The Free Speech Calculus Text

Various authors, Gnu Free Documentation License (see notes)

November 22, 2004

Contents

Introduction

Discussion.

[author=garrett,style=friendly,label=introduction_to_whole_work,version=1, file =text_files/introduction_to_whole]
Relax. Calculus doesn't have to be hard, and its basic ideas can be understood by anyone. So, why does it have a reputation? Well, Calculus can be hard. Huh? That's right, it doesn't have to be hard, but it can be.

Calculus itself just involves two new processes, *differentiation* and *integration*, and *applications* of these new things to solution of problems that would have been impossible otherwise.

For better or for worse, most Calculus classes increase the overall level of difficulty above pre-Calculus, while teaching you the subject matter. What this means is that in addition to learning how to take the derivative, you'll set up word problems, solve some equations, interpret the results etc. Most of this is *algebra*, but the pieces are all held together by the Calculus.

The three hardest parts about a typical Calculus class are (in my opinion): (1) Setting problems up (reading word problems, setting up equations etc), (2) manipulating functions and equations in algebraic way, (3) keeping track of a bunch of parts of the problem and putting it together. Note that (1) and (3) should be really important for anyone who is going to use their head for a living. Note that none of these "hard parts" is taking derivatives or integrating. Of course, I could be biased since I teach the class!

Another thing to think about as you take this course is the role of the calculator. On the one hand, since we have graphing calculators, the way we use calculus should be a little different than how it was used in the past. However, math teachers are still figuring out what parts should change and what parts should stay the same. So please forgive us if we don't have it perfect quite yet.

On the other hand, some of the parts that have changed are now *harder*. So, in the past, it might have been a useful exercise (in a practical sense)for a student to learn how to graph $y = e^x/x$ by hand, now the use of this exercise is probably one of the following: (1) purely for the sake of learning, (2) because we have to be able to double check/understand what the calculators are telling us, (3) as a warm up for the problems that the calculator can't solve. Option (3) might mean being able to analyze all functions of the form $y = ae^{bx}/x$ where a and b are constants, but we don't know what they are ahead of time. Note that the calculator can't graph $y = ae^{bx}$; we can (and will) learn to do it, but this problem is a little harder than graphing $y = e^x/x$.

Discussion.

[author=duckworth,style=formal,label=introduction_to_whole_work,version=2, file =text_files/introduction_to_whole]
This text book aims to provide both insight into the essential problems of Calculus (and the related field of mathematical analysis) and a rigorous proof of all of the standard material in a Calculus class. However, we will not put rigor in the way of leisure or explanation. Thus, while we will prove everything, we will not always do so in the most sophisticated or efficient manner manner.

One of the special features of this text is to include discussion of the historical controversies and so-called paradoxes which made Calculus such an exciting and hard-won mathematical field.

Discussion.

[author=duckworth,style=middle,label=introduction_to_whole_work,version=3, file =text_files/introduction_to_whole]
This text will attempt to introduce the student to all of the varied roles which Calculus plays in science and academia. Calculus is an applied subject which forms the basis of elementary calculations in physics, biology, psychology, statistics, engineering, etc. Calculus is the first math class that most people have taken where they have to learn concepts that are not immediate generalizations of arithmetic or geometric intuition. Finally, and related to both of the above, Calculus is the first math class that many people take where statements are given that are not exercises in proof, like in geometry, but still need to be proven.

It is of course not easy to satisfy all of the above goals at the same time, so we will have to take a middle-of-the-road approach: we will offer a little bit of material towards each goal.

In addition to the elusive goals just layed out, there is the difficulty of having a widely defined audience: in a typical college Calculus class roughly one third of the students have seen Calculus before, and remember the material fairly well. Another third of the class has seen Calculus before, but did not absorb a significant part of the course. The final third of the class has not seen Calculus before.

It is of course not easy to write a book addressed to all of the above parts of the audience, so we will again take a middle-of-the-road approach: we will offer enough explanation that a student who has never seen the material before can, *with diligence*, learn Calculus. The phrase "with diligence" is supposed to suggest that such a student will have to expect to spend more time figuring out examples and discussion in this book than they needed for previous math classes. Such a student might also want to access extra material from study guides.

Chapter 1

Background

Discussion.

[author=duckworth, file =text_files/introduction_to_background]
In this chapter we gather together bacground material. This material actually comes in two varieties: the stuff that is really necessary to have a good chance of passing the class, and the stuff that it's ok to look up, or learn as you go along. Often Calculus books, and teachers, seem to say that you should know everything in this chapter before you start. Well, maybe the ideal Calculus student would, but most of us aren't ideal, and most of us can still pass a calculus class. In any case, I will try to make it clear what is really necessary to know from what is merely helpful.

1.1 The numbers

Discussion.

[author=duckworth,author=livshits, file =text_files/basics_about_numbers]
The natural numbers are symbolized by $\mathbb{N}$. These are the numbers 1, 2, 3, 4, ...

The integers are symbolized by $\mathbb{Z}$ (from the German word "Zahl"). These are the natural numbers, together with their negatives and together with 0. In other words, these are the numbers 0, ±1, ±2, ±3, ±4, ...

The rational numbers are symbolized by $\mathbb{Q}$. These are all the fractions you can make integers on the top and bottom. In other words, these are the numbers of the form $\frac{a}{b}$ with a and b integers. Every integer is a rational number because $\frac{a}{1} = a$. A decimal number is a rational number if and only if the decimal digits have a repeating pattern.

The real numbers are symbolized by $\mathbb{R}$. These include the rational numbers. You can think of the real numbers as being all the points on a number line. The real numbers form the heart of calculus; everything we do in calculus involves them and depends intimately upon their properties.

We can think of the real numbers as the set of all decimals numbers. This includes decimal numbers that extend infinitely to the right. Even decimal numbers with an infinite number of digits can be approximated with decimals numbers having only finitely many digits. For example, although π has an infinite number of digits, we can approximate it as $\pi \approx 3.1415926\cdots$. This is how our calculators and computers work: they approximate the set of all real numbers using only numbers with finitely many digits. This is why their answers are sometimes wrong; because they're based on approximations.

Discussion.

[author=duckworth,uses=complex_numbers,uses=extended_reals,uses=hyperreals, file =text_files/basics_about_numbers]
It is sometimes convenient to add some extra numbers to the real numbers. When we do so we go beyond many people's intuition, and this might make some students uncomfortable. Good! This discomfort is a sign of something interesting; I encourage you to explore any topic here which you think is strange, or suspicious. I'll just briefly say now that everything we do here can be rigorously justified, and that it's great fun to introduce new objects into your mathematics. With these new objects you can do things that were previously "forbidden": take the square root of negative numbers and divide by 0,

The complex numbers are denoted by $\mathbb{C}$ and are obtained by taking the real numbers $\mathbb{R}$ and joining them with the "imaginary" number i, which satisfies $i^2 = -1$. By "joining" I mean that you also take all sums, differences, products and quotients of things in $\mathbb{R}$ together with i. In other words, every complex number can be written in the form $a + bi$ where a and b are any real number. The arithmetic is $\mathbb{C}$ is defined by the two rules:

$$\begin{aligned} (a+bi)+(c+di) &= (a+c)+(b+d)i && \text{for all real numbers } a,b,c,d \\ (a+bi)(c+di) &= (ac-bd)+(ad+bc)i && \text{for all real numbers } a,b,c,d \end{aligned}$$

The extended real numbers do not have a standard symbol. They are obtained by taking the real numbers $\mathbb{R}$ and joining them with infinity ∞. Again, "joining" means taking all sums, differences, products, and quotients of things in $\mathbb{R}$ with ∞. The arithmetic in the extended real numbers is (loosely speaking) defined by the following rules:

$$\begin{array}{ll} -\infty \text{ is also an extended real number} & \\ a \pm \infty = \pm\infty & \text{for every real number } a \\ a(\pm\infty) = \pm\infty & \text{for every real number } a > 0 \\ -\infty < a < \infty & \text{for every real number } a \\ \frac{a}{0} = \pm\infty & \text{for every real number } a \neq 0 \\ \frac{a}{\pm\infty} = 0 & \text{for every real number } a \\ 0\infty \text{ is undefined} & \\ \infty - \infty \text{ is undefined} & \\ \frac{0}{0} \text{ and } \frac{\infty}{\infty} \text{ are undefined} & \end{array}$$

The hyperreal numbers are similar to the extended real numbers. They are obtained by taking the real numbers $\mathbb{R}$ and joining them with ∞, as well as an infinitesimal ϵ. The arithmetic in the hyperreals is (loosely speaking) defined by the following rules:

$-\epsilon$ and $-\infty$ are also hyperreals
$a \pm \infty = \pm\infty$ for every real number a
$a(\pm\infty) = \pm\infty$ for every real number $a > 0$
$-\infty < a < \infty$ for every real number a
$0 < a\epsilon < b$ for all real numbers $a, b > 0$
$\frac{a}{b} = \pm\infty$ for $b = 0, \epsilon$ and every real number $a \neq 0$
$\frac{a}{\pm\infty} = \pm\epsilon$ for every real number a
$b\infty$ is undefined for $b = 0, \epsilon$
$\infty - \infty$ is undefined
$\frac{a}{b}$ and $\frac{\infty}{\infty}$ are undefined (where $a, b = 0, \epsilon$)

These three extended systems of the real numbers have quite different uses, and mathematicians view them quite differently. The complex numbers are seen as a "simple" extension of the real numbers. They are used almost exactly the same way that real numbers are; to solve equations, to define polynomials, exponentials, logarithms, trigonometry, derivatives and anti-derivatives. The extended real numbers are viewed as a notational convenience. They allow one to writ things like $\frac{5}{\infty} = 0$ which is useful when calculating limits. The hyperreals are a modern version of the ideas that Newton and Leibnitz first used to develop Calculus. They are mathematically rigorous, deep, and can be used to prove all the results of Calculus we will use later. They also seem more abstract or foreign to many students than the complex numbers or the extended reals.

Discussion.

[author=wikibooks, file =text_files/rules_of_basic_algebra]
The following rules are always true and the basis of all algebra that we do in this class (and in other classes, like Linear Algebra, Abstract Algebra, etc.)

Algebraic Axioms for the Real Numbers 1.1.1.

[author=wikibooks,uses=algebraic_axioms_for_reals,label=algebra_axiom_for_real_number_field, file =text_files/rules_of_basic_algebra]
The following axioms, or rules, are satisfied by the real numbers.

- Addition is commutative: $a + b = b + a$
- Addition is associative: $(a + b) + c = a + (b + c)$
- Defining property of zero: $0 + a = a$ for all numbers a
- Defining property of negatives: For each number a, there is a unique number, which we write as $-a$, such that $a + (-a) = -a + a = 0$
- Defining property of subtraction: $a - b$ means $a + (-b)$ where $-b$ is defined as above.
- Multiplication is a commutative: $a \cdot b = b \cdot a$
- Multiplication is associative: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
- Defining property of one: $1 \cdot a = a$ for all numbers a

- Defining property of inverses: For every number a, except $a = 0$, there is a unique number, which we write as $\frac{1}{a}$, such that $a\frac{1}{a} = \frac{1}{a}a = 1$.
- Defining property of division: $\frac{a}{b}$ means $a \cdot \frac{1}{b}$

Comment.

[author=wikibooks,author=duckworth, file =text_files/rules_of_basic_algebra]
The above laws are true for all a, b, and c. This also means that the laws are true if a, b and c represent unknowns, or combinations of unknowns; in other words a, b and c can be variables, functions, formulas, etc. All the algebra we do in this class (or any other class), follows from these rules (as well as the rules of logic, and the rule that you can do the same thing to both sides an equation and you will still have an equation). Of course, all of us know lots of other algebraic rules, but each of these other rules must be built up, or derived, from the simple ones above.

Example 1.1.1.

[author=wikibooks,author=duckworth, file =text_files/rules_of_basic_algebra]
When you want to cancel or simplify something, if you're not sure what rule you're trying to use, look up the rule. For instance, occasionally people do the following, which is incorrect

$$\frac{2 \cdot (x+2)}{2} = \frac{2}{2} \cdot \frac{x+2}{2} = \frac{x+2}{2}.$$

So, how do Axioms **??** apply to this situation? Well first, let's review our rules for multiplying fractions. So, can we figure out $\frac{a}{b} \cdot \frac{c}{d}$ from Axioms **??**? Well, $\frac{a}{b}$ doesn't appear in Axioms **??**. In fact, $\frac{a}{b}$ is shorthand notation for $a \cdot \frac{1}{b}$, which does appear in Axioms **??**. So $\frac{a}{b} \cdot \frac{c}{d}$ really equals $ac\frac{1}{b}\frac{1}{d}$. Ok, now what. Now I claim that $\frac{1}{b}\frac{1}{d}$ must equal $\frac{1}{bd}$. Why? Well, by Axiom 1.1.1 there is a unique number which is the inverse of bd, and that number has the unique property that when you multiply it by bd you get 1. Well,

$$\begin{aligned} (\tfrac{1}{b}\tfrac{1}{d})bd &= (\tfrac{1}{b}b)(\tfrac{1}{d}d) && \text{by Axioms 1.1.1 and 1.1.1} \\ &= 1 \cdot 1 && \text{by Axiom 1.1.1} \\ &= 1 && \text{by Axiom 1.1.1} \end{aligned}$$

Therefore, $\frac{1}{b}\frac{1}{d}$ equals the inverse of bd, thus $\frac{1}{b}\frac{1}{d} = \frac{1}{bd}$. Therefore, $\frac{a}{b}\frac{c}{d} = a\frac{1}{b}c\frac{1}{d} = ac\frac{bd}{=}\frac{ac}{bd}$.

Note: I would never suggest that you go through these steps every time. We have just shown how to multiply two fractions, from now on, I would always just use the property we just derived.

Ok, now that we know how to multiply two fractions, we can straighten out the mistake above. It is *not* the case that $\frac{2(x+2)}{2} = \frac{2}{2}\frac{x+2}{2}$. Rather, We should have $\frac{2(x+2)}{2} = \frac{2}{2}\frac{x+2}{1} = 1(x+2) = x+2$.

Example 1.1.2.

[author=wikibooks, file =text_files/rules_of_basic_algebra]

For example, if you're not sure whether it's ok to cancel the $x+3$ in the following expression $\frac{(x+2)(x+3)}{x+3}$ you could justify the steps as follows:

$$\begin{aligned} \frac{(x+2)(x+3)}{x+3} &= (x+2)(x+3)\cdot\frac{1}{x+3} && \text{(Division definition)} \\ &= (x+2)\cdot 1 && \text{(Associtive law and Inverse law)} \\ &= x+2 && \text{(One law)} \end{aligned}$$

Discussion.

[author=duckworth,label=discussion_of_what_less_than_means, file =text_files/inequalities]
The real numbers are split in half; the positive numbers are on the right half of the number line and the negative numbers are on the left half.

For any real numbers a and b we say $a < b$ if a is to the left of b on the real number line. This is equivalent to having $b - a$ be positive.

Next, we're going to review basic facts and arithmetic about positive and negative numbers, and inequalities.

Order Axioms for the Real Numbers 1.1.2.

[author=duckworth,label=order_axioms_for_reals,label=order_axioms_for_reals, file =text_files/inequalities]
In additon to the algebraic axioms for the real numbers (see 1.1), we also have the following order axioms:

- The trichotomy law: for all real numbers a and b we have $a < b$ or $b < a$ or $a = b$.
- Transitivity: if $a \leq b$ and $b \leq c$ then $a \leq c$.
- Addition preserves order: if $a \leq b$ and c is any real number then $a+c \leq b+c$.
- Multiplication by positives preserves order: if $a \leq b$ and $c \geq 0$ then $ac \leq bc$.

Rule 1.1.1.

[author=garrett,label=rules_for_multiplying_pos_negatives, file =text_files/inequalities]
First, a person must remember that the *only* way for a product of numbers to be *zero* is that one or more of the individual numbers be zero. As silly as this may seem, it is indispensable.

Next, there is the collection of slogans:

- positive times positive is positive
- negative times negative is positive
- negative times positive is negative

- positive times negative is negative

Or, more cutely: the product of two numbers *of the same sign* is *positive*, while the product of two numbers *of opposite signs* is *negative*.

Extending this just a little: for a *product* of real numbers to be *positive*, the number of *negative* ones must be *even*. If the number of negative ones is *odd* then the product is *negative*. And, of course, if there are any zeros, then the product is zero.

Notation.

[`author=wikibooks`, file =`text_files/interval_notation`]
The notation used to denote intervals is very simple, but sometimes ambiguous because of the similarity to ordered pair notation

Let a and b be any real numbers, or $\pm\infty$, with $a \leq b$. We define the following sets, called **intervals**, on the real line:

$$\begin{aligned}
[a, b] &= \text{those } x \text{ of the form } a \leq x \leq b \\
(a, b) &= \text{those } x \text{ of the form } a < x < b \\
[a, b) &= \text{those } x \text{ of the form } a \leq x < b \\
(a, b] &= \text{those } x \text{ of the form } a < x \leq b
\end{aligned}$$

Unfortunately the notation (a, b) is the same notation as is used for x, y points. I'm sorry but mathematicians re-use notation and hope that the context makes it clear which meaning is intended.

There is also notation for combining intervals. The union notation $\cup$ means combine the intervals. Thus $(1, 2) \cup (3, 4)$ means the set of numbers that are in $(1, 2)$ or in $(3, 4)$.

Note: the use of the word "or" here is sometimes confusing. You might think of $(1, 2) \cup (3, 4)$ as equalling the interval $(1, 2)$ **and** the interval $(3, 4)$. You're not wrong if you think this way. But, mathematicians have learned through experience that it's best, linguisticly, to talk about a single number x rather than infinite sets of numbers. Thus, a single number x is in $(1, 2) \cup (3, 4)$ if x is in $(1, 2)$ **or** x is in $(3, 4)$.

Exercises

1. Find the intervals on which $f(x) = x(x - 1)(x + 1)$ is positive, and the intervals on which it is negative.

2. Find the intervals on which $f(x) = (3x - 2)(x - 1)(x + 1)$ is positive, and the intervals on which it is negative.

3. Find the intervals on which $f(x) = (3x - 2)(3 - x)(x + 1)$ is positive, and the intervals on which it is negative.

1.2 Functions

Definition 1.2.1.

[`author=duckworth,label=definition_of_function`, file =`text_files/what_is_a_function`]
A **function** is something which takes a set of numbers as inputs, and converts each input into exactly one output.

Comment.

[`author=duckworth,label=comment_explaining_functions`, file =`text_files/what_is_a_function`]
In our definition of function, “something” means rule or algorithm or procedure. The most familiar “something” is a formula like x^2 or $x + 3$.

The function $\sin(x)$ gives an example of something which you might think of as a formula, but actually depends upon a procedure. To find $\sin(.57)$ one “draws” a right triangle which contains the angle .57, and then $\sin(.57)$ equals the ratio of the opposite side over the hypotenuse. People are often bothered by this definition when the first learn it, because it’s not a formula. Eventually, time and experience make people more comfortable with $\sin(x)$ and we actually start to view it as one of our basic functions, as if we knew it’s formula.

Comment.

[`author=duckworth,label=comment_what_kind_of_functions_to_expect`, file =`text_files/what_is_a_function`]
Some of our basic “formulas” that we are familiar with, are actually defined by rules, like $\sin(x)$. In Calculus we will not add new basic functions, although later we will learn a rule which creates new functions from old, possibly without giving a formula for the new one.

For the time being, all functions that we will see will be given by one of the following:

1. With a formula involving basic functions like $\sin(x)$, x^2, e^x, etc. .
2. Piecwise: Giving more than one formula and piecing them together.
3. Graphically: Giving a graph with inputs on one axis and outputs on the other.
4. Numericaly: Listing a table of numbers for the inputs and outputs.
5. Implicitly: Describing the rule verbally or in a problem, or in a formul not solved for y.

Definition 1.2.2.

[`author=garrett,author=duckworth,label=definition_domain_and_range`, file =`text_`

files/what_is_a_function]
The collection of all possible inputs is called the **domain** of the function. The collection of all possible outputs is the **range**.

If the domain has not been stated explicitly, then we assume that the domain equals all real numbers which make the function defined. In this case it is usually easy, with a little work, to find an explicit description of the domain. The range is not usually explicitly stated and it is sometimes difficult to find an explicit description of it.

Discussion.

[author=garrett,label=discussion_what_to_look_for_in_domain, file =text_files/what_is_a_function]
If the domain of a function has not been explicitly stated, then here is how we can find it. We start by asking: What be used as inputs to this function without anything bad happening?

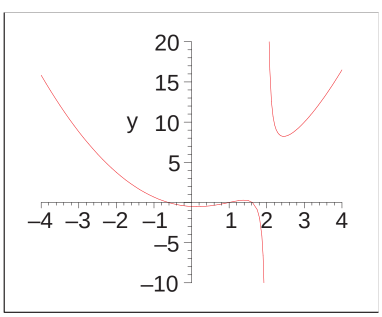
For our purposes, 'anything bad happening' just refers to one of

- trying to take the square root of a negative number
- trying to take a logarithm of a negative number
- trying to divide by zero
- trying to find *arc-cosine* or *arc-sine* of a number bigger than 1 or less than -1

(We note that some of these things aren't so bad if one is willing to work with the complex numbers, or the hyperreals.)

Discussion.

[author=duckworth,label=discussion_finding_range, file =text_files/what_is_a_function]
Finding the range of a function is generally harder than finding the domain. We should memorize the range of e^x, sin, cos, x, x^2, x^3 etc. For other functions we will learn various techniques later in this course that will help us find the range. Sometimes, we may have to graph the function.



simple_rational_function

Figure 1.1:

Example 1.2.1.

[author=duckworth,label=example_finding_domain_simple_rational_function, file =text_files/what_is_a_function]
Find the domain of $f(x) = \frac{1}{x-2} + x^2$. The only problem with plugging any number into this function comes from the division in the fraction. The only way we could have division by zero is is $x = 2$. Thus, the domain is all numbers except 2. This agrees with what we see in figure 1.1, namely that the graph does not exist at $x = 2$.

Example 1.2.2.

[author=garrett,label=example_finding_domain_sqrt_x^2-1, file =text_files/what_is_a_function]
For example, what is the domain of the function

$$y = \sqrt{x^2 - 1}?$$

Well, what could go wrong here? No division is indicated at all, so there is no risk of dividing by 0. But we are taking a square root, so we must insist that $x^2 - 1 \geq 0$ to avoid having complex numbers come up. That is, a preliminary description of the 'domain' of this function is that it is the set of real numbers x so that $x^2 - 1 \geq 0$.

