both injective and surjective, it is a *bijective* function.

Bijective property A function is bijective if it is both injective and surjective. When a function $f : A \to B$ has both the injective and surjective properties, it defines a *one-to-one correspondence* between the numbers of the source set A and the numbers of the target set B. This means for every input value x, there is exactly one corresponding output value y, and for every output value y, there is exactly one input value x such that $f(x) = y$. An example of a bijective function is the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ (see Figure 1.17). For every input x in the source set $\mathbb{R}$, the corresponding output y is given by $y = f(x) = x^3$. For every output value y in the target set $\mathbb{R}$, the corresponding input value x is given by $x = \sqrt[3]{y}$.

A function is not bijective if it lacks one of the required properties. Examples of non-bijective functions are $f(x) = \sqrt{x}$, which is not surjective and $f(x) = x^2$, which is neither injective nor surjective.

Counting solutions Another way to understand the injective, surjective, and bijective properties of functions is to think about the solutions to the equation $f(x) = b$, where b is a number in the target set B. The function f is injective if the equation $f(x) = b$ has *at most one* solution for every number b. The function f is surjective if the equation $f(x) = b$ has *at least one* solution for every number b. If the function f is bijective then it is both injective and surjective, which means the equation $f(x) = b$ has *exactly one* solution.

Inverse function

We used inverse functions repeatedly in previous chapters, each time describing the inverse function informally as an "undo" operation. Now that we have learned about bijective functions, we can give a the precise definition of the inverse function and explain some of the details we glossed over previously.

Recall that a *bijective* function $f : A \to B$ is a *one-to-one correspondence* between the numbers in the source set A and numbers in the target set B: for every output y, there is exactly one corresponding input value x such that $f(x) = y$. The *inverse function*, denoted f^{-1}, is the function that takes any output value y in the set B and finds the corresponding input value x that produced it $f^{-1}(y) = x$.

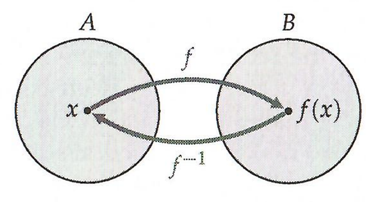


Figure 1.18: The inverse f^{-1} undoes the operation of the function f.

For every bijective function $f : A \to B$, there exists an inverse function $f^{-1} : B \to A$ that performs the *inverse mapping* of f. If we start from some x, apply f, and then apply f^{-1}, we'll arrive—full circle—back to the original input x:

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

In Figure 1.18 the function f is represented as a forward arrow, and the inverse function f^{-1} is represented as a backward arrow that puts the value $f(x)$ back to the x it came from.

Similarly, we can start from any y in the set B and apply f^{-1} followed by f to get back to the original y we started from:

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

In words, this equation tells us that f is the "undo" operation for the function f^{-1}, the same way f^{-1} is the "undo" operation for f.

If a function is missing the injective property or the surjective property then it isn't bijective and it doesn't have an inverse. Without the injective property, there could be two inputs x and x' that both produce the same output $f(x) = f(x') = y$. In this case, computing $f^{-1}(y)$ would be impossible since we don't know which of the two possible inputs x or x' was used to produce the output y. Without the surjective property, there could be some output y' in B for which the inverse function f^{-1} is not defined, so the equation $f(f^{-1}(y)) = y$ would not hold for all y in B. The inverse function f^{-1} exists only when the function f is bijective.

Wait a minute! We know the function $f(x) = x^2$ is not bijective and therefore doesn't have an inverse, but we've repeatedly used the square root function as an inverse function for $f(x) = x^2$. What's going on here? Are we using a double standard like a politician that espouses one set of rules publicly, but follows a different set of rules in their private dealings? Is mathematics corrupt?

Don't worry, mathematics is not corrupt—it's all legit. We can use inverses for non-bijective functions by imposing *restrictions* on the source and target sets. The function $f(x) = x^2$ is not bijective when defined as a function $f : \mathbb{R} \to \mathbb{R}$, but it *is* bijective if we define it as a function from the set of nonnegative numbers to the set of nonnegative numbers, $f : \mathbb{R}_+ \to \mathbb{R}_+$. Restricting the source set to $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$ makes the function injective, and restricting the target set to $\mathbb{R}_+$ also makes the function surjective. The function $f : \mathbb{R}_+ \to \mathbb{R}_+$ defined by the equation $f(x) = x^2$ is bijective and its inverse is $f^{-1}(y) = \sqrt{y}$.

It's important to keep track the of restrictions on the source set we applied when solving equations. For example, solving the equation $x^2 = c$ by restricting the solution space to nonnegative numbers will give us only the positive solution $x = \sqrt{c}$. We have to manually add the negative solution $x = -\sqrt{c}$ in order to obtain the complete solutions: $x = \sqrt{c}$ or $x = -\sqrt{c}$, which is usually written $x = \pm\sqrt{c}$. The possibility of multiple solutions is present whenever we solve equations involving non-injective functions.

Function composition

We can combine two simple functions by chaining them together to build a more complicated function. This act of applying one function after another is called *function composition*. Consider for example the composition:

$$f \circ g\,(x) = f(\,g(x)\,) = z.$$

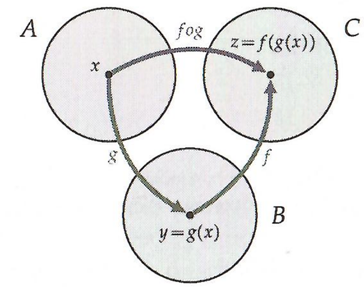


Figure 1.19: The function composition $f \circ g$ describes the combination of first applying the function g, followed by the function f: $f \circ g\,(x) = f(g(x))$.

Figure 1.19 illustrates the concept of function composition. First, the function $g : A \to B$ acts on some input x to produce an intermediary value $y = g(x)$ in the set B. The intermediary value y is then passed through the function $f : B \to C$ to produce the final output value $z = f(y) = f(g(x))$ in the set C. We can think of the *composite function* $f \circ g$ as a function in its own right. The function $f \circ g : A \to C$ is defined through the formula $f \circ g\,(x) = f(g(x))$.

Don't worry too much about the "$\circ$" symbol—it's just a convenient math notation I wanted you to know about. Writing $f \circ g$ is the same as writing $f(g(x))$. The important takeaway from Figure 1.19 is that functions can be combined by using the outputs of one function as the inputs to the next. This is a very useful idea for building math models. You can understand many complicated input-output transformations by describing them as compositions of simple functions.

Example 1 Consider the function $g\colon \mathbb{R}_+ \to \mathbb{R}_+$ given by $g(x) = \sqrt{x}$, and the function $f\colon \mathbb{R} \to \mathbb{R}_+$ defined by $f(x) = x^2$. The composite function $f \circ g\,(x) = (\sqrt{x})^2 = x$ is defined for all nonnegative reals. The composite function $g \circ f$ is defined for all real numbers, and we have $g \circ f\,(x) = \sqrt{x^2} = |x|$.

Example 2 The composite functions $f \circ g$ and $g \circ f$ describe different operations. If $g(x) = \ln(x)$ and $f(x) = x^2$, the functions

$g \circ f(x) = \ln(x^2)$ and $f \circ g(x) = (\ln x)^2$ have different domains and produce different outputs, as you can verify using a calculator.

Using the notation "$\circ$" for function composition, we can give a concise description of the properties of a bijective function $f : A \to B$ and its inverse function $f^{-1} : B \to A$:

$$(f^{-1} \circ f)(x) = x \qquad \text{and} \qquad (f \circ f^{-1})(y) = y,$$

for all x in A and all y in B.

Function names

We use short symbols like $+$, $-$, $\times$, and $\div$ to denote most of the important functions used in everyday life. We also use the squiggle notation $\sqrt{\ }$ for square roots and superscripts to denote exponents. All other functions are identified and denoted by their *name*. If I want to compute the *cosine* of the angle 60° (a function describing the ratio between the length of one side of a right-angle triangle and the hypotenuse), I write cos(60°), which means I want the value of the cos function for the input 60°.

Incidentally, the function cos has a nice output value for that specific angle: $\cos(60°) = \frac{1}{2}$. Therefore, seeing cos(60°) somewhere in an equation is the same as seeing $\frac{1}{2}$. To find other values of the function, say cos(33.13°), you'll need a calculator. All scientific calculators have a convenient little cos button for this very purpose.

Handles on functions

When you learn about functions you learn about the different "handles" by which you can "grab" these mathematical objects. The main handle for a function is its **definition**: it tells you the precise way to calculate the output when you know the input. The function definition is an important handle, but it is also important to "feel" what the function does intuitively. How does one get a feel for a function?

Table of values

One simple way to represent a function is to look at a list of input-output pairs: $\{\{\text{in} = x_1, \text{out} = f(x_1)\}, \{\text{in} = x_2, \text{out} = f(x_2)\}, \{\text{in} = x_3, \text{out} = f(x_3)\}, \ldots\}$. A more compact notation for the input-output pairs is $\{(x_1, f(x_1)), (x_2, f(x_2)), (x_3, f(x_3)), \ldots\}$, where the first number of each pair represents an input value and the second represents the output value given by the function.

We can also build a **table of values** by writing the input values in one column and recording the corresponding output values in a second column. You can choose inputs at random or focus on the important-looking x values in the function's domain.

input $= x$	$\to$	$f(x) =$ output
0	$\to$	$f(0)$
1	$\to$	$f(1)$
55	$\to$	$f(55)$
x_4	$\to$	$f(x_4)$

Table 1.1: Table of input-output values of the function $f(x)$. The input values $x = 0$, $x = 1$ and $x = 55$ are chosen to "test" what the function does.

You can create a table of values for any function you want to study. Follow the example shown in Table 1.1. Use the input values that interest you and fill out the right side of the table by calculating the value of $f(x)$ for each input x.

Function graph

One of the best ways to feel a function is to look at its graph. A graph is a line on a piece of paper that passes through all input-output pairs of a function. Imagine you have a piece of paper, and on it you draw a blank *coordinate system* as in Figure 1.20.

The horizontal axis is used to measure x. The vertical axis is used to measure $f(x)$. Because writing out $f(x)$ every time is long and tedious, we use a short, single-letter alias to denote the output value of f as follows:

$$y = f(x) = \text{output}.$$

Think of each input-output pair of the function f as a point (x, y) in the coordinate system. The graph of a function is a representational drawing of everything the function does. If you understand how to interpret this drawing, you can infer everything there is to know about the function.

