Lecture 1: algebra and functions

Calculus I, section 10
September 6, 2022

1. Syllabus

The syllabus can be found here. We'll take some time to go over it in class.

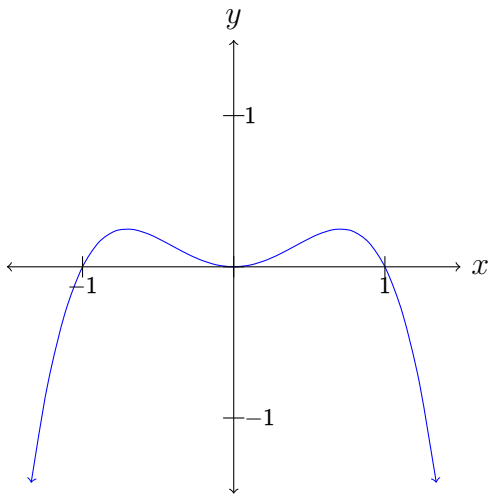
For Thursday, there is a survey posted on the course website. **If you intend to take this class, please fill it out.** You can turn it in either on the website, on paper in person, or by emailing me; if you are currently on the waitlist, you'll probably need to do one of the latter options.

2. Overview

Before we dive into algebra, let's start by giving an overview of what we'll talk about in this course.

We're going to introduce a lot of complicated machinery, and whenever we do that there should be a good justification for it. For calculus, there is a pretty strong justification: there are many problems (which can be *stated* without calculus) which can only be solved using calculus, or at least which can be solved much more easily that way.

For example, consider the graph $y = x^2 - x^4$, which looks like this:



What is the highest point on this graph? In other words, what is the maximum of the function $f(x) = x^2 - x^4$, and where does it occur?

Without calculus, this is a very tricky problem; I am not sure it is even possible (although you are welcome to try). With calculus, we'll see that once we develop enough machinery it becomes doable, and in fact you'll solve this problem later in the course.

This is a neat problem, but it doesn't have any obvious relevance to anything in particular, and although we'll end up using calculus to solve it it's not clear where that comes in. So

let's give another example, this time drawn from physics: I'm holding a ball four feet above the ground, and throw it upwards at 48 feet per second (that's about 33 miles per hour, or a little under 15 meters per second). Physics tells us (and later in the course we can even see how to derive this) that the height of the ball after t seconds is $y = f(t) = -16t^2 + 48t + 4$. There are a few questions we could ask here:

(a) After how many seconds does the ball hit the ground?

This is an *algebra* question. We can solve it by setting $y = 0 = -16t^2 + 48t + 4$. From there, we can apply the quadratic formula to get $t = \frac{-48 \pm \sqrt{48^2 - 4 \cdot 4 \cdot (-16)}}{2 \cdot (-16)}$; we could make our lives a little bit easier by noticing that the right-hand side is divisible by 4, so if we divide both sides by 4 first we instead have $-4t^2 + 12t + 1 = 0$, and so the quadratic formula is just

$$t = \frac{-12 \pm \sqrt{12^2 - 4 \cdot (-4)}}{-8} = \frac{-12 \pm \sqrt{160}}{-8} = \frac{3 \pm \sqrt{10}}{2}.$$

Since $\sqrt{10} > 3$ and we need the answer to be positive, the only possibility is $t = \frac{3+\sqrt{10}}{2} \approx 3.081$.

(b) At what point is the ball highest? How high does it go?

This is similar to the question above, but in this case we can actually solve it using algebra. (We could also solve it with calculus, but we don't know how yet.) How? Well, if we were to allow negative time, we would have two solutions to the question above, $t = \frac{3 \pm \sqrt{10}}{2}$. The point at which the ball is highest is between these two solutions, so we take the average: $\frac{3-\sqrt{10}}{2} + \frac{3+\sqrt{10}}{2}$, and we see that the $\frac{\sqrt{10}}{2}$ terms vanish and we're left with just $\frac{3}{2}$, so the ball is highest after $\frac{3}{2} = 1.5$ seconds. Plugging that into our equation for height, we get that the ball reaches a height of 40 feet.

(c) What is the average speed of the ball over those first 1.5 seconds?

We're moving towards a calculus question, but we can still answer this concretely using algebra. The ball reaches a height of 40 feet after 1.5 seconds, and it started at a height of 4 feet, so it moved 36 feet total in 1.5 seconds; that means that on average, it moved at a rate of $\frac{36}{3/2} = \frac{36}{3} \cdot 2 = 24$ feet per second.

