MAT 1330A: Calculus I for the Life Sciences

Fall 2017

Lecture Notes

Prof: Monica Nevins

These are sketches of my lecture notes, based in very large part on the handwritten lecture notes of Frithjof Lutscher. You might choose to read them before class (although updates may occur as late as the night before); you may choose to print them for class and use them to reduce the writing you need to do; you may choose to ignore them until you're rereading your notes and find that something you've written makes no sense, in which case looking at what I wrote might illuminate the problem. Or not.

These notes are not a substitute for attending class; reading math is less fun than participating in it in class. Beware of typos, errors, omissions and miscalculations. For more details, and a different perspective, read the corresponding section of our textbook — a very good textbook with very few typos and excellent, illustrated explanations and examples.

These notes have been developed from multiple sources. I am particularly grateful to Dheera Venkatraman, who developed the online graphing tool FooPlot (fooplot.com) which I have used to make most of the graphs in these notes.

Last update: November 30, 2017

Contents

Chapter 1

Introduction

My apologies for inevitable typos; these notes are just intended to help supplement what we said in class and of course everything is explained more clearly in the textbook, with nice examples and graphs.

1.1 Lecture 1: Syllabus and course information

In our first class we will present the syllabus for the course, which includes information about various dates and deadlines, and the calculation of your final grade. It also includes a list of exercises, which you should do after each lecture to be sure you are keeping up with the material. A selection of these exercises will be discussed in the DGD. The syllabus also includes an approximate timeline for the course, outlining which sections of the textbook we will be covering each lecture.

The timeline and the list of exercises are also available on the public website for the course, which is also where I will post announcements and additional information. Please check the course website regularly.

Note that the lectures proceed at a rapid pace. Ideally, you will learn quite a bit of what you need in the lecture, through the theory and examples presented there; and the rest you will fill in when you do the exercises after class. The homework assignments give you an opportunity for additional practice and to obtain directed feedback; however, as they are usually fairly short, homework assignments alone do not provide you with sufficient practice to ensure success.

There are many resources available to you as part of this course; it is up to you to choose to use them:

Lectures Two lectures per week, each 80 minutes long. Theory and examples and a sense of the big picture.

DGDs Once per week, 80 minutes, led by a graduate student Teaching Assistant (TA). Lots of examples, smaller group size, more opportunities to ask questions. Plus some quizzes, for some immediate low-stress feedback on the progress of your learning.

Math Help Centre Open five days a week (in a new temporary location in TBT), staffed by graduate student TAs. Come with your exercises (but not your homework problems) and the TAs will help you work through them. If your question is more theoretical in nature, come to office hours instead.

Office Hours As scheduled. Come with your questions: theory, examples, exercises, etc.

Virtual Campus Look for announcements, updates to exercises and remarks about each lecture, etc.

Textbook Has an index at the back, and a good table of contents at the front, to help you find things. Well-written, with lots of examples and excellent graphics. The exercises are from the textbook and are chosen from among those whose answer is at the back.

Homework We will have weekly assignments using MapleTA, an interactive platform that you can use to measure your progress and see what topics you need to focus more attention on. There is a "How did I do?" button that tells you when you have the right answer — so start the homework early, and if you run into difficulties, go back to the book, read examples, and take the time to learn what you've missed. **Note how this strategy can't work if you only start the homework at the last minute before the deadline.**

Peer help groups There are various mentorship and peer help groups on campus; this can be a good way to learn and keep motivated. These groups are not overseen by the instructor and are not part of the course — so please do confirm any rumours about course policies or test content directly with your professor!

Friends Teaching, by trying to explain an example to a friend, can be the most effective way of learning. Working with friends can also be very motivating — everyone's in this together.

We will review some algebra, trigonometry and functions in our first two lectures to ensure we have a common vocabulary and to preview things to come. **Experience has shown that mastering these pre-Calculus skills is 100% essential for success in Calculus.** Bring your questions to the drop-in center or office hours; we will be very happy to help.

Now, for the first two lectures, which are on background material for this course.[1]

1.2 Lecture 1: Algebra

Solid algebraic skills are essential — they are the rules of the language of mathematics, and you will be using them in all your courses and any analytic or quantitative work you do in your field. Let's recap some rules that you need to know in order to do the exercises attached to this section (and indeed, that you will be using throughout the course, without comment). These represent simplication steps in a calculation that would typically be skipped on the blackboard.[2]

[1] I will not be discussing all this material in class; I include it here for your information and to help you remember what you need to do the problems. After Lecture 2, these lecture notes will be much closer to what I present in class.

[2] For solutions to the problems in this section, come to office hours, or the math help centre.

1.2.1 Algebraic manipulations I: powers, roots, exponentials, logarithms

Recall that if n is a positive integer and $a \in \mathbb{R}$ then

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

Therefore, if n, m are two positive integers, we have

$$a^m a^n = (\underbrace{a \times a \times \cdots \times a}_{m \text{ times}}) \times (\underbrace{a \times a \times \cdots \times a}_{n \text{ times}}) = \underbrace{a \times a \times \cdots \times a}_{m+n \text{ times}} = a^{m+n},$$

whereas

$$(a^m)^n = \underbrace{a^m \times a^m \times \cdots \times a^m}_{n \text{ times}} = \underbrace{(\underbrace{a \times a \times \cdots \times a}_{m \text{ times}}) \times \cdots \times (\underbrace{a \times a \times \cdots \times a}_{m \text{ times}})}_{n \text{ times}} = a^{mn}.$$

Now assume $a > 0$. Then these two rules continue to hold for m and n arbitrary real numbers. For example:

$$a^{1/2} = \sqrt{a} \quad \text{since} \quad (a^{1/2})^2 = a^{\frac{1}{2} \times 2} = a.$$

(Note that this would be false if $a < 0$, which is why we insist on a positive base.)

In general, we have

$$a^{m/n} = (a^m)^{1/n} = \sqrt[n]{a^m} = (\sqrt[n]{a})^m$$

and

$$a^{-m/n} = \frac{1}{a^{m/n}}.$$

(To define a^x for a real number x that is not a fraction, like $x = \pi$, requires the notion of limits from Calculus. But your calculator can approximate it very well.)

Therefore, to solve $x^{3/5} = 8$, we can take both sides to the power $5/3$ (since $\frac{3}{5} \times \frac{5}{3} = 1$) to get

$$(x^{3/5})^{5/3} = 8^{5/3} \quad \Rightarrow \quad x = (\sqrt[3]{8})^5 = 2^5 = 32.$$

On the other hand, when the variable is in the exponent, such as in the equation $3^x = 25$, we need to use the logarithm function.

Recall that the logarithm of base $a > 0$ is the inverse function of the exponential: for $x > 0$

$$y = \log_a x \quad \text{iff} \quad a^y = x.$$

When $a = 10$, then we just write log; when $a = e$ (Euler's number) then it's called the *natural logarithm* and we write ln. Of all the logarithm functions, it's $\ln(x)$ that is the easiest to work with in Calculus, but $\log(x)$ is very common in the natural sciences (e.g. measuring pH values).

The rules for logarithms are therefore the inverse of the rules for exponentials. Where exponentials take sums to products and products to exponents, logarithms do the opposite: if $x, y > 0$ then

$$\log_a(xy) = \log_a(x) + \log_a(y), \quad \log_a(x^t) = t \log_a(x).$$

In other words, taking the logarithm of an expression can make it simpler, which is part of the reason they are so useful. When you have a variable in an exponent, you will very often use a logarithm to simplify the expression. If you cannot get your terms over a common base, the default is to use $\ln(x)$.

Problems to try:

- simplify $\frac{\sqrt{x^{1/2}y^5}}{x^5y^{1/4}}$, where $x, y > 0$;
- simplify $\frac{(\sqrt{x}\sqrt[3]{y})^{-1/2}}{\sqrt[3]{x^2y}}$ where $x, y > 0$;
- solve for x: $2^{x+3} = 16^{2x-1}$;
- solve for x: $2^{2x+3} = 3^{4x-1}$;
- solve for x: $\frac{1}{x} + \frac{1}{y} = \frac{1}{z}$;
- solve for x: $\log(x+5) - \log(x-1) = \log(x+1)$;
- solve for x: $2\ln(x) - \ln(x+4) = \ln(2)$.

1.2.2 Algebraic manipulations II: simplify multiple fractions, rationalize a denominator

When faced with a complex algebraic expression, like the exercises below, it's helpful to identify the "worst" part and work from there.

Simplifying fractions always uses the rule:

$$\frac{a}{b} = \frac{a}{b} \times 1 = \frac{a}{b} \times \frac{c}{c} = \frac{ac}{bc}$$

for some nonzero number c. How to choose c? You generally want your denominator to be as simple as possible, because the rules of addition of fractions say:

$$\frac{a+b}{d} = \frac{a}{d} + \frac{b}{d}, \quad \text{(but remember:} \quad \frac{s}{t+u} \neq \frac{s}{t} + \frac{s}{u}\text{!!!).}$$

A particularly clever trick, that applies only to expressions with square roots, is called rationalization, and it's based on the identity

$$(a+b)(a-b) = a^2 - b^2.$$

So, for example,

$$\frac{1}{\sqrt{3}-2} = \frac{1}{\sqrt{3}-2} \times \frac{\sqrt{3}+2}{\sqrt{3}+2} = \frac{\sqrt{3}+2}{3-4} = -(\sqrt{3}+2).$$

Please note that $(\sqrt{x})^2 = |x|$; more on this in Section 1.2.5, below.

Problems to try:

- simplify $\frac{4+\frac{1}{k}}{\frac{5}{k}-2}$;
- simplify $\frac{z^{-1}+3}{z^{-2}+2}$;
- rationalize the denominator $\frac{1}{\sqrt{10}-3}$;
- simplify $\frac{\sqrt{6}-\sqrt{8}}{\sqrt{6}+\sqrt{8}}$.

1.2.3 Polynomials: solve quadratic equations, factor a polynomial, long division

You know how to factor quadratic polynomials. Some equations don't appear to be quadratics until you simplify — which is another reason to always try to simplify an expression before doing the next step.

When you have a polynomial of degree ≥ 3, like

$$x^3 - 7x + 6$$

then if it has a factorization with integer roots, those integers must divide the constant term. So in this case, the possible roots are $\pm 1, \pm 2, \pm 3, \pm 6$. Plug in values until you find one root. In this case, for example, you might find that $x = 1$ is a root, whence $(x - 1)$ is a factor. Then use long division

$$\begin{array}{r|rrrr} & x^2 & +x & -6 & \\ \hline x-1 & x^3 & & -7x & +6 \\ & -x^3 & +x^2 & & \\ \cline{2-3} & & x^2 & -7x & \\ & & -x^2 & +x & \\ \cline{3-4} & & & -6x & +6 \\ & & & 6x & -6 \\ \cline{4-5} & & & & 0 \end{array}$$

to get

$$\frac{x^3 - 7x + 6}{x - 1} = x^2 + x - 6 = (x - 2)(x + 3)$$

so $x^3 - 7x + 6 = (x - 1)(x - 2)(x + 3)$. If your polynomial has fractional coefficients, consider factoring out the common denominator, which you can leave as a constant in front of your factored expression.

Example 1.1. Factor $p(x) = x^3 - x^2 - x - 2$.

Solution: Again, any integer roots need to divide -2. We try some values but $p(1) \neq 0$, $p(-1) \neq 0$, ... but $p(2) = 8 - 4 - 2 - 2 = 0$, so 2 is a root, so $(x - 2)$ is a factor.

We don't find any other roots, so it is time to use long division.

$$
\begin{array}{r|l}
 & x^2 + x + 1 \\
\hline
x-2 & x^3 - x^2 - x - 2 \\
 & -x^3 + 2x^2 \\
\hline
 & x^2 - x \\
 & -x^2 + 2x \\
\hline
 & x - 2 \\
 & -x + 2 \\
\hline
 & 0
\end{array}
$$

Therefore, $p(x) = x^3 - x^2 - x - 2 = (x-2)(x^2 + x + 1)$. Using the quadratic formula, we conclude that

$$p(x) = (x-2)(x - \frac{1}{2}(-1 + \sqrt{3}))(x - \frac{1}{2}(-1 - \sqrt{3})).$$

□

Problems to try:

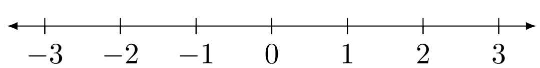
- solve for x: $\frac{1}{x} + \frac{1}{x^2} = 1$;
- solve for m: $m = \sqrt{m+6}$;
- solve for x: $\frac{4x}{1+x} = 3x$;
- factor $x^3 + 1000$;
- divide $x^3 + x^2 + \frac{5}{4}x + 3$ by $x + \frac{3}{2}$;
- find all values of k for which the quadratic $x^2 + 2kx + 9k - 8 = 0$ has only one solution for x.

1.2.4 Solving inequalities

The important thing to remember:

Multiplying an inequality by a negative value changes its direction. Remember,

$$2 < 3 \quad \Leftrightarrow \quad -2 > -3.$$



Example 1.2. Solve for x:

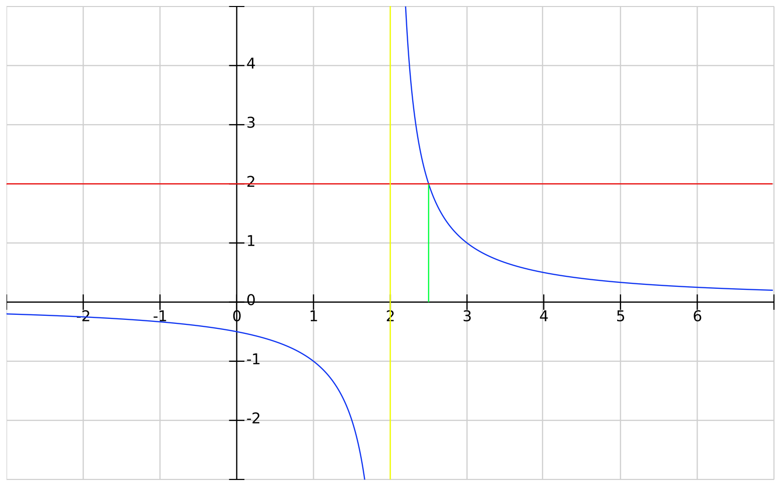
$$\frac{1}{x-2} = 2.$$

We multiply both sides by $x - 2$, to get $1 = 2(x-2)$ or $1 = 2x - 4$ or $2x = 5$ or $x = \frac{5}{2}$. So this is the unique solution. □

Example 1.3. Solve for x:

$$\frac{1}{x-2} < 2.$$

Wrong answer: $1 < 2(x-2) = 2x - 4$ so $x > \frac{5}{2}$. Why is it wrong? Let's look at the graph!



$y = 1/(x-2)$ is in blue; $y = 2$ is in red; the answer should be **all** x-values for which the blue graph lies below the red line: so all $x < 2$ and all $x > 5/2$.

What went wrong? We multiplied both sides of the inequality by $x-2$, which is sometimes positive and sometimes negative. So when $x - 2 > 0$ (meaning $x > 2$) then our reasoning holds, and we find that the only values of $x > 2$ which satisfy the inequality are $x > 5/2$.

But when $x - 2 < 0$, meaning $x < 2$, then we have to change the direction of the inequality when we multiply:

$$(x < 2) : \frac{1}{x-2} < 2 \Leftrightarrow 1 > 2(x-2) = 2x - 4 \Leftrightarrow 5 > 2x \Leftrightarrow x < \frac{5}{2}.$$

What?! This is saying: when $x < 2$, the condition to satisfy the inequality is that x must be less than $5/2$ — which is always true. Therefore every $x < 2$ satisfies the condition.

Our answer: all $x < 2$ and all $x > 5/2$: this is written as

$$(-\infty, 2) \cup (5/2, \infty).$$

□

Example 1.4. Another way to solve $\frac{1}{x-2} < 2$: avoid the problematic multiplication by a variable term.

$$\begin{aligned} \frac{1}{x-2} < 2 &\Leftrightarrow \frac{1}{x-2} - 2 < 0 \\ &\Leftrightarrow \frac{1}{x-2} - \frac{2(x-2)}{x-2} < 0 \\ &\Leftrightarrow \frac{5-2x}{x-2} < 0 \\ &\Leftrightarrow \frac{2x-5}{x-2} > 0 \end{aligned}$$

which holds only if the numerator and the denominator have the same sign. Now $2x - 5 > 0$ iff $x > 5/2$; and $x-2 > 0$ iff $x > 2$. So both are positive if $x > 5/2$ since $5/2 > 2$. Similarly, $2x-5 < 0$ iff $x < 5/2$, and $x - 2 < 0$ iff $x < 2$, so both are negative iff $x < 2$, since $2 < 5/2$. This again gives

$$(-\infty, 2) \cup (5/2, \infty).$$

□

Let's do another example.

Example 1.5. Find all x for which

$$\frac{2}{x} + 5 < 3.$$

Solution: We want to isolate the x. We begin with

$$\frac{2}{x} + 5 < 3 \Leftrightarrow \frac{2}{x} < -2 \Leftrightarrow \frac{2}{-2x} > 1$$

(where in this last step we divided both sides by -2, which is negative, so the inequality changed direction). So our inequality is actually

$$\frac{-1}{x} > 1.$$

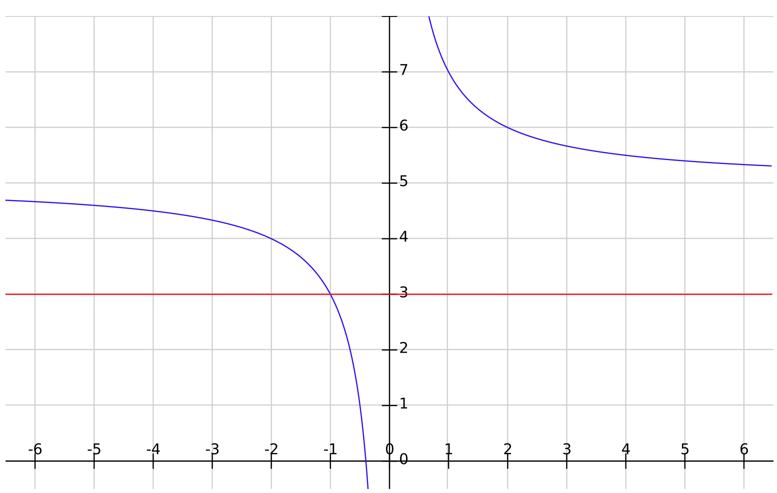
Two cases:

If $x > 0$ then we can multiply both sides by x, to get $-1 > x$, or $x < -1$. But there are **no** positive values of x which are less than -1! We conclude that there are no solutions arising from this case.

If $x < 0$, then we can still multiply both sides by x, but this time the direction of the inequality changes. We thus get $-1 < x$. The negative values of x that satisfy the inequality are those between -1 and 0.

Our total answer: the solution set is $(-1, 0) = \{x \in \mathbb{R} \mid -1 < x < 0\}$. □

We can check our answer against reality by sketching the graph.



$y = 2/x + 5$ is in blue; $y = 3$ is in red; the answer should be **all** x-values for which the blue graph lies below the red line: so just the region between -1 and 0.

Remark 1.6. Another way to solve the previous question:

$$\frac{2}{x} + 5 < 3 \Leftrightarrow \frac{2}{x} + 2 < 0 \Leftrightarrow \frac{2}{x} + \frac{2x}{x} < 0 \Leftrightarrow \frac{2+2x}{x} < 0 \Leftrightarrow \frac{2(x+1)}{x} < 0.$$

This fraction is < 0 iff the numerator and the denominator have opposite signs. If the numerator is negative (so $x < -1$) then the denominator is forced to be negative, so that doesn't work. If the numerator is positive (so $x > -1$) and the denominator is negative (so $x < 0$), then the quotient is negative. So that's the only solution interval.

You might also have learned to use a table to decide this question: since the key points are -1 and 0, it suffices to test values of x which are $x < -1$, $-1 < x < 0$ and $x > 0$ and see which ones give the right sign.

1.2.5 Absolute values: how to handle them and how to solve equations

The important thing to remember:

The absolute value is a function, defined by the formula

$$|x| = \begin{cases} x & \text{if } x \geq 0; \\ -x & \text{if } x < 0. \end{cases}$$

Thus for example:

$$|x-3| = \begin{cases} x-3 & \text{if } x-3 \geq 0, \text{ i.e. } x \geq 3 \\ -(x-3) & \text{if } x-3 < 0, \text{ i.e. } x < 3 \end{cases}$$

and

$$|x^2-3| = \begin{cases} x^2-3 & \text{if } x^2-3 \geq 0 \\ -(x^2-3) & \text{if } x^2-3 < 0. \end{cases}$$

In the latter case, to get a useful formula for $|x^2 - 3|$, we need to solve the conditions $x^2 - 3 \geq 0$ (and $x^2 - 3 < 0$). Well,

$$x^2 - 3 \geq 0 \Leftrightarrow x^2 \geq 3 \Leftrightarrow x \geq \sqrt{3} \text{ or } x \leq \sqrt{3}$$

(which you could alternately solve by saying

$$x^2 - 3 \geq 0 \Leftrightarrow (x - \sqrt{3})(x + \sqrt{3}) \geq 0$$

and considering what happens in each interval on the number line.)

Example 1.7. Solve $|x - 3| = 5$.

Solution: since

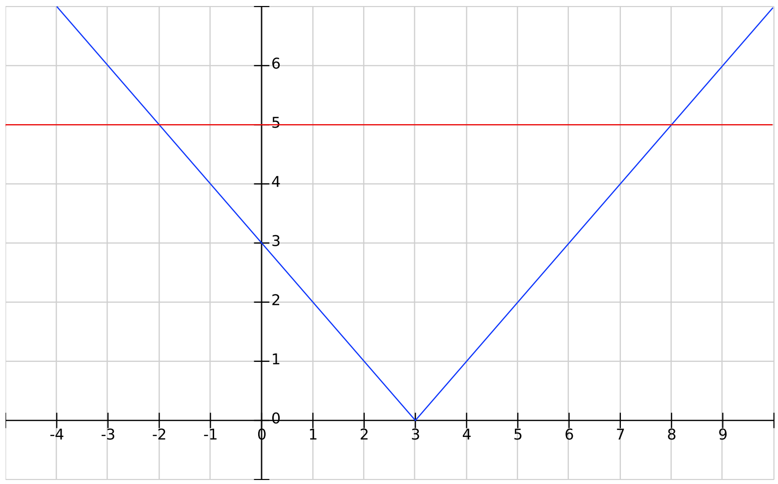
$$|x-3| = \begin{cases} x-3 & \text{if } x-3 \geq 0, \text{ i.e. } x \geq 3 \\ -(x-3) & \text{if } x-3 < 0, \text{ i.e. } x < 3 \end{cases}$$

we have two cases to consider.

If $x \geq 3$, then the equality is actually $x - 3 = 5$, whose solution is $x = 8$. Since $8 \geq 3$ this is a valid solution on this interval. We check: $|8 - 3| = 5$ ✓.

If $x < 3$, then the equality is actually $-(x - 3) = 5$, whose solution is $-x + 3 = 5$ or $x = -2$; since $-2 < 3$, this is a valid solution on this interval. We check: $|-2-3| = |-5| = 5$ ✓.

Therefore, the solutions to $|x - 3| = 5$ are -2 and 8. See the graph below.



$y = |x - 3|$ is in blue; $y = 5$ is in red; the answer is the x-values for the points of intersection, which are -2 and 8.

□

Example 1.8. Solve $|x^2 - 3| = 1$.

Solution: We have

$$|x^2 - 3| = \begin{cases} x^2 - 3 & \text{if } x^2 - 3 \geq 0 \\ -(x^2 - 3) & \text{if } x^2 - 3 < 0. \end{cases}$$

Therefore there are two cases to consider.

1st case: $x^2 - 3 \geq 0$. This is $(x - \sqrt{3})(x + \sqrt{3}) \geq 0$, which happens if $x \geq \sqrt{3}$ or $x \leq -\sqrt{3}$. In this case, our equality is

$$x^2 - 3 = 1 \Leftrightarrow x^2 = 4 \Leftrightarrow x = \pm 2.$$

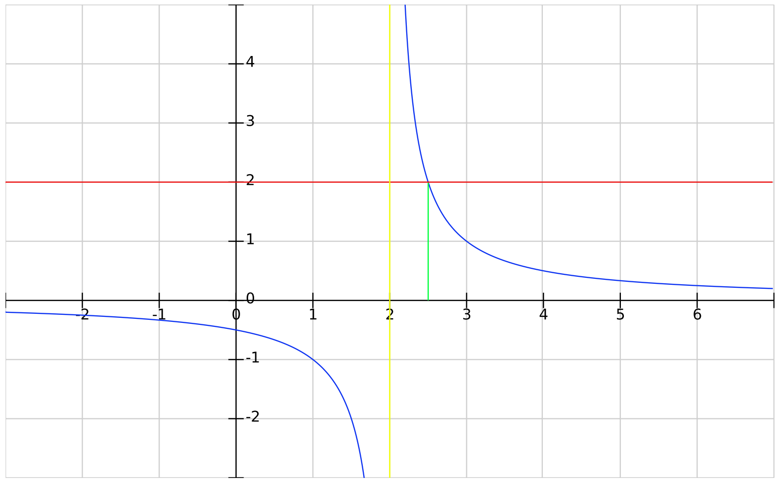
Since $2 > \sqrt{3}$, this is a solution in our interval. Since $-2 < -\sqrt{3}$, this is also valid. So both ± 2 are solutions. Check: if $x = 2$ or $x = -2$, we get $|x^2 - 3| = |4 - 3| = 1$, ✓.

2nd case: $x^2 - 3 < 0$. This happens if $x \in (-\sqrt{3}, \sqrt{3})$. In this case, the equality is

$$-(x^2 - 3) = 1 \Leftrightarrow -x^2 + 3 = 1 \Leftrightarrow x^2 = 2 \Leftrightarrow x = \pm\sqrt{2}.$$

Since both $\pm\sqrt{2} \in (-\sqrt{3}, \sqrt{3})$ both solutions are valid. Check: if $x = \sqrt{2}$ or $x = -\sqrt{2}$, then $|x^2 - 3| = |2 - 3| = |-1| = 1$, ✓.

Our answer: there are four solutions: $x \in \{\pm\sqrt{2}, \pm 2\}$. See the graph.



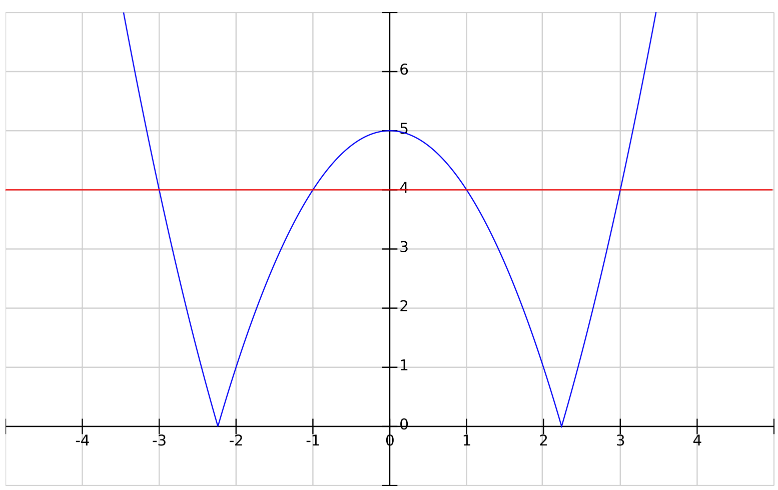
$y = |x^2 - 3|$ is in blue; $y = 1$ is in red; the answer is the x-values for the points of intersection, which are ± 2 and $\pm\sqrt{2}$.

□

We use the same techniques to solve inequalities with absolute values.

Example 1.9. Solve for all x such that $|x^2 - 5| < 4$.

Solution:



$y = |x^2 - 5|$ is in blue; $y = 4$ is in red; the answer is the x-values such that the blue curve lies below the red line, which are the intervals $(-3, -1)$ and $(1, 3)$.

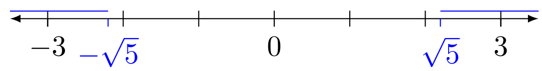
To find this algebraically, we begin by understanding the expression on the left. We have that

$$|x^2 - 5| = \begin{cases} x^2 - 5 & \text{if } x^2 - 5 \geq 0 \\ -(x^2 - 5) & \text{if } x^2 - 5 < 0. \end{cases}$$

1st case: suppose $x^2 - 5 \geq 0$. That means $x^2 \geq 5$, whence $x \leq -\sqrt{5}$ or $x \geq \sqrt{5}$. Then $|x^2 - 5| = x^2 - 5$, so our inequality is just

$$x^2 - 5 < 4 \Leftrightarrow x^2 < 9 \Leftrightarrow -3 < x < 3.$$

But we can only accept solutions within our case (that is, larger than $\sqrt{5}$ in absolute value):



So that leaves us with

$$-3 < x \leq -\sqrt{5} \quad \text{and} \quad \sqrt{5} \leq x < 3.$$

2nd case: suppose $x^2 - 5 < 0$. That means $x^2 < 5$ or $|x| < \sqrt{5}$. In this interval, $|x^2 - 5| = -(x^2 - 5) = 5 - x^2$, so our inequality is

$$5 - x^2 < 4 \Leftrightarrow -x^2 < -1 \Leftrightarrow x^2 > 1 \Leftrightarrow |x| > 1.$$

Again, restricting to the x values in our case, this leaves us with

$$-\sqrt{5} < x < 1 \quad \text{and} \quad 1 < x < \sqrt{5}.$$

Putting all these intervals together, we deduce the final answer is the union of two intervals:

$$(-3, -1) \cup (1, 3) = \{x \in \mathbb{R} \mid -3 < x < -1 \text{ or } 1 < x < 3\}$$

as we had guessed from the picture. □

A slightly easier example is as follows.

Example 1.10. Solve for all x such that $|x^2 - 5| \geq 4$.

We consider the graph as in the preceding example, and infer that the solution will be

$$(-\infty, -3] \cup [-1, 1] \cup [3, \infty),$$

a union of 3 intervals.

To find this algebraically, we divide the absolute value into cases.

1st case: if $x^2 - 5 \geq 0$ (meaning $x^2 \geq 5$ or $|x| \geq \sqrt{5}$), the inequality is simply

$$x^2 - 5 \geq 4 \Leftrightarrow x^2 \geq 9 \Leftrightarrow |x| \geq 3.$$

Since $3 > \sqrt{5} \simeq 2.2$, all of these answers fall within our case, and we conclude that any value $x \geq 3$ or $x \leq -3$ will do.

2nd case: if $x^2 - 5 < 0$ (meaning $x^2 < 5$ or $|x| < \sqrt{5}$), then the inequality becomes

$$-(x^2 - 5) \geq 4 \Leftrightarrow -x^2 + 5 \geq 4 \Leftrightarrow -x^2 \geq -1 \Leftrightarrow x^2 \leq 1.$$

Since $1 < \sqrt{5}$, all of these solutions lie in the interval we are considering for the 2nd case, and so we conclude that any value $-1 \leq x \leq 1$ will do.

Our final answer is that the solutions are

$$\{x \in \mathbb{R} \mid x \leq -3 \text{ or } -1 \leq x \leq 1 \text{ or } x \geq 3\} = (-\infty, -3] \cup [-1, 1] \cup [3, \infty).$$

□

Notice that the answers of the two previous examples were complements of one another, meaning, each $x \in \mathbb{R}$ lies in exactly one of the two solution sets. This makes sense: either $|x^2 - 5| < 4$ or $|x^2 - 5| \geq 4$ (and never both).

When you want to solve something with an absolute value, like $|f(x)| = 3x + 2$ or $|f(x)| < 3x+2$, you have to split into two cases: the case where $|f(x)| = f(x)$ (which happens exactly when $f(x) \geq 0$); and the case where $|f(x)| = -f(x)$ (which happens exactly when $f(x) < 0$).
So you end up solving two problems, so you have a multi-step logical process:

1. Start by asking (and answering) : "for what x values am I in case 1 (or case 2)?".
2. in case 1 ($f(x) \geq 0$), you have $|f(x)| = f(x)$, so your original problem is now much simpler, and you can solve it. (But only answers that fall into case 1 are valid.)
3. in case 2 ($f(x) < 0$), you have $|f(x)| = -(f(x))$, so again your problem is simpler (but definitely different); you solve it, and accept all answers that fall within case 2.
4. then, put these two sets together, to produce the "total answer": the set of all x that solve your original problem. This total answer may not even include the values that defined your cases.

Problems to try:

- easy: solve for all x such that $\frac{x}{2} - 3 > 5$;
- harder: solve for all x such that $\frac{2}{x} - 3 > 5$;
- solve for t : $\frac{1}{3}t + 4 > -1$;
- solve for t: $\frac{3}{x} + 4 < -1$;
- solve for x: $|\frac{x}{2} - 3| > 5$;
- solve for x: $|x^2 - 5| = 1$;
- solve for x: $|x^2 - 16| = 9$;
- solve for x: $|\frac{x}{2} - 3| > 5$;
- solve for y: $|12 - x^2| > 3$.

End of lecture # 1

Note that $|x - 3| \neq |x| + 3$; more generally,

$$|x \pm y| \leq |x| + |y|$$

(but not always equal).

1.3 Lecture 2: Review of functions

A *function* is a rule which assigns to each element in the *domain* a unique element in the *range*. In this course, the domain and range will always be subsets of the real numbers. One way to specify the domain D explicitly is to write such a function f with domain $D \subseteq \mathbb{R}$ as

$$\begin{aligned} f\colon D &\to \mathbb{R} \\ x &\to f(x). \end{aligned}$$

For example,

$$\begin{aligned} f\colon (0,\infty) &\to \mathbb{R} \\ x &\to \frac{1}{\sqrt{x}}. \end{aligned}$$

Very often, we just write something like

$$f(x) = \frac{x-2}{\sqrt{x-5}}$$

from which you may deduce that the *natural domain* or *domain of definition* is all those x for which the formula is well-defined. Here, since $\sqrt{x-5}$ is only defined if $x - 5 \geq 0$, and since $x = 5$ is excluded because we can't divide by 0, we conclude that the domain of definition of f is $\{x \mid x > 5\}$ (that is, $D = (5, \infty)$).

We sometimes restrict the function to a smaller domain (for example, so that it becomes one-to-one); in this case we say we specify the function on a *given domain*.

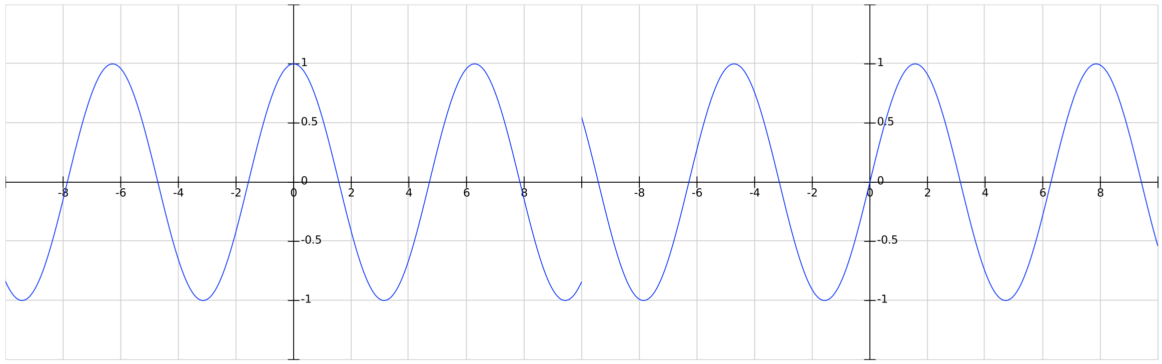
The name of the function is f; its value on a point x is $f(x)$.

1.3.1 Potential characteristics of functions

You can plug values into a function to get some values, but to understand a function is to understand its behaviour as a whole. Calculus is about understanding this behaviour, but even without Calculus, you can often identify key characteristics of functions that dictate its graph.

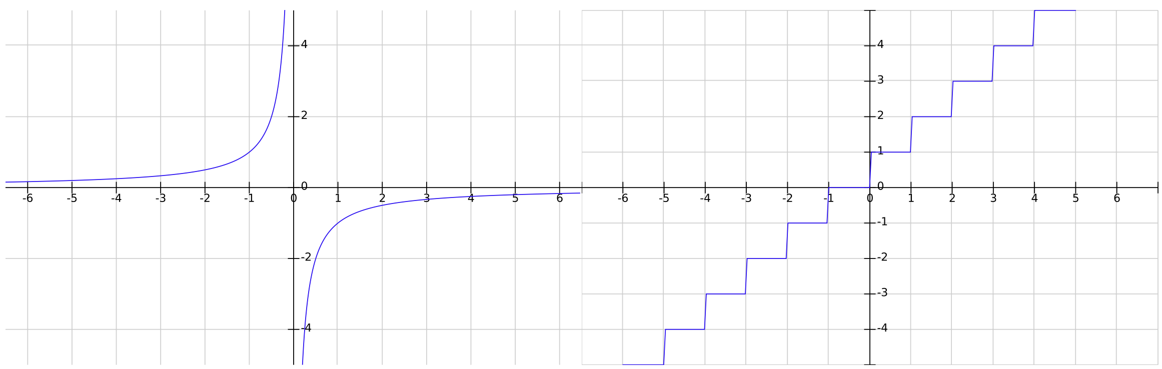
Even and odd functions Recall that a function is *even* if for all x in the domain, $-x$ is also in the domain and $f(x) = f(-x)$; its graph will be symmetric via reflection in the y-axis. A function is *odd* if instead $f(-x) = -f(x)$, which means the symmetry is via reflection in both axes. Most functions are neither even nor odd.

Examples of even functions: x^2, $\cos(x)$; examples of odd functions: x^3, $\sin(x)$.



$y = \cos(x)$ is on the left, it is even; $y = \sin(x)$ is on the right; it is odd.

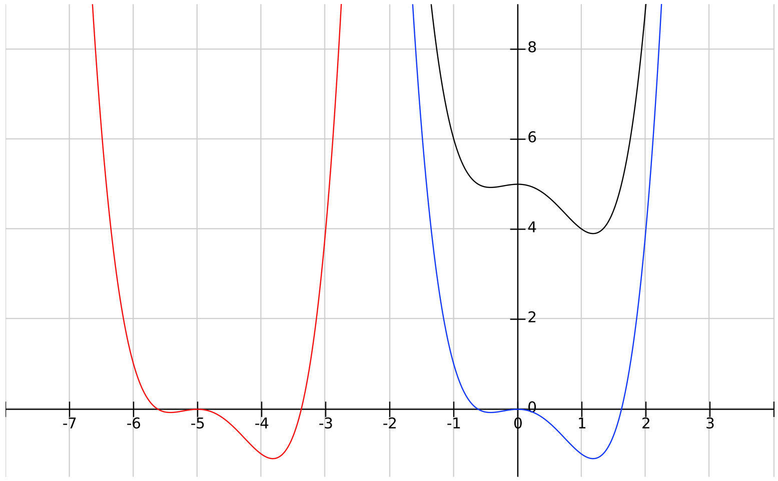
Increasing and decreasing functions A function f is said to be *increasing on an interval* I in its domain if for every $x_1, x_2 \in I$ with $x_1 < x_2$, we have $f(x_1) \leq f(x_2)$. (It is called *strictly increasing* if you in fact have the stronger condition $f(x_1) < f(x_2)$.) The function f is an *increasing function* if it is increasing on every interval of its domain.



$y = -1/x$ is on the left; it is increasing on $(-\infty, 0)$ and on $(0, \infty)$ but not defined at 0, so is an increasing function. (A break in the domain means you are allowed to reset.) The graph of an increasing function which is not strictly increasing is on the right.

We can defined a decreasing function similarly.

Linear transformations We can transform functions by linear operators, which retain the general shape of its graph. For example, the graph of $y = f(x) + 1$ is obtained from the graph of $y = f(x)$ by shifting one unit up; the graph of $y = f(x+1)$ is obtained from the graph of $y = f(x)$ by shifting one unit to the left.



The graph of $y = f(x) = x^4 - x^3 - x^2$ is in blue; the graph of $y = f(x) + 5$ is in black; and the graph of $y = f(x + 5)$ is in red.

Composition More generally, we can *compose* functions, which means to evaluate them sequentially (rather than in parallel, as one does for multiplication). The important point is that composition is not commutative, meaning, order matters.

For example, if $f(x) = x^5 + 3x$ and $g(x) = x^2$, then the composition $f \circ g$ is

$$(f \circ g)(x) = f(g(x)) = f(x^2) = x^{10} + 3x^2$$

whereas composing them in the opposite direction gives

$$(g \circ f)(x) = g(f(x)) = g(x^5 + 3x) = (x^5 + 3x)^2 = x^{10} + 6x^6 + 9x^2.$$

Notice that neither is equal to $f(x)g(x) = x^7 + 3x^3$.

Splicing functions together = piecewise-defined functions Another way to create a new function is to "splice" two or more functions together.

Example 1.11. The recommended maximum dosage, in mg, of a certain pain reliever, as a function of age x in years, is given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq 2 \\ 250 & \text{if } 2 < x \leq 12 \\ 500 & \text{if } x > 12. \end{cases}$$

This function is obtained by splicing three constant functions together. Some of the most interesting medical questions about this dosage model concern the transition points; these are also the most interesting points from the point of view of Calculus. □

Example 1.12. Another drug's maximum daily dosage, in mg, is expressed as a function of mass, in kg, is given by

$$g(x) = \begin{cases} 0 & \text{if } x < 30 \\ 4x - 120 & \text{if } x > 30. \end{cases}$$

In this case, there is no large jump at the transition point, but the transition is still sharp; we say the graph has a "cusp". Again, this cusp is the most interesting point, from both the applications and the mathematical point of view. □

Inverse functions If f is one-to-one on an interval I in its domain (equivalently, it passes the *horizontal* line test there), then there exists a function, denoted f^{-1}, which is the *inverse function* taking values in I. The inverse function is characterized by the relation

$$y = f^{-1}(x) \qquad \Leftrightarrow \qquad x = f(y) \quad \text{and } x \in I.$$

In other words, given a function $y = f(x)$, if you solve for x in terms of y (and get a unique answer!) then the formula you get is called the inverse function.

Notice that we swap variables to defined the function. The convention is always to use x as the independent variable, and y as the dependent variable.

Example 1.13. Find the inverse function of $y = \dfrac{5x+1}{3x-2}$.

Solution: To make sure the inverse function exists, we could sketch the graph and verify that f is one-to-one; or else we can try to solve for x in terms of y, and if we get a unique solution for each y (as opposed to having choices) then our function is one-to-one and we will have found a formula for the inverse. So let's do that:

$$\begin{aligned}
(3x-2)y &= 5x+1 \\
\Leftrightarrow \quad 3xy - 2y &= 5x+1 && \text{expand} \\
\Leftrightarrow \quad 3xy - 5x &= 2y+1 && \text{group terms with } x \text{ together} \\
\Leftrightarrow \quad x(3y-5) &= 2y+1 \\
\Leftrightarrow \quad x &= \frac{2y+1}{3y-5}
\end{aligned}$$

therefore the inverse function is

$$y = f^{-1}(x) = \frac{2x+1}{3x-5}.$$

□

Some important inverse functions: $\ln(x)$ is the inverse function of e^x (and vice versa); $\sqrt{x}$ is the inverse function of x^2 (on $I = (0, \infty)$); $\arcsin(x)$ is the inverse function of $\sin(x)$ (on $I = (-\pi/2, \pi/2)$).

Now let's look at several important classes of functions.

1.3.2 Polynomial functions

A polynomial in x is a function of the form

$$1 + 2x + 4x^2 \quad \text{or} \quad -\pi x + \frac{1}{47}x^7.$$

Most generally, it would look like

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

where $n \in \mathbb{N}$ (the natural numbers, $\mathbb{N} = \{0, 1, \cdots\}$) and each of the a_i are real numbers.

If $n = 1$, then the function is a linear function, like

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

We know that the graph of f is a straight line with slope m and y-intercept b.

Linear functions are the simplest kind of relationship that we can express between two variables. Very often, in an experiment, you will hope your points fall on a straight line, confirming a linear relationship.

If $n = 2$, then the function is a quadratic function, of the form

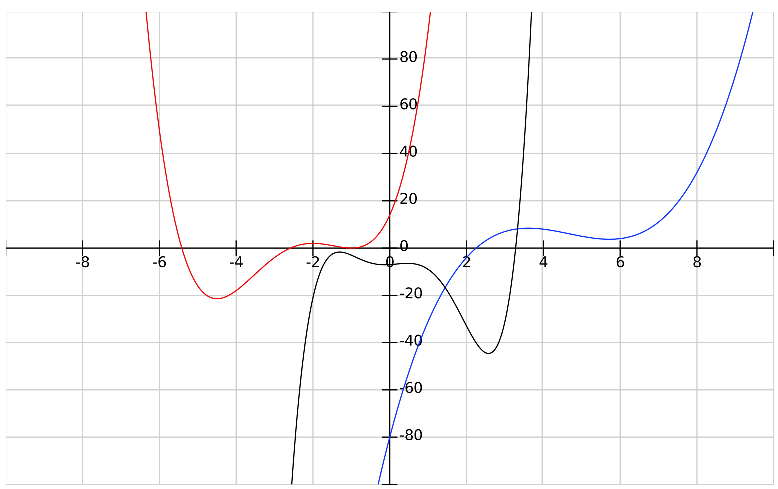
$$f(x) = ax^2 + bx + c.$$

We know that the graph of f is a parabola which is concave up $\cup$ if $a > 0$ and concave down $\cap$ if $a < 0$. Its x-intercepts, if any, are given by factoring, or by applying the quadratic formula:

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

Quadratic functions show up in some logistic growth (= limited resource) population models: the rate of growth relative to population size hits a peak, then decreases.

If $n \geq 3$, then this is a higher-degree polynomial. We recognize their shapes, but it is more difficult to find intercepts algebraically. Moreover, these functions have many more features (local maxima and minima) which are easiest to find using Calculus.



$y = (x-5)^3 + (x-5)^2 - 3(x-5) + 5$ is in blue; $y = (x+2)^4 + 2(x+2)^3 - 5(x+2)^2 + 2$ is in red; $y = x^5 - 2x^4 - 5x^3 + 3x^2 + x - 7$ is in black. Note the limits as $x \to \pm\infty$, as well as the number of local minima and maxima, in relation to the degree of the polynomial.

Given a polynomial of degree ≥ 3, if we know one root, we can use long division to simplify and perhaps find other roots. See Section 1.2.3.

1.3.3 Rational functions

A rational function is one of the form

$$f(x) = \frac{p(x)}{q(x)}$$

where $p(x), q(x)$ are polynomials — so a fraction with polynomials instead of numbers. Since we can never divide by zero, the domain of such a function is the set of all x such that $q(x) \neq 0$.

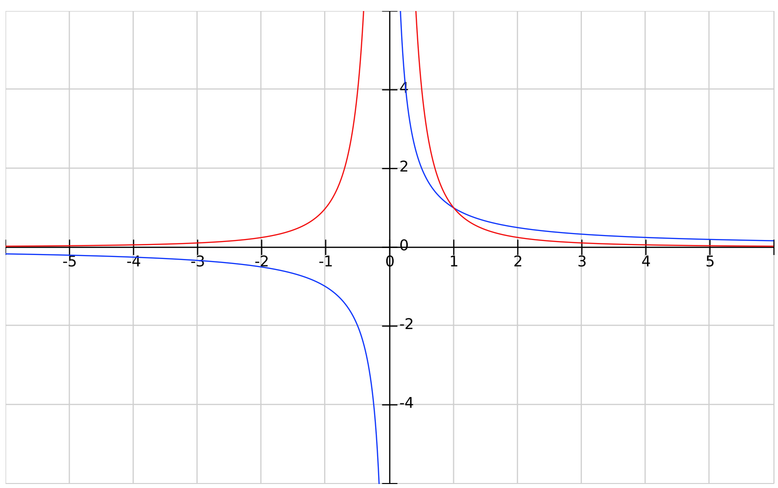
These kinds of functions commonly arise from an inverse-proportional relationship, such as in animal populations where the probability of an individual being eaten is inversely proportional to the size of the population of the prey (and directly proportional to the size of the population of predators). Rational functions can also give more realistic models than polynomials, because one can arrange for various kinds of asymptotes, and for the graph to remain positive for all $x > 0$. They can also arise when there is a natural vertical asymptote (such as trying to achieve the speed of light, or absolute zero in temperature).

Example 1.14. The domain of

$$f(x) = \frac{x^3}{(x-2)(x-1)}$$

is $D = (-\infty, 1) \cup (1, 2) \cup (2, \infty)$, that is, all real numbers except 1 and 2, which are the roots of the denominator. □

Very often (that is, unless the root of the denominator is also a root of the numerator), a root of the denominator induces a vertical asymptote of the graph of f.



The graph of $y = 1/x$ is in blue; the graph of $y = 1/x^2$ is in red. Notice that $y = 1/x$ is an odd function whereas $y = 1/x^2$ is even. Both have a vertical asymptote at $x = 0$.

Exercise 1.15. Find the domain of $f(x) = \frac{x+1}{x^2-2}$, and of $g(x) = \frac{2}{x-1} + \frac{5}{x^2+3}$.

1.3.4 Root or radical functions

A function like

$$f(x) = \sqrt{x^2+2}, \quad g(x) = \sqrt{x-5}, \text{ or } h(x) = \sqrt[3]{x-1}$$

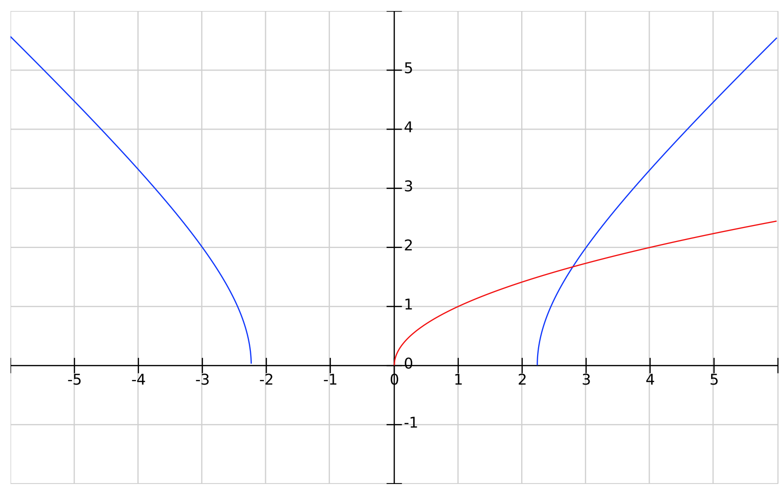
is a root or *radical* function. If the root is even (such as a square root) then the domain is the set of all x such that the expression of which we are taking the root is positive, because you cannot take the square (or fourth, or sixth) root of a negative number. (If the root is odd, like a cube root, there is no problem; for example, $\sqrt[3]{-8} = -2$.)

These kinds of functions commonly show up as part of a distance formula; for example, the distance between $(x, 1)$ and $(1, 2)$ is $d(x) = \sqrt{(x-1)^2 + 1}$. They are also common in various biological applications. For example, Kleiber's law states that the metabolic rate is proportional to $m^{3/4}$ where m is the individual's mass. For another example: heart rate has been observed to be proportional to $m^{1/4}$.

Example 1.16. Find the domain of $g(x) = \sqrt{x^2 - 5}$.

Solution: $\sqrt{x^2 - 5}$ is defined only if $x^2 - 5 \geq 0$, or $x^2 \geq 5$. So the domain is the set $(-\infty, -\sqrt{5}) \cup (\sqrt{5}, \infty)$. □

The square root function $y = \sqrt{x}$ in particular is only defined on half of the real line.