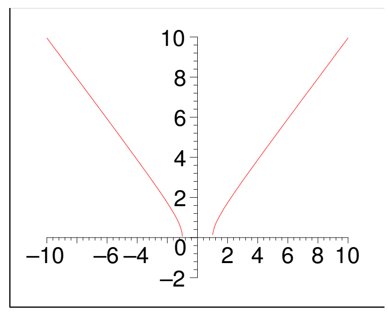
But we can be clearer than this: we know how to solve such inequalities. Often it's simplest to see what to *exclude* rather than *include*: here we want to *exclude* from the domain any numbers x so that $x^2 - 1 < 0$ from the domain.

We recognize that we can factor

$$x^2 - 1 = (x - 1)(x + 1) = (x - 1)\ (x - (-1))$$

This is negative exactly on the interval $(-1, 1)$, so this is the interval we must prohibit in order to have just the domain of the function. That is, the domain is the union of two intervals:

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

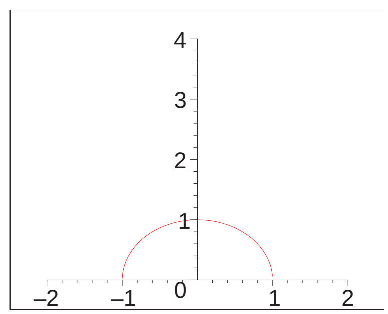


sqrt_x_squared_minus_1

You can also verify our answer by looking at the graph. Of course, we will always try to solve problems algebraically when possible, rather than just relying upon the graph. In any case, on the graph we don't see any points between $x = -1$ and $x = 1$, which is equivalent to saying that the domain equals what we described above.

Example 1.2.3.

[author=wikibooks,label=example_finding_domain_top_half_of_circle, file =text_files/what_is_a_function]
Let $y = \sqrt{1 - x^2}$ define a function. Then this formula is only defined for values of x between -1 and 1, because the square root function is not defined (in the world of real numbers) for negative values. Thus, the domain would be $[-1, 1]$. This agrees with the fact that the graph is the top half of a circle, and not defined outside of $[-1, 1]$.

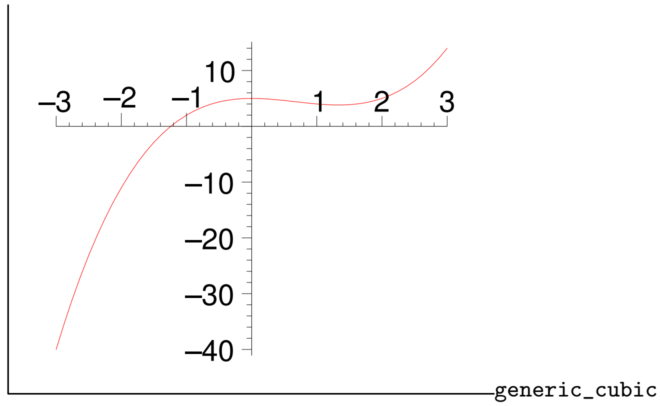


top_half_of_unit_circle

In this case it is easy to see that $\sqrt{1 - x^2}$ can only equal values from 0 to 1. Thus, the range of this function is $[0, 1]$.

Example 1.2.4.

[author=duckworth,label=example_function_given_by_graph, file =text_files/what_is_a_function]
Let $f(x)$ be defined by the graph below.



To determine a function value from the graph we read the y-value (off the vertical axis) which corresponds to some given x-value (on the horizontal axis). For example given the input of $x = 3$ the output is $y = 14$.

A graph shows us lots of information about the function, and much of what we learn later will be how to find this information without relying upon the graph. For example, we can see that there is a certain type of maximum at $x = 0$.

In problems like this, that depend upon the graph, we will generally not require very accurate answers. The answers only need to be accurate enough to show that we've read the graph correctly.

Example 1.2.5.

[author=duckworth,label=example_function_given_by_numbers, file =text_files/what_is_a_function]
Let the table of numbers below define a function, where x is the input and y is the output.

x	1	2	2.5	2.9	3.1	3.5	4
y	2.1	3.72	3.88	4.42	4.36	4.1	2.7

For example, given an input of $x = 1$, the output is $y = 2.1$. Given an input of $x = 4$ the output is $y = 2.7$. However, we can't say for sure what happens to an input of $x = 1.5$. We could make a leap of faith and guess that the corresponding output is somewhere between 2.1 and 3.72. For lots of functions this might be a reasonable assumption, but if we don't know anything else about this function we really can't be sure about this, or even if the output is defined. (Technically, if all we've been given is this table, then the output is definitely not defined. But in practice, we usually think that the table gives us a handful of values of a function which is defined for more numbers than shown.) Similarly, we can't be sure that this function has a maximum around $x = 3$, even though we probably all think that it should.

Example 1.2.6.

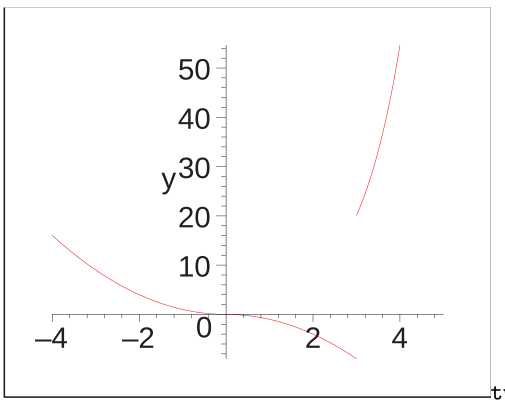
[author=duckworth,label=example_piecewise_function, file =text_files/what_is_a_function]
Let y be defined by the following formulas, each applying to just one range of

inputs.

$$y = \begin{cases} x^2 & \text{if } x \le 0 \\ -x^2 & \text{if } 0 < x \le 3 \\ e^x & \text{if } 3 < x \end{cases}$$

Which formula you use depends upon which x-value you are plugging in. To plug in $x = -1$ we use the first formula. So an input of $x = -1$ has an output of $(-1)^2 = 1$. To plug in $x = 2$ we use the second formula, so the output is $-2^2 = -4$. Similarly an input of $x = 4$ has an output of $y = e^4$.

We can also graph y. In this case it looks like x^2 on the left (i.e. for $x \le 0$); it looks like $-x^2$ in the middle (for $0 < x \le 3$) and it looks like e^x on the right (for $x > 3$). Notice that the graph looks "unnatural," especially at $x = 3$ where it is discontinuous.



two_parbs_and_exponential

Example 1.2.7.

[author=duckworth,label=example_function_implicit, file =text_files/what_is_a_function]

Let y be defined as a function of x, $x < 0$, by the equation:

$$x^3 + y^3 = 6xy$$

It is difficult (but not impossible) to find an explicit equation for y as a function of x. However, for each negative x-value, it is possible to compute a corresponding y-value, which is all we need for an *abstract* definition of function. For example, if $x = -.5$ I can have my calculator solve

$$(-.5)^3 + y^3 = 6 \cdot (-.5)y$$

for y (actually I'll probably have to enter it in the calculator using x instead of y!) to find $y \approx 0.04164259578$. Similarly, I could do this for *any* negative value for x; this is how y can be viewed as a function of x (only for negative values of x though).

To make this more concrete, but still not rely upon a formula, I could fill in a small table of numbers:

x	-1	$-.75$	$-.5$	$-.25$	0
y	0.1659	0.0936	0.0416	0.0104	0

What happens when we try to plug in a positive value for x like $x = 1$? There is more than one solution for y. This means that y is not a function of x for $x > 0$.

Discussion.

[author=livshits,uses=function_extensions,label=discussion_extension_restriction_of_functions, file =text_files/what_is_a_function]
We think of a *function* as a rule by which we can figure out $f(x)$ from x. Strictly speaking, we have to specify what objects x are being used, the collection of all these objects is called the *(definition) domain* of the function.

The home address is a real life example of a function. This function is defined for all the people that have home address, in other words, the definition domain of the home address is the collection of all the people who live at home. The home address is not defined for the homeless people. On the other hand, some homeless individuals pick up their mail at the post office and therefore have their postal addresses. For people who live at home their postal address and their home address coincide.

We say that the postal address is an *extension* of the home address to the homeless individuals who pick up their mail at the post office.

We also say that the home address is a *restriction* of the postal address to the individuals who live at home.

The notions of restriction and extension of functions are central to our approach to *differentiation*.

Discussion.

[author=duckworth,label=discussion_types_of_basic_functions, file =text_files/list_of_basic_functions]
In practice, in this class, we don't have that many basic functions. Here's most of them.

Polynomials These are positive powers of x, combined with addition and multiplication by numbers. We call the highest power that appears the **degree** of the polynomial. The numbers which are multiplied by x are called the **coefficients**. The **leading coefficent** is the coefficent of the highest power of x. The **constant term** is the number which has no power of x.

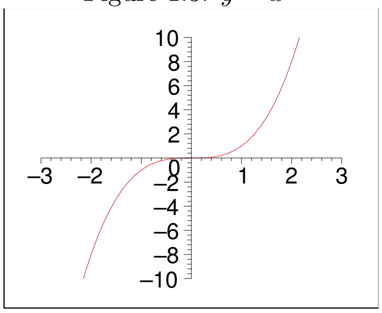
We can write a general expression for a polynomial, but since we don't now exactly what the degree will be, we need to use a letter to represent it; we will use n. Since we don't know how big the degree is, we can't write all the terms, thus we will leave out some number of terms in the middle, and will write "…" in their place. Similarly, we will need to use letters to represent the coefficients. The number of coefficients equals the degree varies with the degree, so we don't know how many letters we'll need. For this reason we don't usually write a general polynomial with letters of the form a, b, c, …, but rather we use a_0, a_1, a_2, etc. We summarize this terminology and show some examples in figure 1.2

Trigonometric Functions $\sin(x)$, $\cos(x)$, $\tan(x)$, $\sin^{-1}(x)$, $\cos^{-1}(x)$, $\tan^{-1}(x)$, $\csc(x)$, $\sec(x)$, $\cot(x)$

Figure 1.2: Polynomial examples

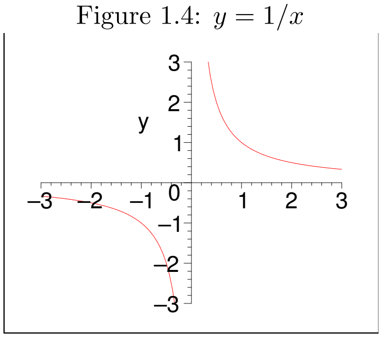
polynomial	degree	coefficients	leading coefficient	constant term
$x^3 + 5x + 6$	3	$1, 5, 6$	1	6
$10x^9$	9	10	10	0
$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$	n	$a_n, a_{n-1}, \ldots, a_1, a_0$	a_n	a_0

Figure 1.3: $y = x^3$



x_cubed_-3_to_3_manual

Figure 1.4: $y = 1/x$



1_over_x_-3_to_3_manual_fit

Exponential and Logarithm e^x, $\ln(x)$

We show some graphs of some of these functions in the next few figures.

Notation.

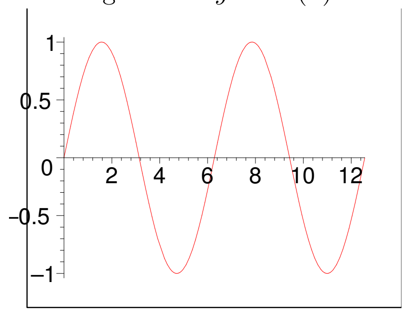
[author=duckworth,label=function_notation, file =text_files/function_notation]

Very Important

All functions use the following notation. When we write "$f(x)$" it means the following: f is the name of a function, x is the input (**anything** which comes inside of the parentheses ( ) is the input), $f(x)$ is the output you get when you plug in x. We read the notation "$f(x)$" as "f of x". We call x the input, or the independent variable, or the argument to f (this sounds somewhat old-fashioned, but it is still what inputs are called in computer science).

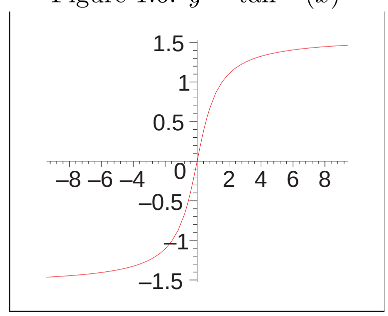
There is one family of exceptions to this notation. Out of laziness, or if you prefer, efficiency, many people write things like $\sin\pi$ instead of $\sin(\pi)$. People do

Figure 1.5: $y = \sin(x)$



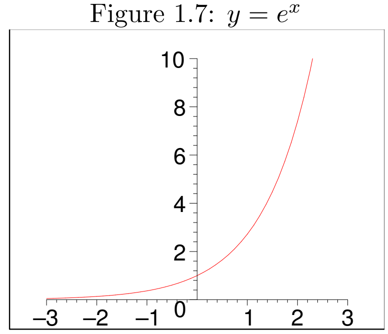
sin_0_to_4pi

Figure 1.6: $y = \tan^{-1}(x)$



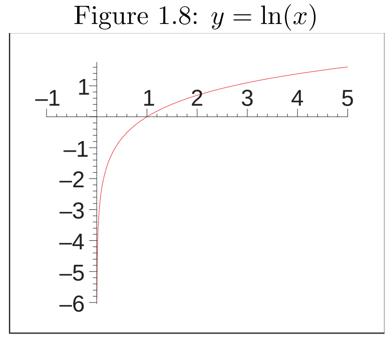
tan_inverse_-3pi_to_3pi

Figure 1.7: $y = e^x$



e_to_the_x_-3_to_3_manual

Figure 1.8: $y = \ln(x)$



ln_of_x_neg_1_to_5

the same thing with ln, cos, tan and the other trig functions. In this book, we will always use parentheses for these functions, unless the notation becomes too complicated and and it seems that leaving out some parenthesis would simplify it.

Example 1.2.8.

[`author=wikibooks,label=example_simple_function_notation`, file =`text_files/function_notation`]
For example, if we write $f(x) = 3x+2$, then we mean that f is the function which, if you give it a number, will give you back three times that number, plus two. We call the input number the argument of the function, or the independent variable. The output number is the value of the function, or the dependent variable.

For example, $f(2)$, (i.e. the value of f when given argument 2) is equal to $f(2) = 3 \cdot 2 + 2 = 6 + 2 = 8$.

Example 1.2.9.

[`author=duckworth`, file =`text_files/function_notation`]
Let f be the function given by $f(x) = x^2$. Then x represents the input, the output is x^2. For instance $f(2) = 4$.

Discussion.

[`author=wikibooks,author=duckworth,label=pros_cons_function_notation`, file =`text_files/function_notation`]
Function notation has great advantages over using $y = \ldots$ notation, but these advantages bring with them the need to be more careful and thoughtful about *exactly* what is being written. **Very Important**

Firstly, we can give different names to different functions. For example we coud say $f(x) = x^2$ and $g(x) = 3\sin(x)$ and then talk about f and g.

Another advantage of function notation as that it clearly labels inputs and outputs. In some of the previous function examples we had to use many phrases of the form "given the input $x = 3$ the output is $y = 7$". In function notation this becomes much more compact: "$f(3) = 7$".

Furthermore, it is possible to replace the input variable with any mathematical expression, not just a number. For instance we can write things like $f(7x)$ or $f(x^2)$ or $f(g(x))$; we'll talk more about what these mean below.

This last point brings up what we need to be careful and thoughtful about in function notation. This brings up a really important point. The variable "x" doesn't always mean x. It just stands for the input. So the function $f(x) = x^2$ could have been described this way "f is the function which takes an input and square it." Why do I care? Because we need to know how to calculate things like $f(3x)$ and $f(x+3)$.

If you get too focused on thinking that f squares $\underline{\ x\ }$, then you might think that $f(3x) = 3x^2$. No! The function squares any input, and in the case of $f(3x)$, the input is $3x$. So the output is $(3x)^2$.

Now, if you really understand the notation, you should be able to say what $f(x+3)$ is without a moment's hesitation. . . . . I hope you said $(x+3)^2$, but if not, keep practicing!

Example 1.2.10.

[**author=duckworth,label=example_function_notation_sin,uses=sin**, file =**text_files/function_notation**]
Let $f(x) = \sin(x)$. Then $f(\pi/2) = \sin(\pi/2)$. Now it so happens that $\sin(\pi/2)$ equals 1, so we can say that $f(\pi/2) = 1$. Similarly, $f(\pi/2+1) = \sin(\pi/2+1)$. Believe it or not, I don't know what $\sin(\pi/2+1)$ equals. It is *not* equal to $\sin(\pi/2)+1$. According to my calculator, $\sin(\pi/2+1)$ is approximately equal to 0.54.

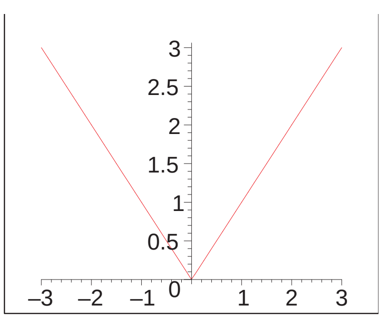
Examples 1.2.11.

[**author=wikibooks,label=example_various_functions**, file =**text_files/function_examples**]
Here are some examples of various functions.

1. $f(x) = x$. This takes an input called x and returns x as the output.
2. $g(x) = 3$. This takes an input called x, ignores it, gives 3 as an output.
3. $f(x) = x+1$. This takes an input called x, and adds one to it.
4. $h(x) = \begin{cases} 1, & \text{if } x > 0 \\ -1, & \text{if } x < 0 \end{cases}$.

   This gives 1 if the input is positive, and -1 if the input is negative. Note that the function only accepts negative and positive numbers, not 0. In other words, 0 is not in the "domain" of the function.
5. $g(y) = y^2$. This function takes an input called y and squares it.
6. $f(y) = \begin{cases} \int_{-y}^{y} e^{x^2}\,dx, & \text{if } y > 0 \\ 0, & \text{if } y \le 0 \end{cases}$.

   This function takes an input called y, and uses it as boundary values for an integration (which we'll learn about later).



abs_value_function

Example 1.2.12.

[**author=livshits,label=example_absolute_value_function,uses=absolute_value**, file =**text_files/function_examples**]
Here's one way to define the absolute value function:

$$|x| = \begin{cases} x & \text{if } x \text{ is already positive, or } 0 \\ -x & \text{if } x \text{ is negative} \end{cases}$$

You can think of this function as "making x positive" or "stripping the sign from x". You can also think of $|x|$ as the distance from x to 0 on the real number line;

this is a nice way to think about it, because it's geometric, and because the main reason we use absolute values is to give a mathematical expression to distances.

Discussion.

[`author=wikibooks,label=discussion_arithmetic_with_functions`, file =`text_files/combining_functions`]
Functions can be manipulated in the same manner as any other variable they can be added, multiplied, raised to powers, etc. For instance, let $f(x) = 3x + 2$ $g(x) = x^2$.

We define $f + g$ to be the function which takes an input x to $f(x) + g(x)$. If you completely understand function notation then you know what the formula for $f(x) + g(x)$ is .... $f(x) + g(x) = (3x + 2) + (x^2)$. Of course, this formula can be simplified to $f(x) + g(x) = x^2 + 3x + 2$.

Similarly,

- $f(x) - g(x) = (3x + 2) - (x^2) = -x^2 + 3x + 2$.
- $f(x) \cdot g(x) = (3x + 2) \cdot (x^2) = 3x^3 + 2x^2$.
- $f(x)/g(x) = (3x + 2)/(x^2) = \frac{3}{x} + \frac{2}{x^2}$.

However, there is one particular way to combine functions which is not like the usual arithmetic we do with variables: you can plug one function inside of the other! This is possibility really opens the door to many wonderful areas of mathematics way beyond Calculus, but for now we won't go there.

Definition 1.2.3.

[`author=wikibooks,label=definition_function_composition`, file =`text_files/combining_functions`]
Plugging one function inside of another is called **composition**. Composition is denoted by $f \circ g = (f \circ g)(x) = f(g(x))$. In this case, g is applied first, and then f is applied to the output of g.

Example 1.2.13.

[`author=wikibooks,label=example_simple_function_composition`, file =`text_files/combining_functions`]
For instance, let $f(x) = 3x + 2$ and $g(x) = x^2$ then $h(x) = f(g(x)) = f(x^2) = 3(x^2) + 2 = 3x^2 + 2$. Here, h is the composition of f and g. Note that composition is not commutative$f(g(x)) = 3x^2 + 2 \neq 9x^2 + 12x + 4 = (3x + 2)^2 = g(3x + 2) = g(f(x))$. Or, more obviously stated $f(g(x)) \neq g(f(x))$.

Examples 1.2.14.

[`author=duckworth,label=example_various_combinations_of_two_functions`, file =`text_`

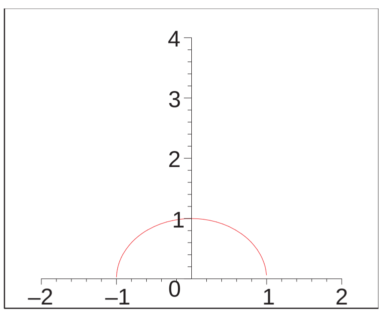
files/combining_functions]
Let $f(x) = x^2 + 1$ and $g(x) = \sin(x)$.

1. Find a formula for $f(x) + g(x)$.
2. Find $f(1) - g(1)$
3. Find a formula for $f(x)/g(x)$
4. Find a formula for $f(g(x))$.
5. Find $f(g(2))$.

Definition 1.2.4.

[author=duckworth,label=definition_one_to_one_function, file =text_files/one_to_one_functions_and_inverses]
A function $f(x)$ is **one-to-one** if it does not ever take to different inputs to the same output. In symbols: if $a \neq b$ then $f(a) \neq f(b)$.

Example 1.2.15.



top_half_of_unit_circle

[author=wikibooks,label=example_circle_is_not_one_to_one, file =text_files/one_to_one_functions_and_inverses]
The function $f(x) = \sqrt{1 - x^2}$ is not one-to-one, because both $x = 1/2$ and $x = -1/2$ result in $f(x) = \sqrt{3/4}$. You can see this graphically as the fact that the horizontal line $y = \sqrt{3/4}$ crosses the graph twice.