(d) What was the average velocity of the ball between $t = 1$ and $t = 1.5$?

This is very similar: at $t = 1$, the ball was at 36 feet, and at $t = 1.5$ we've already computed that it was at 40 feet, so it rises 4 feet over .5 seconds for an average speed of 8 feet per second.

(e) What is the exact speed of the ball at $t = 1$?

This is a true calculus question! How can we solve it?

Well, we already have an approximation: we know that the *average* velocity between $t = 1$ and $t = 1.5$ is 8 feet per second. For a better approximation, we could take an average over a smaller range: how about the average velocity between $t = 1$ and $t = 1.1$?

We can compute (I won't make you do computations this bad, for the most part) that at $t = 1.1$ the ball is at height 37.44, so over that .1 seconds it moves 1.44 feet for an average speed of 14.4 feet per second. This is much larger than the 8 feet per second we got earlier, suggesting the true value may be even larger.

Let's try a much closer approximation: $t = 1.01$. In this case it turns out that the height of the ball at t is 36.1584, so the ball travels 0.1584 feet over that 0.01 seconds, for an average speed of 15.84 feet per second. This is larger than our previous approximation, but not by too much, so maybe we're getting close.

Let's try again with $t = 1.001$. This gives a height of 36.015984, so the ball moves 0.015984 feet over 0.001 seconds for an average of 15.984 feet per second. Again, this is a little bit larger, but very little, so it really seems like we're getting close. We can keep going like this: at $t = 1.0001$, we get an average velocity of 15.9984 feet per second; at $t = 1.00001$, we get 15.99984 feet per second; and so on. From this, we might guess that the answer is 16 feet per second; but we've only done approximations, we (so far) have no way of saying exactly what the answer is.

Let's take a moment to think about what we're doing. For each value of t (e.g. $t = 1.01$), we're evaluating our function $f(t)$ at two points, at t and at 1, taking the difference, and dividing by the difference between t and 1, i.e.

$$\frac{f(t) - f(1)}{t - 1}.$$

What we then want to do is look at the value of this expression as t gets closer and closer to 1; but we can't just plug in $t = 1$, or we'd get zero in the denominator.

What we need is called a *limit*: our answer is the limit of this expression as t approaches 1. We'll talk about these in a lot more detail, but in this example it just means the value that all our approximations are approaching: as t gets closer and closer to 1, this expression gets closer and closer to 16.

This expression is not an arbitrary one, either. The process we did above, where we take this expression for the "average velocity" and take the limit, is a very general process one can do to many different kinds of functions, and in many ways is the basis of all of calculus: it is called the derivative. We'll also talk about it much more, and there are many ways of computing it exactly, rather than just via approximations like we did above.

We've now seen some of the fundamental ideas of this course. In the first unit, we'll study limits; in order to do so, we first need a really solid understanding of functions, which are the most fundamental objects in this kind of math, and really the thing that we care about and that we're doing all this to understand. In the second unit, we'll apply our understanding of limits to study derivatives, which will give us a whole new way to understand functions and are our key tool in this class. Next, we'll look at applications of derivatives, such as the optimization problem we looked at above; and finally we'll look at a new kind of construction,

called an integral, which is first defined in an entirely separate way using limits but turns out to be closely connected to derivatives.

3. Functions

The first concept of calculus is the function. A function is a machine which takes in one piece of information and spits out another. For us, these will usually be numbers, but in principle they can be anything. For example, we could have a function f which takes in words and spits out their first character, so for example $f(\text{elephant}) = \text{e}$. For an example more like those we'll usually encounter, $f(x) = x^2$ takes in a number x and spits out its square, so for example $f(-1) = (-1)^2 = 1$.

Formally, we write a function as $f : A \to B$. (You don't necessarily need to be able to produce this notation, but you should be able to understand it.) There are three pieces of information here:

- The first thing is the name of the function, f. It's very common to call functions f (short for "function"), but there's nothing special about this letter; you can call your functions anything, such as g, sin, or Fred.
- The next part is the set A. This is called the *domain*; it is the set of things which we are putting into the machine f. For us it'll often be the real numbers, like 0, 8, $-\sqrt{2}$, or π, which is written as $\mathbb{R}$, but in principle it could be anything, such as $\mathbb{R} - \{6\}$ (i.e. all real numbers except 6) or the set of characters in the English alphabet.
- Finally, there is the set B, called the *codomain*. This is the set of things that f can output. Again, it will often be $\mathbb{R}$, but could be anything.