Facts and properties

Another way to feel a function is by knowing the function's properties. This approach boils down to learning facts about the function

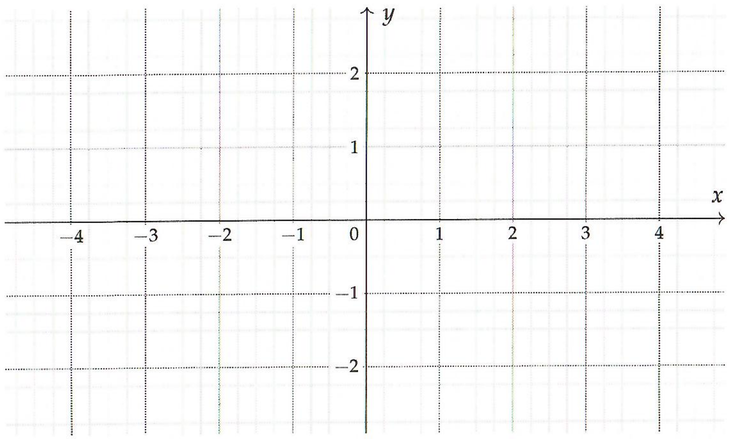


Figure 1.20: An empty (x, y)-coordinate system that you can use to draw function graphs. The graph of $f(x)$ consists of all the points for which $(x, y) = (x, f(x))$. See Figure 1.12 on page 37 for the graph of $f(x) = x^2$.

and its connections to other functions. An example of a mathematical connection is the equation $\log_B(x) = \frac{\log_b(x)}{\log_b(B)}$, which describes a link between the logarithmic function base B and the logarithmic function base b.

The more you know about a function, the more "paths" your brain builds to connect to that function. Real math knowledge is not about memorization; it is about establishing a network of associations between different areas of information in your brain. See the concept maps on page v for an illustration of the paths that link math concepts. Mathematical thought is the usage of these associations to carry out calculations and produce mathematical arguments. For example, knowing about the connection between logarithmic functions will allow you compute the value of $\log_7(e^3)$, even though calculators don't have a button for logarithms base 7. We find $\log_7(e^3) = \frac{\ln e^3}{\ln 7} = \frac{3}{\ln 7}$, which can be computed using the ln button.

To develop mathematical skills, it is vital to practice path-building between concepts by solving exercises. With this book, I will introduce you to some of the many paths linking math concepts, but it's on you to reinforce these paths through practice.

Example 3 Consider the function f from the real numbers to the real numbers ($f\colon \mathbb{R} \to \mathbb{R}$) defined as $f(x) = x^2 + 2x - 3$. The value of f when $x = 1$ is $f(1) = 1^2 + 2(1) - 3 = 0$. When $x = 2$, the output is $f(2) = 2^2 + 2(2) - 3 = 5$. What is the value of f when $x = 0$? You can

use algebra to rewrite this function as $f(x) = (x+3)(x-1)$, which tells you the graph of this function crosses the x-axis at $x = -3$ and at $x = 1$. The values above will help you plot the graph of $f(x)$.

Example 4 Consider the exponential function with base 2 defined by $f(x) = 2^x$. This function is crucial to computer systems. For instance, RAM memory chips come in powers of two because the memory space is exponential in the number of "address lines" used on the chip. When $x = 1$, $f(1) = 2^1 = 2$. When x is 2 we have $f(2) = 2^2 = 4$. The function is therefore described by the following input-output pairs: $(0,1)$, $(1,2)$, $(2,4)$, $(3,8)$, $(4,16)$, $(5,32)$, $(6,64)$, $(7,128)$, $(8,256)$, $(9,512)$, $(10,1024)$, $(11,2048)$, $(12,4096)$, etc. Recall that any number raised to exponent 0 gives 1. Thus, the exponential function passes through the point $(0,1)$. Recall also that negative exponents lead to fractions, so we have the points $(-1, \frac{1}{2})$, $(-2, \frac{1}{4})$, $(-3, \frac{1}{8})$, etc. You can plot these $(x, f(x))$ coordinates in the Cartesian plane to obtain the graph of the function.

Discussion

To describe a function we specify its source and target sets $f\colon A \to B$, then give an equation of the form $f(x) =$ "expression involving x" that defines the function. Since functions are defined using equations, does this mean that functions and equations are the same thing? Let's take a closer look.

In general, any equation containing two variables describes a *relation* between these variables. For example, the equation $x - 3 = y - 4$ describes a relation between the variables x and y. We can isolate the variable y in this equation to obtain $y = x + 1$ and thus find the value of y when the value of x is given. We can also isolate x to obtain $x = y - 1$ and use this equation to find x when the value of y is given. In the context of an equation, the relationship between the variables x and y is symmetrical and no special significance is attached to either of the two variables.

We also can describe the same relationship between x and y as a function $f : \mathbb{R} \to \mathbb{R}$. We choose to identify x as the input variable and y as the output variable of the function f. Having identified y with the output variable, we can interpret the equation $y = x + 1$ as the definition of the function $f(x) = x + 1$.

Note that the equation $x - 3 = y - 4$ and the function $f(x) = x + 1$ describe the same relationship between the variables x and y. For example, if we set the value $x = 5$ we can find the value of y by solving the equation $5 - 3 = y - 4$ to obtain $y = 6$, or by computing the output of the function $f(x)$ for the input $x = 5$, which gives us the same answer $f(5) = 6$. In both cases we arrive at the same answer, but modelling the relationship between x and y as a function allows us to use the whole functions toolbox, like function composition and function inverses.

* * *

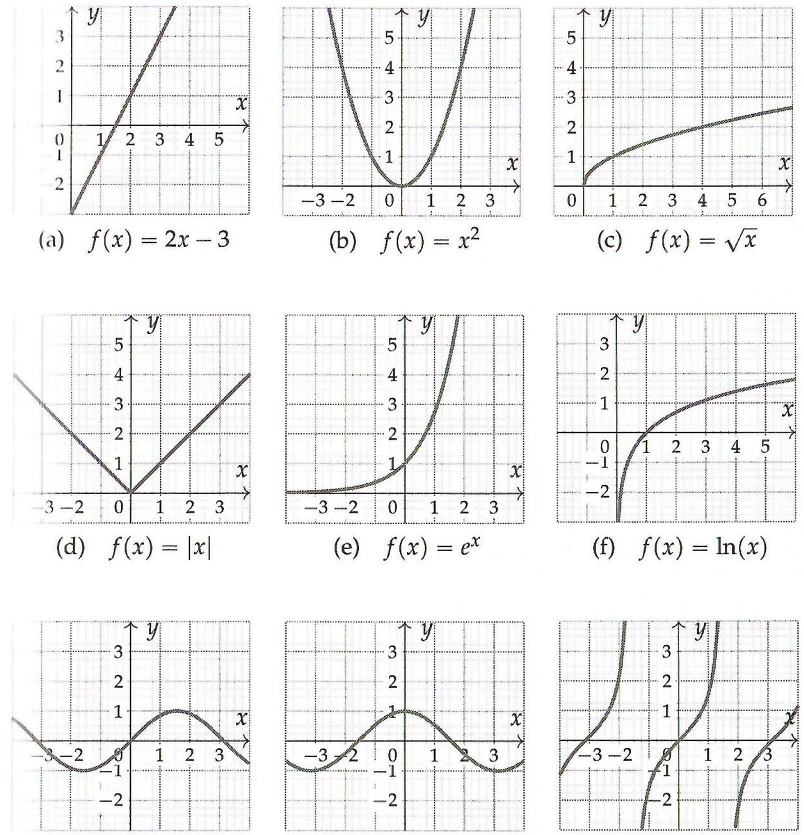
In this section we talked a lot about functions in general but we haven't said much about any function specifically. There are many useful functions out there, and we can't discuss them all here. In the next section, we'll introduce 10 functions of strategic importance for all of science. If you get a grip on these functions, you'll be able to understand all of physics and calculus and handle *any* problem your teacher may throw at you.

1.9 Functions reference

Your *function vocabulary* determines how well you can express yourself mathematically in the same way your English vocabulary determines how well you can express yourself in English. The following pages aim to embiggen your function vocabulary, so you'll know how to handle the situation when a teacher tries to pull some trick on you at the final.

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

To build mathematical intuition, it's essential you understand functions' graphs. Memorizing the definitions and properties of functions gets a lot easier with visual accompaniment. Indeed, remembering what the function "looks like" is a great way to train yourself to recognize various types of functions. Figure 1.21 shows the graphs of some of the most important functions we'll use in this book.



(a) $f(x) = 2x - 3$ (b) $f(x) = x^2$ (c) $f(x) = \sqrt{x}$

(d) $f(x) = |x|$ (e) $f(x) = e^x$ (f) $f(x) = \ln(x)$

(g) $f(x) = \sin(x)$ (h) $f(x) = \cos(x)$ (i) $f(x) = \tan(x)$

Figure 1.21: We'll see many types of function graphs in the next pages.

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 is the value of the function when $x = 0$.

Consider what relationship the equation of $f(x)$ describes for different values of m and b. What happens when m is positive? What happens when m is negative?

Graph

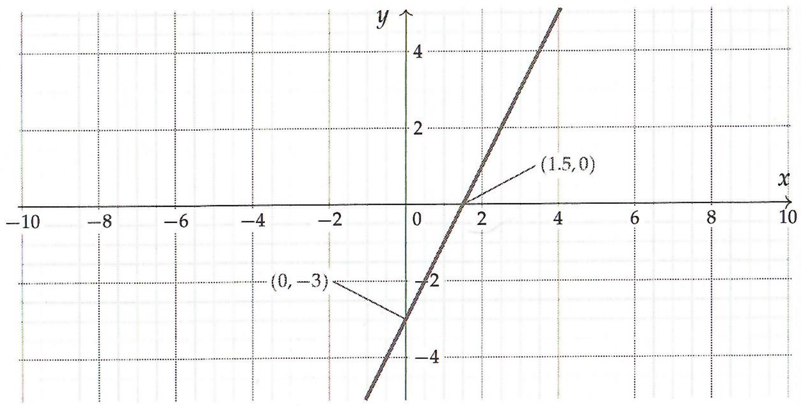


Figure 1.22: The graph of the function $f(x) = 2x - 3$. The slope is $m = 2$. The y-intercept of this line is $b = -3$. The x-intercept is at $x = \frac{3}{2}$.

Properties

- Domain: $\mathbb{R}$. The function $f(x) = mx + b$ is defined for all reals.
- Image: $\mathbb{R}$ if $m \neq 0$. If $m = 0$ the function is constant $f(x) = b$, so the image set contains only a single number $\{b\}$.
- $x = -b/m$: the x-intercept of $f(x) = mx + b$. The x-intercept is obtained by solving $f(x) = 0$.
- The inverse to the line $f(x) = mx + b$ is $f^{-1}(x) = \frac{1}{m}(x - b)$, which is also a line.

General equation

A line can also be described in a more symmetric form as a relation:

$$Ax + By = C.$$

This is known as the *general* equation of a line. The general equation for the line shown in Figure 1.22 is $2x - 1y = 3$.

Given the general equation of a line $Ax + By = C$ with $B \neq 0$, you can convert to the function form $y = f(x) = mx + b$ by computing the slope $m = \frac{-A}{B}$ and the y-intercept $b = \frac{C}{B}$.

