Math 8 — Functions and Modeling

Neil Donaldson

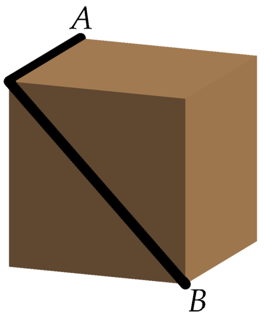
Spring 2026

Introduction

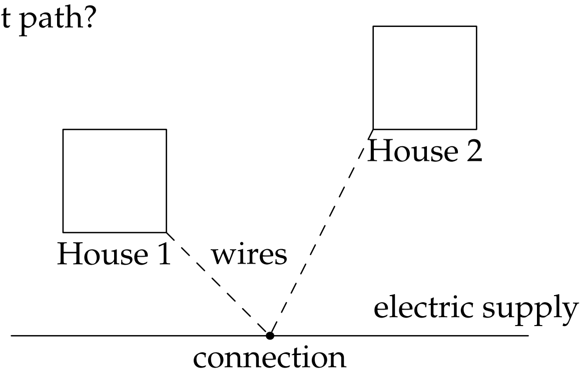
This course aims refresh and reinforce some of the conceptual foundations of grade-school mathematics for future teachers. Teaching is often a matter of *selection*: choosing examples and explanations suited to the level of your students. To do this effectively, you must understand concepts at a higher level than you teach. The *mathematics* in this course shouldn't be difficult for math majors. Instead this is an opportunity to think about content you might one day teach. How would *you* explain the material? How does it connect to other ideas? Can you anticipate students' questions, and how would you respond?

We start with two motivational problems.[1]

1. How can we travel across the surface of a cube between two opposite vertices so that the path is as short as possible? Should you follow the path indicated? If yes, explain why. If not, how should you find the shortest path?



Problem 1



Problem 2

2. Two houses are to be connected to the electricity supply using a single connection. How should we determine where to place the connection so as to minimize the required length of wire? What information do you need in order to find the connection point?

Don't just think about the right answer! Consider how you might discuss these problems with students of different ability levels. Why might calculus *not* be a sensible approach? Are there any similarities between the two problems? Brainstorm some strategies...

---

[1] We are grateful to materials from UT Austin's UTeach program for suggesting several of the examples in this course including these motivational problems.

1 Sets & Functions

1.1 Basic Definitions

When did you first encounter a function? Middle-school? Elementary school? Earlier…How would you define *function* to someone with limited mathematical knowledge? Would you use words like *rule, assign, element, domain, vertical line test*, etc.? How helpful are these to your audience?

Examples 1.1. How would you explain the idea that the following do or do not represent functions?

1. $y = x^2$
2. Mon: fish, Tue: pork, Wed: fajitas, Thur: carbonara, Fri: pizza
3. $(3,5), (2,6), (4,2), (3,1)$.
4. $x^2 = y^2$

After considering the examples, perhaps you settle on a semi-formal definition:

A function f is rule assigning exactly one output $f(x)$ to each input x

Is this a useful definition? In what ways is it imprecise? Does the imprecision matter?

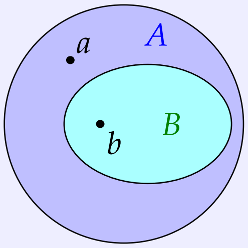
Of course the answers to these questions depend on your audience! What ideas do you want to convey to your students, and can you do so without overburdening and intimidating them? To begin working towards a more complete picture, we need to think about what the *inputs* and *outputs* are allowed to be. This introduces some vocabulary and requires a small amount of set notation.

Definition 1.2. A *set* A is a collection of objects, or *elements*.[2] The notation $a \in A$ means that a is an element of A, sometimes read 'a lies in A.' In the abstract, sets are often written upper case and their elements in lower case.

A set B is a *subset* of a set A, written $B \subseteq A$, if every element of B is also an element of A: that is,

$$b \in B \implies b \in A \qquad \text{(“If } b \text{ lies in } B\text{, then } b \text{ lies in } A\text{”)}$$

The picture illustrates sets A, B and elements a, b for which $B \subseteq A$, $a \in A$, $b \in B$ and $a \notin B$ (a does not lie in B).



The reason we need this language when discussing functions is that the inputs and outputs of a function are *elements of sets*.

---

[2] This is enough for our purposes, though a course in set theory should convince you that this definition is not without its own problems. Selection is always at work…

Example 1.3. Suppose the elements of a set A are the numbers 1, 3, 5, 7 and 9. The simplest way to write this is using *roster notation*: we list the elements (in any order) between braces

$$A = \{1, 3, 5, 7, 9\}$$

Subsets are commonly expressed using *set-builder notation*. Here is a subset of A:

$$B = \{a \in A : 2 < a < 8\} \qquad \text{(“The set of } a \text{ in } A \text{ such that } a \text{ lies strictly between 2 and 8”)}$$

In roster notation, $B = \{3, 5, 7\}$. Can you express B in other ways using set-builder notation?

Several familiar sets of numbers can be expressed using informal combinations of these notations.

Natural numbers $\mathbb{N} = \{1, 2, 3, 4, \dots\}$. For instance, $5 \in \mathbb{N}$ but $-3 \notin \mathbb{N}$. Most students think this is the meaning of *number* until they've done *lots* of math.

Integers $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, 3, \dots\}$. For instance, $-4 \in \mathbb{Z}$ but $\frac{4}{5} \notin \mathbb{Z}$.

Rational numbers (fractions) $\mathbb{Q} = \{\frac{p}{q} : p \in \mathbb{Z},\ q \in \mathbb{N}\}$. For instance $-\frac{6}{7} \in \mathbb{Q}$.

Real numbers $\mathbb{R}$. For instance, $\sqrt{2} \in \mathbb{R}$. A formal definition is difficult, though we often informally visualize $\mathbb{R}$ as a *ruler*. *Intervals* are particularly important subsets, e.g.,

$$[-4, \pi) = \{x \in \mathbb{R} : -4 \leq x < \pi\} \qquad \text{(a half-open interval)}$$

You should also have informally encountered the notion of *irrationality*: for instance, $\sqrt{2}$ and π are real numbers but not rational numbers.

The subset relationships between these sets are in the order listed: $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.

We now turn to a *very* formal definition of function.

Definition 1.4. The *Cartesian product* of sets A and B is the set of *ordered pairs*

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

A *function* from A to B is a non-empty subset $f \subseteq A \times B$ which satisfies the *vertical line test*:

$$\text{For each } a \in A \text{, there is a \emph{unique} } b \in B \text{ such that } (a, b) \in f \qquad (*)$$

Instead of writing $f \subseteq A \times B$ and $(a, b) \in f$, we use the more familiar notation

$$f : A \to B \quad \text{and} \quad f(a) = b$$

To a function $f : A \to B$ are associated three useful sets:

- Domain: $\operatorname{dom} f = A$ is the set of *inputs*.
- Codomain: $\operatorname{codom} f = B$ is the set of *possible outputs*.
- Range: $\operatorname{range} f = \{b \in B : b = f(a) \text{ for some } a \in A\}$ is the set of *realized outputs*.

This probably isn't the definition you should give to 10th graders! How much of this is helpful in a given context? Is it important that *you've* been exposed to this definition?

Example (1.1.2 cont.). We revisit our food-based example in this formal setting. To properly view this as a function $f : A \to B$, we have to carefully label the constituent sets.

$$A = \{\text{Mon, Tue, Wed, Thu, Fri}\}, \qquad B = \{\text{carbonara, fajitas, fish, pizza, pork}\},$$

$$f = \{(\text{Mon, fish}),\ (\text{Tue, pork}),\ (\text{Wed, fajitas}),\ (\text{Thu, carbonara}),\ (\text{Fri, pizza})\}$$

The domain A should be clear, but we had to make a choice for the codomain B: in this case we made it to equal to *range.* Can you suggest a different choice for B? Note how the function is simply the list of ordered pairs (day, food). Try the other examples yourself.

Representing Functions

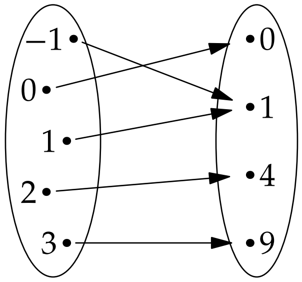
Functions can be represented in various ways. We illustrate a few in an example.

Example 1.5. We consider the familiar *formula/rule* $f(x) = x^2$ in several contexts.

Table This presentation is most helpful when the domain is very small.

x	-1	0	1	2	3
$f(x)$	1	0	1	4	9

Arrows Also useful when the domain is small. Both this and the table describe f when $\operatorname{dom} f = \{-1, 0, 1, 2, 3\}$ and $\operatorname{range} f = \{0, 1, 4, 9\}$.