The function $f(x) = x + 2$ is one-to-one because for every possible value of $f(x)$ we never have two different inputs going to the same output. In symbols: if $a \neq b$ then $a + 2 \neq b + 2$, and therefore $f(a) \neq f(b)$.

Definition 1.2.5.

[author=duckworth,label=definition_function_inverses, file =text_files/one_to_one_functions_and_inverses]
Let $f(x)$ be a function. We say that another function $g(x)$ is the **inverse function** of $f(x)$ if $f(g(x)) = x$ and $g(f(x)) = x$ for all x. This means that f and g cancel each other.

An equivalent definition is that $f(a) = b$ if and only if $g(b) = a$. This means that g reverses inputs and outputs compared to f.

An equivalent way to think about this is that $g(x)$ is the answer to the question: f of what equals x?

Another equivalent way to think about this is that $f(x)$ has an inverse function if and only $f(x)$ is one-to-one.

Example 1.2.16.

[`author=wikibooks,label=example_function_inverse`, file =`text_files/one_to_one_functions_and_inverses`]
For example, the inverse of $f(x) = x + 2$ is $g(x) = x - 2$. To verify this note that $f(g(x)) = f(x - 2) = (x - 2) + 2 = x$.

The function $f(x) = \sqrt{1 - x^2}$ has no inverse (as we saw above). The function $\sqrt{1 + x^2}$ is close, but it works only for positive values of x. To verify this note that $f(\sqrt{1+x^2}) = \sqrt{1 - (\sqrt{1+x^2})^2} = \sqrt{1 - (1 + x^2)} = \sqrt{x^2} = |x|$ where $|x|$ is the absolute value of x.

Example 1.2.17.

[`author=duckworth,label=example_function_inverse_e^x_and_ln,uses=e^x,uses=ln`, file =`text_files/one_to_one_functions_and_inverses`]
Let's consider e^x and $\ln(x)$. These functions are inverses. So, for example, since $e^2 \approx 7.39$ we must have $\ln(7.39) \approx 2$. Another way to state this is that $e^{\ln(7.39)} = 7.39$ and $\ln(e^2) = 2$.

However, these numerical examples are not really how we use the fact that $\ln(x)$ and e^x are inverses. The following would be a much more common example.

Suppose that the amount of money in someone's bank account is given by $1000e^{.05t}$ where t is measured in years. Find out how many years it will take before they have \$3000.

This means that we want to solve $3000 = 1000e^{.05t}$. Dividing both sides by 1000 we get a new eqation

$$3 = e^{.05t}.$$

Now we can take $\ln(x)$ of both sides:

$$\ln(3) = \ln(e^{.05t}).$$

By the inverse property this means

$$\ln(3) = .05t$$

whence

$$t = \frac{\ln(3)}{.05}$$

This is a perfectly good expression for the final answer. Of course, some readers would rather get an explicit number for t; this is understandable, but you should practice being comfortable with answers that are formulas.

Example 1.2.18.

[`author=duckworth,label=example_function_inverse_cos,uses=cos`, file =`text_files/one_to_one_functions_and_inverses`]
Let's consider $\cos(x)$ and $\cos^{-1}(x)$. If I write $x = \cos(y)$ (and x is between $-\pi/2$ and $\pi/2$) this is (mathematically) equivalent to writing $y = \cos^{-1}(x)$. In other words, the two equations will be satisfied by exactly the same values of x and y.

Thus, saying that $\cos(\pi/2) = 0$ is equivalent to saying that $\pi/2 = \cos^{-1}(0)$.

(We will use this idea later to find the derivative of $\cos^{-1}(x)$. We will start with $y = \cos(x)$, solve this for $x = \cos^{-1}(y)$ and then apply implicit derivatives.)

Discussion.

[author=duckworth,label=discussion_how_to_find_inverses, file =text_files/one_to_one_functions_and_inverses]
Many of us learned how to find inverses by following these steps: given an equation $y = \ldots$, (1) reverse x and y, (2) solve the new equation for y.

I think this procedure sometimes makes people confused. To clear up the confusion, I hope you realize that step (1) is purely cosmetic. In other words, the only part of this step that matters is step (2), the reason we do step (1) is because we're not used to having a function of the form $x = \ldots$.

Let's illustrate. The equation which translates Farhenheit into Celsius is $C = \frac{5}{9}(F - 32)$. The inverse of this equation will translate Farhenheit into Celsius. We find the inverse by solving for F:

$$C = \frac{5}{9}(F - 32) \longrightarrow \frac{9}{5}C = F - 32 \longrightarrow F = \frac{9}{5}C + 32$$

Now, wasn't that simple?

If you follow the same steps for the equation $y = \frac{5}{9}(x-32)$ you get $x = \frac{9}{5}y+32$, and *this* is the sort of equation that step (1) was meant to prevent.

The moral of this story should be: don't get too hung up on the roles of x and y, they just represent two numbers. If you get too fixated on which what the input and output should look like, then you will sometimes have *extra* work to do, to sort out purely cosmetic problems.

Exercises

1. Find the domain of the function

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

That is, find the largest subset of the real line on which this formula can be evaluated meaningfully.

2. Find the domain of the function

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

3. Find the domain of the function

$$f(x) = \sqrt{x(x-1)(x+1)}$$

4. What is the graph of the function $y = -x$?

5. What is the graph of the function $y = |x|$?

6. What is the range of $f(x) = |x|$?

7. What is the range of the function $u(x) = 5$?

8. The function is defined by the formula $h(x) = |x|$, the domain of h is all the numbers x such that $-10 \leq x \leq 5$. What is the range of h?

9. What is the graph of this function?

10. What is fg? What is $1/f$? What are their graphs? What is the domain of $1/f$? What is the range of $1/f$?

11. $v(x) = x - 3$, what is the graph of $|v|$?

12. $u(x) = (x + 1)/(x - 1), q(x) = (x^2 - 1)/(x - 1)$. Find the domains of u and q.

13. $p(x) = x^2 + 2x + 5$, what is the range of p?

14. What is the degree of the product fg of 2 polynomials? Hint: What is the highest degree term of fg?

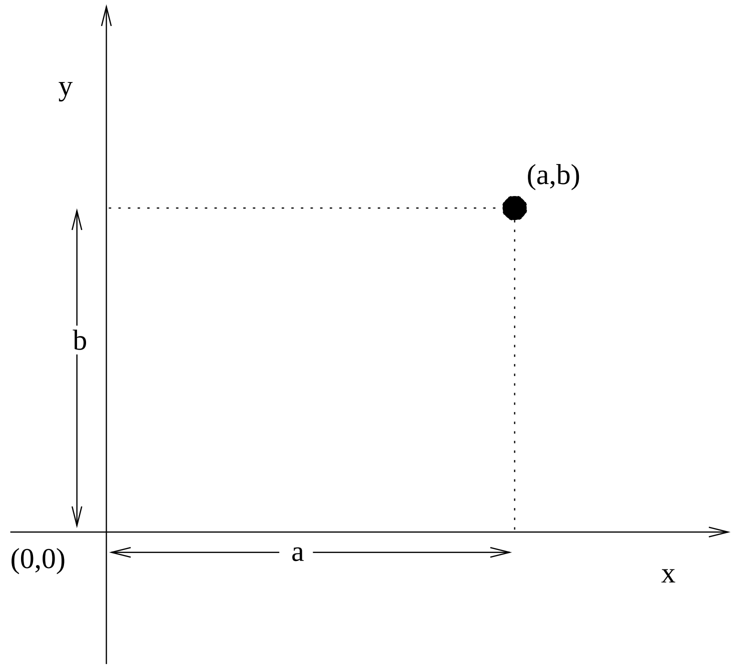
15. Let f and g be 2 nonzero polynomials. Can fg be zero? Hint: What is the leading term of fg?

16. Find the domain of $r(x)$. Check that $r(x) = u(x) = (x+1)/(x-1)$ for $x \neq 0$.

17. Find the domain of $z(x) = 1/(1/x)$. Check that for $x \neq 0$ $z(x) = x$.

18. Extend the function $q(x) = (x^2 - 1)/(x - 1)$ to $x = 1$ by a polynomial; in other words, find a polynomial $p(x)$ such that $p(x) = q(x)$ for $x \neq 1$.

Figure 1.9: A point in the x, y-plane



point_in_plane

1.3 Using, applying, and Manipulating functions and equations

Discussion.

[author=duckworth,label=intro_to_manipulating_functions, file =text_files/intro_to_manipulating_functions]
In this section we lay out some of the basic tools for using functions. This is sort of grab-bag of techniques.

We start by reviewing how equations can represent lines, circles, and other geometric objects.

We review some applications of functions to model real-world data.

We review how to solve some equations and inequalities.

Discussion.

[author=duckworth,label=discussion_of_point_in_xy_plane, file =text_files/cartesian_coords_and_graphs]
Recall that the x, y-plane refers to an ideal mathematical plane labelled with an x-axis which is horizontal, a y-axis which is vertical and the **origin** which is where the two axes intersect. Every point in the plane can be labelled with x and y coordinates which measure the horizontal and vertical distance respectively between the point and the origin, see figure 1.9

Definition 1.3.1.

[author=duckworth,label=definition_graph_of_function, file =text_files/cartesian_coords_and_graphs]
The **graph** of a function $f(x)$ is the set of points (x, y) such that x is in the domain and y equals $f(x)$. Given any equation involving x and y the graph of the equation is the set of points (x, y) which satisfy the equation.

Discussion.

[author=wikibooks,label=discussion_of_how_to_graph, file =text_files/cartesian_coords_and_graphs]
Functions may be graphed by finding the value of f for various x, and plotting the points $(x, f(x))$ in the x, y-plane.

Plotting points like this is laborious (unless you have your calculator do it). Fortunately, many functions's graphs fall into general patterns, and we can learn these patters. For example, consider a function of the form $f(x) = mx$. The graph of $f(x)$ is a straight line, and m controls how steeply angled the line is. Similarly we can learn about the graphs of our other basic functions, and later we will learn how to find out useful information about more complicated graphs as well.

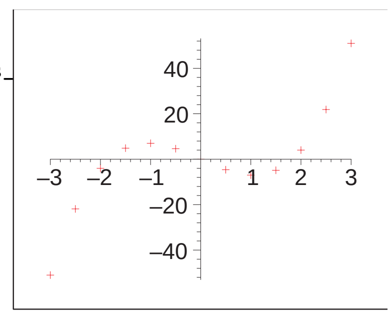
Example 1.3.1.

[author=duckworth,label=example_plotting_points, file =text_files/cartesian_coords_and_graphs]
Draw a picture of the graph of $f(x) = 3x^3 - 10x$ by plotting a few points.

First, we calculate some points:

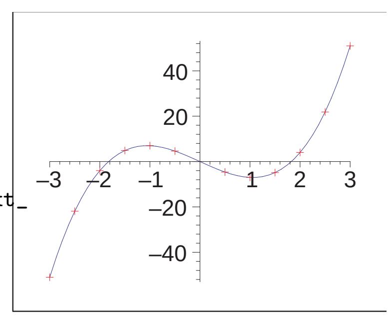
x	-3	-2.5	-2	-1.5	-1	$-.5$	0	$.5$	1	1.5	2	2.5	3
$f(x)$	-51	-21.9	-4	4.9	7	4.6	0	-4.6	-7	-4.9	4	21.9	51

These points are shown in Figure 1.10 Note, we could have saved some effort if we had only calculated half of these points. This function is odd (in a technical sense that we'll define later), and this would have told us that the values in the right half of our table would equal the negative of the values in the left half. Now we draw a smooth line through the points and get the graph shown in figure 1.11.



cubic_plot_points

Figure 1.10:



cubic_plot_points_curve

Figure 1.11:

Definitions 1.3.2.

[author=garrett,author=duckworth,label=defintion_slopes_equations_lines, file =text_files/lines_and_circles]
The simplest graphs are straight **lines**. The main things to remember are:

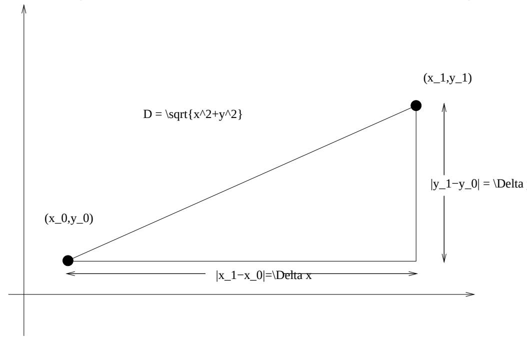
- The **slope** of a line is the ratio

$$m = \frac{\Delta y}{\Delta x}$$

where Δy and Δx are the vertical and horizontal change between two points. If the two points have coordinatens (x_0, y_0) and (x_1, y_1) then we have

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

Figure 1.12: Distance formula triangle

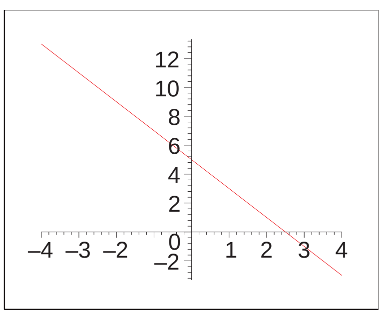


distance_formula_triangle

- A vertical line has equation $x = a$ for some number a. A horizontal line has equation $y = c$ for some number c.
- The **slope-intercept** form of the equation of a line is $y = mx + b$. This form is convenient for graphing by hand, but it is not as convenient for some other purposes.
- The **point-slope** form of the equation of a line with slope m and containing a point (x_0, y_0) is given by

  $$y = m(x - x_0) + y_0.$$

  This is by far the most convenient form of the equation of a line for us to use in Calculus.



straight_line

Example 1.3.2.

[author=duckworth, file =text_files/lines_and_circles]
The line $y = -2x + 5$ has slope of -2 and a y-intercept of 5. It's graph is shown in the margin figure .

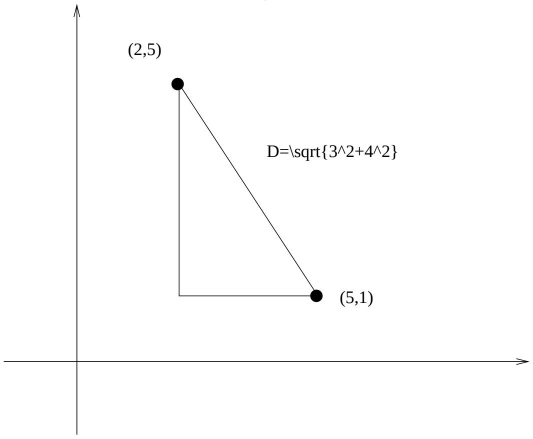
Definition 1.3.3.

[author=duckworth,label=definition_distance_formula, file =text_files/lines_and_circles]
Given two points (x_0, y_0) and (x_1, y_1), in the x, y-plane, their distance apart can be computed by drawing a right triangle that contains them and applying the Pythagorean theorem (see figure 1.12). This gives distance as

$$d = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

Example 1.3.3.

Figure 1.13: Example of distance formula



`example_of_distance_formula`

[`author=duckworth,label=example_distance_between_two_points`, file =`text_files/lines_and_circles`]
The distance between the points $(2,5)$ and $(5,1)$ (see figure 1.13) is

$$\sqrt{(2-5)^2+(5-1)^2}=\sqrt{9+16}=25$$

Definition 1.3.4.

[`author=duckworth,label=definition_equation_of_circle`, file =`text_files/lines_and_circles`]
The equation for distance can be immediately translated into the equation for a circle.

The equation of a circle centered at the origin and with radius r (see figure 1.14) is given by

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

This can be put into function form by solving for y; if we do this we get two values of y, so we need two functions

$$y_1=\sqrt{r-x^2} \text{ and } y_2=-\sqrt{r-x^2}.$$

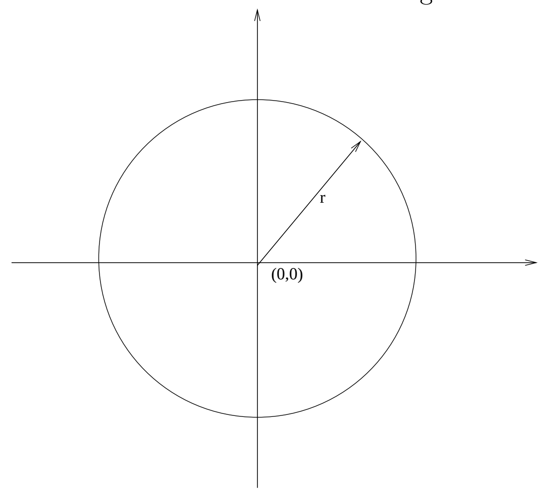
The equation of a circle with center at the point (a,b) and radius r (see figure **??**) is given by

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

Example 1.3.4.

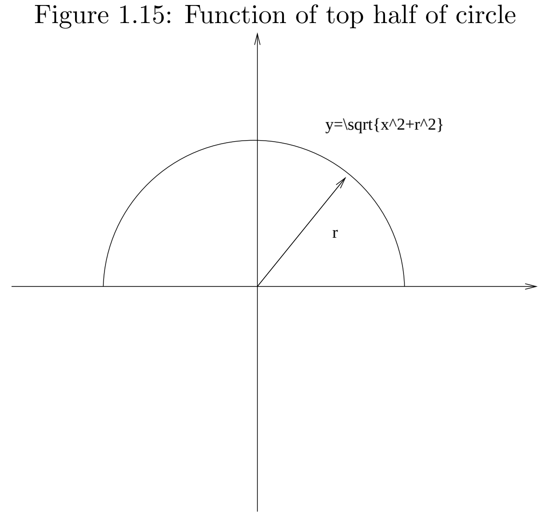
[`author=livshits,label=example_graph_of_circle_and_lines`, file =`text_files/lines_and_circles`]

Figure 1.14: Generic circle centered at origin with radius r



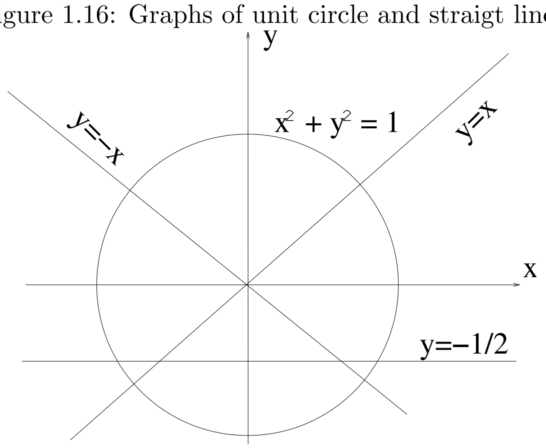
generic_circle

Figure 1.15: Function of top half of circle



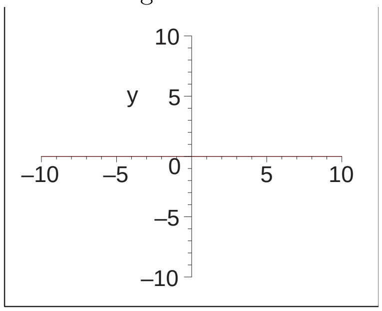
top_half_of_circle

Figure 1.16: Graphs of unit circle and straigt lines



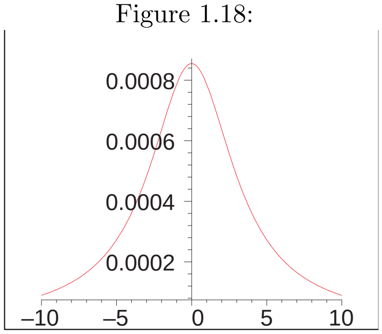
circle_and_axes

Figure 1.17:



`short_rational_fuction_standard_window`

Figure 1.18:



`short_rational_fuction_fit`

In figure 1.16 we show the graphs of the unit circle $x^2 + y^2 = 1$, and the straight lines $y = x$, $y = -x$ and $y = -1/2$.

Discussion.

[`author=duckworth,label=discussion_graphing_on_calculators_not_always_easy`, file =`text_files/complicated_graphs`]
Using calculators does not always make it perfectly easy to graph a function. We collect now a few examples of things which take some work to graph.

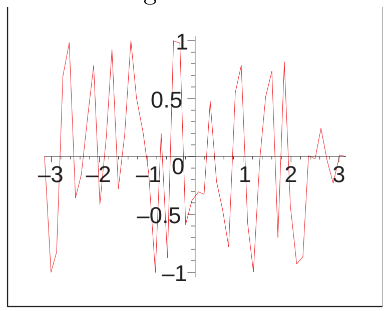
Example 1.3.5.

[`author=duckworth,label=example_graph_squished_rational_function`, file =`text_files/complicated_graphs`]
Using your calculator, graph $y = \frac{1}{100x^2+1170}$.

In figure1.17 we show a standard view (i.e. $-10 \le x \le 10$, $-10 \le y \le 10$) of this graph. This view doesn't help much.

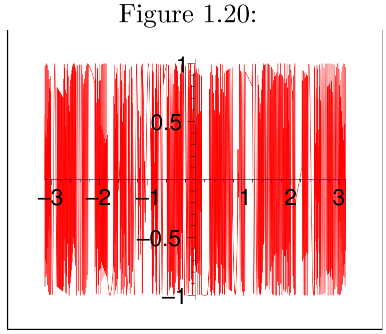
`ZoomFit` is a very useful feature on the calculators for graphs like this. To use this feauter *you must specify the x range*, and then the calculator will fit the y-values of the window to the graph. The result is shown in figure 1.18.

Figure 1.19:



`sin_of_5000_x_calculator`

Figure 1.20:



`sin_of_5000_x_normal_sample`

Example 1.3.6.

[`author=duckworth,label=example_sin_of_5000x`, file =`text_files/complicated_graphs`] Using your calculator, graph $y = \sin(5000x)$.

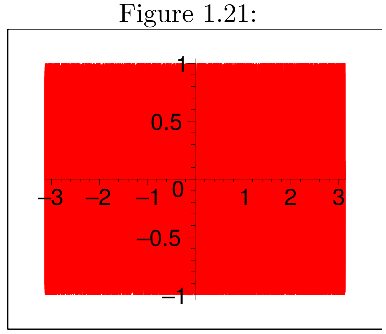
What this figure looks like will depend upon your machine. On your calculator, with an x-range of $-\pi \leq x \leq \pi$, it might look like the graph in figure 1.19.

In the computer package `Maple` the default graph is shown in figure 1.20.

In `Maple` we can increase the number of points used to sample the graph. The result of using a large number of points is shown in figure 1.21.