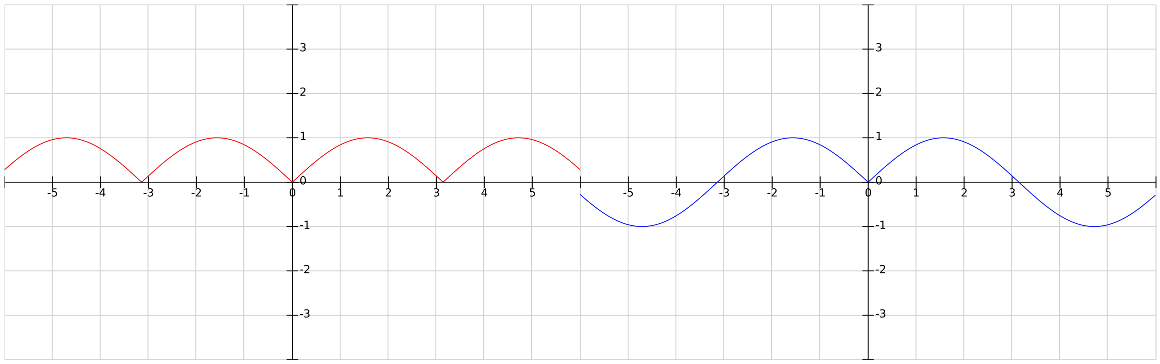
The graph of $y = \sqrt{x^2 - 5}$ is in blue; the graph of $y = \sqrt{x}$ is in red. Notice how they are not defined on the entire real line.

1.3.5 Absolute value function

The absolute value function is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0; \\ -x & \text{if } x < 0. \end{cases}$$

Given $y = f(x)$, the graph of $y = |f(x)|$ is obtained by reflecting any parts below the x-axis upwards. On the other hand, the graph of $y = f(|x|)$ is obtained by erasing the parts to the left of the y-axis and instead reflecting the graph to make it even.



The graph of $y = |\sin(x)|$ is on the left; the graph of $y = \sin(|x|)$ is on the right.

For examples of solving equations and inequalities involving the absolute value, see Section 1.2.5.

1.3.6 Trigonometric functions

The basic trigonometric functions are $\sin(x)$ and $\cos(x)$. They define the six trigonometric functions

$$\tan(x) = \frac{\sin(x)}{\cos(x)}, \quad \csc(x) = \frac{1}{\sin(x)}, \quad \sec(x) = \frac{1}{\cos(x)}, \quad \cot(x) = \frac{\cos(x)}{\sin(x)}.$$

These kinds of functions are crucial for modeling periodic phenomena, like the swinging of a pendulum, or population cycles, or circadian rhythms (eg, sleep-wake cycle).

A wave function is characterized by its mean, amplitude, period and phase. For example, the function

$$f(x) = M + A\cos\left(\frac{2\pi}{T}(x - \varphi)\right)$$

has:

- mean M (meaning, this is the level around which it oscillates),
- amplitude A (meaning, this is its maximum distance from the mean, in absolute value),
- period T (meaning, the graph repeats after time T; more precisely, $f(x) = f(x+T)$ for all x), and
- phase φ (meaning, the left-right displacement of the wave; here, the peak will be at $x = \varphi$ rather than the usual $x = 0$ which it would be for $\cos(x)$).

There are many excellent trigonometric identities, but the key one is

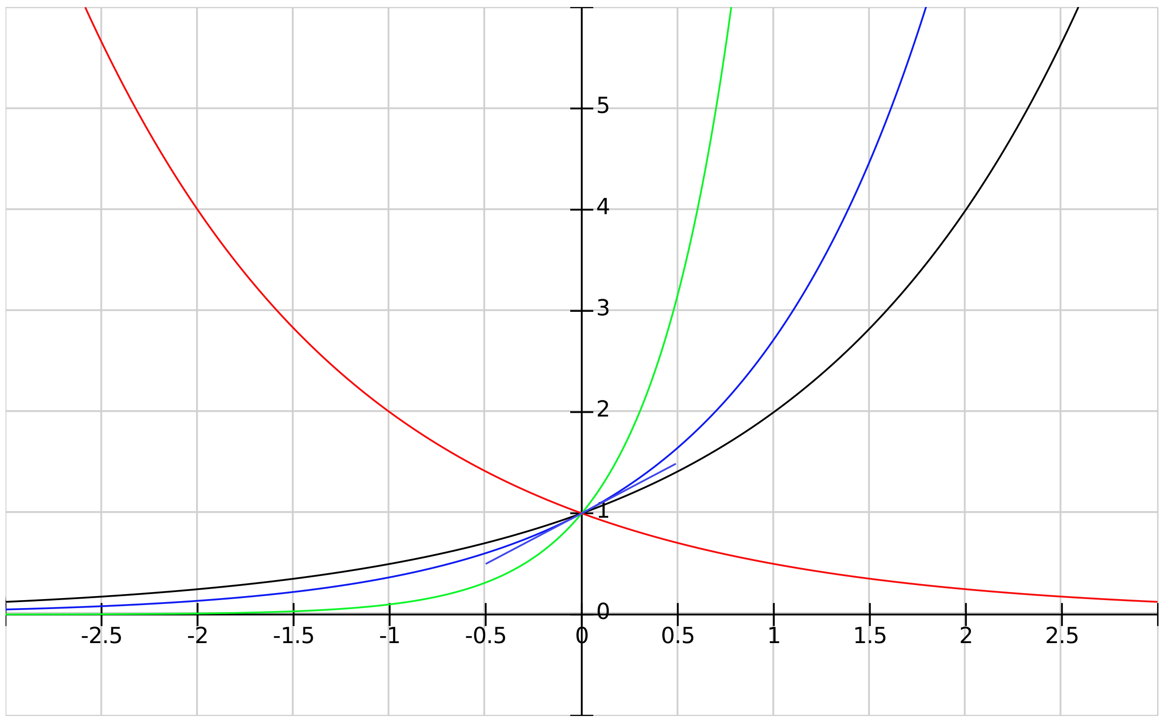
$$\sin^2(x) + \cos^2(x) = 1.$$

The next most useful thing to know: the values of $\sin(x)$ and $\cos(x)$ at $x = 0, \pi/6, \pi/3, \pi/2, 2\pi/3, \pi$.

Remember to use radians in Calculus.

1.3.7 Exponential and logarithmic functions

See also Section 1.2.1. For any number $a > 0$ we define the exponential function with base a as $f(x) = a^x$. The domain is all of $\mathbb{R}$, but its range is only the nonnegative real numbers (because $a^x < 0$ can never happen). The base $a = e \simeq 2.718...$, called Euler's number, is the distinguished base in Calculus, called the *natural base*. We sometimes write $f(x) = e^x = \exp(x)$, especially when the exponent is bulky and would be hard to write above the e.



The graphs of various exponential functions. We have $y = e^x$ in blue, $y = 2^x$ in black (note that $1 < 2 < e$), $y = \left(\frac{1}{2}\right)^x = 2^{-x}$ in red (note that $0 < \frac{1}{2} < 1$), and $y = 10^x$ in green (note that $10 > e$). A short line segment of slope one is tangent to the graph of $y = e^x$ at $x = 0$; the other exponential functions have slope $\ln(a)$ at $x = 0$.

Exponential functions arise in a huge variety of applications, including bacterial growth, radioactive decay, and continuously compounded interest. It is one of the most important functions used in biology.

The laws of exponents: for all $a > 0$ and $x, y \in \mathbb{R}$, we have:

- $a^{x+y} = a^x a^y$;
- $a^{-x} = \frac{1}{a^x}$;
- $a^{xy} = (a^x)^y$;
- $a^0 = 1$;
- $a^1 = a$.

For any $a > 0$ we also have the inverse function of the exponential, which is called the logarithm. It is defined by the relation

$$\log_a(x) = y \Leftrightarrow a^y = x$$

which is saying that these are *inverse functions*. In particular, note that the domain of the log function is only the positive part of the real line, $(0, \infty)$, but its range is all of $\mathbb{R}$.

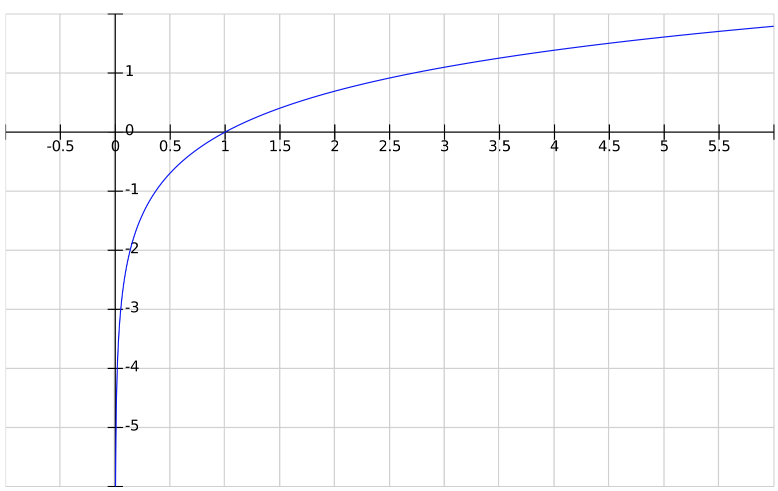
The *natural logarithm* is $\ln(x) = \log_e(x)$. In the physical sciences, $\log_{10}(x)$ is very common (such as for the pH scale); so if we write $\log(x)$ without specifying the base, we mean base 10.[3] Logarithms come up a great deal because exponential functions do, and to solve an equation like

$$300 = 200 + 5e^{x-7}$$

you need the logarithm.

The laws of logarithms (let's state them for ln, but similar rules hold for all logarithms): for every $x, y > 0$ we have

- $\ln(xy) = \ln(x) + \ln(y)$
- $\ln(x^y) = y\ln(x)$
- $\ln(e^x) = x$
- $e^{\ln(x)} = x$
- $\ln(e) = 1$
- $\ln(1) = 0$
- $\ln(0)$ is undefined (because you can't solve $e^x = 0$).



The graph of $y = \ln(x)$.

[3]In computer science, the most common is base 2, so in that subject log() means by default $\log_2(x)$.

The logarithm is a key tool when your measurements will vary by orders of magnitude (from 10^{-7} to 10^7, for example), and so it is actually the exponent you are interested in, rather than the digits of the value. In Calculus, the logarithm is a wonderful function that allows you to simplify complex expressions so they are easier to deal with. For example, for all $x > 3$ we have

$$\ln\left(\frac{x^2(x-3)}{(x+2)^7}\right) = 2\ln(x) + \ln(x-3) - 7\ln(x+2).$$

End of lecture # 2

Chapter 2

Discrete Time Dynamical Systems (DTDS)

Let's now put these functions to their intended use: mathematical modeling.

2.1 Overview

Change is what we study in science, and the life sciences are full of examples. Individuals grow and die; the size of a population varies; individuals physically move within their environment; individuals can change; wounds heal; hearts beat regularly; the immune system responds to threats; diseases spread through populations; drugs are absorbed into the bloodstream; ...

One key goal is to be able to predict future states from the present state, based on understanding the mechanisms of the change. For example, if we know how an organism's life cycle depends on the external temperature, we can predict future developments under climate change.

Experiments can sometimes tell us what the future could bring, by allowing us to extrapolate from past to future — but experiments can be costly, risky, impractical, or have large time requirements.

Mathematical models are invaluable tools to help prediction. Based on experimental data and/or our understanding of the mechanisms of change, mathematical models are used in a huge variety of applications. We use them to try to predict the weather, the stock market, and to regulate species harvesting and management.

Prediction of events with a mathematical model is a three-step process:

1. from life science to mathematics (modeling)
2. mathematical understanding (analysis)
3. from mathematics to life sciences (interpretation).

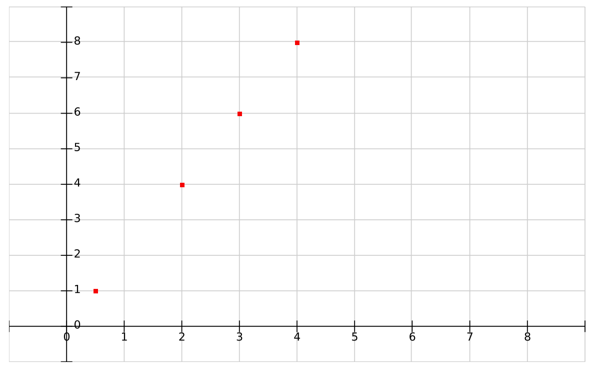
In this course, we will work through examples of doing each of the three steps, with a very large emphasis on the second step (analysis).

2.2 First examples

2.2.1 Bacterial growth

Suppose we grow bacteria in four petri dishes, measuring the amount of bacteria in cm^2 (that is, by area) before and after a 24-hour growth period. We collect the following empirical data.

Dish	Before	After
1	3	6
2	0.5	1
3	2	4
4	4	8



Plot of bacterial population, in cm^2, after a 24 hour period, as a function of the initial population, in cm^2, at the beginning of the 24 hour period.

From these data, we develop our best guess about the growth of these bacteria: The bacteria double in 24 hours.

Let's write this as a formula. If x_{today} is the area covered by bacteria today, then

$$x_{\text{tomorrow}} = 2x_{\text{today}}.$$

Excellent! This is a simple mathematical model.

Predicting one day into the future is nice, but how can we go further? Answer: repeat !

$$x_{\text{in two days}} = 2x_{\text{tomorrow}} = 2(2x_{\text{today}}) = 4x_{\text{today}}.$$

$$x_{\text{in three days}} = 2(x_{\text{in two days}}) = 8x_{\text{today}}$$

and so forth.

Let's formalize this idea, and also give ourselves a cleaner and clearer notation.

2.2.2 Definition of a DTDS

Definition 2.1. A *discrete time dynamical system (DTDS)* consists of:

- a quantity whose change is tracked at time t : x_t;

- a time step (eg, days, years — the units in which t is measured);
- an *updating function* f that describes the change during a *single time step*.

The DTDS is then

$$x_{t+1} = f(x_t).$$

Example 2.2. (Bacteria)

The updating function in our bacteria example was $f(x) = 2x$, where x is the area of bacteria and the time step is 24 hours. The DTDS is

$$x_{t+1} = f(x_t) = 2x_t.$$

□

Example 2.3. (Tree height)

Bamboo is one of the fastest-growing plants on Earth, with a rate of 3 cm/hour (!). So let us take x as the length of bamboo in cm and t the time measured in hours (so the time step is 1 hour). At the end of each hour, x has increased by 3. Therefore the updating function is $f(x) = x + 3$. The DTDS is $x_{t+1} = x_t + 3$. □

Example 2.4. (Medication)

Suppose you prescribe your patient pain medication as a daily dose. The daily dose increases the amount of drug in their body, but at over the day, the body is absorbing and eventually eliminating the drug.

Let us use a time step of 1 day, and let x_t be the amount of drug in the body in mg *just after* the daily dose is administered. What is the updating function in this case?

We think of what happens over the course of a day. If the patient began with x_t mg in their bloodstream, then over the day they eliminate a certain percentage of the drug; then just before the next measurement, a constant amount of drug is added. We could write this mathematically as

$$x_t \xrightarrow{\textit{elimination}} rx_t \xrightarrow{\textit{intake}} rx_t + c$$

where $0 \leq r < 1$ is the percentage (meaning: fraction) of drug left in the bloodstream after one day (so $1 - r$ is the rate of elimination) and $c > 0$ is the amount of drug administered.

Therefore our updating function is $f(x) = rx + c$ and the DTDS is $x_{t+1} = rx_t + c$. □

Example 2.5. In the preceding example, if a certain drug is cleared at a rate of 65% per day and the daily dose is 45mg, then $r = 1 - .65 = 0.35$ and $c = 45$, giving a DTDS of

$$x_{t+1} = 0.35x_t + 45.$$

□

Exercise 2.6. Here are some exercises for you to try.

1. Suppose instead that we measure the amount of drug in the body each day immediately *before* the daily dose. What is the DTDS in this case?

2. Suppose the initial drug level in the bloodstream was 8 mg (that is, immediately after the first dose). What is the amount of drug in the body after the next daily dose? In two days? Plot these points. In your opinion, how does the amount of drug in the bloodstream vary over the course of the day? Should we just connect the dots, or will the graph be much spikier? Does this model capture the maximum amount of drug in the bloodstream per day, or the minimum? What about if we model it as the exercise above?

3. Assume you borrow \$ 1000 and agree to pay back \$ 50 per month. The bank charges 0.5% in interest per month.[1] Write the updating function for this DTDS, where x_t is the amount that you owe immediately *after* making the tth payment, and t is measured in months.

What else does the updating function of a DTDS give you? The updating function f of a DTDS tells you what happens to your measured value from one time step to the next. Therefore it can also see several steps into the future or the past!

Mathematics		Life sciences	
composition:	$(f \circ f)(x) = f(f(x))$	two time steps:	$x_{t+2} = f(f(x_t))$
inverse:	$x = f^{-1}(y)$	previous time:	$x_t = f^{-1}(x_{t+1})$

Exercise 2.7. 1. Consider $f(x) = \frac{1}{2}x + 2$. Calculate the two-time step map: $f \circ f$. Calculate the previous time step map: f^{-1}.

2. Suppose our DTDS models bacterial growth by $x_{t+1} = 3x_t$, where x_t is measure in cm^2 and t is measured in intervals of 6 hours. Thinking of 24 hours as iterating the DTDS four times, give the DTDS for bacterial growth where t is instead measured in days (that is, each interval is 24 hours). Starting with an initial condition of 10 cm^2 of bacteria, check your answer by finding x_4 in the first (6-hour) model and finding x_1 in the second (24-hour) model.

3. Using the DTDs of the previous exercise: how much bacteria did we start with if after 12 hours we have 100 cm^2? Check by evaluating the original DTDS on the value you got.

2.3 Solutions of a DTDS

We now know what a DTDS is. What is the "solution" of a DTDS? The goal was to predict future and past events using the present (initial condition) and the short-term mechanism for change (the updating function). So one answer is: the solution of a DTDS is the sequence of all future values.

Definition 2.8. The *solution* of the DTDS $x_{t+1} = f(x_t)$ with *initial value* x_0 is the sequence

$$\{x_0, x_1, x_2, x_3, \dots\}$$

where each x_t satisfies $x_{t+1} = f(x_t)$.

Example 2.9. (Bacteria) The solution of $x_{t+1} = 2x_t$ with $x_0 = 20$ is $\{20, 40, 80, 160, \dots\}$. □

[1] This is because you have a good relationship with your bank. Your credit card would charge you closer to 1.6% per month.

Example 2.10. (Bamboo) The solution of $x_{t+1} = x_t + 3$ with $x_0 = 0$ is $\{0, 3, 6, 9, 12, \dots\}$. $\square$

Remark 2.11. So a solution is not a single number, nor is it a finite set of numbers — it is an entire sequence, consisting of infinitely many numbers. Here, we are talking about the solution to a dynamical system, rather that the solution to an equation.

So if you have written down the solution up to the 20th element of the sequence[2], then you can easily answer "What is x_t?" for each t from 0 to 20.

That said, wouldn't it be nice to have a formula for each element of the sequence? Then we wouldn't have to write a long list, but could instead answer "What is x_t?" using the formula.

2.4 A general solution formula for a DTDS with a *linear* updating function

Assume that we are given a *linear DTDS*, that is, a DTDS of the form

$$x_{t+1} = rx_t + c$$

where r and c are some constants, that is, real numbers that will not vary over time. For example, for bacterial growth, we take $r = 2$ and $c = 0$; for bamboo growth, we take $r = 0$ and $c = 3$; for medication levels, our constants will satisfy $0 \leq r < 1$ and $c > 0$, with actual values depending on the drug.

Let's calculate the solution.

$$\begin{aligned}
&x_0 \\
&x_1 = rx_0 + c \\
&x_2 = rx_1 + c = r(rx_0 + c) + c = r^2x_0 + c(r + 1) \\
&x_3 = rx_2 + c = r(r^2x_0 + c(r + 1)) + c = r^3x_0 + cr(r + 1) + c = r^3x_0 + c(r^2 + r + 1) \\
&x_4 = \cdots = r^4x_0 + cr(r^2 + r + 1) + c = r^4x_0 + c(r^3 + r^2 + r + 1)
\end{aligned}$$

Now we see the pattern:

$$x_t = r^t x_0 + c(r^{t-1} + r^{t-2} + \cdots + r^2 + r + 1). \tag{2.1}$$

This is wonderful! To find x_t, we can use (2.1) directly, instead of iterating our updating function t times.

We can simplify further, by summing up the expression

$$r^{t-1} + \cdots + r^2 + r + 1.$$

Namely, if $r = 1$, then this is just $1 + \cdots + 1$ (t times) so equals t. Also, $r^t = 1$, so our general solution is

$$x_t = x_0 + ct \tag{2.2}$$

[2]Let's call x_0 the 0th element of the sequence, for simplicity.

On the other hand, if $r \neq 0$, then we can use the geometric series. Recall that

$$(r-1)(r^{t-1} + r^{t-2} + \cdots + r^2 + r + 1) = r^t - 1$$

(which you can check by multiplying out the left side and seeing how it cancels). Therefore, if $r \neq 1$, we can divide both sides by $r - 1$ to get the formula

$$r^{t-1} + \cdots + r^2 + r + 1 = \frac{r^t - 1}{r - 1}.$$

This gives the general formula for the solution to be

$$x_t = r^t x_0 + c\frac{r^t - 1}{r - 1}. \tag{2.3}$$

Actually, we can simplify this a bit further, which will be very useful next class. Start by noticing that

$$\frac{r^t - 1}{r - 1} = \frac{r^t}{r - 1} - \frac{1}{r - 1} = -\frac{r^t}{1 - r} + \frac{1}{1 - r},$$

(where used $(r - 1) = -(1 - r)$), so that

$$x_t = r^t x_0 - c\frac{r^t}{1 - r} + c\frac{1}{1 - r} = r^t(x_0 - \frac{c}{1 - r}) + \frac{c}{1 - r}.$$

So if we set

$$x^* = \frac{c}{1 - r} \tag{2.4}$$

then (2.3) becomes

$$x_t = r^t(x_0 - x^*) + x^*.$$

We can summarize what we have found by writing it as a theorem.

Theorem 2.12. *Let $x_{t+1} = rx_t + c$ be a linear DTDS, with initial condition x_0. Then the general solution formula is*

- $x_t = x_0 + ct$, *if* $r = 1$, *and*
- $x_t = r^t(x_0 - x^*) + x^*$ *if* $r \neq 1$.

We call x^* the fixed point *or* equilibrium point *of this linear DTDS*[3].

End of lecture # 3

Example 2.13. Suppose $x_{t+1} = \frac{1}{2}x_t + 2$. Then $r = \frac{1}{2} \neq 1$, and $c = 2$, so $x^* = \frac{2}{1-\frac{1}{2}} = 4$. Therefore the general solution to the DTDS is

$$x_t = \left(\frac{1}{2}\right)^t (x_0 - x^*) + x^* = \left(\frac{1}{2}\right)^t (x_0 - 4) + 4.$$

[3]See the next section for a more general definition of an equilibrium point, for any DTDS.

For example, if the initial condition is $x_0 = 10$ then the first few terms of the solution are

$$\{10, 7, 5.5, 4.75, \cdots\}$$

which (check!) we can also get by plugging $t = 0, 1, 2, 3$ into the formula

$$x_t = \frac{1}{2^t}(10 - 4) + 4 = \frac{6}{2^t} + 4.$$

□

Example 2.14. (Bacterial model) The DTDS was $x_{t+1} = 2x_t = 2x_t + 0$ so it is a linear DTDS with $r = 2$ and $c = 0$. This gives $x^* = 0/(1-2) = 0/(-1) = 0$, and general solution formula

$$x_t = 2^t(x_0 - 0) + 0 = 2^t x_0.$$

We are greatly relieved to see this is the general solution, as it was exactly what we expected to see. □

Take away messages:

- There is a general solution formula for any *linear* DTDS, given by Theorem 2.12.
- When you plug in values for r, c and x_0, it gives you a formula for x_t as a function of t.
- We derived this formula by iterating until we saw a pattern, and then using the *geometric series identity*:

$$r^{n-1} + r^{n-2} + \cdots + r^2 + r + 1 = \frac{r^n - 1}{r - 1} \quad \text{for any } r \neq 1.$$

  Therefore, if we forget the theorem (eg on a test), we can always figure it out again.

2.5 Behaviour of general DTDS: cobwebbing

So we defined a DTDS as a system of the form $x_{t+1} = f(x_t)$, where x_t is the value of the object of interest at time t, and a time step for t (and units for x_t) have been given.

When the updating function f is linear, that is, $f(x) = ax + b$ for some constants a and b, then we found an *explicit solution*, that is, a formula that immediately tells us the value of x_t for every t, without having to iterate through all the preceding values. (See Example 2.13 for instance.)

Our goal now: Visualize the behaviour of solutions, even if f is not linear.

Method 1: Calculate several terms of the solution, and plot them

First, a familiar example:

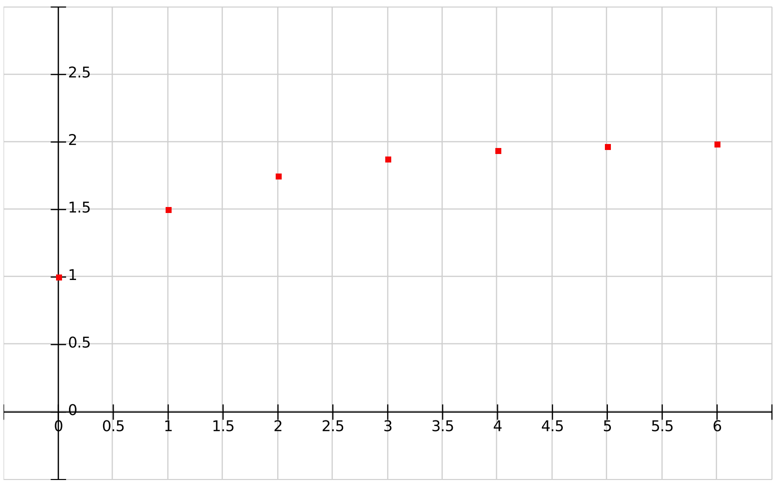
Example 2.15. Suppose our DTDS is $x_{t+1} = \frac{1}{2}x_t + 1$, and the initial condition is $x_0 = 1$.

To make this concrete, we could say that x_t is the concentration of a drug in a patient's bloodstream at time t, given that the patient receives a daily dose that counteracts the natural elimination of the drug. So t is measure in units of days and x_t is the concentration right after the daily dose.

We calculate:

$$
\begin{aligned}
x_1 &= .5(1) + 1 = 1.5 \\
x_2 &= .5(1.5) + 1 = 1.75 \\
x_3 &= .5(1.75) + 1 = 1.875 \\
\ldots
\end{aligned}
$$

And we plot these on a graph of t versus x_t:



Plot of x_t versus t for $t = 0, 1, 2, \ldots, 6$, for the DTDS $x_{t+1} = \frac{1}{2}x_t + 1$.

Note: we *only* plot the values for $= 0, 1, 2, \ldots$, not the times in between.
Our discrete time dynamical systems are often models about very specific times (eg, just after a drug dose) and not for all the times in between.

This plot shows us that the concentration of the drug is increasing over time, but that this concentration seems to be levelling off to about $x^* = 2$.

(Check for yourself that this fits perfectly with the general solution formula we derived in Theorem 2.12: since $c = 1$ and $r = \frac{1}{2}$, $x^* = 1/(1 - \frac{1}{2}) = 2$ and so $x_t = (0.5)^t(x_0 - 2) + 2$.) □

Now for a new example:

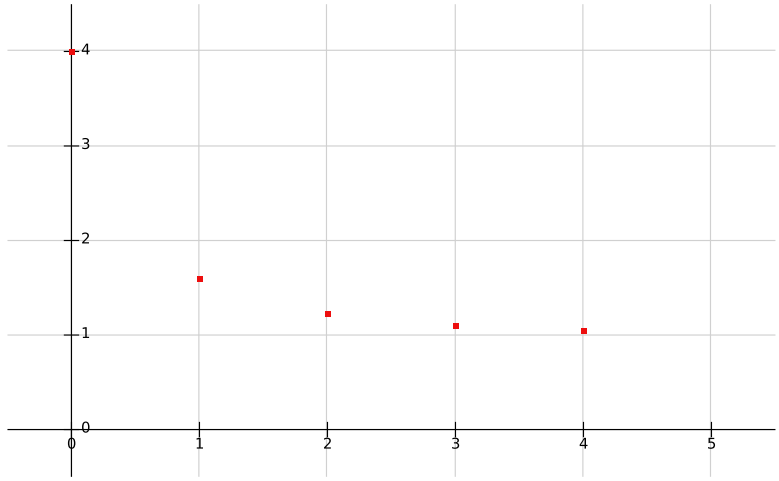
Example 2.16. Suppose our DTDS is $x_{t+1} = \dfrac{2x_t}{1 + x_t}$ and our initial condition is $x_0 = 4$. (This is an example of a *limited population model*. It expresses that the birth rate is inversely proportional to 1+ current population, an effect often seen as individuals compete for resources.[4])

[4]For a similar example with more realistic numbers, see Example 3.3.3 in the book.

We calculate:

$$x_1 = \frac{8}{5} = 1.6$$
$$x_2 = \frac{3.2}{2.6} \simeq 1.23$$
$$x_3 \simeq 1.10$$
$$x_4 \simeq 1.05$$

(Careful: you have to carry many more digits in your intermediate answers, or else the rounding error on your results will be significant.)



Plot of x_t versus t for $t = 0, 1, 2, 3, 4$, for the DTDS $x_{t+1} = \dfrac{2x_t}{1 + x_t}$.

□

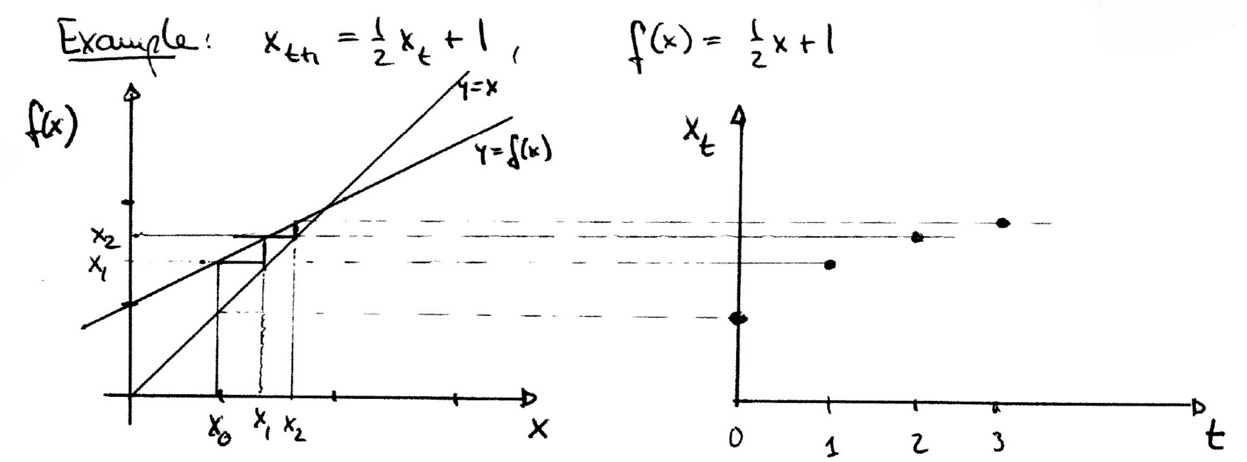
This method is tedious and prone to round-off errors. Good for computers, but not for us.

Method 2: Cobwebbing

Here is the *algorithm* (or recipe) for cobwebbing:

1) Graph the updating function $y = f(x)$ and also the diagonal line $y = x$.
2) Start with x_0 on the x-axis and go vertically to $x_1 = f(x_0)$.
3) Go horizontally to the diagonal.
4) Repeat steps 2) and 3).
5) Draw the solution as the points (t, x_t) on a graph of t versus x, if required.

Example 2.17. Let's apply this algorithm to our familiar example of a linear DTDS.



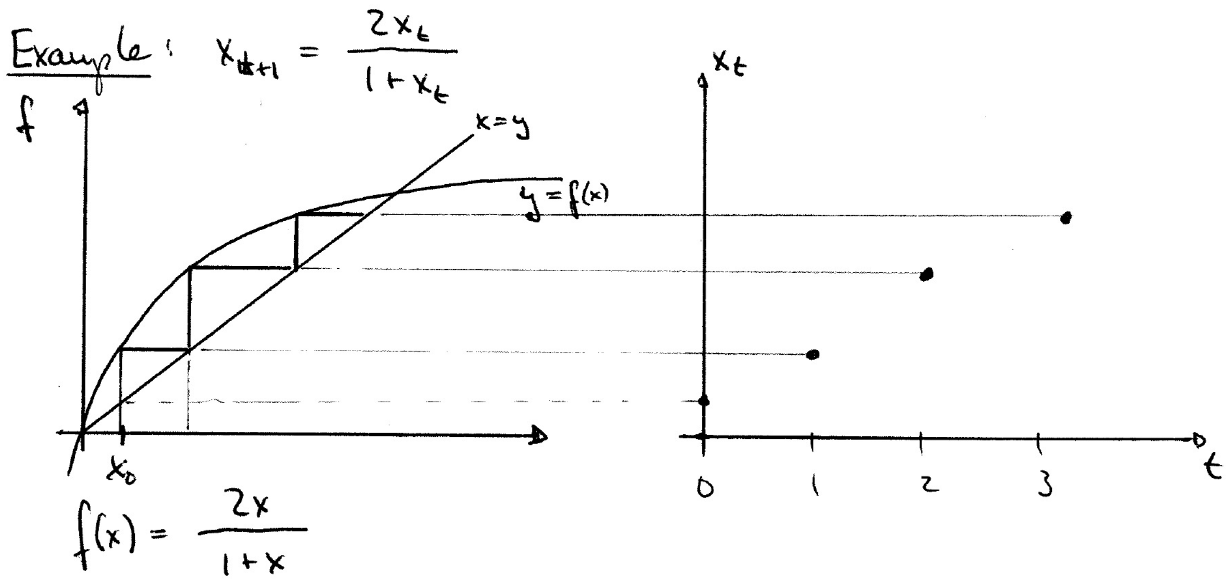
An example of a cobweb for the DTDS $x_{t+1} = \frac{1}{2}x_t + 1$. The graph on the left shows the cobweb diagram, starting from some point x_0; the corners along the graph of the updating function f give the values x_1, x_2, ... To graph the general solution, we plot $(0, x_0)$, $(1, x_1)$, $(2, x_2)$, ... in a graph of t versus x_t, at right.

The solution (at right) is the same as the graph plotted in Example 2.15. □

Example 2.18. Now let's do Example 2.16 using cobwebbing. The updating function is

$$f(x) = \frac{2x}{1+x}$$

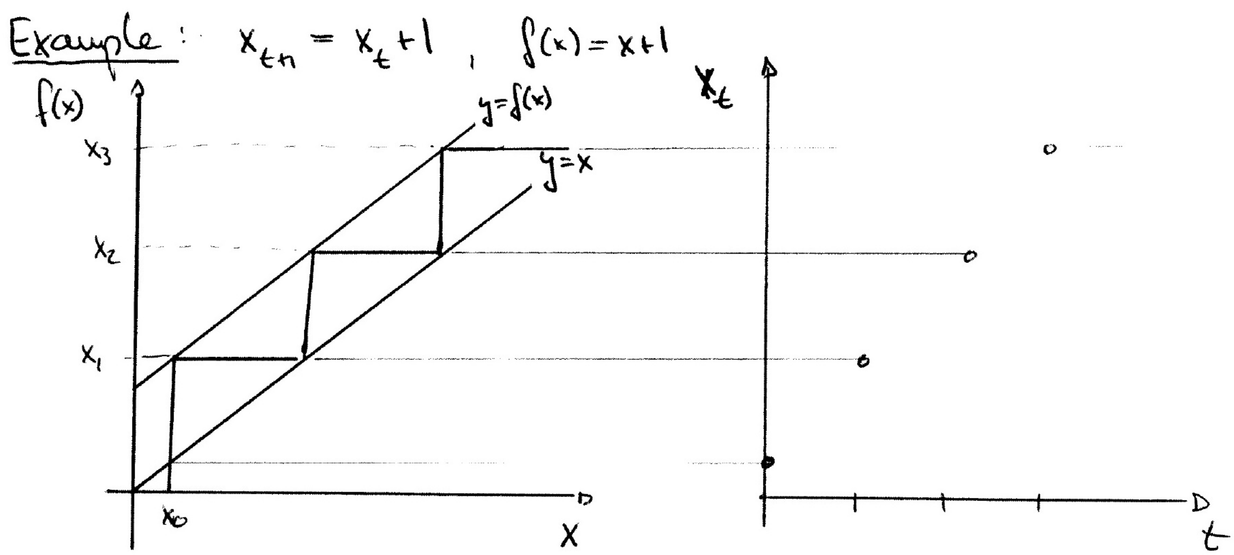
which we can graph using the tools of Calculus. We will remind ourselves of the techniques for doing so later; for now, note that its graph starts at $(0, 0)$ and has a horizontal asymptote at $y = 2$.



An example of a cobweb for the DTDS $x_{t+1} = \dfrac{2x_t}{1+x_t}$. The graph on the left shows the cobweb diagram, starting from some point x_0; the corners along the graph of the updating function f give the values x_1, x_2, ... To graph the general solution, we plot $(0, x_0)$, $(1, x_1)$, $(2, x_2)$, ... in a graph of t versus x_t, at right.

The solution (at right) is the same as the graph plotted in Example 2.16. □

Example 2.19. Another linear DTDS:



An example of a cobweb for the DTDS $x_{t+1} = \dfrac{2x_t}{1+x_t}$. The graph on the left shows the cobweb diagram, starting from some point x_0; the corners along the graph of the updating function f give the values x_1, x_2, ... To graph the general solution, we plot $(0, x_0)$, $(1, x_1)$, $(2, x_2)$, ... in a graph of t versus x_t, at right.

□

This geometric approach gives us a lot more qualitative information about the DTDS; plugging values into the updating function repeatedly gave us more quantitaive information.[5] Both are doing the same thing: iterating f.

In the next sections, we infer properties of the DTDS and its solutions from these cobweb diagrams.

2.6 Steady states

Suppose that our initial value x_0 happens to be where $y = f(x)$ intersects the diagonal line $y = x$. We do the cobweb algorithm... but nothing happens! We stay put.

These points are very important, both mathematically and for the application we are modeling, so they get a special name (actually, many special names for the same thing).

Definition 2.20. A point x^* is called a *steady state* (or *fixed point* or *equilibrium*[6]) of the DTDS $x_{t+1} = f(x_t)$ if

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

So if $x_0 = x^*$ then $f(x_0) = x_0$ so iterating the updating function f gives the same value over and

[5]Motivation #1 for Calculus: How can we sketch updating functions and thereby do cobwebbing, rather than using a calculator and being surprised by how the numbers turn out? Answer: Calculus tells you the shapes of curves.

[6]Plural of equilibrium: equilibria.

over: $x_1 = f(x_0) = x_0$, $x_2 = f(x_1) = f(x_0) = x_0$.... in other words, the dynamics will not change over time.

2.6.1 Examples: finding equilibria

Example 2.21. Suppose $x_{t+1} = \frac{1}{2}x_t + 1$. Find all the equilibria.

Solution: The updating function is $f(x) = \frac{1}{2}x + 1$. A steady state is a solution to $x^* = f(x^*)$, in other words, to

$$x^* = \frac{1}{2}x^* + 1.$$

Subtract $\frac{1}{2}x^*$ from both sides to get

$$\frac{1}{2}x^* = 1 \quad \Leftrightarrow \quad x^* = 2.$$

So there is only one steady state, and it is $x^* = 2$.

We check our answer: $f(2) = \frac{1}{2}(2) + 1 = 1 + 1 = 2$; yes, it's a fixed point of the DTDS. □

Remark 2.22. That is very good news: this is the same point $x^* = \frac{c}{1-r} = \frac{1}{1-\frac{1}{2}} = 2$ that we had earlier called the "fixed point".

Example 2.23. Suppose $x_{t+1} = \dfrac{2x_t}{1+x_t}$. Find all steady states of this DTDS.

Solution: The updating function is $f(x) = \frac{2x}{1+x}$. A steady state is a solution to the equation $x^* = f(x^*)$. In this case, it means solving:

$$\begin{aligned} & x^* = \frac{2x^*}{1+x^*} \\ \Leftrightarrow \quad & x^*(1+x^*) = 2x^* \\ \Leftrightarrow \quad & x^* + (x^*)^2 = 2x^* \\ \Leftrightarrow \quad & (x^*)^2 - x^* = 0 \\ \Leftrightarrow \quad & x^*(x^*-1) = 0. \end{aligned}$$

Therefore this DTDS has exactly two steady states: $x^* = 0$ and $x^* = 1$.

We check: $f(0) = \frac{0}{1} = 0$ and $f(1) = \frac{2}{1+1} = 1$. Yes! □

Example 2.24. Suppose $x_{t+1} = x_t + 1$. Find all fixed points.

Solution: The updating function is $f(x) = x + 1$. A steady state is a solution to the equation $x^* = f(x^*)$. Here, this means solving

$$x^* = x^* + 1$$

but this equation has no solution. Therefore this DTDS does not have any equilibria. □

Exercise 2.25. Compare the answers found here with the cobweb graphs of the previous section, to agree that we have found all the equilibria.

Remark 2.26. When solving for steady states, always sketch the graph, so that you can see how many solutions to expect.

Example 2.27. Let's consider a general linear DTDS of the form $x_{t+1} = rx_t + c$ with r, c real constants and $r \neq 1$. A fixed point is a solution to

$$x^* = rx^* + c \quad \Leftrightarrow \quad x^*(1-r) = c \quad \Leftrightarrow \quad x^* = \frac{c}{1-r}$$

which is the same formula as what we used in Theorem 2.12. □

2.6.2 Stability of equilibria

Our second observation from our cobweb examples is that sometimes our solutions approach an equilibrium, and sometimes they move away from an equilibrium.

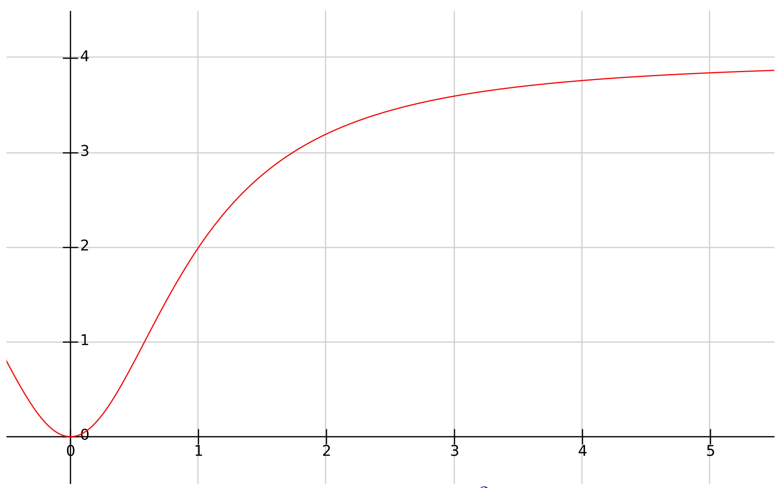
To explore this phenomenon more fully, let's consider a richer example: a population exhibiting the Allee effect[7].

Example 2.28. (Population showing Allee effect)

Consider the DTDS

$$x_{t+1} = \frac{4x_t^2}{1+x_t^2}.$$

The updating function is $f(x) = \frac{4x^2}{1+x^2}$ and the graph of the updating function is characterized by an S-shaped curve (as below).



A plot of $y = \dfrac{4x^2}{1+x^2}$.

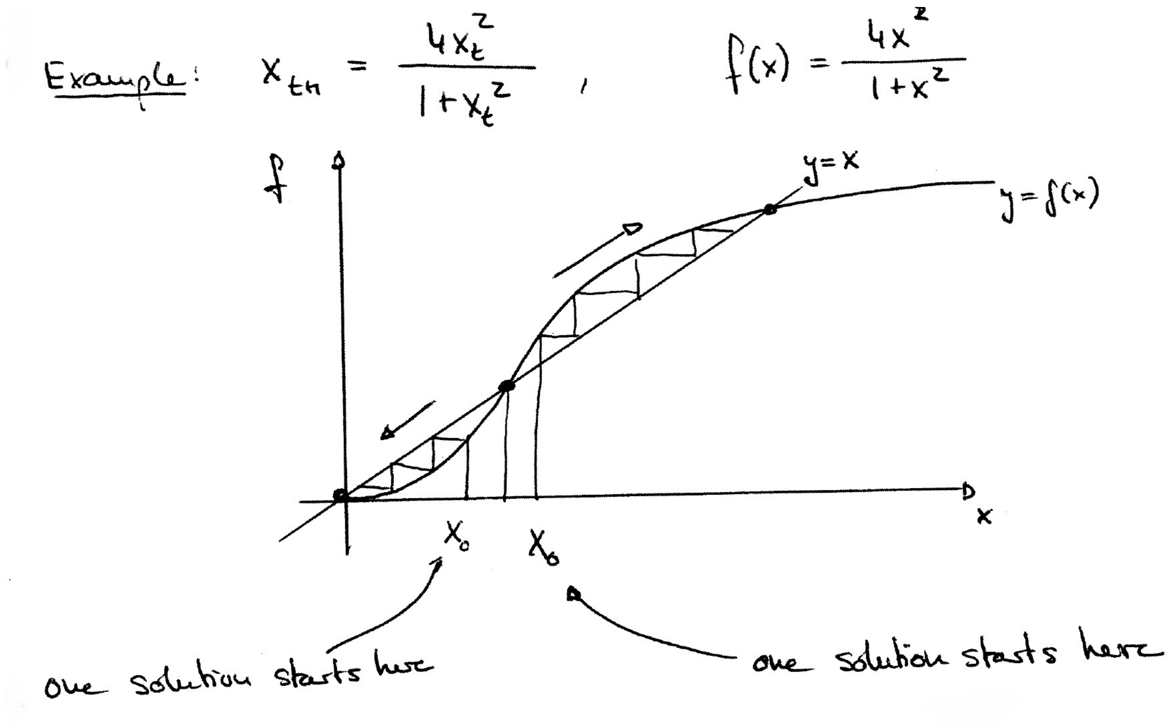
If this DTDS represents a population, then what this updating represents is:

- If the population is low, then the scarcity of mates and general low fitness of the population implies a low reproductive rate.

[7] From Wikipedia: "The Allee effect is a phenomenon in biology characterized by a correlation between population size or density and the mean individual fitness (often measured as per capita population growth rate) of a population or species."

- If the population is large enough, then the reproductive rate is high.
- If the population is too high, then the limited resources reduce the reproductive rate.

Let us apply the cobweb algorithm to some different initial values x_0 and see what happens.



A cobweb diagram applied to the updating function $f(x) = \dfrac{4x^2}{1+x^2}$.

We see that the middle equilibrium is not approached by any cobweb diagram, but the two other fixed points are approached by cobweb diagrams from various initial states.

□

Definition 2.29. A fixed point x^* is called *stable* if all nearby initial conditions give solutions that approach x^*. A fixed point x^* is called *unstable* if there is at least one initial condition near x^* such that the solution does not approach x^*.

Example 2.30. In the preceding example, the middle fixed point is unstable. □

Example 2.31. In the linear DTDS $x_{t+1} = \frac{1}{2}x_t + 1$, the fixed point $x^* = 2$ was stable. □

Exercise 2.32. Calculate the fixed points of $x_{t+1} = \dfrac{4x_t^2}{1+x_t^2}$ explicitly. Prove that the smallest and the largest of these fixed points is stable, by drawing cobwebs starting on either side of each fixed point.

Exercise 2.33. Is the fixed point in Example 2.18 stable or unstable?

Exercise 2.34. Do a cobweb for the linear DTDS $x_{t+1} = -\frac{1}{2}x_t + 2$. Identify any fixed points and classify by stability.

Explore cobwebbing and graphing using the Excel file provided: `linearDTDS.xls`.

End of lecture # 4

2.7 Stability in linear models: a theorem

Recall:

- Given a DTDS $x_{t+1} = f(x_t)$, a *fixed point* is a number x^* such that $f(x^*) = x^*$. There may be none, one, or many fixed points for a given DTDS.
- There are two kinds of fixed points (=equilibria, steady states): stable and unstable.
  - A fixed point x^* is *stable* if all nearby x_0 give solutions that approach x^*;
  - x^* is *unstable* if at least one nearby initial state x_0 gives a solution that does not approach x^*.

The stability of its steady states tells us about the long-term behaviour of a DTDS.

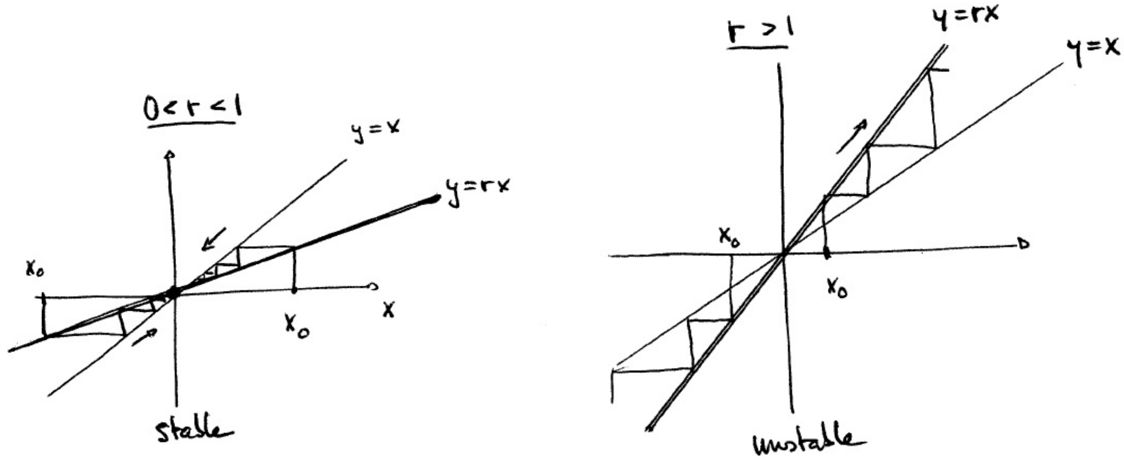
So let's figure out the stability of fixed points in linear models.[8]

Let us begin with simple examples of linear DTDS:

$$x_{t+1} = rx_t, \quad x^* = 0.$$

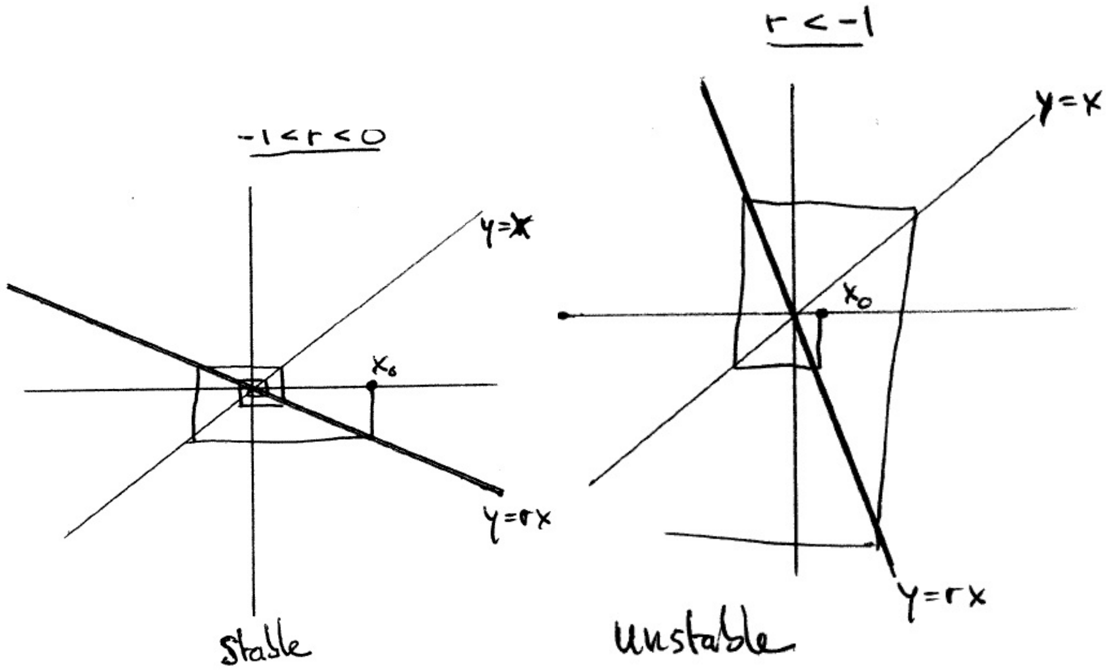
For different values of r, to decide stability we **must** draw at least two cobwebs: one for an initial condition $x_0 > x^*$ and one for an initial condition $x_0 < x^*$. The fixed point is stable only if **all** nearby initial states give solutions that approach it.

