5 Functions

5.1 An informal definition of a function

Informally a function is a rule that assigns, to each of a set of possible inputs, an unambiguous output. Two running examples we'll use are:

Example 1 Given a real number, square it and subtract 1, and

Example 2 Add 1 to the input, subtract 1 from the input, multiply the two answers to get the output.

In **Example 1**, the input 7 leads unambiguously to the output 48, as does the input -7 (there's no rule that says that different inputs must lead to *different* outputs). In **Example 2**, the input 3 leads unambiguously to the output $(3+1)(3-1)$ or 8.

Functions can be much more complex than this; for example we might input a natural number n, and output the nth digit after the decimal point of π, *if* that digit happens to be odd; and output the n digit after the decimal point of $\sqrt{n}$ otherwise. It's not easy to calculate specific values of this function, but you will agree that it is unambiguous[68].

As an[69] example of an ambiguous function, consider the rule "for input a positive number x, output that number y such that $y^2 = x$". What is the output associated with input 4? We have no way of knowing from the rule whether it is intended to be $+2$ or -2, so this rule doesn't define a function.

Every function has a

- **Domain**: the set of all possible inputs,

and a

- **Range**: the set of all outputs, as the inputs run over the whole domain.

For **Example** 1 the domain is the set of all real numbers. The range is less obvious, but it shouldn't be too surprising to learn that it is the set of all reals that are at least -1, or $\{x : x \geq -1\}$.

For **Example** 2 the domain is unclear. But we have the following universally agreed upon

> **Convention**: If the domain of a function of real numbers is not specified, then the domain is taken to be the largest set of reals for which the rule makes sense (i.e., does not involve dividing by zero, taking the square roof of a negative number, or evaluating 0^0); this set is called the *natural domain* of the function.

[68] Or is it?

[69] Possible "another"; see footnote above!

Based on this convention, the domain for **Example 2** is the set of all real numbers. The range is again $\{x : x \geq -1\}$.

In general it is pretty easy to determine the natural domain of a function — just throw out from the reals all values where the rule define the function leads to problems — but usually the range is far from obvious. For example, it's pretty clear that the rule that sends x to $(x^2+1)/(x^2-1)$ (call this **Example 3**) has domain $\mathbb{R} - \{-1, 1\}$, but there is no clear reason why its range is $(-\infty, -1] \cup (1, \infty)$.

This last example, by the way, makes it clear that we need some better notation for functions than "the rule that ...". If we have a compact, easily expressible rule that determines a function, and we know the domain X and range Y of the function, there is a standard convention for expressing the function, namely

$$\begin{gathered} f : X \to Y \\ x \mapsto \text{whatever expression describes the rule in question.} \end{gathered}$$

For **Example 1** we might write

$$\begin{gathered} f : \mathbb{R} \to [-1, \infty) \\ x \mapsto x^2 - 1. \end{gathered}$$

When using this notation, we will also use "$f(x)$" to indicate the output associated with input x, so $f(7) = 48$ and $f(-1) = 0$. But of course we can also do this for generic input x, and write $f(x) = x^2 - 1$; and since this is enough to completely specify what the function does on every possible input, we will often present an expression like this as the definition of the particular function f.

This convention is particularly convenient when we are not specifying the domain of the function we are working with, but instead taking it to have its natural domain. So we might completely specify **Example 3** by writing

$$\text{“the function } \tilde{r}_7(x) = (x^2+1)/(x^2-1)\text{”.}$$

That fully pins down the function, since we can (easily) compute the domain and (with difficulty) compute the range. (I'm deliberately using a wacky name here, $\tilde{r}_7$ rather than the more conventional f, or g, or h, to highlight that the name of a function can be *anything*).

A problem with the above notation is that it involves knowing the range, which is often very difficult to compute. We get over this by introducing the notion of

- **Codomain**: any set that *includes* the set of all possible outputs, but is not necessarily equal to the set of all possible outputs.

We then extend the notation above: if we know the domain X of a function, and also know *a* codomain Y, we can write

$$\begin{gathered} f : X \to Y \\ x \mapsto \text{whatever expression describes the rule in question.} \end{gathered}$$

So **Example 2** could be written as

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto (x+1)(x-1). \end{aligned}$$

(Notice that when working with real numbers, we can *always* resort to a worst-case scenario and take all of $\mathbb{R}$ as a codomain).