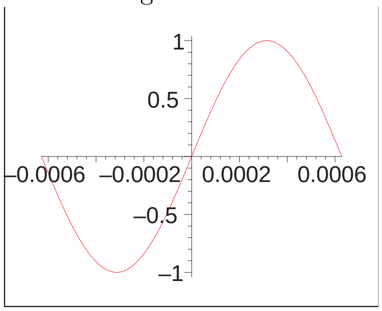
None of these pictures really shows the graph properly. To get a good graph,

Figure 1.21:



`sin_of_5000_x_massive_sample`

Figure 1.22:



`sin_of_5000_x_narrow_range`

we should use our knowledge of $\sin(x)$. We know that $\sin(x)$ oscillates. It turns out that $\sin(ax)$ still oscillates, but it oscillates *faster* if a is greater than 1. To show a graph that is oscillating faster, we need a *smaller* window. Roughly speaking, to graph $\sin(5000x)$ we should use a graph that is 5000 times smaller than usual. Thus, we try graphing with the range $\frac{-\pi}{5000} \leq x \leq \frac{\pi}{5000}$. The results are shown in figure1.22.

Discussion.

[`author=duckworth,label=discussion_graphing_with_calculators_wrap_up`, file =`text_files/complicated_graphs`]
We will see later examples of functions that are even more difficult to graph than the ones we have shown here. In fact, functions which are impossible to graph well are quite common; most of the graphs in Calculus textbooks are artifically nice and well-behaved.

We will also see later examples of problems that show how our calculators and computers can lead us to incorrect solutions.

Definition 1.3.5.

[`author=duckworth,label=definition_mathematical_models`, file =`text_files/mathematical_models`]
A mathematical model is a function that is used to describe a real-world set of data. Sometimes this can be done by exactly solving equations for various parameters. Sometimes this we can only find the model which comes closest to matching some data; in this case we usually need to use our calculators or computers.

Example 1.3.7.

[`author=duckworth,label=example_modelling_population,uses=e^x,uses=ln`, file =`text_files/mathematical_models`]
Use an exponential model (i.e. $P = Ce^{kt}$) to match the following populations

(where $t = 0$ corresponds to 1980), and predict the population in 2020:

Year	Population
1980	4 billion,
2000	5 billion

We wish to find k and C such that the following two equations are satisfied:

$$\begin{aligned} 4 &= Ce^0 \\ 5 &= Ce^{20k} \end{aligned}$$

From the first equaton we see that $C = 4$. Plugging this into the second equation we get

$$5 = 4e^{20k}.$$

As soon as you see an equation with an unknown as an exponent, you can be sure that we will use $\ln(x)$ to find that unknown. In this case, I'll divide by 4 first:

$$5/4 = e^{20k}$$

and then take $\ln(x)$ of both sides (using the cancelling property of $\ln(x)$ and e^x, see Example 1.2)

$$\ln(5/4) = 20k$$

whence

$$k = \frac{1}{20}\ln(5/4).$$

Example 1.3.8.

[author=garrett,label=example_solving_polynomial_inequality,version=1, file =text_files/solving_inequalities]
Solve the following inequality:

$$5(x-1)(x+4)(x-2)(x+3) < 0$$

The roots of this polynomial are $1, -4, 2, -3$, which we put in order (from left to right)

$$\ldots < -4 < -3 < 1 < 2 < \ldots$$

The roots of the polynomial P break the numberline into the intervals

$$(-\infty, -4),\ (-4, -3),\ (-3, 1),\ (1, 2),\ (2, +\infty)$$

On each of these intervals the polynomial is either positive all the time, or negative all the time, since if it were positive at one point and negative at another then it would have to be zero at some intermediate point!

For input x to the right (larger than) all the roots, all the factors $x+4$, $x+3$, $x-1$, $x-2$ are positive, and the number 5 in front also happens to be positive. Therefore, on the interval $(2, +\infty)$ the polynomial $P(x)$ is *positive*.

Next, moving *across* the root 2 to the interval $(1, 2)$, we see that the factor $x-2$ changes sign from positive to negative, while all the other factors $x-1$, $x+3$, and $x+4$ do *not* change sign. (After all, if they would have done so, then they would have had to be 0 at some intermediate point, but they *weren't*, since we know where they *are* zero...). Of course the 5 in front stays the same sign.

Therefore, since the function was *positive* on $(2, +\infty)$ and just one factor changed sign in crossing over the point 2, the function is *negative* on $(1, 2)$.

Similarly, moving *across* the root 1 to the interval $(-3, 1)$, we see that the factor $x - 1$ changes sign from positive to negative, while all the other factors $x - 2$, $x + 3$, and $x + 4$ do *not* change sign. (After all, if they would have done so, then they would have had to be 0 at some intermediate point). The 5 in front stays the same sign. Therefore, since the function was *negative* on $(1, 2)$ and just one factor changed sign in crossing over the point 1, the function is *positive* on $(-3, 1)$.

Similarly, moving *across* the root -3 to the interval $(-4, -3)$, we see that the factor $x + 3 = x - (-3)$ changes sign from positive to negative, while all the other factors $x - 2$, $x - 1$, and $x + 4$ do *not* change sign. (If they would have done so, then they would have had to be 0 at some intermediate point). The 5 in front stays the same sign. Therefore, since the function was *positive* on $(-3, 1)$ and just one factor changed sign in crossing over the point -3, the function is *negative* on $(-4, -3)$.

Last, moving *across* the root -4 to the interval $(-\infty, -4)$, we see that the factor $x + 4 = x - (-4)$ changes sign from positive to negative, while all the other factors $x - 2$, $x - 1$, and $x + 3$ do *not* change sign. (If they would have done so, then they would have had to be 0 at some intermediate point). The 5 in front stays the same sign. Therefore, since the function was *negative* on $(-4, -3)$ and just one factor changed sign in crossing over the point -4, the function is *positive* on $(-\infty, -4)$.

In summary, we have

$$\begin{aligned}
5(x-1)(x+4)(x-2)(x+3) &> 0 \text{ on } (2, +\infty) \\
5(x-1)(x+4)(x-2)(x+3) &< 0 \text{ on } (1, 2) \\
5(x-1)(x+4)(x-2)(x+3) &> 0 \text{ on } (-3, 1) \\
5(x-1)(x+4)(x-2)(x+3) &< 0 \text{ on } (-4, -3) \\
5(x-1)(x+4)(x-2)(x+3) &> 0 \text{ on } (-\infty, -4).
\end{aligned}$$

There's another way to write this. The polynomial is negative on $(1, 2) \cup (-4, -3)$. (The notation $(1, 2) \cup (-4, -3)$ means all those x-values between 1 and 2, together with all those x-values betwen -4 and -3.)

Example 1.3.9.

[`author=garrett,label=solving_polynomial_inequality,version=2`, file =`text_files/solving_inequalities`]
As another example, let's see on which intervals

$$P(x) = -3(1 + x^2)(x^2 - 4)(x^2 - 2x + 1)$$

is positive and and on which it's negative. We have to factor it a bit more: recall that we have nice facts

$$x^2 - a^2 = (x - a)\,(x + a) = (x - a)\,(x - (-a))$$

$$x^2 - 2ax + a^2 = (x - a)\,(x - a)$$

so that we get

$$P(x) = -3(1 + x^2)(x - 2)(x + 2)(x - 1)(x - 1)$$

It is important to note that the equation $x^2 + 1 = 0$ has no *real* roots, since the square of any real number is non-negative. Thus, we can't factor any further than this over the real numbers. That is, the roots of P, in order, are

$$-2 << 1 \text{ (twice!)} \quad < 2$$

These numbers break the real line up into the intervals

$$(-\infty, -2), \ (-2, 1), \ (1, 2), \ (2, +\infty)$$

For x larger than all the roots (meaning $x > 2$) all the factors $x + 2$, $x - 1$, $x - 1$, $x - 2$ are *positive*, while the factor of -3 in front is *negative*. Thus, on the interval $(2, +\infty)$ $P(x)$ is *negative*.

Next, moving *across* the root 2 to the interval $(1, 2)$, we see that the factor $x - 2$ changes sign from positive to negative, while all the other factors $1 + x^2$, $(x - 1)^2$, and $x + 2$ do *not* change sign. (After all, if they would have done so, then they would have be 0 at some intermediate point, but they *aren't*). The -3 in front stays the same sign. Therefore, since the function was *negative* on $(2, +\infty)$ and just one factor changed sign in crossing over the point 2, the function is *positive* on $(1, 2)$.

A *new feature* in this example is that the root 1 occurs *twice* in the factorization, so that crossing over the root 1 from the interval $(1, 2)$ to the interval $(-2, 1)$ really means crossing over *two* roots. That is, *two* changes of sign means *no* changes of sign, in effect. And the other factors $(1 + x^2)$, $x + 2$, $x - 2$ do not change sign, and the -3 does not change sign, so since $P(x)$ was *positive* on $(1, 2)$ it is *still* positive on $(-2, 1)$. (The rest of this example is the same as the first example).

Again, the point is that each time a root of the polynomial is *crossed over*, the polynomial changes sign. So if *two* are crossed at once (if there is a double root) then there is really *no* change in sign. If *three* roots are crossed at once, then the effect is to *change* sign.

Generally, if an *even* number of roots are crossed-over, then there is *no* change in sign, while if an *odd* number of roots are crossed-over then there *is* a change in sign.

Exercises

1. Write the equation for the line passing through the two points $(1, 2)$ and $(3, 8)$.

2. Write the equation for the line passing through the two points $(-1, 2)$ and $(3, 8)$.

3. Write the equation for the line passing through the point $(1, 2)$ with slope 3.

4. Write the equation for the line passing through the point $(11, -5)$ with slope -1.

1.4 End of chapter problems

Exercises

1. Two mathematicians (A and B) are taking a walk and chatting.
   A: I have 3 children.
   B: How old are they?
   A: The product of their ages is 36.
   B: I can't figure out how old they are.
   A: The number on the house that we are passing is the sum of their ages.
   B: I still can't figure it out.
   A: My oldest child is having a soccer match tomorrow.
   B: Now I can figure it out!
   How old are the children?

   Make a list of the possible ages whose product is 36.

   The only possibilities for these three ages are $1, 1, 36$; $1, 6, 6$; $1, 2, 18$; $1, 3, 12$; $2, 2, 9$; $2, 3, 6$; $3, 3, 4$. The sums of these ages are 38, 13, 21, 16, 13, 12, 10. If the house number had been any of these sums except 13, mathematician B would have known the ages, so the street number must have been 13. If there is only one oldest child, then $2, 2, 9$ must be the ages.

2. You have 2 identical ropes, a scissors and a box of matches. Each rope, when ignited at one of its ends, burns for 1 hour. Fugure out how to measure off 45 minutes by burning these ropes. Notice: the ropes may be not uniform, so they can burn in starts and stops, not at a constant speed.

   Ignited from both ends sinultaneously, how long will a rope burn?

3. Repeat example **??**, where you replace the leaky cone-shaped bucket with a leaky cylindrical bucket.

   The surface area A will be proportional to H^2, i.e. $A = cH^2$ the equation $dH/dt = -(a/A)v(H) = -(a/A)\sqrt{2gH}$ still holds, stick the expression for A into it and try to work out the rest.

Chapter 2

Limits

2.1 Elementary limits

Discussion.

[author=duckworth,label=discussion_overview_of_limits,style=historical, file =text_files/limits_overview]
Before we begin to learn limits, it might be worth describing how the way we use limits today is the reverse of how they came to be developed historically.

Almost all modern Calculus courses (see exceptions below) start with the definition of limit, and then everything which follows is built upon this definition: vertical and horizontal asymptotes are described using limits; derivatives are defined in terms of limits, as are definite integrals; sequences and series, Taylor polynomials, L'Hospital's rule, all deal directly with limits. So the modern approach is limits first, and then everything else.

But the modern approach reverses the order of history! Newton and Leibnitz invented a lot of what we think of as Calculus, and they never used the concept of limits. In fact, their work was finished around 1700, but it wasn't until around 1850's that limits were carefully and precisely defined. Even then, it took about another 100 years for Calculus books to base everything on limits. (Over this 100 year period the use of limits gradually trickled from more advanced subjects down to a college freshman level Calculus course.)

So, if you find it a little difficult to understand exactly what limits are, how they are used, and why we discuss them so much, don't feel bad! Geniuses like Newton, Leibnitz, Euler, Gauss, Lagrange, the Bernoulli's, etc. didn't understand them either! On the other hand, by now, limits have been re-worked and simplified so much that anyone can use them, but they still take work. The moral: don't feel bad if they don't make sense at first, but don't give up or decide that you just can't get it; keep working hard.

So, are there alternatives to a limits based Calculus? Yes. In the 1960's the infinitesimal approach was put on rigorous grounds, and this made it acceptable to mathematicians to write Calculus books which based their results on this approach. Infinitesimals are very similar to how Leibnitz thought about derivatives and integrals. They involve doing calculation with infinitely small quantities; this is a strange idea and the strangeness of it is why mathematicians didn't feel that it was rigorously justified until the work in the 1960's mentioned above was complete. (For more about this approach and the the one which is described next see the section on further reading.) Interestingly, even in an infinitesimal based Calculus course, limits are still discussed and used, but not as a foundation for everything else.

Another approach to Calculus has been developed recently by a variety of authors. This approach uses division of functions to define derivatives and uses piecewise-linear functions to work with integrals. In this approach, something similar to limits is always lurking in the background; for example, showing that the exact value for the area under a curve is always between two piecewise-linear approximations which can be made infinitely close to each other. However, limits are never explicitly used, and instead one must use various clever calculations of bounds and inequalities.

Perhaps this discussion leaves one more question unanswered: if limits take some hard work to understand, and if there are alternatives, and if historically limits weren't used for the first 150 of Calculus, why do we learn them now? Well, for 150 years, belief in Calculus required a certain amount of faith: mathematicians would make arguments which mentioned things like "divide two infinitely small quantities", or "take the ratio of the quantities just before they become zero" and these arguments were used to justify the formulas they had. So, to prove that a falling object had a certain speed, or to prove that the planets orbited the sun under the influence of gravity, or to prove a hundred other things, one had to refer to these arguments which were not rigorous (although the answers seemed to work). Mathematically speaking there were other problems when Calculus was not rigorous. Maybe the answers obtained in simple cases were correct, but what about when things got more complicated? Maybe we believe, for example, that the derivative of $\sin(x)$ is $\cos(x)$ (we'll learn this later). But what if x is a *complex number*; is this result still true? What if we want to do mathematics in 3 dimensions, or 4, or 100? Here we can't draw pictures, and our intuition breaks down, so we can't claim that "the answers seem to work"; can we still *prove* anything about derivatives? What about strange counter-examples that mathematicians had discovered; this examples showed that our intuition about functions and derivatives can be completely wrong (see below), so how do we know that the simpler problems, whose "answers seemed right" were really right?

Well, the use of limits (with work) answers all these questions. Finally, everything that we think is correct can be rigorously proven; the proofs work for complex numbers and for mathematics in 100 dimensions. We can see exactly which part of the proof fails in various counter-examples, and we can prove that some things in math are correct, but defy our intuition. Finally, the best part is that once we learn limits we can go far beyond the pictures and numerical arguments in Freshman Calculus. We can do Calculus in infinite-dimensional space, we can prove statements in the space-time universe of Einstein's General Theory of Relativity, we can prove things about geometric spaces that exist beyond anyone's ability to intuitively understand.

What are some of the counter-examples I mentioned above? Well, to fully understand them you have to first understand essentially all of Freshman Calculus, but I can give you some idea of what they are about here.

Here's an intuitive idea: functions are mostly differentiable, and the derivative can only fail to exist at a handful of points. For example, the absolute value function $y = |x|$ is differentiable everywhere except at $x = 0$, where it has a corner. Well, we could make a worse example which has a bunch of corners, but still, most of the points don't have corners, right? Wrong. Weierstrass showed that there exist continuous functions which have infinitely many corners, and in fact the corners are infinitely close to each other! Can this be true? Well, picture a line which zig-zags up and down, and then imagine that if you magnify the picture, that there are more zig-zags that were too small for you to see before; and if you magnify the picture again, there are more zig-zags, etc. This example is correct; to prove that it is correct, you need to understand limits and continuity; but more importantly, it shows that you cannot rely on intuition to say things like "it's clear that a continuous function is differentiable".

Here's another idea: integrals are calculated by finding anti-derivatives. For example, the area between the curve $y = x^2$ and $x = 0$ and $x = 1$ is calculated by finding the anti-derivative $\frac{1}{3}x^3$, and plugging in $x = 1$ and $x = 0$ to get the area of $\frac{1}{3}$. But what about the function e^{-x^2}? Does that have an anti-derivative? Well, it turns out that there is no *formula* for the anti-derivative of e^{-x^2}. So, how can we calculate areas under this curve? Well, with *limits* we can *define* the integral $\int_0^b e^{-x^2}\,dx$ and we can show that the limit exists, and therefore the integral exists, even though we cannot write down a formula for it.

Discussion.

[`author=duckworth,label=overview_of_limits,style=middle`, file =`text_files/limits_overview`]
This section picks two problems to act as guiding examples for the rest of the chapter: finding the slope of a tangent line and finding the instantaneous velocity. In both cases we look at a fraction as the bottom gets smaller and smaller. (Using later notation we could say that we were approximating $\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}$, and $\lim_{\Delta t\to 0}\frac{\Delta d}{\Delta t}$.)

Example 2.1.1.

[`author=duckworth,label=example_glimpse_of_deriv_as_limit`, file =`text_files/limits_overview`]
Here's a brief glimpse of something that's coming later. We show it now because it's so important; in fact, it's the whole reason we introduced limits now! Let $f(x) = x^2$. Then the derivative of $f(x)$ at $x = 3$ will be defined (later) to be

$$f'(x) = \lim x \to 3\frac{f(x) - f(3)}{x - 3}$$

. We *interpret* the derivative to be the slope of the tangent line at $x = a$, or the instantaneous velocity.

Example 2.1.2.

[`author=duckworth,label=example_naive_approach_to_inst_vel`, file =`text_files/limits_overview`]
It can be shown through experiments (as Galileo did in the early 1600's) that an object thrown off of a building of height 100 m and with an intial velocity of 23 m/s has a position given by the formula

$$h(t) = -9.8t^2 + 23t + 100.$$

Find (without using derivatives yet) the velocity of such an object at $t = 4$.

The point of this exercise is to see the steps that we're about to do as leading to the idea of limits, which we'll define in the next section, and that limits will allow us to define derivatives.

The definition of average velocity is $\frac{\Delta \mathrm{h}}{\Delta \mathrm{t}}$ where Δh is the change in height h and Δt is the change in time t. The probleb is that the example did not tell us to find the velocity from $t = 4$ to, say, $t = 6$. We were given only one point in time, and so Δt appears to be 0. We can't plug 0 into our definition of velocity or we would be dividing by zero.

The elementary way out of this dilemma, is to find the average velocity from $t = 4$ to $t = 4.1$, and figure that this is pretty close to the instantaneous velocity at $t = 4$. We have:

$$\text{velocity from } t = 4 \text{ to } t = 4.1 = \frac{h(4.1) - h(4)}{4.1 - 4} = \frac{-7.938}{.1} = -79.38\,\mathrm{m}s$$

(I've done all the calculations in my calculator using $y_1 = -9.8x^2 + 23x + 100$ and enterying $y_1(4.1)$, etc.) Now, this answer is probably pretty close to the correct value. But, to make sure, we should probably compute a few more velocities over shorter intervals of time; this should get closer to the correct answer at $t = 4$.

$$\begin{array}{ll} t = 4 \text{ to } t = 4.01 & \frac{h(4.01)-h(4)}{4.01-4} = -78.498 \\ t = 3.999 \text{ to } t = 4 & \frac{h(4)-h(3.999)}{4-3.999} = -78.3902 \end{array}$$

This makes it pretty clear that the "real" answer should be somewhere around $-78\,\mathrm{m}s$. We can't be sure how accurate our calculations are until we learn later how to get the exact expression.

The idea of limit will be to take the calculations just done, and try to figure out what the limit as t approaches 4 of the velocity function $\frac{h(t)-h(4)}{t-4}$ is.

Discussion.

[author=duckworth,label=discussion_looking_forward_to_derivatives, file =text_files/limits_overview]
Looking ahead to the chapter on derivatives: Once we decide that formulas for derivatives are more useful than finding the derivitive at lots of randomly chosen numbers, we want to get a list of shortcuts. To prove that these shortcuts are correct we need to use the long definition given above. But we only have to do that once for each shortcut and then we will always use the shortcut.

Discussion.

[author=wikibooks,label=discussion_introducing_section_on_limits, file =text_files/basic_limits]
Now that we have done a review of functions, we come to the central idea of calculus, the concept of limit.

Example 2.1.3.

[author=wikibooks,label=example_removable_discontinuity_leading_to_limits, file =text_files/basic_limits]

Let's start with a function, $f(x) = x^2$. Now we know that $f(2) = 4$. But let's be a bit mischevious and create a gap at 2. We can do this by creating the function

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

Now this truly is a mischevious function. It's equal to x^2 everywhere except at $x = 2$, where it has no well-defined value. Now, one fact about the funny function is that as x gets closer to 2, then $f(x)$ gets closer to 4. This is a useful fact, and we can express this in symbols as

$$\lim_{x \to 2} f(x) = 4.$$

Notice it doesn't matter what $f(x)$ is at $x = 2$, in this case we have left it undefined, but it could be 2 or 15 or $1,000,000$. The idea of the limit is that that you can talk about how a function behaves as it gets closer and closer to a value, without talking about how it behaves at that value. Now using variables we can say that L is the "limit' of the function $f(x)$ as x approaches c if $f(x) \approx L$ whenever $x \approx c$.

Definition 2.1.1.

[author=duckworth,label=definition_of_a_limit,style=informal, file =text_files/basic_limits]
The notation $\lim_{x \to a} f(x) = L$ means any of the following equivalent statements (choose whichever one makes the most sense to you):

1. If $\left\{ \begin{array}{l} x \text{ is close} \\ \text{to } a (\text{but } \neq a) \end{array} \right.$ then $\left\{ \begin{array}{l} f(x) \text{ is close} \\ \text{to } L \end{array} \right.$
2. If $x \neq a$ and $|x - a|$ is small then $|f(x) - L|$ is small
3. If $x \neq a$ and $|x - a| < \delta$ then $|f(x) - L| < \epsilon$ where ϵ can be chosen as small as we want.
4. If $x \neq a$ and $a - \delta < x < a + \delta$ then $L - \epsilon < f(x) < L + \epsilon$.