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.

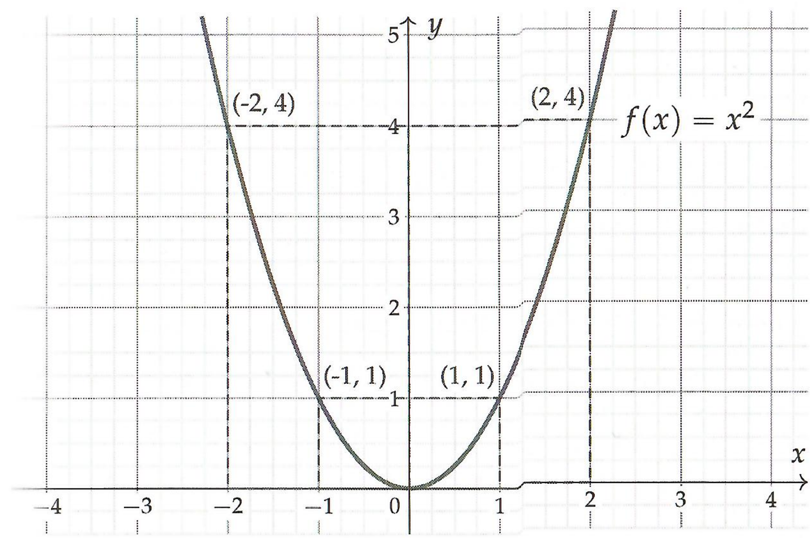


Figure 1.23: Plot of the quadratic function $f(x) = x^2$. The graph of the function passes through the following (x, y) coordinates: $(-2, 4)$, $(-1, 1)$, $(0, 0)$, $(1, 1)$, $(2, 4)$, $(3, 9)$, etc.

Properties

- Domain: $\mathbb{R}$. The function $f(x) = x^2$ is defined for all numbers.
- Image: $\mathbb{R}_+ = \{y \in \mathbb{R} \mid y \geqslant 0\}$. The outputs are nonnegative numbers since $x^2 \geqslant 0$, for all real numbers x.
- 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.

The set expression $\{y \in \mathbb{R} \mid y \geqslant 0\}$ that we use to define the non-negative real numbers ($\mathbb{R}_+$) is read "the set of real numbers that are greater than or equal to zero."

Square root

The square root function is denoted

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

The square root $\sqrt{x}$ is the inverse function of the square function x^2 when the two functions are defined as $f : \mathbb{R}_+ \to \mathbb{R}_+$. 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

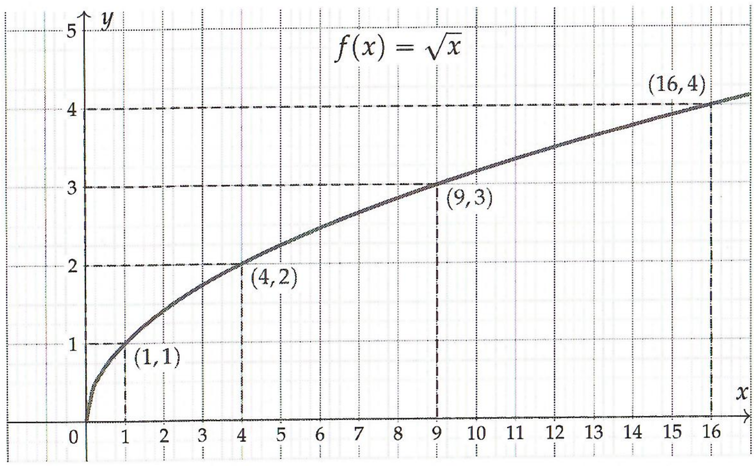


Figure 1.24: The graph of the function $f(x) = \sqrt{x}$. The domain of the function is $\mathbb{R}_+$ because we can't take the square root of a negative number.

Properties

- Domain: $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$. The function $f(x) = \sqrt{x}$ is only defined for nonnegative inputs. 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: $\mathbb{R}_+ = \{y \in \mathbb{R} \mid y \geqslant 0\}$. The outputs of the function $f(x) = \sqrt{x}$ are nonnegative numbers since $\sqrt{x} \geqslant 0$.

In addition to *square* root, there is also *cube* root $f(x) = \sqrt[3]{x} = x^{\frac{1}{3}}$, 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* function tells us the 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 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:

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

Graph

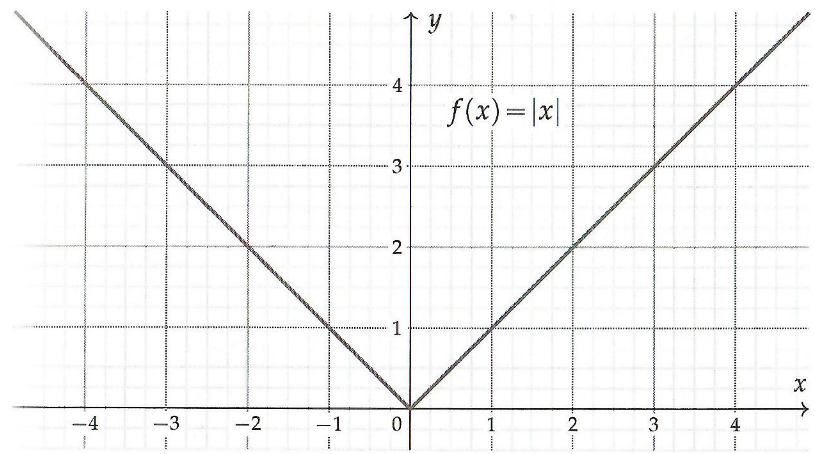


Figure 1.25: The graph of the absolute value function $f(x) = |x|$.

Properties

- Domain: $\mathbb{R}$. The function $f(x) = |x|$ is defined for all inputs.
- Image: $\mathbb{R}_+ = \{y \in \mathbb{R} \mid y \geqslant 0\}$
- The combination of squaring followed by square-root is equivalent to the absolute value function:

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

since squaring destroys the sign.

Polynomials

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

The general equation for a polynomial function of degree n is

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_nx^n.$$

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

Parameters

- x: the variable
- a_0: the constant term
- a_1: the *linear* coefficient, or *first-order* coefficient
- a_2: the *quadratic* coefficient
- a_3: the *cubic* coefficient
- a_n: the n^{th} order coefficient
- n: the *degree* of the polynomial. The degree of $f(x)$ is the largest power of x that appears in the polynomial.

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

Properties

- Domain: $\mathbb{R}$. Polynomials are defined for all inputs.
- The roots of $f(x)$ are the values of x for which $f(x) = 0$.
- The image of a polynomial function depends on the coefficients.
- The sum of two polynomials is also a polynomial.

The most general first-degree polynomial is a line $f(x) = mx + b$, where m and b are arbitrary constants. The most general second-degree polynomial is $f(x) = a_2x^2 + a_1x + a_0$, where again a_0, a_1, and a_2 are arbitrary constants. We call a_k the *coefficient* of x^k, since this is the number that appears in front of x^k. Following the pattern, a third-degree polynomial will look like $f(x) = a_3x^3 + a_2x^2 + a_1x + a_0$.

In general, a polynomial of degree n has the equation

$$f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0.$$

You can add two polynomials by adding together their coefficients:

$$\begin{aligned} f(x) + g(x) &= (a_nx^n + \cdots + a_1x + a_0) \ + \ (b_nx^n + \cdots + b_1x + b_0) \\ &= (a_n + b_n)x^n + \cdots + (a_1 + b_1)x + (a_0 + b_0). \end{aligned}$$

The subtraction of two polynomials works similarly. We can also multiply polynomials together using the general algebra rules for expanding brackets.

Solving polynomial equations

Very often in math, you will have to *solve* polynomial equations of the form

$$A(x) = B(x),$$

where $A(x)$ and $B(x)$ are both polynomials. Recall from earlier that to *solve*, we must find the values of x that make the equality true.

Say the revenue of your company is a function of the number of products sold x, and can be expressed as $R(x) = 2x^2 + 2x$. Say also the cost you incur to produce x objects is $C(x) = x^2 + 5x + 10$. You want to determine the amount of product you need to produce to break even, that is, so that revenue equals cost: $R(x) = C(x)$. To find the break-even value x, solve the equation

$$2x^2 + 2x = x^2 + 5x + 10.$$

This may seem complicated since there are xs all over the place. No worries! We can turn the equation into its "standard form," and then use the quadratic formula. First, move all the terms to one side until only zero remains on the other side:

$$\begin{aligned} 2x^2 + 2x \quad - x^2 &= \cancel{x^2} + 5x + 10 \quad -\cancel{x^2} \\ x^2 + 2x \quad - 5x &= \cancel{5x} + 10 \quad -\cancel{5x} \\ x^2 - 3x \quad - 10 &= \cancel{10} \quad -\cancel{10} \\ x^2 - 3x - 10 &= 0. \end{aligned}$$

Remember, if we perform the same operations on both sides of the equation, the resulting equation has the same solutions. Therefore, the values of x that satisfy $x^2 - 3x - 10 = 0$, namely $x = -2$ and $x = 5$, also satisfy $2x^2 + 2x = x^2 + 5x + 10$, which is the original problem we're trying to solve.

This "shuffling of terms" approach will work for any polynomial equation $A(x) = B(x)$. We can always rewrite it as $C(x) = 0$, where $C(x)$ is a new polynomial with coefficients equal to the difference of the coefficients of A and B. Don't worry about which side you move all the coefficients to because $C(x) = 0$ and $0 = -C(x)$ have exactly the same solutions. Furthermore, the degree of the polynomial C can be no greater than that of A or B.

The form $C(x) = 0$ is the *standard form* of a polynomial, and we'll explore several formulas you can use to find its solution(s).

Formulas

The formula for solving the polynomial equation $P(x) = 0$ depends on the *degree* of the polynomial in question.

For a first-degree polynomial equation, $P_1(x) = mx + b = 0$, the solution is $x = \frac{-b}{m}$: just move b to the other side and divide by m.

For a second-degree polynomial,

$$P_2(x) = ax^2 + bx + c = 0,$$

the solutions are $x_1 = \frac{-b+\sqrt{b^2-4ac}}{2a}$ and $x_2 = \frac{-b-\sqrt{b^2-4ac}}{2a}$.

If $b^2 - 4ac < 0$, the solutions will involve taking the square root of a negative number. In those cases, we say no real solutions exist.

There is also a formula for polynomials of degree 3 and 4, but they are complicated. For polynomials with order $\geqslant 5$, there does not exist a general analytical solution.

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 SymPy solve the standard-form equation $C(x) = 0$, call the function `solve(expr,var)`, where the expression `expr` corresponds to $C(x)$, and `var` is the variable you want to solve for. For example, to solve $x^2 - 3x + 2 = 0$, type in the following:

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

The function `solve` will find the solutions to any equation of the form `expr = 0`. In this case, we see the solutions are $x = 1$ and $x = 2$.