Often the rule that defines a function is best expressed in pieces, as in

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } x \geq 0. \end{cases}$$

We've seen this before, for example with the absolute value function.

5.2 The formal definition of a function

Going back to **Example 1** and **Example 2**, we might ask, are they different functions? Given our informal definition, the answer has to be "yes". The two rules — "square and subtract 1", and "add 1, subtract 1, multiply results" — are *different* rules. But we really would like the two examples to lead to the *same* function — they both have the same domains, and, because $x^2 - 1 = (x+1)(x-1)$ for all reals, for any given input each function has the same output.

This highlights one shortcoming of the informal definition we've given of a function. Another shortcoming is that it is simply too vague; what exactly do we mean by a "rule"? And without a precise formation of what is and isn't a rule, can we do any mathematics with functions?

We now give the *formal* definition of a function, which is motivated by the fact that all that's needed to specify a function is the information of what the possible inputs are, and what output is (unambiguously) associated with each input.

- A **function** is a set of *ordered pairs* (pairs of the form (a, b), where the order in which a and b are written down matters), with the property that each a which appears as the first co-ordinate of a pair in the set, appears as the first co-ordinate of exactly *one* pair.

Think of a as a possible input, and b as the associated output. The last part of the definition is what specifies that to each possible input there is an *unambiguous* assignment of output.

As an example, the function whose domain is all integers between -2 and 2 inclusive, and which is informally described by the rule "square the input", would formally be

$$f = \{(-2, 4), (-1, 1), (0, 0), (1, 1), (2, 4)\}.$$

We write "$f(-1) = 1$" as shorthand for $(-1, 1) \in f$ ($(-1, 1)$ is one of the pairs that makes up f). With this formal definition, the functions in **Example 1** and **Example 2** become the same function, because the sets of pairs (a, b) in both functions is the same.

In the context of this formal definition, we can now formally define domain, range and codomain.

- The **domain** of a function f, written Domain(f), is the set of all first co-ordinates of pair in the function;
- the **range** of f, written Range(f), is the set of all second co-ordinates of pair in the function; and
- a (not "the" — it's not unique) **codomain** of f is any set that contains the range as a subset.

Notice that although you have probably long been used to using the notation "$f(x)$" as the name for a generic function, with this formal definition it ok (and in fact more correct) to just use "f". A function is a set of pairs, and the name we use for the set (a.k.a. the name for the function) doesn't need to, and indeed shouldn't, use a variable. The expression "$f(x)$" should be understood not as a stand-in for the function, but (informally) as the output of the function when the input is x and (formally) the second co-ordinate of that pair in f whose first co-ordinate is x, if there is such a pair.

Having said that, in the future we will frequently use informality like "the function $f(x) = 3x - 2$" to specify a function, rather than the formal but more cumbersome

$$f = \{(x, 3x - 2) : x \in \mathbb{R}\}.$$

5.3 Combining functions

If f, g, h, et cetera, are all real functions (meaning: functions whose domains and ranges are all subsets of the real numbers), we can combine them to form other functions.

Addition and subtraction Informally the function $f+g$ is specified by the rule $(f+g)(x) = f(x) + g(x)$. Of course, this only makes sense for those x for which both $f(x)$ and $g(x)$ make sense; that is, for those x which are in both the domain of f and the domain of g. Formally we define

$$f + g = \{(a, b + c) : (a, b) \in f, (a, c) \in g\},$$

and observe that

$$\text{Domain}(f + g) = \text{Domain}(f) \cap \text{Domain}(g).$$

Notice that $f + g$ really is a function. It's a set of ordered pairs certainly. And suppose that a is a first co-ordinate of some pair (a, d) in $f + g$. It's in the set because there a b with $(a, b) \in f$ and a c with (a, c) in g; but by the definition of function (applied to f and g) we know that b and c are unique, so d can only be $b + c$.

Informally $f - g$ is defined by $(f - g)(x) = f(x) - g(x)$. You should furnish the formal definition for yourself as an exercise, verify that $f - g$ is indeed a function, and verify that Domain$(f - g)$ = Domain$(f) \cap$ Domain(g) (so is the same as Domain$(f + g)$).

Notice that just like ordinary addition, addition of functions is commutative. This follows quickly from the commutativity of ordinary addition. We give the proof of this fact here; take it as a template for other, similarly straightforward facts that will be left as exercises.