We also have variations on this definition if $x \to a^+$ (i.e. x approaches a from the right), $x \to a^-$ (i.e. x approaches a from the left), $a = \pm\infty$ (i.e. we are finding horizontal asymptotes) or $L = \pm\infty$ (i.e. we are finding vertical asymptotes).

Strategy.

[author=duckworth,label=strategy_how_we_find_limits, file =text_files/basic_limits]
We can find a limit in one of the following ways.

1. Graph $f(x)$, look at y-values as x gets close to a.
2. Make a table of numbers for x and $f(x)$ as x gets close to a and look for the pattern of y-values.
3. Simplify $f(x)$, if neccessary, and plug in $x = a$. (I call this the algebraic approach).

Once you've found the limit, you still might be asked to verify that it satisfies the definition. In particular, you might be given $f(x)$, L, a and ϵ and asked to find δ. Essentially, you do this graphically as follows: find the closest x-value corresponding to $y = L \pm \epsilon$ and δ is the distance from this x-value to $x = a$.

Discussion.

[author=wikibooks,label=discussion_of_limits_after_definition, file =text_files/basic_limits]
Now this idea of talking about a function as it approaches something was a major breakthrough, because it lets us talk about things that we couldn't before. For example, consider the function $1/x$. As x gets very big, $1/x$ gets very small. In fact $1/x$ gets closer and closer to zero, the bigger x gets. Now without limits its very difficult to talk about this fact, because $1/x$ never actually gets to zero. But the language of limits exists precisely to let us talk about the behavior of a function as it approaches something, without caring about the fact that it will never get there. So we can say

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

Notice that we could use "=" instead of saying "close to". Saying that the limit *equals* 0 already means that $1/x$ is *close to*.

Exercises

1. Find $\lim_{x\to5} 2x^2 - 3x + 4$.
2. Find $\lim_{x\to2} \frac{x+1}{x^2+3}$.
3. Find $\lim_{x\to1} \sqrt{x+1}$.

2.2 Formal limits

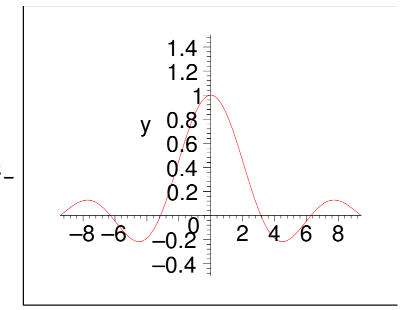
Discussion.

[author=wikibooks,label=discussion_intro_to_formal_limits, file =text_files/formal_limits]
In preliminary calculus, the definition of a limit is probably the most difficult concept to grasp. If nothing else, it took some of the most brilliant mathematicians 150 years to arrive at it.

The intuitive definition of a limit is adequate in most cases, as the limit of a function is the function of the limit. But what is our meaning of "close"? How close is close? We consider this question with the aid of an example.

Example 2.2.1.

[author=duckworth,label=example_limit_sin_over_x,uses=sin,uses=limits, file =text_files/formal_limits]
Consider the function $f(x) = \frac{\sin(x)}{x}$. What happens to $f(x)$ as x gets close to 0? Well, if you try to plug $x = 0$ in, you get $f(0) = 0/0$, and this is undefined. But if you graph the function you figure 2.1. It seems clear that the y-value "at" (or near) $x = 0$ should be 1.



sin_over_x

Figure 2.1:

How do we convert that intuition into a rigorous statement? What do I mean by "rigorous statement"? Well, we need a statement that *doesn't* depend on looking at graphs. Why can we not depend on graphs? Well, we need to be able to find limits of functions like x^n, without *knowing what n is*! So if we don't know n, how can we graph the function? Also, we need a statement that will work for other kinds of limits, like those we will use when we define definite integrals, and like those we will use when we do calculus in three (or higher) dimensions, where we can't rely on a graph. Finally, sometimes graphs, can be misleading, or even wrong. See Section 1.3 for examples of this.

Discussion.

[author=duckworth,label=discussion_limit_means_infinitely_close,uses=sin,uses=limits, file =text_files/formal_limits]
So, to say, for example, that $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$, how close does $\frac{\sin(x)}{x}$ have to get to 1? Infinitely close. In mathematics, we usually want answers that are exactly correct, not just "close enough" (actually, there are many parts of math where "close enough" is of interest, but if it's possible, then exactly right is always better). So, how can we define infinitely close? The human brain doesn't deal well with "infinite" statements. So in fact, we translate infinite statements into finite ones.

A first attempt at this might give something like $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$ means that $\lim_{x\to 0} \frac{\sin(x)}{x}$ is closer to 1 than any other real number. This attempt has the problem that it's circular, we explained what "$\lim_{x\to 0} \frac{\sin(x)}{x}$" means by talking about "$\lim_{x\to 0} \frac{\sin(x)}{x}$" itself.

No, we need to describe what this limit means by refering only to $\frac{\sin(x)}{x}$. What should this be doing? It should be close to 1. How close? Infinitely close. How can

I state this using only "finite" concepts? By saying something like the following: "for every distance you want to pick, $\frac{\sin(x)}{x}$ will get at least that close to 1". The formal definition of limit merely names "distance" with the letter ϵ.

Definition 2.2.1.

[author=wikibooks,label=definition_of_limit_formal,style=formal,uses=limits, file =text_files/formal_limits]
Let $f(x)$ be a function. We write

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

if for every number ϵ, there exists a number δ such that $|x - a| < \delta$ and $x \neq a$ implies that $|f(x) - L| < \epsilon$.

Comment.

[author=wikibooks,label=comment_about_what_limit_definition_means, file =text_files/formal_limits]
Note that instead of saying $f(x)$ approximately equals L, the formal definition says that the difference between $f(x)$ and L is less than any number epsilon.

Definition 2.2.2.

[author=duckworth,label=defintion_of_one_sided_limits,uses=limits,style=formal, file =text_files/formal_limits]
Let $f(x)$ be a function. We write

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

if for every number ϵ, there exists a number δ such that $a < x < a+\delta$ implies that $|f(x) - L| < \epsilon$. We write

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

if for every number ϵ, there exists a number δ such that $a-\delta < x < a$ implies that $|f(x) - L| < \epsilon$.

Comment.

[author=wikibooks,label=comment_how_to_read_one_sided_limits, file =text_files/formal_limits]
We read $\lim_{x\to a^-} f(x)$ as the limit of $f(x)$ as x approaches a from the left, and $\lim_{x\to a^+} f(x)$ as x approaches a from the right.

Fact.

[author=wikibooks,label=fact_limit_implies_equality_of_two_sided_limits, file =text_files/formal_limits]

$\lim_{x\to a} f(x) = L$ if and only if $\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) = L$

Example 2.2.2.

[author=wikibooks,label=example_find_limit_of_x+7,style=formal, file =text_files/formal_limits]
What is the limit of $f(x) = x + 7$ as x approaches 4?

There are two steps to answering such a question first we must determine the answer – this is where intuition and guessing is useful, as well as the informal definition of a limit. Then, we must prove that the answer is right. For this problem, the answer happens to be 11. Now, we must prove it using the definition of a limit

Informally, 11 is the limit because when x is close to 4, then $f(x) = x + 7$ is close to $4 + 7$, which equals 11.

Here's the formal approach. (Note: please keep in mind that this example is to *practice* the formal approach; this example is so simple that you might feel that there is no need for the formal approach, but we will need it later to prove more general statements.) We need to prove that no matter what value of ϵ is given to us, we can find a value of δ such that $|f(x) - 11| < \epsilon$ whenever $|x - 4| < \delta$.

For this particular problem, letting δ equal ϵ works. (We'll talk more later about how to pick δ in different problems.) Now, we have to prove $|f(x) - 11| < \epsilon$ given that $|x - 4| < \delta = \epsilon$. Since $|x - 4| < \epsilon$, we know $|f(x) - 11| = |x + 7 - 11| = |x - 4| < \epsilon$, which is what we wished to prove.

Example 2.2.3.

[author=duckworth,label=example_lengthy_calculation_of_easy_limit, file =text_files/formal_limits]
Suppose we want to look at the definition of the limit as applied to the function $y = 2x + 1$. If we want to show that it's continuous (which it certainly should be judging from the graph) we would like to show that as x gets very close to a what happens to the y-values is what we would expect, that $y(x)$ gets very close to $y(a)$. By the way, one hard part about all this is that it's hard to say exactly how*fast* these y-values should be getting close to $y(a)$. On the other hand we don't care if this is happening really quickly or not, as long as the y-values do what they're supposed to eventually. Another way to say all this is that we can make $y(x)$ as close to $y(a)$ as we want by making x close enough to a. **procedure** (1) **FIRST** we decide how close we want to make $y(x)$ to $y(a)$.

(2) **THEN** we figure out how close we need x to be to a to accomplish (1)

For example, ouppose that I want to make sure that we'll get $|y(x) - y(a)| < .0000001 = 10^{-6}$ when we get close enough to $x = a$. Well, after some guesswork and some graphing we might figure out that this will always happen if we just look at x's which satisfy $|x - a| < 10^{-7}$. By the way, this was certainly overkill; we don't absolutely **need** for x to be this close to a, it just makes it easy.

But how do we do this in general? What if someone asked me to make $y(x)$ a million times closer to $y(a)$? I would like an argument which will help me out no matter how close I'm supposed to get to $y(a)$.

Let $\epsilon > 0$ be some (small) number, and suppose that we want to get our y values to a distance within ϵ of $2a+1$. In other words we want $|2x+1-(2a+1)| < \epsilon$ when we make x close enough to a. How close do we need x to be to a?

We want

$$\begin{aligned} |2x+1-(2a+1)| &< \epsilon \\ |2x-2a| &< \epsilon \\ |2(x-a)| &< \epsilon \\ 2|(x-a)| &< \epsilon \\ |(x-a)| &< \epsilon/2 \end{aligned}$$

Ah ha!! Whatever ϵ is we can guarantee that $|y(x)-y(a)| < \epsilon$ if we pick x's with $|x-a| < \epsilon/2$.

We have just proven that $\lim_{x\to a} 2x+1 = 2a+1$. Note that this applies to stuff we did before with the difference quotient where we simplified an expression down to something like $\lim_{h\to 0} 2h+1 = 1$

Example 2.2.4.

[author=wikibooks,label=example_limit_of_x^2,style=formal, file =text_files/formal_limits]

What is the limit of $f(x) = x^2$ as x approaches 4?

Informal reasoning suggests that the limit should be 16. Again, we'll try to prove this formally.

Let ϵ be any positive number. Define δ to be $\delta = \sqrt{(\epsilon+16)} - 4$. Note that δ is always positive for positive ϵ. Now, we have to prove $|x^2-16| < \epsilon$ given that $|x-4| < \delta = \sqrt{\epsilon+16} - 4$.

We know that $|x+4| = |(x-4)+8| \le |x-4|+8 < \delta+8$ (because of the triangle inequality), thus

$$\begin{aligned} |x^2-16| &= |x-4|\cdot|x+4| \\ &< \delta\cdot(\delta+8) \\ &= (\sqrt{16+\epsilon}-4)\cdot(\sqrt{16+\epsilon}+4). \\ &= (\sqrt{16+\epsilon})^2-4^2 \\ &= \epsilon \end{aligned}$$

Example 2.2.5.

[author=wikibooks,label=example_limit_of_sin_of_1_over_x_dne,style=formal, file =text_files/formal_limits]
Show that the limit of $\sin(1/x)$ as x approaches 0 does not exist.

We will proceed by contradiction, thus, suppose the limit exists and is L. We show first that $L \neq 1$ is a contradiction, the case $L = 1$ is similar. Choose $\epsilon = L-1$, then for every $\delta > 0$, there exists a large enough n such that $0 < x_0 = \frac{1}{\pi/2+2\pi n} < \delta$, but $|sin(1/x_0) - l| = |L-l| = \epsilon$ a contradiction.

The function $\sin(1/x)$ is known as the **topologist's sine curve**.

Example 2.2.6.

[author=wikibooks,label=example_limit_x_times_sin_1_over_x, file =text_files/formal_limits]
What is the limit of $x \sin(1/x)$ as x approaches 0?

We will prove that the limit is 0. For every $\epsilon > 0$, choose $\delta = \epsilon$ so that for all x, if $0 < |x| < \delta$, then $|x \sin x - 0| <= |x| < \epsilon$ as required.

2.3 Foundations of the real numbers

Discussion.

[author=duckworth,label=discussion_logical_foundations, file =text_files/foundations_of_reals]
In this section we present the logical foundations of calculus. We note that it is possible to study calculus without first studying these logical foundations. The advantage of such a study is that it leads immediately to "doing" calculus, to applications, and it does not bewilder the beginning student with the harder work required for mathematical rigor. However, skipping the foundations also skips learning *why* calculus works the way mathematicians say it does, it skips the chance to stretch your mind and exercise your deductive reasoning, and it skips developing the skills and techniques needed to study higher mathematics (like calculus in n-dimensions, differential geometry, theoretical physics, etc).

In rigorous mathematics *everything* starts with axioms. Axioms are simple statements, that are hopefully somewhat intuitive, and which one accepts as logically true if one wants to continue with calculus (if you want to debate the axioms, that's worth studying too, but then you are doing logic, or metamathematics, or model theory, but not calculus).

After the axioms, the first assertions are proven using only the axioms. By proof we mean a finite set of logical steps, each of which can be justified completely, which start with the axioms and which finish with the assertion to be proven. Finally, later assertions are proven using the first assertions or the axioms.

There's one more ingrediant in rigorous mathematics: definitions. Logically, definitions play no essential role; the only important things are axioms, assertions, and deductive proofs. But practically speaking definitions are crucial for the way we think about things: essentially, they just give names to certain properties, formulas, or statements. So, logically, we wouldn't have to define "continuous", we could merely repeat it's definition in every assertion that used the property of "continuous". Of course, in practice such a text would be unreadable.

So, the ingrediants of a rigorous approach to calculus (or any mathematical subject) are: axioms, assertions, deductive proofs, definitions.

Discussion.

[author=wikibooks,label=discussion_recalling_basic_axioms_of_reals, file =text_files/foundations_of_reals]
Recall that we have already assumed certain basic properties of the real numbers (see Section 1.1, Axioms 1.1 and 1.1). The real numbers have addition, multiplication, and a relation $\leq$

Definition 2.3.1.

[author=wikibooks,label=definition_upper_lower_bounds, file =text_files/foundations_of_reals]
A subset E of the real numbers $\mathbb{R}$ is **bounded above** if there exists a number M which is $\geq$ every number in E. Any M which satisfies this condition is called an **upper bound** of the set E. We say that M is the **least upper bound** if it is the smallest number which is an upper bound of E.

Similarly, E is bounded below of there exists a number M which is $\leq$ every number in E.

Least Upper bound axiom 2.3.1.

[author=wikibooks,label=axiom_least_upper_bound, file =text_files/foundations_of_reals]
Every non-empty set E of real numbers which is bounded above has a least upper bound in $\mathbb{R}$.

Comment.

[author=duckworth,label=comment_on_least_upper_bound_axiom, file =text_files/foundations_of_reals]
The least upper bound axiom is the most subtle axiom in all of Calculus (and in a lot of other mathematics for that matter!). This axiom is what distinguishes the real numbers (which satisfy the axiom) from the rational numbers (which do not). Historically, it was this axiom which gave the first rigorous approach to the real numbers. One way to think about what this axiom means, is that the real number line does not have any holes. Because if it had a hole, it would have to be infinitely small (since the real number line contains $\mathbb{Q}$), and then you could let E be set of all real numbers to the left of this hole. The axiom would then say that the real numbers contain a least upper bound of E; this least upper bound would have to be the number where the hole was located!

Of course an axiom is assumed, so it's not immediately clear how this axiom would contribute to a rigorous study. Here's how: before this, people made all kinds of assertions about what "continuous" meant, what "convergent" meant, what was different between the real numbers and the rational numbers. Some of these assertions were "clear", some were complicated, all appeared a little different, and actually most were not even clearly articulated, but rather implicitly used without specefic mention. In contrast, the least upper bound axiom (after you look at a few pictures) is fairly clear and starting with it you can derive all the various other assertions people made. So at the very least you've replaced a variety of implicit assumptions, with one, clear assertion.

Now, if you still don't like this axiom, that's ok. You *can* try to develop a theory of calculus without it, and you can see how far you get. Seriously, that would be a fun exercise. But, if you want, you can simply preface all the statements later in calculus with the invisible statement "*If* the least upper bound axiom holds, then . . . " where ". . . " might be some rule about limits, or some rule about derivatives, or some rule about max and mins of a function. In this way, all the statements which follow are *hypothetical* statements, which are logically perfect, and then one can debate if they are "really" true, which is to say, does the least upper bound axiom "really" hold!

Comment.

[author=duckworth,label=comment_that_lub_implies_glb, file =text_files/foundations_of_reals]
The least upper bound axiom is not symmetric, in that it talks only about upper bounds and not lower ones. However, the real numbers are quite symmetric, and multiplying by -1 turns lower bounds into upper bounds and vice versa. The following theorem makes this more precise.

Theorem 2.3.1.
[author=wikibooks, label=theroem_ existence_ of_ glb, file =text_files/foundations_of_reals]
Every non-empty set of real numbers which is bounded below has an greatest lower bound.

Proof.
[author=duckworth,label=proof_that_existence_of_lub_implies_glb, file =text_files/foundations_of_reals]
Let E be a non-empty set of of real numbers which is bounded below. Then $-E$ is bounded above (check this assertion). Let M be a least upper bound for $-E$. Then $-M$ is a greatest lower bound for E (check this assertion). □

Notation.

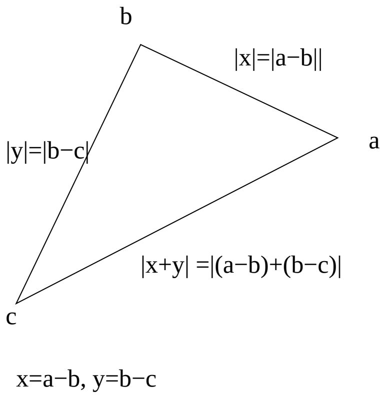
[author=duckworth,label=notation_for_glb_and_lub, file =text_files/foundations_of_reals]
Let E be a nonempty subset of the real numbers.

The greatest lower bound of E is denoted by $\inf E$ ("inf stands for the Latin word *infimum* which was used historically in this context).

The least upper bound of E is denoted by $\sup E$ ("sup" stands for the Latin word *supremum* which was used historically in this context).

Lemma 2.3.1.
[author=wikibooks, label=lemma_ facts_ about_ infs_ and_ sups_ and_ subsets, file =text_files/foundations_of_reals]
Let A and B be two nonempty subsets of the real numbers. The following hold:

1. $A \subseteq B \Rightarrow \sup A \leq \sup B$
2. $A \subseteq B \Rightarrow \inf A \geq \inf B$



`triangle_inequality`

Figure 2.2:

3. $\sup A \cup B = \max(\sup A, \sup B)$

4. $\inf A \cup B = \min(\inf A, \inf B)$

Triangle Inequality 2.3.2.
[author= `duckworth`*,* file =`text_files/foundations_of_reals`*]*
For all real numbers x, y we have $|x+y| \leq |x| + |y|$. This inequality can be picture as in figure **??**.

Comment.

[`author=duckworth`, file =`text_files/foundations_of_reals`]
The previous lemma is called the triangle inequality because it can be pictured thus: sides

---

Proof.
[`author=duckworth`, file =`text_files/foundations_of_reals`]
Case 1: x and y are positive. Then $|x+y| = x+y$ and $|x|+|y| = x+y$.

Case 2: x is positive, y is negative, and $x+y$ is positive. Then $|x+y| = x+y$ and $|x|+|y| = x-y$. Now we calculate:

$$\begin{aligned} x+y \leq x-y &\iff y \leq -y \\ &\iff 2y \leq 0 \\ &\iff y \leq 0 \\ &\quad\ \text{which is true} \end{aligned}$$

The other cases are similar. □

2.4 Continuity

Discussion.

[`author=garrett,label=discussion_of_limits_as_usually_easy`, file =`text_files/continuity`]
The idea of **limit** is intended to be merely a slight extension of our *intuition*. The so-called ε, δ-definition was invented after people had been doing calculus for hundreds of years, in response to certain relatively pathological technical difficulties. For quite a while, we will be entirely concerned with situations in which we can either 'directly' see the value of a limit *by plugging the limit value in*, or where we *transform* the expression into one where we *can* just plug in.

So long as we are dealing with functions no more complicated than polynomials, most *limits* are easy to understand: for example,

$$\lim_{x \to 3} 4x^2 + 3x - 7 = 4 \cdot (3)^2 + 3 \cdot (3) - 7 = 38$$

$$\lim_{x \to 3} \frac{4x^2 + 3x - 7}{2 - x^2} = \frac{4 \cdot (3)^2 + 3 \cdot (3) - 7}{2 - (3)^2} = \frac{38}{-7}$$

The point is that we just substituted the '3' in and *nothing bad happened*. This is the way people evaluated easy limits for hundreds of years, and should always be the first thing a person does, just to see what happens.

Definition 2.4.1.

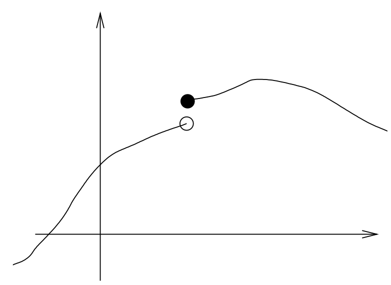
[author=wikibooks,label=definition_of_continuity_at_point, file =text_files/continuity]
We say that $f(x)$ is at c if $\lim_{x \to c} f(x) = f(c)$.

Discussion.

[author=duckworth,label=discussion_of_continuous_definition, file =text_files/continuity]
The definition of continuous is a technical version of something that is supposed to be intuitive. This is not done to make an easy thing seem hard. Rather, it is done so that results can be rigorously proven. In fact, in every technical field it is common to take an intuitive idea, often an idea that that exists outside of the field, and translate it into a technical statement that can be used within the field.