[8] Why linear models? First: because linear updating functions give DTDS for which we have a general solution formula and so can answer all questions completely. Second: we will see that this answer gives us the hint we need on how to predict stability in general.



Cobwebs applied to the updating function $y = rx$, with two initial conditions each: $x_0 < 0$ and $x_0 > 0$. At left, the slope of the updating function is $0 < r < 1$; at right, the slope of the updating function is $r > 1$. We see that in the former x^* is a stable fixed point; in the latter, x^* is an unstable fixed point.

What if $r < 0$?



Cobwebs applied to the updating function $y = rx$, with two initial conditions each: $x_0 < 0$ and $x_0 > 0$. At left, the slope of the updating function is $-1 < r < 0$; at right, the slope of the updating function is $r < -1$. We see that in the former x^* is a stable fixed point; in the latter, x^* is an unstable fixed point.

Exercise 2.35. Show that if $r = 0$ then the fixed point is stable. Decide if this makes sense, given that the DTDS is $x_{t+1} = 0x_t = 0$.

Exercise 2.36. Examine the stability of the fixed points when $r = 1$ or $r = -1$. Argue (making reference to the DTDS, and to the kinds of cobwebs you obtain) that these two borderline cases are not biologically-relevant — they are just too extreme to occur in nature.

Well, that pattern seems clear enough: the fixed point is stable if $-1 < r < 1$ and is unstable if $|r| > 1$. In fact, it is true even with a general linear updating function; let's see why.

Theorem 2.37. *Let $x_{t+1} = rx_t + c$ be a linear DTDS with $r \neq \pm 1$. Then the fixed point*

$$x^* = \frac{c}{1-r}$$

is stable if $|r| < 1$ and is unstable if $|r| > 1$.

Why is this true? Remember our general solution formula:

$$x_t = r^t(x_0 - x^*) + x^*.$$

If $|r| < 1$ then r^t gets smaller and smaller, approaching[9] 0 as $t \to \infty$. So as $t \to \infty$, $x_t \to 0(x_0 - x^*) + x^* = x^*$. So x^* is stable (and it doesn't even matter where our initial condition x_0 was, or the value of c!).

If $|r| > 1$, then the powers of r will grow (in absolute value!), and in fact $|r^t| \to \infty$ as $t \to \infty$. In particular, $r^t \not\to 0$, so $x_t \not\to x^*$, so the fixed point is unstable. □

Example 2.38. Consider the medication model $x_{t+1} = \frac{1}{2}x_t + 1$, with constant dose 1. The fixed point is

$$x^* = \frac{c}{1-r} = \frac{1}{1-\frac{1}{2}} = 2.$$

Since $|r| = \frac{1}{2} < 1$, this fixed point is stable. Therefore, if we continue this regular daily dose, the concentration in the bloodstream will stabilize to close to $x^* = 2$. □

Example 2.39. Suppose we want to change the daily dose so that the steady state is $x^* = 3$. Your first guess might be: add 1 to the daily dose. Let's try that:

$$x_{t+1} = \frac{1}{2}x_t + 2 \quad \Rightarrow \quad x^* = \frac{2}{1-\frac{1}{2}} = 4.$$

Oops! No, that was not the correct approach, because we forgot that some of the extra daily dose is being kept in the system from one day to the next (and thus we overdosed our patient).

Solution: we want to choose c so that the linear DTDS $x_{t+1} = \frac{1}{2}x_t + c$ has a steady state of $x^* = 3$. So we solve

$$\frac{c}{1-\frac{1}{2}} = 3 \quad \Leftrightarrow \quad c = \frac{3}{2}.$$

So our answer is: the daily dose should be increased to 1.5 from 1 to achieve a steady state of $x^* = 3$. □

Exercise 2.40. Suppose the DTDS for a different drug is $x_{t+1} = \frac{2}{3}x_t + 4$. Find x^* and explain why it is stable, using two different arguments. (Hint: cobweb, theorem). Now suppose we instead want a steady state of $x^* = 10$. What should the new daily dosage be?

[9] Motivation #2 for Calculus: what does it mean that $r^t \to 0$ as $t \to \infty$? How could we decide this if the expression were more complicated? Answer: limits.

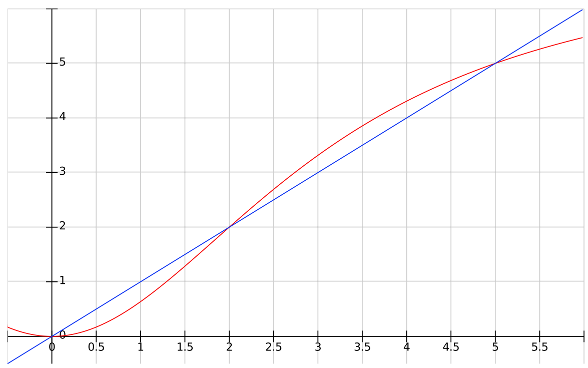
2.8 Stability in nonlinear models: examples

Nonlinear models are very common in nature. Our eventual goal[10] is a simple criterion that could tell us (mathematically) whether a given fixed point is stable or not.

In Life Sciences: only stable states are visible in nature! (You would have to conduct a controlled experiment to observe the initial few solutions, given any biological process.) So the long term behaviour is about the steady states.

Example 2.41. (Allee effect; see Example 2.28)

$$x_{t+1} = \frac{0.7x_t^2}{1+0.1x_t^2}$$



The graph of the updating function $f(x) = 0.7x^2/(1+0.1x^2)$ of a DTDS displaying the Allee effect, in red. The diagonal $y = x$ is in blue. The fixed points are the solutions to $f(x^*) = x^*$, which are 0, 2 and 5; these are the intersections of the two graphs.

When we cobweb on the updating function $f(x) = \frac{0.7x^2}{1+0.1x^2}$, we see that the fixed points 0 and 5 are stable, whereas the middle fixed point, at $x^* = 2$, is unstable.

See also the Excel file AlleeDTDS.xls, to vary the parameters and see the effect on the graph of the updating function and on the long-term behaviour. □

We make some observations from the examples we have seen so far:

- When the graph of f crosses the diagonal from below to above, then x^* is unstable.
- When the graph of f crosses from above to below with positive slope), then x^* is stable.

Actually, here we see Motivation #3 for Calculus: we'd love to talk about and calculate the slope of f at x^*, and compare it to 1.

Let's now do several more in-depth examples.

Example 2.42. (Alcohol absorption dynamics)

[10] Motivation #3 for Calculus!

In the homework, we used a simple model for alcohol elimination in the bloodstream. In reality, the rate at which a person's body absorbs or eliminates alcohol in the bloodstream depends on the alcohol level: the more alcohol in the body, the smaller the *fraction* that can be absorbed and eliminated.

Let t be time in hours, and c_t the concentration of alcohol in the blood at time t. (The units in this model are such that $c = 7$ corresponds to one drink for an average size person.)

First model: pure absorption (no new alcohol added to the body)

$$c_{t+1} = c_t - r(c_t)c_t$$

where $0 < r(c_t) < 1$ is the fraction absorbed over one unit of time, which depends on c_t.

Using empirical data, we establish that

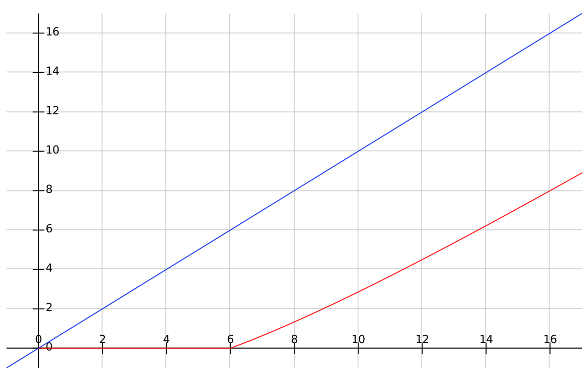
$$r(c) = \frac{10}{4+c} \quad \text{when } c > 6$$

(and $r(c) = 1$ for $c < 6$, which says that a body can completely eliminate the alcohol in one hour if the concentration is below 6 at the beginning of the hour).

Therefore, our model is

$$c_{t+1} = \left(1 - \frac{10}{4+c_t}\right)c_t,$$

for $c_t > 6$. The graph of the updating function $f(x) = (1 - \frac{10}{4+x})x = \frac{x(x-6)}{x+4}$, for $x > 6$, and $f(x) = 0$ for $x < 6$ is given below.



The graph of the updating function $f(x) = \frac{x(x-6)}{x+4}$ (for $x > 6$, and $f(x) = 0$ for $x < 6$) of a DTDS for alcohol elimination, in red. The diagonal $y = x$ is in blue. Zero is the steady state (as you can see by cobwebbing).

We only sketched part of the graph; are we sure there isn't another fixed point way off the graph? (And what would it mean?) So we check, for $x > 6$:

$$x = f(x) \quad \Leftrightarrow \quad x = \frac{x(x-6)}{x+4}.$$

So if $x > 6$, $x \neq 0$ so we can divide by x to conclude

$$x + 4 = x - 6$$

which has no solution. Therefore there are no fixed points with $x > 6$ (and only the obvious fixed point $x = 0$ in the region $x \le 6$).

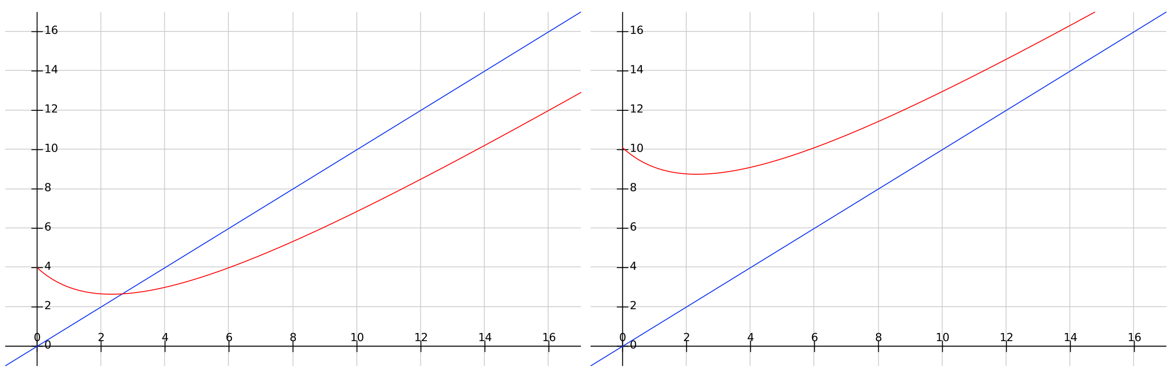
Second model: absorption plus drinking. Assume subject raises their blood alcohol level by d each hour through drinking. Then we have

$$\begin{aligned} c_{t+1} &= c_t - r(c_t)c_t + d \\ &= \left(1 - \frac{10}{4 + c_t}\right) c_t + d = f(c_t). \end{aligned}$$

Let's solve for the steady states, which we call c^*, that is, we solve $f(c^*) = c^*$:

$$\begin{aligned} c^* &= c^* - r(c^*)c^* + d \\ \Leftrightarrow \quad r(c^*)c^* &= d \\ \Leftrightarrow \quad \frac{10c^*}{4 + c^*} &= d \\ \Leftrightarrow \quad 10c^* &= 4d + c^*d \\ \Leftrightarrow \quad (10 - d)c^* &= 4d \\ \Leftrightarrow \quad c^* &= \frac{4d}{10 - d}. \end{aligned}$$

What a strange answer! As with all Life Science applications, it is helpful to ask ourselves: when is this positive, and when is it negative?



The graphs of the updating function $f(x) = x - r(x)x + d)$ of a DTDS for alcohol elimination with drinking, in red; on the left, $d < 10$ and on the right, $d > 10$. The diagonal $y = x$ is in blue. By cobwebbing, we see that the fixed point in the case $d < 10$ is stable. When $d > 10$, the fixed point is negative and is unstable.

We see that the steady state is only biologically meaningful when $d < 10$, in which case $c^* > 0$, a steady level of alcohol in the bloodstream in the long run. We conclude that the body remains intoxicated, but at some steady level (which is not d, but rather $4d/(10 - d)$).

When $d > 10$, we see that the fixed point is negative; doing cobwebbing, we see that it is also unstable and that the concentration of alcohol over time climbs without bound (until death).

You can experiment with this model (changing the parameter d and the initial value) in the excel file provided : AlcoholDTDS.xls.

□

End of lecture # 5

Let's contrast this with another drink: coffee.

Example 2.43. (Caffeine absorption)

Caffeine absorption/elimination is essentially independent of concentration. Let t be time in hours, c_t the concentration of caffeine in the body, and d the amount of caffeine concentration added to the body each hour. Then the DTDS is

$$c_{t+1} = 0.87c_t + d.$$

We solve for the steady state:

$$c^* = 0.87c^* + d \quad \Rightarrow \quad c^* = \frac{d}{1 - 0.87} = \frac{d}{0.13}.$$

This steady state always exists (and is meaningful). This is a linear DTDS, and the slope $r = 0.87$ satisfies $|r| < 1$, so the fixed point is stable.

Our conclusion: the more coffee one drinks per unit time, the greater the concentration in the body (as d increases, the steady state increases); but at a constant rate of drinking coffee, the level of caffeine in the body levels off and stabilizes. □

Now let's consider a famous population model: logistic growth. It captures the phenomenon that reproductive rate can decline with increasing population.

Example 2.44. (The logistic equation)

Let t represent time in years and x_t the population at time t, normalized so that 1 represents the maximum population that the resources can sustain. Then the logistic DTDS is

$$x_{t+1} = rx_t(1 - x_t) \quad \text{for some } 0 < r < 4$$

where the per capita growth rate $\frac{x_{t+1}}{x_t}$ is proportional to $1 - x_t$ with factor r. This means that the rate of growth declines with the density of the population, due to intraspecific[11] competition for resources.

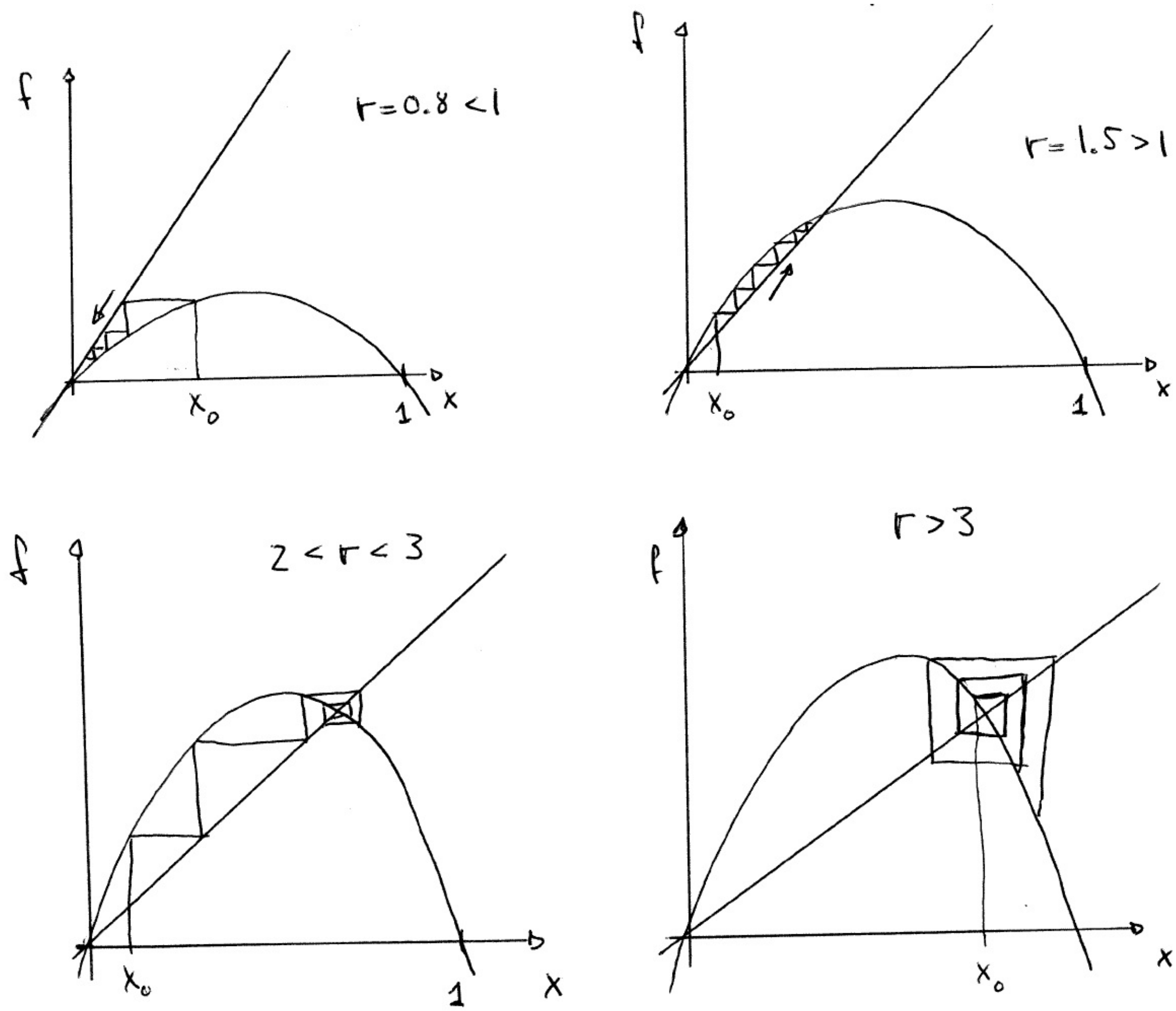
Let's solve for the fixed points, then determine their stability using cobwebbing. We have $f(x) = rx(1 - x)$ so a fixed point x^* satisfies

$$x^* = rx^*(1 - x^*) \quad \Leftrightarrow \quad r(x^*)^2 + (1 - r)x^* = 0 \quad \Leftrightarrow \quad x^*(rx^* + 1 - r) = 0$$

so we have two fixed points: $x^* = 0$ and $x^* = \frac{r-1}{r}$.

Now to set up the cobweb, we need to choose r; and it turns out that the behaviour is very different depending on the value of r!

[11]intraspecific: between individuals of a single species



Cobweb diagrams for the logistic equation with $0 < r < 1$, $1 < r < 2$, $2 < r < 3$ and $3 < r < 4$.

What we conclude is:

- if $0 < r < 1$, then 0 is the only nonnegative equilibrium, and it is stable : r is too small and the population dies out;
- if $1 < r < 3$, then there is a positive equilibrium, and it is stable, though if $r < 2$ the population climbs up to the steady state and if $r > 2$ it fluctuates around the steady state until it stabilizes;
- if $r > 4$, then the population fluctuates more and more wildly, in boom and bust cycles, until it dies out; the equilibrium is unstable.

Compare this to the linear model we analysed in Section 2.7, to notice that the stability is similarly related to the "slope of the tangent line of f at x^*".

Try it out yourself: LogisticDTDS.xls

□

The variety of interesting applications of DTDS to the life sciences is huge. Check out more examples in the textbook. Of particular interest is a sophisticated model of the heartbeat, using a discontinuous updating function that lets one understand arrhythmia from an electrical viewpoint (also available on pages 25–27 of Professor Lutscher's handwritten lecture notes).

Chapter 3

Limits and the path to Calculus

The most important scientific discovery of the second millenium was the discovery of Calculus: it changed natural philosophers into scientists: able to quantify not only the observations about the *state* of matter, but also about its *change*.

The breakthrough result was the understanding of the concept of a limit. Isaac Newton formulated limits as a theory of *infinitesimals* — theoretical "numbers" so small that when you square them you get zero — but our modern version expresses itself as : extrapolate what $f(x)$ wants to be from what you get as an answer as you get infinitely close to your target value of x. Where this becomes Calculus is when you use your understanding of the function at hand to do so.

> Weird fact: There is no number that is "right after" or "right next to" 0. If you choose a number that's close, like 0.000000001, there are always a ton of numbers that are even closer to 0, like 0.000000000134345243098. Just like there is no largest number, there is no smallest positive number.

3.1 Limits of functions: the concept

The goal: characterize the behaviour of a function at a point where it might not be defined.

Example 3.1. (Motivating example #1) The average rate of change of a function g over an interval $[x, a]$ in its domain is the rise over the run. Thinking of a as being a fixed number, we could create a new function

$$f(x) = \frac{g(x) - g(a)}{x - a} \quad \text{if } x \neq a$$

which is the slope of all the possible secant lines through $(a, g(a))$. The instantaneous rate of change of g at the point x would be obtained by choosing x to be equal to a (but $f(a)$ is illegal, involving division by 0), or "right next to a" (there is no such x). □

Example 3.2. (Motivating example #2) In our study of DTDS, we wanted to know the long term behaviour of the general solution, which is a function of the variable t (like $x_t = 4(\frac{1}{3})^t - 6$). "Long term behaviour" means "when t is ∞" — but ∞ is not a number, so the general solution function is not defined there; we can't "plug in" $t = \infty$. □

In both of these motivating examples, the solution is to take the limit of the function, in the first case as x goes to a (today's lecture) and in the second case as t goes to ∞ (tomorrow's lecture).

A first try. Suppose $g(x) = x^3$, and $a = 1$. Then we are trying to understand the function

$$f(x) = \frac{x^3 - 1}{x - 1} \quad x \neq 1,$$

as x approaches 1. Using a calculator, we could make the following tables of values (approaching from the left and from the right):

x	0.9	0.99	0.999
$f(x)$	2.71	2.9701	2.997...

x	1.1	1.01	1.001
$f(x)$	3.31	3.0301	3.003...

We infer that the closer x gets to 3, the closer $f(x)$ gets to 3. We formalize this idea in the following definition.

Definition 3.3. We say that the *limit of a function* f as x approaches a is equal to a number L, and we write

$$\lim_{x \to a} f(x) = L,$$

if we can make $f(x)$ as close to L as we wish by choosing x very close to a.

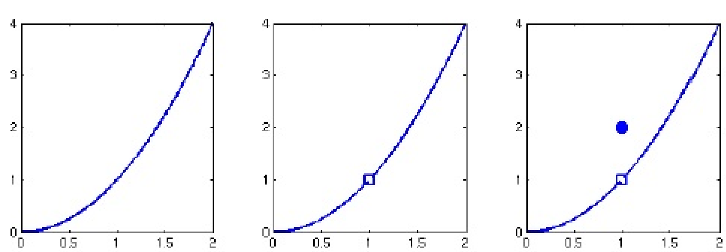
This seems to be the case in our example above, that is, we want to say

$$\lim_{x \to 1} \frac{x^3 - 1}{x - 1} = 3$$

(which is in fact true); but limits can be tricky and sometimes deceiving, so let's start slowly.

A key observation: independence of $f(a)$. The first thing to notice is that it doesn't matter what $f(a)$ is, or even if f is defined at the point a. The limit is inferring what f would *like to be* at a, not necessarily what it is.

Example 3.4. Consider three functions, whose graphs are drawn below.



The graphs of three different functions, each having the same limit as x goes to 1, despite being different at the point $x = 1$.

The first graph represents $f_1(x) = x^2$. The second represents

$$f_2(x) = \frac{x^2 - x}{x - 1} \quad x \neq 2.$$

The third represents

$$f_3(x) = \begin{cases} x^2 & \text{if } x \neq 1; \\ 2 & \text{if } x = 1. \end{cases}$$

For example, f_3 could represent the rule for winnings in a game, where the rules have an exception that when you hit 1 exactly on the nose, your winnings double.

In all three cases,

$$\lim_{x \to 1} f(x) = 1,$$

because from looking at the graph, we see that we can pick a y-value c as close to 1 as we want, and go back and find an x value b such that $f(b) = c$. This is what the definition of the limit asks us to verify. □

Another observation: disagreement possible It can happen that the function does not have a limit as x approachest a.

Example 3.5. A function may not tend toward any single number.

Consider the function

$$f(x) = \sin(\pi/x),$$

whose graph is amazing!

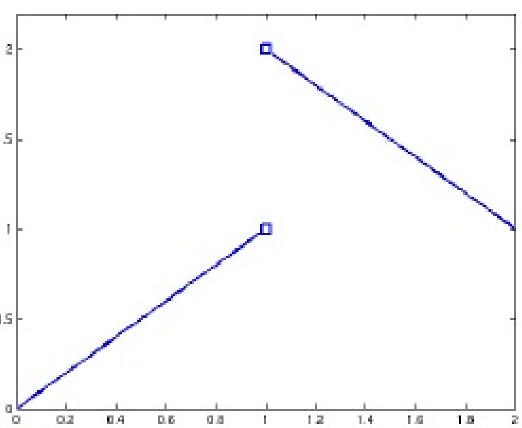
As x goes towards 0, the function oscillates more and more wildly between 0 and 1, and most certainly does not settle down to a limit. □

Example 3.6. A function may look different whether you approach from the left or the right.

Consider the function

$$f(x) = \begin{cases} x & \text{if } x < 1; \\ 3 - x & \text{if } x > 1. \end{cases}$$

whose graph is:



The graph of $y = f(x)$.

As we approach 1 from the left, our y-values are approaching 1. But as we approach 1 from the right, our y-values are approaching 2. Again, the function has not settled on a value, so the limit does not exist. □

This last example suggests we might sometimes do well to also consider one-sided limits.

Definition 3.7. We say that the limit of a function f as x approaches a *from above* (or *from the right*) is equal to the number L, and write

$$\lim_{x \to a^+} f(x) = L$$

if we can make $f(x)$ as close to L as we wish be choosing x very close to a and larger than a. Similarly, we say that the limit of a function f as x approaches a *from below* (or *from the left*) is equal to the number L, and write

$$\lim_{x \to a^-} f(x) = L$$

if we can make $f(x)$ as close to L as we wish be choosing x very close to a and smaller than a. These limits are called the *one-sided limits* of the function f at a; when needed we call them the *right-hand limit* and *left-hand limit*, respectively.

One-sided limits are also the correct thing to consider when the function is only defined on one side of the point a.

Example 3.8. Consider $f(x) = \sqrt{x}$, whose domain of definition is the interval $[0, \infty)$. In this case, we cannot ask about $\lim_{x \to 0^-} f(x)$, because f is not defined on any number less than 0. But it is reasonable to ask about the right-hand limit:

$$\lim_{x \to 0^+} \sqrt{x} = 0$$

which we reason from the graph. □

So in fact we can put this together into a concrete test.

Proposition 3.9 (Existence test)**.** *We say that the limit of f as x approaches a exists if the two one-sided limits exist and*

$$\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x).$$

So for example, we would say that $\lim_{x \to 0} \sqrt{x}$ does not exist, because the left-hand limit is not defined.

On the other hand,

$$\lim_{x \to 0} \sqrt{|x|}$$

does exist, because it is defined on both sides of 0, and the limit from either side is 0.

3.2 Evaluating limits

We have three options for evaluating limits:

(a) by calculator. But this can go wrong (see below) and is generally not accepted in this course.

(b) by reading the graph (see above). But this means sketching the accurate graph, which usually involves already knowing the limit!

(c) by limit laws and algebraic manipulation. This is the most elegant method, and the only precise way, to determine limits — and is the method required in this course.

Motivation: Examples of how a calculator can fail.

Example 3.10. Consider the function

$$f(x) = \frac{\sqrt{x^6 - 25} - 5}{x^6}.$$

If we make a table of values, we get

x	1	0.1	0.01	0.001
$f(x)$	0.099	0.1	0	0

(with my calculator)

whereas in fact, as we will show later, the limit is 0.1. The problem here was the round-off error inherent to calculators. □

Example 3.11. Consider the function $f(x) = \sin(\pi/x)$ we drew the graph of earlier. If we make the following table of values

x	1	0.1	0.01	0.001
$f(x)$	0	0	0	0

we might erroneously think that the limit was 1. But the problem here was our choice of numbers x approaching 0; we could have chosen a different sequence of x-values to make a table, like

x	3	0.3	0.03	0.003
$f(x)$	0.5	-0.866	-0.866	-0.866

and think the limit was $-\sqrt{3}/2$ (!!). □

The problem with using a calculator is that we can't possibly test *every single way* that x approaches 0, and by focussing just on some numbers, we might miss the big picture completely.

3.3 Evaluating limits with limit laws and algebraic manipulations

The following at the indisputable laws of limits.

1. $\lim_{x \to a} c = c$. ("If the function doesn't depend on x, neither does its limit.")

2. $\lim_{x \to a} x = a$. ("If x goes to a, then x goes to a.")

For the rest of the laws, suppose you already know that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist (eg, from Laws #1 and #2). Then

3. $\lim_{x \to a}(f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$;
4. $\lim_{x \to a}(f(x)g(x)) = \left(\lim_{x \to a} f(x)\right)\left(\lim_{x \to a} g(x)\right)$;
5. if $\lim_{x \to a} g(x) \neq 0$ then $\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$.

Example 3.12. Let's use the limit laws to determine

$$\lim_{x \to 2}(x^3 - 3x + 5).$$

Assuming for the moment that all the limits exist, we could use the sum rule to write

$$\lim_{x \to 2}(x^3 - 3x + 5) = \lim_{x \to 2} x^3 + \lim_{x \to 2}(-3x) + \lim_{x \to 2} 5.$$

We can apply the product rule to each of $x^3 = x \times x \times x$ and $-3x = (-3) \times x$ (and apply the constant rule #1 to the limit of -3) to get

$$= \left(\lim_{x \to 2} x\right)^3 + (-3) \lim_{x \to 2} x + \lim_{x \to 2} 5.$$

Finally, applying laws #1 and #2, we see that yes, in fact, all the limits exist, so we conclude

$$= (2)^3 - 3(2) + 5 = 7.$$

Therefore $\lim_{x\to 2}(x^3 - 3x + 5) = 7$. □

What this example shows is that if your function $f(x)$ is a polynomial function, then

$$\lim_{x \to a} f(x) = f(a),$$

because you can repeatedly apply the limit laws until you're just taking the limit of x as x goes to a (and the constant limit).

Are there other functions that behave this perfectly? Certainly.

Theorem 3.13 (Direct substitution rule)**.** *If your function $f(x)$ is a polynomial, a rational function, a trigonometric function, an exponential function, a logarithm function, a root function, or any composition of these* **and** *a is in the domain of f, then $\lim_{x \to a} f(x) = f(a)$.*

Example 3.14. The limit of the rational function

$$\lim_{x \to 2} \frac{5x^4 - 3x + 4}{x^2 - 7}$$

can be evaluated by the direct substitution rule because 2 is in the domain. Therefore,

$$\lim_{x \to 2} \frac{5x^4 - 3x + 4}{x^2 - 7} = \frac{5(2)^4 - 3(2) + 4}{(2)^2 - 7} = \frac{78}{-3} = -26.$$

But we cannot use the direct substitution rule to determine its limit as $x \to \sqrt{7}$ or $x \to -\sqrt{7}$, because these values are not in the domain. □

This rule is very helpful in many cases, but the real cases of interest are those for which the rule cannot be applied (like in motivating example #1). So the strategy in such cases (where a is not in the domain) is: use algebraic manipulation to transform your function into another form where a *is* in the domain.

Example 3.15. (Simplify)

Consider $\lim_{x\to 1} \frac{x^3-1}{x-1}$. In this case, we can use long division to see that

$$\frac{x^3-1}{x-1} = \frac{(x-1)(x^2+x+1)}{x-1} = x^2+x+1 \text{ except at } x=1.$$

That is, these two functions are identical everywhere except at $x = 1$, where the former function is not defined but the latter function is. Therefore their limits as $x \to 1$ are the same; and since 1 is in the domain of the latter function, we can evaluate by the Direct Substitution Rule:

$$\lim_{x\to 1} \frac{x^3-1}{x-1} = \lim_{x\to 1}(x^2+x+1) = 3.$$

□

Example 3.16. (Rationalize)

When the problem has a difference of square roots, rationalisation can transform it into something very different that might be easier to deal with. For example

$$\begin{aligned}
\lim_{x\to 0} \frac{\sqrt{x^6+25}-5}{x^6} &= \lim_{x\to 0} \left(\frac{\sqrt{x^6+25}-5}{x^6} \cdot \frac{\sqrt{x^6+25}+5}{\sqrt{x^6+25}+5} \right) \\
&= \lim_{x\to 0} \frac{x^6+25-25}{x^6(\sqrt{x^6+25}+5)} \\
&= \lim_{x\to 0} \frac{1}{\sqrt{x^6+25}+5} \\
&= \frac{1}{\sqrt{0+25}+5} = \frac{1}{10} \quad \text{by direct substitution.}
\end{aligned}$$

□

The key is: if when you plug the value into a quotient and it comes out as "0/0", then what you can hope is that there is secretly a common factor between the numerator and denominator that could be cancelled by algebraic manipulation.

Example 3.17. (Simplify: find a common factor)

$$\lim_{t\to 0} \frac{(2+t)^2-4}{t} = \lim_{t\to 0} \frac{4+4t+t^2-4}{t} = \lim_{t\to 0} \frac{4t+t^2}{t} = \lim_{t\to 0}(4+t) = 4.$$

□

End of lecture # 6

Example 3.18. (Simplify: find a common factor)

$$\lim_{z\to 3}\frac{z-3}{z^2-9}=\lim_{z\to 3}\frac{z-3}{(z-3)(z+3)}=\lim_{z\to 0}\frac{1}{z+3}=\frac{1}{6}.$$

□

Sometimes, your function is just a mess, and simplifying is about cleaning it up.

Example 3.19. (Simplify)

$$\lim_{x\to 0}\frac{\frac{3}{x}-x^2+4}{5+\frac{1}{x}}=\lim_{x\to 0}\frac{3-x^3+4x}{5x+1}=\frac{3}{1}=3$$

□

You can also use the limit laws to evaluate one-sided limits.

Example 3.20. Find $\lim_{x\to 0} f(x)$, if it exists, when

$$f(x)=\begin{cases} x^2-4x & \text{if } x>0 \\ e^x & \text{if } x\le 0.\end{cases}$$

In this case, we do not have a single formula that is valid on both sides of 0; therefore we have no choice but to consider the one-sided limits. Namely,

$$\begin{aligned}\lim_{x\to 0^+} f(x) &= \lim_{x\to 0+}(x^2-4x) \quad \text{because that's the formula for } x>0 \\ &= 0 \quad \text{by Direct Substitution Rule.}\end{aligned}$$

whereas

$$\begin{aligned}\lim_{x\to 0^-} f(x) &= \lim_{x\to 0^-} e^x \quad \text{because that's the formula for } x<0 \\ &= e^0=1 \quad \text{by Direct Substitution Rule.}\end{aligned}$$

In this case, the left-hand and right-hand limits are different, so $\lim_{x\to 0} f(x)$ does not exist. □

Remember that the absolute value function is one of these piecewise-defined functions in disguise!

Example 3.21. (The sign function)

The sign function $\operatorname{sgn}(x)$ is

$$\operatorname{sgn}(x)=\frac{x}{|x|}\quad \text{for } x\neq 0.$$

What is $\lim_{x\to 0}\operatorname{sgn}(x)$, if it exists? To figure this out, we have to deal with the absolute value, which means we need to separate when its argument is positive or negative.

$$\begin{aligned}\lim_{x\to 0^+}\frac{x}{|x|} &= \lim_{x\to 0^+}\frac{x}{x} \quad \text{since } x>0 \text{ means } |x|=x \\ &= \lim_{x\to 0^+} 1 = 1.\end{aligned}$$

whereas

$$\begin{aligned}\lim_{x\to 0^-}\frac{x}{|x|} &= \lim_{x\to 0^-}\frac{x}{-x} \quad \text{since } x<0 \text{ means } |x|=-x\\ &= \lim_{x\to 0^-}(-1) = -1.\end{aligned}$$

Since the two limits are different, $\lim_{x\to 0}\frac{x}{|x|}$ does not exist.

That's kind of obvious from the graph! Because in fact, $\operatorname{sgn}(x)$ is the function that returns 1 if $x > 0$ and -1 if $x < 0$. □

Notice that when we compute $\lim_{x\to a} f(x)$, we only care about values of x near a. That is a handy observation when the function has many strange features.

Example 3.22. Find $\lim_{x\to -2} f(x)$ where

$$f(x) = \begin{cases} -4x-8 & \text{if } x < -2;\\ \sin(\pi x) & \text{if } -2 \le x \le 2;\\ -4x+8 & \text{if } x > 2.\end{cases}$$

Again, we use one-sided limits.

$$\lim_{x\to -2^-} f(x) = \lim_{x\to -2^-}(-4x-8) = -4(-2)-8 = 0.$$

On the right side, it is not true that $f(x) = \sin(\pi x)$ for all $x > -2$, but this equality *does* hold for all $x > -2$ and *close* to -2 (namely, $x < 2$). So we may still write

$$\lim_{x\to -2^+} f(x) = \lim_{x\to -2^+}\sin(\pi x) = \sin(\pi(-2)) = 0,$$

by the Direct Substitution Rule. □

Just because a function is piecewise-defined, doesn't mean you have to use one-sided limits.

Example 3.23. Find $\lim_{x\to 4}\operatorname{sgn}(x)$.

In this case, we notice that near 4 (namely, for all $x > 0$), the function is given by $\operatorname{sgn}(x) = 1$. Therefore $\lim_{x\to 4}\operatorname{sgn}(x) = \lim_{x\to 4} 1 = 1$. □

3.4 Limits involving infinity

There are several ways that a limit question might involve ∞. That said, it is really important to remember:

> Infinity is **NOT** a number. Infinity is a concept. Arithmetic with ∞ does not follow the rules of arithmetic.

Example 3.24. So it is very reasonable to say $\infty + \infty = \infty$, or $(\infty)(-\infty) = -\infty$. But it is completely insane and false to say $\infty - \infty = 0$, or $\infty/\infty = 1$. Look for examples of these fallacies in this section. □

3.4.1 Limits that diverge to infinity: vertical asymptotes

Definition 3.25. We say that the limit of a function f as x approaches a is infinity (or: *diverges to infinity*), and we write

$$\lim_{x\to a} f(x) = \infty$$

if we can make $f(x)$ as large as we wish by choosing x close enough to a. Similarly, we say the limit of f as x approaches a is negative infinity (or: *diverges to negative infinity*), and we write

$$\lim_{x\to a} f(x) = -\infty$$

if we can make $f(x)$ as large a negative value as we wish by choosing x close enough to a.

We also apply this definition to one-sided limits. Geometrically, a one-sided limit which diverges to ∞ or $-\infty$ corresponds to a *vertical asymptote* on the graph.

Example 3.26. $f(x) = \frac{1}{x}$; what is $\lim_{x\to 0} f(x)$?

We consider the one-sided limits. If $x > 0$ but very small, then $\frac{1}{x}$ gets increasingly large. In fact, if we want $f(x) > 10^n$, then we should take $x < 10^{-n}$. (Example, to get $f(x) > 1000$, choose $0 < x < 0.001$.) Thus we can make $f(x)$ arbitrarily large by taking x close enough to 0, so

$$\lim_{x\to 0^+} \frac{1}{x} = \infty.$$

On the other hand, if $x < 0$ and is close to zero, then $1/x$ will be very large negative. For example, to get $f(x) < -10^n$, we should choose $0 > x > 10^{-n}$. So we have verified

$$\lim_{x\to 0^-} \frac{1}{x} = -\infty.$$

In this case, the two one-sided limits are different; this often happens. □

Other rational functions can be treated similarly.

In general, if substitution of $x = a$ gives $c/0$ (for some number $c \neq 0$), then you can reason that the function is going to grow very large as $x \to a$, and can reason whether it is going to ∞, $-\infty$, or oscillating in between (in which case we just say it diverges).

Example 3.27. We have

$$\lim_{x\to \frac{\pi}{2}^+} \tan(x) = \lim_{x\to \frac{\pi}{2}^+} \frac{\sin(x)}{\cos(x)} = -\infty,$$

because when $x > \pi/2$, $\cos(x) < 0$ but very close to 0, while the numerator is near the positive number 1. Therefore the quotient is a large negative number, and so in the limit goes to $-\infty$.

On the other hand, since $\cos(x) > 0$ if $x < \pi/2$ and x is close to $\pi/2$ (and $\sin(x)$ is still near $1 > 0$), we conclude

$$\lim_{x\to \frac{\pi}{2}^-} \tan(x) = \infty.$$

□

We also know that

$$\lim_{x \to 0^+} \ln(x) = -\infty,$$

and in this case we cannot take the left-hand limit because $\ln(x)$ is not defined for $x \leq 0$.

3.4.2 Limits as x goes to ∞: horizontal asymptotes and long-term behaviour of functions

The kind of limit we encountered in DTDS were those where $t \to \infty$, where in that case the function was x_t (depending on the variable t). We define these limits as follows.

Definition 3.28. We say that the limit of a function f as x goes to ∞ is equal to the number L, and write

$$\lim_{x \to \infty} f(x) = L,$$

if we can make $f(x)$ as close to L as we wish by choosing x arbitrarily large. We can define each of the following expressions in a similar way:

$$\lim_{x \to -\infty} f(x) = L, \quad \lim_{x \to \infty} f(x) = \infty, \quad \lim_{x \to -\infty} f(x) = \infty, \quad \ldots.$$

When the limit as $x \to \infty$ or $x \to -\infty$ is a number (as opposed to not existing, or giving $\pm\infty$) then geometrically this corresponds to a *horizontal asymptote*.

Example 3.29.

$$\lim_{x \to \infty} \frac{1}{x} = 0$$

because as x grows huge (like 10^4), its reciprocal (10^{-4}) gets closer and closer to 0, and we can get $\frac{1}{x}$ as close to zero as we like by choosing x sufficiently large.

Similarly,

$$\lim_{x \to -\infty} \frac{1}{x} = 0$$

as well. □

Another favourite function with horizontal asymptote is e^x — but only as $x \to -\infty$:

$$\lim_{x \to -\infty} e^x = 0.$$

Many functions grow without bound as x does. For example,

$$\lim_{x \to \infty} e^x = \infty, \quad \text{and} \quad \lim_{x \to \infty} \ln(x) = \infty,$$

and all polynomials functions of degree at least 1 give $\lim_{x \to \infty} = \pm\infty$, where the signs are determined by the coefficient of the highest degree term.

A function need not have a limit, or diverge to ∞, as $x \to \infty$. Key examples to remember are the trigonometric functions. For example,

$$\lim_{x \to \infty} \sin(x) \text{ does not exist}$$

because the value of $\sin(x)$ oscillates between -1 and 1 as x goes larger and larger, never settling to any single value L.

There are some standard techniques for working out limits as $x \to \pm\infty$.

Example 3.30. (Divide numerator and denominator by highest power term in the denominator)

$$\lim_{x\to\infty} \frac{3x^2+2x+1}{-x^2+3} = \lim_{x\to\infty} \frac{3+2/x+1/x^2}{-1+3/x^2} = \frac{3}{-1} = -3$$

since the other terms in the numerator and the denominator are going to zero. Similarly

$$\lim_{x\to\infty} \frac{3x^3+2x+1}{-x^2+3} = \lim_{x\to\infty} \frac{3x+2/x+1/x^2}{-1+3/x^2} = -\infty$$

because the numerator is growing without bound towards ∞ while the denominator is staying very close to -1. At the other extreme,

$$\lim_{x\to\infty} \frac{3x^2+2x+1}{-x^3+3} = \lim_{x\to\infty} \frac{3/x+2/x^2+1/x^3}{-1+3/x^3} = 0$$

since this time the numerator is going to 0 while the denominator is staying constant near 1. □

This technique also works with other rapidly-growing functions, such as exponentials.

Example 3.31.

$$\lim_{x\to\infty} \frac{e^x-1}{4+5e^x} = \lim_{x\to\infty} \frac{1-e^{-x}}{4e^{-x}+5} = \frac{1}{5}$$

since $\lim_{x\to\infty} e^{-x} = 0$. □

3.5 Continuous functions

Definition 3.32. A function f is called *continuous at a point a (in its domain)* if $\lim_{x\to a} f(x)$ exists and is equal to $f(a)$. If $\lim_{x\to a} f(x)$ does not exist, or is not equal to $f(a)$, then the function is called discontinuous (at a). A function that is continuous at every point in its domain is simply called *continuous*.

In other words, a function f is continuous at a if and only if you may apply the Direct Substitution Rule to evaluate the limit of f as x goes to a.

Example 3.33. The following functions are continuous (that is, continuous on every point of their domain):

- polynomial, rational, exponential, logarithmic, trigonometric, inverse trigonometric, absolute value, and root functions;
- if f and g are continuous, so is their sum, difference, product and quotient;[1]

[1]Don't forget that the domain of the quotient f/g excludes any point where $g(x) = 0$.

- if f and g are continuous, so is their composition $f \circ g$.

Thus, for example, we can conclude that the following functions are continuous:

$$\frac{1}{x}, \quad \ln(e^x + x^2), \quad \sin(3 \arccos(x)),$$

since these are formed from continuous functions by a combination of the above constructions. □

Example 3.34. Many very interesting functions are discontinuous. For example,

- the extended sign function given by

$$\operatorname{sgn}(x) = \begin{cases} 0 & \text{if } x = 0; \\ \frac{x}{|x|} & \text{if } x \neq 0. \end{cases}$$

  is discontinuous at 0 [2] because (as we saw) $\lim_{x \to 0} \operatorname{sgn}(x)$ does not exist.

- The nearest integer function

$$\lfloor x \rfloor = \text{round } x \text{ down to the nearest integer that's} \leq x$$

  has a staircase-like graph. It is discontinuous at each integer value of x.

- In our model of drug absorption in the body, with a daily dose, we agreed that "connecting the dots" of the general solution was not a good model of what actually happened over the course of the day. Instead, we could model the concentration of drug in the body as a discontinuous function: decreasing linearly but with a jump discontinuity with each daily dose that makes the concentration suddenly jump much higher.

□

> The most common occurrence of a discontinuity is at the junctions of a piecewise defined function. The points of discontinuity, and the points excluded from the domain of f, are often where some of the most critical features of a function are to be found.

We sometimes want to splice two functions together in such a way that the result is continuous.

Example 3.35. Find the value of the parameter c, if it exists, such that the following function is continuous:

$$f(x) = \begin{cases} x + c & \text{if } x < 0; \\ \cos(x) & \text{if } x \geq 0. \end{cases}$$

Since $\lim_{x \to 0^+} f(x) = \lim_{x \to 0^+} \cos(x) = 1$ and $\lim_{x \to 0^-} f(x) = \lim_{x \to 0^-} x + c = c$, we deduce that we need to choose $c = 1$ so that these two one-sided limits will coincide (and equal $f(0)$). □

End of lecture # 7

[2] no matter how we chose the value of sgn(0), in fact.

Theorem 3.36 (Continuity and exchanging limits)**.** *If f is continuous and g is a function such that* $\lim_{x\to a} g(x) = b$ *exists, then*

$$\lim_{x\to a} f(g(x)) = f\left(\lim_{x\to a} g(x)\right) = f(b).$$

The new thing here is that it is not important if g is even defined at a, or continuous at a; we just need the innermost limit to exist. This also works for limits as $x \to \pm\infty$.

Example 3.37. Since the exponential function is continuous,

$$\lim_{x\to\infty} e^{1/x} = e^{\lim_{x\to\infty} 1/x} = e^0 = 1.$$

□

Example 3.38. We want to evaluate

$$\lim_{x\to\infty} (\ln(x+3) - \ln(x-1))$$

but plugging in ∞ for x gives us $\infty - \infty$, which is meaningless. So instead we simplify (always the first thing to do) using rules of logarithms, which gives

$$= \lim_{x\to\infty} \ln\left(\frac{x+3}{x-1}\right)$$

and since ln is continuous, we can exchange it with the limit to say

$$= \ln\left(\lim_{x\to\infty} \frac{x+3}{x-1}\right) = \ln\left(\lim_{x\to\infty} \frac{3+3/x}{1-1/x}\right) = \ln(3).$$

□

Chapter 4

The Derivative

One of the two central notions in Calculus is that of the derivative. The derivative of a function at a point is its instantaneous rate of change at that point; if we know the derivative of f at every point x, this gives us a new function $f'(x)$. Finding this function, and understanding what it tells us about f, is the object of this chapter.

4.1 The definition

The *average rate of change* of a function f over an interval $[a, b]$ in its domain is the rise over the run:

$$f_{av} = \frac{f(b) - f(a)}{b - a}$$

which is the slope of a *secant line* of the curve. The average rate of change tells us something about the function. For example:

- if $f(t)$ represents the reading on your odometer at time t, then f_{av} is your average speed from time $t = a$ to time $t = b$;
- if $f(t)$ represents the population of an organism at time t, then f_{av} is the average net growth rate from time $t = a$ to time $t = b$;
- if $f(x)$ is the amount in mg of a chemical (or drug) absorbed by the lungs when the amount in each breath is x (which varies as x varies! [1]), then f_{av} is the average marginal rate of absorption of the drug as the concentration rises from $x = a$ to $x = b$.

This information is crude, however: in the case of your odometer, it doesn't tell you if you drove within the speed limit during that interval; in the case of chemical absorption, it only gives a kind of rule of thumb (eg: "when x increases from a to b, you are probably absorbing half of the extra drug with each breath"). This is not precise enough to do science (or avoid a speeding ticket).

[1] This kind of variation is called functional response. In Chemistry, you might call it Michaelis-Menten or Monod reaction kinetics; it's the effect that your lungs (or any absorbing substance) reach saturation and can't absorb more. See also Absorption Functions in your textbook for varied examples.

What we want is the *instantaneous rate of change* of f at the point x in its domain. In the case of your odometer, the instantaneous rate of change means the value on your spedometer (your instantaneous speed); in the case of chemical absorption, the instantaneous rate of change tells you about the precise sensitivity of your lungs to the uptake of the drug as a function of concentration, which can let you correctly prescribe an increased dosage that will have the effect you want.

Definition 4.1. Let f be a function defined on a interval around x. Then f is called *differentiable* at x if the following limit exists:

$$\lim_{u \to x} \frac{f(u) - f(x)}{u - x}$$

in which case we denote the limit $f'(x)$, and call it the *derivative* of f at x. If f is differentiable at every point in an interval, then this defines a function f' and we say that f is differentiable, and f' is the derivative (function).

It's often easier to make a new variable $h = u - x$ and then notice that as $u \to x$, we have $h \to 0$. This gives the equivalent definition:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

Both mean the same thing: the slope of the *tangent line* to the curve $y = f(x)$ at the point $(x, f(x))$.

Remark 4.2. Other notation for the derivative: if $y = f(x)$ we can write

$$f'(x) = \frac{df}{dx} = \frac{dy}{dx} = y'.$$

Also: when necessary, we call the quotient $\frac{f(x+h)-f(x)}{h}$ or $\frac{f(u)-f(x)}{u-x}$ the *difference quotient*.

4.2 Examples of using the definition

Example 4.3. Consider the function $f(x) = mx + b$. This is a straight line with slope m; the derivative should therefore come out to be equal to m. Let's check:

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{m(x+h) + b - (mx+b)}{h} \\ &= \lim_{h \to 0} \frac{mx + mh + b - mx - b}{h} = \lim_{h \to 0} \frac{mh}{h} = \lim_{h \to 0} m = m, \end{aligned}$$

so yes, indeed, the derivative of a linear function is its slope. □

Example 4.4. Consider a quadratic function like $f(x) = 5x^2$. The slope of this curve varies with x, so we expect a more interesting answer. Indeed:

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{5(x+h)^2 - 5x^2}{h} \\ &= \lim_{h \to 0} \frac{5(x^2 + 2xh + h^2) - 5x^2}{h} = \lim_{h \to 0} \frac{5x^2 + 10xh + 5h^2 - 5x^2}{h} \\ &= \lim_{h \to 0} \frac{10xh + 5h^2}{h} = \lim_{h \to 0} (10x + 5h) = 10x. \end{aligned}$$

We look at this formula and judge that it makes sense: as x gets larger positive, $y = f(x)$ gets steeper (larger slope), and when $x = 0$, the slope of $y = 5x^2$ is 0, and when x is large negative, then the slope is a large negative number, too. We could also sketch the graph of $y = f(x)$ carefully and measure the slope of the tangent line at each point to compare. □

Example 4.5. The number 5 was almost just a decoration in the preceding calculation. We could do the same thing for a general abstract quadratic function $f(x) = ax^2 + bx + c$, where a, b, c are parameters. Then we compute

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{a(x+h)^2 + b(x+h) + c - \big(ax^2 + bx + c\big)}{h} \\ &= \lim_{h\to 0} \frac{a(x^2 + 2xh + h^2) + bx + bh + c - ax^2 - bx - c}{h} = \lim_{h\to 0} \frac{2axh + ah^2 + bh}{h} \\ &= \lim_{h\to 0} (2ax + ah + b) = 2ax + b. \end{aligned}$$

This gives us the general formula

$$f'(x) = 2ax + b,$$

which says the derivative of any quadratic function is a linear function. Note that it coincides with the particular case of $a = 5$, $b = c = 0$ of the preceding example. □

More complex functions generally take more work, since we do have to solve for the limit.