Claim 5.1. *For any two real functions f and g, $f + g = g + f$.*

Proof: Suppose $(a, d) \in f + g$. That means there is a unique real b and a unique real c such that $(a, b) \in f$, $(a, c) \in g$, and $d = b + c$. But by commutativity of addition, we have $d = c + b$. This says that $(a, d) \in g + f$.

By the same reasoning, if $(a, d) \in g + f$ then $(a, d) \in f + g$. So as sets of ordered pairs, $f + g = g + f$. □

As a first exercise in similar manipulations, you should verify also that addition of real functions is associative.

Multiplication and division The product of two functions f, g is defined informally by $(fg)(x) = f(x)g(x)$, and formally by

$$fg = \{(a, bc) : (a, b) \in f, (a, c) \in g\}.$$

As with addition, $\text{Domain}(fg) = \text{Domain}(f) \cap \text{Domain}(g)$, and multiplication is commutative and associative. Moreover multiplication distributes across addition: $f(g + h) = fg + fh$.

We can also define the product of a function with a real number. If f is a function and c is a real number then cf is defined informally by $(cf)(x) = c(f(x))$, and formally by

$$cf = \{(a, cb) : (a, b) \in f\}.$$

(Notice that we never write fc — it's conventional to put the constant *in front* of the function name).

We can define $-f$ to mean $(-1)f$ (and easily check that this creates no clash with the previous use of "$-$" in the context of functions — $f + (-g) = f - g$).

Division of a function f by a function g is defined informally by $(f/g)(x) = f(x)/g(x)$, and formally by

$$f/g = \{(a, b/c) : (a, b) \in f, (a, c) \in g\}.$$

We have to be a little careful about the domain of f/g, as we not only have to consider whether f and g make sense at possible input x, but also whether the expression f/g makes sense (i.e., we have to make sure that we are not dividing by 0). We have

$$\text{Domain}(f/g) = (\text{Domain}(f) \cap \text{Domain}(g)) - \{x : (x, 0) \in g\},$$

that is, the domain of f/g is all things in the domain of both f and g, other than those things which get send to 0 by g.

Rational functions Two very important special functions are the

- **constant function**: $f(x) = 1$ for all x, formally $\{(x, 1) : x \in \mathbb{R}\}$,

and the

- **linear function**: $f(x) = x$ for all x, formally $\{(x, x) : x \in \mathbb{R}\}$,

both with domains all of $\mathbb{R}$.

Combining these two functions with repeated applications of addition, multiplication and multiplication by constants, we can form the family of

- **polynomial functions**: functions of the form $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a + 0$, where $a_0, \ldots, a_{n-1}$ are all real constants, and a_n is a non-zero constant.

Such a polynomial is said to have *degree* n, and the numbers $a_0, a_1, \ldots, a_n$ are said to be the *coefficients* of the polynomial. We will see a lot more of polynomials as the course progresses; for now we will just say that the domain of any polynomial all of $\mathbb{R}$.

Combining polynomial functions with division, we can form the family of

- **rational functions**: functions of the form $f(x) = P(x)/Q(x)$, where P and Q are polynomials, and Q is not the identically (or constantly) 0 function, $\{(x, 0) : x \in \mathbb{R}\}$.

The domain of a rational function P/Q is $\mathbb{R}$ minus all those places where Q is 0.

In the discussion above we've talked about the domains of the functions we have been building. In general it is *very difficult* to pin down ranges of functions. In fact, it's a theorem[70] that if the degree of a polynomial is an even number six or greater, and the coefficients of the polynomial are rational numbers, it is in general not possible to express the range of the polynomial using rational numbers together with addition, subtraction, multiplication, division and taking nth roots; and for a rational function P/Q, if the degree of Q is five or greater, it is in general not even possible to express the *domain* of the function succinctly!

We know that there are many more functions beyond polynomials and rational functions. Familiar examples include $\sqrt{\cdot}$, sin, log, and exp. These will be introduced formally as we go on. For now, we'll use them for examples, but won't use them in the proofs of any theorems.

5.4 Composition of functions

There's one more very important way of building new functions from old: *composition*. Informally, giving two functions f and g, the composition $f(g(x))$ means exactly what it says:

[70] A quite difficult one, using something called Galois theory.

first apply g to x, and then apply f to the result. As an example, suppose $f(x) = \sin x$ and $g(x) = x^2 + 1$. Then the composition would be $f(g(x)) = \sin(x^2 + 1)$.