Here's two intuitive translations of the definition continuity:

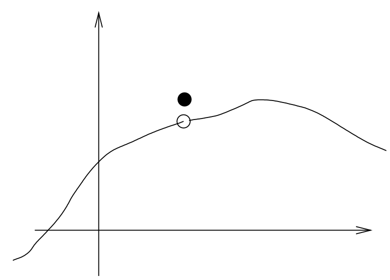
1. To take the limit, just plug the number in.
2. What you get when you plug the number c in is what you get for numbers near c.



jump_discontinuity

The main intuitive ideas of continuity that this definition is supposed to capture are these:

1. Continuous should mean that there are no holes in the graph. If you think about it there are two types of holes, and in both cases, what is happening at the number $x = c$ is different than what is happening near the number.
2. The slope between $x = c$, $y = f(c)$ and an other point on the curve of $f(x)$ is bounded, i.e. the absolute value of this slope does not become infinitely large. The same pictures we drew showing holes in a discontinuos function should also show you that the slopes become infinite. It will take us some time to prove that the slope is bounded for a continous function.



removable_discontinuity

Discussion.

[author=livshits,label=discussion_of_continuity_as_bounded_accuracy, file =text_files/continuity]
Here is a real-life way to think about continuity. How much accuracy do we need in x in order to get a certain accuracy in $f(x)$?" Or, to put it more precisely, how many accurate decimal places in x do we need to get a certain number of accurate decimal places in $f(x)$? You can view continuity as saying that it's possible to get a nice function which relates the accuracy of $f(x)$ to the accuracy of x.

Here are a few examples.

1. $f(x) = 100x + 7$, then by taking $n + 2$ accurate decimal places in x we get n accurate decimal places in $f(x)$, no matter what x is.
2. $f(x) = x^2$ and assume $|2x| < 10^k$, then by taking $n + k$ accurate decimal places in x we get n accurate decimal places in $f(x)$.

3. $f(x) = \sqrt(x)$, $x = 0$, then we need $2n$ accurate decimal places in x to get n accurate decimal places in $f(x)$, and it will work for $x > 0$ as well.

4. $f(x) = 1/x$ and $|x| > 10^{-k}$, then we can get n accurate decimal places in $f(x)$ by taking $n + 2k$ accurate decimal places in x.

5. $f(x) = sin(x)$, then we can get n accurate places in $f(x)$ by taking n accurate places in x.

The examples above suggest that as long as x stays away from the "bad" values (such as $x = 0$ for $f(x) = 1/x$) and from infinity (which means that there is an estimate of the form $|x| < A$, like in example 2), we can answer the question in a satisfactory manner. In othes words, given n, we can, by taking enough (but still a finite number) of accurate decimal places in x get n accurate decimal places in $f(x)$.

Definition 2.4.2.

[author=duckworth,label=definition_of_continuous_on_interval, file =text_files/continuity]
We say that a function $f(x)$ is continous on an interval $[a, b]$ if it is continous at each number c in the interval.

Discussion.

[author=livshits, file =text_files/continuity]
Actually there are two brands of continuity.

If we fix x first and then worry about the question, we get continuity at this particular x.

If we consider the whole range of values for x and then worry about the question (?n), we get the uniform continuity (for this particular range of values of x).

This brand of continuity is more important for practical purposes.

There is a theorem by E. Heine (1872) that says that if a function f is continuous at every x such that $a \leq x \leq b$, then f is uniformly continuous on the whole closed interval $[a, b]$ (which is the set of numbers x such that $a \leq x \leq b$).

This theorem becomes wrong if we replace one of the $\leq$ signs with the $<$ sign. We can understand why by inspecting in more detail the function $1/x$ from example 4. It is continuous at every x of the interval $(0, 1]$, but not uniformly continuous on this interval

We will mostly deal with continuous functions on closed intervals and by continuity will mean the uniform continuity. Continuity at a given point will be less important. In fact the whole notion of a given point becomes problematic when we deal only with the finite accuracy approximations, but it is still handy for the theory.

Discussion.

[author=livshits, file =text_files/continuity]
Continuous functions are rather reasonable, in particular, *continuous functions*

don't jump, in other words, *if f is a continuous function defined on an interval (a, b) and $f(x) = 0$ for all $x \neq c$ then $f(c) = 0$ too.*

Indeed, there is $d \neq c$ such that $a < d < b$ and $d - c$ is as small as we want, but $f(d) = 0$, therefore $f(c) \approx 0$ with any accuracy we want, therefore we must have $f(c) = 0$.

The following properties of continuous functions are immediate.

1. A sum of two continuous functions is continous.
2. A constant multiple of a continuous function is continuous.

It follows that our approach to differentiation (see section 2.1) works for continuous functions, i.e. the rule that $f'(a)$ is $(f(x) - f(a))/(x - a)$ evaluated at $x = a$ defines $f'(a)$ unambiguously if the division is carried out in the class of continuous functions.

It follows from the observation that any 2 continuous functions g and h such that $(x - a)(g(x) - h(x)) = 0$ must be equal because they are equal for $x \neq a$ as well as for $x = a$ ($g - h$ can't jump).

Discussion.

[`author=wikibooks`, file =`text_files/finding_limits`]
Now we will concentrate on finding limits, rather than proving them. In the proofs above, we started off with the value of the limit. How did we find it to even begin our proofs?

First, if the function is continuous at a particular point c, that the limit is simply the value of the function at c, due to the definition of continuity. All polynomial, trigonometric, logarithmic, and exponential functions are continuous over their domains.

If the function is not continuous at c, then in many cases (as with rational functions) the function is continuous all around it, but there is a discontinuity at that isolated point. In that case, we want to find a similar function, except with the hole is filled in. The limit of this function at c will be the same, as can be seen from the definition of a limit. The function is the same as the previous except at a point c. The limit definition depends on f(x) only at the points where $0 < |x - c| < \delta$. When $x = c$, that inequality is false, and so the limit at c does not depend on the value of the function at c. Therefore, the limit is the same. And since our new function in continuous, we can now just evaluate the function at c as before.

Lastly, note that the limit might not exist at all. There are a number of ways that this can occur There a is gap (more than a point wide) in the function where the function is not defined.

Example 2.4.1.

[`author=wikibooks`, file =`text_files/finding_limits`]
As an example, in

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

f(x) does not have any limit when $-4 \leq x \leq 4$. There is no way to "approach" the middle of the graph. Note also that the function also has no limit at the endpoints of the two curves generated (at x=-4 and x=4). For the limit to exist, the point must be approachable from both the left and the right. Note also that there is no limit at a totally isolated point on the graph.

Discussion.

[author=wikibooks, file =text_files/finding_limits]
Let's take a closer look at different types of discontinuities.

Jump discontinuities It follows from the previous discussion that if the graph suddenly jumps to a different level (creating a discontinuity, where the function is not continuous), there is no limit. This is illustrated in the floor function (in which the output value is the greatest integer not greater than the input value).

Asymptotic discontinuities In $f(x) = \frac{1}{x^2}$ the graph gets arbitrarily high as x approaches 0. There is no limit.

Infinite Oscillation These next two can be tricky to visualize. In this one, we mean that a graph continually rises above and below a horizontal line. In fact, it does this infinitely often as you approach a certain x-value. This often means that there is no limit, as the graph never homes in on a particular value. However, if the height (and depth) of each oscillation diminishes as the graph approaches the x-value, so that the oscillations get arbitrarily smaller, then there might actually be a limit.

The use of oscillation naturally calls to mind trigonometric functions. And, indeed, a simply-defined example of this kind of nonlimit is $f(x) = \sin 1x$.

In the plain old sine function, there are an infinite number of waves as the graph heads out to infinity. The 1/x takes everything that in $(1, \infty)$ and squeezes it into (0, 1). There we have it infinite oscillation over a finite interval of the graph.

Incomplete graph Let us consider two examples. First, let f be the constant function $f(q) = 2$ defined for some arbitrary number q. Let q_0 be an arbitrary value for q.

We can show that f is continuous at q_0. Let $\delta > 0$ then if we pick any $\epsilon > 0$, then whenever q is a real number within ϵ of q_0, we have $|f(q_0) - f(q)| = |2 - 2| = 0 < \delta$. So f is indeed continuous at q_0.

Now let g be the similar-looking function defined on the entire real line, but we change the value of the function based on whether q is rational or not.

$$g(q) = \begin{cases} 2, & \text{if } q \text{ is rational} \\ 0, & \text{if } q \text{ is irrational} \end{cases}$$

Now g is continuous nowhere! For let x be a real number we show that g isn't continuous at x. Let $\delta = 2$ then if g were continuous at x, there'd be a number ϵ such that whenever y was a real number at distance less than ϵ, we'd have $|g(x) - g(y)| < 1$. But no matter how small we make ϵ we can find a number y within ϵ of x such that $|g(x) - g(y)| = 2$ for if x is rational, just pick y irrational and if x is irrational, pick x rational. Thus g fails to be continuous at every real number!

Discussion.

[author=garrett, file =text_files/limits_cancellation]
But sometimes things 'blow up' when the limit number is substituted:

$$\lim_{x\to 3} \frac{x^2-9}{x-3} = \frac{0}{0} \quad ?????$$

Ick. This is not good. However, in this example, as in *many* examples, doing a bit of simplifying algebra first gets rid of the factors in the numerator and denominator which cause them to vanish:

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

Here at the very end we *did* just plug in, after all.

The lesson here is that some of those darn algebra tricks ('identities') are helpful, after all. If you have a 'bad' limit, *always* look for some *cancellation* of factors in the numerator and denominator.

In fact, for hundreds of years people *only* evaluated limits in this style! After all, human beings can't really execute infinite limiting processes, and so on.

Definition 2.4.3.

[author=wikibooks, file =text_files/discontinuities]
A **discontinuity** is a point where a function is not continuous. The discontinuity is said to be removable if we can define or redefine a single value of the function to make it continuous.

Example 2.4.2.

[author=wikibooks, file =text_files/discontinuities]
For example, the function $f(x) = \frac{x^2-9}{x-3}$ is considered to have a "removable discontinuity' at $x = 3$.

In particular, we can divide the function to get $f(x) = x + 3$, except at $x = 3$. If we let f(x) be 6 at that point, we will get a continuous function

$$g(x) = \begin{cases} x+3, & \text{if } x \neq 3 \\ 6, & \text{if } x = 3 \end{cases}$$

But $x + 3 = 6$ for x = 3, and so we can simplify the function to simply g(x) = x + 3. (This is not the same as the original function, in that it has an extra point at (3, 6).) Thus the limit at x = 3 is 6. In fact, this kind of simplification is always possible with a removable discontinuity in a rational function. When the denominator is not 0, we can divide through to get a function which is the same. When it is 0, this new function will be identical to the old except for new points where previously we had division by 0. And above it was proved that the limit of this function (since it is continuous) is the same at the limit of the old function.

Exercises

1. Find $\lim_{x\to 5} 2x^2 - 3x + 4$.
2. Find $\lim_{x\to 2} \frac{x+1}{x^2+3}$.
3. Find $\lim_{x\to 1} \sqrt{x+1}$.
4. Verify the claims in these examples. (Hint: use the fact that the chord is shorter than the corresponding arc to treat the example 5.)
5. Generilize example 2 to $f(x) = x^m$ and example 3 to $x^{1/m}$.
6. Check that $1/x$ is continuous, but not uniformly continuous on $(0, 1]$

   (The following exercises an outline of another approach to continuity using the moduli of continuity, all functions are defined on a closed interval)
7. An increasing function that hits all its intermediate values is continuous.
8. The inverse of an increasing function is continous.
9. Bolzano theorem says that a continouous function defined on $[a, b]$ hits all the values between $f(a)$ and $f(b)$. Derive the following: an increasing function is continuous if and only if it hits all its intermediate values.
10. A continuous function that has an inverse must be monotonic (= increasing or decreasing). (Hint: Use Bolzano).
11. A one-to-one function from an interval onto another interval is continuous if and only if its inverse is continuous.
12. Assume that $|f(x) - f(a)| \leq g(|x - a|)$ with increasing continuous g and $g(0) = 0$. Then f is continuous at a.
13. Let $|f(x + h) - f(x)| \leq g(|h|)$ for any x, with g as in the previous exercise. Then f is uniformly continuous.

2.5 Limits at infinity

Discussion.

[**author=garrett**, file =**text_files/limits_at_infinity**]
On the other hand, what we really mean anyway is *not* that x 'becomes infinite' in some *mystical* sense, but rather that it just 'gets larger and larger'. In this context, the crucial observation is that, as x gets larger and larger, $1/x$ gets smaller and smaller (going to 0). Thus, just based on what we want this all to mean,

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

$$\lim_{x\to\infty} \frac{1}{x^2} = 0$$

$$\lim_{x\to\infty} \frac{1}{x^3} = 0$$

and so on.

This is the essential idea for evaluating simple kinds of limits as $x \to \infty$: rearrange the whole thing so that everything is expressed in terms of $1/x$ instead of x, and then realize that

$$\lim_{x\to\infty} \quad \text{is the same as} \quad \lim_{\frac{1}{x}\to 0}$$

Example 2.5.1.

[**author=garrett**, file =**text_files/limits_at_infinity**]
Next, let's consider

$$\lim_{x\to\infty} \frac{2x+3}{5-x}$$

The hazard here is that ∞ is *not* a number that we can do arithmetic with in the normal way. Don't even try it. So we *can't* really just 'plug in' ∞ to the expression to see what we get.

So, divide numerator and denominator both by *the largest power of x appearing anywhere:*

$$\lim_{x\to\infty} \frac{2x+3}{5-x} = \lim_{x\to\infty} \frac{2+\frac{3}{x}}{\frac{5}{x}-1} = \lim_{y\to 0} \frac{2+3y}{5y-1} = \frac{2+3\cdot 0}{5\cdot 0-1} = -2$$

Discussion.

[**author=garrett**, file =**text_files/limits_at_infinity**]
The point is that we called $1/x$ by a new name, 'y', and rewrote the original limit as $x \to \infty$ as a limit as $y \to 0$. Since 0 *is* a genuine number that we can do arithmetic with, this brought us back to ordinary everyday arithmetic. Of course, it was necessary to rewrite the thing we were taking the limit of in terms of $1/x$ (renamed 'y').

Notice that this is an example of a situation where we used the letter 'y' for something other than the name or value of the vertical coordinate.

Discussion.

[author=garrett, file =text_files/limits_infinity_exponential]
It is important to appreciate the behavior of exponential functions as the input to them becomes a large positive number, or a large negative number. This behavior is different from the behavior of polynomials or rational functions, which behave similarly for large inputs regardless of whether the input is large *positive* or large *negative*. By contrast, for exponential functions, the behavior is radically different for large *positive* or large *negative*.

As a reminder and an explanation, let's remember that exponential notation started out simply as an **abbreviation**: for positive integer n,

$$2^n = 2 \times 2 \times 2 \times \ldots \times 2 \qquad (n \text{ factors})$$

$$10^n = 10 \times 10 \times 10 \times \ldots \times 10 \qquad (n \text{ factors})$$

$$\left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right) \times \left(\frac{1}{2}\right) \times \left(\frac{1}{2}\right) \times \ldots \times \left(\frac{1}{2}\right) \qquad (n \text{ factors})$$

From this idea it's not hard to understand the **fundamental properties of exponents** (they're not *laws* at all):

$$a^{m+n} = \underbrace{a \times a \times a \times \ldots \times a}_{m+n} \qquad (m+n \text{ factors})$$

$$= \underbrace{(a \times a \times a \times \ldots \times a)}_{m} \times \underbrace{(a \times a \times a \times \ldots \times a)}_{n} = a^m \times a^n$$

and also

$$a^{mn} = \underbrace{(a \times a \times a \times \ldots \times a)}_{mn} =$$

$$= \underbrace{\underbrace{(a \times a \times a \times \ldots \times a)}_{m} \times \ldots \times \underbrace{(a \times a \times a \times \ldots \times a)}_{m}}_{n} = (a^m)^n$$

at least for positive integers m, n. Even though we can only easily see that these properties are true when the exponents are positive integers, the *extended* notation is guaranteed (by its *meaning*, not by *law*) to follow the same rules.

Discussion.

[author=garrett, file =text_files/limits_infinity_exponential]
Use of *other* numbers in the exponent is something that came later, and is also just an *abbreviation*, which happily was *arranged* to match the more intuitive simpler version. For example,

$$a^{-1} = \frac{1}{a}$$

and (as consequences)

$$a^{-n} = a^{n \times (-1)} = (a^n)^{-1} = \frac{1}{a^n}$$

(whether n is positive or not). Just to check one example of consistency with the properties above, notice that

$$a = a^1 = a^{(-1)\times(-1)} = \frac{1}{a^{-1}} = \frac{1}{1/a} = a$$

This is not supposed to be surprising, but rather reassuring that we won't reach false conclusions by such manipulations.

Also, fractional exponents fit into this scheme. For example

$$a^{1/2} = \sqrt{a} \qquad a^{1/3} = \sqrt[3]{a}$$

$$a^{1/4} = \sqrt[4]{a} \qquad a^{1/5} = \sqrt[5]{a}$$

This is *consistent* with earlier notation: the fundamental property of the n^{th} root of a number is that its n^{th} power is the original number. We can check:

$$a = a^1 = (a^{1/n})^n = a$$

Again, this is not supposed to be a surprise, but rather a consistency check.

Then for arbitrary *rational* exponents m/n we can maintain the same properties: first, the definition is just

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

One hazard is that, if we want to have only real numbers (as opposed to complex numbers) come up, then we should not try to take square roots, 4^{th} roots, 6^{th} roots, or any *even* order root of negative numbers.

For general *real* exponents x we likewise should *not* try to understand a^x except for $a > 0$ or we'll have to use complex numbers (which wouldn't be so terrible). But the value of a^x can only be defined as a *limit*: let $r_1, r_2, \ldots$ be a sequence of *rational* numbers approaching x, and define

$$a^x = \lim_i a^{r_i}$$

We would have to check that this definition does not accidentally depend upon the sequence approaching x (it doesn't), and that the same properties still work (they do).

Discussion.

[`author=garrett`, file =`text_files/limits_infinity_exponential`]
The number e is not something that would come up in really elementary mathematics, because its reason for existence is not really elementary. Anyway, it's approximately

$$e = 2.71828182845905$$

but if this ever really mattered you'd have a calculator at your side, hopefully.

Discussion.

[`author=garrett`, file =`text_files/limits_infinity_exponential`]

With the definitions in mind it is easier to make sense of questions about **limits** of exponential functions. The two companion issues are to evaluate

$$\lim_{x\to+\infty} a^x$$

$$\lim_{x\to-\infty} a^x$$

Since we are allowing the exponent x to be *real*, we'd better demand that a be a *positive real* number (if we want to avoid complex numbers, anyway). Then

$$\lim_{x\to+\infty} a^x = \begin{cases} +\infty & \text{if} \quad a > 1 \\ 1 & \text{if} \quad a = 1 \\ 0 & \text{if} \quad 0 < a < 1 \end{cases}$$

$$\lim_{x\to-\infty} a^x = \begin{cases} 0 & \text{if} \quad a > 1 \\ 1 & \text{if} \quad a = 1 \\ +\infty & \text{if} \quad 0 < a < 1 \end{cases}$$

To remember which is which, it is sufficient to use 2 for $a > 1$ and $\frac{1}{2}$ for $0 < a < 1$, and just let x run through positive integers as it goes to $+\infty$. Likewise, it is sufficient to use 2 for $a > 1$ and $\frac{1}{2}$ for $0 < a < 1$, and just let x run through negative integers as it goes to $-\infty$.

Exercises

1. Find $\lim_{x\to\infty} \frac{x+1}{x^2+3}$.
2. Find $\lim_{x\to\infty} \frac{x^2+3}{x+1}$.
3. Find $\lim_{x\to\infty} \frac{x^2+3}{3x^2+x+1}$.
4. Find $\lim_{x\to\infty} \frac{1-x^2}{5x^2+x+1}$.
5. Find $\lim_{x\to\infty} e^{-x^2}$

Chapter 3

Derivatives

3.1 The idea of the derivative of a function

Discussion.

[author=wikibooks, file =text_files/derivatives_intro]
Historically, the primary motivation for the study of 'differentiation' was to solve a problem in mathematics known as the tangent line problem for a given curve, find the slope of the straight line that is tangent to the curve at a given point.

The solution is obvious in some cases for example, a straight line, y = mx + c, is its own tangent so the slope at any point is m. For the parabola $y = x^2$, the slope at the point (0,0) is 0 (the tangent line is flat). In fact, at any vertex of any smooth function the slope is zero, because the tangent line slopes in opposite directions on either side.

But how does one find the slope of, say, $y = \sin(x) + x^2$ at $x = 1.5$?

The easiest way to find slopes for any function is by differentiation. This process results in another function whose value for any value of x is the slope of the original function at x. This function is known as the derivative of the original function, and is denoted by either a prime sign, as in $f'(x)$ (read "f prime of x"), the quotient notation, $\frac{df}{dx}$ or $\frac{d}{dx}[f]$ (which is more useful in some cases), or the differential operator notation, $D_x[f(x)]$, which is generally just written as $Df(x)$.

Most of the time, the brackets are not needed, but are useful for clarity if we speak of something like D (fg) for a product.

Example 3.1.1.

[author=wikibooks, file =text_files/derivatives_intro]
For example, if $f(x) = 3x + 5$, then $f'(x) = 3$, no matter what x is. If f(x) = —x—, the absolute value function, then

$$f'(x) = \begin{cases} -1, & x < 0 \\ \text{undefined}, & x = 0 \\ 1, & x > 0 \end{cases}$$

The reason f'(x) is undefined at 0 is that the slope suddenly changes at 0, so there is no single slope at 0 - it could be any slope from -1 to 1 inclusive.

Definition 3.1.1.

[author=wikibooks, file =text_files/derivatives_intro]
The definition of slope between two points (x_1, y_1) and (x_2, y_2) is $m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$.

If the two points are on a function $f(x)$ and $f(x_i) = y_i$.

If we let $h = \Delta x = x_2 - x_1$ then $x_2 = x_1 + h$ and $y_2 = f(x_2) = f(x_1 + h)$ and of course $y_1 = f(x_1)$

we find that by substituting these into the former equation, we can express it in terms of two variables (h and x_1)

$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{f(x_1 + h) - f(x_1)}{h}$.

We then, to find the slope at a single point, let $x_2 \to x_1$ to become any point x. This also defines $h \to 0$. By defining h and x, we have defined the slope - or derivative - at any single point x as the [[CalculusLimits—limit]]

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

Definition 3.1.2.

[author=duckworth,, file =text_files/derivatives_intro]
After we have absorbed the idea of the derivative at a single point $x = a$, we will start looking for formulas which will work for any value of a. In this context, we don't know what a is and so we will write x instead of a. The derivative of $f(x)$ is the following *function*:

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

We write this as $f'(x)$ or $\frac{df}{dx}$ or $\frac{d}{dx} f(x)$.