Example 4.6. Let $f(x) = \sqrt{x}$. Then if $x > 0$, we have:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \\ &= \lim_{h\to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \cdot \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}} = \lim_{h\to 0} \frac{x+h-x}{h(\sqrt{x+h} + \sqrt{x})} \\ &= \lim_{h\to 0} \frac{h}{h(\sqrt{x+h} + \sqrt{x})} = \lim_{h\to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} = \frac{1}{2\sqrt{x}}. \end{aligned}$$

If $x = 0$, this formula fails — with good reason.

First: since $\sqrt{x}$ is not defined on both sides of $x = 0$, we are not allowed to define the derivative of $\sqrt{x}$ at 0. The rule is: the function must be defined on both sides of the point; we have to take the two-sided limit.

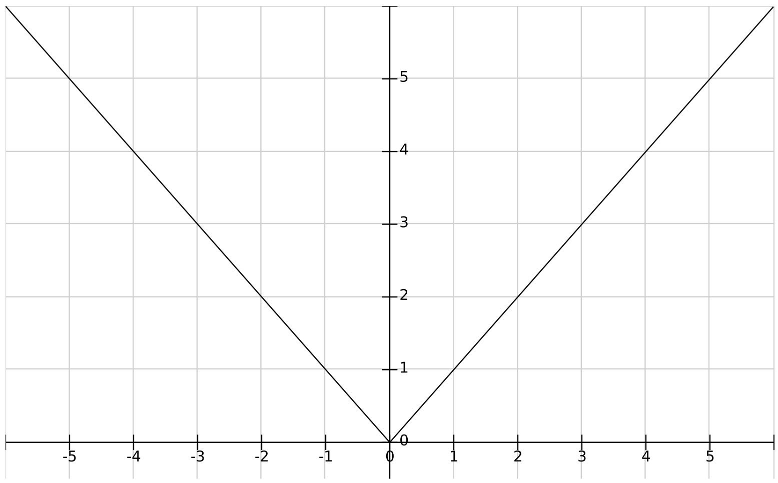
Secondly: as $x \to 0^+$, the slope of the curve is increasing to ∞, so the instantaneous rate of change is growing without bound. Although it seems reasonable to do so, no, we don't say "~~$f'(0) = \infty$~~"; we say "$f'(0)$ does not exist". □

4.3 Five ways a function can fail to be differentiable at x

There are several ways that a function could fail to be differentiable. Each one indicates that the behaviour of the function at that point is in some way unpredictable, which means it will be an important point to understand — it will be the place where the function acts in an interesting way.

The graph has a corner or a cusp at x. Consider for example

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



The graph of $y = |x|$. Its slope is -1 if $x < 0$ and $+1$ if $x > 0$, and the two do not agree at $x = 0$.

Since this is a piecewise defined function, to calculate the derivative at $x = 0$ we need to compute the two one-sided limits. As $h \to 0^+$ we have $h > 0$ so therefore

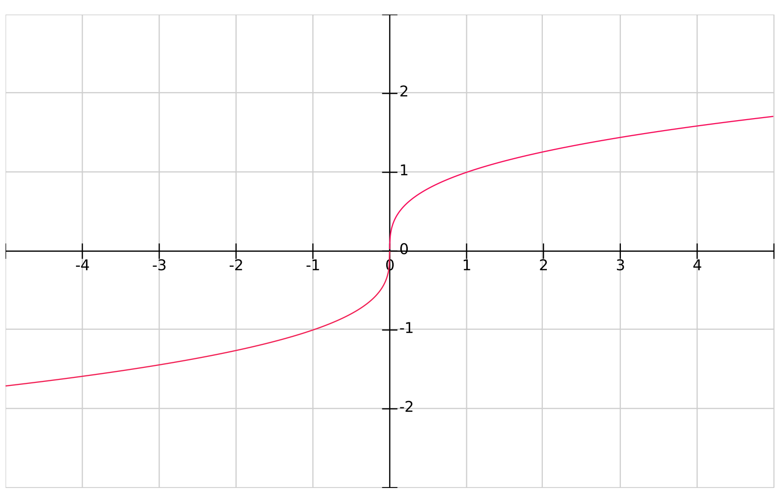
$$\lim_{h \to 0^+} \frac{|0+h| - |0|}{h} = \lim_{h \to 0^+} \frac{\cancel{h}}{\cancel{h}} = 1$$

whereas

$$\lim_{h \to 0^-} \frac{|0+h| - |0|}{h} = \lim_{h \to 0^-} \frac{-h}{h} = -1.$$

Since the two one-sided limits disagree, the (two-sided) limit does not exist, so f is not differentiable at $x = 0$.

The graph is vertical at x. Consider for example $f(x) = \sqrt[3]{x} = x^{1/3}$, which is the inverse function to $y = x^3$. Its graph is below.



The graph of $y = x^{1/3}$. Its slope increases to ∞ as you approach 0.

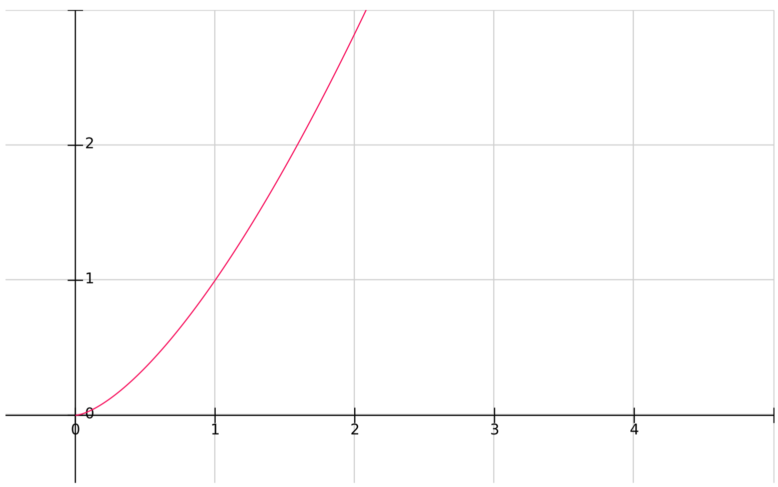
Note that $f(x)$ is still a function — it passes the vertical line test — but at the instant it passes zero its tangent line is vertical. We can see this from the definition as well:

$$\lim_{h\to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h\to 0} \frac{h^{1/3}}{h} = \lim_{h\to 0} \frac{1}{h^{2/3}} = \infty.$$

So $f'(0)$ does not exist.

The function does not exist on both sides of x. This is part of the requirement for the derivative, and it reflects the idea that it only makes sense to talk about the instantaneous rate of change at a point if you can pass through that instant.

For example, consider $f(x) = x^{3/2} = \sqrt{x^3}$, which is only defined for $x \geq 0$. Its graph is drawn below.



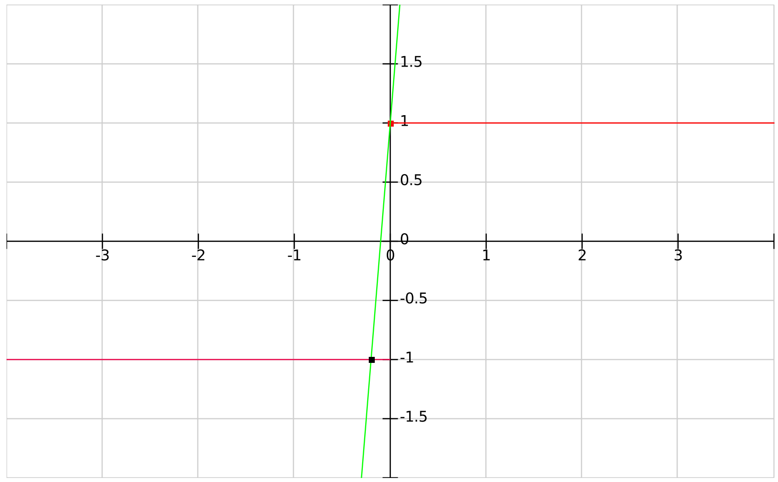
The graph of $y = x^{3/2}$. The derivative of this function is not defined at 0, because the function is not defined on both sides of 0.

So $f'(0)$ does not exist. In this case, however, we might reasonably as about $\lim_{x\to 0} f'(x)$, which would be answering the question "What is the limit of the slope of f as x approaches 0?". You can verify that $f'(x) = \frac{3}{2}\sqrt{x}$, so the limit is 0, which is quite reasonable-looking from the graph.

The function is discontinuous at x. For example, consider a function like

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ -1 & \text{if } x < 0. \end{cases}$$

This is discontinuous at 0 and we claim that the derivative at 0 does not exist. We look at the graph, below.



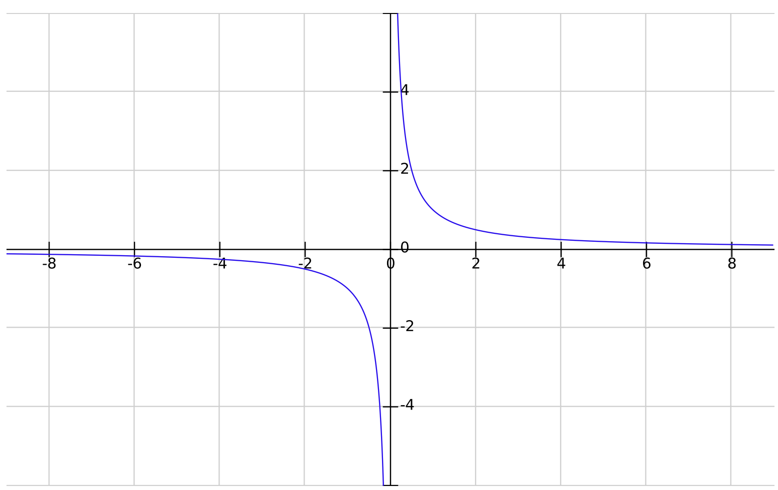
The graph of $y = \operatorname{sgn}(x)$ in blue. A secant line of the function going through $(0, f(0))$ but starting from the left, is drawn in green. The derivative of this function is equal to 1 on either side of 0, but the tangent line at 0 does not exist because the function abruptly jumps, causing the slopes of the secant lines to go to ∞ as you approach 0 from the left.

This is a general fact, which we can state in two equivalent ways, as follows.

Theorem 4.7. *Every function f that is differentiable at a point x is automatically continuous there. If a function f is not continuous at a point x, then it cannot be differentiable there.*

The function f is not defined at x. If f is not defined at x, we cannot even write down the formula for the derivative, because we have no value for $f(x)$. At best, we could be asking for the limit of the derivative as we approach x.

An interesting example might be $f(x) = \frac{1}{x}$, which is undefined at 0.



The graph of $y = 1/x$ in blue. It is undefined at 0 so we cannot draw a secant line through $(0, f(0))$; moreover, any average rate of change across an interval containing 0 is meaningless.

Here, f is not defined at 0 so neither is f'. We can encode this observation as follows.

Observation 4.8. *The domain of f' can never be bigger than the domain of f, but it can be smaller.*

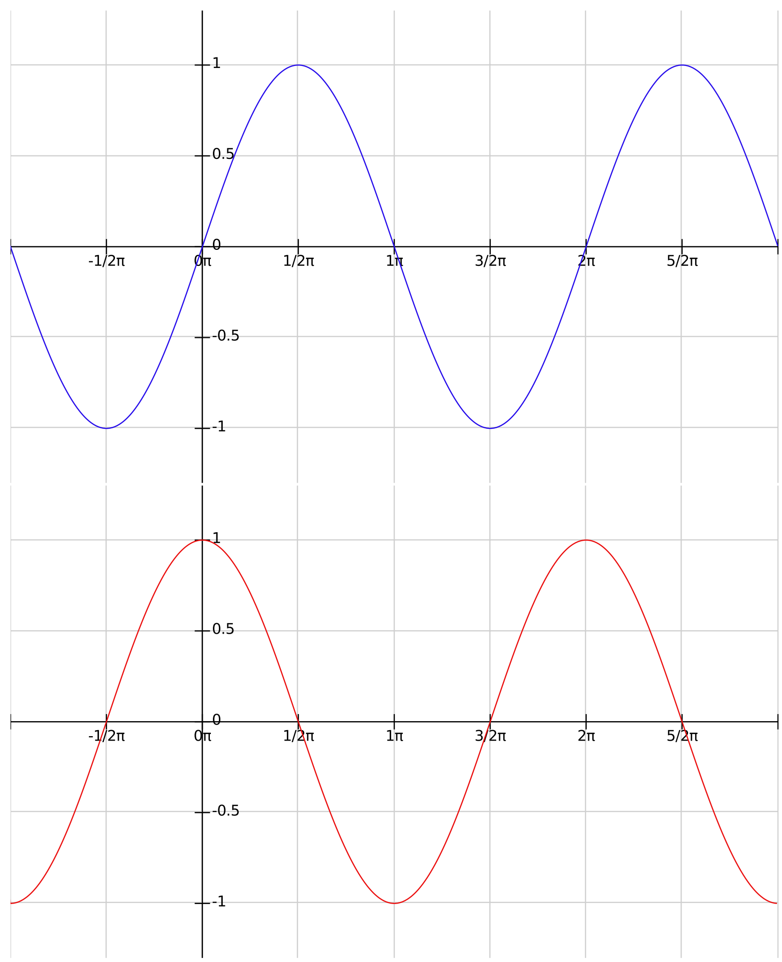
We saw an example where the domain of f' is smaller than the domain of f, for $f(x) = x^{2/3}$.

4.4 What f' tells you about f

Since we defined the derivative in terms of the tangent line, there is a nice correspondence between the graph of a function and the properties of its derivative:

$$\begin{array}{rcc} f'(x) > 0 & \Leftrightarrow & f \text{ is increasing} \\ f'(x) < 0 & \Leftrightarrow & f \text{ is decreasing} \\ f'(x) = 0 & \Leftrightarrow & f \text{ has a horizontal tangent} \end{array}$$

So for example, by looking at the graph of a function f, we can sketch the graph of f':



The graph of $f(x)$ in blue, and the inferred graph of f' in red. Where f has a horizontal tangent, f' is 0; where f is increasing, f' is positive; where f is decreasing, f' is negative.

We can also reverse this process, that is, sketch f from the graph of f'; but the answer will not be unique. The derivative determines the *shape* of the function, but not where exactly it is.[2] Thus $f(x)$ and $f(x)+c$ for any constant c will have the same derivative.

One of the key things the derivative can tell us is where the function does something interesting. We call these critical points.

Definition 4.9. A number c in the domain of a function f is a *critical point* of f if either $f'(c) = 0$ or else $f'(c)$ is undefined.

Example 4.10. The function $f(x) = 2x + 3$ has derivative $f'(x) = 2$ everywhere, so it has no critical points. The function $f(x) = 5x^2$ has derivative $f'(x) = 10x$, which is 0 at $x = 0$, so $c = 0$ is a critical point of f. The derivative of the function $f(x) = |x|$ is not defined at $x = 0$, so 0 is a critical point of f. In each case, the critical point identifies the existence of an interesting feature of the graph. □

End of lecture # 8

4.5 Differentation Rules: The basics

Although the definition of the derivative can be used to compute derivatives, this is quite tedious. Thankfully, over the years since the discovery of the derivative, people have figured out a number of simple rules that, taken together, can be used to evaluate the derivative of almost function that is given by a formula. The definition of the derivative then only needs to be used if one is given a function as a graph, or as a table of data — that is, where you don't have a formula for f.

Theorem 4.11. *The following rules of derivatives hold:*

1. *(Power rule) If $f(x) = x^n$ for some $n \in \mathbb{R}$, then $f'(x) = nx^{n-1}$. In particular, the derivative of the constant function 1 is 0.*
2. *(Constant multiple rule) If f is differentiable and c is a constant, then $g(x) = cf(x)$ is differentiable and $g'(x) = cf'(x)$.*
   Suppose now that f and g are differentiable functions. Then:
3. *(Sum/difference rule) $h(x) = f(x) \pm g(x)$ is differentiable and $h'(x) = f'(x) \pm g'(x)$.*
4. *(Product rule) $h(x) = f(x)g(x)$ is differentiable and $h'(x) = f'(x)g(x) + f(x)g'(x)$.*
5. *(Quotient rule) $h(x) = \dfrac{f(x)}{g(x)}$ is differentiable and*
$$h'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}.$$
6. *(Chain rule) $h(x) = f(g(x))$ is differentiable and $h'(x) = f'(g(x))g'(x)$.*

[2]For example, if you know exactly what speed you were driving at every instant of a day, you could figure out how far you'd travelled in that day — but not where you were!

Please memorize these formulas, in whatever way that works for you. I remember the quotient rule as : "the bottom times the derivative of the top, minus the top times the derivative of the bottom, all over the bottom squared." Others write $h(x) = \frac{u}{v}$ and remember $vdu - udv$ over v^2.

For the chain rule: remember that f was evaluated at $g(x)$, so that is where you have to evaluate f': it's $f'(g(x))$ NOT $f'(x)$ in the chain rule.

Example 4.12. If $f(x) = x^{47.5}$ then $f'(x) = 47.5x^{46.5}$ by the power rule. □

Example 4.13. If $f(x) = \sqrt[3]{x}$ then rewrite this as $f(x) = x^{1/3}$. So by the power rule,

$$f'(x) = \frac{1}{3}x^{\frac{1}{3}-1} = \frac{1}{3}x^{-2/3} = \frac{1}{3x^{2/3}}.$$

□

Example 4.14. If $f(x) = \frac{1}{x^3}$, then rewrite this as $f(x) = x^{-3}$. So by the power rule,

$$f'(x) = -3x^{-4} = \frac{-3}{x^4}.$$

□

These rules can be combined to compute derivatives of all rational functions.

Example 4.15. If $f(x) = 5x^2 + 4x + 2$ then by the power, sum and constant multiple rules, $f'(x) = 5(2x) + 4(1) + 2(0) = 10x + 4$. □

Example 4.16. If $f(x) = (2x + 1)(3x + 4)$ then by the product rule, we have

$$f'(x) = (2)(3x + 4) + (2x + 1)(3) = 12x + 11.$$

We could also have gotten this answer by multiplying out $f(x) = 6x^2 + 11x + 4$ and applying the sum and constant multiple rules. □

Example 4.17. If

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

then by the quotient rule, we have

$$f'(x) = \frac{(x + 4)(2x) - x^2(1)}{(x + 4)^2} = \frac{x^2 + 8x}{(x + 4)^2}.$$

□

Using the chain rule requires you to be strongly aware of the composition of functions. For example, here is a table of some compositions of functions:

inner function $g(x)$	outer function $f(u)$	composition $F(x) = f(g(x))$
$4x+1$	$\sqrt{u}$	$\sqrt{4x+1}$
$\sqrt{x}$	$4u+1$	$4\sqrt{x}+1$
$1+x^2$	$1/u$	$\dfrac{1}{1+x^2}$

In each case, we apply the chain rule to find the derivative as $F'(x) = f'(g(x))g'(x)$:

- If $g(x) = 4x+1$ then $g'(x) = 4$; if $f(u) = \sqrt{u} = u^{1/2}$ then $f'(u) = \frac{1}{2}u^{-1/2}$. Therefore $F(x) = f(g(x)) = \sqrt{4x+1}$ has derivative
$$F'(x) = \frac{1}{2}(4x+1)^{-1/2} \cdot 4 = \frac{2}{\sqrt{4x+1}}.$$
- If $g(x) = \sqrt{x}$ and $f(u) = 4u+1$, then $g'(x) = \frac{1}{2}x^{-1/2}$ and $f'(u) = 4$, so if $F(x) = f(g(x)) = 4\sqrt{x}+1$ then
$$F'(x) = 4 \cdot \frac{1}{2}x^{-1/2} = \frac{2}{\sqrt{x}}$$
as we can see by applying the constant multiple rule directly.
- If $g(x) = 1+x^2$ and $f(u) = 1/u$ then $g'(x) = 2x$ and $f'(u) = -u^{-2}$, so if $F(x) = f(g(x)) = 1/(1+x^2)$ we have
$$F'(x) = -(1+x^2)^{-2} \cdot (2x) = \frac{-2x}{(1+x^2)^2}$$
as we can check directly with the quotient rule (but this way is faster).

Example 4.18. If $f(x) = \frac{1}{3x^4+x}$ then by the quotient rule

$$f'(x) = \frac{(3x^4+x)0 - 1(12x^3+1)}{(3x^4+x)^2} = \frac{-12x^3-1}{(3x^4+x)^2}.$$

Alternately, we could write $f(x) = (3x^4+x)^{-1}$ and then use the chain rule

$$f'(x) = -(3x^4+x)^{-2}(12x^3+1) = \frac{-12x^3-1}{(3x^4+x)^2},$$

which of course comes out the same. □

It is hugely important to practice these rules! Over the coming sections, we will be adding the rules for differentiating more functions, and combining functions in better ways. These rules get easier to use the more you practice with them. Like knowing your multiplication tables by heart, being able to differentiate easily will make everything we do after this point make better sense and go more easily.

4.5.1 Why the power rule is true

We can explain why the power rule

$$\boxed{\frac{d}{dx}x^n = nx^{n-1}}$$

is true for the case that n is a positive integer. (To prove it holds for all values of n requires using the exponential and logarithm functions.)

Recall the *binomial theorem*, which says that

$$(x+h)^n = \sum_{k=0}^{n} \binom{n}{k} x^k h^{n-k} = x^n + nx^{n-1}h + \cdots + nxh^{n-1} + h^n.$$

In particular, each term is divisible by h except the first; and after the second term, each term is divisible by h^2.

Thus when we calculate

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = \lim_{h\to 0} \frac{\cancel{x^n} + nx^{n-1}h + \cdots + nxh^{n-1} + h^n - \cancel{x^n}}{h}$$

we can divide what is left evenly by h, leaving

$$= \lim_{h\to 0} nx^{n-1} + \frac{n(n-1)}{2}x^{n-2}h + \cdots + nxh^{n-2} + h^{n-1} = nx^{n-1}$$

as we expected.

4.5.2 Why the product rule is true

The product rule says that

$$\boxed{\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x).}$$

In particular, the derivative of a product is NOT the product of the derivatives.

The reason for the strange mixed term comes from geometry. Algebraically, the way we have to look at the difference is as follows:

$$\begin{aligned} f(x+h)g(x+h) - f(x)g(x) &= f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x) \\ &= (f(x+h)-f(x))g(x+h) + f(x)(g(x+h)-g(x)). \end{aligned}$$

So now we can see what happens as we take the limit as $h \to 0$:

$$\begin{aligned}
\lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} &= \lim_{h\to 0} \left(\frac{f(x+h)-f(x)}{h} g(x+h) - f(x)\frac{g(x+h)-g(x)}{h} \right) \\
&= \lim_{h\to 0} \left(\frac{f(x+h)-f(x)}{h} g(x+h) \right) - \lim_{h\to 0} \left(f(x)\frac{g(x+h)-g(x)}{h} \right) \\
&= \lim_{h\to 0} \left(\frac{f(x+h)-f(x)}{h} \right) \lim_{h\to 0} g(x+h) - f(x) \lim_{h\to 0} \left(\frac{g(x+h)-g(x)}{h} \right) \\
&= f'(x)g(x) - f(x)g'(x)
\end{aligned}$$

where we have used the continuity of g at x to conclude that $\lim_{h\to 0} g(x+h) = g(x)$, and we have used the definition of the derivative in the other two cases.

4.5.3 Why the chain rule is true

The chain rule says that

$$\boxed{\frac{d}{dx} f(g(x)) = f'(g(x))g'(x).}$$

Notice that we differentiate each function once, and multiply the result.

Example 4.19. $f(h)$ = distance travelled in km as a function of time h measured in hours; $g(t) = t/60$ is the function that converts minutes to hours, $f(g(t))$ measures distanced travelled in km as a function of time measure in minutes. So if you travel 10 km/h (so $f'(h) = 10$ for all h) then you are travelling

$$\frac{d}{dt} f(g(t)) = f'(g(t))g'(t)$$

km per minute. Since $g'(t) = \frac{1}{60}$, this gives 10 km/h $\times \frac{1}{60}$ h/min $= \frac{1}{6}$ km/min. □

This example was boring because the derivatives were all constants. To see why the rule holds in general, here is an argument.

So let's write $g(x) = y$ and for each h, define a new variable k by the formula $k = g(x+h) - g(x)$. So $g(x+h) = y + k$ for some small value k. Since g is continuous at x, we see that as $h \to 0$, we also have $k \to 0$. That lets us write

$$\begin{aligned}
\lim_{h\to 0} \frac{f(g(x+h)) - f(g(x))}{h} &= \lim_{h\to 0} \frac{f(y+k) - f(y)}{h} \\
&= \lim_{h\to 0} \left(\frac{f(y+k)-f(y)}{k} \cdot \frac{k}{h} \right) \\
&= \lim_{k\to 0} \left(\frac{f(y+k)-f(y)}{k} \right) \lim_{h\to 0} \frac{g(x+h)-g(x)}{h} \\
&= f'(y)g'(x) \\
&= f'(g(x))g'(x).
\end{aligned}$$

(What we didn't allow for in this formula was the possibility that g is constant, so that $k = 0$; but there are other ways to deduce the same formula even in these weird cases.)

4.6 Derivatives of exponential functions

Consider an exponential function

$$f(x) = a^x, \quad \text{where } a > 0.$$

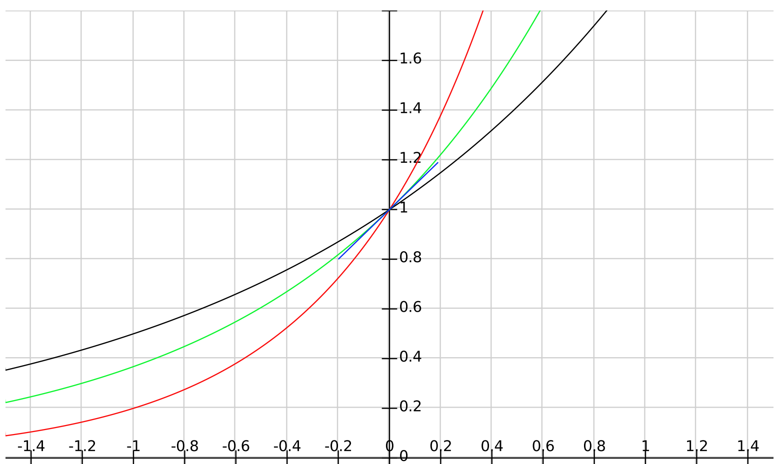
To find its derivative, lacking any other ideas, we use the definition:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{a^{x+h} - a^x}{h} = \lim_{h \to 0} a^x \left(\frac{a^h - 1}{h} \right) = a^x \lim_{h \to 0} \frac{a^h - 1}{h} = a^x f'(0). \tag{4.1}$$

We have gone in a bit of a circle here: we wanted the derivative of $f(x) = a^x$ at some random point x, but instead figured out that the derivative will satisfy

$$f'(x) = a^x f'(0).$$

Well, at least that is a simpler problem: just figure out the derivative at 0, which should be the slope of the tangent line to the curve $y = f(x)$ at $x = 0$.



Graphs of various exponential functions : $y = 2^x$ in black, $y = 5^x$ in red, and $y = e^x$ in green. The short line segment at in blue has slope exactly 1, and is tangent to the graph of e^x.

This is one way to define the natural base e (Euler's number): it is the base of the exponential function $f(x) = e^x$ for which $f'(0) = 1$. To three decimal places, $e \simeq 2.718$. Leonhard Euler (1707–1783) calculated e to 18 decimal places (!).

So $f(x) = e^x$, the natural exponential, is the one that satisfies, for any x, $f'(x) = e^x$, that is,

$$\boxed{\frac{d}{dx} e^x = e^x}$$

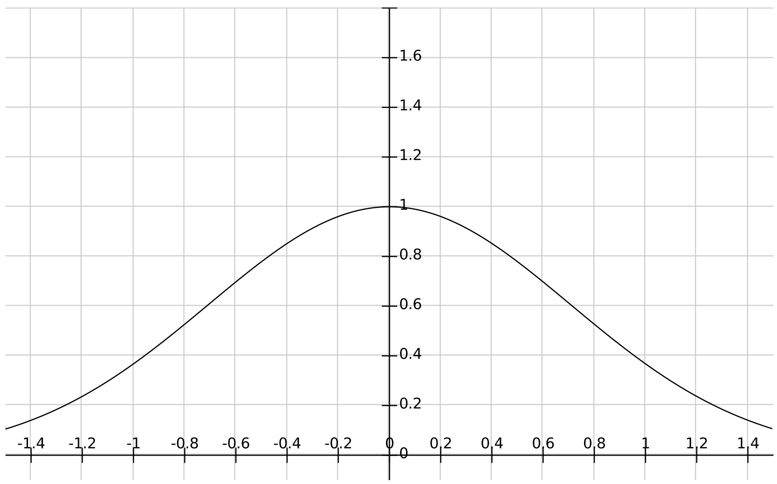
That's an incredible property, for a function to be equal to its own derivative; in fact, the only functions which this property are those of the form Ke^x for some constant K.

Example 4.20. Problem: If $g(x) = x^2 e^x$ find $g'(x)$. Solution: This is a product of two functions, so we use the product rule.

$$g'(x) = (2x)e^x + x^2(e^x) = 2xe^x + x^2e^x = (2x + x^2)e^x = x(x + 2)e^x.$$

(Tip: this last form is the most useful for finding roots of $g'(x)$, but if we needed to compute $g''(x)$, leave it in the second-last form for efficiency.) □

Example 4.21. Problem: the *normal distribution* describes a standard bell curve, and the basic form is $h(x) = e^{-x^2}$.



The graph of $y = e^{-x^2}$, which describes a standard bell curve in Statistics.

Find $h'(x)$.

Solution: This is a composition of two functions, so we apply the chain rule. We have $h(x) = f(g(x))$ where $g(x) = -x^2$ is the innermost function and $f(u) = e^u$ is the outermost function; therefore

$$h'(x) = f'(g(x))g'(x) = e^{g(x)}g'(x) = e^{-x^2}(-2x) = -2xe^{-x^2}.$$

□

Example 4.22. Consider $F(x) = x^n e^{-x}$, where n is some fixed number; this is related to another important function in Statistics, called the Gamma Distribution. Then

$$\frac{d}{dx}F(x) = nx^{n-1}e^{-x} + x^n(-e^{-x}) = (n - x)x^{n-1}e^{-x}.$$

□

Example 4.23. Problem: Let $a > 0$. Find the derivative of $f(x) = a^x$.

Solution: We rewrite a^x as an exponential to base e:

$$y = a^x \Leftrightarrow \ln(y) = \ln(a^x) \Leftrightarrow \ln(y) = x\ln(a) \Leftrightarrow e^{\ln(y)} = e^{x\ln(a)} \Leftrightarrow y = e^{x\ln(a)}.$$

In other words, the above argument proves:

$$a^x = e^{x\ln(a)}$$

Therefore, by the chain rule (remembering that $\ln(a)$ is just a number, because a is some fixed number):

$$f'(x) = e^{x\ln(a)}\ln(a) = a^x\ln(a),$$

that is,

$$\frac{d}{dx}a^x = a^x\ln(a)$$

□

In the previous example, we have actually found the mysterious value from the beginning of this section, in (4.1): the derivative of a^x at $x = 0$! That is, since $f'(0) = \ln(2)$ we have shown

$$\lim_{h\to 0}\frac{a^h - 1}{h} = \ln(a).$$

4.7 Derivatives of logarithms

In the previous section, we found the derivatives of exponential functions, by using the definition and discovering the number e for which the definition gives a nice limit. Now we want to differentiate $f(x) = \ln(x)$ (or more generally $\log_a(x)$ for some fixed constant a). We *could* use the definition, but now we actually have more tools available, so there is an easier way.

We start with the identity

$$e^{\ln(x)} = x.$$

The left hand side is a function $F(x) = e^{\ln(x)}$ which is the composition $f(g(x))$ where $f(u) = e^u$ and $g(x) = \ln(x)$. We do not know what $g'(x)$ is, but we reason: the left hand side is a function equal to the function on the right hand side at every point x, and so their graphs are the same and their derivatives are the same. So this should give is an equation to find $g'(x)$!

The derivative of the left hand side is

$$f'(g(x))g'(x) = e^{g(x)}g'(x)$$

whereas the derivative of the right hand side is 1. Therefore, we have the equation

$$e^{g(x)}g'(x) = 1$$

which says that

$$g'(x) = \frac{1}{e^{g(x)}}.$$

Now $g(x) = \ln(x)$ so $e^{\ln(x)} = x$; thus we conclude:

$$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$$

This is an incredible formula! Notice that it fills a gap in our differentiation tables: the derivative of x^n is nx^{n-1}, so there was no function that gave us derivative x^{-1}. Now we've found it — it's the natural logarithm, which you can think of as the slowest growing function that goes to ∞ as $x \to \infty$.

Example 4.24. Problem: Find the derivative of $\ln(x^2+1)$.

Solution: This is a composition of functions, so we apply the chain rule.

$$\frac{d}{dx}(\ln(x^2+1)) = \frac{1}{x^2+1}(2x) = \frac{2x}{x^2+1}.$$

□

Example 4.25. Problem: Find the derivative of $\sqrt{\ln(x)+4}$.

Solution: This is a composition of functions so we apply the chain rule:

$$\frac{d}{dx}(\sqrt{\ln(x)+4}) = \frac{1}{2}(\ln(x)+4)^{-1/2} \cdot \frac{1}{x} = \frac{1}{2x\sqrt{\ln(x)+4}}.$$

□

Example 4.26. Problem: Let $a > 0$ be a constant. Find the derivative of $f(x) = \log_a(x)$.

Solution: We only know the derivative of $g(x) = \ln(x)$, so we need to change base. We use the standard method:

$$y = \log_a(x) \Leftrightarrow a^y = x \Leftrightarrow \ln(a^y) = \ln(x) \Leftrightarrow y\ln(a) = \ln(x) \Leftrightarrow y = \frac{1}{\ln(a)}\ln(x),$$

which says

$$\log_a(x) = \frac{\ln(x)}{\ln(a)}$$

Therefore

$$\frac{d}{dx}(\log_a(x)) = \frac{d}{dx}(\frac{\ln(x)}{\ln(a)}) = \frac{1}{\ln(a)} \cdot \frac{1}{x} = \frac{1}{x\ln(a)},$$

that is

$$\frac{d}{dx}(\log_a(x)) = \frac{1}{x\ln(a)},$$

where we have remember that a is just a constant, so $\ln(a)$ is just a number. □

4.8 Derivatives of functions like $f(x)^{g(x)}$

We have seen that $\frac{d}{dx}x^n = nx^{n-1}$, if n is a constant. We have also seen that $\frac{d}{dx}a^x = a^x \ln(a)$, if a is a positive constant. So what is

$$\frac{d}{dx}x^x?$$

Woudl it be $x(x^{x-1})$ "power rule" or $x^x \ln(x)$ "exponential rule"? Answer: NEITHER. You may only apply a rule under the hypotheses in which it was derived, and in this case, both are wrong.

Instead, we go back and recall how we solved for the derivative of a^x: we converted it to base e. That process will work here as well:

$$y = x^x \Leftrightarrow \ln(y) = \ln(x^x) \Leftrightarrow \ln(y) = x\ln(x) \Leftrightarrow y = e^{x\ln(x)}.$$

Fabulous! This is now a function that we can differentiate, using the exponential and chain rules:

$$\frac{d}{dx}x^x = \frac{d}{dx}e^{x\ln(x)} = e^{x\ln(x)}\frac{d}{dx}(x\ln(x)) = e^{x\ln(x)}(1\cdot\ln(x) + x\cdot\frac{1}{x}) = x^x(\ln(x)+1).$$

(We used that $e^{x\ln(x)} = x^x$ to simplify the expression in the last step.)

End of lecture # 9

Example 4.27. Find the derivative of $h(x) = (x^2+1)^{\ln(x)}$.

Solution: We use identity

$$a^b = e^{b\ln(a)} \quad \text{or, here:} \quad f(x)^{g(x)} = e^{g(x)\ln(f(x))} \quad \star$$

with $a = f(x) = x^2+1$ and $b = g(x) = \ln(x)$ to rewrite f as

$$(x^2+1)^{\ln(x)} = e^{\ln(x)\ln(x^2+1)}.$$

(Since we want base e, it isn't useful to simplify this to $x^{\ln(x^2+1)}$; notice what a crazy function this is, with so many equivalent forms.)

Now we use the chain rule:

$$\begin{aligned} h'(x) &= e^{\ln(x)\ln(x^2+1)}\left(\frac{1}{x}\ln(x^2+1) + \ln(x)\frac{1}{x^2+1}(2x)\right) \\ &= (x^2+1)^{\ln(x)}\left(\frac{\ln(x^2+1)}{x} + \frac{2x\ln(x)}{x^2+1}\right). \end{aligned}$$

□

Therefore, we have a general rule:

$$\frac{d}{dx}f(x)^{g(x)} = f(x)^{g(x)}\left(g'(x)\ln(f(x)) + \frac{g(x)f'(x)}{f(x)}\right)$$

but this is too ridiculous to memorize; instead we remember the technique $\star$.

Incidentally, you can use this method to go back and prove the power rule for any power $n \in \mathbb{R}$, not just positive integers, by rewriting $x^n = e^{n\ln(x)}$ and simplifying your answer.

4.9 Derivatives of sine and cosine

We go back to the definition to understand the derivative of $f(x) = \sin(x)$:

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x)\cos(h) + \sin(h)\cos(x) - \sin(x)}{h} \\ &= \lim_{h\to 0} \left(\sin(x)\frac{\cos(h)-1}{h} + \cos(x)\frac{\sin(h)}{h} \right) \\ &= \sin(x) \lim_{h\to 0} \frac{\cos(h)-1}{h} + \cos(x) \lim_{h\to 0} \frac{\sin(h)}{h}. \end{aligned}$$

So it all comes down to understanding these two limits — which are exactly the derivatives of $\cos(x)$ and of $\sin(x)$ at $x = 0$.

In fact, we have:

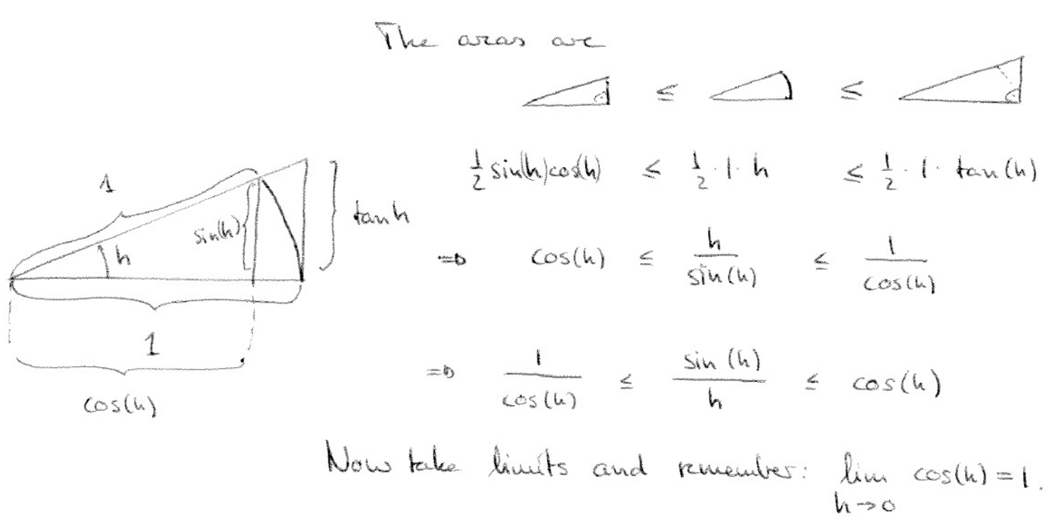
$$\boxed{\lim_{h\to 0} \frac{\cos(h)-1}{h} = 0 \quad \text{and} \quad \lim_{h\to 0} \frac{\sin(h)}{h} = 1}$$

(see below). Thus $\sin'(x) = \cos(x)$. A similar process with the definition of the derivative of cosine comes down to the same two limits. and after some work we conclude that

$$\boxed{\frac{d}{dx}\sin(x) = \cos(x) \quad \text{and} \quad \frac{d}{dx}\cos(x) = -\sin(x).}$$

4.9.1 Why $\lim_{h\to 0} \frac{\sin(h)}{h} = 1$

We can make a geometric argument.



4.9.2 Why $\lim_{h\to 0} \frac{\cos(h)-1}{h} = 0$

So

$$\cos'(0) = \lim_{h\to 0} \frac{\cos(0+h)-\cos(h)}{h} = \lim_{h\to 0} \frac{\cos(h)-1}{h}.$$

We know that

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

(This is more correctly written as $(\sin(x))^2 + (\cos(x))^2 = 1$.) Therefore we can differentiate both sides to give

$$2(\sin(x))\sin'(x) + 2\cos(x)\cos'(x) = 0.$$

When $x = 0$, this tells us that

$$0 + 2\cos'(0) = 0$$

so $\cos'(0) = 0$. Hence the limit.

Notice: we know that the graph of $y = \cos(x)$ at $x = 0$ has a horizontal tangent, so we expect the derivative to be 0 there.

4.10 Derivatives of other trigonometric functions

The other trigonometric functions are

$$\boxed{\tan(x) = \frac{\sin(x)}{\cos(x)}, \quad \cot(x) = \frac{\cos(x)}{\sin(x)}, \quad \csc(x) = \frac{1}{\sin(x)}, \quad \sec(x) = \frac{1}{\cos(x)}.}$$

Therefore we can just apply the quotient rule to deduce their derivatives of those of $\sin(x)$ and $\cos(x)$.

Example 4.28. Find $\frac{d}{dx}\tan(x)$.

Solution: we apply the quotient rule

$$\frac{d}{dx}\tan(x) = \frac{d}{dx}\frac{\sin(x)}{\cos(x)} = \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{\cos^2(x)} = \frac{1}{\cos^2(x)} = \sec^2(x).$$

□

In this example, we applied the key trigonometric identity:

$$\boxed{\sin^2(x) + \cos^2(x) = 1}$$

Example 4.29. Find $\frac{d}{dx}\csc(x)$.

Solution: we apply the chain rule to $\csc(x) = (\sin(x))^{-1}$. This gives

$$\frac{d}{dx}\csc(x) = -(\sin(x))^{-2}\cos(x) = \frac{-\cos(x)}{\sin^2(x)} = -\frac{\cos(x)}{\sin(x)}\frac{1}{\sin(x)} = -\cot(x)\csc(x).$$

□

Exercise 4.30. Use the quotient rule and standard identities to find the derivatives of $\sec(x)$ and of $\cot(x)$.

It is very useful to memorize the *derivatives of the six standard trigonometric functions*:

$$\begin{aligned}
\frac{d}{dx}\sin(x) &= \cos(x), & \frac{d}{dx}\cos(x) &= -\sin(x)\\
\frac{d}{dx}\tan(x) &= \sec^2(x), & \frac{d}{dx}\cot(x) &= -\csc^2(x)\\
\frac{d}{dx}\sec(x) &= \sec(x)\tan(x), & \frac{d}{dx}\csc(x) &= -\csc(x)\cot(x)
\end{aligned}$$

4.11 Inverse trigonometric functions and their derivatives

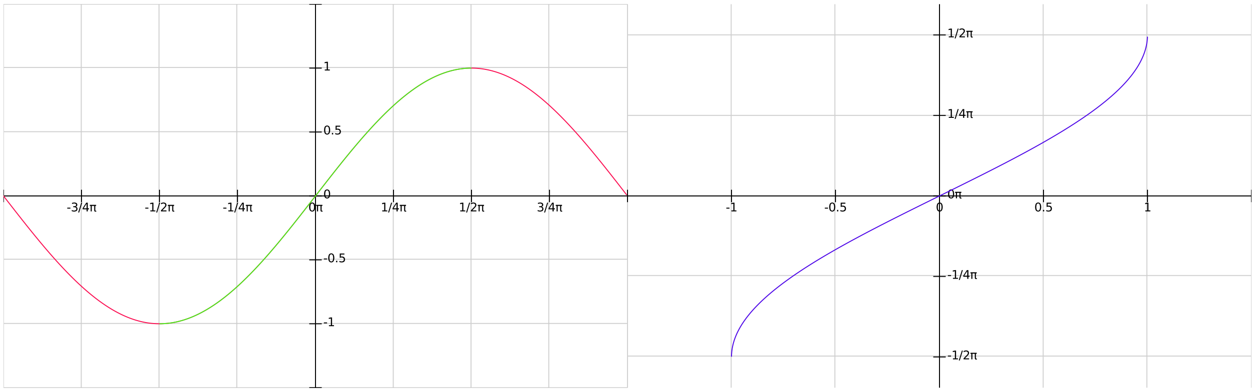
Inverse trigonometric functions have a special place in Calculus, because their derivatives are such astonishingly normal-looking functions. This means that inverse trigonometric functions sometimes pop up when you need to find anti-derivatives (see "integrals", later in this course) *even when there are no trigonometric functions in sight!* (This is a bit how the logarithm $\ln(x)$ shows up as an anti-derivative of $1/x$, a rational function.)

We also need inverse trigonmetric functions whenever we want to solve an equation like $\cos(x) = 0.3$ or $\tan(x) = 17$.

4.11.1 The inverse sine function, $\arcsin(x)$ or $\sin^{-1}(x)$

Both notations are acceptable but you must recall that $\sin^{-1}(x)$ means the inverse function of sine, NOT $\csc(x)$, DESPITE the suggestive -1. The "-1" is intended to evoke "inverse function" NOT reciprocal.

So we sketch the graph of $y = \sin(x)$; this is not one-to-one; therefore, like we did for $y = x^2$, we have to agree on a portion of the domain of $y = \sin(x)$ to which we can restrict the function. We have universally agreed on $[-\pi/2, \pi/2]$.



The graph of $y = \sin(x)$ on left, with the portion over $[-\pi/2, \pi/2]$ on which it is one-to-one highlighted in green, together with the graph of $y = \arcsin(x)$, which is the inverse of sin restricted to $[-\pi/2, \pi/2]$, and thus has domain $[-1, 1]$. Note the scales on the axes.

So we conclude that:

- sine is one-to-one on domain $[-\pi/2, \pi/2]$ with image $[-1, 1]$; so
- arcsine is defined on domain $[-1, 1]$ with image $[-\pi/2, \pi/2]$.

They are related by:

$$y = \sin(x) \Leftrightarrow x = \arcsin(y) \quad \text{for all } x \in [-\pi/2, \pi/2]$$

Example 4.31. Find all x such that $\sin(x) = 0$.

Solution: exactly one solution is given by $x = \arcsin(0) = 0$. To find all the others, we look at the graph of $y = \sin(x)$ and see that $k\pi$, for any integer k, is also a solution. □

Example 4.32. Find all x such that $\sin(x) = \frac{1}{2}$.

Solution: exactly one solution is given by $x = \arcsin(\frac{1}{2}) = \pi/6$. To find all others, we look at the graph of $y = \sin(x)$, or we use the identities:

$$\sin(x + 2\pi k) = \sin(x) \quad \text{for any integer } k$$

and

$$\sin(x) = \sin(\pi - x).$$

We see that these account for all the possible solutions, so our final answer is:

$$x = \pi/6 + 2\pi k, \quad \text{or} \quad x = (\pi - \pi/6) + 2\pi k = 5\pi/6 + 2\pi k$$

for any integer k. □

The derivative of $\arcsin(x)$

We consider the identity

$$\sin(\arcsin(x)) = x \qquad \text{for all } x \in [-1, 1].$$

So we can differentiate both sides with respect to x, using the chain rule on the left, to get

$$\cos(\arcsin(x))\frac{d}{dx}(\arcsin(x)) = 1$$

or

$$\frac{d}{dx}(\arcsin(x)) = \frac{1}{\cos(\arcsin(x))}.$$

This looks somewhat hideous: but let's simplify it.

So $\arcsin(x)$ is the angle θ, with $-\pi/2 \leq \theta \leq \pi/2$, such that $\sin(\theta) = x$. Now we know

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

and moreover, on $-\pi/2 \leq \theta \leq \pi/2$, $\cos(\theta) \geq 0$, so we can conclude

$$\cos(\arcsin(x)) = \cos(\theta) = \sqrt{1 - \sin^2(\theta)} = \sqrt{1 - x^2}.$$

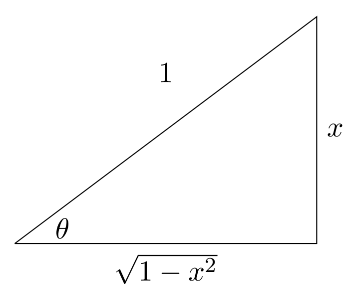
Therefore:

$$\boxed{\frac{d}{dx}(\arcsin(x)) = \frac{1}{\sqrt{1-x^2}} \qquad \text{for all } x \in (-1, 1)}$$

A quick reality check: indeed, this function is defined only for $-1 < x < 1$, which is what you'd expect for the derivative of $\arcsin(x)$. It is, nonetheless, a little shocking that the derivative of this function isn't another inverse trig function — but notice that the formula is **defintely** related to trigonometry, which is more obvious from the following example.

Example 4.33. Let's prove that $\cos(\arcsin(x)) = \sqrt{1-x^2}$ using triangles. We pretend that $0 < \arcsin(x) < \pi/2$ but the argument can be adapted for $-\pi/2 < \arcsin(x) \leq 0$ to give the same answer as well.

Draw a right angled triangle with base angle $\theta = \arcsin(x)$. Then $\sin(\theta) = x = \frac{x}{1} = \frac{opp}{hyp}$, so up to similarity, our triangle has opposite of length x and hypotenuse of length 1. By the Pythagorean theorem, the adjacent side has length $\sqrt{1-x^2}$. Therefore $\cos(\arcsin(x)) = \cos(\theta) = \frac{adj}{hyp} = \sqrt{1-x^2}$.



□

Now that we have a formula for the derivative of $\arcsin(x)$, we can use it to differentiate any function involving the arcsine function; we don't need to rederive it each time.