Notice that unlike previous ways of combining functions,

composition is not commutative!!!.

Indeed, if you are familiar with the sin function then you will know that in the example above, since $g(f(x)) = (\sin x)^2+1$ and this is definitely a different function from $f(g(x)) = \sin(x^2+1)$, we have an example already of a pair of function f, g for which $f(g(x)) \neq g(f(x))$, in general. For a more prosaic example, consider $a(x) = x^2$ and $b(x) = x + 1$; we have

$$a(b(x)) = (x + 1)^2 = x^2 + 2x + 1 \neq x^2 + 1 = b(a(x)),$$

then inequality in the middle being witnessed by any x other than $x = 0$.

Because composition is not commutative, we have to be very careful with the informal language we use to describe composition. By convention, "f composed with g (applied to x)" means "$f(g(x))$". Notice that in this convention there is an inherent order among the functions: "f composed with g" means something quite different from "g composed with f". It is sometimes tempting to use language like "the composition of f and g", but *this is ambiguous, and should be avoided!.*

Along with the language "f composed with g", it's also common to see "f of g" and "f after g". Both of these last two an inherent order, and the last is particularly suggestive: if f is *after* g, then the action of g gets performed *first*.

What is the domain of the composition of f with g? The composed function makes sense exactly for those elements of the domain of g, for which the outputs of g are themselves in the domain of f. Consider, for example, the function given by the rule that x maps to $\sqrt{(x + 1)/(x - 1)}$. This is the composition of the square root function (call it sq), with the function (call it f) that maps x to $(x+1)/(x-1)$. Now the domain of f is all reals except 1; but since the domain of sq is non-negative numbers, the domain of the composition is exactly those real x that are not 1, and that have $(x + 1)/(x - 1) \geq 0$. It's an easy exercise that $(x + 1)/(x - 1) \geq 0$ precisely when either $x \leq -1$ or $x > 1$; so the domain of the composition is $\{x : x \leq -1 \text{ or } x > 1\}$, which we can also write as $(-\infty, -1] \cup (1, \infty)$.

Formally, we use the notation "$\circ$" (read "composed with", "after") to indicate composition:

$$f \circ g = \{(a, c) : (a, b) \in g \text{ for some } b,\ (b, c) \in f\}$$

Although composition is not commutative, it is associative; proving this is just a matter of unpacking the definition:

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

while

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

so indeed $(f \circ (g \circ h))(x) = ((f \circ g) \circ h)(x)$ for every x. To finish we just need to check that the domains of $f \circ (g \circ h)$ and $(f \circ g) \circ h$ are the same; but it's easy to check that x is in the domain of $f \circ (g \circ h)$ exactly when

- x is in the domain of h,
- $h(x)$ is in the domain of g, and
- $g(h(x))$ is in the domain of f,

and these are also exactly the conditions under which x is in the domain of $(f \circ g) \circ h$.

5.5 Graphs

Note: I haven't included any pictures in my first pass through this section. I *strongly* encourage you to read this section with desmos open on a browser, so that you can create pictures as you go along. Spivak (Chapter 4) covers the same material, and has plenty of pictures.

In this section we talk about representing functions as graphs. It's important to point out from the start, though, that a graphical representation of a function should only ever be used as an aid to thinking about a function, and to provide intuition; considerations of graphs should *never* serve as part of a proof. The example of $f(x) = \sin(1/x)$ below gives an illustration of why not, as does the graph of Dirichlet's function (again, see below).

To start thinking about graphs, first recall the real number line, a graphical illustration of the real numbers. The line is usually drawn horizontally, with an arbitrary spot marked in the center representing 0, and an arbitrary spot marked to the right of 0, representing 1. This two marks define a unit distance — the length of the line segment joining them. Relative to this unit distance, the positive number x is represented by the spot a distance x to the right of 0, while the negative number x' is represented by the spot a distance x' to the left of 0. In this way all real numbers are represented by exactly on point on the number line (assuming the line is extended arbitrarily far in each direction), and the relation "$a < b$" translates to "a is to the left of b" on the line.