Example 3.1.2.

[author=duckworth, file =text_files/derivatives_intro]
For example, let $f(x) = x^2$ and suppose we are interested in the derivative at $a = 2$, 3 and 4. We can show that $f'(2) = 4$ and $f'(3) = 6$ and $f'(4) = 8$. But it is more compact to say that $f'(x) = 2x$ and let anyone who wants to plug numbers into the formula.

Discussion.

[author=wikibooks, file =text_files/velocity_problem_as_limit]
To see the power of the limit, let's go back to the moving car we talked about at the introduction. Suppose we have a car whose position is linear with respect to time (that is, that a graph plotting the position with respect to time will show a staight line). We want to find the velocity. This is easy to do from algebra, we just take a slope, and that's our velocity.

But unfortunately (or perhaps fortunately if you are a calculus teacher), things in the real world don't always travel in nice straight lines. Cars speed up, slow down, and generally behave in ways that make it difficult to calculate their velocities. (figure 2)

Now what we really want to do is to find the velocity at a given moment. (figure 3) The trouble is that in order to find the velocity we need two points, while at any given time, we only have one point. We can, of course, always find the average speed of the car, given two points in time, but we want to find the speed of the car at one precise moment.

Here is where the basic trick of differential calculus comes in. We take the average speed at two moments in time, and then make those two moments in time closer and closer together. We then see what the limit of the slope is as these two moments in time are closer and closer, and as those two moments get closer and closer, the slope comes out to be closer and closer to the slope at a single instant.

Discussion.

[author=garrett, file =text_files/derivative_idea]
First we can tell what the *idea* of a derivative is. But the issue of *computing* derivatives is another thing entirely: a person can understand the *idea* without being able to effectively *compute*, and vice-versa.

Suppose that f is a function of interest for some reason. We can give f some sort of 'geometric life' by thinking about *the set of points* (x, y) *so that*

$$f(x) = y$$

We would say that this describes a *curve* in the (x, y)-plane. (And sometimes we think of x as 'moving' from left to right, imparting further intuitive or physical content to the story).

For some particular number x_o, let y_o be the value $f(x_o)$ obtained as output by plugging x_o into f as input. Then the point (x_o, y_o) is a point on our curve. The **tangent line** to the curve **at** the point (x_o, y_o) is a line passing through (x_o, y_o) and 'flat against' the curve. (As opposed to *crossing it at some definite angle*).

The *idea* of the derivative $f'(x_o)$ is that it is *the slope of the tangent line* at x_o to the curve. But this isn't the way to *compute* these things...

Example 3.1.3.

[author=livshits, file =text_files/derivative_idea]
A troublemaker on the seventh floor dropped a plastic bag filled with water. It took the bag 2 seconds to hit the ground. How fast was the bag moving at that moment? The distance the bag drops in t seconds is $s(t) = 16t^2$ feet.

The average velocity of the bag between time t and time 2 is $(s(t)-s(2))/(t-2)$. If we take $t = 2$ the expression becomes $0/0$ and it is undefined. To make sense out of it we should use the formula for $s(t)$. When we plug it in, we get $16(t^2-2^2)/(t-2)$. The numerator is divisible by the denominator because $t^2 - 2^2 = (t+2)(t-2)$, therefore the expression can be rewritten as $16(t+2)$, and it makes sense for $t = 2$ too. The problem is solved; the velocity of the bag when it hits the ground is $16(2 + 2) = 64$ ft/sec. More generally, the velocity at time t will be $32t$ (exercise).

Was it just luck? Not at all! The reason for our success is that the numerator is a polynomial in t that vanishes at $t = 2$, so the numerator is divisible by $t - 2$ (see section 1.2); the ratio, which is $16(t + 2)$, is a polynomial in t and is defined for $t = 2$. Now we can see that the trick will work when $s(t)$ is any polynomial whatsoever.

But is our trick good only for polynomials? No, as we can see from the following problem.

Example 3.1.4.

[`author=livshits`, file =`text_files/derivative_idea`]
The area of a circular puddle is growing at π square feet per second. How fast is the radius of the puddle growing at time T? Assume that the area was 0 at time 0 when the puddle started growing.

Let us denote by $r(t)$ the radius of the puddle at time t. Then the area of the puddle at time t is $\pi r(t)^2$, which must be equal to πt. Therefore $r(t) = \sqrt{\pi t/\pi} = \sqrt{t}$. Now we have to make sense out of the expression $(\sqrt{t} - \sqrt{T})/(t - T)$ for $t = T$. To do so we can multiply both the numerator and the denominator by $\sqrt{t} + \sqrt{T}$, then we get $(t - T)/(\sqrt{t} + \sqrt{T})(t - T) = 1/(\sqrt{t} + \sqrt{T})$ which makes sense for $t = T$. We conclude that at time T the radius r is growing at $1/(2\sqrt{(T)})$ feet per second.

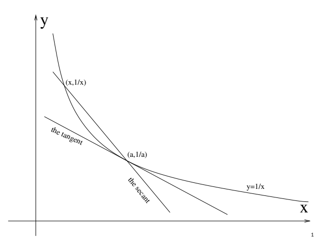
You may notice that it is the same trick "upside down", because if we put $z = \sqrt{t}$ and $Z = \sqrt{T}$, the undefined expression to take care of becomes $(z - Z)/(z^2 - Z^2)$ which is the same as $(z - Z)/((z - Z)(z + Z))$.

Here is one more similar problem that is easy enough to do "with the bare hands".

Example 3.1.5.

[`author=livshits`, file =`text_files/derivative_idea`]
the slope of the tangent line to the hyperbola $y = 1/x$ at the point $x = a, y = 1/a$.

The slope of the secant line that passes through the points $(a, 1/a)$ and $(x, 1/x)$ is $(1/x - 1/a)/(x - a)$ which is an expression that is not defined for $x = a$, but we can rewrite it in the form $-(x - a)/(xa)/(x - a)$ which becomes (after we cancel $x - a$) $-1/(xa)$ which is defined for $x = a$ and is $-1/a^2$.



1_over_x_tangent_secant

Definition 3.1.3.

[`author=wikibooks`, file =`text_files/derivatives_definition`]

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

This is known as the definition of a derivative. The more visual explanation of this formula is that the slope of the tangent line touching one point is the limit of the slopes of the secant lines intersecting two points near that point, as the two points merge to one.

Example 3.1.6.

[author=wikibooks, file =text_files/derivatives_definition]
Let us try this for a simple function

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

$f'(x) = \lim_{h\to 0} \frac{\frac{x}{2}+\frac{h}{2}-\frac{x}{2}}{h} = \lim_{h\to 0} \frac{1}{2} = \frac{1}{2}$

This is consistent with the definition of the derivative as the slope of a function.

Example 3.1.7.

[author=wikibooks, file =text_files/derivatives_definition]
Sometimes, the slope of a function varies with x. This is easily demonstrated by the function $f(x) = x^2$,

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

Though this may seem surprising, because $y = x^2$ fits $y = mx + c$ if $m = x$ and $c = 0$, it becomes intuitive when one realizes that the slope changes twice as fast as with $m = x$ because there are two xs that vary.

Discussion.

[author=livshits, file =text_files/derivatives_definition]
In each of the three problems that we dealt with so far we had a function, let us call it now $f(x)$, and we had to make sense out of the ratio $q(x, a) = (f(x) - f(a))/(x - a)$ (which is called the difference quotient) for $x = a$. The difference quotient $q(x, a)$ is well defined for $x \neq a$, but when $x = a$ both the numerator and the denominator vanish, so $q(a, a)$ is undefined if we treat it as the quotient of numbers (it is clear that any number c can be considered as the quotient $0/0$ because $c \cdot 0 = 0$ for any c).

Our approach was to rewrite the expression for $q(x, a)$ in a form $p(x, a)$ that is well defined for $x = a$ and that agrees with $q(x, a)$ for $x \neq a$. For example, in the third problem $f(x) = 1/x$, $q(x, a) = (1/x - 1/a)/(x - a)$ which is undefined for $x = a$, $p(x, a) = -1/(xa)$ which is well defined for $x = a$, also $q(x, a) = p(x, a)$ for $x \neq a$.

The key idea is to consider the numerator and the denominator in $q(x, a)$, as well as $p(x, a)$, as functions of a certain class, not as numbers, to disambiguate the ambiguous expression $0/0$.

For example, in the first problem our class of functions is the polynomials of t, in the second problem it is the class of rational functions of $\sqrt{t}$ and $\sqrt{T}$, while in the third problem it is the class of rational functions of x and a.

Why do we need a special class of functions? Why can't we consider all functions whatsoever? Because the class of all functions is too wide to disambiguate the ambiguous ratio 0/0. Indeed, if we allow $p(x, a)$ to be any function such that $q(x, a) = p(x, a)$ for $x \neq a$, we can get no information about $p(a, a)$ because $p(a, a)$ can be changed to any number if we admit all the functions into the game. We see that some restrictions on the functions that we treat are inevitable.

The following property of the functions we treated so far was crucial for our success: any 2 of such functions that are defined for $x = a$ and coincide for all $x \neq a$ also coincide for $x = a$. It means that the value $p(a, a)$ is defined unambiguously by the condition that $p(x, a) = q(x, a)$ for $x \neq a$ (see the last paragraph of section **??**).

Later on we will describe some other classes of functions, much more general than the ones we dealt with so far, but still nice enough for our machinery to work.

To summarize briefly, the function f is *differentiable* if the increment $f(x)-f(a)$ factors as $f(x) - f(a) = (x - a)p(x, a)$ and the function $p(x, a)$ is well defined for $x = a$. The *derivative* $f'(a) = p(a, a)$.

In the next section we will consider some elementary properties (the rules) of differentiation that will be handy in calculations.

Discussion.

[`author=wikibooks`, file =`text_files/derivative_notation`]
The derivative notation is special and unique in mathematics. The most common use of derivatives you'll run into when first starting out with differentiating, is the $\frac{dy}{dx}$ differentiation. You can think of this as "change in y divided by change in x". You can also think of it as "infinitesimal value of y divided by infinitesimal value of x". Either way is a good way of thinking. Often, in an equation, you will see just $\frac{d}{dx}$, which literally means "derivative with respect to x". You can safely assume that this is the equivalent of $\frac{dy}{dx}$ for now.

Also, later, as you advance through your studies, you will see that dy and dx can act as separate entities that can be multiplied and divided (to a certain degree). Eventually you will see derivatives such as $\frac{dx}{dy}$, which sometimes will be written $\frac{d}{dy}$, or you'll see a derivative in polar coordinates marked up as $\frac{d\theta}{dr}$.

Notation.

[`author=livshits`, file =`text_files/derivative_notation`]
The standard notation (due to Lagrange) for the derivative of f for $x = a$ is $f'(a)$. We can also consider it as a function of a and then differentiation becomes the operation of passing from a function f to its derivative f' (which is also a function of x).

The other notation for f' (due to Leibniz) is df/dx. In particular, we can say that we calculated $s'(2)$ in our first problem, $r'(T)$ in our second problem and $dy/dx(a)$ in our third problem. We can also write the results we got so far as $(16t^2)' = 32t$, $\sqrt{t}' = 1/(2\sqrt{t}$ and $d(1/x)/dx = -1/x^2$.

Newton used dots on top of the letters denoting functions as the differentiation sign; for example, by solving problem 1, we got $\dot{s}(t) = 32t$. This notation is still popular in mechanics.

3.2 Derivative Shortcuts

Example 3.2.1.

[`author=livshits`, file =`text_files/deriv_of_polys`]
In this section we will discuss the *division* of polynomials and prove that $p(x)$ is divisible by $x-a$ if and only if $p(a)=0$. From this fact we will derive that a polynomial of degree d can not have more than d zeroes. It will follow that 2 rational functions that coincide on an infinite set also coincide wherever both of them are defined.

Polynomials can be *divided with remainder* pretty much the same way as integers. Let us start with an example that will make the general rule clear. We will divide $3x^7+5x^4+x^2+1$ by $x-3$.

$$
\begin{array}{rllllllll}
 & 3x^6+ & 9x^5+ & 27x^4+ & 86x^3+ & 258x^2+ & 775x+ & 2325 & \\
x-3) & \overline{3x^7+} & & & 5x^4+ & & x^2+ & & 1 \\
 & 3x^7- & 9x^6 & & & & & & \\
 & & 9x^6+ & & 5x^4+ & & x^2+ & & 1 \\
 & & 9x^6- & 27x^5 & & & & & \\
 & & & 27x^5+ & 5x^4+ & & x^2+ & & 1 \\
 & & & 27x^5- & 81x^4 & & & & \\
 & & & & 86x^4+ & & x^2+ & & 1 \\
 & & & & 86x^4- & 258x^3 & & & \\
 & & & & & 258x^3+ & x^2+ & & 1 \\
 & & & & & 258x^3- & 774x^2 & & \\
 & & & & & & 775x^2+ & & 1 \\
 & & & & & & 775x^2- & 2325x & \\
 & & & & & & & 2325x+ & 1 \\
 & & & & & & & 2325x- & 6975 \\
 & & & & & & & & 6976
\end{array}
$$

On each step we multiplied the divisor by the monomial to kill the leading term of the remainder obtained at the previous step. This way the degree of the remainder dropped by one every step of the process. The process stops when the degree of the remainder is less than the degree of the divisor. The remainder in our example is 6976. On the other hand, $p(3)=6976$ too. Is it a coincidence? No, because the result of the division can be written as

$$p(x)=3x^7+5x^4+x^2+1=(3x^6+9x^5+27x^4+86x^3+258x^2+775x+2325)(x-3)+6976$$

and we can plug $x=3$ into this formula to see that $p(3)=6976$. In general, $p(a)$ is the remainder of the division of $p(x)$ by $x-a$, in particular, $p(a)=0$ if and only if $x-a$ divides $p(x)$ *evenly*, i.e. with zero remainder.

This is a very important fact. Assume that $a_1, ..., a_k$ are the roots of $p(x)$. Then each $x - a_j$ divides $p(x)$, whence $p(x) = (x - a_1)...(x - a_k)g(x)$, so the degree of p is at least k. It follows that a polynomial of degree d can not have more than d different roots. In particular, no nonzero polynomial can have infinite number of roots; in other words, if a polynomial has an infinite number of roots, it is zero. Also two polynomial functions that coincide on an infinite set must coincide everywhere (consider their difference!).

We can also see that any rational function is well defined for all the values of the argument except for the finite number of values at which some denominator involved in this function vanishes.

It also follows that a rational function can have at most a finite number of zeroes, in particular, any two rational functions that coincide on an infinite set coincide wherever they are both defined (exercise!).

We can use this fact to check our algebraic manipulations. For example, if we rewrite some formula in a different form, to catch a mistake it is usually enough to plug in some random number into both formulas and see if they give different results. The probability that this approach fails is zero.

Discussion.

[`author=garrett`, file =`text_files/deriv_of_polys`]
There are just four simple facts which suffice to take the derivative of any polynomial, and actually of somewhat more general things.

Rule 3.2.1.

[`author=garrett`, file =`text_files/deriv_of_polys`]
First, there is the rule for taking the derivative of a **power function** which takes the nth power of its input. That is, these functions are functions of the form $f(x) = x^n$. The formula is

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

That is, the exponent comes down to become a coefficient in front of the thing, and the exponent is decreased by 1.

Rule 3.2.2.

[`author=garrett`, file =`text_files/deriv_of_polys`]
The second rule, which is really a special case of this power-function rule, is that *derivatives of constants are zero*:

$$\frac{d}{dx}c = 0$$

for any constant c.

Rule 3.2.3.

[`author=garrett`, file =`text_files/deriv_of_polys`]

The third thing, which reflects the innocuous role of constants in calculus, is that for *any* function f of x we have

$$\frac{d}{dx} c \cdot f = c \cdot \frac{d}{dx} f$$

The fourth is that for *any* two functions f, g of x, *the derivative of the sum is the sum of the derivatives*:

$$\frac{d}{dx}(f+g) = \frac{d}{dx} f + \frac{d}{dx} g$$

Rule 3.2.4.

[`author=garrett`, file =`text_files/deriv_of_polys`]
Putting these four things together, we can write general formulas like

$$\frac{d}{dx}(ax^m + bx^n + cx^p) = a \cdot mx^{m-1} + b \cdot nx^{n-1} + c \cdot px^{p-1}$$

and so on, with more summands than just the three, if so desired. And in any case here are some examples with numbers instead of letters:

$$\frac{d}{dx} 5x^3 = 5 \cdot 3x^{3-1} = 15x^2$$

$$\frac{d}{dx}(3x^7 + 5x^3 - 11) = 3 \cdot 7x^6 + 5 \cdot 3x^2 - 0 = 21x^6 + 15x^2$$

$$\frac{d}{dx}(2 - 3x^2 - 2x^3) = 0 - 3 \cdot 2x - 2 \cdot 3x^2 = -6x - 6x^2$$

$$\frac{d}{dx}(-x^4 + 2x^5 + 1) = -4x^3 + 2 \cdot 5x^4 + 0 = -4x^3 + 10x^4$$

Even if you do catch on to this idea right away, it is wise to practice the *technique* so that not only can you do it *in principle*, but also *in practice*.

Rule 3.2.5.

[`author=livshits`, file =`text_files/deriv_of_polys`]
Sums Rule: $(f+g)'(x) = f'(x) + g'(x)$

Multiplier Rule: $(cf)'(x) = cf'(x)$ when c is a constant

Both rules together say that differentiation is a *linear* operation. These rules are sort of obvious. For example, to calculate $(f+g)'(a)$ we consider the difference quotient $(f(x)+g(x)-(f(a)+g(a)))/(x-a)$ which can be rewritten as $(f(x)-f(a))/(x-a)+(g(x)-g(a))/(x-a)$. Since both additive terms make sense for $x=a$ and produce $f'(a)$ and $g'(a)$, we are done.

Examples 3.2.2.

[author=garrett, file =text_files/deriv_powers]
It's important to remember some of the other possibilities for the exponential notation x^n. For example

$$x^{1/2} = \sqrt{x}$$
$$x^{-1} = \frac{1}{x}$$
$$x^{-1/2} = \frac{1}{\sqrt{x}}$$

and so on. The good news is that the rule given just above for taking the derivative of powers of x still is correct here, even for exponents which are negative or fractions or even real numbers:

$$\frac{d}{dx} x^r = r\, x^{r-1}$$

Thus, in particular,

$$\frac{d}{dx}\sqrt{x} = \frac{d}{dx} x^{1/2} = \frac{1}{2} x^{-1/2}$$
$$\frac{d}{dx}\frac{1}{x} = \frac{d}{dx} x^{-1} = -1 \cdot x^{-2} = \frac{-1}{x^2}$$

When combined with the sum rule and so on from above, we have the obvious possibilities:

Example 3.2.3.

[author=garrett, file =text_files/deriv_powers]

$$\frac{d}{dx}(3x^2 - 7\sqrt{x} + \frac{5}{x^2} = \frac{d}{dx}(3x^2 - 7x^{\frac{1}{2}} + 5x^{-2}) = 6x - \frac{7}{2}x^{-1/2} - 10x^{-3}$$

Comment.

[author=garrett, file =text_files/deriv_powers]
The possibility of expressing square roots, cube roots, inverses, etc., in terms of exponents is a very important idea in algebra, and can't be overlooked.

Discussion.

[author=wikibooks, file =text_files/derivative_rules]
The process of differentiation is tedious for large functions. Therefore, rules for differentiating general functions have been developed, and can be proved with a little effort. Once sufficient rules have been proved, it will be possible to differentiate a wide variety of functions. Some of the simplest rules involve the derivative of linear functions.

Rule 3.2.6.

[author=wikibooks, file =text_files/derivative_rules]

Constant rule $\frac{d}{dx}c = 0$

Linear functions $\frac{d}{dx}mx = m$

The special case $\frac{dy}{dx} = 1$ shows the advantage of the d/dx notation - rules are intuitive by basic algebra, though this does not constitute a proof, and can lead to misconceptions to what exactly dx and dy actually are.

Constant multiple and addition rules Since we already know the rules for some very basic functions, we would like to be able to take the derivative of more complex functions and break them up into simpler functions. Two tools that let us do this are the constant multiple rules and the addition rule.

The constant multiple rule is $\frac{d}{dx}cf(x) = c\frac{d}{dx}f(x)$

The reason, of course, is that one can factor the c out of the numerator, and then of the entire limit, in the definition.

Example 3.2.4.

[author=wikibooks, file =text_files/derivative_rules]
Example we already know that $\frac{d}{dx}x^2 = 2x$ Suppose we want to find the derivative of $3x^2$ $\frac{d}{dx}3x^2 = 3\frac{d}{dx}x^2 = 3 \times 2x = 6x$

Rule 3.2.7.

[author=wikibooks, file =text_files/derivative_rules]

Addition rule $\frac{d}{dx}\left(f(x) + g(x)\right) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$

Subtraction Rule $\frac{d}{dx}\left(f(x) - g(x)\right) = \frac{d}{dx}f(x) - \frac{d}{dx}g(x)$

Example 3.2.5.

[author=wikibooks, file =text_files/derivative_rules]
Example what is $\frac{d}{dx}\left(3x^2 + 5x\right)$

$\frac{d}{dx}\left(3x^2 + 5x\right) = \frac{d}{dx}3x^2 + \frac{d}{dx}5x$

$= 6x + \frac{d}{dx}5x$

$= 6x + 5$

Comment.

[author=wikibooks, file =text_files/derivative_rules]
The fact that both of these rules work is extremely significant mathematically because it means that differentiation is **linear**. You can take an equation, break

it up into terms, figure out the derivative individually and build the answer back up, and nothing odd will happen.

Rule 3.2.8.

[author=wikibooks, file =text_files/derivative_rules]
The Power Rule $\frac{d}{dx}x^n = nx^{n-1}$ - that is, bring down the power and reduce it by one.

Example 3.2.6.

[author=wikibooks, file =text_files/derivative_rules]
For example, in the case of x^2, the derivative is $2x^1 = 2x$, as was established earlier.

Example 3.2.7.

[author=wikibooks, file =text_files/derivative_rules]
The power rule also applies to fractional and negative powers, therefore

$$\frac{d}{dx}\left[\sqrt{x}\right] = \frac{d}{dx}\left[x^{1/2}\right] = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}} = \frac{\sqrt{x}}{2x}$$

Comment.