Graph This is the set of ordered pairs $\{(x, x^2) : x \in \operatorname{dom} f\}$. In the context of the formal definition (1.4), *the graph is the function!*

For formulæ whose inputs and outputs are real numbers, two conventions are often observed:

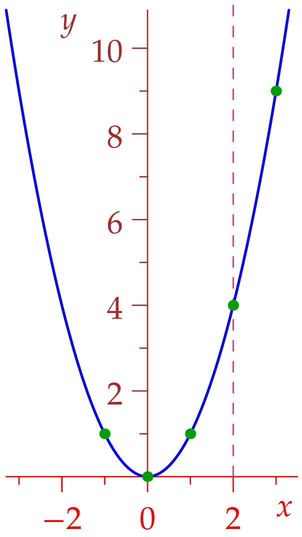
- The domain is *implied* to be all values for which the formula makes sense (don't do silly stuff like dividing by zero!).
- The codomain is the set of real numbers.

If no other information is provided, we'd assume that the formula $f(x) = x^2$ represents a function $f : \mathbb{R} \to \mathbb{R}$, in which case the range (the set of realized outputs) is the interval

$$\operatorname{range} f = \{x^2 \in \mathbb{R} : x \in \mathbb{R}\} = [0, \infty)$$

For 'calculus functions' like this, the vertical line test $(*)$ really does involve vertical lines; every vertical line intersects the graph in precisely one point.

In the picture, the dots are the graph when the domain is the set $\{-1, 0, 1, 2, 3\}$ (as described in the table/arrow-diagram).



Are there other ways to represent a function? How might your audience influence your choice?

We'll have many more chances to analyze and represent functions as we move on. The challenge here is primarily one of vocabulary. A student has perhaps 15 years—elementary school to early college—to become comfortable with all this language. How would you help a student, regardless of where they are on that journey, to take the next step?

Exercises 1.1. *Key concepts: Function, Domain, Range, Codomain, Representations of Functions, Multiple descriptions/selection*

1. Let d represent the cost in millions of dollars to produce n cars, where n is measured in 1000s. As clearly as you can, explain what is meant by $d(25) = 431$.

2. A movie theater seats 200 people. For any particular show, the amount of money the theater takes in is a function of the number of people n in attendance. If a ticket costs \$25, describe the domain and range of the function using set notation.

3. Temperature readings T were recorded every two hours from midnight to noon. Time t was measured in hours from midnight.

t	0	2	4	6	8	10	12
T (°F)	82	75	74	75	84	90	93

   Plot the readings and use them to sketch a rough graph of T as a function of t. Now use your graph to estimate the temperature at 10:30 a.m..

4. State parts 1, 3 and 4 of Example 1.1 using the formal language of Definition 1.4. If you have a function, state the domain and range and explain how you know you have a function. If you don't have a function, explain why not.

   (*Since insufficient information is provided, there is no single correct answer!*)

5. (a) Let $A = \{1, 3, 5, 7, 9\}$. Explain in words what is meant by the set

   $$B = \{x \in A : x^2 > 10\}$$

   Now state B in roster notation.

   (b) Find the set $C = \{x \in \mathbb{N} : (x-1)^2 < 16\}$ in roster notation.

   (c) Find the Cartesian product $B \times C$ in roster notation. Is it the same as $C \times B$?

6. Suppose that $f : \{-2, -1, 0, 1, 2\} \to \mathbb{R}$ is defined by the formula $f(x) = x^3 - 4x + 1$.

   Describe f using a table, an arrow diagram and a graph.

7. Find the implied domain and range (as subsets of the real numbers $\mathbb{R}$) for the functions defined by each rule:

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

   (b) $g(x) = \sqrt{x^2 - 16x}$

   (c) $h(x) = \frac{1}{x}\sqrt{4x - x^2}$

   (*What is the largest set of real numbers for which the formula makes sense*?)

8. You should be familiar with the Cartesian plane: $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$. This is the set of ordered pairs, typically thought of as co-ordinates.

    (a) Draw a picture of the set $A = \{(x, \sqrt{x}) : x \geq 0\}$. Is this the graph of a function? Explain.

    (b) Draw a picture of the set $B = \{(y^2, y) : y \in \mathbb{R}\}$. Is this the graph of a function? Explain.

    (c) Draw a picture of the *set* $C = \{3, 4\} \times (3, 4)$. Be careful to distinguish between an ordered pair and an interval. Consider how you would explain this to a student.

9. You ask your students to determine the range of the function f defined by the rule $f(x) = x^2$ with domain the interval $[-5, 2]$. You obtain various responses, including $[25, 4]$, $[4, 25]$, and $[-25, 4]$.

    What is going wrong? What is the correct answer, and how would you explain it to your students?

    More generally, if $\operatorname{dom} f = [a, b]$ (where $a \leq b$), what is $\operatorname{range} f$?

10. The unit circle is often represented by the implicit equation $x^2 + y^2 = 1$.

    (a) Draw the circle and explain why the full circle isn't the graph of a function.

    (b) Describe *two* functions $f : [-1, 1] \to \mathbb{R}$ and $g : [-1, 1] \to \mathbb{R}$ whose graphs together comprise the circle. What are the ranges of each function?

11. Functions become more difficult when they take functions as inputs! You've already encountered this in calculus. Here are two examples.

    (a) The differential operator $D = \frac{\mathrm{d}}{\mathrm{d}x}\Big|_{x=0}$ can be thought of as a function: $D(f) = f'(0)$ takes as its input a function f and returns its derivative evaluated at $x = 0$.

        Give an example of a function f which does not lie in the (implied) domain of D.

    (b) (If you've done integral calculus) Let I be the integral operator $I = \int_0^x$ whose domain is the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$. Otherwise said, if f is a continuous function with domain $\mathbb{R}$, then $I(f) = \int_0^x f(t)\,\mathrm{d}t$.

        State at least one function that is *not* in the range of I. Explain your answer!

1.2 Linear Polynomials

Perhaps the simplest functions are those whose graphs are straight lines. These are the *linear polynomials*,

$$y = f(x) = mx + c, \tag{$*$}$$

where m, c are constants. Such functions make for very simple models:[3]regardless of the initial input, if we increase x by Δx, then the output changes by a constant multiple $\Delta y = m\Delta x$.

By this stage of your mathematical career, you should be *very* familiar with linear polynomials. From a teaching viewpoint this can be dangerous. The real challenge is how to explain your understanding to those who don't have it. For instance, some of your earliest forays into algebra likely involved finding equations of the straight line through two given points. Ask a college freshman how to solve such a problem and they'll likely offer some version of the *point-slope formula*. But what is this, and why does it work? More importantly, could you teach it?

Abstract concepts in mathematics are very often just a way of writing down all possible examples simultaneously. In such a case, it is usually better to start with an example first before sharing any general formulæ. Here are two questions you could use to introduce abstract concepts around linear polynomials.

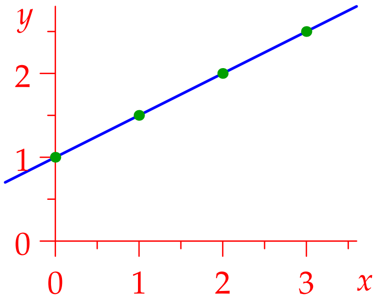
1. How to graph $y = mx + c$? Why is the graph a straight line?

Example 1.6. Suppose we want to graph $y = f(x) = \frac{1}{2}x + 1$.

One way to start is to consider the most obvious point on the graph: plug in $x = 0$ to obtain $y = 1$ and the point $(0, 1)$. Next, consider other values of x, for instance integers, and plot these:

x	0	1	2	3	4
$f(x)$	1	1.5	2	2.5	3

Joining these points, it certainly appears as if the graph is a straight line. At this stage you may have enough to convince a middle-school student.



In the abstract, to graph $y = f(x) = mx + c$,

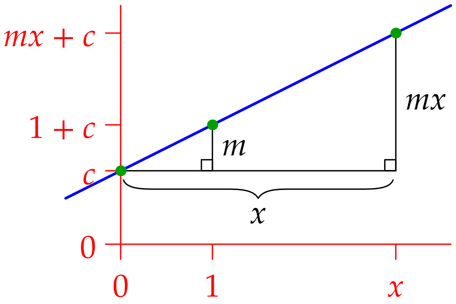
- Find the y-intercept: $x = 0 \Longrightarrow y = f(0) = c$ yields the point $(0, c)$.
- Plot a few more points $(1, m + c)$, $(2, 2m + c)$, $(3, 3m + c)$ and draw the line through them.