Recall that after introducing the absolute value function, we commented that the (positive) number $|a - b|$ encodes a notion of the "distance" between a and b. This interpretation of absolute value makes it quite easy to represent on the number line solutions to inequalities involving absolute value. For example:

- the set of x satisfying $|x - 7| < 3$ is the set of x whose distance from 7 is at most 3; that is, the set of x which on the number line are no more than (and not exactly) 3 units above 7 and no less than (and not exactly) 3 units below 7; that is, the *open* interval of numbers between $7 - 3$ and $7 + 3$ ("open" meaning that the end-points are not in the interval); that is, the interval $(4, 10)$; and, more generally

- for fixed real x_0 and fixed $\delta > 0$, $\{x : |x - x_0| < \delta\} = (x_0 - \delta, x_0 + \delta)$.

This general example will play a major role in the most important definition of the semester, the definition of a *limit* (coming up soon).

Now we move on to graphing functions. The *coordinate plane* consists of two copies of the number line, called *axes* (singular: *axis*), perpendicular to each other, with the point of intersection of the lines (the *origin* of the plane) being the 0 point for both axes. Traditionally one of the axes is horizontal (the "x-axis"), with the right-hand direction being positive, and the other is vertical (the "y-axis"), with the upward direction being positive. It's also traditional for the location of 1 on the x-axis to be the same distance from the origin as the distance from 1 to the origin along the y-axis.

A point on the co-ordinate plane represents an *ordered pair* of numbers (a, b), with a (the "x-coordinate") being the perpendicular distance from the point to the y-axis, and b (the "y-coordinate") being the perpendicular distance from the point to the x-axis. In the other direction, each ordered pair (a, b) has associated with a unique point in the coordinate plane: to get to that point from the origin, travel a units along the x-axis (so to the right if a is positive, and to the left if a is negative), and then travel b units in a direction parallel to the y-axis (so up if b is positive, and down if b is negative).

(Some notation:

- the *first quadrant* of the coordinate plane is the top right sector consisting of points (a, b) with a, b positive;
- the *second quadrant* is the top left sector consisting of points (a, b) with a negative, b positive;
- the *third quadrant* is the bottom left sector consisting of points (a, b) with a, b negative; and
- the *fourth quadrant* is the bottom right sector consisting of points (a, b) with a positive, b negative.)

Since functions are (formally) nothing more or less than ordered pairs of numbers, the coordinate plane should be an ideal tool for representing them. Formally, the *graph* of a function is precisely the set of points on the coordinate plane that represent the pairs that make up the function. Informally, we think of the graph as encode the output for every input — to see the output associated with input x, travel x units along the x-axis, then move parallel to the y-axis until the graph is hit. Notice:

- if the graph is not hit by the line parallel to the y-axis, that passes through the point at distance x from the origin along the x-axis, then we can conclude that x is not in the domain of the function;

- the line parallel to the y-axis may need to be scanned in both directions (up and down) to find the graph; if one has to scan up, then the function is positive at x, and if one has to scan down, then it's negative at x; and
- if the line parallel to the y-axis hits the graph, it must hit it at a *single* point; otherwise the output of the function at input x is ambiguous. This leads to the
  - **Vertical line test**: A collection of points in the coordinate plane is the graph of a function, if and only if every vertical line in the plane (line parallel to the y-axis) meets the collection of points *at most once*.

A graph can only provide an imperfect representation of a function of the reals, at least if the function has infinitely many points in its domain, because we can only every plot finitely many points. Even the slickest computer, that renders lovely smooth images of graphs, is only actually displaying finitely many points — after all, there are only finitely many pixels on a screen. Except for the very simplest of graphs (e.g. straight line graphs) the best we can ever do is to plot a bunch of points, and make our best guess as to how to interpolate between the points. We can never be *certain*, just from looking at the graph, that weird things don't in fact happen in the places where we have interpolated. This is the main reason why we won't use graphs to reason about functions (but as we'll see in a while, there are other reasons).

Nonetheless, it behooves us to be familiar with the graphs of at least some of the very basic functions, and how these graphs change as the function changes slightly. The best way to become familiar with the shapes of graphs, is to draw lots of them.

The tool that I recommend for drawing graphs is `desmos.com`. After you hit "Start Graphing", you can enter a function in the box on the left, in the form

$$\text{“}f(x) = \text{something to do with } x''$$

(e.g., $f(x) = x^2 - 3\sqrt{x}$). The graph of the function (or at least, a good approximation to it) will appear on the right, where you can zoom in or out, and/or move to different parts of the graph. You can enter multiple functions (just give then different names), and they will helpfully appear in different colors (the color of the graph on the right matching the color of the text specifying the function on the left). This allows you to compare the graphs of different functions.