Example 4.34. Let $y = \arcsin(e^x + x^2)$. Then

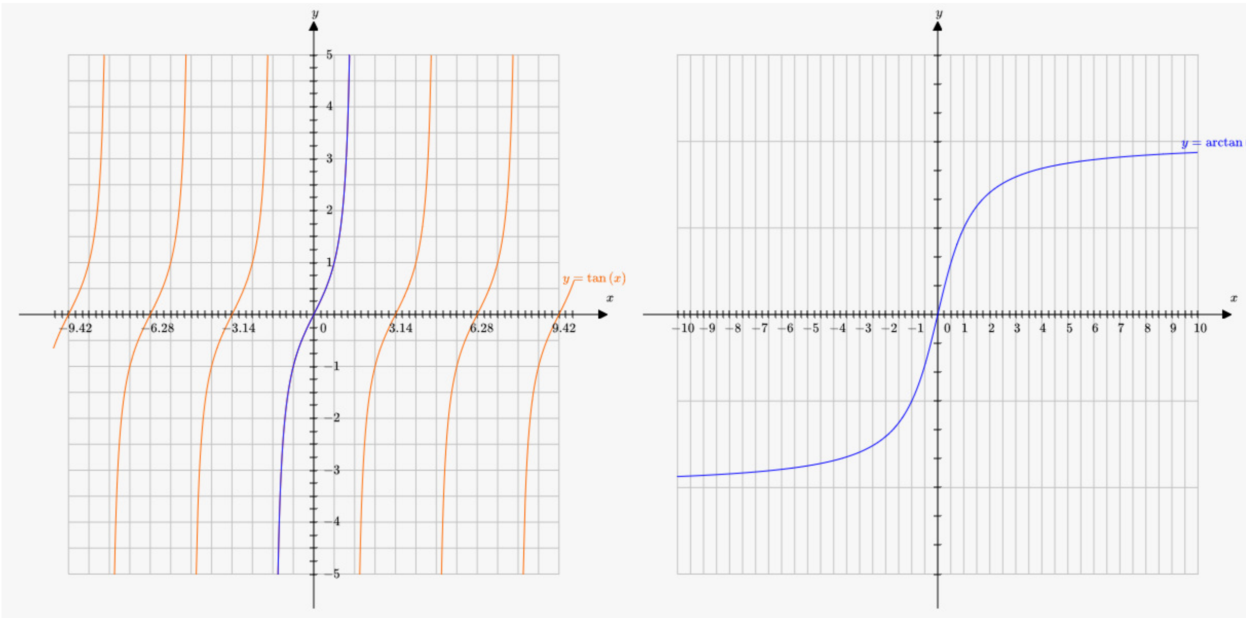
$$y' = \frac{1}{\sqrt{1-(e^x+x^2)^2}} \cdot (e^x + 2x)$$

□

Remark 4.35. Notice that when arcsin is the outermost function of a composition, it does not occur in the derivative.

4.11.2 The inverse tangent function $\arctan(x) = \tan^{-1}(x)$

Sketching the graph of $y = \tan(x)$, we see that our natural choice for restricting the domain is similar to that for $\sin(x)$; except we must exclude endpoints because of the vertical asymptotes.



The graph of $y = \tan(x)$, with a maximal portion selected on which it is one-to-one, together with the graph of $y = \arctan(x)$, which is the inverse of tan restricted to $(-\pi/2, \pi/2)$, which has domain all of $\mathbb{R}$.

- tan is one-to-one on domain $(-\pi/2, \pi/2)$ with image $\mathbb{R}$; so
- arctan is defined on domain $\mathbb{R}$ with image $(-\pi/2, \pi/2)$.

They are related by:

$$y = \tan(x) \Leftrightarrow x = \arctan(y) \quad \text{for all } x \in (-\pi/2, \pi/2)$$

Example 4.36. Find all solutions to $\tan(x) = 1$.

One solution is $\arctan(1) = \pi/4$. The graph of $y = \tan(x)$ is periodic with period π, and so in fact, we simply have that all solutions are

$$x = \pi/4 + \pi k$$

for some integer k. □

The derivative of $\arctan(x)$

Suppose $y = \arctan(x)$; we want $y' = \frac{d}{dx}\arctan(x)$. Then we have the identity $\tan(\arctan(x)) = x$; differentiating on both sides gives

$$\sec^2(\arctan(x))\frac{d}{dx}\arctan(x) = 1$$

which gives $y' = \cos^2(\arctan(x))$.

Now if $\arctan(x) = \theta$, then we deduce (by a trig identity, or using triangles) that

$$\sec^2(\theta) = 1 + \tan^2(\theta) = 1 + \tan(\arctan(x))^2 = 1 + x^2$$

so

$$\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2} \qquad \text{for all } x.$$

Again, this is reasonable; this function is defined on all of $\mathbb{R}$ and goes to 0 as x goes to $\pm\infty$, as you'd expect from the graph of $\arctan(x)$.

We can use this formula to differentiate functions involving arctan.

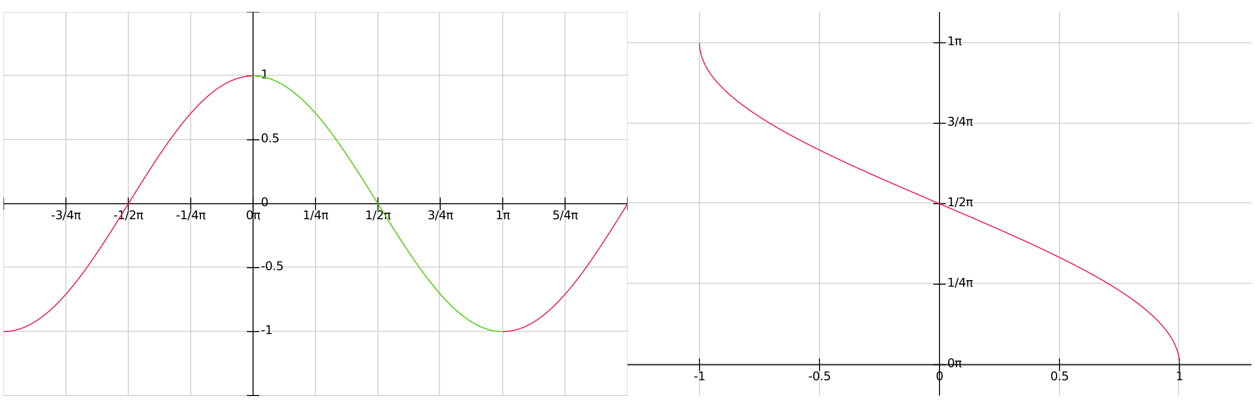
Example 4.37. Let $y = \frac{1}{\arctan(x)}$. Then $y = (\arctan(x))^{-1}$ so

$$y' = -(\arctan(x))^{-2}\frac{1}{1+x^2}.$$

(This function, you'll notice, is completely unrelated to $\sec^2(x)$, because $1/\arctan(x)$ is completely unrelated to $\tan(x)$.) ☐

4.11.3 The remaining inverse trig functions

We can do the same with $\arccos(x)$.



The graph of $y = \cos(x)$ on left, with the portion over $[0, \pi]$ on which it is one-to-one highlighted in green, together with the graph of $y = \arccos(x)$, which is the inverse of cos restricted to $[0, \pi]$, and thus is defined on domain $[-1, 1]$. Note the scales on the axes.

- cosine is one-to-one on domain $[0, \pi]$ with image $[-1, 1]$; so
- arccosine is defined on domain $[-1, 1]$ with image $[0, \pi]$.

They are related by:

$$y = \cos(x) \Leftrightarrow x = \arccos(y) \quad \text{for all } x \in [0, \pi]$$

Example 4.38. Find all solutions to $\cos(x) = \sqrt{3}/2$.

One solution is $\arccos(\sqrt{3}/2) = \pi/6$; this is the only solution in the interval $[0, \pi]$ since $\cos(x)$ is one-to-one there. To find all other solutions, we use the identities:

$$\cos(x + 2\pi k) = \cos(x) \quad \text{for all integers } k$$

and

$$\cos(-x) = \cos(x).$$

We see from the graph that these give us all other solutions. Therefore the answer is

$$x = \pi/6 + 2\pi k, \quad \text{or} \quad x = -\pi/6 + 2\pi k$$

for any integer k. □

The derivative of $\arccos(x)$

We could repeat the argument used above, but instead we might stare at the graph of $\arccos(x)$ and realize that it has a very similar shape to that of $\arcsin(x)$, because they are each portions of a sinusoidal curve in y. How can we use this?

We can relate the portions of the sine and cosine graphs on which we took the inverse functions. We know that for $0 \leq x \leq \pi$,

$$\cos(x) = \sin(\pi/2 - x)$$

with $-\pi/2 \leq \pi/2 - x \leq \pi/2$. So

$$y = \arccos(x) \Rightarrow x = \cos(y) = \sin(\pi/2 - y) \Rightarrow \pi/2 - y = \arcsin(x)$$

so that

$$\arccos(x) = \pi/2 - \arcsin(x).$$

It follows that

$$\boxed{\frac{d}{dx}(\arccos(x)) = \frac{-1}{\sqrt{1 - x^2}}}$$

So that was a bit boring.

The remaining inverse trigonometric functions

We don't tend to worry about the rest, because they can be expressed in terms of the ones we know.

For example,

$$y = \operatorname{arccsc}(x) \Rightarrow x = \csc(y) = \frac{1}{\sin(y)} \Rightarrow \sin(y) = \frac{1}{x} \Rightarrow y = \arcsin(1/x)$$

so that

$$\text{arccsc}(x) = \arcsin(1/x).$$

Similarly,

$$\text{arcsec}(x) = \arccos(1/x)$$

and

$$\text{arccot}(x) = \arctan(1/x).$$

They're interesting, but since most calculators don't even have these functions as buttons, it's kind of pointless to use them. You can compute their derivatives using the chain rule (exercise).

End of lecture # 10

4.12 Implicit differentiation

A function $y = f(x)$ is a particular kind of relation between the variables x and y — one whose graph passes the vertical line test. If our variables satisfy a relation like

$$x^2 + y^2 = 9$$

then the corresponding graph is not a function, and does not pass the vertical line test. However, we know we can decompose this graph into pieces, such that each piece is a function; in this case, the graph is the union of the graphs of

$$y = \sqrt{9 - x^2} \qquad \text{and} \qquad y = -\sqrt{9 - x^2}.$$

Now here's the clever idea: if we want to find the slope of the tangent line to the circle at a certain point, do we really *need* to solve for y in terms of x? After all, if we know that y is a function of x near each point, then we can differentiate y with respect to x. For example,

$$\frac{d}{dx}(y^2) = 2y\frac{dy}{dx}.$$

What this implies is that sometimes we can solve for y' without first having to solve for y.

For example, if $x^2 + y^2 = 9$ then near any point we can think of both sides as being functions of x; since they're equal, their derivatives are equal. So we have

$$2x + 2yy' = 0$$

and we can solve for

$$y' = -\frac{x}{y}.$$

We check that at various points (x, y) on the circle, this formula does indeed give the correct slope of the tangent line.

Remark 4.39. Notice that our answer for the derivative in this case contains both x and y! That's because unlike a function, where x is enough to determine the point on the graph, here we'll need both coordinates because there might be more than one point with that x-coordinate.

Example 4.40. Find the equation of the tangent line to the curve $x^2 + y^2 = 9$ at the point $(1, -\sqrt{8})$.

Solution: this is indeed a point on the curve, and by the preceding, the slope of the tangent line at that point is

$$m = y' = \frac{-x}{y} = \frac{-1}{-\sqrt{8}} = \frac{1}{\sqrt{8}} = \frac{\sqrt{2}}{4}.$$

A line is $y = mx + b$; given the point $(x, y) = (1, -\sqrt{8})$ and the slope $m = \frac{\sqrt{2}}{4}$ we solve to get

$$b = y - mx = -\sqrt{8} - \frac{\sqrt{2}}{4} = -2\sqrt{2} - \frac{1}{4}\sqrt{2} = -\frac{9}{4}\sqrt{2}.$$

Thus the equation of the tangent line is $y = \frac{\sqrt{2}}{4}x - \frac{9}{4}\sqrt{2}$, which you can judge to be about right (using a calculator to find out what these values are like). □

So:

- At a theoretical level, implicit differentiation is saying that if near a certain point y is a differentiable function of x, then the derivative exists *even if we can't actually find a formula for y in terms of x.* As a consequence of the chain rule, you can therefore differentiate the relation itself to deduce a formula for y'.
- At a mechanical level, implicit differentiation is saying that if you have an equation with variables which depend on x, then you can differentiate both sides with respect to x using the chain rule (remembering that $dx/dx = 1$, but $dy/dx = y'$, for example).

Example 4.41. Find the derivative of $y = \arctan(x)^{x^2+1}$.

We could rewrite this as $y = e^{(x^2+1)\ln(\arctan(x))}$ but here is an equivalent way.

Apply ln to both sides to get

$$\ln(y) = (x^2 + 1)\ln(\arctan(x))$$

and now differentiate with respect to x:

$$\frac{1}{y}y' = 2x\ln(\arctan(x)) + \frac{x^2+1}{\arctan(x)} \cdot \frac{1}{x^2+1} = 2x\ln(\arctan(x)) + \frac{1}{\arctan(x)}$$

whence, upon multiplying both sides by y we get

$$y' = \arctan(x)^{x^2+1}\left(2x\ln(\arctan(x)) + \frac{1}{\arctan(x)}\right)$$

which is the same as we'd get by the other method, of course. □

Exercise 4.42. Find the derivative of $y = (\sec(x^2))^{x^2}$ using two methods: (1) be rewriting the function as $y = e^{x^2\ln(\sec(x^2))}$; and (2) by writing $\ln(y) = x^2\ln(\sec(x^2))$ and differentiating implicitly.

More examples, not done in class

Example 4.43. Find the equation of the tangent line to the curve

$$x^{2/3} + y^{2/3} = 5$$

at the point $(1, 8)$.
Solution: we sketch the curve (called an astroid) and note that $(1, 8)$ is indeed a point on the curve. We differentiate both sides with respect to x:

$$\frac{2}{3}x^{-1/3} + \frac{2}{3}y^{-1/3}y' = 0$$

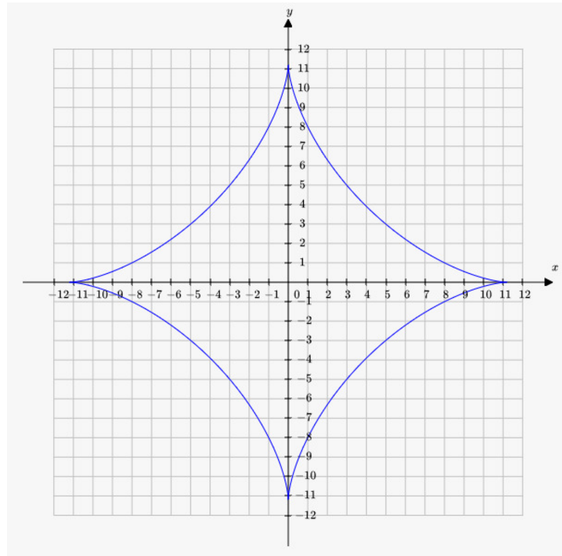
whence

$$y' = -\frac{x^{-1/3}}{y^{-1/3}} = \left(\frac{-y}{x}\right)^{1/3}.$$

At the point $(1, 8)$, we see $y' = (-8/1)^{1/3} = -2$. Therefore the equation fo the tangent line is

$$y - 8 = -2(x - 1) \qquad \text{or} \qquad y = -2x + 10$$

which looks about right from the graph. □



The graph of $x^{2/3} + y^{2/3} = 5$. This shape is called an *astroid* and is cut out by a small circle rolling along the inside of a large circle: `http://mathworld.wolfram.com/Astroid.html`.

We can also do second derivatives in the same way.

Example 4.44. For the astroid of the previous example, compute y'' at the point $(1, 8)$ and infer the concavity of the graph at that point.

Solution: We have $y' = (-y/x)^{1/3}$. Therefore, differentiating both sides with respect to x and remembering that $\frac{d}{dx}y = y'$, we have

$$\begin{aligned} y'' &= \frac{1}{3}\left(\frac{-y}{x}\right)^{-2/3}\frac{x(-y') - (-y)(1)}{x^2} \\ &= \frac{1}{3}\left(\frac{x}{-y}\right)^{2/3}\frac{y - xy'}{x^2}. \end{aligned}$$

(One could further simplify by plugging in the formula for y', which would give an answer completely in terms of x and y.) So at the point $(x, y) = (1, 8)$, where $y' = -2$, we have

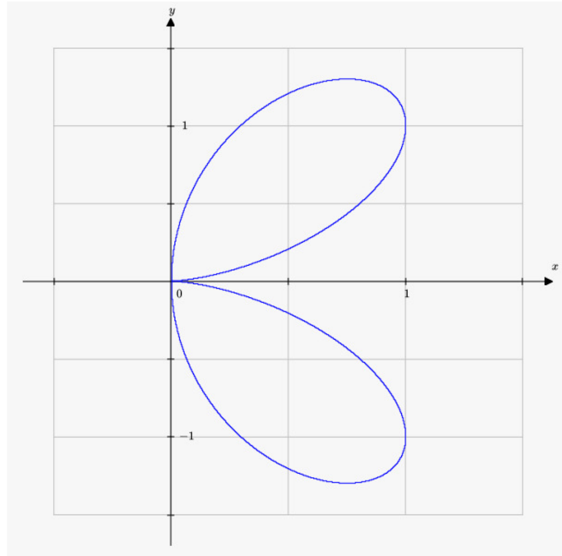
$$y'' = \frac{1}{3}(-1/8)^{2/3}\frac{8 - (-2)}{1} = \frac{1}{3} \cdot \frac{1}{4} \cdot 10 = \frac{5}{6},$$

representing a curve that is concave up (with concavity more shallow that $y = \frac{1}{2}x^2$, for example). □

Example 4.45. The *bifolium* has equation

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

Three points on the graph are $(0,0)$, $(1,1)$ and $(3/4, \sqrt{3}/4)$. Find the slope of the tangent line, when defined.



Solution: We differentiation both sides with respect to x to get

$$2(x^2 + y^2)(2x + 2yy') = 4y^2 + 4x(2yy').$$

Now we isolate and solve for y':

$$4x(x^2 + y^2) + 4y(x^2 + y^2)y' = 4y^2 + 8xyy'$$

whence

$$(4y(x^2 + y^2) - 8xy)y' = 4y^2 - 4x(x^2 + y^2),$$

or

$$y' = \frac{y^2 - x(x^2 + y^2)}{y(x^2 + y^2) - 2xy}.$$

At $(0,0)$, this is $0/0$ so undefined. On the graph we see that there's a huge mess at the origin; of course there's no tangent line.

At $(1,1)$ this is $-2/0$ so again undefined; but on the graph we see that in fact there's a vertical tangent line at this point. (So it's not a function of x there; rather, x is a function of y.)

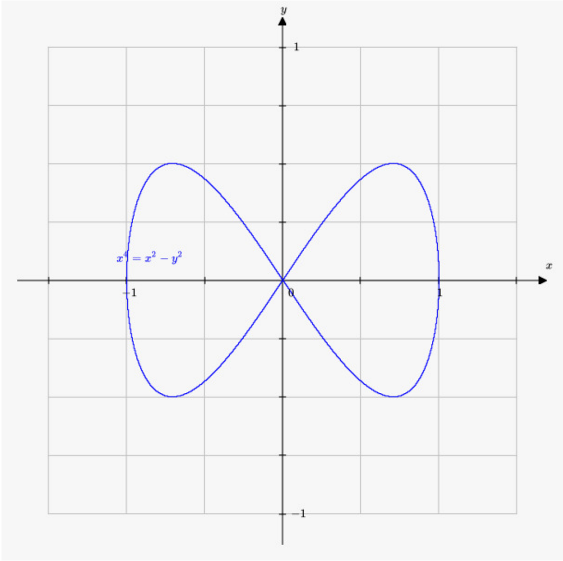
At $(3/4, \sqrt{3}/4)$, we have

$$y' = \frac{3/16 - (3/4)(3/4)}{(\sqrt{3}/4)(3/4) - (3\sqrt{3}/8)} = 2\frac{\sqrt{3}}{3}$$

which looks reasonable from the graph. □

You can even find the second derivative this way.

Example 4.46. Consider $x^4 = x^2 - y^2$. The set of all points (x, y) satisfying this equation is a lemniscate (pictured at right). We want to find y' and y'' at the point $(-\frac{1}{2}, -\frac{\sqrt{3}}{4})$.
We differentiate once, to get:



$$4x^3 = 2x - 2yy'$$

or $y' = \frac{1}{y}x(1 - 2x^2)$, after simplifying. At $(-\frac{1}{2}, -\frac{\sqrt{3}}{4})$, this is $\frac{1}{\sqrt{3}}$.
Now differentiate the relation $4x^3 = 2x - 2yy'$, noting that $\frac{d}{dx}(yy') = y'y' + yy''$ by the product rule, to get:

$$12x^2 = 2 - 2y'y' - 2yy''$$

or $y'' = (1 - (y')^2 - 6x^2)/y$ after simplifying. Pluggin in the point $(x, y) = (-\frac{1}{2}, -\frac{\sqrt{3}}{4})$ and the first derivative $y' = \frac{1}{\sqrt{3}}$ at this point yields

$$y'' = \frac{4}{3\sqrt{3}}.$$

We compare with the graph, and agree that the slope is positive and around 0.6 at that point; we agree that the curve is concave up with less concavity than a parabola of the form $y = x^2$. □

Example 4.47. Suppose $\sqrt{x}+\sqrt{y} = 1$. Find y' and y'' at the point $(1/4, 1/4)$.
We have

$$\frac{1}{2\sqrt{x}} + \frac{y'}{2\sqrt{y}} = 0$$

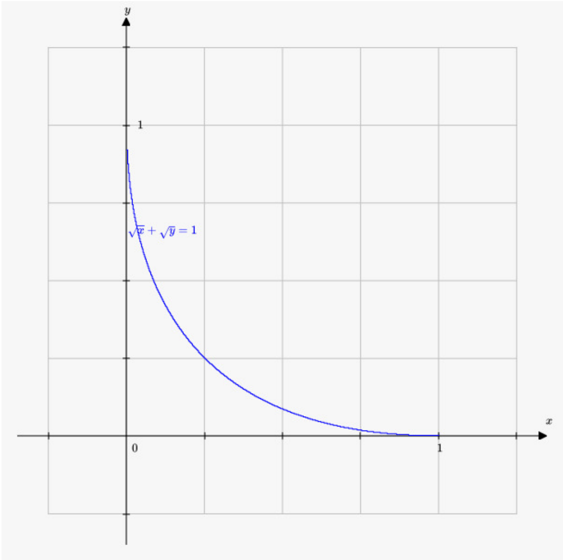
whence

$$y' = -\frac{\sqrt{y}}{\sqrt{x}}.$$

So at that point the slope is -1. The second derivative is

$$y'' = -\frac{\sqrt{x}\frac{y'}{2\sqrt{y}} - \sqrt{y}\frac{1}{2\sqrt{x}}}{|x|} = \frac{1}{|x|}\left(\frac{1}{2} + \frac{\sqrt{y}}{2\sqrt{x}}\right)$$

after simplifying with $y' = -\frac{\sqrt{y}}{\sqrt{x}}$. Therefore at that point the concavity is $y'' = 4$; it's concave up. This makes sense from the graph. □



Chapter 5

Applications of the Derivative

In the previous chapter, we learned how to differentiate everything. If a function has a formula that you recognize, then you can differentiate, or else can tell where there is a cusp or a discontinuity and therefore can't differentiate there.

Now let's explore what the derivative can give us.

5.1 The second derivative

Given a function $f(x)$, then its *second derivative* of f is the derivative of $f'(x)$, which we denote $f''(x)$. The third derivative is the derivative of $f''(x)$, and we usually write $f^{(3)}(x)$ rather than $f'''(x)$ just because it's confusing otherwise. We can define the nth derivative of a function, denoted $f^{(n)}(x)$.

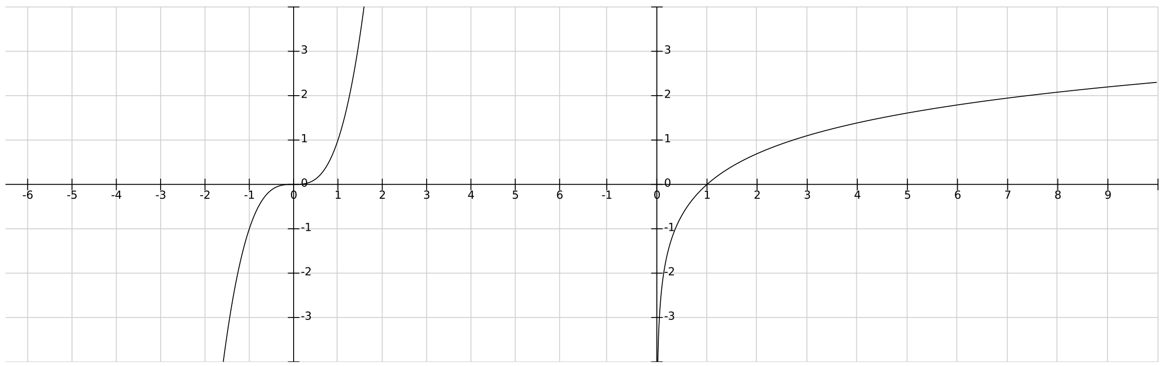
5.1.1 Concavity

The first derivative of f tells us about the rate of change of f. If f is increasing, then $f'(x) > 0$; if f is decreasing, then $f'(x) < 0$.

Therefore, the second derivative of f tells us about the rate of change of f'.

- If f' is increasing, meaning the slope of the tangent line to f is increasing (say, from negative to positive, or from positive to a larger number, or from a large-negative number to a small-negative number), then $f''(x) > 0$.
  On this kind of graph, the tangent line is under the curve.
  We call this shape *concave up*, and the shape is like that of a cup $\cup$ (or a smile).
- Similarly, if f' is decreasing, then $f''(x) < 0$.
  Since the slope is decreasing, the tangent line is on top of the curve.
  We call this shape *concave down*, and the shape is $\cap$ like that of a frown ☹.

What does this look like?

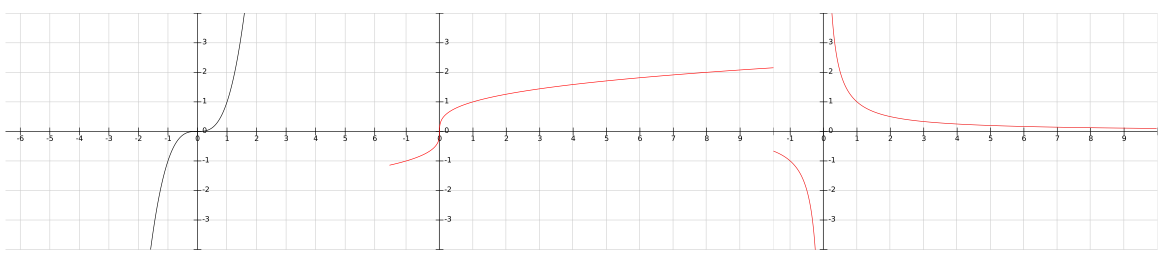


The graph of $f(x) = x^3$ on the left; its derivative is $3x^2$, which is decreasing on $(-\infty, 0)$ and increasing on $(0, \infty)$. Its second derivative is $6x$, which is negative if $x < 0$ and positive if $x > 0$. The graph is concave down when $x < 0$ and concave up when $x > 0$. So $(0, 0)$ is an inflection point of the graph.
The graph of $\ln(x)$ is on the right. Its derivative is $1/x$, on the domain $x > 0$, where it is a decreasing function. Its second derivative is $-1/x^2$, which is always negative. Thus the function is concave down on $(0, \infty)$.

A good example to know are the power functions.

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

There are three kinds of shapes for the part with $x > 0$:



The graph of $f(x) = x^3$ on the left, which on $x > 0$ is like all x^p with $p > 1$; the middle is $y = x^{1/3}$, which for $x > 0$ is similar to all x^p with $0 < p < 1$; the right one is $y = x^{-1}$, which has a shape similar to x^p for all $p < 0$.

We compute

$$f(x) = x^p, \quad f'(x) = px^{p-1}, \quad f''(x) = p(p-1)x^{p-2}.$$

So when $x > 0$, we see:

	$f(x)$ x^p	$f'(x)$ px^{p-1}	$f''(x)$ $p(p-1)x^{p-2}$	**on** $x > 0$ **only**, because $x < 0$ may not be in the domain, or be different
if $p > 1$ (like x^3 or $x^{3/2}$) :	+	+	+	increasing and concave up
if $0 < p < 1$ (like $\sqrt{x}$ or $\sqrt[3]{x}$):	+	+	-	increasing and concave down
if $p < 0$ (like $1/x$ or $1/\sqrt{x}$):	+	-	+	decreasing and concave up

5.1.2 First and second derivative tests

Recall that a *critical point* of f is a point x at which either $f'(x) = 0$ or is undefined. These are the points at which extrema (maxima or minima) of a function can occur.

Definition 5.1. A function f has a *local maximum* at $x = c$ if there is an open interval containing c which is in the domain of f and such that for every x in that interval, we have $f(x) \leq f(c)$. A function f has a *local minimum* at $x = c$ if there is an open interval containing c which is in the domain of f and such that for every x in that interval, we have $f(x) \geq f(c)$. A local maximum or a local minimum is also called a *local extremum*.

Proposition 5.2 (First Derivative Test)**.** *Suppose c is a critical point of f and f is continuous at c. Then if in a small interval on both sides of c we have*

- *if $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$, then $(c, f(c))$ is a local minimum of f;*
- *if $f'(x) > 0$ for $x < c$ and $f'(x) < 0$ for $x > c$, then $(c, f(c))$ is a local maximum of f.*

This just says that you increase to a maximum and then decrease afterwards, and vice-versa for a minimum. Thinking about concavity gives us a second test that is simpler, but sometimes doesn't apply or gives no answer:

Proposition 5.3 (Second Derivative Test)**.** *Suppose c is a critical point of f such that $f'(c) = 0$. Then*

- *if $f''(c) > 0$, then $(c, f(c))$ is a local minimum of f;*
- *if $f''(c) < 0$, then $(c, f(c))$ is a local maximum of f;*
- *if $f''(c) = 0$, then anything can happen.*

Example 5.4. Find and classify the critical points of $f(x) = x^2 e^{-x}$. This is an example of a Gamma distribution, which is used to model the expected length of time it will take for something to occur (eg: for three synaptic impulses to occur; for you to receive four phone calls) when the average time between these random events is known.

Solution: We first calculate the derivative.

$$f'(x) = 2xe^{-x} + x^2 e^{-x}(-1) = (2x - x^2)e^{-x}.$$

This is defined everywhere; since $e^{-x} \neq 0$ for any x, it is zero only when $2x - x^2 = 0$, which happens when $x = 0$ or $2 - x = 0$ so $x = 2$. To classify the critical points, we try the second derivative test:

$$f''(x) = (2 - 2x)e^{-x} + (2x - x^2)e^{-x}(-1) = (x^2 - 4x + 2)e^{-x}.$$

So $f''(0) = 2 > 0$, which tells us that is a local minimum, and $f''(2) = 4 - 8 + 2 = -2 < 0$, which tells us this is a local maximum. □

5.1.3 Inflection points

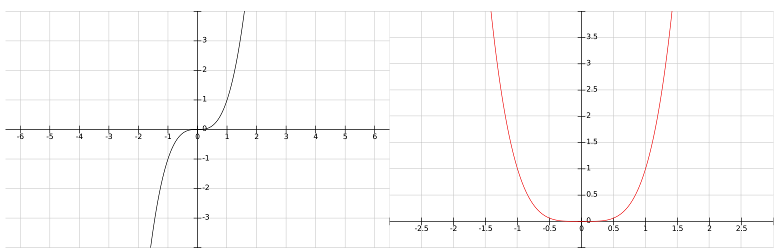
Definition 5.5. An *inflection point* on the graph of $y = f(x)$ is a point (x, y) on the curve where the concavity of the curve changes: from concave up to concave down, or vice versa.

Again, power functions give good examples to bear in mind.

Example 5.6. Consider $f(x) = x^3$ and $g(x) = x^4$. We compute

	$f(x) = x^3$	$g(x) = x^4$	
	$f'(x) = 3x^2$	$g'(x) = 4x^3$	so $x = 0$ is the only critical point
$x < 0$:	+	-	
$x > 0$:	+	+	
So:	0 is not a local extremum	0 is a local minimum	
	$f''(x) = 6x$	$g''(x) = 12x^2$	so vanishes at 0 (2nd derivative test fails!)
$x < 0$:	-	+	
$x > 0$:	+	+	
	0 is an inflection point	0 is not an inflection point	

You can see what happened from the graph:



The graphs of $f(x) = x^3$ and $g(x) = x^4$. Both have a critical point at $x = 0$ and also the second derivative is 0. We can see that f changes concavity at 0, so $(0, 0)$ is an inflection point of the graph; whereas g does not change concavity, so this is not an inflection point (but it is a local minimum).

□

5.2 Graphing functions

Our knowledge of limits and derivatives lets us graph functions by identifying the most important features, rather than by plotting points and connecting the dots. This is important for two reasons: (1) when you plot points and connect the dots, you may easily miss key features of the graph that would completely change a cobwebbing, for example; (2) when you consider functions of more that one variable next term, you definitely can't plot points anymore, because the graph is 3D (or more!).

Here are the things to look for when you are graphing a function:

1. The domain and the zeros of f;
2. The limit of f as x approaches a point not in the domain (eg, asymptotes) as well as $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$;
3. The derivative;
4. The critical points of f (i.e. where $f'(x) = 0$ or undefined) and the intervals of increase or decrease (i.e. the sign of f' between critical points);
5. The second derivative;
6. Where $f''(x) = 0$ or undefined, and the sign of f'' between these points : to tell us concavity of f, and to identify any inflection points of the graph.
7. Consistency! Make sure these clues all fit together and make sense. (Otherwise: check your work.)

Example 5.7. Graph the function $f(x) = x^2e^{-x}$.

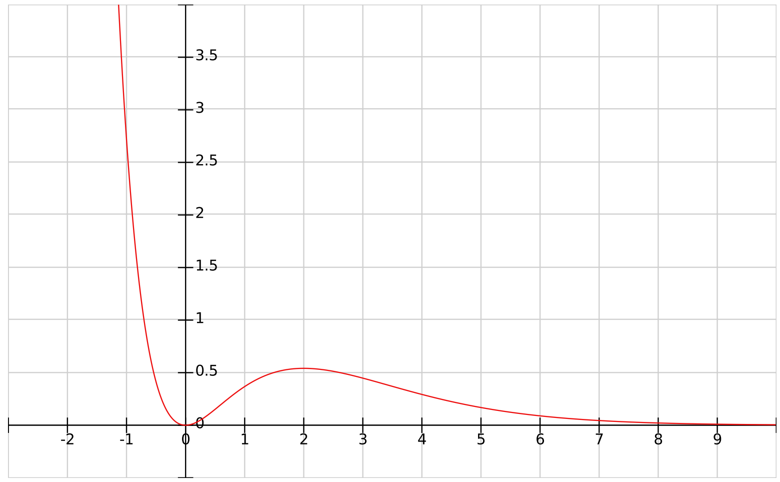
1. The domain is $\mathbb{R}$, and $f(x) = 0$ only when $x = 0$
2. $\lim_{x\to-\infty} f(x) = \infty$ since both $x^2 \to \infty$ and $e^{-x} \to \infty$; $\lim_{x\to\infty} f(x)$ — hard to say, since $x^2 \to \infty$ but $e^{-x} \to 0$ (so set this aside until we learn l'Hospital's rule); can plug in a number to guess limit is 0
3. $f'(x) = 2xe^{-x} + x^2e^{-x}(-1) = (2x - x^2)e^{-x}$.
4. Only critical points are 2 and 0 :

	$x < 0$	$0 < x < 2$	$x > 2$
$f'(x)$:	-	+	-
$f(x)$ is	decreasing	increasing	decreasing

5. $f''(x) = (2 - 2x)e^{-x} + (2x - x^2)e^{-x}(-1) = (x^2 - 4x + 2)e^{-x}$
6. This is zero only if $x^2 - 4x + 2 = 0$, so when $x = \frac{1}{2}(4 \pm \sqrt{16 - 8}) = 2 \pm \sqrt{2}$:

	$x < 2 - \sqrt{2}$	$2 - \sqrt{2} < x < 2 + \sqrt{2}$	$x > 2 + \sqrt{2}$
$f'(x)$:	+	-	+
$f(x)$ is	$\cup$	$\cap$	$\cup$

7. This gives the following graph, once you put all the pieces together.



Label all the features on your graph. The above details will give you the rough shape; to get an accurate graph, find the coordinates of key points like critical points.

□

Example 5.8. We saw that the updating function for one kind of limited population model was

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

Graph this function.

1. The domain is all $x \neq -1$, and it is zero at $x = 0$ only.

2. $\lim_{x\to-\infty} f(x) = 2 = \lim_{x\to\infty} f(x)$;
   $\lim_{x\to-1^-} f(x) = \infty$ and $\lim_{x\to-1^+} f(x) = \infty$ (in both cases: we observe the numerator is constant and the denominator goes to zero; then we look at the signs to decide if it is going to $+\infty$ or $-\infty$).

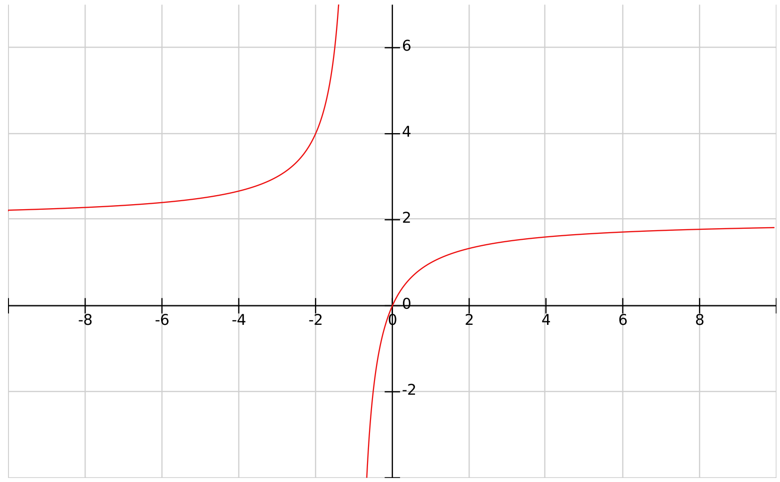
3. $f'(x) = \frac{(1+x)(2)-2x(1)}{(1+x)^2} = \frac{1}{(1+x)^2}$

4. $f'(x) > 0$ for all x, so f is increasing on each interval of its domain

5. $f''(x) = -2(1+x)^{-3}$

6. If $x < -1$ then $(1+x)^3 < 0$ so $f''(x) > 0$ and the function is concave up;
   if $x > -1$ then $(1+x)^3 > 0$ so $f''(x) < 0$ and the function is concave down

7. Since it never reaches a local maximum, but goes to a horizontal asymptote, it must approach the asymptote from one side. We sketch the graph:



□

Example 5.9. The updating function for a population showing the Allee effect was

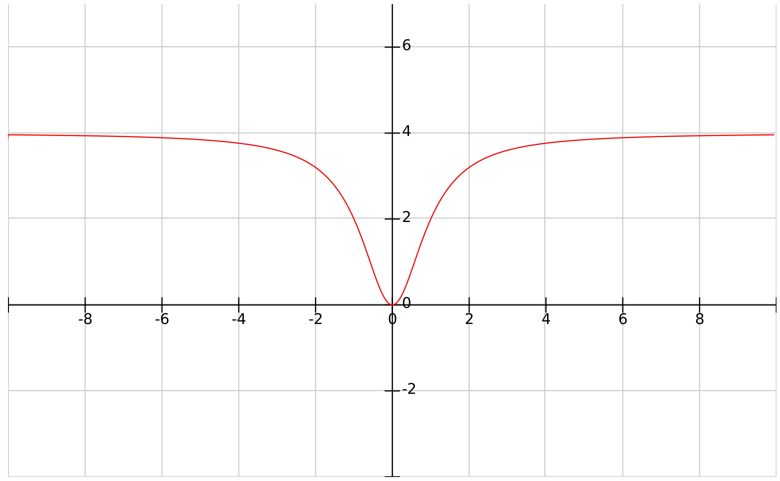
$$f(x) = \frac{4x^2}{1+x^2}$$

Let's graph it.

1. Domain $\mathbb{R}$, only 0 at 0
2. $\lim_{x\to\pm\infty} f(x) = 4$ (divide by highest power)
3. $$f'(x) = \frac{(1+x^2)(8x) - 4x^2(2x)}{(1+x^2)^2} = \frac{8x}{(1+x^2)^2}$$
4. The only critical point is 0; $f'(x) < 0$ when $x < 0$, so f is decreasing there; but $f'(x) > 0$ for $x > 0$, so f is increasing there;
5. $f'(x) = 8x(1+x^2)^{-2}$ so
$$f''(x) = 8(1+x^2)^{-2} - 16x(1+x^2)^{-3}(2x) = \left(8(1+x^2) - 32x^2\right)(1+x^2)^{-3} = \frac{8-24x^2}{(1+x^2)^3}$$
6. $f''(x) = 0$ when $24x^2 = 8$ or $x = \pm\sqrt{1/3}$. We do a table for concavity:

	$x < -1/\sqrt{3}$	$-1/\sqrt{3} < x < 1/\sqrt{3}$	$x > 1/\sqrt{3}$
$f''(x)$:	-	+	-
$f(x)$ is	concave down	concave up	concave down

7. Putting this together yields



□

End of lecture # 11

5.3 Extrema

5.3.1 Local vs global extrema

When we graphed a function f, we were particularly interested in local extrema. Recall:

Definition 5.10. A function f has a *local maximum* at $x = c$ if there is an open interval containing c which is in the domain of f and such that for every x in that interval, we have $f(x) \leq f(c)$. A function f has a *local minimum* at $x = c$ if there is an open interval containing c which is in the domain of f and such that for every x in that interval, we have $f(x) \geq f(c)$. A local maximum or a local minimum is also called a *local extremum*.

So a local maximum (respectively, minimum) is a part of the curve that's not an endpoint, and where if you zoom in close enough, $f(c)$ is the largest (respectively, the smallest) y-value of your function in that neighbourhood. It's an interesting feature of the graph.

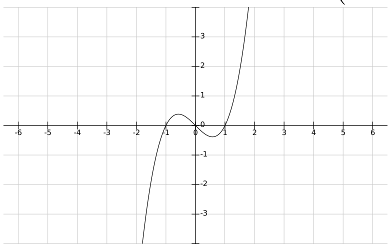
Contrast this with global, or absolute, extrema:

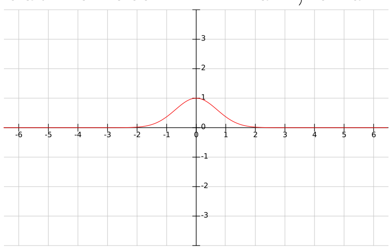
Definition 5.11. A function f attains a *global* or *absolute maximum* at a point c in its domain if $f(x) \leq f(c)$ for all x in the domain of f. Similarly, we say f attains a *global* or *absolute minimum* at a point c in its domain if $f(x) \geq f(c)$ for all x in the domain.

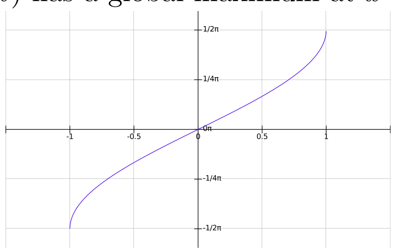
So a global maximum or minimum is the very largest or smallest y-value of the graph.

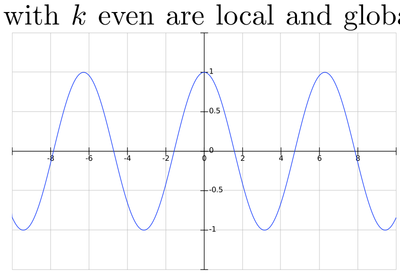
Some examples:

Example 5.12. Consider $f(x) = x(x-1)(x+1)$. This function does not attain a global maximum or a global minimum on its domain, since $\lim_{x\to\infty} x^3 = \infty$ and $\lim_{x\to-\infty} x^3 = -\infty$. It does, however, attain local extrema at $x = \pm 1/\sqrt{3}$.

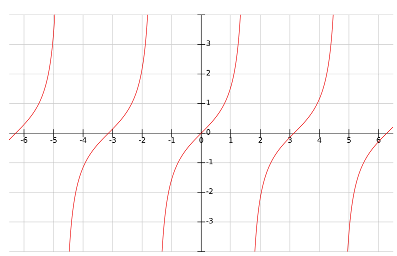
Summary: $y = x^3 - x$ has a local max (at $x = -1/\sqrt{3}$), a local min (at $x = 1/\sqrt{3}$), but no global max or min:  $\square$

Example 5.13. Consider the bell curve function e^{-x^2}. It has a local and global maximum at $x = 0$, but no local minimum, and no global minimum (since there is no value of x that gives $y = 0$):  $\square$

Example 5.14. $y = \arcsin(x)$ has a global maximum at $x = 1$, a global minimum at $x = -1$, but no local maxima or minima:  $\square$

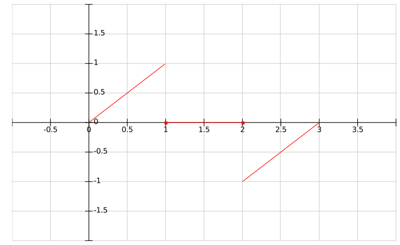
Example 5.15. $y = \cos(x)$ has infinitely many local and global maxima and minima, occuring at each peak $x = k\pi$ for $k \in \mathbb{Z}$. The ones with k even are local and global maxima, while the ones with k odd are local and global minima:  $\square$

Example 5.16. Consider $\sin(x)$. Then sin attains an absolute maximum at $x = \frac{\pi}{2}$ because $\sin(\pi/2) = 1$ and for all $x \in \mathbb{R}$, $\sin(x) \leq 1$. This is also a local maximum. Sine also attains an absolute maximum at $x = 5\pi/2$, and at $9\pi/2$ (since it also takes value 1 at those points). $\square$

Example 5.17. $y = \tan(x)$ has no local or global extrema:  $\square$

Example 5.18.

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

is not continuous, has no local extrema, and does not attain its global max or min: 

□

Summary: Not every function attains a global maximum or global minimum, and even if it does, the x-value where the extremum is attained need not be unique. Not every local extremum is a global extremum; not every global extremum is a local extremum.

5.3.2 Finding local extrema

To find local extrema, we defined critical points of f. Recall:

Definition 5.19. A number c in the domain of a function f is a *critical point* of f if either $f'(c) = 0$ or else $f'(c)$ is undefined.

We used the fact that local extrema can only occur at critical points. Recall:

Theorem 5.20 (Fermat's[1] Theorem)**.** *If f has a local extremum at c, and $f'(c)$ exists, then $f'(c) = 0$.*

Another way to say it:

If f is a continuous function, then its local extrema, if any, can only occur at critical points (but not all critical points give local extrema).

Please note: Fermat's theorem only goes **one way**. That is, just because $f'(c) = 0$ you **cannot** deduce that c gives a local extremum. Think of $f(x) = x^3$, which has a critical point at 0 but no extremum there.

Example 5.21. $f(x) = x^2$ has $f'(x) = 2x$ so 0 is a critical point; since $f''(x) = 2 > 0$, by the second derivative test, 0 is a local minimum. □

[1] Pierre Fermat had a lot of theorems; his most famous was called Fermat's Last Theorem, about solutions to equations like $x^3 + y^3 = z^3$. He'd written the theorem in the margin of a book in 1637, with a little note to the effect that he had a "marvelous proof, but the margin is too small to contain it." No one ever found the proof, but Andrew Wiles famously finally proved the theorem by other means in 1995.

Example 5.22. $f(x) = |x|$ is the function

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

so its derivative is

$$f'(x) = \begin{cases} 1 & \text{if } x > 0 \textbf{ note that we exclude } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

We know from the graph that f is not differentiable at 0. Thus the only critical point is $x = 0$. We cannot apply the second derivative test, but by the first derivative test we infer this is a local minimum. □

Example 5.23. $f(x) = x(x+1)(x-1) = x^3 - x$ has $f'(x) = 3x^2 - 1$ so two critical points where $f'(x) = 0$ or $3x^2 = 1$ or $x = \pm\frac{1}{\sqrt{3}}$. We have $f''(x) = 6x$ so by the second derivative test, $x = 1/\sqrt{3}$ gives a local minimum and $x = -1/\sqrt{3}$ gives a local maximum. This is consistent with our sketch of this cubic function (check).

So $y = x(x-1)(x+1)$ has a local maximum at $x = -\frac{1}{\sqrt{3}}$, where $f(x) = 1/\sqrt{3} - 1/3 \simeq 0.244$. But you can find many points x such that $f(x) > 0.244$, like $x = 10$ for which $f(10) = 990$ — so this local maximum is not a *global* maximum. □

Remember: A local extremum is an interesting feature of the graph of $y = f(x)$, but it need not represent an extreme value of the function.

Example 5.24. $f(x) = \sqrt{x}$ is defined only on $x \geq 0$. We have that $f'(x) = \frac{1}{2\sqrt{x}}$, which is not defined at 0. This time, although $y = f(x)$ has a minimum at $x = 0$, we do not call it a local minimum, because the function is not defined to the left of 0. A minimum is local if the function exists in a (small) interval on both sides of the point. □

5.3.3 The Extreme Value Theorem

There is a common case where we can be sure that an absolute maximum and an absolute minimum exist.

Theorem 5.25 (Extreme Value Theorem)**.** *Suppose that f is a continuous function. Then for any closed interval $[a, b]$ in the domain of f, f attains both a global maximum and a global minimum on $[a, b]$.*

This seems kind of obvious! But it's helpful to consider why we needed "continuous" and "closed interval" to make the theorem work.

Example 5.26. If f is not continuous, then it can avoid attaining an absolute extremum. For example consider

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

Sketch the graph of this function. It is not continuous. It is defined on a closed interval $[0, 3]$. Nevertheless, there is no value of x for which $f(x) = 1$ or $f(x) = -1$ (although we can get arbitrarily close!) so f does not attain an absolute max or an absolute min on $[0, 3]$. □

Example 5.27. If f is continuous, but the interval is open, then f can avoid attaining an absolute extremum. For example, consider $f(x) = 2x$ on the interval $(1, 3)$. We can get arbitrarily close to 2 and to 6, but there is no value of x such that $f(x) = 2$ or $f(x) = 6$, so f does not attain its max or min. □

Example 5.28. If f is continuous, but we consider an infinite interval like $(0, \infty)$, then f can avoid attaining an absolute extremum. For example, consider $f(x) = 1/x$. □

> Principle: If f is continuous on a closed interval, then its absolute max and absolute min must occur at critical points or endpoints of the interval.

This gives us a method.

5.3.4 Method for finding the absolute maximum and absolute minimum of a function

We can't evaluate the function at all of the points in its domain; there are infinitely many. But we can use Calculus to reduce the question to consideration of a few points. Suppose f is continuous.

1. Find all critical points of f.
2. Check each critical point (continuity, sign changes of derivative, value).
3. Evaluate f at each boundary point.
4. Compare.

If f is not continuous, or if the interval is not closed, then you should also check the limit as f approaches these bad points; but in this case, you are not guaranteed to find a value of x such that $f(x)$ is maximal (or minimal). We focus on the continuous case in this course.

Example 5.29. Let $f(x) = \sqrt{x}e^{-x}$. Find all local and global max and min.

Solution:

1. $f'(x) = \left(\frac{1}{2\sqrt{x}} - \sqrt{x}\right) e^{-x}$. The critical points are $x = 0$ (since $f'(0)$ is undefined) and where $\sqrt{x} = \frac{1}{2\sqrt{x}}$ or $x = \frac{1}{2}$.
2. Since $f'(x) > 0$ when $0 < x < 1/2$ and $f'(x) < 0$ when $x > 1/2$ (which we can see since $1/2\sqrt{x}$ is a decreasing function, and $\sqrt{x}$ is an increasing function, so their difference goes from positive to negative as we pass through the point where they are equal), we see that f is a local maximum. We have $f(1/2) = e^{-1/\sqrt{2}}/\sqrt{2} > 0$.

3. The endpoint is 0, and $f(0) = 0$.
4. Since $f(x) \geq 0$ for all $x \geq 0$, f attains a global (but not local) minimum at 0. Since $y = f(x)$ decreases everywhere after 1/2, 1/2 must be a global maximum on $(0, \infty)$.

□

Example 5.30. Consider $f(x) = |x|$ on the interval $-1 \leq x \leq 2$. Find its global maximum and minimum on this domain.

Solution:

1. The only critical point is $x = 0$.
2. This is a local minimum, by the first derivative test; we have $f(0) = 0$. This function has no other local extrema.
3. The endpoints are -1 and 2. $f(-1) = |-1| = 1$ and $f(2) = |2| = 2$.
4. Comparing these values, and knowing that the function is continuous between these points, we conclude that f attains a global max at $x = 2$ and a global min at $x = 0$.

□

Example 5.31. Consider $f(x) = \frac{\sqrt{x}}{1+x}$ on $x \geq 0$. Find its local and global extrema.