Another way to solve the equation is to factor the polynomial $C(x)$ using the function `factor` like this:

```
>>> factor(x**2 - 3*x + 2)              # usage: factor(expr)
(x - 1)*(x - 2)
```

We see that $x^2 - 3x + 2 = (x-1)(x-2)$, which confirms the two roots are indeed $x = 1$ and $x = 2$.

Substitution trick

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

$$x^4 - 7x^2 + 10 = 0.$$

Imagine this problem is on your exam, where you are not allowed to use 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

$$y^2 - 7y + 10 = 0,$$

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 $x = \pm\sqrt{\alpha}$ 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:

```
>>> solve(y**2 - 7*y + 10, y)
[2, 5]
>>> solve(x**4 - 7*x**2 + 10, x)
[sqrt(2), -sqrt(2), sqrt(5), -sqrt(5)]
```

Note how the second-degree polynomial has two roots, while the fourth-degree polynomial has four roots.

Even and odd functions

The polynomials form an entire family of functions. Depending on the choice of degree n and coefficients $a_0, a_1, \ldots, a_n$, a polynomial function can take on many different shapes. 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 1.26.

In the remainder of this book, we'll meet the function $\sin(x)$ many times. We'll define the function $\sin(x)$ more formally as a trigonometric ratio in Section 1.11. In Section 1.13 we'll use $\sin(x)$ and $\cos(x)$ (another trigonometric ratio) to work out the *components* of vectors.

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 on 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

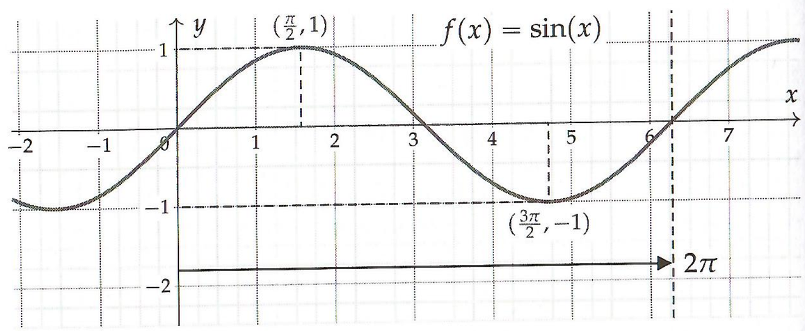


Figure 1.26: The graph of the function $y = \sin(x)$ passes through the following (x,y) coordinates: $(0,0)$, $(\frac{\pi}{6},\frac{1}{2})$, $(\frac{\pi}{4},\frac{\sqrt{2}}{2})$, $(\frac{\pi}{3},\frac{\sqrt{3}}{2})$, $(\frac{\pi}{2},1)$, $(\frac{2\pi}{3},\frac{\sqrt{3}}{2})$, $(\frac{3\pi}{4},\frac{\sqrt{2}}{2})$, $(\frac{5\pi}{6},\frac{1}{2})$, and $(\pi,0)$. For x between π and 2π, the function's graph has the same shape it has for x between 0 and π, but with negative values.

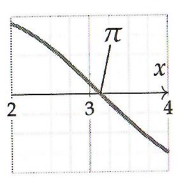


Figure 1.27: The function $f(x) = \sin(x)$ crosses the x-axis at $x = \pi$.

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 minimum value of -1 at $x = \frac{3\pi}{2}$. It then travels up again to cross the x-axis at $x = 2\pi$. This 2π-long cycle repeats after $x = 2\pi$. This is why we call the function *periodic*—the shape of the graph repeats.

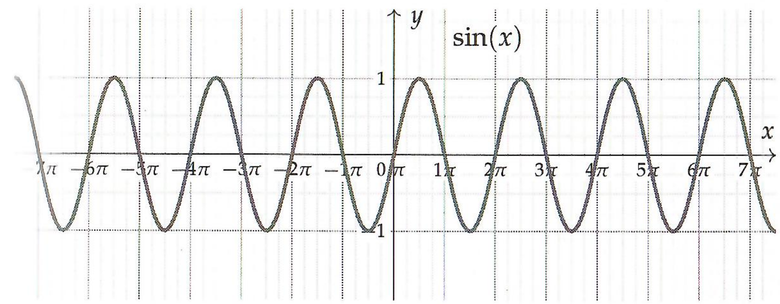


Figure 1.28: The graph of $\sin(x)$ from $x = 0$ to $x = 2\pi$ repeats periodically everywhere else on the number line.

Properties

- Domain: $\mathbb{R}$. The function $f(x) = \sin(x)$ is defined for all input values.
- Image: $\{y \in \mathbb{R} \mid -1 \leqslant y \leqslant 1\}$. The outputs of the sine function are always between -1 and 1.
- Roots: $\{\ldots, -3\pi, -2\pi, -\pi, 0, \pi, 2\pi, 3\pi, \ldots\}$. The function $\sin(x)$ has roots at all multiples of π.
- The function is periodic, with period 2π: $\sin(x) = \sin(x + 2\pi)$.
- The sin function is *odd*: $\sin(x) = -\sin(-x)$
- Relation to cos: $\sin^2 x + \cos^2 x = 1$
- Relation to csc: $\csc(x) = \frac{1}{\sin x}$ (csc is read *cosecant*)
- The inverse function of $\sin(x)$ is denoted as $\sin^{-1}(x)$ or $\arcsin(x)$, not to be confused with $(\sin(x))^{-1} = \frac{1}{\sin(x)} = \csc(x)$.
- The number $\sin(\theta)$ is the length-ratio of the vertical side and the hypotenuse in a right-angle triangle with angle θ at the base.

Links

[See the Wikipedia page for nice illustrations]
http://en.wikipedia.org/wiki/Sine

Cosine

The cosine function is the same as the sine function *shifted* by $\frac{\pi}{2}$ to the left: $\cos(x) = \sin(x + \frac{\pi}{2})$. Thus everything you know about the sine function also applies to the cosine function.

Graph

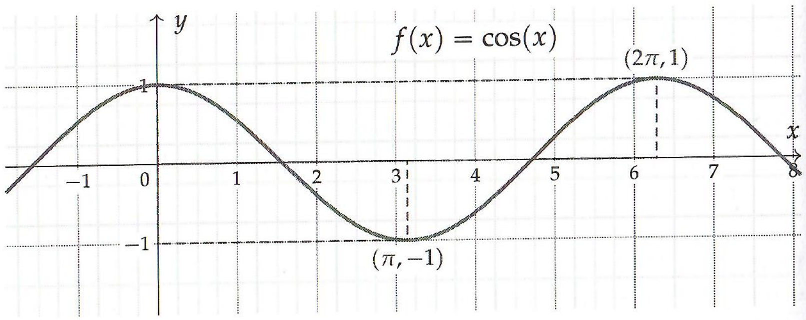


Figure 1.29: The graph of the function $y = \cos(x)$ passes through the following (x, y) coordinates: $(0,1)$, $(\frac{\pi}{6}, \frac{\sqrt{3}}{2})$, $(\frac{\pi}{4}, \frac{\sqrt{2}}{2})$, $(\frac{\pi}{3}, \frac{1}{2})$, $(\frac{\pi}{2}, 0)$, $(\frac{2\pi}{3}, -\frac{1}{2})$, $(\frac{3\pi}{4}, -\frac{\sqrt{2}}{2})$, $(\frac{5\pi}{6}, -\frac{\sqrt{3}}{2})$, and $(\pi, -1)$.

The cos function starts at $\cos(0) = 1$, then drops down to cross the x-axis at $x = \frac{\pi}{2}$. Cos continues until it reaches its minimum value at $x = \pi$. The function then moves upward, crossing the x-axis again at $x = \frac{3\pi}{2}$, and reaching its maximum value again at $x = 2\pi$.

Properties

- Domain: $\mathbb{R}$
- Image: $\{y \in \mathbb{R} \mid -1 \leqslant y \leqslant 1\}$
- Roots: $\{\ldots, -\frac{3\pi}{2}, -\frac{\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots\}$
- Relation to sin: $\sin^2 x + \cos^2 x = 1$
- Relation to sec: $\sec(x) = \frac{1}{\cos x}$ (sec is read *secant*)
- The inverse function of $\cos(x)$ is denoted $\cos^{-1}(x)$ or $\arccos(x)$.
- The cos function is *even*: $\cos(x) = \cos(-x)$
- The number $\cos(\theta)$ is the length-ratio of the horizontal side and the hypotenuse in a right-angle triangle with angle θ at the base

Tangent

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

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

Graph

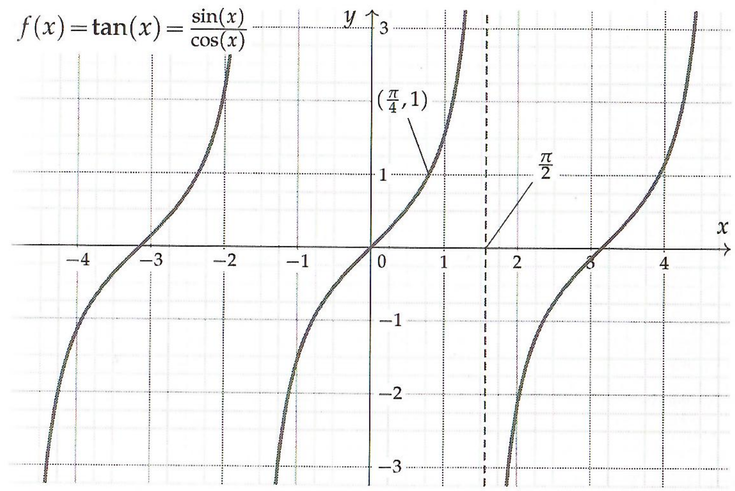


Figure 1.30: The graph of the function $f(x) = \tan(x)$.

Properties

- Domain: $\{x \in \mathbb{R} \mid x \neq \frac{(2n+1)\pi}{2}$ for any $n \in \mathbb{Z}\}$
- Image: $\mathbb{R}$
- 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 denoted $\tan^{-1}(x)$ or $\arctan(x)$.
- 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

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

Graph

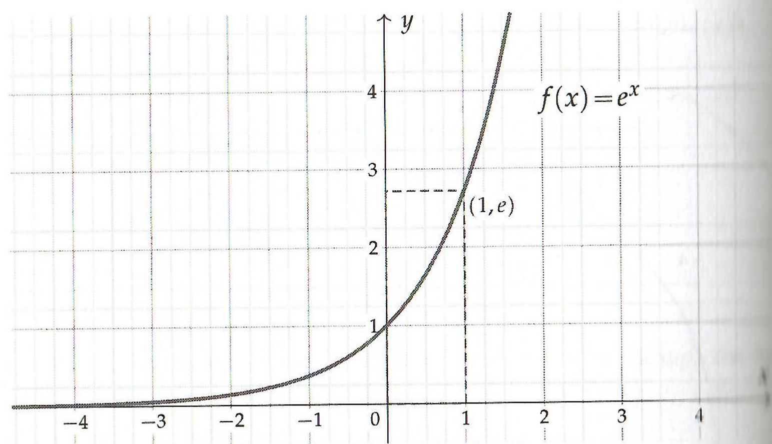


Figure 1.31: The graph of the exponential function $f(x) = e^x$ passes through the following points: $(-2, \frac{1}{e^2})$, $(-1, \frac{1}{e})$, $(0,1)$, $(1,e)$, $(2,e^2)$, $(3,e^3)$, $(4,e^4)$, etc.