You can enter variables into the specification of a function, and you be able to create a "slider" that lets you change the specific value assigned to the variable. For example, entering "$f(x) = ax^2 + bx + c$" and creating sliders for each of a, b, c, allows you to explore how the graph of the general quadratic equation changes as the coefficients change.

In these notes, I won't go over all the graphs that might be of interest to us, and laboriously describe their properties. That would be pointless, mainly because (at the risk of beating a dead horse) *we will never use our understanding of a graph to prove something; we will only use it to aid intuition.* Instead, I invite you to go to desmos.com, and explore these families, discovering their properties for yourself. I'll provide a list of suggested functions, and leading questions (with some answers):

- The constant function, $f(x) = c$ for real c. What happens to the graph as c changes?

- The linear function through the origin, $f(x) = mx$ for real m. What happens to the graph as m changes? What's the difference between positive m and negative m? Do all straight lines through the origin occur, as m varies over the reals? (The answer to this last question is "no". Think about the vertical line test).

- The linear function, $f(x) = mx + c$ for real m, c. What happens to the graph as c changes?

  Evidently, the graphs of the linear functions are straight lines. The number m is the *slope* of the line. It measures the ratio of the change in the y-coordinate brought about by change in the x-coordinate: if x is changed to $x + \Delta x$ (change Δx) then the output changes from mx to $m(x+\Delta x)$ for change $m\Delta x$, leading to ratio $m\Delta x/\Delta x = m$. Notice that this is *independent* of x — the linear functions are the only functions with this property, that the ration of the change in the y-coordinate to change in x-coordinate is independent of the particular x that one is at.

  This leads to an easy way to calculate the slope of a line, given two points (x_0, y_0) and (x_1, y_1) on the line: just calculate the ratio of change in y-coordinate to change in x-coordinate as one moves between these points, to get

$$m = \frac{y_1 - y_0}{x_1 - x_0}.$$

  And it also gives an easy way to calculate the precise equation of a line, given two (different) points (x_0, y_0) and (x_1, y_1): since the slope is independent of the x-value, consider a generic point $(x, f(x))$ on the line, and equate the calculations of the slope using the pair $(x, f(x))$, (x_0, y_0) and the pair $(x, f(x))$, (x_1, y_1), to get

$$\frac{f(x) - y_0}{x - x_0} = \frac{f(x) - y_1}{x - x_1},$$

  then solve for $f(x)$.

- The quadratic function, $f(x) = ax^2 + bx + c$ for real a, b, c. What is the general shape of the graph? What happens to the graph as a, b, c change? In particular, how does the *sign* of a (whether it is positive or negative) affect the shape?

  The shape of the graph of a quadratic function is referred to as a parabola. Parabolas have very clean geometric interpretations:

  - a *parabola* is the set of points in the coordinate plane that are equidistant from a fixed line and a fixed point.

We illustrate by considering the horizontal line $f(x) = r$, and the point (s,t), where we'll assume $t \neq r$. The (perpendicular, shortest) distance from a point (x,y) to the line $f(x) = r$ is $|y - r|$, and the (straight line) distance from (x,y) to (s,t) is, by the Pythagorean theorem

$$\sqrt{(x-s)^2 + (y-t)^2}.$$ [71]

So the points (x,y) that are equidistant from the line and the point are exactly those that satisfy

$$|y - r| = \sqrt{(x-s)^2 + (y-t)^2}$$

which, because both sides are positive, is equivalent to

$$(y-r)^2 = (x-s)^2 + (y-t)^2$$

or

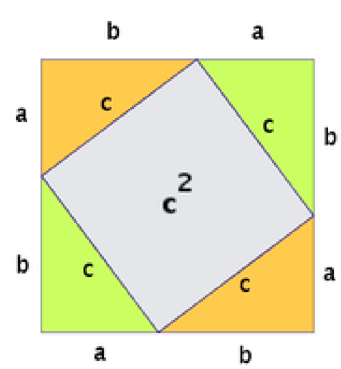
$$y = \frac{x^2}{2(t-r)} - \frac{sx}{(t-r)} + \frac{s^2 + t^2 - r^2}{2(t-r)},$$

so the graph of the set of points is the graph of a specific quadratic equation.

Of course, there are far more parabolas than graphs of quadratic equations: by drawing some lines and points in the plane, and roughly sketching the associated parabolas, you will quickly see that a parabola is only the graph of a quadratic (that is, only passes the vertical line test) if the line happens to be parallel to the x-axis.