But what if a student is not convinced? What if you are teaching higher-level students? Just because the graph *looks like* a line doesn't guarantee that it is a line. At this point you might have to appeal to some elementary geometry: *similar triangles*.

---

[3] Given experimental data or a physical situation relating two quantities x and y, a *linear model* is an linear polynomial $(*)$. In practice, models only *approximate* to real-world data. Later in the course we'll consider what should be meant by, and how to find, a 'good' linear model for approximately linear data.

Plot three points: the y-intercept $(0, c)$, the point $(1, 1 + c)$, and a general point $(x, mx + c)$, and complete two right-triangles as pictured. These triangles are *similar.* Why?



- Perhaps your students observe that both are right-angled with the same slope:

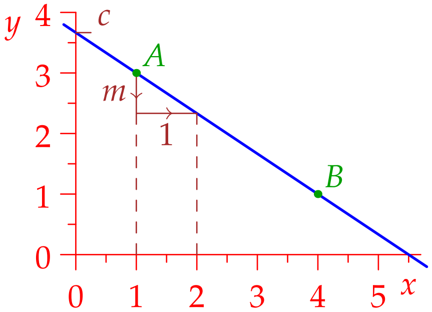
$$\frac{\text{rise}}{\text{run}} = \frac{m}{1} = \frac{mx}{x}$$

- Perhaps they've studied trigonometry and say that the left angle in both triangles is $\arctan m$.

Both observations are just rephrasings of the SAS triangle similarity theorem.[4] The upshot is that both triangles share an angle at c, whence the three points are collinear.

2. How to find the equation of the line through two points?

Example 1.7. We find the equation of the linear polynomial $y = mx + c$ passing through the points $A = (1, 3)$ and $B = (4, 1)$.



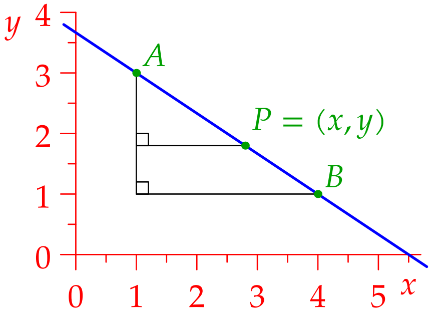
Starting by plotting the points and sketching the line. The natural approach is to substitute both given points into the equation to find two relationships between m and c

$$\begin{cases} 3 = m + c \\ 1 = 4m + c \end{cases}$$

Your students might be comfortable solving this system of equations algebraically for m and c. But algebra is only convincing if you're already comfortable with the language! Instead consider what you are trying to motivate: the *point-slope formula*. We want the slope m, so...

Ask the students how to get from point A to point B. Hopefully you get some version of *along three and down two*. But they've just told you the slope!

$$m = \frac{\text{rise}}{\text{run}} = \frac{2\text{ down}}{3\text{ along}} = -\frac{2}{3}$$



To get to any other point $P = (x, y)$ on the line, we're back to similar triangles:

$$-\frac{2}{3} = m = \frac{\text{rise}}{\text{run}} = \frac{y-3}{x-1}$$
$$\implies y = 3 - \frac{2}{3}(x-1) = \frac{1}{3}(11 - 2x)$$

---

[4] If two triangles have two pairs of sides in a common ratio (here m) and the angle between them congruent (here 90°), then the triangles are similar. This is *much* older (c. 2300 years) than graphs and equations (400 years).

This process works for any two given points with distinct x co-ordinates.

Theorem 1.8 (Point-slope formula). *If $x_A \neq x_B$, then the straight line through points $A = (x_A, y_A)$ and $B = (x_B, y_B)$ has slope*

$$m = \frac{\Delta y}{\Delta x} = \frac{y_B - y_A}{x_B - x_A}$$

and equation

$$y = y_A + m(x - x_A) = y_A + \frac{y_B - y_A}{x_B - x_A}(x - x_A)$$

The details are in Exercise 5. Depending on their level, the abstract point-slope formula might be too intimidating for your students. Even if they are capable of applying it, you might feel that its derivation is beyond them. Indeed it might be so for some students, but consider several counterpoints:

- Once a student has developed comfort with concrete examples, the point-slope formula and its proof help to summarize and unify what they've learned. An abstract discussion can help build confidence by convincing a student that any such problem can be solved the same way.
- The most helpful elementary proofs are those which replicate an examples. Exercise 5 involves no trickery; it simply *reinforces the core technique* by applying it in general.
- Encouraging students to think abstractly is an overarching learning outcome of mathematics. You might get push-back, but it's part of the job...

Another common issue is that students resist examples and clamor for the point-slope formula: "So we just need to memorize the formula, right?" In the moment, the desire for *efficiency* often dominates student thinking: they have many demands on their time and are tempted by a short-cut, the promising of knowing how to do a problem without putting in the work necessary to understand how or why it works. Fighting against this is hard, but consider the following.

- Effective learning is rarely efficient. Your goal as a teacher is often to force students to think slowly about something, even when they don't want to.
- Memorizing formulae and algebraic algorithms sacrifices geometric intuition. Mathematics is a language for modeling and describing the real world. The pictures, the *geometry*, are what is real.
- Mathematics is interconnected. In this case, the fundamental underlying concept of the point-slope formula is SAS similarity. Linking the discussion to this helps motivate the study of and reinforce foundational geometry.
- You are doing a student no favors by letting them get away with a *see this do that* approach to mathematics. By the end of high-school, there are too many concepts to memorize this way. Your job is to build the scaffolding, not paint the wall for them...

Modeling with straight lines

Often the challenge of modeling lies in converting a word problem into algebra—don't underestimate how hard students find this! Here is a simple, if disguised, straight line model.

Example 1.9. Beaker A contains a 300 ml solution of a 2% acid. Beaker B contains 400 ml of the same acid but of unknown concentration. When the beakers are mixed together the result is an acid with concentration 6%. What was the concentration in beaker B?

It should seem natural to denote the unknown concentration (beaker B) by x. After mixing:

- The total volume of the solution is $300 + 400 = 700$ ml.
- The solution contains $300 \times \frac{2}{100} + 400x = 6 + 400x$ ml of pure acid.

Putting this together, the concentration of the resulting solution is a linear polynomial function of x:

$$C(x) = \frac{6 + 400x}{700}$$

The problem is now easily solved: $C(x) = \frac{6}{100} \Longrightarrow x = \frac{9}{100} = 9\%$.

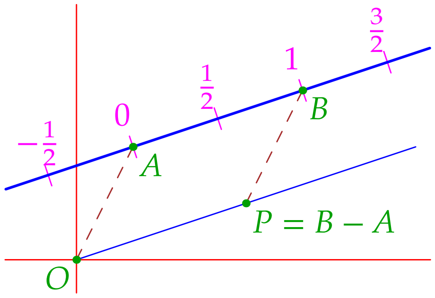
While the algebra in the example is straightforward to those with sufficient experience, consider what natural mistakes grade-school students might make with this problem...

Parametrized Lines

Here is an alternative visualization of a straight line. Imagine placing a *ruler* so that its zero point is at the origin $O = (0, 0)$ and its "1" lies at a point $P = (p, q)$. If t (a real number) is the number on the ruler, then the points on the line have co-ordinates

$$tP = (tp, tq) \tag{$*$}$$

To describe the line through points A and B, place a virtual ruler so that 0 corresponds to A and 1 to B. Slide the ruler so that A moves to the origin O: this amounts to *subtracting* the co-ordinates of A from all points on the line. We obtain a *parallel line* through the origin, with B transformed to $P = B - A$. Putting this together with $(*)$ results in a parametrized description of the line:



$$\begin{aligned}(x, y) &= A + tP = A + t(B - A)\\ &= (1 - t)A + tB\end{aligned}$$

The points corresponding to various values of t are marked.

Contrast this description with the linear polynomial approach. One challenge is that a line may be parametrized using infinitely many different rulers (choose *any* two points on the line!), whereas the linear polynomial description is unique. Does the parametrized approach have any advantages? Which description is easier to understand or to work with?

In the Exercises we make sure that the two descriptions of a line correspond. The discussion is little more than the generalization of an example.

Example 1.10. The line through points $A = (3,6)$ and $B = (-1,4)$ may be parametrized

$$(x,y) = (1-t)(3,6) + t(-1,4) = \big(3-4t, 6-2t\big)$$

To convert this to a linear polynomial, first solve for t in terms of x,

$$x = 3-4t \implies t = \frac{1}{4}(3-x)$$

Now substitute this into our expression for y:

$$y = 6-2t = 6 - \frac{2}{4}(3-x) = \frac{1}{2}x - \frac{9}{2}$$

Exercises 1.2. *Key concepts: Linear Polynomial, Relationship with a straight line, Point-slope formula, Parametrizing a line using a ruler*

1. The cost of gasoline is \$4.20 per gallon on January 1st and \$4.90 on March 1st. State a *linear* function/model for the cost of gasoline as a function of time.
2. You have a choice of three different cell-phone plans.
   (a) No monthly charge and 10¢ per minute for all calls.
   (b) \$10 per month and 5¢ per minute for all calls.
   (c) \$30 per month, regardless of how many calls you make.