Properties

- Domain: $\mathbb{R}$
- Image: $\{y \in \mathbb{R} \mid y > 0\}$
- $f(a)f(b) = f(a+b)$ since $e^a e^b = e^{a+b}$

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 1.31. 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`

Natural logarithm

The natural logarithm function is denoted

$$f(x) = \ln(x) = \log_e(x).$$

The function $\ln(x)$ is the inverse function of the exponential e^x.

Graph

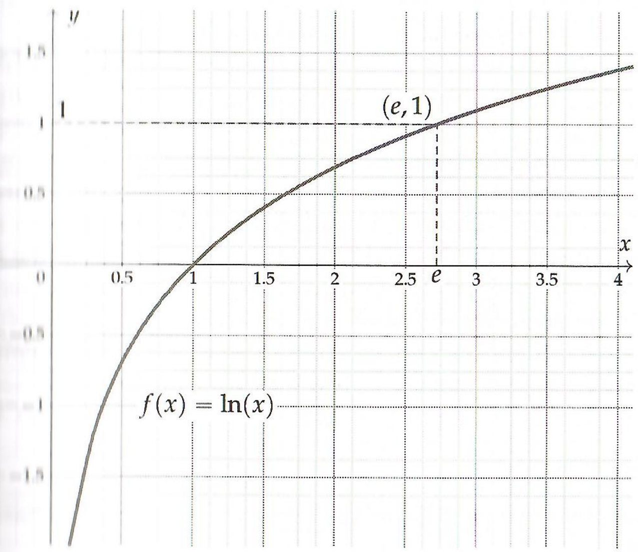


Figure 1.32: The graph of the function $\ln(x)$ passes through the following coordinates: $(\frac{1}{e^2}, -2)$, $(\frac{1}{e}, -1)$, $(1,0)$, $(e,1)$, $(e^2,2)$, $(e^3,3)$, $(e^4,4)$, etc.

Properties

- Domain: $\{x \in \mathbb{R} \mid x > 0\}$
- Image: $\mathbb{R}$

Exercises

E1.10 Find the domain, the image, and the roots of $f(x) = 2\cos(x)$.

E1.11 What are the degrees of the following polynomials? Are they even, odd, or neither?

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

E1.12 Solve for x in the following polynomial equations.

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

Exponential

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

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

Graph

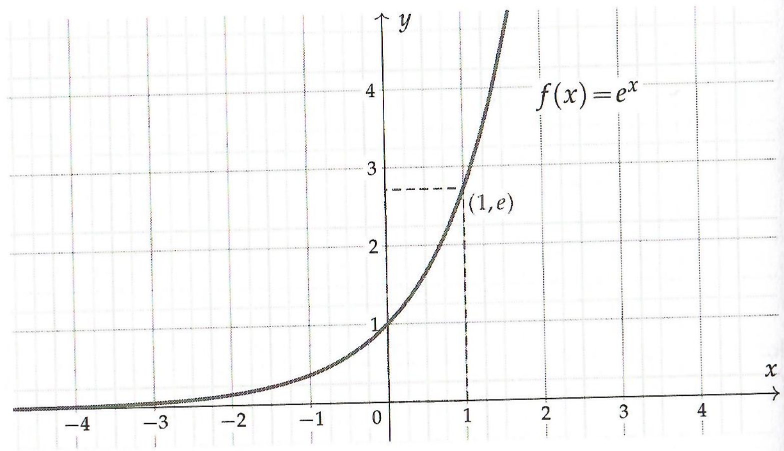


Figure 1.31: The graph of the exponential function $f(x) = e^x$ passes through the following points: $(-2, \frac{1}{e^2})$, $(-1, \frac{1}{e})$, $(0,1)$, $(1,e)$, $(2,e^2)$, $(3,e^3)$, $(4,e^4)$, etc.

Properties

- Domain: $\mathbb{R}$
- Image: $\{y \in \mathbb{R} \mid y > 0\}$
- $f(a)f(b) = f(a+b)$ since $e^a e^b = e^{a+b}$

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 1.31. 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`

Natural logarithm

The natural logarithm function is denoted

$$f(x) = \ln(x) = \log_e(x).$$

The function $\ln(x)$ is the inverse function of the exponential e^x.

Graph

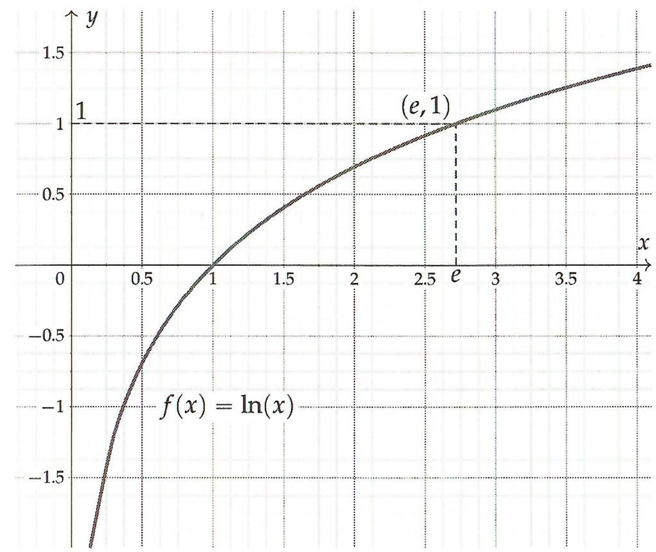


Figure 1.32: The graph of the function $\ln(x)$ passes through the following coordinates: $(\frac{1}{e^2}, -2)$, $(\frac{1}{e}, -1)$, $(1,0)$, $(e,1)$, $(e^2,2)$, $(e^3,3)$, $(e^4,4)$, etc.

Properties

- Domain: $\{x \in \mathbb{R} \mid x > 0\}$
- Image: $\mathbb{R}$

Exercises

E1.10 Find the domain, the image, and the roots of $f(x) = 2\cos(x)$.

E1.11 What are the degrees of the following polynomials? Are they even, odd, or neither?

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

E1.12 Solve for x in the following polynomial equations.

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

1.10 Geometry

The word "geometry" comes from the Greek roots *geo*, which means "earth," and *metron*, which means "measurement." This name is linked to one of the early applications of geometry, which was to measure the total amount of land contained within a certain boundary region. Over the years, the study of geometry evolved to be more abstract. Instead of developing formulas for calculating the area of specific regions of land, mathematicians developed general area formulas that apply to *all* regions that have a particular shape.

In this section we'll present formulas for calculating the perimeters, areas, and volumes for various shapes (also called "figures") commonly encountered in the real world. For two-dimensional figures, the main quantities of interest are the figures' areas and the figures' perimeters (the length of the walk around the figure). For three-dimensional figures, the quantities of interest are the surface area (how much paint it would take to cover all sides of the figure), and volume (how much water it would take to fill a container of this shape). The formulas presented are by no means an exhaustive list of everything there is to know about geometry, but they represent a core set of facts that you want to add to your toolbox.

Triangles

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

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

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

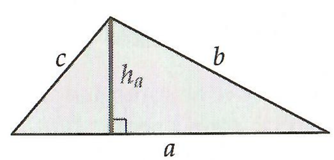


Figure 1.33: A triangle with side lengths a, b, and c. The height of the triangle with respect to the side a is denoted h_a.

The perimeter of a triangle is given by the sum of its side lengths:

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

Interior angles of a triangle rule The sum of the inner angles in any triangle is equal to 180°. Consider a triangle with internal angles α, β and γ as shown in Figure 1.34. We may not know the values of

the individual angles α, β, and γ, but we know their sum is $\alpha + \beta + \gamma = 180°$.

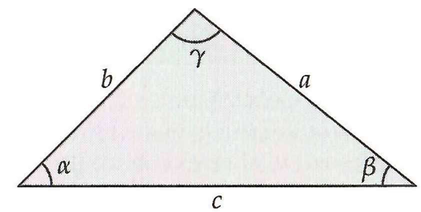


Figure 1.34: A triangle with inner angles α, β, and γ and sides a, b, and c.

Sine rule The sine rule states the following equation is true:

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

where α is the angle opposite to side a, β is the angle opposite to side b, and γ is the angle opposite to side c, as shown in Figure 1.34.

Cosine rule The cosine rules states the following equations are true:

$$\begin{aligned} a^2 &= b^2 + c^2 - 2bc\cos(\alpha), \\ b^2 &= a^2 + c^2 - 2ac\cos(\beta), \\ c^2 &= a^2 + b^2 - 2ab\cos(\gamma). \end{aligned}$$

These equations are useful when you know two sides of a triangle and the angle between them, and you want to find the third side.

Circle

The circle is a beautiful shape. If we take the centre of the circle at the origin $(0, 0)$, the circle of radius r corresponds to the equation

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

This formula describes the set of points (x, y) with a distance from the centre equal to r.

Area

The area enclosed by a circle of radius r is given by $A = \pi r^2$. A circle of radius $r = 1$ has area π.

Circumference and arc length

The circumference of a circle of radius r is

$$C = 2\pi r.$$

A circle of radius $r = 1$ has circumference 2π. This is the total length you can measure by following the curve all the way around to trace the outline of the entire circle. For example, the circumference of a circle of radius 3 m is $C = 2\pi(3) = 18.85$ m. This is how far you'll need to walk to complete a full turn around a circle of radius $r = 3$ m.

What is the length of a part of the circle? Say you have a piece of the circle, called an *arc*, and that piece corresponds to the angle $\theta = 57°$ as shown in Figure 1.35. What is the arc's length ℓ? If the circle's total length $C = 2\pi r$ represents a full 360° turn around the circle, then the arc length ℓ for a portion of the circle corresponding to the angle θ is

$$\ell = 2\pi r \frac{\theta}{360}.$$

The arc length ℓ depends on r, the angle θ, and a factor of $\frac{2\pi}{360}$.

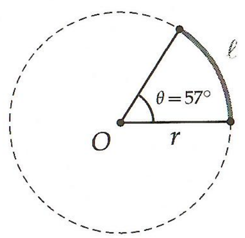


Figure 1.35: The arc length ℓ equals $\frac{57}{360}$ of the circle's circumference $2\pi r$.

Radians

While scientists and engineers commonly use degrees as a measurement unit for angles, mathematicians prefer to measure angles in *radians*, denoted rad.