Solution:

1. $f'(x) = \frac{1}{(1+x)^2}(\frac{1}{2\sqrt{x}}(1+x) - \sqrt{x})$, so the critical points are $x = -1$ (not in the domain), and where $(1+x)/(2\sqrt{x}) = \sqrt{x}$ or $1 + x = 2x$ or $x = 1$.
2. We apply the first derivative test to the critical point 1, because calculating $f''(x)$ looks hard. In fact, we can reason as follows:

$$f'(x) > 0 \Leftrightarrow \frac{1+x}{2\sqrt{x}} > \sqrt{x} \Leftrightarrow 1 + x > 2x \;\; (\text{since } \sqrt{x} > 0\ !) \;\; \Leftrightarrow x < 1.$$

   Thus f is increasing before $x = 1$ and decreasing after, so this is a local max. In fact it must also be a global max.
3. The only endpoint is 0 and $f(0) = 0$.
4. Comparing values, we see that f has a global max at 1 and a global min at 0.

□

5.4 Optimization

Finding extreme values of functions is the goal of optimization. Respecting our constraints and our goals, we want the highest yield, the minimum cost, the highest temperature or the lowest dosage. These are all the maximum or the minimum of a function.

5.4.1 Maximization with trade-offs

The yield of crop in agriculture changes with the amount of fertilizer (for example, nitrogen) applied. When nitrogen levels in the soil are low, then adding some nitrogen will greatly increase yield. When nitrogen levels are already very high, however, adding more might *decrease* yield. Assume that yield Y as a function of the amount of nitrogen in the soil N is given by the equation

$$Y(N) = \frac{N}{1+N^2}.$$

What is the optimal level of nitrogen in the soil?

Solution. We want to choose N so as to maximize Y, so we are looking for the absolute maximum value of the function $Y(N)$.

We compute

$$Y'(N) = \frac{(1+N^2) - N(2N)}{(1+N^2)^2} = \frac{1-N^2}{(1+N^2)^2},$$

so the only critical points are where $1 - N^2 = 0$ or $N = \pm 1$. Since N is the amount of nitrogen, $N \geq 0$, and there is only one critical point to consider. Since $Y'(N) > 0$ if $0 \leq N < 1$ and $Y'(N) < 0$ if $N > 1$, we conclude that $N = 1$ gives a local and global maximum.

Thus the maximum yield is $Y(1) = \frac{1}{2}$ at it occurs at $N = 1$. □

5.4.2 Areas and volumes:

Minimize the material used to produce a cylindrical can of a fixed volume $V = 355\ cm^3$.

Solution. We draw a picture, and introduce some variables by labeling the important parts of the picture. We obtain equations by relating the variables. Once we have thus translated the question into math, we can decide what we need to optimize.

Denote by r the radius of the bottom of the can and by h its height, in cm. Then the volume is $V = \pi r^2 h = 355\ cm^3$ and the surface area is $A = 2\pi r h + 2\pi r^2$.

So we want to minimize A — but right now this is a function of two variables, r and h. We need to use the information that V is fixed; this allows us to solve for h in terms of V and r:

$$h = 355/(\pi r^2)$$

so that we rewrite

$$A = 2\pi r \cdot \frac{355}{\pi r^2} + 2\pi r^2 = \frac{710}{r} + 2\pi r^2$$

which expresses A as a function of one variable, r (as well as a parameter V, which is fixed).

Now differentiate A with respect to r to get

$$A' = -710 r^{-2} + 4\pi r$$

whose zeros are $4\pi r = 710r^{-2}$ or $r^3 = \frac{355}{2\pi}$. This has only one root, at about $r \simeq 3.8372$ cm.

End of lecture # 12

Caution: We chose more than 3 significant figures at this point because we will need our final answer to be accurate to the same precision as the data given: 3 significant figures, and round-off error comes in whenever we use estimates.

Since $A'' = 1420r^{-3} + 4\pi > 0$ for all $r > 0$, this function is always concave up, and so our critical point is a local and global minimum. The dimensions of the cylinder having volume 355 cm^3 and with minimal surface area is thus $r = 3.84$ and $h = 7.68$ (cm); the minimal surface area is $A = 277\ cm^2$. (Note all these answers were only rounded to 3 significant digits at the last step.) □

Note that if we used a parameter V in place of 355 in the above computation, we could have come up with a formula for the minimum dimensions as a function of volume: $r = \sqrt[3]{V/2\pi}$ so $h = 2r$ and $A = 6\pi r^2$.

5.4.3 Distances

Find the distance of the line $y = 1 + 2x$ from the origin and find the point on the line that is closest to the origin.

Solution. Draw a picture, assign variable names and decide what we are trying to minimize.

Here, a point on the line is (x, y), and its distance to the origin is $d = \sqrt{x^2 + y^2}$. Again, there are 3 variables and we need to cut it down to two; again, there is an equation relating x and y, namely $y = 1 + 2x$.

In this case: minimizing the distance is equivalent to minimizing the square of the distance, since the square root function is an increasing function. So let's minimize

$$D = x^2 + y^2 = x^2 + (1 + 2x)^2$$

instead, because it is easier. We have $D' = 2x + 2(1 + 2x)(2) = 10x + 4$, which has a unique critical point at $x = -\frac{2}{5}$. Since $D'' = 10 > 0$, the function is concave up everywhere, so this is a local minimum and must also be a global minimum (by concavity).

So the minimum distance occurs when $x = -\frac{2}{5}$, and thus $y = \frac{1}{5}$; the distance is

$$d = \sqrt{(-\frac{2}{5})^2 + (\frac{1}{5})^2} = \frac{1}{\sqrt{5}}.$$

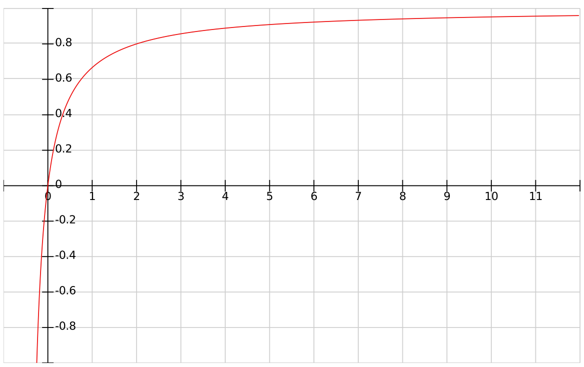
□

5.4.4 Optimize food intake by adjusting residence time.

Suppose that a bee remains at each flower for a fixed amount of time before it travels to the next flower. If that residence time is small, then the bee might leave valuable nectar behind. If it is large, then it might be depleting all the nectar and getting less than if it went to look for the next flower. What is the optimal residence time?

Approach: To answer this question, we need to know how much food the bee collects in t time units while at one flower, measured from $t = 0$ its arrival at the flower. Let's call this function $F(t)$.

Typically, F should be a positive, non-decreasing function, but would probably be concave down after some point, as the bee has to work harder or wait longer to get more nectar. An example is $F(t) = t/(t + 0.5)$:



Suppose now that the bee takes on average τ time units to fly to the next flower. Then if it spent time t at the flower and time τ flying, and gained $F(t)$ units of nectar over that time, then the rate of nectar collection is

$$R(t) = \frac{F(t)}{t + \tau}.$$

The bee wants to maximize R as a function of t, the amount of time it spends at one flower.

Solution for the specific function $F(t) = \frac{t}{t+\frac{1}{2}}$. We have

$$R(t) = \frac{t}{(t + \frac{1}{2})(t + \tau)} \quad \text{where } \tau \text{ is a positive constant parameter.}$$

Thus

$$R'(t) = \frac{(t + \frac{1}{2})(t + \tau) - t(2t + \tau + \frac{1}{2})}{(t + \frac{1}{2})^2(t + \tau)^2} = \frac{t^2 + (\tau + \frac{1}{2})t + \frac{1}{2}\tau - 2t^2 - (\tau + \frac{1}{2})t}{(t + \frac{1}{2})^2(t + \tau)^2}$$

thus

$$R'(t) = \frac{\frac{1}{2}\tau - t^2}{(t + \frac{1}{2})^2(t + \tau)^2}.$$

This is undefined when $t = -\frac{1}{2}$ and $t = -\tau$ (neither of which are biologically relevant); and it is zero when $t^2 = \frac{1}{2}\tau$ or $t = \pm\sqrt{\frac{\tau}{2}}$. Again, only the positive root is relevant.

Therefore there is only one biologically relevant critical point, $t = \sqrt{\frac{\tau}{2}}$. The denominator is positive; the sign of $R'(t)$ is positive when $0 < t < \sqrt{\frac{\tau}{2}}$ and is negative if $t > \sqrt{\frac{\tau}{2}}$, so this is a local (and in fact, on this domain) global max.

Conclusion: in this model, the bee should spend $\sqrt{\frac{\tau}{2}}$ time at each flower to maximize its average yield. For example, if $\tau = 50$ then $t = 5$; if $\tau = 2$ then $t = 1$. Funny: if the flowers are closer together, it spends less time at each flower (but spends a larger percentage of its time on flowers). □

What does it all mean? Sometimes, it's helpful to work with a more general formula so we can better see the patterns.

General Solution. So now just suppose that F is increasing and concave down and $R(t) = F(t)/(t+\tau)$ for some constant parameter τ. We differentiate R:

$$R'(t) = \frac{(t+\tau)F'(t) - F(t)}{(t+\tau)^2}.$$

This is zero when $(t+\tau)F'(t) = F(t)$ or $F'(t) = \frac{F(t)}{t+\tau} = R(t)$.

Wow: this means something! The critical point is where the *average rate of nectar collection* ($R(t)$) is equal to the *instantaneous rate of nectar collection* ($F'(t)$) on the flower.

Since F is concave down, F' is decreasing but F is increasing, so we conclude that $R' > 0$ before the critical point and $R' < 0$ after, meaning it is a local (and in fact global) maximum.

The result is the *marginal value theorem*: the bee should leave the lower if the instantaneous food intake falls below the average food intake. □

This gives us an alternative solution:

Solving the $F(t) = \frac{t}{t+\frac{1}{2}}$ model with the general solution. So, specifically, if $F(t) = \frac{t}{0.5+t}$, then $F'(t) = ((0.5+t)-t)/(t+.5)^2 = 0.5/(t+0.5)^2$. So now we solve for t by setting $F'(t) = R(t)$ (the marginal value theorem):

$$\frac{0.5}{(t+0.5)^2} = \frac{t}{(t+0.5)(t+\tau)} \Leftrightarrow 0.5(t+\tau) = t(t+0.5) \Leftrightarrow \frac{1}{2}\tau = t^2$$

so when $t = \sqrt{\frac{1}{2}\tau}$ (since t in this case must be ≥ 0). □

5.4.5 Maximize yield in a DTDS.

Assume that a population grows logistically and is being harvested regularly. In this case, we model "harvesting" by saying that at each time t, a certain fraction h of the population is removed. (Thus, a higher population yields a higher harvest, but a smaller population yields a smaller harvest.)

Actually, the way we determine harvest is relative to *last month's* population: We start with population x_t, we let it grow over the course of a month, then harvest a total of hx_t individuals. [2]

[2] So in a slow-growing population, $h < 1$, but in a fast-growing population, you could have $h > 1$ — harvesting more than what was left after the last harvest.

We model our logistic population with harvesting according to the DTDS

$$x_{t+1} = 2.5x_t(1 - x_t) - hx_t$$

where $h > 0$ is a parameter which denotes the intensity of harvesting. Let us assume that this DTDS has a stable positive steady state x^*. Then in the long term, the yield of the harvest is $Y(h) = hx^*$.

How should we choose h so that the steady state yield Y is maximized?

Solution. We notice that our function $Y(h)$ depends on h and on x^*, but not on x_t or t. So in fact, the first step is to find the equilibria of the DTDS and decide what x^* is.

Recall that x^* is an equilibrium of a DTDS with updating function f if $x^* = f(x^*)$. In this case, we have updating function

$$f(x) = 2.5x(1 - x) - hx$$

so $f(x) = x$ means

$$2.5x(1 - x) - hx = x \Leftrightarrow 2.5x - 2.5x^2 - hx = x \Leftrightarrow 2.5x^2 = (1.5 - h)x.$$

Therefore we have two equilibria, $x^* = 0$ (of course) and

$$x^* = \frac{1.5 - h}{2.5} = \frac{3/2 - h}{5/2} = \frac{3 - 2h}{5}.$$

This is positive if $3 - 2h > 0$ or $h < 3/2$. We need $h > 0$ so the range giving positive equilibria is $0 < h \leq 1.5$. In the question, we are told that this is a stable steady state[3].

Thus we now have a function for $Y(h)$ just in terms of h:

$$Y(h) = hx^* = h\frac{(3 - 2h)}{5} = \frac{1}{5}(3h - 2h^2).$$

This is a parabola which is concave down, so has a unique global maximum at its critical point. We compute $Y'(h) = \frac{3}{5} - \frac{4}{5}h$ so the unique critical point is where

$$\frac{3}{5} - \frac{4}{5}h = 0 \Leftrightarrow 4h = 3 \Leftrightarrow h = 3/4.$$

Thus the optimate harvesting rate is $h = 0.75$ and the maximum steady state harvest will be

$$Y^* = Y(\frac{3}{4}) = \frac{3}{4}(\frac{3 - \frac{3}{2}}{5}) = \frac{9}{40}$$

which is measured in the units of our original population. □

So for example if this is an annual harvest, then by harvesting 75% of the population each year our population continues to grow to a positive steady state of $x^* = 0.3$ (30% of the maximum population possible given the limited resources, as per the logistic growth model) and we harvest 75% of this.

[3] which we will learn to verify in a few lectures

5.4.6 Minimal perimeter for area

Let's start with a farmer wanting to build a rectangular yard for his sheep. The fence costs 20$ per meter, and he wants to enclose $100m^2$. What should the dimensions be to minimize cost?

Proof. We draw a picture of a rectangle and label the sides a, b, a, b, which represent length in meters. So the perimeter of the fence is $2a+2b$ and so the cost is $C = 20 \times (2a+2b) = 40a+40b$ \$.

Now C is currently a function of two variables, a and b, so we are not ready yet. We look back and see that we haven't yet taken into account that the area should be 100 m^2. The area of the rectange is ab; so the equation is $ab = 100$ or $b = 100a^{-1}$.

That gives $C = 40a+4000a^{-1}$. Excellent: cost as a function of the length of one side; let's minimize.

$$C' = 40 - 4000a^{-2}$$

so $C' = 0$ when $4000a^{-2} = 40$ or $100 = a^2$ or $a = 10$. When $a = 10$, $b = 100/a = 10$. It's a square!

To decide if it is a minimum, we compute

$$C'' = 8000a^{-3} > 0$$

so the function is concave up at $a = 10$ (and in fact on $(0, \infty)$) so this is a global minimum on the domain $(0, \infty)$.

The minimum cost is $C(10) = 400 + 400 = 800$\$. The minimal perimeter is $P = 40$. □

So the optimal shape for minimizing perimeter for fixed area (or: for maximizing area given the perimeter) if you start with a rectangle is a square: the most symmetric one.

Next question: consider the different regular polygons: equilateral triangle, square, regular pentagon, regular hexagon, etc. Could the farmer do better? With a perimeter of $P = 40$, could he get more area with a different shape? (See textbook.) Answer: a circle.

Bees in fact deal with this problem, wanting to use the least amount of material to build their honeycombs but have the most space for honey. But they don't want just one cell; they want to create dozens of them stacked together, so if there are gaps, they would waste space (and material). The only regular polygons that tile the plane are triangles, squares and hexagons — so bees use hexagons!

5.4.7 Optimal age of reproduction

Semelparous organisms, like Pacific salmon, reproduce only once in their lifetime, and then die. Typically, they can produce more female offspring as they get older, which is an advantage for population growth. But if they wait too long, then they might die before they reproduce.

What then is the optimal age of reproduction?

To answer this, we need a mathematical model and we need to refine our question.

Fact:[4] If we denote by $\ell(x)$ the probability that an individual lives to age x and by $m(x)$ the average number of female offspring of an individual at age x, then the average annual reproduction is given by

$$r(x) = \frac{\ln(\ell(x)m(x))}{x}.$$

We[5] want to maximize r as a function of x.

Specific problem: Suppose that our semelparous organism is such that $\ell(x) = e^{-ax}$ and $m(x) = bx^c$, for some positive constants a, b, c. Find the value of x that maximizes r as above.

Solution. We are given r as a function of x, so this question is purely an extreme value problem. Let's simplify r before differentiating:

$$r(x) = \frac{1}{x}(\ln(\ell(x)) + \ln(m(x))) = \frac{1}{x}(\ln(e^{-ax}) + \ln(bx^c)) = \frac{1}{x}(-ax + \ln(b) + c\ln(x))$$

so

$$r(x) = -a + \frac{\ln(b)}{x} + c\frac{\ln(x)}{x}.$$

So $r'(x) = -\ln(b)x^{-2} + c\frac{x\frac{1}{x} - \ln(x)}{x^2} = \frac{1}{x^2}(-\ln(b) + c - c\ln(x)) = \frac{1}{x^2}(c - \ln(bx^c))$. The critical points are $x = 0$ and where $c = \ln(bx^c)$. We solve this:

$$e^c = bx^c \Leftrightarrow x^c = \frac{e^c}{b} \Leftrightarrow x = eb^{-c}.$$

The critical point $x = 0$ is irrelevant, as it means a lifetime of length 0. The other critical point is positive, since $b > 0$. Since ln is an increasing function, as is x^c, we deduce that $c - \ln(bx^c)$ is decreasing, so goes from positive to negative. Thus eb^{-c} is a local maximum (and in fact global maximum).

We deduce that $x = eb^{-c}$ is a formula for the optimal age of reproduction for this species.

□

5.4.8 Optimal clutch size.

If an organism produces only few offspring, then each has a high probability of survival; if there are many offspring then the survival probability individually declines[6].

At how many offspring is the total number of survivors maximized?

Again, to get to a mathematical question, we need to convert this concept to an equation using a mathematical model.

[4]For example, see Vaupel JW, Missov TI, Metcalf CJE (2013) Optimal Semelparity. PLoS ONE8(2): e57133. https://doi.org/10.1371/journal.pone.0057133

[5]Why isn't it just $\ell(x)m(x)$? Because that's just for one individual; we have to average over the whole population size, so there are more individuals if they reproduce more often. This formula takes all that into account; see MAT2379 intro to biostatistics.

[6]This is often called r vs K strategy; see Wikipedia, for example

Let R denote the total resources (per adult female) available for reproduction and N the clutch size. Then the amount of resources per offspring is $x = R/N$. Denote the survival probability of an offspring having resources x as $f(x)$. This function should be positive (between 0 and 1) and non-decreasing (since more resources should not *decrease* survival). Then the expected number of surviving offspring is

$$w(x) = Nf(x) = \frac{R}{x}f(x).$$

We want to maximize the number of offspring w and we have expressed this number as a function of x, the amount of resources per offspring.

Let us solve this question with

$$f(x) = \frac{x^2}{x^2 + k^2}$$

for some constant $k > 0$.

Solution. So $w'(x) = R(xf'(x) - f(x))/x^2$, which gives critical points $x = 0$ (not relevant) and $xf'(x) = f(x)$.

Here,

$$f'(x) = \frac{(x^2 + k^2)(2x) - x^2(2x)}{(x^2 + k^2)^2} = \frac{2xk^2}{(x^2 + k^2)^2}$$

so

$$xf'(x) = f(x) \Leftrightarrow \frac{2x^2k^2}{(x^2 + k^2)^2} = \frac{x^2}{x^2 + k^2} \Leftrightarrow 2k^2 = x^2 + k^2 \Leftrightarrow x^2 = k^2$$

gives the only critical points as $x = \pm k$; only $x = k$ is biologically relevant.

We compute $w''(x)$ to classify this critical point. To do so efficiently, let's write

$$w'(x) = \frac{-R}{x^2}f(x) + \frac{R}{x}f'(x)$$

so that

$$w''(x) = \frac{2R}{x^3}f(x) - \frac{R}{x^2}f'(x) - \frac{R}{x^2}f'(x) + \frac{R}{x}f''(x) = \frac{2R}{x^3}(f(x) - 2Rxf'(x)) + \frac{R}{x}f''(x),$$

and at the critical point, the first term in this last expression is 0 since $xf'(x) = f(x)$. Great! So the concavity at the critical point is comes down to the sign of $\frac{R}{x}f''(x)$, which is the same as the sign of $f''(x)$.

We compute

$$\begin{aligned} f''(x) &= \frac{(x^2 + k^2)^2(2k^2) - 2xk^2(2(x^2 + k^2)(2x))}{(x^2 + k^2)^4} \\ &= \frac{(x^2 + k^2)(2k^2) - 2xk^2(2(2x))}{(x^2 + k^2)^3} \\ &= \frac{2x^2k^2 + 2k^4 - 8x^2k^2}{(x^2 + k^2)^3} \end{aligned}$$

so that $f''(k) = -4k^4/(2k^2)^3 < 0$ and thus the critical point is a local maximum.

Since the function $w(x)$ is concave down at $x = k$, we deduce that f is increasing before k and decreasing after k; and since there are no other critical points, this must therefore be a global max. □

End of lecture # 13

5.5 L'Hospital's rule for finding limits

5.5.1 Recall: Algebra of limits with infinity

For infinite limits, some arithmetic is valid; for example, let $c > 0$ be any finite number then

$$\begin{aligned}
\infty \pm c &= \infty & (-1)\infty &= -\infty \\
\infty + \infty &= \infty & \infty \cdot \infty &= \infty \\
c \cdot \infty &= \infty & & \\
\frac{c}{0+} &= \infty & \frac{c}{0-} &= -\infty \\
\frac{c}{\infty} &= 0 & \frac{\infty}{c} &= \infty \\
\frac{\infty}{0+} &= \infty & \frac{\infty}{0-} &= -\infty
\end{aligned}$$

(along with many other variations).

Example 5.32.

$$\lim_{x\to 3-} \frac{x+4}{x-3} = \text{``}\frac{7}{0-}\text{''} = -\infty$$

$$\lim_{x\to\infty} e^{2x}(\frac{1}{x} + 4) = \text{``}\infty \cdot 4\text{''} = \infty$$

□

> But the following are called *indeterminate forms* and their value cannot be assessed without analyzing the functions involved:
>
> $$\frac{0}{0} \qquad \frac{\infty}{\infty}$$
>
> $$\infty - \infty \qquad 0 \cdot \infty$$

Example 5.33.

$$\lim_{x\to 2} \frac{x^2 - 4}{x - 2}$$

is an indeterminate form of type $0/0$; to find the limit we have to algebraically manipulate the function (here, simplify) to deduce the true value. The point is that the indeterminate form holds no information about the value of the limit. □

5.5.2 Recall: evaluating limits

Earlier in the course, we talked about the limit — but then we jumped to the theorem that all our favourite functions are continuous, and in that case you can evaluate the limit

$$\lim_{x\to a} f(x)$$

by just plugging in a for x. In many cases this works when $a = \pm\infty$ as well.

Example 5.34.

$$\lim_{x\to(\pi/2)^+} \tan(x) = -\infty \quad \text{whereas} \quad \lim_{x\to(\pi/2)^-} \tan(x) = -\infty$$

$$\lim_{x\to\infty} e^x = \infty \quad \text{whereas} \quad \lim_{x\to-\infty} e^x = 0$$

So therefore

$$\lim_{x\to\pi/2^+} e^{\tan(x)} = 0$$

since as $x \to \pi/2^+$, we have $\tan(x) \to -\infty$, so $e^{\tan(x)} \to 0$. □

Example 5.35. Find $\lim_{x\to 0+} \frac{\ln(x)}{x}$.

We know that $\lim_{x\to 0+} \ln(x) = -\infty$ (so there's a vertical asymptote at 0).

So as $x \to 0+$, the numerator goes to $-\infty$ and the denominator goes to 0 (on the positive side). Dividing by a very small positive number makes you bigger — so we see that the fraction $\ln(x)/x$ also tends to $-\infty$, by our rules of "algebra with infinity."

So $\lim_{x\to 0+} \frac{\ln(x)}{x} = -\infty$. □

5.5.3 Indeterminate forms and l'Hospital's rule

Let's consider

$$\lim_{x\to\infty} \frac{\ln(x)}{x}.$$

This time we cannot reason it out so easily: both $\ln(x)$ and x go to ∞ as $x \to \infty$. We call this an indeterminate form of type ∞/∞ — indeterminate because we can't determine the answer without thinking more about the functions involved.

Another one we've previously encountered was

$$\lim_{x\to 0} \frac{\sin(x)}{x}$$

which we solved geometrically (back when we worked out the derivative of $y = \sin(x)$).

So what's going on with these types of limits? For $\ln(x)/x$ it's a question of who "gets to infinity" fastest, and for $\sin(x)/x$, it's about how quickly $\sin(x)$ goes to 0 as compared with x going to 0. In other words: if they are both headed for zero or both headed for ∞, then what we need to do is compare their rates — that is, their derivatives.

Theorem 5.36 (L'Hospital's rule)**.** *Let f and g be differentiable functions such that $g'(x)$ is nonzero around a. If*

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

is an indeterminate form of type $\frac{0}{0}$ or $\pm\frac{\infty}{\infty}$ then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

This formula is also valid for one-sided limits, and limits $x \to \pm\infty$.

Example 5.37. Type: $\frac{\infty}{\infty}$

$$\begin{aligned} \lim_{x\to\infty} \frac{\ln(x)}{x} &= \text{“}\frac{\infty}{\infty}\text{”}, \quad \text{so we can apply l'Hospital's rule} \\ &=^{L'H} \lim_{x\to\infty} \frac{1/x}{1} = 0 \end{aligned}$$

□

Example 5.38. Type: $\frac{0}{0}$

$$\begin{aligned} \lim_{x\to 0} \frac{\sin(x)}{x} &= \text{“}\frac{0}{0}\text{”}, \quad \text{so we can apply l'Hospital's rule} \\ &=^{L'H} \lim_{x\to 0} \frac{\cos(x)}{1} = \cos(0) = 1 \end{aligned}$$

Note this was a lot easier than our geometric argument — but we had to use the geometric argument back then because we were trying to find out the derivative of $\sin(x)$! □

Example 5.39. What about

$$\lim_{x\to 0+} \frac{e^x}{x} \;?$$

It is WRONG to say this $\lim_{x\to 0} \frac{e^x}{1} = 1$. Why? BECAUSE IT WASN'T AN INDETERMINATE FORM!!!

$$\lim_{x\to 0+} \frac{e^x}{x} = \text{“}\frac{1}{0+}\text{”} = \infty$$

and it's quite clear that yes, there should be a vertical asymptote at $x = 0$, and this answer is the correct one.

You CANNOT apply l'Hospital's rule UNLESS it's an indeterminate form of type $\frac{0}{0}$ or $\pm\frac{\infty}{\infty}$.

□

5.5.4 Other indeterminate forms

Type $0 \cdot \infty$

This doesn't mean actually 0, it means a limit that's tending to 0 times something that's growing to infinity.

Find

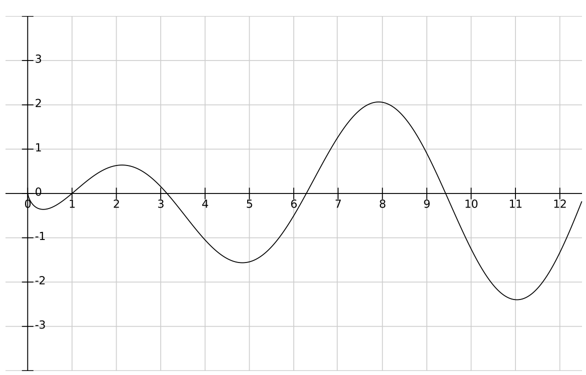
$$\lim_{x\to 0+} \sin(x)\ln(x).$$

Since $\sin(x) \to 0$ and $\ln(x) \to -\infty$, this is an indeterminate form of type $0 \cdot \infty$. Does the function $\sin(x)$ go to zero faster than $\ln(x)$ goes to $-\infty$? or vice versa? Or do their rates match and cancel?

Convert $0 \cdot \infty$ into $\frac{0}{0}$ or $\frac{\infty}{\infty}$ by the identity : $ab = \frac{b}{a^{-1}}$.

We have

$$\begin{aligned}
\lim_{x\to 0+} \sin(x)\ln(x) &= \lim_{x\to 0+} \frac{\ln(x)}{\csc(x)} \text{ since } \frac{1}{\sin(x)} = \csc(x) \\
&= \text{“}\frac{-\infty}{\infty}\text{”} \quad \text{so we can apply l'Hospital's rule} \\
&=^{L'H} \lim_{x\to 0+} \frac{1/x}{-\csc(x)\cot(x)} \\
&= \lim_{x\to 0+} \frac{1/x}{-\cos(x)/\sin^2(x)} \\
&= \lim_{x\to 0+} \frac{\sin^2(x)}{-x\cos(x)} \\
&= \text{“}\frac{0}{0}\text{”} \quad \text{so we can apply l'Hospital's rule} \\
&=^{L'H} \lim_{x\to 0+} \frac{2\sin(x)\cos(x)}{-\cos(x) + x\sin(x)} \\
&= 0
\end{aligned}$$



The graph of $y = \sin(x)\ln(x)$, which indeed tends to 0 as $x \to 0$, even though $\ln(x) \to -\infty$.

Type $\infty - \infty$

To solve, put over a common denominator. Sometimes you have to use tricks like rationalization to "create" a fraction.

$$\begin{aligned}
\lim_{x\to 0+}(\csc(x)-\cot(x)) &= \text{“}\infty-\infty\text{”, an indeterminate form} \\
&= \lim_{x\to 0+}\left(\frac{1}{\sin(x)}-\frac{\cos(x)}{\sin(x)}\right) \\
&= \lim_{x\to 0+}\frac{1-\cos(x)}{\sin(x)} \\
&= \text{“}\frac{0}{0}\text{”} \quad \text{so we can apply l'Hospital's rule} \\
&=^{L'H} \lim_{x\to 0+}\frac{\sin(x)}{\cos(x)} = 0
\end{aligned}$$

Once it's over a common denominator, plug in the values again, and decide if you've got an indeterminate form (so apply l'Hospital's rule) or else one that you can reason out.

Types ∞^0, 0^0 and 1^∞

These are all indeterminate; examples are

$$\begin{aligned}
\lim_{x\to\infty} x^{1/x} &\quad \text{type } \infty^0 \\
\lim_{x\to 0} x^x &\quad \text{type } 0^0 \\
\lim_{x\to\infty}(1+\frac{1}{x})^x &\quad \text{type } 1^\infty
\end{aligned}$$

In all three cases, it's a exponent where we have two competing rules that each say the opposite:

- "∞^0 should be ∞ because the base is ∞; but it should be 1 because any number to the power 0 is 1" (but remember that ∞ is not a number and it's only the limit as $x \to 0$, not really 0)
- "0^0 should be 0 because the base is 0, but it should be 1 because the exponent is 0"
- "1^∞ should be 1 because 1 to any power is 1 but it should be ∞ if the base is a bit bigger that 1 because anything bigger than 1 to the power ∞ is ∞"

These lines of reasoning are fine: the fact that you get contradictory answers is what says you don't have enough information, and that this is an indeterminate form.

The solution is the same for all exponential types like this:

$$y = f(x)^{g(x)} \Rightarrow \ln(y) = g(x)\ln(f(x)) \Rightarrow y = e^{g(x)\ln(f(x))}$$

So

$$\lim_{x\to\infty} x^{1/x} = \lim_{x\to\infty} e^{\ln(x)/x} = e^0 = 1 \text{(see earlier)}$$
$$\lim_{x\to 0} x^x = \lim_{x\to 0} e^{x\ln(x)} = e^{\lim_{x\to 0} x\ln(x)}$$
$$\lim_{x\to\infty} (1+\frac{1}{x})^x = \lim_{x\to\infty} e^{x\ln(1+\frac{1}{x})}$$

In the last two cases, we see that the exponent is an indeterminate form of type $0\cdot\infty$, and so we apply the techniques of the preceding section to solve them. For example,

$$\begin{aligned}\lim_{x\to\infty} x\ln(1+\frac{1}{x}) &= \lim_{x\to\infty} \frac{\ln(1+\frac{1}{x})}{\frac{1}{x}}\\ &= \lim_{x\to\infty} \frac{\frac{1}{1+1/x}(-\frac{1}{x^2})}{-1/x^2}\\ &= \lim_{x\to\infty} \frac{1}{1+1/x} = 1\end{aligned}$$

So that

$$\lim_{x\to\infty} (1+\frac{1}{x})^x = \lim_{x\to\infty} e^{x\ln(1+\frac{1}{x})} = e^1 = e.$$

This is one of the coolest formulas for e.

Summary

To find limits of continuous functions, try plugging in the values. When the result is an indeterminate form, either simplify it algebraically to find the answer, or else turn the limit into an indeterminate form of type $0/0$ or ∞/∞ so that you can apply l'Hospital's rule (as many times as it takes).

5.5.5 Using L'Hospital's rule to graph even more interesting functions

One of the reasons we need to evaluate limits is to sketch the graph of a function.

Example 5.40. Sketch the graph of $f(x) = \dfrac{\ln(x)}{x}$.

Solution:

Information from $f(x)$:

- domain is $x > 0$.
- x-intercept is $\ln(x) = 0$ or $x = 1$.
- $\lim_{x\to 0+} f(x) = -\infty$ (vertical asymptote)
- $\lim_{x\to\infty} f(x) = 0$ (by previous calculation) (horizontal asymptote)

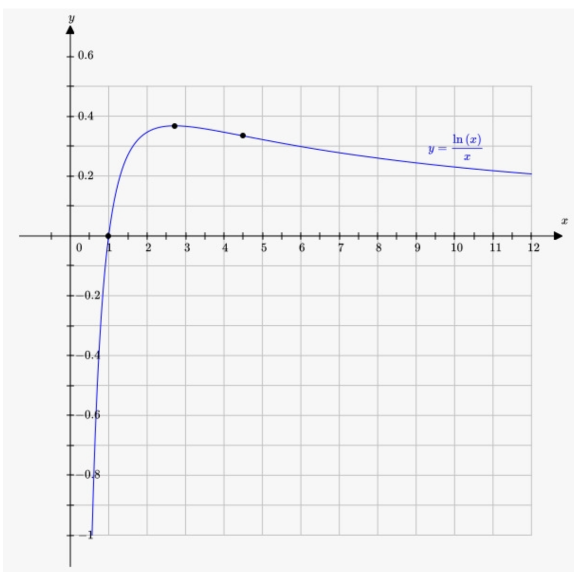
Information from $f'(x) = \dfrac{x(1/x) - \ln(x)}{x^2} = \dfrac{1 - \ln(x)}{x^2}$:

- $f'(x)$ is defined on the entire domain of f
- $f'(x) = 0$ when $1 = \ln(x)$ or $x = e$
- so only one critical point: $x = e$
- $f'(x) > 0$ if $x < e$ so increasing there
- $f'(x) < 0$ if $x > e$ so decreasing there
- conclude that there's a local maximum at $x = e$: $(e, 1/e)$, which must in fact be the absolute maximum.

Information from $f''(x) = \dfrac{x^2(-1/x) - 2x(1 - \ln(x))}{x^4} = \dfrac{-3 + 2\ln(x)}{x^3}$:

- $f''(x)$ is defined on the entire domain of f
- $f''(x) = 0$ when $3 = 2\ln(x)$ or $x = e^{3/2}$
- $f''(x) < 0$ when $x < e^{3/2}$, so concave down there — good, because our critical point in on this interval and we'd said it was a maximum!
- $f''(x) > 0$ when $x > e^{3/2}$, so concave up there
- conclude that we change concavity at $(e^{3/2}, 1.5 * e^{-3/2})$, so this is an inflection point

Putting these clues together gives the graph, and our sketch will be quite accurate.



Graph of $y = \ln(x)/x$. The intercepts, critical point and point of inflection are marked. Notice the properties of this graph are consistent will all clues from the function and its derivatives.

□

Let's do another example of graphing a function using all the information we can readily glean from the function itself, and its first and second derivatives.

Example 5.41. Let $f(x) = e^{1/x}$.

- f has domain all real numbers except $x = 0$.
- f is always positive.
- Since f is undefined at 0, we need to find the limit of f as x approaches 0, that is, what does f look like near $x = 0$? So:

  $$\lim_{x\to 0+} e^{1/x} = \text{“}e^{\infty}\text{”} = \infty$$

  since for $x < 0$, $1/x > 0$ and as $x \to 0+$ we have $1/x \to \infty$; so $e^{1/x} \to infty$ also. On the other hand

  $$\lim_{x\to 0-} e^{1/x} = \text{“}e^{-\infty}\text{”} = 0$$

  since as $x \to 0-$, $1/x \to -\infty$; but we know that $\lim_{z\to-\infty} e^z = 0$ so we deduce $e^{1/x} \to 0$. This is a weird answer; so we check, but it's right.
- Finally, we would like to know if there are horizontal asymptotes, or more generally, how the graph of f behaves as $x \to \infty$ and $x \to -\infty$:

  $$\lim_{x\to\infty} e^{1/x} = \text{“}e^{1/\infty}\text{”} = e^0 = 1$$

  and

  $$\lim_{x\to-\infty} e^{1/x} = \text{“}e^{-1/\infty}\text{”} = e^0 = 1$$

  so there are horizontal asymptotes at both ends.

Remember: a graph can cross a horizontal asymptote! (Look at our graph for $y = \ln(x)/x$, for example.)

Ok, this has given us quite a few details about the graph but now we look for the bumps and valleys, the local extrema, which really start to define the shape of the curve in between the points we've figured out so far.

We calculate

$$f'(x) = \frac{-1}{x^2}e^{1/x}$$

and then:

- $f'(x)$ is undefined only at $x = 0$, which is not in the domain of f anyway
- $f'(x) = 0$ is never true, since $f'(x)$ is a product of two functions and neither one is ever zero.
- So f has no critical point on its domain, meaning it attains no local extrema

- We see that $e^{1/x} > 0$ for all $x \neq 0$, and $-1/x^2 < 0$ for all $x \neq 0$. So $f'(x) < 0$ for all $x \neq 0$.
- This says f is decreasing **on every connected component of its domain**. That is, f is decreasing on $(-\infty, 0)$ and also on $(0, \infty)$.

Remark 5.42. We might also want to know

$$\lim_{x\to 0-} \frac{-1}{x^2} e^{1/x}$$

since that tells us the angle at which we will be approaching $(0, 0)$. (We don't need to ask about the horizontal asymptotes (obviously the graph flattens out to horizontal at infinity) or the vertical asymptote (obviously the graph gets steeper and steeper). The above limit is an indeterminate form of type $0/0$ so we use l'Hospital's rule

$$\lim_{x\to 0-} \frac{-1}{x^2} e^{1/x} =^{L'H} \lim_{x\to 0-} \frac{e^{1/x} x^{-2}}{2x} = \lim_{x\to 0-} \frac{e^{1/x}}{2x^3}$$

YUCK! This is worse than what we started with!! So let's flip things around and see if it improves:

$$\begin{aligned}
\lim_{x\to 0-} \frac{-1}{x^2} e^{1/x} &= \lim_{x\to 0-} \frac{-x^{-2}}{e^{-1/x}} \quad \text{indeterminate form } \frac{\infty}{\infty} \\
&=^{L'H} \lim_{x\to 0-} \frac{2x^{-3}}{e^{-1/x}(-x^{-2})} \\
&= \lim_{x\to 0-} \frac{-2x^{-1}}{e^{-1/x}} \quad \text{indeterminate form } \frac{\infty}{\infty} \\
&=^{L'H} \lim_{x\to 0-} \frac{2x^{-2}}{e^{-1/x}(-x^{-2})} \\
&= \lim_{x\to 0-} -2e^{1/x} = 0
\end{aligned}$$

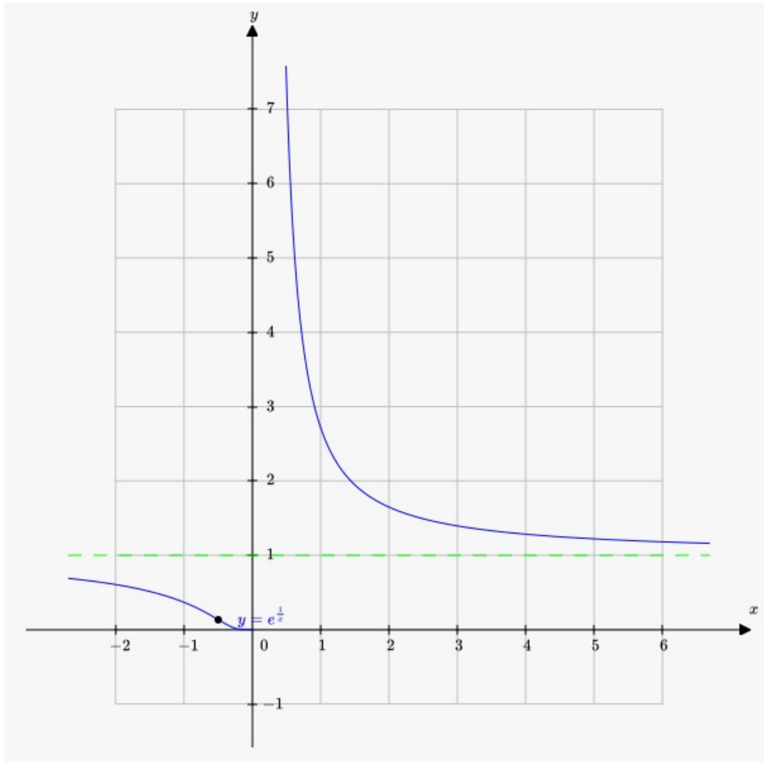
so the graph comes in to $(0, 0)$ from the left at a shallow angle.

Now for the second derivative:

$$f''(x) = \frac{x^2(-e^{1/x}(-x^{-2})) - (-e^{1/x})(2x)}{x^4} = e^{1/x} \frac{1 + 2x}{x^4}$$

- $f''(x)$ is undefined only for $x = 0$.
- $f''(x) = 0$ when $1 + 2x = 0$ or $x = -\frac{1}{2}$
- For $x < -\frac{1}{2}$, $f''(x) < 0$ so the graph is concave down
- For $-\frac{1}{2} < x < 0$, $f''(x) > 0$ so the graph is concave up
- For $x > 0$, $f''(x) > 0$ so the graph is concave up
- We see that $(-0.5, e^{-2})$ is an inflection point since the graph changes concavity there

From these details, we piece together a very good sketch of the graph.



Graph of $y = e^{1/x}$. The point of inflection and horizontal asymptote are marked. Notice the properties of this graph are consistent will all clues from the function and its derivatives and the values of the limits as $x \to 0^{\pm}$ and $x \to \pm\infty$.

□

You should do practice problems with functions that make you fearful — because they will challenge you! The hard parts are finding the derivatives and then finding the critical points of f and of f' — once you have those, it becomes a fun puzzle of connecting the dots.

Fun ones include $f(x) = xe^{-1/x}$ and $f(x) = \dfrac{x}{\sqrt{x^2+1}}$.

End of lecture # 14

5.6 Approximating functions with polynomials

Having the full graph of a function, or a formula for it, is wonderful. But many functions are difficult to work with and to evaluate (such as $\sqrt{x}$, $\ln(x)$, e^x) compared to polynomial functions (such as linear, quadratic functions). Can we locally approximate the more complex functions with polynomials?

This kind of approximation is used to simplify complex mathematical calculations. But it is also important when you need to make complex mathematics accessible. For example, if the absorption model of a drug is given by a complicated function, but you want those who use it to understand the effect of modifying their dosage, then you are better off estimating it with a linear function that can be explained in words.

As an example of how we routinely estimate instead of using the actual function: if we want to estimate $\sqrt{150}$, we would say: $\sqrt{144} = 12$ and $\sqrt{169} = 13$ so since $\sqrt{x}$ is a monotone (increasing) function, we know that

$$12 < \sqrt{150} < 13.$$

Can we get a better estimate, without a calculator?

5.6.1 Linear approximation

The point of Calculus and the derivative is that locally near any point, the graph of a differentiable function is quite close to its tangent line. In practice, this means you can estimate $f(x)$ for x near a by its *linear approximation*, meaning, the function for this tangent line.

That is: think of the tangent line to f at a as another function (a far simpler function!!!) which is pretty close to f near a.

What is a formula for the tangent line to f at $(a, f(a))$?

$$y - f(a) = f'(a)(x - a)$$

or

$$y = f(a) + f'(a)(x - a).$$

So given a function $f(x)$ and a base point a, the tangent line to f at a is $$T(x) = f(a) + f'(a)(x - a)$$

and it is a linear approximation to f at a in the sense that

- it's a linear function, and
- $T(x)$ is close to $f(x)$ for x near a, and
- $T(a) = f(a)$, that is, the functions coincide at a.

Example 5.43. Consider $f(x) = \sqrt{x}$ at $a = 100$. We have $f'(x) = \frac{1}{2\sqrt{x}}$. Then

$$T(x) = f(100) + f'(100)(x - 100) = \sqrt{100} + \frac{1}{2\sqrt{100}}(x - 100) = 10 + \frac{1}{20}(x - 100).$$

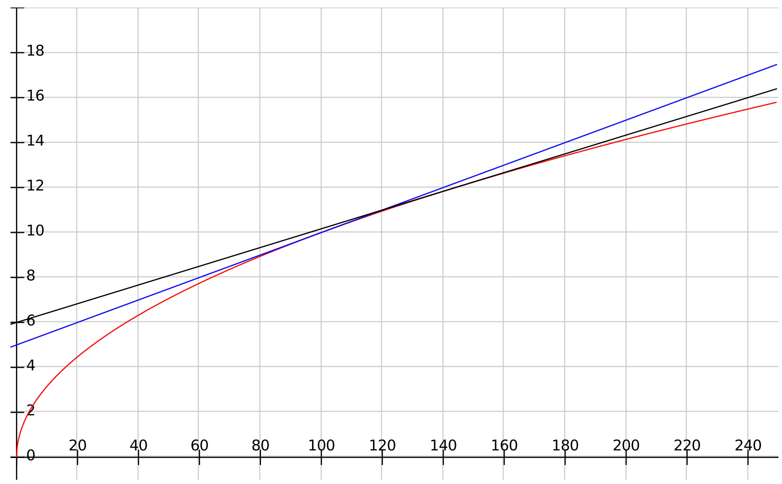
So $\sqrt{150} \sim T(150) = 10 + \frac{1}{20}(50) = 12.5$. Check: $\sqrt{150} \sim 12.247$, not bad. $\square$

We should get a better approximation if we choose a base point closer to 150.

Example 5.44. Consider $f(x) = \sqrt{x}$ with base point $a = 144$. Then

$$T(x) = f(144) + f'(144)(x - 144) = 12 + \frac{1}{24}(x - 144)$$

so $\sqrt{150} \sim T(150) = 12 + \frac{1}{24}(6) = 12.25$, which is **really** close. $\square$



The graph of $y = \sqrt{x}$ is in red. Its tangent line at the base point $a = 100$ is given in blue and its tangent line at the base point $a = 144$ is given in black. Compare the closeness of the approximation at $x = 150$.

Example 5.45. Find the linear approximation to $f(x) = e^x$ near $x = 0$.

Solution: We write down the equation of the tangent line.

$$T(x) = f(a) + f'(a)(x - a) = e^0 + e^0(x - 0) = 1 + x.$$

We compare:

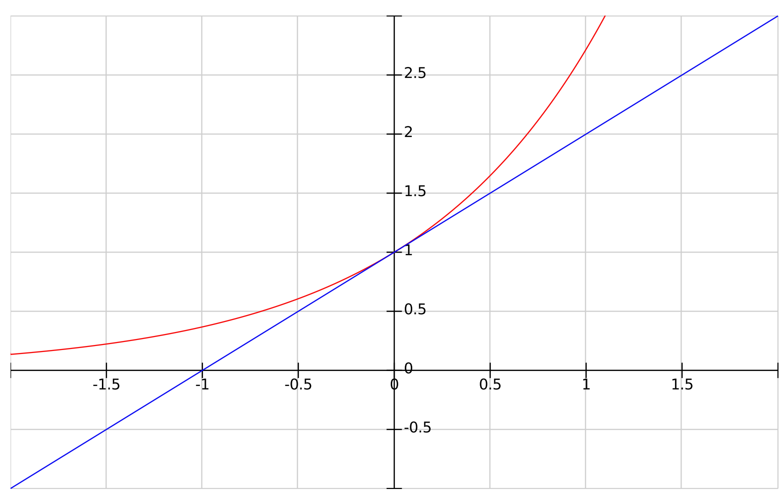
$$f(0.1) = e^{0.1} = 1.10517$$

whereas

$$T(0.1) = 1 + 0.1 = 1.1$$

and we needed a calculator to find $f(0.1)$ whereas we didn't for $T(0.1)$.

However, if x is not near a, then your linear approximation won't be very good: for example $e^1 \sim 2.718$ but $T(1) = 1 + 1 = 2$. This is obvious from the graph.



The graph of $y = e^x$ and its tangent line $T(x) = 1 + x$ at the origin.

□

The linear approximation $L(x) = f(a) + f'(a)(x - a)$ is useful anytime you know $f(a)$ and $f'(a)$ and want to estimate $f(x)$ for x near a.
Also: if your function is given numerically, then $L(x)$ is really just some linear extrapolation from your data.

Remark 5.46. In Physics, it is routine to replace $\sin(x)$ or $\cos(x)$ with their linearizations at 0 when the physical problem involves smaller angles. For example, if

$$f(x) = \sin(x), \qquad a = 0$$

then its linear approximation is

$$L(x) = \sin(0) + \cos(0)(x - 0) = 0 + 1(x) = x.$$

Replacing $\sin(x)$ by x makes it easier to solve for x in certain formulas, such as approximating the period of oscillation of a pendulum, or in the defraction of light through a lens.

5.6.2 Taylor polynomials

Linear approximation is fine, but we could do better. The second derivative tells us about the concavity of the function, so it we took it into account, we could find a simple function that matches both the slope of f and the concavity of f at a point a.

Example 5.47. Our linear approximation of $\sqrt{150}$ was too high, because the graph of $y = \sqrt{x}$ is concave down near $x = 150$. Our linear approximation of $e^{0.1}$ was too low, because the graph of $y = e^x$ is concave up near $x = 0.1$. □

Theorem 5.48. *Let f be a function*[7] *and a a base point. Then for any $n > 0$ there is a polynomial of degree n, called the* Taylor polynomial of degree n *and denoted $T_n(x)$, with the property that*

$$T_n(a) = f(a), \quad T_n'(a) = f'(a), \quad T_n''(a) = f''(a), \quad \cdots \quad T_n^{(n)}() = f^{(n)}(a)$$

where $f^{(n)}(a)$ denotes the nth derivative of x evaluated at the number a.

That's quite cool, but even cooler is the *formula for $T_n(x)$*:

$$T_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

Note that:

- $f^{(n)}(a) = \frac{d^n}{dx^n} f(x)$ at $x = a$, the n-th derivative;
- $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot (n-1) \cdot n$, called "$n$-factorial" : for example $3! = 1 \cdot 2 \cdot 3 = 6$, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$ and by convention we say $0! = 1$;
- $T_1(x) = T(x)$, the linear approximation = the equation of the tangent line to f at a;

[7] such that the first n derivatives of f exist at a

- You get the nth approximation from the $(n-1)$st approximation but just adding one term : $f^{(n)}(a)/n!(x-a)^n$.