   How should you determine which plan to purchase?
3. Revisit Exercise 1.1.3. Find an approximate linear model $T(t) = mt + c$ for this data.

   (*There is no perfect answer*)
4. Revisit the beakers problem (Example 1.9). This time suppose we know that the concentration in beaker B is 9%. How much from beaker B should we pour into beaker A to obtain an acid with concentration 5%? Would you consider this a linear polynomial problem? Why/why not?
5. Suppose points $A = (x_A, y_A)$ and $B = (x_B, y_B)$ are given.
   (a) If $x_B \neq x_A$, use the method of Example 1.7 to prove the point-slope formula (Theorem 1.8). Otherwise said: find the equation $y = mx + c$ of the line through these points.

(b) Find a parametrized description of the same line (page 10) where A corresponds to 0 and B to 1. If $x_B \neq x_A$, make things match up with your answer to part (a).

What parametrization do you get if $A = (0, c)$ and $B = (1, m + c)$?

6. A straight line is sometimes described as the set of points $(x, y) \in \mathbb{R}^2$ satisfying an equation of the form

$$ax + by = c$$

where a, b, c are constants and a, b are not both zero. How does this approach differ from our use of linear polynomials? Are these descriptions covering exactly the same collection of objects?

7. Throughout mathematics (particularly within *linear algebra*), a function $f : \mathbb{R} \to \mathbb{R}$ is said to be *linear* if it satisfies the condition

$$\text{For all } \lambda, x \in \mathbb{R}, \quad f(\lambda x) = \lambda f(x)$$

Is this the same thing as a linear polynomial? Explain.

1.3 Quadratic Polynomials

Quadratic polynomials are functions of the form $y = f(x) = ax^2+bx+c$ where $a \neq 0$. The simplest is $y = x^2$, the standard parabola opening upwards: it is straightforward to convince students of the shape of its graph just by plotting a few integer points. For more general quadratics, here are some commonly encountered activities:

1. Find the *roots/zeros* of f, the solutions x to the equation $f(x) = 0$.
2. Sketch the *graph* of the function f.
3. Use quadratic functions to model a real-world problem.

You likely know two methods for finding zeros: factorizing and the quadratic formula, each of which has its problems. With experience it is easy to spot that

$$x^2 + 2x - 15 = (x-3)(x+5) = 0 \iff x = 3 \text{ or } x = -5$$

though the required creativity can make this difficult, particularly when coefficients are large. Students often prefer the quadratic formula since it always works, though at the cost of some intimidating algebra. We'll think about factorization shortly. First, we see how *completing the square* lies behind both the quadratic formula and the standard approach to graphing quadratic functions.

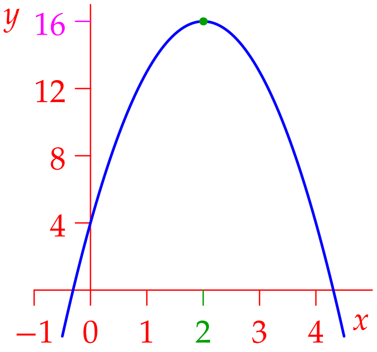
Example 1.11. Describe/graph the parabola $y = -3x^2 + 12x + 4$.

Pay attention to the x terms; $-3x^2 + 12x = -3(x^2 - 4x)$. Observe that

$$-3(x-2)^2 = -3(x^2 - 4x + 4) = -3x^2 + 12x - 12$$

gives most of what we want: note how we *divided the x-coefficient by two.* To finish, just tidy everything up

$$y = (-3x^2 + 12x - 12) + 16 = -3(x-2)^2 + 16$$



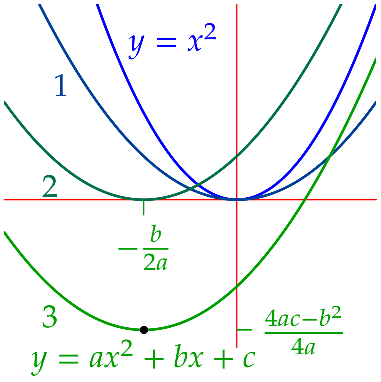
The parabola therefore opens downwards ($-3 < 0$) with its vertex (maximum) at $(x, y) = (2, 16)$. Otherwise said, we translate the parabola $y = -3x^2$ to the right by 2 and up by 16.

This is easy, if intimidating, to repeat in general:

$$\begin{aligned} ax^2 + bx + c &= a\left(x^2 + \frac{b}{a}x\right) + c = a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a^2}\right] + c \\ &= a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a} \end{aligned} \qquad (*)$$

Completing the square is what tells us how to graph the function! Starting with the standard parabola $y = x^2$:

1. Vertically scale by a
2. Shift horizontally by $-\frac{b}{2a}$
3. Shift vertically by $\frac{4ac-b^2}{4a}$



Even more is true, for $(*)$ may easily be solved for x: *completing the square is the quadratic formula.*

Theorem 1.12. *If* $a \neq 0$*, then* $ax^2 + bx + c = 0 \iff x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$

Example (1.11 cont). Our analysis suggests two equivalent methods for finding the roots.

1. Quadratic formula: with $a = -3$, $b = 12$ and $c = 4$, we have

$$x = \frac{-12 \pm \sqrt{12^2 - 4(-3)\cdot 4}}{2(-3)} = \frac{-12 \pm 4\sqrt{3^2+3}}{-6} = 4 \pm \frac{2}{\sqrt{3}}$$

   Note the difficult surd expression and how much work it was to simplify.

2. Use the fact that we've already completed the square:

$$-3(x-2)^2 + 16 = 0 \iff (x-2)^2 = \frac{16}{3} \iff x = 2 \pm \frac{4}{\sqrt{3}}$$

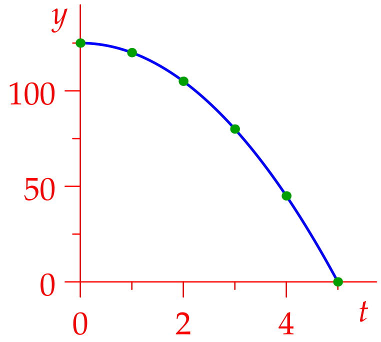
   It is often simpler to complete the square directly!

Quadratics are often used as models. For instance, the motion of a projectile can be modeled using quadratic polynomials, an observation going back to at least to Galileo in the early 1600s: the distance travelled by a falling body is proportional to the *square* of the time taken $y(t) - y(0) \propto t^2$.

Example 1.13. A body is dropped from a height of 125 meters, taking exactly 5 seconds to reach the ground. Its height at time t seconds is given by $y(t) = 125 - 5t^2$ m.

This certainly fits with Galileo's observation: $y(t) - y(0) = -5t^2$. Over each 1s interval, consider how far the body falls:

t	0	1	2	3	4	5
$y(t)$	125	120	105	80	45	0
$y(t)-y(0)$	0	-5	-20	-45	-80	-125
Δy		-5	-15	-25	-35	-45



Since each interval has duration 1 s, each Δy is the *average speed* of the body over that interval.

You've likely met such problems in calculus and are sorely tempted compute the *velocity* $y'(t) = -10t\,\mathrm{m/s}$ and *acceleration* $y''(t) = -10\,\mathrm{m/s^2}$ by differentiation. Historically, and in introductory calculus, it is problems like these that *motivate the definition* of the derivative: note how the last line of the table really suggests that the instantaneous velocity is a linear function!
Armed with calculus, Galileo's observation is that the height $y(t)$ solves a differential equation.

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = -g$$

where g (approximately $32\,\mathrm{ft/s^2}$ or $10\,\mathrm{m/s^2}$) is the constant acceleration due to gravity. This can, of course, be solved by integrating twice:

$$y'(t) = -gt + v_0 \implies y(t) = -\frac{1}{2}gt^2 + v_0t + h_0 \tag{$*$}$$

where the constants of integration are the initial height h_0 and vertical velocity v_0. Unless you are explicitly teaching calculus or Newtonian physics, this is probably *not* the place to start.

Example 1.14. A frisbee is stuck 15 m up a tree. Standing 10 m from the base of the trunk, you throw a ball in the hope of knocking the frisbee out of the tree.