Measuring an angle in radians is equivalent to measuring the arc length ℓ on a circle with radius $r = 1$, as illustrated in Figure 1.36.

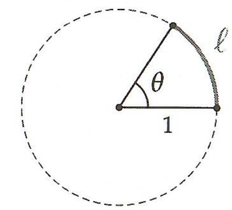


Figure 1.36: The angle θ measured in radians corresponds to the arc length ℓ on a circle with radius 1. The full circle corresponds to the angle 2π rad.

The conversion ratio between degrees and radians is

$$2\pi \text{ rad} = 360°.$$

When the angle θ is measured in radians, the arc length is given by:

$$\ell = r\theta.$$

To find the arc length ℓ, we simply multiply the circle radius r times the angle θ measured in radians.

Note the arc-length formula with θ measured in radians is simpler than the arc-length formula with θ measured in degrees, since we don't need the conversion factor of 360°.

Sphere

A sphere of radius r is described by the equation $x^2 + y^2 + z^2 = r^2$. The surface area of the sphere is $A = 4\pi r^2$, and its volume is given by $V = \frac{4}{3}\pi r^3$.

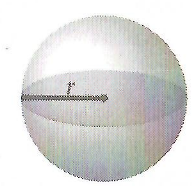


Figure 1.37: A sphere of radius r has surface area $4\pi r^2$ and volume $\frac{4}{3}\pi r^3$.

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

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

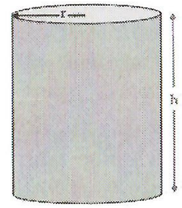


Figure 1.38: A cylinder with radius r and height h has volume $\pi r^2 h$.

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 radius of each piston is 43.75 mm, and the height of each piston is 83.1 mm. Verify the total engine volume is 1998789 $\text{mm}^3 \approx 2$ L.

Cones and pyramids

The volume of a square pyramid with side length a and height h is given by the formula $V = \frac{1}{3}a^2h$. The volume of a cone of radius r and height h is given by the formula $V = \frac{1}{3}\pi r^2 h$. Note the factor $\frac{1}{3}$ appears in both formulas. These two formulas are particular cases of the general volume formula that applies to all pyramids:

$$V = \tfrac{1}{3}Ah,$$

where A is the area of the pyramid's base and h is its height. This formula applies for pyramids with a base that is a triangle (triangular pyramids), a square (square pyramids), a rectangle (rectangular pyramids), a circle (cones), or any other shape.

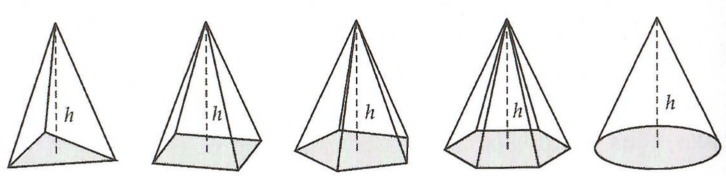
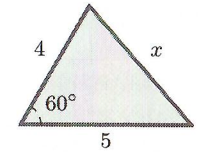


Figure 1.39: The volumes of pyramids and cones are described by the formula $V = \frac{1}{3}Ah$, where A is the area of the base and h is the height.

Exercises

E1.13 Find the length of side x in the triangle below.



Hint: Use the cosine rule.

E1.14 Find the volume and the surface area of a sphere with radius 2.

E1.15 On a rainy day, Laura brings her bike indoors, and the wet bicycle tires leave a track of water on the floor. What is the length of the water track left by the bike's rear tire (diameter 73 cm) if the wheel makes five full turns along the floor?

1.11 Trigonometry

If one of the angles in a triangle 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 concepts like vectors and complex numbers.

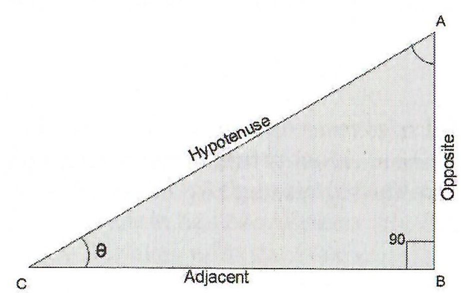


Figure 1.40: A right-angle triangle. The angle at the base is denoted θ and the names of the sides of the triangle are indicated.

Concepts

- A, B, C: the three *vertices* of the triangle
- θ: the angle at the vertex C. Angles can be measured in degrees or radians.
- opp $= AB$: the length of the *opposite* side to θ
- adj $= BC$: the length of side *adjacent* to θ
- hyp $= AC$: the *hypotenuse*. This is the triangle's longest side.
- h: the "height" of the triangle (in this case $h =$ opp $= AB$)
- $\sin\theta = \frac{\text{opp}}{\text{hyp}}$: the *sine* of theta is the ratio of the length of the opposite side and the length of the hypotenuse
- $\cos\theta = \frac{\text{adj}}{\text{hyp}}$: the *cosine* of theta is the ratio of the adjacent length and the hypotenuse length
- $\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{\text{opp}}{\text{adj}}$: the *tangent* is the ratio of the opposite length divided by the adjacent length

Pythagoras' theorem

In a right-angle triangle, the length of the hypotenuse squared is equal to the sum of the squares of the lengths of the other sides:

$$\text{adj}^2 + \text{opp}^2 = \text{hyp}^2.$$

If we divide both sides of the above equation by hyp^2, we obtain

$$\frac{\text{adj}^2}{\text{hyp}^2} + \frac{\text{opp}^2}{\text{hyp}^2} = 1.$$

Since $\frac{\text{adj}}{\text{hyp}} = \cos\theta$ and $\frac{\text{opp}}{\text{hyp}} = \sin\theta$, this equation can be rewritten as

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

This is a powerful *trigonometric identity* that describes an important relation between sine and cosine functions. In case you've never seen this notation before, the expression $\cos^2\theta$ is used to denote $(\cos(\theta))^2$.

Sin and cos

Meet the trigonometric functions, or trigs for short. These are your new friends. Don't be shy now, say hello to them.

"Hello."

"Hi."

"Soooooo, you are like functions right?"

"Yep," sin and cos reply in chorus.

"Okay, so what do you do?"

"Who me?" asks cos. "Well I tell the ratio... hmm... Wait, are you asking what I do as a *function* or specifically what *I* do?"

"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 at Figure 1.41 and you'll see a circle. This is the unit circle because it has a radius of 1. You see it, yes?"

"Yes."

"Now imagine a point P that moves along the circle of radius 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 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* is a circle of radius one centred at the origin. The unit circle consists of all points (x, y) that satisfy the equation $x^2 + y^2 = 1$. A point P 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.

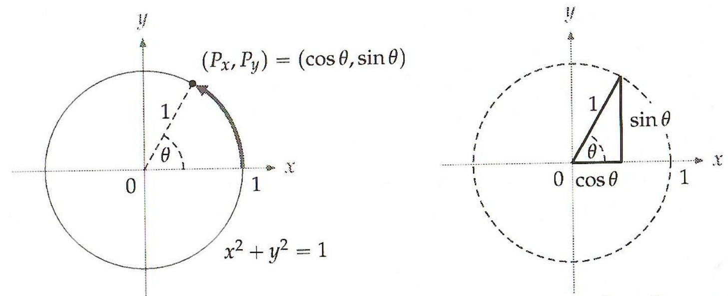


Figure 1.41: The unit circle corresponds to the equation $x^2 + y^2 = 1$. The coordinates of the point P on the unit circle are $P_x = \cos\theta$ and $P_y = \sin\theta$.

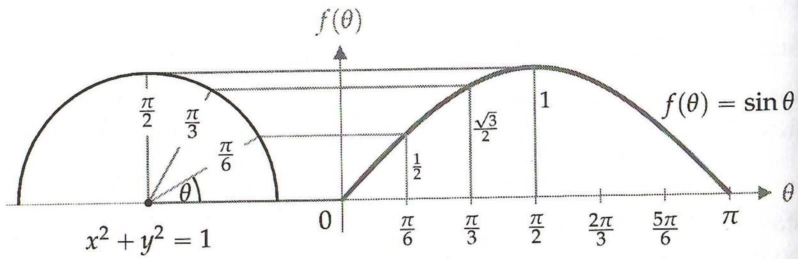


Figure 1.42: The function $f(\theta) = \sin\theta$ describes the vertical position of a point P that travels along the unit circle. The graph shows the values of the function $f(\theta) = \sin\theta$ for angles between $\theta = 0$ and $\theta = \pi$.

Figure 1.42 shows the graph of the function $f(\theta) = \sin\theta$. The values $\sin\theta$ for the angles 0, $\frac{\pi}{6}$ (30°), $\frac{\pi}{3}$ (60°), and $\frac{\pi}{2}$ (90°) are marked. There are three values to remember: $\sin\theta = 0$ when $\theta = 0$, $\sin\theta = \frac{1}{2}$ when $\theta = \frac{\pi}{6}$ (30°), and $\sin\theta = 1$ when $\theta = \frac{\pi}{2}$ (90°). See Figure 1.26 (page 60) for a graph of $\sin\theta$ that shows a complete cycle around the circle. Also see Figure 1.29 (page 62) for the graph of $\cos\theta$.

Instead of trying to memorize the values of the functions $\cos\theta$ and $\sin\theta$ separately, it's easier to remember them as a combined "package" $(\cos\theta, \sin\theta)$, which describes the x- and y-coordinates of the point P for the angle θ. Figure 1.43 shows the values of $\cos\theta$ and $\sin\theta$ for the angles 0, $\frac{\pi}{6}$ (30°), $\frac{\pi}{4}$ (45°), $\frac{\pi}{3}$ (60°), and $\frac{\pi}{2}$ (90°). These are the most common angles that often show up on homework and exam questions. For each angle, the x-coordinate (the first number in the bracket) is $\cos\theta$, and the y-coordinate (the second number in the bracket) is $\sin\theta$.

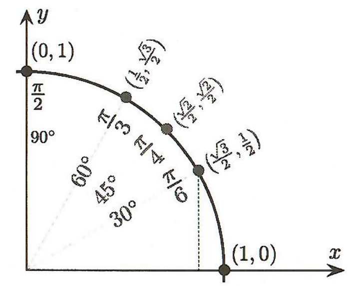


Figure 1.43: The combined $(\cos\theta, \sin\theta)$ coordinates for the points on the unit circle at the most common angles: 0, $\frac{\pi}{6}$ (30°), $\frac{\pi}{4}$ (45°), $\frac{\pi}{3}$ (60°), and $\frac{\pi}{2}$ (90°).