Example 5.49. So let $f(x) = \sqrt{x}$, so that $f'(x) = \frac{1}{2\sqrt{x}}$, $f''(x) = \frac{-1}{4x^{3/2}}$. At $a = 100$ we have $\sqrt{a} = 10$ and $a^{3/2} = 10^3 = 1000$, so

$$\begin{aligned} T_2(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 \\ &= 10 + \frac{1}{20}(x-100) + \frac{-1}{2 \cdot 4000}(x-100)^2 \\ &= 10 + \frac{1}{20}(x-100) - \frac{1}{8000}(x-100)^2 \end{aligned}$$

(which will indeed be less than our linear approximation, as desirec) and we can evaluate

$$T_2(150) = 10 + \frac{50}{20} - \frac{50^2}{8000} = 10 + \frac{5}{2} - \frac{5}{16} = 12.1875.$$

This time we are a little too small — but much closer. □

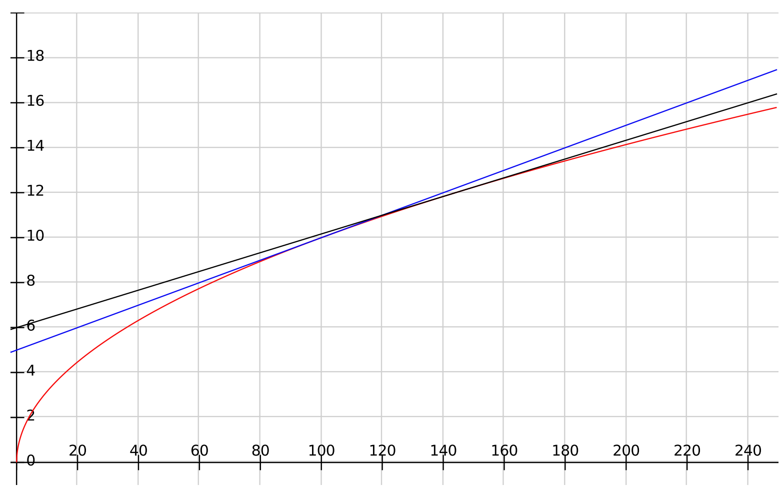
We can always do better by choosing the base point a closer to the x at which we want the approximation.

Example 5.50. Find $T_2(x)$ for $f(x) = \sqrt{x}$ at $a = 144$, and use it to estimate $\sqrt{150}$.

Solution: as above, we need to know $\sqrt{144} = 12$, $144^{3/2} = 12^3$ so

$$T_2(x) = 12 + \frac{1}{24}(x-144) - \frac{1}{2 \cdot 4 \cdot 12^3}(x-144)^2 = 12.25 - \frac{36}{8 \cdot 12^3} = 12.25 - \frac{1}{384} \sim 12.247$$

which is correct to three decimal places! □



The graph of $y = \sqrt{x}$ is in red. Its quadratic Taylor approximation at the base point $a = 100$ is given in blue and its quadratic Taylor approximation at the base point $a = 144$ is given in black. Compare the closeness of the approximation over a large interval to that of the linear approximation.

So we can get a better approximation by choosing a base point closer to x; or we can get a better approximation by choosing a higher order Taylor polynomial.

Example 5.51. Find $T_3(x)$ for $f(x) = \sqrt{x}$ at base point $a = 100$, and use it to estimate $\sqrt{150}$.

Solution:

$$\begin{aligned} T_3(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 \\ &= \sqrt{a} + \frac{1}{2\sqrt{a}}(x-a) + \frac{-1}{2 \cdot 4(a)^{3/2}}(x-a)^2 + \frac{1}{6} \cdot \frac{3}{8a^{5/2}}(x-a)^3 \\ &= 10 + \frac{1}{20}(x-100) - \frac{1}{8000}(x-100)^2 + \frac{1}{16 \cdot 10^5}(x-100)^3 \\ &= 12.1875 + \frac{5}{64} \sim 12.265. \end{aligned}$$

□

So we can approximate $\sqrt{x}$ with a linear, quadratic, cubic or even higher order polynomial.

What happens if we do the cubic Taylor approximation of a cubic function?

Example 5.52. Find the cubic Taylor polynomial of $f(x) = x^3$ at the point $a = 8$ and use it to estimate 8.1^3.

Solution: We have $f(x) = x^3$ so $f'(x) = 3x^2$, $f''(x) = 6x$ and $f'''(x) = 6$. Thus $f(a) = 8^3 = 512$, $f'(a) = 3(64) = 196$, $f''(a) = 6(8) = 48$ and $f'''(a) = 6$. Then

$$T_3(x) = 512 + 192(x-8) + \frac{48}{2}(x-8)^2 + \frac{6}{3!}(x-8)^3 = 512 + 192(x-8) + 24(x-8)^2 + (x-8)^3$$

which gives

$$T_3(8.1) = 512 + 19.2 + 0.24 + 0.001 = 531.441$$

which is exactly spot on!

Amazing, right? Well, not so amazing. If you multiply out the formula for $T_3(x)$ you get simply x^3. So Taylor's theorem in this case just gives us a really cool refactorisation of x^3 which makes it easy to evaluate near the base point. □

5.6.3 More examples of Taylor approximations

Example 5.53. Find the 4th Taylor polynomial of $f(x) = \cos(x)$ at the base point $a = 0$.

Solution: We make a table:

	function	function evaluated at $a = 0$	Taylor polynomial coefficient $\frac{1}{n!}f^{(n)}(a)$
$f(x)$	$\cos(x)$	1	1
$f'(x)$	$-\sin(x)$	0	0
$f''(x)$	$-\cos(x)$	-1	$\frac{-1}{2}$
$f'''(x)$	$\sin(x)$	0	0
$f^{(4)}(x)$	$\cos(x)$	1	$\frac{1}{4!} = \frac{1}{24}$

and conclude that

$$T_4(x) = 1 - \frac{1}{2}(x-0)^2 + \frac{1}{24}(x-0)^4 = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4.$$

□

Remark 5.54. In fact, you can continue this pattern infinitely to get the *Taylor series* of f at $a = 0$:

$$\cos(x) = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 - \frac{1}{6!}x^6 + \frac{1}{8!}x^8 \pm \cdots$$

and yes, if you really could add all these infinitely many terms, the answer would completely equal $\cos(x)$. Basically, this is how your calculator evaluates $\cos(x)$ (with x in RADIANS!!); it works better than drawing triangles and measuring ratios...

Example 5.55. Find the third Taylor polynomial of $\ln(x)$ at base point $a = 1$ (you can't use $a = 0$ here!!) and use it to estimate $\ln(1.1)$.

Solution: We make a table

	function	function evaluated at $a = 1$	Taylor polynomial coefficient $\frac{1}{n!}f^{(n)}(a)$
$f(x)$	$\ln(x)$	0	0
$f'(x)$	$1/x$	1	1
$f''(x)$	$-1/x^2$	-1	$\frac{-1}{2}$
$f'''(x)$	$2/x^3$	2	$\frac{2}{3!} = \frac{1}{3}$

so we have

$$T_3(x) = 0 + (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3$$

and thus we estimate $\ln(1.1)$ by

$$T_3(1.1) = 0.1 - \frac{1}{2}(0.01) + \frac{1}{3}(0.001) = 0.0953.$$

With a calculator, we check that $\ln(1.1) = 0.0953101798$; pretty fabulous approximation to have done by hand. □

Again, you can infer a pattern from the derivatives and figure out the Taylor series in this case.

Exercise 5.56. Estimate $\sqrt{6}$ using the cubic Taylor polynomial of $f(x) = \sqrt{x}$ at the base point $a = 4$.

Exercise 5.57. Estimate $\sin(1)$ (radians!!!!) using the quintic Taylor polynomial of $f(x) = \sin(x)$ at the base point $a = 0$.

5.6.4 Estimating a function using a secant line

In some cases (eg, when you are providing a simple approximation to a drug absorption model for people to understand), you want to have an approximation that is good over an entire interval, rather than just good close to one point.

In these circumstances, we might approximate our function by a *secant line approximation*. That is, given a function $f(x)$ and two points a, b defining an interval in the domain of the function, the secant line of f from a to b is the straight line from

$$(a, f(a)) \quad \text{to} \quad (b, f(b)).$$

The slope of the secant line is thus

$$m = \frac{f(b) - f(a)}{b - a}$$

and the equation for the line is

$$y = f(a) + m(x - a) = f(a) + \left(\frac{f(b) - f(a)}{b - a}\right)(x - a).$$

We usually leave this is factored form, because it's quite natural to evaluate $x - a$.

Example 5.58. Give the secant line approximation of the function $\sqrt{x}$ on the interval $[100, 169]$, and use it to approximate $\sqrt{150}$.

Solution: We have $(a, f(a)) = (100, 10)$ and $(b, f(b)) = (169, 13)$ so

$$m = \frac{f(b) - f(a)}{b - a} = \frac{3}{69} = \frac{1}{23}$$

and thus the secant line is given by

$$y = 10 + \frac{1}{23}(x - 100).$$

When $x = 150$, this gives $y = 10 + \frac{50}{23} \sim 12.17$, compared with $\sqrt{150} \sim 12.247$. □

5.7 The Mean Value Theorem

We know how to get from function to its derivative : take the limit of the slope of the secant line. But soon we will want to go backwards: from the derivative back to the function. The first clue is a deceptively simple theorem which will end up being the key to understanding how things work, called the *Mean Value Theorem*.[8]

Theorem 5.59 (Mean Value Theorem)**.** *Suppose f is a function that satisfies the following hypotheses:*

(a) f is continuous on the closed interval $[a, b]$ (and maybe more); and

(b) f is differentiable on the interval (a, b) (and maybe more).

Then there is at least one (unknown) value c somewhere in the interval $[a, b]$ for which

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

[8]This is one of the grand trio of key theorems about continuous and differentiable functions: the Intermediate Value Theorem, the Mean Value Theorem and the Extreme Value Theorem.

That is, this theorem says that for a nice function like f, at some point in the interval, the instantaneous rate of change of f is equal to the average rate of change over the whole interval (or, the slope of the secant line on that interval).

Let us agree that this is very plausible. If you drive from Ottawa to Montreal (200km) in 2 hours ($f(b) - f(a) = 200$, $b - a = 2$) you could not have driven less than 100km/h (your average speed) the whole time; nor could you have driven more than 100km/h the whole time. At some instant — maybe just ONE instant in the whole 2 hours, maybe for lots of minutes during those two hours, you were driving exactly 100km/h. Any one of those instants could be called c.

End of lecture # 15

5.7.1 Applications of the Mean Value Theorem (optional)

Sometimes we can solve for c.

Example 5.60. Suppose $f(x) = x^3$ on the interval $[0, 1]$. Then the slope of the secant line on that interval is

$$m = \frac{f(b) - f(a)}{b - a} = \frac{1^3 - 0^3}{1 - 0} = 1.$$

The derivative of $f(x)$ is $f'(x) = 3x^2$. So the derivative equals the slope of the secant line when

$$3x^2 = 1 \Leftrightarrow x^2 = \frac{1}{3} \Leftrightarrow x = \pm\frac{1}{\sqrt{3}}.$$

So setting $c = \frac{1}{\sqrt{3}}$ we have $f'(c) = 1$ and $a \leq c \leq b$, as required. □

But where this is particularly interesting is when we *can't* solve for c — the theorem still tells us that c exists.

Example 5.61. Consider $f(x) = \sin(x)\ln(x)$ on the interval $[\pi, 2\pi]$. Since $f(\pi) = 0$ and $f(2\pi) = 0$, the slope of the secant line is 0. Therefore, by the Mean Value Theorem, there is some number c between π and 2π such that $f'(c) = 0$. If we calculate:

$$f'(x) = \cos(x)\ln(x) + \sin(x)/x$$

we are completely stuck: it is impossible to solve $f'(x) = 0$!! But we know a solution exists. □

This special case of the Mean Value Theorem is so common that it has its own name.

Theorem 5.62 (Rolle's theorem)**.** *Let f be a function and let $[a, b]$ be an interval in the domain of f. Suppose that*

- *f is continuous on $[a, b]$;*
- *f is differentiable on (a, b); and*
- *$f(a) = f(b)$.*

Then there is a number $c \in (a,b)$ such that $f'(c) = 0$.

Notice that this is just the Mean Value Theorem in the case that $m = 0$.

Example 5.63. Suppose an object follows a straight line path, and occupies the same position at two different moments in time — that is, there are times $t_1 < t_2$ such that $s(t_1) = s(t_2)$.

It then follows that there was a time in between, say t_3, such that $s'(t_3) = 0$. But this is just saying the velocity was zero at time t_3, such as when it turned around. □

5.7.2 Proof of Rolle's theorem and advanced applications of the Mean Value Theorem (optional)

Proof of Rolle's theorem. If f is a constant function, then in fact $f'(c) = 0$ for all $c \in (a,b)$, so the theorem is true but boring.

So let's assume f is not a constant function. Then since it is continuous, by the Extreme Value Theorem it attains its absolute maximum and its absolute minimum somewhere on $[a,b]$.

If the absolute maximum is $f(a) = f(b)$, then the absolute minimum has to be somewhere in the middle, at a point $c \in (a,b)$, meaning it is a local minimum. Since f is differentiable, then c is a critical point with $f'(c) = 0$. Done.

Otherwise, the absolute maximum is somewhere in the middle, at a point $c \in (a,b)$, meaning it is a local maximum. Again, this means $f'(c) = 0$.

So no matter what: there is some point $c \in (a,b)$ such that $f'(c) = 0$. □

As a more advanced application of the Mean Value Theorem, we can prove the following result about the function $\sin(x)$.

Proposition 5.64. *For all $a, b \in \mathbb{R}$,*

$$|\sin(a) - \sin(b)| \leq |a - b|.$$

Interpretation: the difference in the y-values of the function $y = \sin(x)$ is always less than or equal to the difference in the x-values; i.e. the slope of any secant line is always less than or equal to 1. That sounds good!

Proof. The Mean Value Theorem applies to $f(x) = \sin(x)$, and says that for every $a < b \in \mathbb{R}$ there is a number $c \in [a,b]$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

or, since $f'(c) = \cos(c)$, we can rewrite this as

$$\frac{\sin(b) - \sin(a)}{b - a} = \cos(c).$$

Taking absolute values of both sides, this gives

$$\left|\frac{\sin(b)-\sin(a)}{b-a}\right| = |\cos(c)| \leq 1.$$

Therefore

$$\frac{|\sin(b)-\sin(a)|}{|b-a|} \leq 1$$

which gives

$$|\sin(b)-\sin(a)| \leq |b-a|.$$

Since $|x-y| = |y-x|$, this is equivalent to the inequality we wanted to prove. □

The Mean Value Theorem can also be used to figure out how far from the correct answer your Taylor approximation can be, that is, to estimate the error $|f(x) - T_n(f)|$. As a simple example, we do this for $n = 0$, which is the constant approximation $T_0(x) = f(a)$.

Proposition 5.65. *If for all $x \in [a, b]$, we have $m \leq f'(x) \leq M$, then*

$$m(x-a) \leq f(x) - f(a) \leq M(x-a)$$

for all $x \in [a, b]$.

Proof. By the Mean Value Theorem, $\frac{f(x)-f(a)}{x-a}$ is equal to $f'(c)$ for some c between a and x; by hypothesis this number satisfies $m \leq f'(c) \leq M$. Since $x - a > 0$ in this case, we can multiply both sides by $x - a$ and the inequalities keep the same direction. Doing this for all x gives the same approximation, so the approximation holds for all $x \in [a, b]$. □

5.8 Stability of Discrete Time Dynamical Systems

Recall: A *DTDS* is an iteration

$$x_{t+1} = f(x_t)$$

over discrete time periods $t = 0, 1, 2, \ldots$. A *fixed point* or *equilibrium* or *steady state* is a value x^* that satisfies $x^* = f(x^*)$. A fixed point is called *stable* if all solutions from nearby initial values converge to x^*, meaning

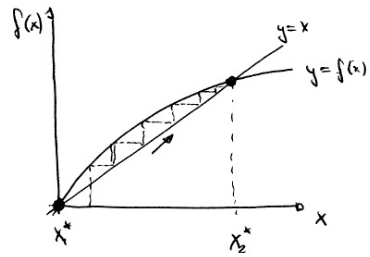
$$\lim_{t\to\infty} x_t = x^* \quad \text{if } x_0 \text{ is near } x^*$$

and *unstable* otherwise.

Our method for distinguishing stable from unstable fixed points was *cobwebbing*. For example, if we consider the DTDS

$$x_{t+1} = f(x_t) = \frac{2x_t}{1+x_t}$$

we see it has two fixed points, $x_1^* = 0$ and $x_2^* = 1$. If we draw the cobweb on some initial value between the two fixed points, we get the following:



which shows that x_1^* is unstable and x_2^* is stable. (Well, we should also check an initial value beyond x_2^* — exercise.)

Goal: Find an analytical way to distinguish stable and unstable fixed points, that is, a method that doesn't rely on graphing and cobwebbing, or on numerical tests.

5.8.1 Stability of linear DTDS (recall)

A linear DTDS is a special kind of DTDS, where the updating function $f(x)$ is a linear function $f(x) = rx + c$. If the slope $r \neq 1$, then the corresponding DTDS has exactly 1 fixed point

$$x^* = rx^* + c \Leftrightarrow (1-r)x^* = c \Leftrightarrow x^* = \frac{c}{1-r}.$$

We had examined the stability of the fixed point in Section 2.7. Our conclusion was:

For a linear DTDS $x_{t+1} = rx_t + c$, with slope $r \neq 1$, the fixed point $x^* = \frac{c}{1-r}$ is stable if the slope satisfies $\lvert r\rvert < 1$ and unstable if the slope satisfies $\lvert r\rvert > 1$.

5.8.2 Stability of general DTDS

We saw in the previous section that we can approximate a function near a point like $a = x^*$ by its linearization. So the stability of the fixed point should be determined by the stability of the fixed point of the corresponding linear DTDS.

Example 5.66. Let's check in our previous example. We have

$$f(x) = \frac{2x}{1+x}$$

so

$$f'(x) = \frac{(1+x)2 - 2x(1)}{(1+x)^2} = \frac{2}{(1+x)^2}$$

and we note that $|f'(0)| = |2| > 1$, which suggests 0 is unstable, whereas $|f'(1)| = |2/4| = \frac{1}{2} < 1$, which suggests 1 is stable. □

Theorem 5.67. *Suppose $x_{t+1} = f(x_t)$ is a DTDS and x^* is a fixed point. Then*

- *if $|f'(x^*)| < 1$, then x^* is a stable fixed point; and*

- *if $|f'(x^*)| > 1$, then x^* is an unstable fixed point.*

If $|f'(x^*)| = 1$, we can't use this test; see further courses on differential equations.

The idea of the proof. So let's write down the linearization of f at $a = x^*$:

$$L(x) = f(x^*) + f'(x^*)(x - x^*) = x^* + f'(x^*)x - f'(x^*)x^* = c + f'(x^*)x$$

where $c = x^*(1 - f'(x^*))$ is a constant. Then, if x_t is close to x^*, we have

$$\begin{aligned} x_{t+1} &= f(x_t) \\ &\simeq L(x_t) \\ &= c + f'(x^*)x_t \end{aligned}$$

which is saying that x_t approximately satisfies the linear DTDS

$$x_{t+1} = c + f'(x^*)x_t$$

whose fixed point is x^* (check!).

So if $|f'(x^*)| < 1$, x_{t+1} will be closer to x^*, and we can repeat the argument, and it follows that $\lim_{t\to\infty} x_t = x^*$, and the fixed point is stable.

But if $|f'(x^*)| > 1$, then x_{t+1} will be further away from x^*, and the approximation will get worse, not better, so $\lim_{t\to\infty} x_t \neq x^*$, and the fixed point is unstable.

□

5.8.3 Example: Allee effect

Consider a population displaying the Allee effect, and described by the DTDS

$$x_{t+1} = f(x_t) = \frac{3x_t^2}{1 + x_t^2}.$$

Find its fixed points and classify them according to their stability.

Solution: First, we find the fixed points. We solve

$$x = f(x) = \frac{3x^2}{1 + x^2} \Leftrightarrow x(1 + x^2) = 3x^2$$

so either $x = 0$ (the usual fixed point we expect) or $1 + x^2 = 3x$ which means

$$x^2 - 3x + 1 = 0 \Leftrightarrow x = \frac{3 \pm \sqrt{9 - 4}}{2} = \frac{3}{2} \pm \sqrt{5}2.$$

We note that $\sqrt{5} < 3$ (or use a calculator) to see that this gives a total of three positive fixed points:

$$x_1 = 0, \quad x_2 = \frac{3}{2} - \frac{1}{2}\sqrt{5} \simeq 0.382, \quad x_3 = \frac{3}{2} + \frac{1}{2}\sqrt{5} \simeq 2.618.$$

Now, we discuss stability of the fixed points. For that, we need to find $f'(x)$:

$$f'(x) = \frac{(1+x^2)(6x) - 3x^2(2x)}{(1+x^2)^2} = \frac{6x}{(1+x^2)^2}.$$

Therefore:

- At $x_1 = 0$, we have $f'(0) = 0$. Since $|f'(0)| < 1$, this is a stable fixed point. (So if our population is too small, it dies out.)
- At x_2, we have

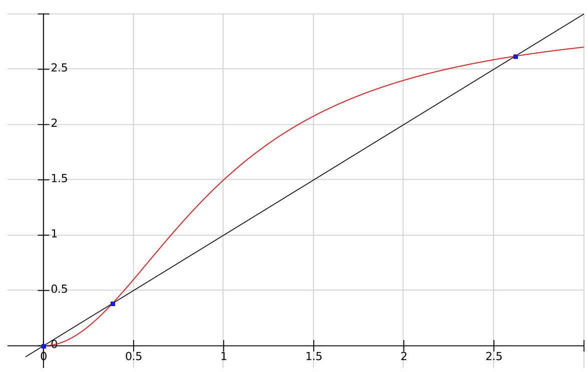
$$\begin{aligned} f'(x_2) &= \frac{6x_2}{(1+x_2^2)^2} \\ &= \frac{6x_2}{(3x_2)^2} \quad \text{since } x_2 \text{ is a root of } 1+x^2 = 3x \text{, see above} \\ &= \frac{6}{3x_2} = \frac{2}{x_2} \simeq 5.236 > 1 \end{aligned}$$

  so that x_2 is an unstable fixed point.
- At x_3, the same argument gives us

$$f'(x_3) = \frac{2}{x_3} \simeq 0.764 < 1$$

  so that x_3 is a stable fixed point.

We can confirm this by cobwebbing on the graph:



5.8.4 Logistic growth

A population growing under logistic growth follows the dynamics of a DTDS of the form

$$x_{t+1} = rx_t(1 - x_t)$$

for some parameter r satisfying $r > 0$. Let's find the fixed points and classify their stability.

The updating function is

$$f(x) = rx(1-x).$$

To find the fixed points, we solve $x = f(x)$, which gives

$$x = rx(1-x)$$

so the fixed points satisfy either $x = 0$ or else

$$1 = r(1-x) \Leftrightarrow 1 = r - rx \Leftrightarrow rx = r - 1 \Leftrightarrow x = \frac{r-1}{r}.$$

This second fixed point is thus positive (and relevant) only if $r > 1$.

To determine stability, we compute the derivative of f; since $f(x) = rx - rx^2$ this is

$$f'(x) = r - 2rx.$$

So the fixed point $x = 0$ is stable only if $0 < r < 1$ and unstable if $r > 1$.

The fixed point $x^* = \frac{r-1}{r}$ (for $r > 1$) gives $f'(x^*) = r - 2(r-1) = 2 - r$. This is stable when $-1 < 2 - r < 1$, or

$$-3 < -r < -1 \Leftrightarrow 1 < r < 3.$$

It is thus unstable for $r > 3$. Putting these together give several different cobwebbing scenarios that we have previously described.

5.8.5 The Ricker equation

The logistic model incorporates a diminishing per capita rate of reproduction $r(1 - x_t)$, reflecting that the rate of reproduction goes down as the size of the population increases. But this model for the reproduction rate is only valid for $x_t < 1$ (or else it becomes negative).

A more sophisticated model would model the per capita reproduction rate with a function like

$$e^{r(1-x_t)}$$

for some positive constant r. This function is always positive and decreasing. This leads to the *Ricker model* which is often used in modeling the population in fisheries:

$$x_{t+1} = x_t e^{r(1-x_t)}.$$

So here, $f(x) = xe^{r(1-x)}$ is the updating function.

Find the fixed points and classify they by their stability.

We need to solve $x = f(x) = xe^{r(1-x)}$; one solution is $x = 0$, the other must satisfy

$$1 = e^{r(1-x)} \Leftrightarrow 0 = r(1-x) \Leftrightarrow x = 1.$$

To discuss stability, we differentiate:

$$f'(x) = e^{r(1-x)} + x(-r)e^{r(1-x)} = (1 - rx)e^{r(1-x)}.$$

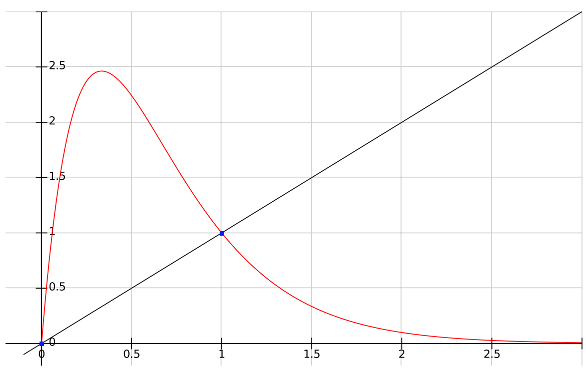
Thus $f'(0) = e^r > 1$ for $r > 0$; this is always unstable and that small populations always grow.

For $x^* = 1$, we compute $f'(1) = (1-r)e^{r(1-1)} = 1 - r$. So this fixed point is stable

$$\Leftrightarrow -1 < 1 - r < 1$$
$$\Leftrightarrow -2 < -r < 0 \quad \text{(subtracted 1 from all terms)}$$
$$\Leftrightarrow 0 < r < 2 \quad \text{(multiplied all terms by } -1\text{, and changed the direction of inequalities).}$$

So this fixed point is stable if $0 < r < 2$ and unstable if $r > 2$.

The graph of $f(x) = xe^{3(1-x)}$ is given below. Try cobwebbing to see what kind of instability we have at $x^* = 1$ in this case. Both fixed points being unstable essentially drives the population into chaos. In terms of what's happening in the fishery: the fish reproduce so much one year that they annihilate the resources in their environment, leading to a population crash the following year. The population swings can seem completely random — the dynamics of this population are in fact an example of the mathematical concept of chaos.



5.8.6 Harvesting and optimization : DTDS

Recall in Section 5.4.5 that we considered a population undergoing logistic growth, but being harvested regularly at a rate h, which led to a DTDS of the form

$$x_{t+1} = 2.5x_t(1 - x_t) - hx_t$$

where $h > 0$ is a parameter which denotes the intensity of harvesting.

We started by assuming that this DTDS had a stable positive steady state x^*, so that we could talk about the yield of the harvest in the long term, which would be $Y(h) = hx^*$.

Now, we can complete the problem by determining if the level of harvest that we chose gives a steady state that is stable.

We begin by finding the steady states by solving $x = f(x) = 2.5x(1-x) - hx$. This gives one solution $x = 0$ and the other satisfies

$$1 = 2.5(1-x) - h \Leftrightarrow 1 = 2.5 - 2.5x - h \Leftrightarrow 2.5x = 1.5 - h \Leftrightarrow x = \frac{1.5 - h}{2.5}.$$

This is positive if $1.5 - h > 0$ or $h < 1.5$. Since h is a rate of harvesting, $h \geq 0$. So the region of interest is

$$0 \leq h < 1.5.$$

Next, we discuss stablity. We calculate $f(x) = 2.5x - 2.5x^2 - hx$ and differentiate with respect to x to get

$$f'(x) = 2.5 - 5x - h.$$

When $x^* = \frac{1.5-h}{2.5}$ this gives

$$f'(x^*) = 2.5 - 5\left(\frac{1.5-h}{2.5}\right) - h = 2.5 - 2(1.5-h) - h = 2.5 - 3 + 2h - h = -0.5 + h$$

so the steady state arising from harvesting is stable only if

$$-1 < -0.5 + h < 1$$

or (adding 0.5 from all three terms):

$$-\frac{1}{2} < h < \frac{3}{2}.$$

We conclude that, among the harvesting rates that give positive steady states ($0 \leq h < \frac{3}{2}$), all give stable steady states.

(In particular, the harvesting rate we found, of $h = 0.75$, falls in the good range.)

Exercise 5.68. In the above example, the range of values of h for which the nonzero steady state was stable was wider than the range of values of h for which the nonzero steady state was positive. Find the range of $h \geq 0$ that make the nonzero steady state x^* of the following DTDS (a) positive (b) stable:

$$x_{t+1} = 2x_t(2 - x_t) - hx_t$$

and notice that this time it is possible to choose h to give a nonstable positive steady state. This would be a very problematic rate of harvesting, as it would give a different yield each year, in some kind of chaotic dynamics.

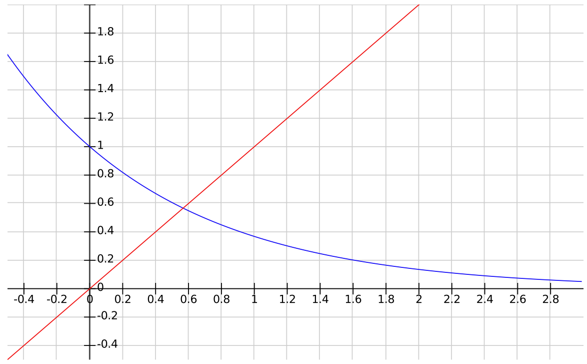
End of lecture # 16

5.9 Solving equations: the Intermediate Value Theorem and Newton's Method

The final application of differentiation we'll consider is about solving equations. We know a great many algebraic techniques, but occasionally come across equations that are impossible to solve algebraically.

Motivation: Find the fixed points of the DTDS $x_{t+1} = e^{-x_t}$.

Solution: We have to solve $x = e^{-x}$. But applying ln gives $\ln(x) = -x$, which is just as difficult. We look at the graph:



The graphs of $y = e^{-x}$ (in blue) and $y = x$ (in red). They have a unique point of intersection near $x = 0.55$.

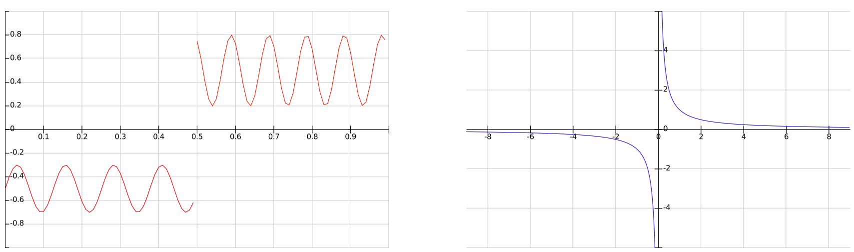
We agree that there is a solution. Actually, we can do better than that.

Strategy: Form the function $f(x) = x - e^{-x}$. Then

$$f(0) = 0 - e^0 = -1 < 0 \quad \text{and} \quad f(1) = 1 - e^{-1} > 0,$$

so there should be a number c, with $0 < c < 1$, such that $f(c) = 0$, right?

Caution: Imagine our function f had a graph like one of the following.

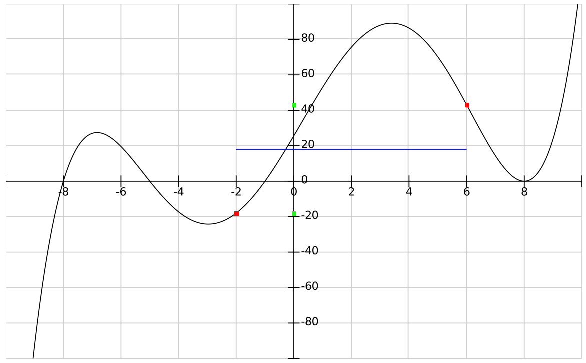


On the left, the graph of a discontinuous function $y = f(x)$ which satisfies that $f(0) < 0$ and $f(1) > 0$ but there is no number c between 0 and 1 such that $f(c) = 0$. On the right, the graph of $f(x) = 1/x$, which satisfies $f(-1) < 0$ and $f(1) > 0$ but there is no x for which $f(x) = 0$.

It is important that the function be continuous on your interval for this strategy to work!

Theorem 5.69 (Intermediate Value Theorem)**.** *Suppose that f is a function which is continuous on the interval $[a, b]$, and that y is a number between $f(a)$ and $f(b)$. Then there is a number $c \in [a, b]$ such that $f(c) = y$.*

The idea of the theorem is: if your function is continuous on $[a, b]$, then if you draw the curve from $(a, f(a))$ to $(b, f(b))$, you have to cross every horizontal line (every y-value) between $f(a)$ and $f(b)$. You can't avoid solving $f(c) = y$.



The graphs of a continuous function $f(x)$ and two marked points in red $(-2, -18)$ and $(6, 43.12)$, whose y-values are marked on the y-axis in green. The Intermediate Value Theorem says: every horizontal line between $y = f(a)$ and $y = f(b)$ has to cross the graph at least once, at an x-value lying between a and b.

Remark 5.70. This is a different application of an idea we have used before in a completely different context: if we have an interval (a, b) and f has no critical points in that interval (meaning $f'(x)$ is never zero or undefined) then necessarily f' is either always positive or always negative on that interval. In other words: if f' changes sign on an interval, then that interval must contain a critical point of f.

5.9.1 Classic: Bisection Method

So let's apply this to our example, to solve $x = e^{-x}$.

The function $f(x) = x - e^{-x}$ is defined and continuous on all of $\mathbb{R}$, so in particular on any subinterval. **Tip: in general, watch out for asymptotes!!**

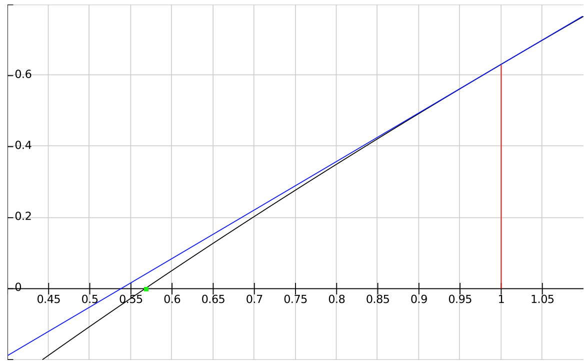
- We have $f(0) < 0$ and $f(1) > 0$ so there is a root of f (that is, a value x such that $f(x) = 0$) between 0 and 1.
- We compute $f(0.5) = -0.11 < 0$. So there is a root of f between 0.5 and 1.
- We compute $f(0.75) = 0.28 > 0$. So there is a root of f between 0.5 and 0.75.
- We compute $f(0.625) = 0.09 > 0$. So there is a root of f between 0.5 and 0.625.
- We compute $f(0.57) = 0.004$, which is equal to zero to two decimal places, which we think is good enough.

This method is a good classic, but it is slow.

5.9.2 Sophisticated: Newton's Method

(So named because he used it very effectively.)

Let's use information about the derivative and the tangent line to make better guesses at a root of $f(x)$. The idea is in the following picture:



The graph of $f(x)$ in black. An initial guess $x_0 = 1$ gives a point on the curve $y = f(x)$. We draw the tangent line to the curve and see where it intersects the x-axis.

1. Make an initial guess x_0, so that $f(x_0) \sim 0$.
2. Write down the equation of the tangent line of f at $(x_0, f(x_0))$.
3. Let x_1 be where this tangent line intersects the x-axis.
4. Take x_1 as your next guess, and repeat until $f(x_n)$ is as close to zero as you need.

Actually, if we just work this out in general, we'll get a simple formula.

So given $f(x)$, our goal is to solve $f(x) = 0$. We assume we have used the Intermediate Value Theorem to help us make an initial guess x_0.

The equation of the tangent line to f at x_0 is

$$L(x) = f(x_0) + f'(x_0)(x - x_0).$$

Solve for x such that $L(x) = 0$:

$$0 = f(x_0) + f'(x_0)x - f'(x_0)x_0 \Leftrightarrow f'(x_0)x = f'(x_0)x_0 - f(x_0) \Leftrightarrow x = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Therefore:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}, \quad \ldots$$

that is, we have in effect created a DTDS which ought to converge to the root we were looking for!

Newton's method: To solve $f(x) = 0$ with initial guess x_0, use the iterative formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad i = 0, 1, 2, 3, \ldots$$

until your numbers agree to the accuracy you need.

Memorize this formula correctly, or else remember how to derive it. Wrong formula = garbage.

Example 5.71. Solve $x = e^{-x}$ with an accuracy of three decimal places using Newton's method.

Solution: We first have to convert the problem into a root-finding problem. Let

$$f(x) = x - e^{-x}, \quad f'(x) = 1 + e^{-x}.$$

So our Newton's method formula is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n - e^{-x_n}}{1 + e^{-x_n}}.$$

Start with $x_0 = 0.57$, which is the best guess we got with our bisection method. Then

$$x_1 = 0.57 - \frac{0.57 - e^{-0.57}}{1 + e^{-0.57}} = 0.56714181501$$

$$x_2 = x_1 - \frac{x_1 - e^{-x_1}}{1 + e^{-x_1}} = \mathbf{0.567143}2904$$

and we stop, because our solution is already accurate to 5 decimal places! □

Example 5.72. Find $\sqrt{2}$ to 3 decimal places using Newton's method.

Solution: We need to create a nice function with $\sqrt{2}$ as a root; $f(x) = x^2 - 2$ is a good choice. Then we need a first guess; $x_0 = 1$ is close. Our formula is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - 2}{2x_n}$$

so that gives

$$x_1 = 1 - \frac{1-2}{2} = 1.5$$

$$x_2 = 1.5 - \frac{2.25 - 2}{3} = 1.4166667$$

$$x_3 = 1.41421568633$$

$$x_4 = \mathbf{1.41421}356237$$

which is 4 decimals; in fact $x_5 = 1.41421356237 = x_4$ is already the maximum precision my calculator can do. □

5.9.3 Discussion: So why does it work? Can it fail?

The iterative formula for Newton's method is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

So if $f'(x_n)$ is not near zero, but $f(x_n)$ is near zero, then we are correcting our guess by the small amount each time, and by our graph, we see we are getting better and better.

If $f'(x_n)$ is near zero — that is, if we have a critical point near our guess — then Newton's method can just spike off and give ridiculous numbers. You can tell when that happens: x_n is nowhere near x_{n+1}, and your value of $f'(x_n)$ was suspiciously small.

In particular, if x is a root of both f and f' (like a double root of a polynomial) then you shouldn't do Newton's method to f — do it to f' instead!

Another possible failure: if you apply Newton's method to a function with many roots, it could happen that it converges to a different root than the one you wanted. The solution is to make as good a first guess x_0 as possible, so that your function more or less looks like a straight line from the root to your guess (no local maxima or local minima in between).

So: Newton's method can fail, but only for reasons that you should have noticed when you started.

Finally: using a computer, it's very fun and easy to get maximum precision. Use a spreadsheet, for example, and just cut and paste the formula from one line to the next. Therefore, as long as you can differentiate the function, you never have to worry about solving an equation again.

(Newton himself was pretty happy about that.)

End of lecture # 17

Chapter 6

Integration

We have explored lots of applications of the derivative; now it's time to turn the story around and consider anti-derivatives. We begin with some motivation, and setting the stage for why we really want to do this.

6.1 Introduction

6.1.1 Motivation for Differential Equations

Utility companies measure the *rate* of flow of water (or gas, or electricity) into your home in litres/second (respectively, m^3/second, Watts = joules/second). In the end, though, they bill you for the total amount consumed in litres (respectively, m^3, kWh). In other words: they measure the *instantaneous* rate of change and use that to compute the total amount used. This is the *inverse process* of differentiation.

Examples abound:

$f(x)$		$f'(x)$
value	differentiation	rate of change
position	$\longrightarrow$	velocity
mass		growth rate
volume	$\longleftarrow$	flow rate
amount	anti-differentiation (new)	production rate

Let's work through some examples in detail, working to understand the big picture, and then we'll start to focus on techniques.

First kind of example

Suppose that the volume of a cell increases by $2\mu m^3$ per second. What is the volume of the cell after 3 seconds, if the cell starts at a volume of $1\mu m^3$?

Solution: Denote $V(t) =$ volume of the cell. Then we are given that $V'(t) = 2\mu m^3/s$.

Since the derivative is a constant, the function must be linear, with slope 2, which gives:

$$V(t) = 2\frac{\mu m^3}{s} \cdot t + V(0),$$

and this initial value is $V(0) = 1\mu m^3$. So our solution is

$$V(t) = (2t + 1)\mu m^3.$$

Notice that:

- The units work out! Rate in $\mu m^3/s$ times time in s gives μm^3.
- Different initial conditions give different volumes — of course! Without the initial condition, we couldn't tell you $V(t)$ just from the rates of change.

Our answer: after 3 seconds, the volume is $V(3) = 7\mu m^3$.

Second kind of example

The amount of radioactive material decreases by 1% per day. If there are $10g$ initially, when will there be $5g$?

Solution: Denote $x(t) =$ the amound in grams of radioactive material. Then we are told that

$$x'(t) = -0.01x(t).$$

So the derivative of $x(t)$ is a multiple of $x(t)$. We know an example of this: the exponential function!

If $x(t) = e^{\alpha t}$ then $x'(t) = \alpha e^{\alpha t} = \alpha x(t)$; so in our case $\alpha = -0.01$. This can't be quite right: $x(0) = 1$, no matter how we choose α.

Aha: if $x(t) = Ke^{\alpha t}$ for some constant K then $x'(t) = \alpha K e^{\alpha t} = \alpha x(t)$ also, and in this case, $x(0) = K$ is the initial condition.

So our specific solution is $x(t) = 10g \cdot e^{-0.01t}$. What are the units? Since α is $\%/day$, its units are $1/day$, so αt is dimensionless (a ratio), leaving $x(t)$ in grams. Good.

When will $x(t) = 5g$? Solve

$$\begin{aligned} 10e^{-0.01t} &= 5 \\ \Rightarrow e^{-0.01t} &= \frac{1}{2} \\ \Rightarrow -0.01t &= \ln(0.5) \\ \Rightarrow t &= -100\ln(0.5) = 69.3\ days. \end{aligned}$$

What's the difference?

We saw two kinds of examples:

- $V'(t) = 2$ (independent of V)
- $x'(t) = -0.01x(t)$ (dependent on x)

We call the first kind a *pure-time differential equation*. The right hand side may depend on t (the independent variable) but not on V (the dependent variable that we are looking for). (MAT1330)

Examples:
$x'(t) = e^t$, $x(0) = 1$
$y'(t) = \frac{1}{t}$, $y(1) = 0$

The second kind is called an *autonomous differential equation*. The rate of change depends on x (the dependent variable) but not on t (the independent variable). (MAT1332)

Examples:
$w'(t) = 2w(t)$, $w(0) = 1$
$z'(t) = \frac{1}{z(t)}$, $z(0) = 1$

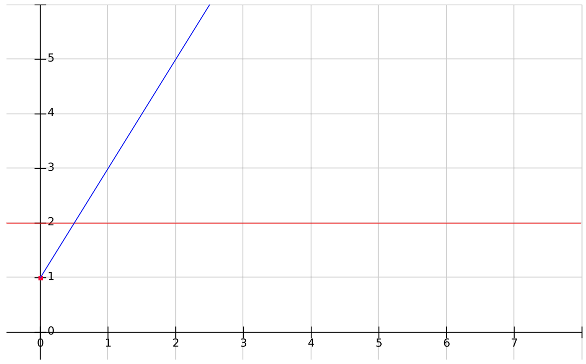
(We can also talk about differential equations that depend on both x and t — but that's for second year courses on differential equations!)

In MAT1330, we are concerned with *pure-time* differential equations.

Graphical solutions

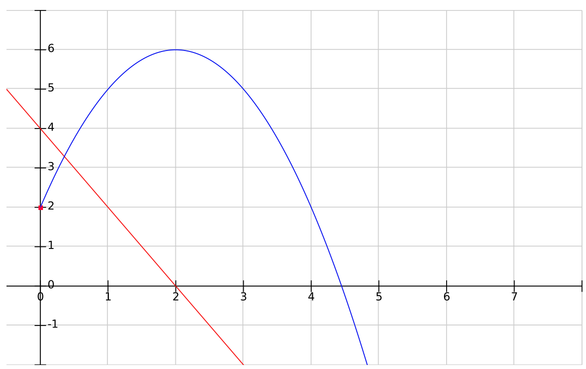
Before proceeding to analytic solutions, let's work with some graphs to see how it's possible to go backwards from the derivative to the function, given an initial condition.

Example 6.1. Suppose $\frac{dV}{dt} = 2$ and $V(0) = 1$. In the graph below, we have sketched the graph of the derivative (the red horizontal line), together with the initial value $(0, 1)$ (red dot). Starting at the red dot, draw a curve (in blue) with slope equal to $V'(t)$ at each point t (which in this case, means a line with constant slope 2); this is the graph of $V(t)$.



□

Example 6.2. Suppose $\dfrac{dV}{dt} = 4 - 2t$ and $V(0) = 2$. In the graph below, we have sketched the graph of the derivative (the red line), together with the initial value $(0, 2)$ (red dot). Starting at the red dot, draw a curve (in blue) with slope equal to $V'(t)$ at each point t (so: initially steep, with slope 2, but this slope decreases as t increases); this is the graph of $V(t)$.



□

6.1.2 Anti-differentiation

Let's begin by being clear about what we are looking for: we are given a function $f(t)$ and want to find a new function $F(t)$ such that $F'(t) = f(t)$.

Definition 6.3. An *antiderivative* of a function $f(t)$ is a function $F(t)$ with the property that $\frac{dF}{dt} = f(t)$. We write

$$F(t) = \int f(t)\, dt$$

which we read out loud as "the *integral* of f of t dt." The function f is called the *integrand*.
If F is one anti-derivative of f, then so is $F + c$, for any constant c. So we write

$$\int f(t)\, dt = F(t) + c$$

and say that $\int f(t)\, dt$ is the *indefinite integral of* f.

So an antiderivative is one function whose derivative is f; the indefinite integral is the set of *all* functions whose derivative is f.

Example 6.4. Suppose $f(t) = 1$. Then $F(t) = t$ as an antiderivative; and $F(t) = t + 5$ is another antiderivative. The indefinite integral is

$$\int f(t)\,dt = \int 1\,dt = t + c \quad \text{with } c \in \mathbb{R}.$$

□

We can turn our differentiation rules into rules for anti-derivatives. Let's start by making a list of anti-derivatives of common functions.

The power rule

Recall that

$$\frac{d}{dt} t^{n+1} = (n+1)t^n;$$

therefore,

$$\int t^n\,dt = \frac{1}{n+1} t^{n+1} + c, \quad \text{if } n \neq -1.$$

Examples:

$$\int t^3\,dt = \frac{1}{4} t^4 + c$$

$$\int x^5\,dx = \frac{1}{6} x^6 + c$$

$$\int \sqrt{x}\,dx = \int x^{1/2}\,dx = \frac{1}{\frac{1}{2}+1} x^{\frac{1}{2}+1} + c = \frac{2}{3} x^{3/2} + c$$

$$\int \frac{1}{t^2}\,dt = \int t^{-2}\,dt = \frac{1}{-2+1} t^{-2+1} + c = -t^{-1} + c.$$

Constant multiple rule and sum rule

If a is a constant, then $\frac{d}{dx}(af(x)) = a\frac{d}{dx}(f(x))$. Therefore,

$$\int af(x)\,dx = a \int f(x)\,dx \quad \text{for any constant } a.$$

We also have that $(f+g)' = f' + g'$, so

$$\int f(x) + g(x)\,dx = \int f(x)\,dx + \int g(x)\,dx.$$

Examples:

$$\int (3x^2 - 7x)\,dx = 3\int x^2\,dx - 7\int x\,dx = 3\frac{1}{3}x^3 - 7\frac{1}{2}x^2 + c = x^3 - \frac{7}{2}x^2 + c.$$

$$\int \frac{3}{\sqrt{x}} + \frac{4}{x^3}\,dx = 3\int x^{-1/2}\,dx + 4\int x^{-3}\,dx = 3\frac{1}{1/2}x^{1/2} + 4\frac{1}{-2}x^{-2} + c = 6\sqrt{x} - \frac{2}{x^2} + c.$$

Special functions

- Since $(e^x)' = e^x$, we have $\displaystyle\int e^t \, dt = e^t + c$;
- Since $(\sin(x))' = \cos(x)$, we have $\displaystyle\int \cos(x) \, dx = \sin(x) + c$
- Since $(\cos(x))' = -\sin(x)$, we have $\displaystyle\int \sin(x) \, dx = -\cos(x) + c$

We also know that $(\ln(x))' = 1/x$ — but this one is annoying. The domain of $\ln(x)$ is $(0, \infty)$ whereas the domain of $1/x$ is all real numbers except 0. But there turns out to be an easy solution.

Example 6.5. The function $f(x) = \ln|x|$ satisfies

$$f(x) = \ln|x| = \begin{cases} \ln(x) & \text{if } x > 0 \\ \ln(-x) & \text{if } x < 0 \end{cases}$$

so if we differentiate, we get

$$f'(x) = \begin{cases} \frac{1}{x} & \text{if } x > 0 \\ \frac{1}{-x}(-1) = \frac{1}{x} & \text{if } x < 0 \end{cases}$$

which is just perfect! □

- So $(\ln|x|)' = \frac{1}{x}$ gives us $\displaystyle\int \frac{1}{x} \, dx = \ln|x| + c$[1]
- Since $(\arctan(x))' = \frac{1}{1+x^2}$, we have $\displaystyle\int \frac{1}{1+x^2} \, dx = \arctan(x) + c$.
- Since $(\arcsin(x))' = \frac{1}{\sqrt{1-x^2}}$, we have $\displaystyle\int \frac{1}{\sqrt{1-x^2}} \, dx = \arcsin(x) + c$.

End of lecture # 18

6.1.3 Applications of Anti-differentiation

Let's explore some examples where anti-differentiation arises naturally.

Example 6.6. In 1981, there were about 340 cses of HIV infection in the USA. In the years following, the number of cases grew by around $500t^2$/year, where $t = 0$ corresponds to 1981 and t is measured in years. How many cases were there in 1991?

Solution: Let $A(t)$ = number of cases, with $A(0) = 340$. We are told that $A'(t) = 500t^2$.

[1] In fact, you can choose a different constant on each half of the domain of $\ln|x|$, and still have the same derivative, but we write $+c$ for short.

Therefore,

$$A(t) = \int A'(t)\ dt = \int 500t^2\ dt = \frac{500}{3}t^3 + c$$

for some constant c. Since $A(0) = 340$, plugging in $t = 0$ gives $c = 340$. Therefore $A(t) = \frac{500}{3}t^3 + 340$.

We want to know the number of cases in 1991, which corresponds to $t = 10$. So plug in:

$$A(10) = \frac{500}{3}(10)^3 + 340 \sim 167,000.$$

□

Example 6.7. A bucket falls from a window cleaner's platform, and experiences constant acceleration due to gravity of $a = -9.8m/s^2$.

Recall: the rate of change of position is velocity, and the rate of change of velocity is acceleration.

1. Suppose the platform is 49m up and the initial speed of the bucket is 0.

(a) Find the equation for the position $p(t)$ of the bucket.

(b) Where is the bucket after 1 second?

(c) When will it hit the ground?

(d) At what speed will it hit the ground?

Solution:

(a) Write $v(t)$ for the velocity of the bucket. So

$$v(t) = \int a\ dt = -9.8t + c$$

and since $v(0) = 0$, we conclude $c = 0$. Next, we have

$$p(t) = \int v(t)\ dt = \int -9.8t\ dt = \frac{-9.8}{2}t^2 + c'$$

and since $p(0) = 49$, we conclude that $c' = 49$. Therefore an equation for the position of the bucket is

$$p(t) = -4.9t^2 + 49.$$

Check the units: $9.8m/s^2$ times t^2 seconds^2 gives m, good.

(b) After 1 second, the position will be $p(1) = -4.9 + 49 = 44.1$ m above the ground.

(c) It will hit the ground when its position is 0. So we solve $p(t) = 0$ for t, which gives

$$0 = -4.9t^2 + 49 \Leftrightarrow 4.9t^2 = 49 \Leftrightarrow t^2 = 10 \Leftrightarrow t = \pm\sqrt{10}s;$$

since we are going forward in time, the positive solution is our answer, giving $t \sim 3.2s$.

(d) Since it hits the ground after $\sqrt{10}$ seconds, its velocity at that time is

$$v(\sqrt{10}) = -9.8(\sqrt{10}) \sim -31m/s$$

(negative because downwards), which is about 112 km/h.

2. Suppose now that the window cleaner tosses the bucket straight upwards towards another platform, but it misses and falls to the ground. If the bucket is thrown with an initial velocity of $10m/s$,

(a) When will it reach its highest point?

(b) How high will this be?

(c) When will it hit the ground?

(d) How fast will it be going?