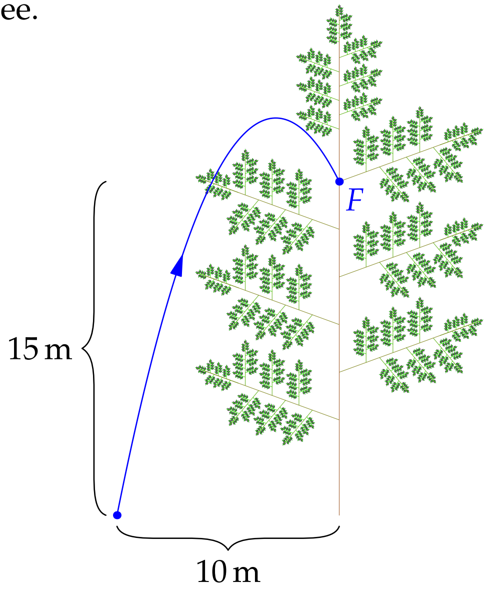
To model this problem, consider the horizontal and vertical motions separately.

Horizontal Since there is no horizontal acceleration, the horizontal speed must be constant: $x(t) = pt + q$ is a *linear function* of time.

Vertical Applying $(*)$ with $g = 10\,\mathrm{m/s^2}$, we see that $y(t) = -5t^2 + rt + s$ is a *quadratic function* of time.

Substituting $t = \frac{x-q}{p}$ into y yields a quadratic function for the trajectory

$$y(x) = ax^2 + bx + c$$



The details of the solution are left to Exercise 6. For the present, consider why you should expect *two answers* to the problem. Can you explain why *without* explicitly finding the solutions?

Exercises 1.3. *Key concepts: Completing the square, Graphing, Quadratic formula, Modeling with quadratics*

1. Complete the square for each quadratic function. Assuming the domain to be the entire real line, use your answer to find the range and to graph the function.
   (a) $f(x) = x^2 - 6x + 5$
   (b) $f(x) = -x^2 + x + 1$
   (c) $f(x) = -3x^2 + 8x + 5$

2. Revisit Example 1.13. For the quadratic function $y = f(x) = 2x^2 - 5x + 7$, produce a table of outputs given the inputs $x \in \{0, 1, 2, 3, 4, 5, 6\}$. What do you observe about Δy?

3. Find the implied domain of the function $f(x) = \frac{1}{\sqrt{4-7x+x^2}}$

4. (a) Find the equations of all quadratic polynomial functions which pass through the points $(1, 3)$ and $(2, 4)$.
   (b) More generally, if $P = (a, b)$ and $Q = (c, d)$ are given, and $c \neq a$, find all quadratic functions whose graphs pass through both P and Q.

5. Describe as best you can how the graph of the function $f(x) = 3x^2 + bx + 2$ depends on b.

6. Consider the frisbee problem (Example 1.14). Assume you're standing at the origin and that the frisbee is located at the point $(10, 15)$.

   (a) Find/describe all trajectories that result in the ball hitting the frisbee.
   (b) (Hard) Find a formula relating the initial speed v and initial slope m of the parabola (the initial speed/direction in which you throw the ball), and apply it to answer the following:
      i. If you throw the ball in such a way that the initial *vertical* speed of the ball is twice its *horizontal* speed, how fast must you throw the ball in order to hit the frisbee?
      ii. What is the *minimum* speed at which you could throw the ball if you want to dislodge the frisbee?

   (*Hint: You'll need some calculus! In the language of the original problem, the initial slope is* $m = \frac{r}{p}$ *and the initial speed* $v = \sqrt{p^2 + r^2}$*; why?*)

1.4 Polynomials, Factorization & the Rational Roots Theorem

Recall our simple example of factorization in the previous section

$$x^2+2x-15=(x-3)(x+5)=0 \iff x=3 \text{ or } x=-5$$

That this approach provides *all* roots relies on several familiar algebraic facts:

1. Factor Theorem: $f(c)=0 \iff x-c$ is a *factor* of $f(x)$.
2. No zero-divisors: $g(x)h(x)=0 \iff g(x)=0$ or $h(x)=0$.
3. A quadratic has *at most two* distinct roots.

We'll examine this more closely at the end of this section. For students first learning factorization, it isn't the *why* that's the challenge, it's the *how.* Multiplying out $(x-3)(x+5)$ is mechanical, but factorizing requires some creativity; we can't really factor without somehow knowing that 3 and -5 are roots! Beyond making a lucky guess, how might we go about this?

Example 1.15. Let's re-examine $f(x)=x^2+2x-15=0$ in a couple of stages.

Integer solutions The simplest type of root would be an *integer* n. If $f(n)=0$, observe that

$$n^2+2n-15=0 \implies n(n+2)=15 \implies 15 \text{ is divisible by } n$$

There are only *eight possible candidates* for n, and it doesn't take long to test them all:

n	1	-1	3	-3	5	-5	15	-15
$n+2$	3	1	5	-1	7	-3	17	-13

Rather than computing $f(n)$ explicitly, we listed all divisors of n in the first, the corresponding $n+2$ in the second, and mentally checked when $n(n+2)=15$. There are precisely two integer solutions, namely $n=3$ and $n=-5$.

Rational Solutions If you already believe that a quadratic polynomial has *at most* two solutions, then you're done. The next simplest possibility, however, is that a solution be a *rational number* $x=\frac{p}{q}$: we may assume this is in *simplest terms*.[5] Substituting into the polynomial, we see that

$$\frac{p^2}{q^2}+2\frac{p}{q}-15=0 \iff p^2+2pq-15q^2=0$$

Remembering that p,q are *integers*, we rearrange this equation in two ways:

$p(p+2q)=15q^2$ Since the left side is a multiple of p, so also is the *right.* Since p,q have no common factors, it follows that p divides into 15 (15 is a multiple of p).

$p^2=q(15q-2p)$ Since the right side is a multiple of q, so also is the *left.* Since p,q have no common factors, we conclude that $q=1$.

The upshot is that the only rational solutions to $f(x)=0$ are the two *integers* we've already found.

Definition 1.16. A *degree n polynomial*[6] is any function of the form

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

where the *coefficients* a_k are constants with $a_n \neq 0$.

Our analysis in Example 1.15 generalizes to a famous result.

Theorem 1.17 (Rational Roots). *Suppose $f(x) = a_n x^n + \cdots + a_0$ has* integer *coefficients where a_n and a_0 are non-zero. If $x = \frac{p}{q}$ is a rational root in simplest terms, then q divides into a_n and p into a_0. In particular, if $a_n = 1$, then the only possible rational roots are integers.*

While you are unlikely to teach the rational roots theorem to grade-school students *as written*, it is the reason why every trick or shortcut used to factorize works!

Proof. Substitute $\frac{p}{q}$ into $f(x)$ and multiply by q^n to obtain an equation where everything is an *integer*

$$\underbrace{a_n p^n + \overbrace{a_{n-1}p^{n-1}q + \cdots + a_1 pq^{n-1}}^{\text{divisible by } q}}_{\text{divisible by } p} + \overbrace{}^{} a_0 q^n = 0$$

By considering the braced terms we see that $a_n p^n$ is divisible by q and $a_0 q^n$ by p. Since p, q have no common factors, we obtain the result. ∎

Examples 1.18. We factorize two polynomials using the rational roots theorem.

1. If $x = \frac{p}{q}$ is a rational root of $f(x) = 2x^2 - x - 3$ in lowest terms, then $q = 1$ or 2 and $p = \pm 1$ or ± 3. The eight possibilities for x are easily checked:

x	1	-1	3	-3	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	$-\frac{3}{2}$
$2x-1$	1	-3	5	-7	0	-2	2	-4

   You may prefer to compute $f(x)$ directly: as in the previous example, since we already know x it is quicker to check whether $x(2x-1) = 3$ rather than $f(x) = 0$. The two roots are indicated; it is easily verified that the polynomial can be factorized

$$f(x) = (2x-3)(x+1)$$

2. If the cubic polynomial $f(x) = x^3 - 2x^2 + 5$ had any rational roots, the only possibilities would be ± 1 or ± 5. It is quickly verified that none of these work,

$$f(1) = 4, \quad f(-1) = 2, \quad f(5) = 80, \quad f(-5) = -170$$

   whence $f(x) = 0$ has no rational roots.

---

[5] I.e., $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ have no common factors: $\gcd(p,q) = 1$.

[6] A quadratic polynomial has degree 2 and a linear polynomial $mx + c$ degree one (if $m \neq 0$). A non-zero constant polynomial has degree zero. By convention, the *zero polynomial* $y \equiv 0$ has degree $-\infty$ so that the theorem $\deg fg = \deg f + \deg g$ makes sense for all polynomials.