- The general polynomial, $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ for real a_i's, $a_n \neq 0$. What is the general shape? In particular, what happens to the graph for very large x

[71] Why is this a reasonable formula for the straight-line distance between (x,y) to (s,t)? This distance formula comes quickly from the Pythagorean theorem, which says that if the hypotenuse of a right-angled triangle has length c, and the other two side lengths are a and b, then $a^2 + b^2 = c^2$. But why is *this* true? There are *many* proofs, going back to Euclid, about 300BC. My favorite proof is conveyed succinctly in the following picture (taken from https://math.stackexchange.com/questions/563359/is-there-a-dissection-proof-of-the-pythagorean-theorem-for-tetrahedra):



The area of the big square is $(a+b)^2$; but it is also $c^2 + 4((1/2)ab)$. So $(a+b)^2 = c^2 + 4((1/2)ab)$, or $a^2 + b^2 = c^2$.

and very small x (i.e., very large negative x), and how does that depend on n and a_n? How many "turns" does the graph have, and how does change as n changes?

We will be able to answer these questions fairly precisely, once we have developed the notion of the derivative.

More generally one may ask,

- How does the graph of $f(cx)$ related to the graph of $f(x)$, for constant c? What's the different between positive and negative c here?
- What about the graph of $cf(x)$?
- and $f(x+c)$?
- and $f(x)+c$?

and one may explore the answers to these questions by plotting various graphs, and seeing what happens as the various changes are made.

One important graph that is not the graph of a function is that of a circle. Geometrically, a *circle* is the set of all points at a fixed distance r (the *radius*) from a given point (a,b) (the *center*), and algebraically the circle is set of all points (x,y) in the coordinate plane satisfying

$$(x-a)^2+(y-b)^2=r^2$$

(using the Pythagorean theorem to compute distance between two points). A circle that will be of special interest to us is the *unit circle centered at the origin*, given algebraically by

$$x^2+y^2=1.$$

The circle is not the graph of a function, because it fails the vertical line test. A circle can be represented as the union of *two* functions, namely

$$f(x)=\sqrt{r^2-(x-a)^2}+b, \quad x\in[a-r,a+r]$$

and

$$f(x)=-\sqrt{r^2-(x-a)^2}+b, \quad x\in[a-r,a+r].$$

Related to the circle is the *ellipse*, a "squashed" circle, which geometrically is the set of all points, the sum of whose distances to two fixed points is a given fixed constant (so when the two points coincide, the ellipse becomes a circle). One also sometimes encounters the *hyperbola*, the set of all points the difference of whose distance from two points is the same. Circles, ellipses, parabolas and hyperbola are all examples of *conic sections*, shapes beloved of ancient mathematicians. In a modern calculus course like the present one, we will not have any need for conic sections, but if you interested there is a chapter in Spivak on the topic.

Two important functions that we will use for examples are the trigonometric functions sin and cos. We'll give a provisional definition here; it won't be until the spring semester, when we have studied the derivative, that we will give a precise definition.

Provisional definition of $\sin$ **and** $\cos$ The points reached on unit circle centered at the origin, starting from $(1,0)$, after traveling a distance θ, measured counter-clockwise, is $(\cos\theta, \sin\theta)$.

The domain of point cos and sin is all of $\mathbb{R}$, since one can travel any distance along the circle. Negative distances are interpreted to mean clockwise travel, and distance greater than 2π (the circumference of the circle) simply traverse the circle many times.

Let's watch the trajectory of cos, as a point travels around the circle:

- at $\theta = 0$, we are at $(1,0)$, and so $\cos 0 = 1$;
- as θ increases from 0 to $\pi/2$ (a quarter of the circle), we go from $(1,0)$ to $(0,1)$, with decreasing x-coordinate, and so $\cos\theta$ decreases from 1 to 0 as θ increases from 0 to $\pi/2$, and $\cos\pi/2 = 0$;
- as θ increases from $\pi/2$ to π, we go from $(0,1)$ to $(-1,0)$, with decreasing x-coordinate, and so $\cos\theta$ decreases from 0 to -1 as θ increases from $\pi/2$ to π, and $\cos\pi = -1$;
- as θ increases from π to $3\pi/2$, we go from $(-1,0)$ to $(0,-1)$, with increasing x-coordinate, and so $\cos\theta$ increases from -1 to 0 as θ increases from π to $3\pi/2$, and $\cos 3\pi/2 = 0$;
- as θ increases from $3\pi/2$ to 2π, we go from $(0,-1)$ to $(1,0)$, with increasing x-coordinate, and so $\cos\theta$ increases from 0 to 1 as θ increases from $3\pi/2$ to 2π, and $\cos 2\pi = 1$.