Solution: Again, the only force exerted on the bucket is gravity, so we have

$$v(t) = \int a(t)\, dt = \int -9.8\, dt = -9.8t + c.$$

Since $v(0) = 10$, we have $c = 10$, so $v(t) - 9.8t + 10$ m/s. Next, we solve

$$p(t) = \int v(t)\, dt = \int -9.8t + 10\, dt = \frac{-9.8}{2}t^2 + 10t + c',$$

and since $p(0) = 49$, we have $c' = 49$. Therefore the equation of motion for the bucket is

$$p(t) = -4.9t^2 + 10t + 49.$$

(a) The highest point is attained where the bucket stops for a moment, that is, when $v(t) = 0$. We solve

$$v(t) = 0 \Leftrightarrow -9.8t + 10 = 0 \Leftrightarrow t = \frac{10}{9.8} \simeq 1.02s.$$

(b) Its position at this time is $p(10/9.8) = -4.9(10/9.8)^2 + 10(10/9.8) + 49 \sim 54.1m$, which is about 5 m above the platform.

(c) It hits the ground when its position is equal to 0:

$$-4.9t^2 + 10t + 49 = 0 \Leftrightarrow t = \frac{-10 \pm \sqrt{100 + 960.4}}{-9.8} = 1.02 \pm 3.32$$

and we choose the positive root, which yields $t \simeq 4.34$s.

(d) Its speed when it hits the ground will be $v(4.34) \simeq -32.6$ m/s, which is about 117 km/h.

□

6.2 Techniques of integration: Substitution

So far we can only calculate indefinite integrals when the integrand is a function whose anti-derivative we already know. This is a fairly small list (see Table 7.2.1 in the textbook, for example).

Today we'll learn and practice a method which is based on undoing the chain rule; we call it the method of substitution.

Recall that the chain rule tells us that

$$\int f'(g(x))g'(x)dx = f(g(x)) + c.$$

Example 6.8. Consider

$$\int 2xe^{x^2} dx.$$

Try $g(x) = x^2$, then $g'(x) = 2x$, so indeed:

$$\int e^{x^2}\ 2x\ dx = \int e^{g(x)}g'(x)\ dx = e^{g(x)} + c = e^{x^2} + c.$$

□

But how would we recognize that our integrand has this special form? This is where the special notation of integrals comes in handy.

6.2.1 The method of substitution:

1. Define a new variable $u = g(x)$ (typically the innermost function).
2. Differentiate and write:

   $$\frac{du}{dx} = g'(x) \Rightarrow du = g'(x)\ dx.$$

   This is *just notation* — a clever way to keep track of things.
3. Write the entire integral in terms of u and du:

   $$\boxed{\int f'(g(x))g'(x)\ dx = \int f'(u)\ du}$$

   (typically start by replacing $g'(x)dx$ with du and then transforming every other occurrence of x into a u). Important: the resulting integral CANNOT have a mix of x and u: it must all be in one variable.
4. Then integrate:

   $$\int f'(u)\ du = f(u) + c$$

5. Then substitute back $u = g(x)$:

   $$= f(g(x)) + c.$$

6. Check your math by differentiating!

Let's do some examples.

Example 6.9. Back to $\int 2xe^{x^2}\ dx$: With the substitution $u = x^2$ we get $du = 2x\ dx$, we get

$$\int 2xe^{x^2}\ dx = \int e^{\boxed{x^2}} \boxed{2x\ dx} = \int e^{\boxed{\text{u}}} \boxed{\text{du}} = \int e^u\ du = e^u + c = e^{x^2} + c.$$

We check by differentiating: yes! $\square$

Example 6.10. Consider

$$\int e^{3x}\ dx.$$

We try $u = 3x$, which gives $du = 3\ dx$. We don't have a 3 in the integral, but it's just a constant, so we can write $dx = \frac{1}{3}\ du$. That gives:

$$\int e^{3x}\ dx = \int e^u\ \frac{1}{3}\ du = \frac{1}{3}\int e^u\ du = \frac{1}{3}e^u + c = \frac{1}{3}e^{3x} + c.$$

$\square$

Example 6.11. Consider

$$\int \frac{1}{1+5x}\ dx$$

We try the substitution $u = 1 + 5x$ which gives $du = 5dx$ or $dx = \frac{1}{5}\ du$ so:

$$\int \frac{1}{1+5x}\ dx = \int \frac{1}{u}(\frac{1}{5})\ du = \frac{1}{5}\int \frac{1}{u}\ du = \frac{1}{5}\ln|u| + c = \frac{1}{5}\ln|1+5x| + c.$$

$\square$

Example 6.12. Given $\int \cos(2\pi(x-1))\ dx$, we try the substitution $u = 2\pi(x-1)$ which gives $du = 2\pi dx$ or $dx = \frac{1}{2\pi}du$ so

$$\int \cos(2\pi(x-1))\ dx = \int \cos(u)\frac{1}{2\pi}du = \frac{1}{2\pi}\sin(u) + c = \frac{1}{2\pi}\sin(2\pi(x-1)) + c.$$

$\square$

Notice that we converted the integral to one of the variable u, and then at the end reverted back to x.

Example 6.13. Find

$$\int \frac{\sec^2(1/x)}{x^2}dx.$$

Again, the trickiest part is $\sec^2(1/x)$ and we know that $\sec^2(u)$ has a nice anti-derivative; so we try

$$u = 1/x \Rightarrow du = -\frac{1}{x^2}dx \quad \Rightarrow \frac{1}{x^2}dx = -du$$

which gives

$$\begin{aligned}\int \frac{\sec^2(1/x)}{x^2}dx &= \int \sec^2(1/x)\frac{1}{x^2}\,dx \\ &= \int \sec^2(u)(-1)du \\ &= -\int \sec^2(u)du \\ &= -\tan(u)+c \\ &= -\tan(1/x)+c.\end{aligned}$$

Check by differentiating: yes! □

6.2.2 Other situations where you might try substitution

Example 6.14. Find

$$\int \frac{e^{-3t}}{e^{-3t}+1}\,dt.$$

We can do the substitution $u=-3t$, but it doesn't get us to a substantially better integral (try it!) There isn't an obvious composition of functions in this case. But the thing which is making this integral difficult is that there is a sum in the denominator, and we notice that the derivative of $u=e^{-3t}+1$ is just $du=-3e^{-3t}dt$, which (up to a constant) is right there in the numerator.

So we get

$$\int \frac{1}{e^{-3t}+1}e^{-3t}\,dt = \int \frac{1}{u}\,(-\tfrac{1}{3})du = -\frac{1}{3}\ln|u|+c = \ln|e^{-3t}+1|+c = \ln(e^{-3t}+1)+c$$

where in the last step we noticed that $e^{-3t}+1>0$ for all t so the absolute value sign was superfluous. □

Substitution is a great thing to try whenever you see both a function and its derivative in the integrand.

Example 6.15. Find

$$\int \frac{\arctan(x)}{1+x^2}\,dx.$$

This time, there is no composition of functions, so it's not obvious that substitution is the thing to do. But we notice that the integrand is a product of a function $u=\arctan(x)$ and its derivative $du=\frac{1}{1+x^2}dx$, so we will just go ahead and make the substitution and see what happens:

$$\int \frac{\arctan(x)}{1+x^2}\,dx = \int u\,du = \frac{1}{2}u^2+c = \frac{1}{2}(\arctan(x))^2+c.$$

We check by differentiation.

So what happened here was: we couldn't see the term $f'(g(x))$ that normally clues us in because in this case $f(u)=\frac{1}{2}u^2$ and so $f'(u)=u$. □

Be willing to try a substitution, to see if it can work out.

Example 6.16. Find
$$\int \frac{\sin(\ln(3x))}{x} dx.$$
The integrand is a composition of functions (in fact, of three functions). We could try $u = 3x$, or $u = \ln(3x)$. Let's choose this second one[2]
$$u = \ln(3x) \quad \Rightarrow \quad du = \frac{1}{3x} 3dx = \frac{1}{x} dx$$
which is wonderful, since we have $\frac{1}{x}dx$ in our integrand. So we go ahead and make the substitution:
$$\int \frac{\sin(\ln(3x))}{x} dx = \int \sin(u) du = -\cos(u) + c = -\cos(\ln(3x)) + c.$$
We check that this is correct by differentation. □

6.2.3 Trying a substitution in the hopes of simplifying a complicated integrand

Sometimes, you try a substitution without actually realizing what $f(u)$ is going to be.

Example 6.17. Find
$$\int \frac{x^3}{\sqrt{x^2+4}} dx.$$
We see that the toughest part is $\sqrt{x^2+4}$. We have a few choices (namely $u = x^2$ or $u = x^2 + 4$); let's try $u = x^2 + 4$. So
$$u = x^2 + 4, \quad \Rightarrow \quad du = 2x\ dx \quad \Rightarrow \quad x\ dx = \frac{1}{2} du$$
So now we try to rewrite our integral in terms of u:
$$\begin{aligned}
\int \frac{x^3}{\sqrt{x^2+4}} dx &= \int \frac{1}{\sqrt{x^2+4}} x^2 \cdot x\ dx \\
&\neq \int \frac{1}{\sqrt{u}} (x^2) \frac{1}{2}\ du \quad \text{INCOMPLETE! you can't mix } x \text{ and } u \\
&= \int \frac{1}{\sqrt{u}} (u-4) \frac{1}{2}\ du \quad \text{SUCCESS! we used } x^2 = u - 4 \\
&= \frac{1}{2} \int \frac{u-4}{u^{1/2}} du \\
&= \frac{1}{2} \int (u^{1/2} - 4u^{-1/2})\ du \\
&= \frac{1}{2} (\frac{1}{3/2} u^{3/2} - \frac{4}{1/2} u^{1/2}) + c \\
&= \frac{1}{2} (\frac{2}{3} (x^2+4)^{3/2} - 8(x^2+4)^{1/2}) + c \\
&= \frac{1}{3} (x^2+4)^{3/2} - 4\sqrt{x^2+4} + c
\end{aligned}$$
and again **we check by differentiating** — this time, even the check is a lot more work! □

[2] although you could also have tried $u = 3x$ first, which works fine but yields $\frac{1}{3} \int \sin(ln(u))/u du$, and you have to do a second substitution $v = \ln(u)$. So it all works out just fine, it just takes a little longer.

Lesson: sometimes you just try a substitution to see if it will work out.

Example 6.18. Find

$$\int \frac{1}{x^{1/3}+1} dx$$

Well, this one seems hopeless but the nasty part is $x^{1/3}$ so let's just try

$$u = x^{1/3} + 1 \quad \Rightarrow \quad du = \frac{1}{3}x^{-2/3}\, dx.$$

We certainly don't have $\frac{1}{3}x^{-2/3}\ dx$ in our integrand; but since $u = x^{1/3} + 1$, it follows that $x^{1/3} = (u-1)$ so $x^{-2/3} = (u-1)^{-2}$. Therefore we can rewrite :

$$du = \frac{1}{3}x^{-2/3}\, dx \Rightarrow du = \frac{1}{3}(u-1)^{-2}dx \Rightarrow 3(u-1)^2 du = dx$$

which means we can in fact perform the substitution:

$$\begin{aligned}
\int \frac{1}{x^{1/3}+1} dx &= \int \frac{1}{u}\ 3(u-1)^2\ du \\
&= 3\int \frac{u^2-2u+1}{u}\ du \\
&= 3\int u - 2 + \frac{1}{u}\ du \\
&= 3(\frac{1}{2}u^2 - 2u + \ln|u|) + c \\
&= \frac{3}{2}(x^{1/3}+1)^2 - 6(x^{1/3}+1) + 3\ln|x^{1/3}+1| + c \\
&= \frac{3}{2}x^{2/3} - 3x^{1/3} + 3\ln|x^{1/3}+1| + c'
\end{aligned}$$

(where we have gathered all the constants into a new c'). We check our answer is correct, by differentiating it:

$$\begin{aligned}
&\frac{d}{dx}\left(\frac{3}{2}x^{2/3} - 3x^{1/3} + 3\ln|x^{1/3}+1| + c'\right) = \\
&= x^{-1/3} - x^{-2/3} + \frac{3}{x^{1/3}+1}(\frac{1}{3}x^{-2/3}) \\
&= \frac{1}{x^{1/3}+1}\left(x^{-1/3}(x^{1/3}+1) - x^{-2/3}(x^{1/3}+1) + x^{-2/3}\right) \\
&= \frac{1}{x^{1/3}+1}\left(1 + x^{-1/3} - x^{-1/3} - x^{-2/3} + x^{-2/3}\right) \\
&= \frac{1}{x^{1/3}+1}
\end{aligned}$$

as required, as if by a minor miracle. □

Sometimes, it pays to be persistent and creative with substitutions.

Example 6.19. Find

$$\int \frac{\sin(x)}{1+\cos^2(x)}dx.$$

In this one, the danger is being too greedy with your substitution. If you try $u = 1+\cos^2(x)$ then $du = -2\cos(x)\sin(x)dx$ which is a disaster.

However, if we just try

$$u = \cos(x) \quad \Rightarrow \quad du = -\sin(x)dx$$

then things go very nicely:

$$\begin{aligned}\int \frac{\sin(x)}{1+\cos^2(x)}dx &= \int \frac{1}{1+u^2}(-1)du \\ &= -\int \frac{1}{1+u^2}du \\ &= -\arctan(u)+c \\ &= -\arctan(\cos(x))+c\end{aligned}$$

which we check by differentiation. □

Example 6.20.

$$\int \frac{\tan(x)}{\ln(\cos(x))}\,dx$$

We do not see what this will be, but there is a composition of ln with $\cos(x)$ as the innermost function, and $\tan(x)$ is related to $\cos(x)$, so we give it a shot.

$$u = \cos(x) \quad \Rightarrow \quad du = -\sin(x)\,dx$$

and although we do not see $\sin(x)$, we realize we can rewrite:

$$\int \frac{\tan(x)}{\ln(\cos(x))}\,dx = \int \frac{\frac{\sin(x)}{\cos(x)}}{\ln(\cos(x))}\,dx = \int \frac{\sin(x)}{\cos(x)\ln(\cos(x))}\,dx$$

and therefore the substitution will work fine:

$$= -\int \frac{1}{u\ln(u)}\,du.$$

Well, this integral is definitely an improvement, but we're still not done. This time, however, we see that there is a $\ln(u)$ in our integral, and also a $\frac{1}{u}\,du$, which is exactly what we'd need to make the substitution for $\ln(u)$! So (and of course we have to take a different letter) we set

$$w = \ln(u) \quad \Rightarrow \quad dw = \frac{1}{u}\,du$$

so that our integral becomes

$$= -\int \frac{1}{\ln(u)}\cdot\frac{1}{u}\,du = -\int \frac{1}{w}\,dw = -\ln|w|+c$$

which we have to devolve back into our original variable x as:

$$= -\ln|\ln(u)|+c = -\ln|\ln(\cos(x))|+c.$$

Wow, we really didn't see that one coming. We check by differentiating:

$$\frac{d}{dx}\left(-\ln|\ln(\cos(x))|\right) = \frac{-1}{\ln(\cos(x))}\left(\frac{1}{\cos(x)}(-\sin(x))\right) = \frac{\tan(x)}{\ln(\cos(x))},$$

which is what we wanted. □

Example 6.21.

$$\int \frac{(4t+2)^2}{t^2}\, dt$$

We look at this integral and think that the most complicated piece is $4t+2$ — but wait, a substitution is the SECOND thing you think of, after "CAN I SIMPLIFY THIS?" because in fact

$$\int \frac{(4t+2)^2}{t^2}\, dt = \int \frac{16t^2+16t+4}{t^2}\, dt = \int 16 + \frac{16}{t} + \frac{4}{t^2}\, dt = 16t + 16\ln|t| - \frac{4}{3x^3} + c$$

as you can check by differentiating. If you had done the substitution, you would have had to rewrite the denominator, making it more complex; but that would be awful: a sum in the numerator is great, because we can simplify, but a sum in the denominator is usually a big headache. □

6.2.4 Tips on substitution

- Not everything needs substitution. Always look to see if you can find an anti-derivative directly, first.
- When in doubt, start small. A little linear substitution like $u = 3x$ or $u = x - 2$ is always possible (since $du = 3dx$ or $du = dx$ can always be done) and it might make the mess look a lot more clear to you.
- Don't be too greedy with your substitution: don't take $u = \ln(\sin(x))$ in one go but instead start with $u = \sin(x)$ and see what happens.
- Be meticulous in your work. Sloppy substitution will give you garbage and is worthless.
- Make sure that you have translated every part of your integral to your new variable. Never write any integral with two different variables in it.
- Be flexible. Try a substitution even if you can't see how it will turn out. If you can't get it to work, or it gives a yuckier integral, don't erase it, but try a different substitution.

6.2.5 Two examples where substitution is not enough

Example 6.22. Consider

$$\int \frac{e^{3t}}{e^{-3t}+1}\, dt.$$

This is very similar to one we had before. We again try $u = e^{-3t} + 1$ which gives $du = -3e^{-3t}\, dt$. This is trickier: if we try directly we get

$$\int \frac{e^{6t}}{3u}\, du \quad \text{incomplete substitution: it's a mix of } t \text{ and } u$$

so now we have to use: $u = e^{-3t} + 1$ which gives $e^{-3t} = u - 1$ so $e^{3t} = \frac{1}{u-1}$ or $e^{6t} = \frac{1}{(u-1)^2}$. Thus the real substitution is

$$\int \frac{e^{3t}}{e^{-3t}+1}\, dt = \int \frac{1}{3u(u-1)^2}\, du.$$

This is fine; but we don't have the necessary techniques to solve this yet. That's for MAT1332. □

Example 6.23. Consider $\int e^{x^2}\, dx$. If we try $u = x^2$, we would need $du = 2x\, dx$. There is no x in the integral. But we could say: $u = x^2$ so $\sqrt{u} = x$. Therefore $dx = \frac{1}{2x}\, du = \frac{1}{2}\sqrt{u}\, du$. That means we have

$$\int e^{x^2}\, dx = \int \frac{1}{2\sqrt{u}} e^u\, du.$$

Nice — but we still don't see an antiderivative. In fact: there is *no* formula for a function whose derivative is e^{x^2} (or $u^{-1/2}e^u$, for that matter). The antiderivative must exist, but it is a brand new function that has no name or formula besides "an anti-derivative of e^{x^2}. □

End of lecture # 19

6.3 Techniques of integration : Integration by Parts

Last time we learned how to use substitution (which is like an anti-chain rule) to change one integral into a (hopefully) simpler integral. The goal is to change our integrand into an elementary function who anti-derivative we know. Today: we will learn how to use integration by parts, which you can think of as the anti-product rule.

Recall that the product rule tells us that

$$\int (f(x)g'(x) + g(x)f'(x))\, dx = f(x)g(x) + c.$$

So of course if your integrand looks like the left side, you can solve it. However, that's not what usually happens. Let's rewrite the above equation by splitting the left side into a sum of two integrals and moving it to the other side of the equation:

$$\int f(x)g'(x)\, dx = f(x)g(x) - \int g(x)f'(x)dx.$$

This tells us: if you have to solve $\int f(x)g'(x)\, dx$, then by using this identity you can reduce the problem to finding $\int g(x)f'(x)\, dx$. This is great if this other integral is easier!

6.3.1 Method of integration by parts

1. Divide your integrand into two pieces: a function that is easy to differentiate, and one that is easy to anti-differentiate.
2. Call the piece you will differentiate $u = f(x)$; call the rest $dv = g'(x)dx$. Then differentiate u to give $du = f'(x)dx$ and choose an antiderivative $v = g(x) = \int dv$. I write this in a little table like

$u = f(x)$	$dv = g'(x)dx$
$du = f'(x)dx$	$v = g(x)$

3. Using the notation of the integral makes the rule easier to remember:

$$\boxed{\int u\,dv = uv - \int v\,du}$$

since we multiply the two functions (on the diagonal in our table) and then subtract the integral of the product across the bottom row.

4. Note: unlike with substitution, your resulting integral is still in terms of x; just solve and check.

Example 6.24. Find

$$\int xe^x\,dx$$

We choose to split our integrand into $u = f(x)$ and $dv = g'(x)dx$, so:

$$\boxed{\begin{array}{ll} u = x & dv = e^x dx \\ du = dx & v = e^x \end{array}}$$

Now using the rule for integration by parts:

$$\begin{aligned}\int xe^x\,dx &= x \cdot e^x - \int e^x\,dx \\ &= xe^x - e^x + c \\ &= (x-1)e^x + c\end{aligned}$$

which we verify is correct by differentiation. □

Example 6.25. Find

$$\int x^2 e^{3x}\,dx.$$

We like to use something that differentiates well for u, so $u = f(x) = x^2$; and the rest should be easy to integrate, $dv = e^{3x}\,dx$:

$$\boxed{\begin{array}{ll} u = x^2 & dv = e^{3x}\,dx \\ du = 2x\,dx & v = \frac{1}{3}e^{3x} \end{array}}$$

(where we solved $\int e^{3x}\,dx$ either by staring really hard, or by using the substitution $t = 3x$). So we have

$$\int x^2e^{3x}\,dx = \frac{1}{3}x^2 \cdot e^{3x} - \int 2x \cdot \frac{1}{3}e^{3x}\,dx$$

$$= \frac{1}{3}x^2e^{3x} - \frac{2}{3}\int xe^{3x}\,dx \quad \text{by parts again:}$$

$$\boxed{\begin{array}{ll} u = x & dv = e^{3x}\,dx \\ du = dx & v = \frac{1}{3}e^{3x} \end{array}}$$

$$\begin{aligned}&= \frac{1}{3}x^2e^{3x} - \frac{2}{3}\left(\frac{1}{3}xe^{3x} - \frac{1}{3}\int e^{3x}\,dx\right) \\ &= \frac{1}{3}x^2e^{3x} - \frac{2}{9}xe^{3x} + \frac{2}{9}\left(\frac{1}{3}e^{3x} + c\right) \\ &= \frac{1}{3}x^2e^{3x} - \frac{2}{9}xe^{3x} + \frac{2}{27}e^{3x} + c'\end{aligned}$$

which we verify is correct by differentiation. □

Example 6.26.

$$\int \ln(x)\ dx$$

We do not know this antiderivative, and there is no substitution to make, since there is only the one function, so we try integration by parts. No choice:

$$\boxed{\begin{array}{ll} u = \ln(x) & dv = dx \\ du = \frac{1}{x} dx & v = x \end{array}}$$

Therefore

$$\int \ln(x)\ dx = x\ln(x) - \int x\frac{1}{x}\ dx = x\ln(x) - \int dx = x\ln(x) - x + c$$

where we have remembered that $\int dx = \int 1\ dx$. We check by differentiating! □

Example 6.27.

$$\int x\ln(x)\ dx$$

This time we have choices. You might be tempted to set $dv = \ln(x)\ dx$, since we now know the integral (try it! it gets messy) but the strategy is: choose u to have a nice derivative, and dv to have a nice integral. So here we go with

$$\boxed{\begin{array}{ll} u = \ln(x) & dv = x\ dx \\ du = \frac{1}{x}\ dx & v = \frac{1}{2}x^2 \end{array}}$$

which gives

$$\int x\ln(x)\ dx = \frac{1}{2}x^2\ln(x) - \int \frac{1}{2}x^2\frac{1}{x}\ dx = \frac{1}{2}x^2\ln(x) - \frac{1}{2}\int x\ dx = \frac{1}{2}x^2\ln(x) - \frac{1}{4}x^2 + c$$

which we again check by differentiating. □

Example 6.28.

$$\int \frac{\ln(x)}{x^2}\ dx$$

Again, doesn't seem to be one we know, or a good candidate for substitution, so we go to integration by parts; again, $\ln(x)$ is the one you'd like to differentiate, because then it turns into a function in the same family as x^{-2}, which will make the integral easier. This means

$$\boxed{\begin{array}{ll} u = \ln(x) & dv = \frac{1}{x^2}\ dx \\ du = \frac{1}{x}\ dx & v = -x^{-1} \end{array}}$$

which gives

$$\int \frac{\ln(x)}{x^2}\ dx = -\frac{\ln(x)}{x} - \int \left(-\frac{1}{x}\right)\frac{1}{x}\ dx = -\frac{\ln(x)}{x} + \int x^{-2}\ dx = -\frac{\ln(x)}{x} - x^{-1} + c$$

which we can check by differentiation. □

Remark 6.29. If these answers seem strangely similar, you might note that

$$-2x^{-2}\ln(x) = x^{-2}\ln(x^{-2}) = u\ln(u) \quad \text{with } u = x^{-2}.$$

Example 6.30.

$$\int \frac{\ln(x)}{x}\,dx$$

Given our success, we just dive right in:

$u = \ln(x)$	$dv = \frac{1}{x}\,dx$
$du = \frac{1}{x}\,dx$	$v = \ln(x)$

which gives[3]

$$\int \frac{\ln(x)}{x}\,dx = \ln(x)\ln(x) - \int \frac{\ln(x)}{x}\,dx$$

(!!!!??!!) But actually: this is marvelous. It's an equation, and the thing we want ($\int \frac{\ln(x)}{x}\,dx$) can be isolated:

$$2\int \frac{\ln(x)}{x}\,dx = \ln(x)^2$$

so

$$\int \frac{\ln(x)}{x}\,dx = \frac{1}{2}(\ln(x))^2 + c$$

where we remember to put $+c$ in the end. We check by differentiating. □

Remark 6.31. Actually, in the previous example, we should have noticed that it was a prime candidate for substitution: set $w = \ln(x)$ then $dw = \frac{1}{x}\,dx$, so

$$\int \frac{\ln(x)}{x}\,dx = \int w\,dw = \frac{1}{2}w^2 + c = \frac{1}{2}(\ln(x))^2 + c.$$

That's easier!

Exercise 6.32. Find $\int t^2 \ln(t)dt$.

Example 6.33. Find

$$\int \arcsin(x)\,dx.$$

We don't know this anti-derivative, and there's no substitution to make, so we try by parts.

$u = \arcsin(x)$	$dv = dx$
$du = \frac{1}{\sqrt{1-x^2}}dx$	$v = x$

which gives

$$\int \arcsin(x)\,dx = x\arcsin(x) - \int \frac{x}{\sqrt{1-x^2}}dx.$$

Now we figure out the resulting integral; since we see an xdx in the numerator, we'll do a substitution $w = 1 - x^2$ and $dw = -2xdx$. (You can use u if you want to, but I don't want to be confusing.)

$$\int \frac{x}{\sqrt{1-x^2}}dx = \int w^{-1/2}\frac{-1}{2}dw = -w^{1/2} + c = -\sqrt{1-x^2} + c$$

Therefore:

$$\int \arcsin(x)\,dx = x\arcsin(x) - (-\sqrt{1-x^2}) + c' = x\arcsin(x) + \sqrt{1-x^2} + c'$$

which we check by differentiation. □

Exercise 6.34. Find $\int \arctan(x)dx$.

[3] We used $\ln(x)$ instead of $\ln|x|$ here because the $\ln(x)$ in the integrand means we are only considering $x > 0$ anyway.

6.3.2 Applying by parts more than once: two different kinds of examples

Example 6.35. Find $\int x^2 \sin(3x)\, dx$.

This is a product of two unrelated functions so a good candidate for by parts.

$$\boxed{\begin{array}{ll} u = x^2 & dv = \sin(3x)dx \\ du = 2xdx & v = -\frac{1}{3}\cos(3x) \end{array}}$$

So

$$\begin{aligned} \int x^2 \sin(3x)\, dx &= -\frac{1}{3}x^2\cos(3x) - \int (-\frac{1}{3})\cos(3x)(2x)dx \\ &= -\frac{1}{3}x^2\cos(3x) + \frac{2}{3}\int x\cos(3x)\, dx. \end{aligned}$$

To work out the resulting integral, we need to use by parts again. So let's solve

$$\int x\cos(3x)\, dx$$

using by parts and then after that we'll plug it back into our equation.

$$\boxed{\begin{array}{ll} u = x & dv = \cos(3x)dx \\ du = dx & v = \frac{1}{3}\sin(3x) \end{array}}$$

which gives

$$\int x\cos(3x)\, dx = \frac{1}{3}x\sin(3x) - \int \frac{1}{3}\sin(3x)dx = \frac{1}{3}x\sin(3x) + \frac{1}{9}\cos(3x) + c$$

Therefore we have:

$$\begin{aligned} \int x^2\sin(3x)\, dx &= -\frac{1}{3}x^2\cos(3x) + \frac{2}{3}\int x\cos(3x)\, dx \\ &= -\frac{1}{3}x^2\cos(3x) + \frac{2}{3}\left(\frac{1}{3}x\sin(3x) + \frac{1}{9}\cos(3x)\right) + c' \\ &= -\frac{1}{3}x^2\cos(3x) + \frac{2}{9}x\sin(3x) + \frac{2}{27}\cos(3x) + c' \end{aligned}$$

which we check by differentiation. □

There's also a stranger way that doing integration by parts twice can pay off.

Example 6.36. Find

$$\int e^{-\theta}\cos(\theta)d\theta.$$

This is a product of two unrelated functions so a good candidate for by parts. We have two choices in this case, since both functions differentiation and integrate as easily; it doesn't matter which one we take. Let's go with:

$$\boxed{\begin{array}{ll} u = e^{-\theta} & dv = \cos(\theta)d\theta \\ du = -e^{-\theta}d\theta & v = \sin(\theta) \end{array}}$$

so

$$\int e^{-\theta}\cos(\theta)d\theta = e^{-\theta}\sin(\theta) - (-1)\int e^{-\theta}\sin(\theta)d\theta.$$

Now the resulting integral looks analogous to the one we had before; it is certainly not easier. But we persevere. We do integration by parts again.

CAREFUL: if at this point you were to choose $u = \sin(\theta)$ and $dv = e^{-\theta}d\theta$, you would just UNDO your first step and get back exactly to where you started. Try this, and then compare what happens with the magic of the following steps.

So we're trying to solve

$$\int e^{-\theta}\sin(\theta)d\theta$$

and we choose

$u = e^{-\theta}$	$dv = \sin(\theta)d\theta$
$du = -e^{-\theta}d\theta$	$v = -\cos(\theta)$

which gives

$$\begin{aligned}\int e^{-\theta}\sin(\theta)d\theta &= -e^{-\theta}\cos(\theta) - \int(-e^{-\theta})(-\cos(\theta))d\theta \\ &= -e^{-\theta}\cos(\theta) - \int e^{-\theta}\cos(\theta)d\theta.\end{aligned}$$

Now let's carefully write out what we've figured out, putting all this together:

$$\int e^{-\theta}\cos(\theta)d\theta = e^{-\theta}\sin(\theta) + (-e^{-\theta}\cos(\theta) - \int e^{-\theta}\cos(\theta)d\theta)$$

and the integral we want to solve for DOES NOT CANCEL OUT. In other words, we can add $\int e^{-\theta}\cos(\theta)d\theta$ to both sides of this equation to get

$$2\int e^{-\theta}\cos(\theta)d\theta = e^{-\theta}\sin(\theta) + (-e^{-\theta}\cos(\theta)$$

or

$$\int e^{-\theta}\cos(\theta)d\theta = \frac{1}{2}e^{-\theta}\sin(\theta) + \frac{1}{2}(-e^{-\theta}\cos(\theta) + c$$

which we check by differentation. □

Note: in this example, we could have swapped the functions we used for u and dv; the answer comes out the same.

6.3.3 Tips for integration by parts

- Good candidates for u: polynomials, exp, log, trig, inverse trig — anything whose derivative is a bit simpler
- Good candidates for dv: polynomials, exp, sine, cosine — functions whose anti-derivative is (a) known and (b) hopefully simpler
- If you do integration by parts twice, don't UNDO the first one.
- Keep CAREFUL TRACK of all signs. There's a minus sign in the formula, and often extra constants floating around. Be meticulous!

- If the result of your by parts doesn't look helpful, don't erase it! Try another combination of udv, or maybe come back and do a second by parts, or look for a substitution .... persistence is key!
- Remember that you have two big methods: substitution and by parts. They sometimes both work on the same integrand, but most of the time, substitution helps you with compositions of functions and most of the time, by parts helps you with products of functions.

End of lecture # 20

6.4 Mixed examples, and applications

6.4.1 More examples with integration by parts and substitution

Example 6.37.

$$\int \sin(\sqrt{x})\ dx$$

Now the messy part is $\sqrt{x}$ so we try a substitution, which effectively turns a composition problem into a product problem. So set $t = \sqrt{x}$. Knowing that we don't have the derivative available, we shortcut to saying $t^2 = x$ so $2t\ dt = dx$ by implicit differentiation. Then we can make the substitution:

$$\int \sin(\sqrt{x})\ dx = \int \sin(t)(2t)\ dt = 2\int t\sin(t)\ dt.$$

This is now a great candidate for by parts:

$u = t$	$dv = \sin(t)\ dt$
$du = dt$	$v = -\cos(t)$

which gives

$$2\int t\sin(t)\ dt = 2\left(-t\cos(t) - \int(-\cos(t))\ dt\right) = -2t\cos(t) + 2\sin(t) + c$$

and then finally come back to our original variable x:

$$= -2\sqrt{x}\cos(\sqrt{x}) + \sin(\sqrt{x}) + c.$$

We check by differentiating; as always with by parts, you see one of the product rule factors of the first cancels off the derivative of the second summand. □

Example 6.38. Find

$$\int \frac{\ln(y)}{\sqrt{y}}dy$$

We can try substitution: $x = \sqrt{y}$ so $dx = \frac{1}{2\sqrt{y}}dy$ and $y = x^2$; so

$$\int \frac{\ln(y)}{\sqrt{y}}dy = \int \ln(x^2)2dx = 2\int 2\ln(x)dx = 4\int \ln(x)dx$$

which is a bit of a surprise, perhaps. Anyway, this last integral we solved before, whence

$$= 4(x \ln(x) - x) + c$$

□

(We could also have solve this last one using by parts $u = \ln(y)$, $dv = \frac{1}{\sqrt{y}} dy$ and no substitution.)

6.4.2 Application examples

Example 6.39. A fish grows in length over time by a function $L(t)$ which obeys the pure-time differential equation

$$L'(t) = 7e^{-0.1t} \; cm/year.$$

Suppose that $L(0) = 0$ (meaning we measure from fertilization); how long until the fish reaches 50 *cm* in length?

Solution: Since $L'(t) = 7e^{-0.1t}$, we have

$$L(t) = \int 7e^{-0.1t} \, dt$$

and we make the substitution $u = -0.1t$, so $du = -0.1 \, dt$ or $dt = -10 \, du$. Thus

$$L(t) = 7 \int e^u \, (-10) \, du = -70e^u + c = -70e^{-0.1t} + c.$$

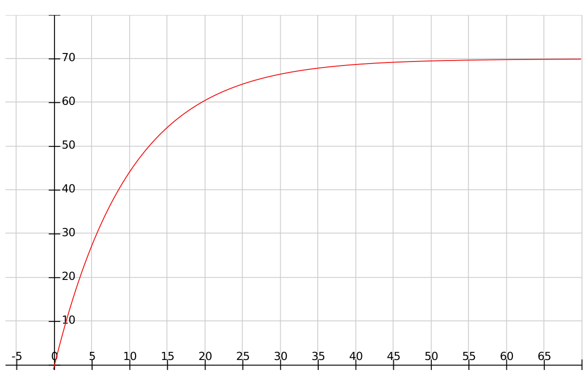
This seems like a stupid answer since the function takes negative values — but let's find c. Since $L(0) = 0$, we have

$$-70e^{-0.1(0)} + c = 0 \Leftrightarrow c = 70,$$

and thus our answer is

$$L(t) = 70 - 70e^{-0.1t} = 70(1 - e^{-0.1t})$$

which makes perfect sense! In fact, we can sketch the graph of $y = L(t)$ to get:



Therefore yes, this is a positive function for $t > 0$ and we compute

$$\begin{aligned} L(t) &= 50 \\ \Leftrightarrow 1 - e^{-0.1t} &= \frac{5}{7} \\ \Leftrightarrow e^{-0.1t} &= 1 - \frac{5}{7} \\ \Leftrightarrow e^{-0.1t} &= \frac{2}{7} \\ \Leftrightarrow -0.1t &= \ln(2/7) \\ \Leftrightarrow t &= \ln(2/7)/(-0.1) = 12.5 \; years. \end{aligned}$$

In its lifetime, it approaches a length of 70cm although it never stops growing.

□

Example 6.40. The mass of a worm, $M(t)$, changes over time according to the pure-time differential equation $M'(t) = ate^{-t}$, for some positive constant a. If $M(0) = 0$, find a formula for $M(t)$ and $\lim_{t\to\infty} M(t)$.

Solution: We need an antiderivative of $M'(t)$, so we have

$$M(t) = \int ate^{-t}\, dt = a \int te^{-t}\, dt.$$

This is a good candidate for by parts:

$u = t$	$dv = e^{-t}\, dt$
$du = dt$	$v = -e^{-t}$

giving

$$M(t) = a\left(-te^{-t} - \int (-e^{-t})\, dt\right) = -ate^{-t} + a\int e^{-t}\, dx = -ate^{-t} - ae^{-t} + c.$$

It seems worrisome that the functions are all negative, but let's plug in the initial condition of $M(0) = 0$. This gives

$$-a(0)e^0 - ae^0 + c = 0 \Leftrightarrow c = a > 0.$$

So our formula is

$$M(t) = a(1 - (t+1)e^{-t}).$$

Since $M'(t) > 0$ for all t and $M(0) = 0$, this is always positive (try it out!) and

$$\lim_{t\to\infty} M(t) = \lim_{t\to\infty} a\left(1 - \frac{t+1}{e^t}\right)$$

The quotient gives an indeterminate form of type ∞/∞ in the limit, so we can apply l'Hospital's rule

$$\lim_{t\to\infty} \frac{t+1}{e^t} = \lim_{t\to\infty} \frac{1}{e^t} = 0,$$

whence $\lim_{t\to\infty} M(t) = a$. The mass of the worm increases over its lifetime to asymptotically approach a. □

6.4.3 Integrals that we still can't solve

Example 6.41.

$$\int e^x/x\,dx$$

It is not a function we recognize. If we do a substitution $w = e^x$, then $x = \ln(w)$ so $dx = \frac{1}{w}\,dw$ and we get

$$\int e^x/x\,dx = \int \frac{w}{w\ln(w)}\,dw = \int \frac{1}{\ln(w)}\,dw$$

which we can't solve, either, and if we now substitution $t = \ln(w)$ then we will end up right back at x.

For integration by parts we have two choices; let's try them.

With

$u = e^x$	$dv = \frac{1}{x}\,dx$
$du = e^x\,dx$	$v = \ln\lvert x\rvert$

we get

$$\int e^x/x\,dx = e^x \ln\lvert x\rvert - \int \ln\lvert x\rvert e^x\,dx$$

which is not better; and if we do by parts again with $u = \ln\lvert x\rvert$ we'll end up back where we started.

With

$u = x^{-1}$	$dv = e^x\,dx$
$du = -x^{-2}\,dx$	$v = e^x$

we get

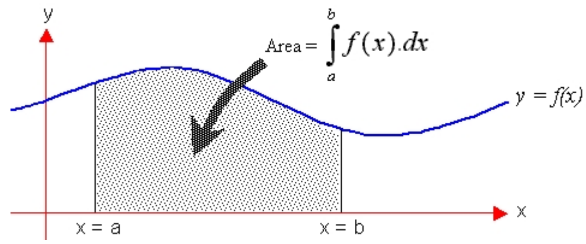
$$\int x^{-1}e^x\,dx = x^{-1}e^x + \int x^{-2}e^x\,dx$$

which is again worse than when we started. This integral again has no elementary function as antiderivative. □

6.5 Definite Integrals

We now switch to discussing what seems, at first, like a totally unrelated problem: finding the area of a region in the plane.

The area problem is as follows. Given a function $y = f(x)$, and two points a and b, find the area of the region bounded by $y = f(x)$, $x = a$, $x = b$ and the x-axis, as in the following picture:



Our strategy is to divide the interval $[a, b]$ into n subintervals, and approximate the area over each subinterval by the area of a rectangle, and add them all together. If we choose n larger and larger, then we should get closer and closer to the actual value we will define the integral as the limit as $n \to \infty$.

This looks like:

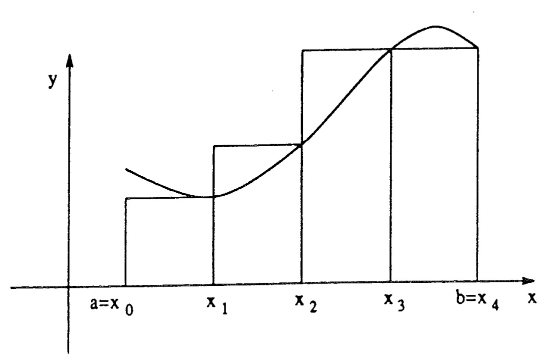
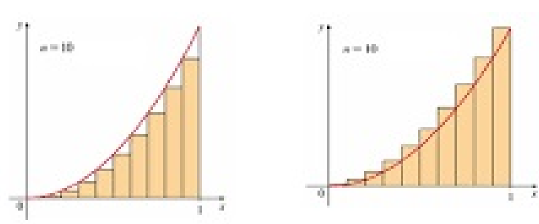


Figure 1

Let us do this in an example first.

Example 6.42. Consider $f(x) = x^2$, between $a = 0$ and $b = 1$.



We have many choices on how to set up our rectangles.

We could use the *left endpoint rule* : choose the height of each rectangle to be the value of the function at the left endpoint of the subinterval. So if we had 4 subintervals, we can draw the picture and the total area would be:

$$L_4 = \frac{1}{4}f(0) + \frac{1}{4}f(1/4) + \frac{1}{4}f(1/2) + \frac{1}{4}f(3/4) = \frac{1}{4}(0) + \frac{1}{4}(\frac{1}{16}) + \frac{1}{4}(\frac{1}{4}) + \frac{1}{4}(\frac{9}{16}) = 0.21875.$$

Here, the factor $\frac{1}{4}$ represents the width of each interval, and $f(x^*)$ the height.

What if we had built our rectangles using the *right endpoint rule* instead? Then with 4 subintervals, we would get

$$R_4 = \frac{1}{4}f(1/4) + \frac{1}{4}f(1/2) + \frac{1}{4}f(3/4) + \frac{1}{4}f(1) = \frac{1}{4}(\frac{1}{16}) + \frac{1}{4}(\frac{1}{4}) + \frac{1}{4}(\frac{9}{16}) + \frac{1}{4}(1) = 0.46875.$$

These answers are very different; but from the graph, we must believe that the correct value for the area lies somewhere between these two values.

Now let's figure out what the left and right rules would give us if we had n subintervals, for some positive integer n. Then the endpoints of all the subintervals are:

$$0 < \frac{1}{n} < \frac{2}{n} < \frac{3}{n} < \cdots < \frac{n-1}{n} < \frac{n}{n} = 1.$$

So if we use the left endpoint rule, our expression becomes

$$\begin{aligned} L_n &= \frac{1}{n}0^2 + \frac{1}{n}\left(\frac{1}{n}\right)^2 + \frac{1}{n}\left(\frac{2}{n}\right)^2 + \cdots + \frac{1}{n}\left(\frac{n-1}{n}\right)^2 \\ &= \frac{1}{n}\left(\frac{1}{n^2}\right)(0^2 + 1^2 + 2^2 + 3^2 + \cdots + (n-1)^2) \end{aligned}$$

and with the right endpoint rule, our expression becomes

$$\begin{aligned} R_n &= \frac{1}{n}\left(\frac{1}{n}\right)^2 + \frac{1}{n}\left(\frac{2}{n}\right)^2 + \frac{1}{n}\left(\frac{3}{n}\right)^2 + \cdots + \frac{1}{n}\left(\frac{n-1}{n}\right)^2 + \frac{1}{n}\left(\frac{n}{n}\right)^2 \\ &= \frac{1}{n}\left(\frac{1}{n^2}\right)(1^2 + 2^2 + 3^2 + \cdots + (n-1)^2 + n^2). \end{aligned}$$

This would be very tedious to work out if we didn't have the formula (which you may have seen in high school, else you can look up in the textbook):

$$1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}.$$

So for example

$$0^2 + 1^2 + 2^2 + 3^2 + \cdots + (n-1)^2 = \frac{(n-1)(n-1+1)(2(n-1)+1)}{6} = \frac{n(n-1)(2n-1)}{6}$$

and $1^2 + 2^2 + 3^2 + \cdots + (n-1)^2 + n^2 = \frac{1}{6}n(n+1)(2n+1)$.

Thus the expression for our estimate of the area using the left hand rule becomes

$$L_n = \frac{1}{n^3}\left(\frac{(n-1)(n)(2(n-1)+1)}{6}\right) = \frac{1}{6}\left(\frac{n-1}{n}\right)\left(\frac{2n-1}{n}\right);$$

you can check that it gives 0.21875 if you plug in $n = 4$. But what is interesting is:

$$\lim_{n\to\infty} L_n = \lim_{n\to\infty} \frac{1}{6}\left(1 - \frac{1}{n}\right)\left(2 - \frac{1}{n}\right) = \frac{2}{6} = \frac{1}{3}.$$

So the area we get by this line of reasoning is $\frac{1}{3}$.

On the other hand,

$$R_n = \frac{1}{n^3}\frac{n(n+1)(2n+1)}{6} = \frac{1}{6}(1+\frac{1}{n})(2+\frac{1}{n})$$

so

$$\lim_{n\to\infty} R_n = \lim_{n\to\infty}\frac{1}{6}(1+\frac{1}{n})(2+\frac{1}{n}) = \frac{2}{6} = \frac{1}{3},$$

which is the same![4] □

6.5.1 The definite integral of a function

We finally give the full definition of the integral, which we call the definite integral from now on; this definition is for any function, not just a positive one.

Definition 6.43. Suppose f is a function defined on $[a, b]$. The *definite integral of f from a to b* is

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} S_n$$

(provided this limit exists) where for each n the S_n are any Riemann sums with n subintervals.

Note:

- The numbers a and b are called the *(lower and upper) limits of integration.*
- Calculating the value of an integral (exactly, not approximately) is called *integration.*
- Think of $\int_a^b$ as a left parenthesis and dx as a right parenthesis; they are always matched up as a pair and dx doesn't mean anything by itself — it's just a way of saying what the name of the variable is.
- So

$$\int_a^b f(t)\,dt = \int_a^b f(x)\,dx = \int_a^b f(z)\,dz \left(\neq \int_a^b f(x)\,dt\right)$$

Example 6.44. So for example, we have calculated:

$$\int_0^1 x^2\,dx = \frac{1}{3}$$

□

Theorem 6.45 (*Fundamental Theorem of Calculus, Part 2* OR *Evaluation Theorem*)**.** *Suppose g is a differentiable function on $[a, b]$. Then*

$$\int_a^b g'(t)\,dt = g(b) - g(a).$$

Let's do some examples to understand this better.

[4] If you look at the expressions for L_n and R_n, you see that they differ only by $1/n$ — so as $n \to \infty$, this difference goes to 0, which is why they approach the same value.

Example 6.46. Suppose $g(x) = 3x$. Then $g'(x) = 3$, so $\int_a^b g'(x)\,dx$ is just the area of a rectangle (draw a picture). Thus $\int_a^b 3\,dx = 3(b-a) = 3b - 3a = g(b) - g(a)$. True. □

Example 6.47. Suppose $g(x) = \frac{1}{3}x^3$. Then $g'(x) = x^2$. We computed $\int_0^1 x^2\,dx = \frac{1}{3}$, and we compare with the FTC which says $\int_0^1 x^2\,dx = g(1) - g(0) = \frac{1}{3}(1)^3 - \frac{1}{3}(0)^3 = \frac{1}{3}$. True again. □

Example 6.48. Suppose $v(t)$ is the velocity function. This is the derivative of the displacement function $s(t)$, that is, $v(t) = s'(t)$. If $v(t)$ represents the instantaneous velocity of a particle at time t, then since each rectangle of the Riemann sum is velocity times time, giving displacement, we see that $\int_a^t v(t)\,dt$ represents the displacement of the object since an initial time $t = a$. We know that $\int_a^b v(t)\,dt$ is the displacement from $t = a$ to $t = b$; but that's precisely $s(b) - s(a)$ — your position at b minus your position at a. True again! □

So if we believe the theorem, then we deduce some very nice consequences:

Example 6.49. Suppose $r(t)$ is the rate of change of your population, where your population is given by $n(t)$. So $r(t) = n'(t)$. Then $\int_a^b r(t)\,dt = n(b) - n(a)$, that is, the area under the rate of change curve is just the total change in population. □

Example 6.50. If $\rho(x)$ represents the linear density of a rod, then each rectangle of the Riemann sum corresponds to density times length, which is mass. So in the limit, $\int_a^x \rho(x)dx$ is the total mass of a piece of rod starting at the point a and ending at x. That is, if $\rho(x)$ is the linear density of a rod, then this is $m'(x)$ where $m(x)$ is the mass of a length b of the rod, measured from any arbitrary starting point. Then $\int_a^b \rho(x)\,dx = m(b) - m(a)$. □

Example 6.51. Suppose $g(x) = f(x) + c$, for a constant c. Then $g'(x) = f'(x)$. Thus

$$\int_a^b g'(x)\,dx = \int_a^b f'(x)\,dx$$

which implies by the Evaluation theorem that

$$g(b) - g(a) = f(b) - f(a)$$

which seems fishy until we check: $g(b) = f(b) + c$, $g(a) = f(a) + c$, so the difference $g(b) - g(a)$ is indeed equal to $f(b) - f(a)$. So the choice of antiderivative doesn't matter. □

6.5.2 Why is the Fundamental Theorem of Calculus true?

Why is the theorem true? To prove Part 2 we're just going to compute the integral as a limit of Riemann sums, and use the Mean Value Theorem to choose our sample points in a fantastic way.

Recall: We defined the definite integral of f from a to b to be

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

(provided this limit exists) where

- $\Delta x = (b - a)/n$;
- for each $i \in \{0, 1, \cdots, n\}$, $x_i = a + i\Delta x$;
- x_i^* is any choice of *sample point* in the interval $[x_{i-1}, x_i]$.

We say that f is *integrable* on $[a, b]$ when $\int_a^b f(x)\, dx$ exists.

Let's divide $[a, b]$ into n subintervals and build a Riemann sum:

$$\sum_{i=1}^{n} g'(x_i^*)\Delta x$$

where $\Delta x = x_i - x_{i-1}$. Choose your sample point x_i^* using the Mean Value theorem, so that we have

$$g'(x_i^*) = \frac{g(x_i) - g(x_{i-1})}{x_i - x_{i-1}} \Rightarrow g(x_i) - g(x_{i-1}) = g'(x_i^*)\Delta x$$

so that our Riemann sum (writing it out in long notation) is just

$$\sum_{i=1}^{n}(g(x_i) - g(x_{i-1})) = (g(x_1) - g(x_0)) + (g(x_2) - g(x_1)) + (g(x_3) - g(x_2)) + \cdots + (g(x_n) - g(x_{n-1})).$$

This kind of sum — where almost everything cancels except in the first and last terms — is called a telescoping sum, and it simplifies to

$$g(x_n) - g(x_0) = g(b) - g(a)$$

as required. So when we choose our sample points using the Mean Value Theorem, our Riemann sum comes out to $g(b) - g(a)$, no matter how big n is. In particular, that means the limit as $n \to \infty$ is $g(b) - g(a)$. □

If that feels like a cheap trick: we put all the cards on the table in MAT2125 : Introduction to Real Analysis.

Remark 6.52. You can remember Part 2 as saying either of the following:

- The integral of the derivative is the net change.
- If you take a function, differentiate it, and then integrate it over an interval, you get the difference of the function at the endpoints of the interval.

It is really quite astonishing that something as complicated as the integral of a function can be evaluated so simply.

So the Fundamental Theorem of Calculus part 2 gives us that

$$\int_a^b f(t)dt = F(b) - F(a)$$

where F is any anti-derivative of f, that is, $F' = f$. The notation we use when solving problems is:

$$\int_a^b f(t)dt = F(t)\Big]_a^b = F(b) - F(a) \qquad \text{where } F' = f.$$

This theorem changes the problem of finding a limit of Riemann sums to the problem of finding an anti-derivative, which is often much easier.

Example 6.53. If $f(x) = 2x$ then we saw $F(x) = x^2$ and so

$$\int_2^7 2x\, dx = x^2\Big]_2^7 = 7^2 - 2^2 = 49 - 4 - 45.$$

□

Index