Unless there are very few candidates, checking all possibilities is time-consuming. The rational roots theorem is typically used in conjunction with factorization by providing options for how to start.

Example 1.19. Given $f(x) = x^3 - x^2 - 7x + 10$, the rational roots theorem offers eight candidates: $x = \pm 1, \pm 2, \pm 5, \pm 10$. It is not difficult to check the first few in your head, e.g.

$$f(2) = 8 - 4 - 14 + 10 = 0$$

By the factor theorem, $x - 2$ is a factor of $f(x)$. The factorization can be performed in various ways. Here are three options: these, and every other method, are versions of the same process.

Long/synthetic division You should have practiced this in high-school.

$$
\begin{array}{rrrrr}
 & x^2 & +x & -5 & \\ \cline{2-5}
x-2\big) & x^3 & -x^2 & -7x & +10 \\
 & -x^3 & +2x^2 & & \\ \cline{2-3}
 & & x^2 & -7x & \\
 & & -x^2 & +2x & \\ \cline{3-4}
 & & & -5x & +10 \\
 & & & 5x & -10 \\ \cline{4-5}
 & & & & 0
\end{array}
\implies x^3 - x^2 - 7x + 10 = (x-2)(x^2 + x - 5)
$$

Multiply out and solve Write $f(x) = (x-2)q(x)$ where $q(x) = ax^2 + bx + c$ is some quadratic:

$$x^3 - x^2 - 7x + 10 = (x-2)(ax^2 + bx + c) = ax^3 + (b - 2a)x^2 + (c - 2b)x - 2c$$

Equating coefficients, we obtain the same factorization as before:

$$a = 1, \qquad b - 2a = -1 \implies b = 1, \qquad -2c = 10 \implies c = -5$$

Term-by-term factorization With practice, you can construct the required quadratic factor term-by-term as you go without showing any work. *Teaching* such an approach is likely a terrible idea unless your students are already very comfortable…

(a) To create x^3, the first term of the quadratic factor must be x^2

$$x^3 - x^2 - 7x + 10 = (x-2)(x^2 + \cdots) = x^3 - 2x^2 + \cdots$$

(b) We have $-2x^2$ but want $-x^2$. To correct this, add x to the quadratic ($x^2 - 2x^2 = -x^2$):

$$(x-2)(x^2 + x + \cdots) = x^3 - x^2 - 2x + \cdots$$

(c) We have $-2x$ but want $-7x$. To fix, subtract 5 from the quadratic ($-5x - 2x = -7x$):

$$(x-2)(x^2 + x - 5) = x^3 - x^2 - 7x + 10$$

(d) Since the last term 10 is correct, the factorization worked!

You might have seen other approaches involving arranging the coefficients in a table. Regardless, the calculations required to complete these methods are exactly those seen above; all these methods are versions of the same thing.

Why Does Factorization Work?

The theory of factorization relies on some algebra. Here is a *brief* treatment.

Theorem 1.20 (Factor Theorem). *Suppose $f(x)$ is a degree n polynomial. Then:*

1. $f(c) = 0$ *if and only if* $f(x) = (x - c)q(x)$ *for some (degree $n - 1$) polynomial* $q(x)$.
2. *The polynomial has* at most *n distinct roots.*

Proof. 1. $(\Leftarrow)$ This is essentially trivial: $f(x) = (x - c)q(x) \Longrightarrow f(c) = (c - c)q(c) = 0$.

$(\Rightarrow)$ This relies on the *division algorithm for polynomials*: if f, g are polynomials, then there are unique polynomials q, r with[7]

$$f(x) = g(x)q(x) + r(x) \quad \text{and} \quad \deg r < \deg g$$

If $g(x) = x - c$ is linear, $r(x)$ must be constant. Evaluate both sides at $x = c$ to obtain

$$f(x) = (x - c)q(x) + f(c) \qquad \qquad (\text{thus } f(c) = 0 \Longrightarrow f(x) = (x - c)q(x))$$

2. Suppose $c_1, \ldots, c_n$ are distinct real roots. By part 1, $f(x) = (x - c_1)q_1(x)$. Since

$$0 = f(c_2) = (c_2 - c_1)q_1(c_2) \implies q_1(c_2) = 0$$

we may factor $x - c_2$ from $q_1(x)$ to obtain

$$f(x) = (x - c_1)(x - c_2)q_2(x), \qquad \deg q_2 = n - 2$$

Repeat this process to factor out all n linear polynomials $x - c_k$:

$$f(x) = (x - c_1) \cdots (x - c_n)q_n, \qquad \deg q_n = n - n = 0$$

whence $q_n \neq 0$ is *constant*. Plainly

$$f(c) = (c - c_1) \cdots (c - c_n)q_n = 0 \Longrightarrow c = c_j$$

for some j, so there are no other roots. ■

---

[7] For a given example, q and r may be found by synthetic division. This is similar (and may be demonstrated similarly) to the more familiar division algorithm for integers: if m, n are integers, then there are unique integers q, r for which

$$m = qn + r \text{ and } 0 \le r < |n|$$

In elementary school, this is typically written $m \div n = q \,\mathrm{r}\, r$ (q remainder r); e.g., $23 \div 4 = 5 \,\mathrm{r}\, 3$ corresponds to $23 = 5 \times 4 + 3$.

Example (1.19 cont). We know that $f(x) = x^3 - x^2 - 7x + 10 = (x-2)(x^2+x-5)$. But then

$$f(x) = 0 \iff x - 2 = 0 \text{ or } x^2 + x - 5 = 0$$

The former gives the root $x = 2$, and the latter can be attacked via the quadratic formula or completing the square; the polynomial therefore has exactly three real roots

$$x = 2, \ \frac{-1 \pm \sqrt{21}}{2}$$

Example 1.21. We finish with a quick example of how long division (or any other factorization method as in Example 1.19) computes the ingredients in the division algorithm.
If $f(x) = x^3 + 7x^2 - 2$ and $g(x) = x^2 - 2$, then

$$\begin{array}{rrrrrl}
 & & x & & +7 & \implies x^3 + 7x^2 - 2 = (x^2-2)(x+7) + (2x+12) \\ \cline{2-5}
x^2 - 2 \big) & x^3 & +7x^2 & & -2 & \\
 & -x^3 & & +2x & & \\ \cline{2-4}
 & & 7x^2 & +2x & -2 & \\
 & & -7x^2 & & +14 & \\ \cline{3-5}
 & & & 2x & +12 &
\end{array}$$

Otherwise said,

$$f(x) = g(x)q(x) + r(x), \text{ where } q(x) = x + 7, \ r(x) = 2x + 12 \text{ and } \deg r = 1 < 2 = \deg g$$

Exercises 1.4. *Key concepts: Rational Root Theorem, Factorizing, Factor Theorem*

1. Apply the rational roots theorem to the polynomial $x^3 + 2x^2 - x - 2$ and use it to factorize the polynomial.
2. Repeat the previous question for the polynomial $6x^2 + x - 2$.
3. Use the rational roots theorem to prove that the polynomial $2x^5 - 3x + 7$ has no rational roots.
4. Factorize the polynomials and thereby find their (real) roots. Explain your steps carefully.
   (a) $f(x) = x^3 + 2x^2 - 3x$
   (b) $f(x) = x^4 - 13x^2 + 36$
   (c) $f(x) = x^3 - 7x - 6$
5. Factorize the polynomial $f(x) = x^6 - 2x^5 - x^4 - 4x^3 - 4x^2 - 4x - 6$, and thus demonstrate that it has exactly two real roots.

6. Students often follow a heuristic when trying to factorize a polynomial $f(x) = 0$:

   > Try some small integer values for x until you find a root, then apply long division.

   Discuss when this might or might not be a good idea. For what types of polynomial $f(x)$ will this approach work? When will it fail?

7. The polynomial $f(x) = 2x^4 - 3x^3 + 2x^2 + 3x - 9$ has only one rational root. Find it and factorize the polynomial as $f(x) = g(x)q(x)$ where $\deg g = 1$.

8. Find unique polynomials $q(x)$ and $r(x)$ for which $f(x) = g(x)q(x) + r(x)$ and $\deg r < \deg g$.
   (a) $f(x) = x^3 + 1$ and $g(x) = x + 2$.
   (b) $f(x) = x^4 + x^3 - 2$ and $g(x) = x^2 + 1$.

9. Let $f(x) = ax^3 + bx^2 + cx + d$ be a cubic polynomial. 'Complete the cube' by finding a constant k such that

   $$f(x) = a(x-k)^3 + p(x-k) + q$$

   has no $(x-k)^2$ term (here p, q are constants).