This gives the familiar graph of cos on the interval $[0, 2\pi]$, and of course, since we are back where we started after traveling fully around the circle, the graph just periodically repeats itself from here on.

Going the other direction, as θ decreases from 0 to $-\pi/2$ (a quarter of the circle, clockwise), we go from $(1,0)$ to $(0,-1)$, with decreasing x-coordinate, and so $\cos\theta$ decreases from 1 to 0 as θ decreases from 0 to $-\pi/2$, and $\cos -\pi/2 = 0$, and continuing in this manner we see the graph also extends periodically on the negative side of the y-axis.

We can play the same game with sin, and discover that this provisional definition[72] yields the expected periodic graph there, too.

The sin function, suitably modified, gives us a ready example of a function whose behavior cannot be understand fully using a graph. Consider $f(x) = \sin(1/x)$ (on domain $\mathbb{R} - \{0\}$) (formally, the composition of sin with the function that takes reciprocal). Just like sin, this is a function that oscillates, but unlike sin the oscillations are not of length (2π in the case of sin. As x comes from infinity to $1/(2\pi)$, $1/x$ goes from 0 to 2π, so f has one oscillation in that (infinite) interval. Then, as x moves down from $1/(2\pi)$ to $1/(4\pi)$, $1/x$ goes from 2π to 4π, so f has another oscillation in that (finite) interval. The next oscillation happens in the shorter finite interval as x moves down from $1/(4\pi)$ to $1/(6\pi)$; the next in the even

[72]Why is this a *provisional* definition? Because it requires understanding length along the curved arc of a circle. To make the notion of length along a curve precise, we need to first study the integral.

shorter interval as x moves down from $1/(6\pi)$ to $1/(8\pi)$. As x gets closer to 0, the oscillations happen faster and faster, until they get to a point where each oscillation is happening in an interval that is shorter than the resolution of the graphing device. Go ahead and graph $f(x) = \sin(1/x)$ on Desmos, and see what happens (in particular as you zoom in to $(0, 0)$). This should convince you that a graph is not always a useful tool to understand a function.

Another function that illustrates the limitations of graphing is *Dirichlet's function*:

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Because the rationals are "dense" in the reals — there are rationals arbitrarily close to any real — and the irrationals are also dense, any attempt at a graph of f is going to end up looking like two parallel straight lines, one along the x-axis (corresponding to the irrational inputs) and the other one unit higher (corresponding to the rational inputs), and this is certainly a picture that fails the vertical line test.

Going back to $f(x) = \sin(1/x)$, let's consider a related function, $g(x) = x\sin(1/x)$ (again on domain $\mathbb{R} = \{0\}$). Again this has oscillations that get arbitrarily close together as x gets close to 0, but now these oscillations are "pinched" by the lines $y = x$ and $y = -x$, so as we get closer to zero, the amplitudes of the oscillations (difference between highest and lowest point reached) get smaller and smaller. We will soon discuss the significant difference between f and g in their behavior close to 0. For now, let's ask the question

> how do f and g behave for very large positive inputs?

It's not hard to see that f should be getting closer to 0 as the input x gets larger — for large x, $1/x$ is close to 0 and $\sin 0 = 0$. It's less clear what happens to g. The $\sin(1/x)$ part is going to 0, while the x part is going to infinity. What happens when these two parts are multiplied together?

- Is the x part going to infinity faster than the $\sin(1/x)$ part is going to 0, leading to the product g going to infinity?
- Or is the x part going to infinity slower than the $\sin(1/x)$ part is going to 0, leading to the product g going to zero?
- Or are they both going to their respective limits at roughly the same rate, so that in the product they balance each other out, and g gets closer to some fixed number?
- Or is g oscillating as x grows, not moving towards some limit?

A look at the graph of g on a graphing calculator suggests the answer. To mathematically pin down the answer, we need to introduce a concept that is central to calculus, and has been central to a large portion of mathematics for the last 200 years, namely the concept of a limit.