Note the values of $\cos\theta$ and $\sin\theta$ for the angles shown in Figure 1.43 are all combinations of the fractions $\frac{1}{2}$, $\frac{\sqrt{2}}{2}$, and $\frac{\sqrt{3}}{2}$. The square roots appear as a consequence of the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$. This identity tells us that the sum of the squared coordinates of each point on the unit circle is equal to one. Let's look at what this equation tells us for the angle $\theta = \frac{\pi}{6}$ (30°). Remember that $\sin(30°) = \frac{1}{2}$ (the length of the dashed line in Figure 1.43). We can plug this value into the equation $\cos^2(30°) + \sin^2(30°) = 1$ to find the value: $\cos(30°) = \sqrt{1 - \sin^2(30°)} = \sqrt{1 - \frac{1}{4}} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$.

The coordinates $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ for the angle $\theta = \frac{\pi}{4}$ (45°) are obtained from a similar calculation. We know the values of $\sin\theta$ and $\cos\theta$ must be equal for that angle, so we're looking for the number a that satisfies the equation $a^2 + a^2 = 1$, which is $a = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$. The values of $\cos(60°)$ and $\sin(60°)$ can be obtained from a symmetry argument. Measuring 60° from the x-axis is the same as measuring 30° from the y-axis, so $\cos(60°) = \sin(30°) = \frac{1}{2}$ and $\sin(60°) = \cos(30°) = \frac{\sqrt{3}}{2}$.

We can extend the calculations described above for all other angles that are multiples of $\frac{\pi}{6}$ (30°) and $\frac{\pi}{4}$ (45°) to obtain the $\cos\theta$ and $\sin\theta$ values for the whole unit circle, as shown in Figure 1.44.

Don't be intimidated by all the information shown in Figure 1.44! You're not expected to memorize all these values. The primary reason for including this figure is so you can appreciate the symmetries of the sine and cosine values that we find as we go around the circle. The values of $\sin\theta$ and $\cos\theta$ for all angles are the same as the values for the angles between 0° and 90°, but one or more of their coordinates has a negative sign. For example, 150° is just like 30°, except its x-coordinate is negative since the point lies to the left of the y-axis. Another use for Figure 1.44 is to convert between angles measured in degrees and radians, since both units for angles are indicated.

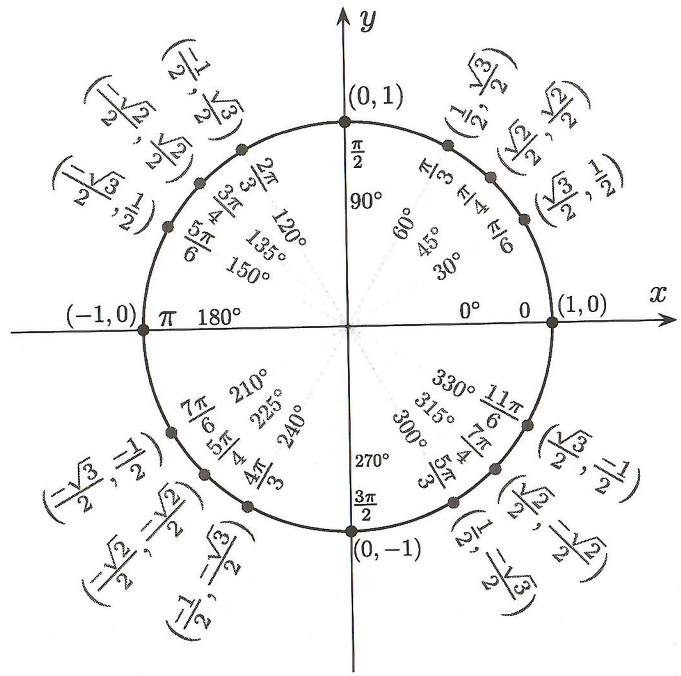


Figure 1.44: The coordinates of the point on the unit circle $(\cos\theta, \sin\theta)$ are indicated for all multiples of $\frac{\pi}{6}$ (30°) and $\frac{\pi}{4}$ (45°). Note the symmetries.

Non-unit circles

Consider a point $Q(\theta)$ at an angle of θ on a circle with radius $r \neq 1$. How can we find the x- and y-coordinates of the point $Q(\theta)$?

We saw that the coefficients $\cos\theta$ and $\sin\theta$ correspond to the x- and y-coordinates of a point on the *unit* circle ($r = 1$). To obtain the coordinates for a point on a circle of radius r, we must *scale* the coordinates by a factor of r:

$$Q(\theta) = (Q_x(\theta), Q_y(\theta)) = (r\cos\theta, r\sin\theta).$$

The take-away message is that you can use the functions $\cos\theta$ and $\sin\theta$ to find the "horizontal" and "vertical" components of any length r. From this point on in the book, we'll always talk about the length of the *adjacent* side as $x = r\cos\theta$, and the length of the *opposite* side as $y = r\sin\theta$. It is extremely important you get comfortable with this notation.

The reasoning behind the above calculations is as follows:

$$\cos\theta = \frac{\text{adj}}{\text{hyp}} = \frac{x}{r} \quad\Rightarrow\quad x = r\cos\theta,$$

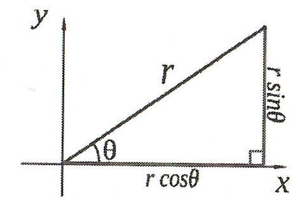


Figure 1.45: The x- and y-coordinates of a point at the angle θ and distance of r from the origin are given by $x = r\cos\theta$ and $y = r\sin\theta$.

and

$$\sin\theta = \frac{\text{opp}}{\text{hyp}} = \frac{y}{r} \quad\Rightarrow\quad y = r\sin\theta.$$

Calculators

Watch out for the units of angle measures when using calculators and computers. Make sure you know what kind of angle units the functions sin, cos, and tan expect as inputs, and what kind of outputs the functions $\sin^{-1}$, $\cos^{-1}$, and $\tan^{-1}$ return.

For example, let's see what we should type into the calculator to compute the sine of 30 degrees. If the calculator is set to degrees, we simply type: [3], [0], [sin], [=], and obtain the answer 0.5.

If the calculator is set to radians, we 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 the calculator.

Try computing $\cos(60°)$, $\cos(\frac{\pi}{3}\text{ rad})$, and $\cos^{-1}(\frac{1}{2})$ using your calculator to make sure you know how it works.

Exercises

E1.16 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?

E1.17 Convert the following angles from degrees to radians.

a) 30° **b)** 45° **c)** 60° **d)** 270°

Links

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

1.12 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. Sine angle sum

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

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

3. Cosine angle sum

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

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 identity 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*:

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

Self-similarity

Sin and cos are periodic functions with period 2π. Adding a multiple of 2π to the function's input does not change the function:

$$\sin(x+2\pi) = \sin(x), \qquad \cos(x+2\pi) = \cos(x).$$

This follows because adding a multiple of 2π brings us back to the same point on the unit circle.

Furthermore, sin and cos have symmetries with respect to zero,

$$\sin(-x) = -\sin(x), \qquad \cos(-x) = \cos(x),$$

within each π half-cycle,

$$\sin(\pi - x) = \sin(x), \qquad \cos(\pi - x) = -\cos(x),$$

and within each full 2π cycle,

$$\sin(2\pi - x) = -\sin(x), \qquad \cos(2\pi - x) = \cos(x).$$

Take the time to revisit Figure 1.26 (page 60), Figure 1.29 (page 62), and Figure 1.44 (page 76) to visually confirm that all the equations shown above are true. Knowing the points where the functions take on the same values (symmetries) or take on opposite values (anti-symmetries) is very useful in calculations.

Sin is cos, cos is sin

It shouldn't be surprising if I tell you that sin and cos are actually $\frac{\pi}{2}$-shifted versions of each other:

$$\cos(x) = \sin\left(x+\tfrac{\pi}{2}\right), \qquad \sin(x) = \cos\left(x-\tfrac{\pi}{2}\right).$$

the solutions to equations of the form $x + m = n$, where $m, n \in \mathbb{N}$. The rational numbers $\mathbb{Q}$ are necessary to solve for x in $mx = n$, with $m, n \in \mathbb{Z}$. To find the solutions of $x^2 = 2$, we need the real numbers $\mathbb{R}$. And in this section, we learned that the solutions to the equation $x^2 = -1$ are complex numbers $\mathbb{C}$. At this point you might be wondering if you're attending some sort of math party, where mathematicians write down complicated equations and—just for the fun of it—invent new sets of numbers to describe the solutions to these equations. Can this process continue indefinitely?

Nope. The party ends with $\mathbb{C}$. The fundamental theorem of algebra guarantees that any polynomial equation you could come up with—no matter how complicated it is—has solutions that are complex numbers $\mathbb{C}$.

Euler's formula

It turns out the exponential function is related to the functions sine and cosine. Lo and behold, we have *Euler's formula*:

$$e^{i\theta} = \cos\theta + i\sin\theta\,.$$

Inputting an imaginary number to the exponential function outputs a complex number that contains both cos and sin. Euler's formula gives us an alternate notation for the polar representation of complex numbers: $z = |z|\angle\varphi_z = |z|e^{i\varphi_z}$.

If you want to impress your friends with your math knowledge, plug $\theta = \pi$ into the above equation to find

$$e^{i\pi} = \cos(\pi) + i\sin(\pi) = -1,$$

which can be rearranged to obtain the equation $e^{i\pi} + 1 = 0$. The equation $e^{i\pi} + 1 = 0$ is called *Euler's identity*, and it shows a relationship between the five most important numbers in all of mathematics: Euler's number $e = 2.71828\ldots$, $\pi = 3.14159\ldots$, the imaginary number i, 1, and zero. It's kind of cool to see all these important numbers reunited in one equation, don't you agree?

One way to understand the equation $e^{i\pi} + 1 = 0$, is to think of $e^{i\pi}$ as the polar representation of the complex number $z = 1e^{i\pi} = 1\angle\pi$, which is the same as 1 rotated counterclockwise by π radians (180°) in the complex plane. We know $e^{i\pi} = 1\angle\pi = -1$ and so $e^{i\pi} + 1 = 0$.

De Moivre's formula

By replacing θ in Euler's formula with $n\theta$, we obtain de Moivre's formula:

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta.$$

De Moivre's 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 78) 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

[Intuitive proof of the fundamental theorem of algebra]
`https://www.youtube.com/watch?v=shEk8sz1oOw`

1.15 Solving systems of linear equations

Solving equations with one unknown—like $2x + 4 = 7x$, for instance—requires manipulating both sides of the equation until the unknown variable is *isolated* on one side. For this instance, we can subtract $2x$ from both sides of the equation to obtain $4 = 5x$, which simplifies to $x = \frac{4}{5}$.

What about the case when you are given *two* equations and must solve for *two* unknowns? For example,

$$x + 2y = 5,$$
$$3x + 9y = 21.$$

Can you find values of x and y that satisfy both equations?

Concepts

- x, y: the two unknowns in the equations
- $eq1, eq2$: a system of two equations that must be solved *simultaneously*. These equations will look like

  $$a_1x + b_1y = c_1,$$
  $$a_2x + b_2y = c_2,$$

  where as, bs, and cs are given constants.