   (*Hint: evaluate* $f(x+k)$)

10. Suppose $\deg f = k$ and $\deg g = l$.
    (a) Show that $\deg(fg) = kl$.
    (b) Is it always the case that $\deg(f+g) = \max(k, l)$? Why/why not?

11. You are given a cubic polynomial $f(x) = x^3 + 7x^2 + px + q$ where p, q are (as yet) unknown. If $f(x) = 0$ has three real roots $x = 1, 2$ and c, find c and the coefficients p, q.

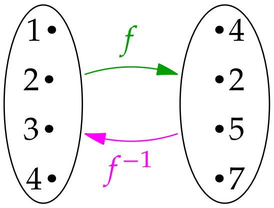
1.5 Inverse Functions & the Horizontal Line Test

The informal idea of an inverse function is that f^{-1} takes the *output* of f and returns its *input* (and vice versa).

Example 1.22. Define a simple function using a table or an arrow diagram

x	1	2	3	4
$f(x)$	4	2	5	7

y	4	2	5	7
$f^{-1}(y)$	1	2	3	4



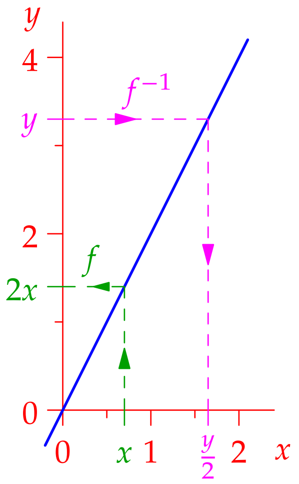
The inverse f^{-1} is the function obtained by *reversing the arrows* or flipping the table upside-down.

Definition 1.23. A function $f : A \to B$ is *invertible* if it has an *inverse*: a function $f^{-1} : B \to A$ which, for all possible values $x \in A$ and $y \in B$, satisfies

$$f^{-1}\big(f(x)\big) = x \quad \text{and} \quad f\big(f^{-1}(y)\big) = y \tag{$*$}$$

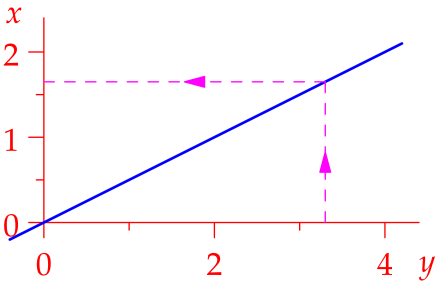
Certainly Example 1.22 satisfies the input–output properties $(*)$. Our concerns are identifying when a function is invertible, how to make it so if not, and how to compute an inverse.

Example 1.24. The simple linear function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x$ has inverse $f^{-1}(y) = \frac{y}{2}$. Its graph admits a similar interpretation to the arrow diagram.

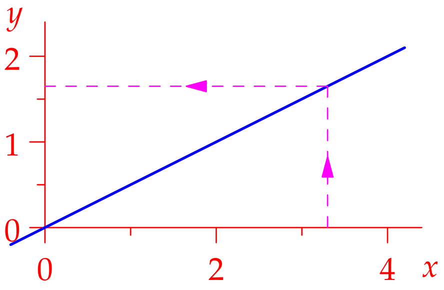


- The function takes an input x, moves it vertically to the graph, then projects horizontally to the y-axis. This interpretation is precisely the vertical line test (Definition 1.4)!
- The inverse function f^{-1} *reverses the arrows*: project a value on the y-axis horizontally to the graph, and then vertically to the x-axis.

It is unnecessary, but you might want to re-draw the graph of the inverse function so that its domain (inputs) are on the horizontal axis. It is entirely up to you whether to also switch the labels x and y.



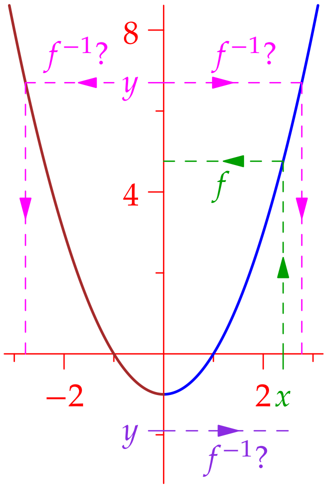
Axes reversed: $f^{-1}(y) = \frac{y}{2}$



Axes and labels reversed: $f^{-1}(x) = \frac{x}{2}$

Since we are interested in the question of *which* functions can be inverted, it is helpful to consider a function where this arrow-reversing procedure fails.

Example 1.25. Consider $f:\mathbb{R}\to\mathbb{R}$ where $f(x)=x^2-1$. This time, when we attempt to move a real number y horizontally to the graph to compute $x=f^{-1}(y)$, we (usually) encounter one of two problems:



(a) If $y>-1$, the horizontal line has two intersections with the graph, producing two choices of x.

(b) If $y<-1$, the horizontal line has no intersection with the graph, so there is no way to choose a suitable x.

The naïve approach of *reversing the arrows* fails to define an inverse. However, a simple remedy arises from staring at the graph:

- Problem (a) goes away if we delete the left half of the graph. Otherwise said, we *restrict the domain* of f to $[0,\infty)$.
- Problem (b) disappears if we insist that $y\ge -1$. Equivalently, we *restrict the codomain* of f to its *range* $[-1,\infty)$.

After making these restrictions so that $f:[0,\infty)\to[-1,\infty)$, we have a nicely defined inverse function. Indeed

$$f^{-1}:[-1,\infty)\to[0,\infty),\quad f^{-1}(y)=\sqrt{y+1}$$

is easily seen to satisfy the input–output conditions $(*)$:

$$\begin{aligned} x\in[0,\infty) &\implies f^{-1}\big(f(x)\big)=\sqrt{(x^2-1)+1}=\sqrt{x^2}=x \qquad \text{(since } x\ge 0\text{!!!)}\\ y\in[-1,\infty) &\implies f\big(f^{-1}(y)\big)=\big(\sqrt{y+1}\big)^2-1=y \end{aligned}$$

We could have made a different choice in part (a) by instead deleting the right side of the graph. This amounts to restricting the domain to $\operatorname{dom} f=(-\infty,0]$ and produces a *different inverse function*. We'll return to this example shortly. The takeaway at present should be the importance of paying attention to *domains* and *ranges*…

What makes a function invertible?

The fixes in the last example can be rephrased succinctly. A function is invertible precisely when it passes the…

Horizontal line test: every horizontal line intersects the graph exactly once

This unpacks to two conditions, each of which addresses one of the problems in the example.

Definition 1.26. Let $f : A \to B$ be a function. We say that f is:

(a) *One-to-one* (1–1) if distinct inputs $x_1 \neq x_2 \in A$ have distinct outputs $f(x_1) \neq f(x_2)$. Otherwise said, there is *at most* one way to achieve a given output:

$$f(x_1) = f(x_2) \implies x_1 = x_2$$

If A, B are sets of real numbers with A along the x-axis, this is the assertion that each horizontal line intersects the graph at most once.

(b) *Onto* if $\operatorname{range} f = B$. Equivalently,

Given $y \in B$, there is some input $x \in A$ for which $y = f(x)$

Again, if $A, B \subseteq \mathbb{R}$ with A along the x-axis, this says that the horizontal line through $y \in B$ intersects the graph at least once.

Putting these ideas together, a function $y = f(x)$ is both 1–1 and onto precisely when every output $y \in B$ corresponds to a *unique input* $x \in A$. In summary:

Theorem 1.27. *A function $f : A \to B$ is invertible if and only if it is both 1–1 and onto.*
In such a case, it inverse is the function $f^{-1} : B \to A$ such that $f^{-1}(y) = x$ whenever $y = f(x)$.
Moreover, $\operatorname{dom} f^{-1} = \operatorname{range} f = B$ *and* $\operatorname{range} f^{-1} = \operatorname{dom} f = A$.

Let us revisit our previous example in this context.

Example (1.25, mk. II). Consider the two properties in the context of the function $f : \mathbb{R} \to \mathbb{R}$ with formula $f(x) = x^2 - 1$:

(a) Suppose x_1, x_2 are real numbers, then

$$\begin{aligned} f(x_1) = f(x_2) &\implies x_1^2 - 1 = x_2^2 - 1 \implies x_1^2 = x_2^2 \\ &\implies x_1 = \pm x_2 \end{aligned}$$

To force f to be 1–1, it is enough to *restrict the domain* so that all x have the same sign: the natural choice is $\operatorname{dom} f = [0, \infty)$.

(b) Compute the range:

$$\operatorname{range} f = \{x^2 - 1 : x \in [0, \infty)\} = [-1, \infty)$$

We therefore force f to be onto by *restricting its codomain* to $[-1, \infty)$.