For example, consider the function $f(x) = \frac{1}{x}$. This is defined for every real number x except for $x = 0$, and the output will also be a nonzero real number. Therefore we could write f as $f : \mathbb{R} - \{0\} \to \mathbb{R} - \{0\}$.

Since every nonzero real number is also just a real number, we could also replace the codomain by $\mathbb{R}$, and just write $f : \mathbb{R} - \{0\} \to \mathbb{R}$. However, we *cannot* replace the domain by $\mathbb{R}$, since $\mathbb{R}$ contains 0 and $f(0)$ is not defined.

In the case where we write it as $f : \mathbb{R} - \{0\} \to \mathbb{R}$, even though the codomain is $\mathbb{R}$, not every real number will be a value of this function. This often happens: for example, if $f(x) = 3$, we can view this as a function $\mathbb{R} \to \mathbb{R}$ since it takes in any real number and spits out a real number. But in fact, in this case the only number this ever outputs is 3; in the previous case, our function $f(x) = \frac{1}{x}$ always outputs a nonzero number. The set of values which f actually outputs is sometimes called the *range* or *image* of f. In these examples, the image of $f(x) = \frac{1}{x}$ is $\mathbb{R} - \{0\}$, the nonzero real numbers, and the image of $f(x) = 3$ is just $\{3\}$.

Given two functions, we can form a new function from them in a few different ways. The most general way to do this is *composition*. Given functions $f : A \to B$ and $g : B \to C$, we can form a new function, called $g \circ f : A \to C$, which works by feeding in an element a of

A into f, which gives an output $f(a)$ in B; and then feeding this output into g, which spits out $g(f(a))$ in C.

For example, let $f : \mathbb{R} - \{0\} \to \mathbb{R}$ be given by $f(x) = \frac{1}{x}$, and $g : \mathbb{R} \to \mathbb{R}$ be given by $g(x) = 2x + 1$. Then $g \circ f$ is given by $(g \circ f)(x) = g(f(x)) = 2 \cdot (\frac{1}{x}) + 1 = \frac{2}{x} + 1$.

It's important that the functions be lined up correctly, i.e. the *codomain* of f has to be the *domain* of g in order for this to work. For example, in the case above we *cannot* form $f \circ g : \mathbb{R} \to \mathbb{R} - \{0\}$. Why not? Formally, we can plug in g to f to get $f(g(x)) = \frac{1}{2x+1}$. But this does not define a function $\mathbb{R} \to \mathbb{R} - \{0\}$, because it is not defined on $x = -\frac{1}{2}$.

Usually, we'll think of functions as $\mathbb{R} \to \mathbb{R}$, even if they're not strictly defined on all of $\mathbb{R}$; we tend to think of the points where they're not defined as "singularities," or somehow "bad" points to be avoided. This usually works out fine, and then we can think of function composition pretty freely, because any two functions $\mathbb{R} \to \mathbb{R}$ have the same domain and codomain, so we can compose them in either order; we then instead have to keep track of where the bad points are. Nevertheless, it's important to understand this more formal way of thinking about functions too.

If we restrict to functions $\mathbb{R} \to \mathbb{R}$ (or away from some bad points), there are other ways we can combine functions as well, using arithmetic operations: for example, given $f, g : \mathbb{R} \to \mathbb{R}$ we can form $f + g$ or $f - g$ by $(f + g)(x) = f(x) + g(x)$ and similarly $(f - g)(x) = f(x) - g(x)$. We can do the same thing with multiplication or division: $(fg)(x) = f(x)g(x)$, and $(f/g)(x) = f(x)/g(x)$ (watching out for division by zero!).

Warning: multiplication and composition are different! This might sound obvious, but it's an easy mistake: $(fg)(x)$ and $(f \circ g)(x)$ look very similar, and even something like $f(x)$ is easy to think of as "f times x" because we're used to putting two symbols next to each other as meaning multiplication, but these are different. For example, if $f(x) = \frac{1}{x}$ and $g(x) = 2x + 1$ like above, then $g \circ f(x) = \frac{2}{x} + 1$, as we saw above, but $(gf)(x) = g(x)f(x) = (2x + 1) \cdot (\frac{1}{x}) = 2x \cdot \frac{1}{x} + 1 \cdot \frac{1}{x} = 2 + \frac{1}{x}$.