Principles

If you have n equations and n unknowns, you can solve the equations *simultaneously* and find the values of the unknowns. There are several different approaches for solving equations simultaneously. We'll show three of these approaches for the case $n = 2$.

Solution techniques

When solving for two unknowns in two equations, the best approach is to *eliminate* one of the variables from the equations. By combining the two equations appropriately, we can simplify the problem to the problem of finding one unknown in one equation.

Solving by substitution

We want to solve the following system of equations:

$$x + 2y = 5,$$
$$3x + 9y = 21.$$

We can isolate x in the first equation to obtain

$$x = 5 - 2y,$$
$$3x + 9y = 21.$$

Now *substitute* the expression for x from the top equation into the bottom equation:

$$3(5 - 2y) + 9y = 21.$$

We just eliminated one of the unknowns by substitution. Continuing, we expand the bracket to find

$$15 - 6y + 9y = 21,$$

or

$$3y = 6.$$

We find $y = 2$, but what is x? Easy. To solve for x, plug the value $y = 2$ into any of the equations we started from. Using the equation $x = 5 - 2y$, we find $x = 5 - 2(2) = 1$.

Solving by subtraction

Let's now look at another way to solve the same system of equations:

$$x + 2y = 5,$$
$$3x + 9y = 21.$$

Observe that any equation will remain true if we multiply the whole equation by some constant. For example, we can multiply the first equation by 3 to obtain an equivalent set of equations:

$$3x + 6y = 15,$$
$$3x + 9y = 21.$$

Why did I pick 3 as the multiplier? By choosing this constant, the x terms in both equations now have the same coefficient.

Subtracting two true equations yields another true equation. Let's subtract the top equation from the bottom one:

$$\cancel{3x} - \cancel{3x} + 9y - 6y = 21 - 15 \quad \Rightarrow \quad 3y = 6.$$

The $3x$ terms cancel. This subtraction eliminates the variable x because we multiplied the first equation by 3. We find $y = 2$. To find x, substitute $y = 2$ into one of the original equations:

$$x + 2(2) = 5,$$

from which we deduce that $x = 1$.

Solving by equating

There is a third way to solve the system of equations

$$x + 2y = 5,$$
$$3x + 9y = 21.$$

We can isolate x in both equations by moving all other variables and constants to the right-hand sides of the equations:

$$x = 5 - 2y,$$
$$x = \frac{1}{3}(21 - 9y) = 7 - 3y.$$

Though the variable x is unknown to us, we know two facts about it: x is equal to $5 - 2y$ and x is equal to $7 - 3y$. Therefore, we can eliminate x by equating the right-hand sides of the equations:

$$5 - 2y = 7 - 3y.$$

We solve for y by adding $3y$ to both sides and subtracting 5 from both sides. We find $y = 2$ then plug this value into the equation $x = 5 - 2y$ to find x. The solutions are $x = 1$ and $y = 2$.

Discussion

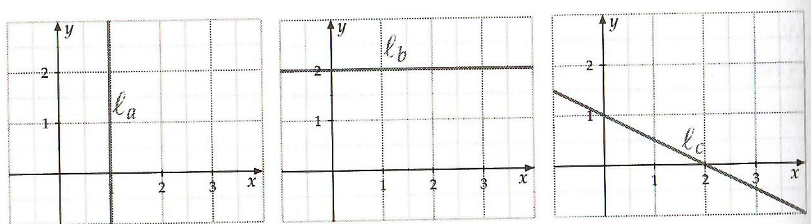
The repeated use of the three algebraic techniques presented in this section will allow you to solve any system of n linear equations in n unknowns. Each time you eliminate one variable using a substitution, a subtraction, or an elimination by equating, you're simplifying the problem to a problem of finding $(n - 1)$ unknowns in a system of $(n - 1)$ equations. In Chapter 3 we'll develop a more advanced, systematic approach for solving systems of linear equations.

Geometric solution

Solving a system of two linear equations in two unknowns can be understood geometrically as finding the point of intersection between two lines in the Cartesian plane. In this section we'll explore this correspondence between algebra and geometry to develop yet another way of solving systems of linear equations.

The algebraic equation $ax + by = c$ containing the unknowns x and y can be interpreted as a *constraint* equation on the set of possible values for the variables x and y. We can visualize this constraint geometrically by considering the coordinate pairs (x, y) that lie in the Cartesian plane. Recall that every point in the Cartesian plane can be represented as a coordinate pair (x, y), where x and y are the coordinates of the point.

Figure 1.59 shows the geometrical representation of three equations. The line ℓ_a corresponds to the set of points (x, y) that satisfy the equation $x = 1$, the line ℓ_b is the set of points (x, y) that satisfy the equation $y = 2$, and the line ℓ_c corresponds to the set of points that satisfy $x + 2y = 2$.



(a) $x = 1$ (b) $y = 2$ (c) $x + 2y = 2$

Figure 1.59: Graphical representations of three linear equations.

You can convince yourself that the geometric lines shown in Figure 1.59 are equivalent to the algebraic equations by considering individual points (x, y) in the plane. For example, the points $(1, 0)$, $(1, 1)$, and $(1, 2)$ are all part of the line ℓ_a since they satisfy the equation $x = 1$. For the line ℓ_c, you can verify that the line's x-intercept $(2, 0)$ and its y-intercept $(0, 1)$ both satisfy the equation $x + 2y = 2$.

The Cartesian plane as a whole corresponds to the set $\mathbb{R}^2$, which describes all possible pairs of coordinates. To understand the equivalence between the algebraic equation $ax + by = c$ and the line ℓ in the Cartesian plane, we can use the following precise math notation:

$$\ell : \{(x, y) \in \mathbb{R}^2 \mid ax + by = c\}.$$

In words, this means that the line ℓ is defined as the subset of the pairs of real numbers (x, y) that satisfy the equation $ax + by = c$.

Figure 1.60 shows the graphical representation of the line ℓ.

You don't have to take my word for it, though! Think about it and convince yourself that all points on the line ℓ shown in Figure 1.60 satisfy the equation $ax + by = c$. For example, you can check that the x-intercept $(\frac{c}{a}, 0)$ and the y-intercept $(0, \frac{c}{b})$ satisfy the equation $ax + by = c$.

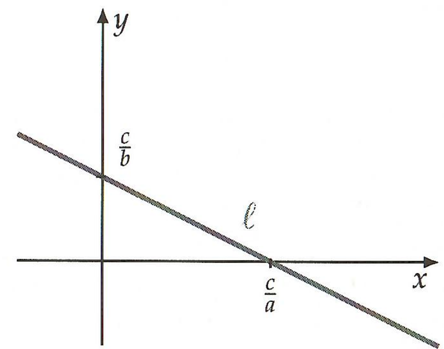


Figure 1.60: Graphical representation of the equation $ax + by = c$.

Solving the system of two equations

$$\begin{aligned} a_1x + b_1y &= c_1, \\ a_2x + b_2y &= c_2, \end{aligned}$$

corresponds to finding the intersection of the lines ℓ_1 and ℓ_2 that represent each equation. The pair (x, y) that satisfies both algebraic equations simultaneously is equivalent to the point (x, y) that is the intersection of lines ℓ_1 and ℓ_2, as illustrated in Figure 1.61.

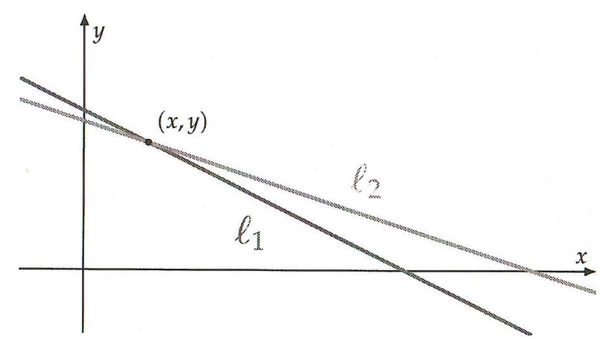


Figure 1.61: The point (x, y) that lies at the intersection of lines ℓ_1 and ℓ_2.

Example Let's see how we can use the geometric interpretation to solve the system of equations

$$x + 2y = 5,$$
$$3x + 9y = 21.$$

We've already seen three different *algebraic* techniques for finding the solution to this system of equations; now let's see a *geometric* approach for finding the solution. I'm not kidding you, we're going to solve the exact same system of equations a fourth time!

The first step is to draw the lines that correspond to each of the equations using pen and paper or a graphing calculator. The second step is to find the coordinates of the point where the two lines intersect as shown in Figure 1.62. The point $(1,2)$ that lies on both lines ℓ_1 and ℓ_2 corresponds to the x and y values that satisfy both equations simultaneously.

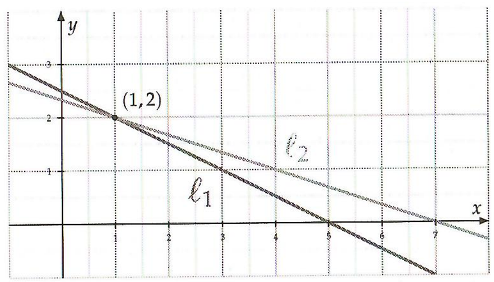


Figure 1.62: The line ℓ_1 with equations $x + 2y = 5$ intersects the line ℓ_2 with equation $3x + 9y = 21$ at the point $(1,2)$.

Visit the webpage at `www.desmos.com/calculator/exikik615f` to play with an interactive version of the graphs shown in Figure 1.62. Try changing the equations and see how the graphs change.

Exercises

E1.23 Plot the lines ℓ_a, ℓ_b, and ℓ_c shown in Figure 1.59 (page 106) using the Desmos graphing calculator. Use the graphical representation of these lines to find: **a)** the intersection of lines ℓ_c and ℓ_a, **b)** the intersection of ℓ_a and ℓ_b, and **c)** the intersection of lines ℓ_b and ℓ_c.

E1.24 Solve the system of equations simultaneously for x and y:

$$2x + 4y = 16,$$
$$5x - y = 7.$$

E1.25 Solve the system of equations for the unknowns x, y, and z:

$$2x + y - 4z = 28,$$
$$x + y + z = 8,$$
$$2x - y - 6z = 22.$$

E1.26 Solve for p and q given the equations $p + q = 10$ and $p - q = 4$.

1.16 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, set operations, and set containment relations. This section introduces all the relevant concepts.

Definitions

- *set*: a collection of mathematical objects
- S, T: the usual variable names for sets
- $s \in S$: this statement is read "s is an element of S" or "s is in S"
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$: some important number sets: the naturals, the integers, the rationals, and the real numbers, respectively.
- $\varnothing$: the *empty set* is a set that contains no elements
- $\{ \ldots\ldots \}$: the curly brackets are used to define sets, and the expression inside the curly brackets describes the set contents.

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 set difference $S \setminus T$ corresponds to the elements of S that are not in T.

Set relations:

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