After making these restrictions, the function $f : [0, \infty) \to [-1, \infty)$ is invertible. Its inverse function is obtained by solving $y = x^2 - 1$ for x:

$$x^2 = y + 1 \implies x = f^{-1}(y) = \sqrt{y + 1}$$

The *non-negative square root* is used since $x \in \operatorname{dom} f = [0, \infty)$.

An algorithm for inverting functions

Our discussions suggest an algorithmic process for making a function $f : A \to B$ invertible and for computing an inverse.

1. Check that f is 1–1. If not, *restrict the domain* to make it so.
2. Check that f is onto. If not, *redefine* $B = \operatorname{range} f$.
3. *Solve* $y = f(x)$ for $x = f^{-1}(y)$.
4. (Optional) *Switch the symbols* x, y and write $y = f^{-1}(x)$.

The last optional step is because mathematical culture tends to prefer x as an input. If $A, B \subseteq \mathbb{R}$, switching $x \leftrightarrow y$ is equivalent to *reflecting the graph* in the line $y = x$. It is safest to do this step *last*, since you will likely have to consider domains in step 3...

Note that step 1 likely involves a *choice*; depending on how you restrict the domain, you might be able to find multiple inverse functions! To make this explicit, we return once more to our example.

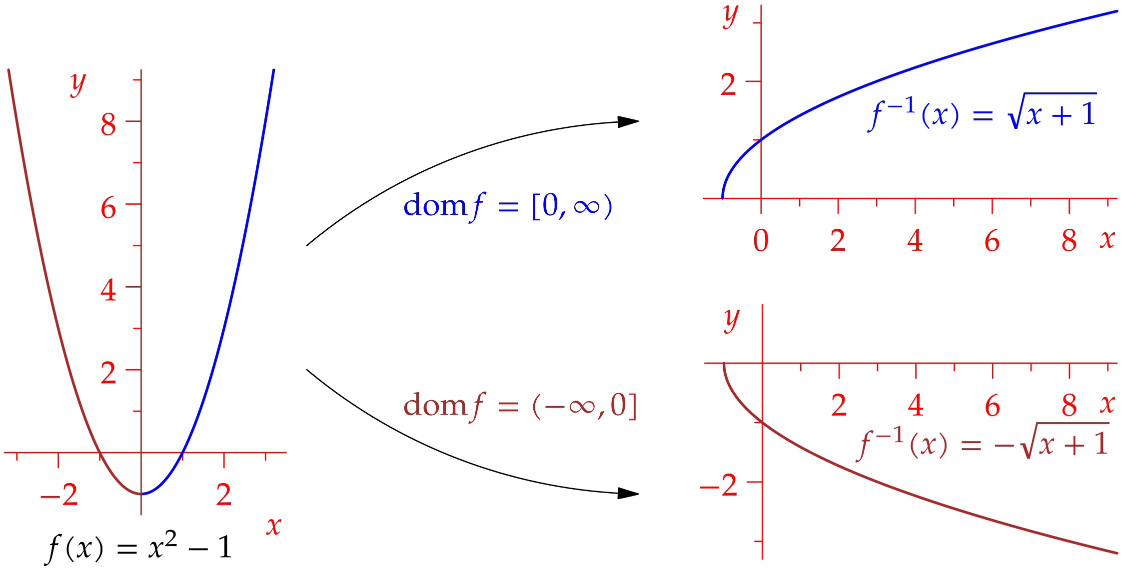
Example (1.25, mk. III). Recall that if $f(x) = x^2 - 1$, then

$$f(x_1) = f(x_2) \implies x_1 = \pm x_2$$

Instead of restricting the domain to $[0, \infty)$, we can instead force f to be 1–1 by taking the left half of the graph; that is by *choosing* $\operatorname{dom} f = (-\infty, 0]$. The range/codomain remains $[-1, \infty)$, but the inverse function is now different:

$$x^2 = y + 1 \implies x = -\sqrt{y+1} \in (-\infty, 0] = \operatorname{dom} f \implies f^{-1}(x) = -\sqrt{x+1}$$

The domain of f forces us to take the *negative square root*.



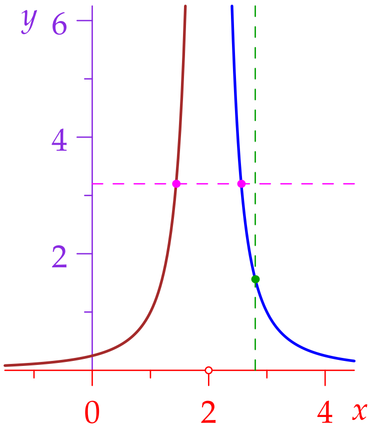
There are other domains on which f is 1–1, but these are the most natural choices. The moral is that you cannot invert a function unless you are precise about its domain and range!

We finish with an algebraically tougher example where you may feel that more detail is justified.

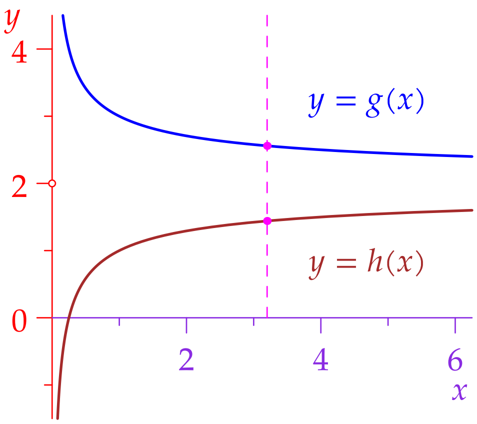
Example 1.28. We compute inverses for the function with formula $y = f(x) = \frac{1}{(x-2)^2}$.
The implied *domain* is $\mathbb{R} \setminus \{2\} = (-\infty, 2) \cup (2, \infty)$: all real numbers except 2.
The *vertical line test* is clearly visible on the graph: every vertical line through a point on the domain (except $x = 2$) intersects the graph exactly once.
The *range* is the interval $\mathbb{R}^+ = (0, \infty)$: since

$$f(x) = y \iff \frac{1}{x-2} = \pm\sqrt{y} \iff x = 2 \pm \frac{1}{\sqrt{y}} \qquad (*)$$

any positive output y may be obtained by evaluating $y = f\big(2 + \frac{1}{\sqrt{y}}\big)$.
The ± in $(*)$ is because f fails the *horizontal line test*: it isn't 1–1.



The function $f(x) = \frac{1}{(x-2)^2}$



Two inverse functions g and h

There are two natural choices for an inverse:

1. Choose $\operatorname{dom} f = (2, \infty)$. Then

$$\pm\sqrt{y} = \frac{1}{x-2} > 0$$

   so we take the *positive* square-root. The inverse function is therefore

$$g : (0, \infty) \to (2, \infty), \quad g(x) = 2 + \frac{1}{\sqrt{y}}$$

2. Choose $\operatorname{dom} f = (-\infty, 2)$. This time

$$\pm\sqrt{y} = \frac{1}{x-2} < 0$$

   We take the *negative* square-root to obtain a second inverse function

$$h : (0, \infty) \to (-\infty, 2), \quad h(x) = 2 - \frac{1}{\sqrt{y}}$$

Exercises 1.5. *Key concepts: Definition of an inverse, Visualizing inverses, Horizontal line test, 1–1 and onto, Domain and range restriction*

1. If $\operatorname{dom} f = \mathbb{R}$, check that $f(x) = x^3 + 8$ passes the horizontal line test. Find f^{-1}.
2. Consider $f(x) = x^2 + 2x - 3$. Similarly to Example 1.25, find *two inverses* of f.
3. Sketch the graph of the function

   $$f(x) = \begin{cases} x & \text{if } 0 \le x < 1 \\ x - 1 & \text{if } 1 \le x < 2 \\ x - 2 & \text{if } 2 \le x < 3 \end{cases}$$

   Find *three* domains on which f is 1–1 and thus compute three distinct inverses.
4. Sketch the graph of the function

   $$f : \mathbb{R} \to (\tfrac{3}{2}, \infty) \qquad f(x) = \begin{cases} 3 - \frac{1}{2}x & \text{if } x \le 2 \\ 2 - \frac{1}{x} & \text{if } x > 2 \end{cases}$$

   Show that f is 1–1 and onto, and compute f^{-1}.
5. (Hard) Find the implied domain and range of

   $$f(x) = \frac{x+1}{1 + \frac{1}{x+1}}$$

   Find an interval on which f is 1–1 and compute its inverse.
6. An astute student observes that Definition 1.23 only describes the properties satisfied by *an* inverse and asks why we keep referring to *the* inverse. How would you respond?