[author=wikibooks, file =text_files/derivative_rules]
Since polynomials are sums of monomials, using this rule and the addition rule lets you differentiate any polynomial.

Example 3.2.8.

[author=wikibooks, file =text_files/derivative_rules]
With the rules in hand we can now find the derivative of any polynomial we want. Rather than write the general formula, let's go step by step through the process $\frac{d}{dx}\left[6x^5 + 3x^2 + 3x + 1\right]$ The first thing we can do is to use the addition rule to split the equation up into terms $\frac{d}{dx}6x^5 + \frac{d}{dx}3x^2 + \frac{d}{dx}3x + \frac{d}{dx}1$ Immediately we can use the linear and constant rules to get rid of some terms $\frac{d}{dx}6x^5 + \frac{d}{dx}3x^2 + 3 + 0$ We use the constant muliplier rule to move the constants outside the derivative $6\frac{d}{dx}x^5 + 3\frac{d}{dx}x^2 + 3$ Then we use the power rule to work with the powers $6(5x^4) + 3(2x) + 3$ And then we do some simple math to get our answer $30x^4 + 6x + 3$

Exercises

1. Find $\frac{d}{dx}(3x^7 + 5x^3 - 11)$

2. Find $\frac{d}{dx}(x^2 + 5x^3 + 2)$
3. Find $\frac{d}{dx}(-x^4 + 2x^5 + 1)$
4. Find $\frac{d}{dx}(-3x^2 - x^3 - 11)$
5. Find $\frac{d}{dx}(3x^7 + 5\sqrt{x} - 11)$
6. Find $\frac{d}{dx}(\frac{2}{x} + 5\sqrt[3]{x} + 3)$
7. Find $\frac{d}{dx}(7 - \frac{5}{x^3} + 5x^7)$

3.3 An alternative approach to derivatives

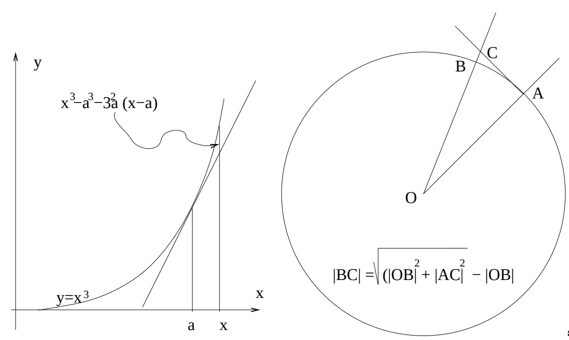
Discussion.

[**author=livshits**, file =**text_files/increasing_function_theorem**]
Our treatment of differentiation in Sections **??** and **??** was rather formal. In this section we will try to understand why a tangent looks like a tangent and why the average velocity over a small time interval is a good approximation for the instantaneous velocity. We will also prove the *Increasing Function Theorem* (IFT). This theorem says that if the derivative of a function is not negative, the function is nondecreasing. To put it informally, it says that if the velocity of a car is not negative, the car will not move backward. IFT will play the major role in our treatment of Calculus.

Example 3.3.1.

[**author=livshits**, file =**text_files/increasing_function_theorem**]
Let us start with a rather typical example. Consider a cubic polynomial $f(x) = x^3$ and the tangent to its graph at the point (a, a^3).

The equation of this tangent is $y = a^3 + 3a^2(x - a)$, and the vertical distance from a point on this tangent to the graph will be $|x^3 - a^3 - 3a^2(x - a)| = |(x - a)(x^2 + xa + a^2) - 3a^2(x - a)| = |(x - a)(x^2 + ax - 2a^2)| = |(x - a)(x^2 - a^2 + a(x - a))| = |(x - a)((x - a)(x + a) + a(x - a))| = |x + 2a|(x - a)^2$.



strange_cubic_tangent_circle

We see that this distance has a factor $(x - a)^2$ in it. The other factor, $|x + 2a|$ will be bounded by some constant K if we restrict x and a to some finite segment $[A, B]$, in other words, if we demand that $A \le x \le B$ and $A \le a \le B$ (in fact we can take $K = 3max\{|A|, |B|\}$).

Now the whole estimate can be rewritten as $|f(x) - f(a) - f'(a)(x - a)| \le K(x - a)^2$ for x and a in $[A, B]$. Here K may depend only on function f and on segment $[A, B]$, but not on x and a. We can also see that $|(f(x) - f(a))/(x - a) - f'(a)| \le K|x - a|$ for $x \ne a$.

The same kind of estimates hold when f is any polynomial or a rational function defined everywhere in $[A, B]$, it is also true if f is *sin* or *cos*

Definition 3.3.1.

[**author=livshits**, file =**text_files/increasing_function_theorem**]
We say that f is *uniformly Lipschitz differentiable* on $[A, B]$ if for some constant

K we have

$$|f(x) - f(a) - f'(a)(x - a)| \leq K(x - a)^2 \tag{3.1}$$

for all x and a in $[A, B]$.

Comment.

[**author=livshits**, file =**text_files/increasing_function_theorem**]
Geometrically speaking, **??** says that the graph $y = f(x)$ is located between the 2 parabolas: it is above the *lower parabola* with the equation

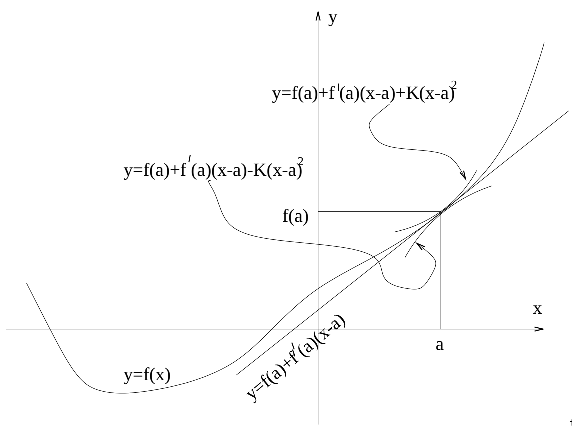
$$y = f(a) + f'(a)(x - a) - K(x - a)^2$$

and below the *upper parabola* with the equation

$$y = f(a) + f'(a)(x - a) + K(x - a)^2.$$

To see this we only have to rewrite Equation 3.1 in the form

$$f(a) + f'(a)(x - a) - K(x - a)^2 \leq f(x) \leq f(a) + f'(a)(x - a) + K(x - a)^2$$



tangent_parabs_bound_graph

We will often use "ULD" as an abbreviation for "uniformly Lipschitz differentiable."

The figure showing the upper and lower parabolas suggests that any ULD function with a positive derivative will be increasing. This is not easy to show, however, if we assume that $f'(x) \geq C$ for some $C > 0$, it becomes easy to demonstrate that f is increasing.

Comment.

[**author=livshits**, file =**text_files/increasing_function_theorem**]
Another motivation for this definition is related to the idea to view differentiation as factoring of functions of a certain class, that was developed in section 2.1. Let us say that we want to deal only with the functions that don't change too abruptly. To insure it we can demand that $|f(x) - f(a)|$ can be estimated in terms $|x - a|$, the simplest estimate of this kind is used in the following definition.

Definition 3.3.2.

[author=livshits, file =text_files/increasing_function_theorem]
A function g defined on $[A, B]$ is *uniformly Lipschitz continuous* if

$$|g(x) - g(a)| \leq L|x - a| \tag{3.2}$$

for all x and a in $[A, B]$.

Comment.

[author=livshits, file =text_files/increasing_function_theorem]
Important: the constant L (which is called *a Lipschitz constant* for g and $[A, B]$) in this definition depends only on the function and the interval, but not on individual x or a.

Definition 3.3.3.

[author=livshits, file =text_files/increasing_function_theorem]
We will often use "ULC" as an abbreviation for "uniformly Lipschitz continuous."

Definition 3.3.4.

[author=livshits, file =text_files/increasing_function_theorem]
Now let us say that $f(x) - f(a)$ factors as $f(x) - f(a) = (x - a)p(x, a)$ where $p(x, a)$ is a ULC function of x and $f'(a) = p(a, a)$. Then the following inequality holds for $x \neq a$:

$$|\frac{f(x) - f(a)}{x - a} - f'(a)| \leq L(a)|x - a|.$$

Here the function $L(a)$ may be rather nasty, but if it is bounded by a constant, that is if $L(a) \leq K$ for all a between A and B, we arrive (by multiplying both sides by $|x - a|$ and replacing $L(a)$ by K) at 3.1.

Increasing Function Theorem 3.3.1.
[author=livshits, file =text_files/increasing_function_theorem]
If f is uniformly Lipschitz differentiable on $[a, b]$ and $f' \geq 0$ then $f(a) \leq f(b)$.

Proof.
[author=livshits, file =text_files/increasing_function_theorem]
Case 1. We assume that if $f'(x) \geq C$ for some $C > 0$ then f is increasing.

It follows from this result that f will be increasing if $f' \geq 0$. Here is how. According to exercise **??**, for any $C > 0$ the function $f(x) + Cx$ will be increasing, i.e. for any $a < b$ we will have $f(a) + Ca \leq f(b) + Cb$, whence $f(b) - f(a) \geq -C(b - a)$, and since C is arbitrary, we must have $f(a) \geq f(b)$.

Case 2. The idea is the most popular one in Calculus: to chop up the segment $[A, B]$ into N equal pieces, use the estimate from our definition on each piece, and then notice what happens when N becomes large.

Let us take $x_n = A + n(B - A)/N$ for $n = 0, \ldots, N$ and let us take $a = x_{n-1}$ and $x = x_n$ in the estimate from the definition. The estimate from **??** can be rewritten as

$$-K(x_n - x_{n-1})^2 \le f(x_n) - f(x_{n-1}) - f'(x_{n-1})(x_n - x_{n-1}) \le K(x_n - x_{n-1})^2.$$

Since $f' \ge 0$ and $x_n \ge x_{n-1}$ and therefore $f'(x_{n-1})(x_n - x_{n-1}) \ge 0$, we can (by also noticing that $x_n - x_{n-1} = (B-A)/N$) get the following estimate:

$$-K(B-A)^2/N^2 = -K(x_n - x_{n-1})^2 \le f(x_n) - f(x_{n-1}).$$

Now let us replace $f(B) - f(A)$ with the following telescoping sum:

$$(f(x_1) - f(x_0)) + (f(x_2) - f(x_1)) + ... + (f(x_N) - f(x_{N-1}))$$

There are N terms in this sum, each one is $\ge -K(B-A)^2/N^2$, therefore the whole sum is $\ge -K(B-A)^2/N$. But the whole sum is equal to $f(B) - f(A)$, therefore

$$-K(B-A)^2/N \le f(B) - f(A)$$

This inequality can hold for all N only if $f(B) - f(A) \ge 0$ (this is called *Archimedes Principle*), therefore $f(A) \le f(B)$. □

Corollary 3.3.1.
[author= `livshits`, file =`text_files/increasing_function_theorem`]
If $f'(x) = 0$ for all x, then f is a constant function.

Proof.
[`author=livshits`, file =`text_files/increasing_function_theorem`]
Let f be ULD on $[A, B]$ and $f' = 0$. IFT tells us that $f(A) \ge f(B)$. But $(-f)' = 0$ too, so $-f(A) \ge -f(B)$, and $f(A) \le f(B)$, therefore $f(A) = f(B)$. Taking $A = u$ and $B = x$, $u \le x$ finishes the proof. □

Corollary 3.3.2.
[author= `livshits`, file =`text_files/increasing_function_theorem`]
From this result we can conclude that any two ULD antiderivatives of the same function may differ only by a constant, and therefore if $F' = f$ then all the ULD antiderivatives of f are of the form $F + C$, where C is a constant.

Theorem 3.3.2.
[author= `livshits`, file =`text_files/increasing_function_theorem`]
The derivative of a ULD function is ULC.

Proof.
[`author=livshits`, file =`text_files/increasing_function_theorem`]
For $x \ne a$, by dividing both sides of **??** by $|x - a|$, we get

$$\left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| \le K|x - a|. \tag{3.3}$$

This estimate may be handy to check your differentiation. If your formula for f' is right, the left side of 3.3 will be small for x close to a (how close – will depend on K), if it is wrong – it will not be so.

Interchanging x and a in formula 3.3 leads to

$$\left| \frac{f(a) - f(x)}{a - x} - f'(x) \right| \le K|a - x|,$$

but

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

and $|a-x| = |x-a|$, so $f'(x)$ and $f'(a)$ are less than $K|x-a|$ away from the same number, and therefore less than $2K|x-a|$ apart, i.e.

$$|f'(x) - f'(a)| \leq 2K|x-a|. \tag{3.4}$$

□

Comment.

[**author=livshits**, file =**text_files/increasing_function_theorem**]
This theorem together with the estimate 3.3 demonstrate that the time derivative of the distance is a reasonable mathematical metaphor for instantaneous velocity if the distance is a ULD function of time. Indeed, in this case the average velocity over a short enough time interval will be close to the time derivative of the distance at any time during this interval.

Comment.

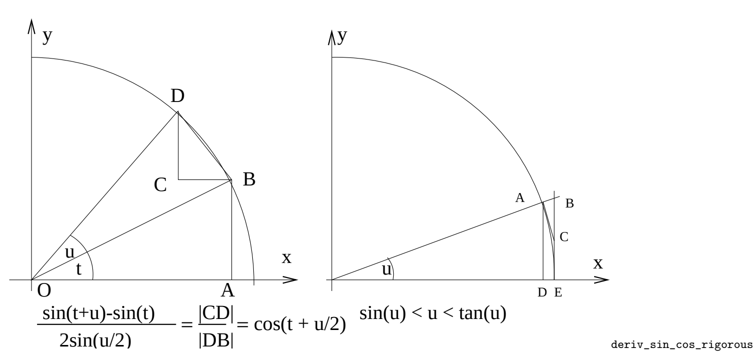
[**author=livshits**, file =**text_files/increasing_function_theorem**]
It is natural to ask whether any ULC function has a ULD primitive. Later on, after taking a closer look at area and integration, we show that it is true. Combining this fact with IFT, we can derive *positivity* of definite integrals that was promised at the end of section **??**.

It is also clear that uniform Lipschitz differentiability is stronger than mere divizibility of $f(x) - f(a)$ by $x - a$ in the class of ULC functions of x. As an example, consider $f(x) = x^2 sin(1/x)$. We have $f(0) = f'(0) = 0$, but the x-axis doesn't look like a tangent, near $x = 0$ it cuts the graph of f (that looks like fuzz) at infinitely many points. However, if f' understood in the spirit of section 2.1 turns out to be ULC, f will be ULD. To prove this fact one needs some rather delicate property of the real numbers (completeness) that will be treated in another chapter.

Derivation.

[**author=livshits**, file =**text_files/increasing_function_theorem**]
Here we give a rigorous proof of the derivative rules for $\sin(x)$ and $\cos(x)$.

Consider the following picture:



Dividing the inequality $\sin(u) < u < \tan(u)$ by u (assuming $\pi/4 > u > 0$), we get

$$\frac{\sin(u)}{u} < 1 < \frac{\tan(u)}{u} = \frac{\sin(u)/u}{\cos(u)},$$

therefore

$$\cos(u) < \frac{\sin(u)}{u} < 1$$

which holds for $-\pi/4 < u < 0$ as well since $\cos(-u) = \cos(u)$ and $\sin(-u) = -\sin(u)$, whence $\sin(-u)/(-u) = sin(u)/u$. Now

$$\frac{\sin(t+u) - \sin(u)}{u} = \frac{\sin(t+u) - \sin(u)}{2\sin(u/2)} \times \frac{2\sin(u/2)}{2(u/2)} = \cos(t+u/2)\sin(u/2)/(u/2).$$

To conclude our proof that $\sin'(u) = \cos(u)$ we have to get an estimate

$$\left|\cos(t) - \cos(t+u/2)\frac{\sin(u/2)}{u/2}\right| \leq K|u|$$

for some K. This is now easy because $|\cos(t)-\cos(t+u/2)| \leq |u|/2$, $|\sin(u/2)/(u/2)-1| \leq |\cos(u/2) - 1| \leq |u|/2$ and $|\cos(t+u/2)| \leq 1$, and by the triangle inequality we get

$$|\cos(t)-\cos(t+u/2)\sin(u/2)/(u/2)| \leq |\cos(t)-\cos(t+u/2)|+|cos(t+u/2)|\times|sin(u/2)/(u/2)-1| \leq |u|/2+|u|/2 \leq |u|$$

that demonstrates that

$$\left|\frac{sin(t+u) - sin(u)}{u} - cos(t)\right| \leq |u|,$$

and therefore $\sin'(u) = \cos(u)$.

This takes care of $\sin'$. To get the formula for $\cos'$ we can observe that $\cos(t) = \sin(\pi/2 - t)$, use the *chain rule* and then remember that $\cos(\pi/2 - t) = \sin(t)$. We leave the details as an exercise.

Exercises

1. The same kind of estimates as in section **??** hold when f is any polynomial or a rational function defined everywhere in $[A, B]$, it is also true if f is *sin* or *cos*

   Prove it (*sin* and *cos* involve some geometry, they will be treated later in this section).

2. Prove all the differentiation rules for ULD functions.

3. Try to show "The figure showing the upper and lower parabolas suggests that any ULD function with a positive derivative will be increasing." it and see that it is not easy.

4. Construct a demonstration that "However, if we assume that $f' \geq C$ for some $C > 0$, it becomes easy to demonstrate that f is increasing."

5. Show that functions with positive derivatives are increasing. Can you use IFT to make the argument easy?

6. Fill in the details of "This theorem together with the estimate 3.3 demonstrate that the time derivative of the distance is a reasonable mathematical metaphor for instantaneous velocity if the distance is a ULD function of time. Indeed, in this case the average velocity over a short enough time interval will be close to the time derivative of the distance at any time during this interval. "

3.4 Derivatives of transcendental functions

Discussion.

[author=garrett, file =text_files/deriv_transcend]
The new material here is just a list of formulas for taking derivatives of exponential, logarithm, trigonometric, and inverse trigonometric functions. Then any function made by composing these with polynomials or with each other can be differentiated by using the chain rule, product rule, etc. (These new formulas are not easy to derive, but we don't have to worry about that).

Rule 3.4.1.

[author=garrett, file =text_files/deriv_transcend]
The first two are the essentials for exponential and logarithms:

$$\begin{aligned} \frac{d}{dx} e^x &= e^x \\ \frac{d}{dx} \ln(x) &= \frac{1}{x} \end{aligned}$$

Rule 3.4.2.

[author=garrett, file =text_files/deriv_transcend]
The next three are essential for trig functions:

$$\begin{aligned} \frac{d}{dx} \sin(x) &= \cos(x) \\ \frac{d}{dx} \cos(x) &= -\sin(x) \\ \frac{d}{dx} \tan(x) &= \sec^2(x) \end{aligned}$$

Rule 3.4.3.

[author=garrett, file =text_files/deriv_transcend]
The next three are essential for inverse trig functions

$$\begin{aligned} \frac{d}{dx} \arcsin(x) &= \frac{1}{\sqrt{1-x^2}} \\ \frac{d}{dx} \arctan(x) &= \frac{1}{1+x^2} \\ \frac{d}{dx} \mathrm{arcsec}(x) &= \frac{1}{x\sqrt{x^2-1}} \end{aligned}$$

Comment.

[author=garrett, file =text_files/deriv_transcend]
The previous formulas are the indispensable ones in practice, and are the only ones that I personally remember (if I'm lucky). Other formulas one *might* like to have seen are (with $a > 0$ in the first two):

Rule 3.4.4.

[author=garrett, file =text_files/deriv_transcend]

$$\begin{array}{rcl}
\frac{d}{dx}a^x & = & \ln\, a \cdot a^x \\
\frac{d}{dx}\log_a\, x & = & \frac{1}{\ln\, a \cdot x} \\
\frac{d}{dx}\sec\, x & = & \tan\, x\, \sec\, x \\
\frac{d}{dx}\csc\, x & = & -\cot\, x\, \csc\, x \\
\frac{d}{dx}\cot\, x & = & -\csc^2\, x \\
\frac{d}{dx}\arccos\, x & = & \frac{-1}{\sqrt{1-x^2}} \\
\frac{d}{dx}\text{arccot}\, x & = & \frac{-1}{1+x^2} \\
\frac{d}{dx}\text{arccsc}\, x & = & \frac{-1}{x\sqrt{x^2-1}}
\end{array}$$

Comment.

[author=garrett, file =text_files/deriv_transcend]
(There are always some difficulties in figuring out which of the infinitely-many possibilities to take for the values of the inverse trig functions, and this is especially bad with arccsc*, for example. But we won't have time to worry about such things).*

Comment.

[author=garrett, file =text_files/deriv_transcend]
To be able to use the above formulas it is *not* necessary to know very many *other* properties of these functions. For example, *it is not necessary to be able to graph these functions to take their derivatives!*

Discussion.

[author=wikibooks, file =text_files/derivative_exponentials]
To determine the derivative of an exponent requires use of the symmetric difference equation for determining the derivative

$\frac{d}{dx}f(x) = \lim_{h\to 0} \frac{f(x+h)-f(x-h)}{2h}$

First we will solve this for the specific case of an exponent with a base of e and then extend it to the general case with a base of a where a is a positive real number.

Derivation.

[author=wikibooks, file =text_files/derivative_exponentials]
First we set up our problem using $f(x) = e^x$

$\frac{d}{dx}e^x = \lim_{h\to 0} \frac{e^{x+h}-e^{x-h}}{2h}$

Then we apply some basic algebra with powers (specifically that $a^b + c = a^b a^c$)

$\frac{d}{dx}e^x = \lim_{h\to 0} \frac{e^x e^h - e^x e^{-h}}{2h}$

Treating e^x as a constant with respect to what we are taking the limit of, we can use the limit rules to move it to the outside, leaving us with

$\frac{d}{dx}e^x = e^x \cdot \lim_{h\to 0}\frac{e^h - e^{-h}}{2h}$

A careful examination of the limit reveals a hyperbolic sine

$\frac{d}{dx}e^x = e^x \cdot \lim_{h\to 0}\frac{\sinh(h)}{h}$

Which for very small values of h can be approximated as h, leaving us with

Derivative of the exponential function $\frac{d}{dx}e^x = e^x$ in which $f'(x) = f(x)$.

Derivation.

[`author=wikibooks`, file =`text_files/derivative_exponentials`]
Now that we have derived a specific case, let us extend things to the general case. Assuming that a is a positive real constant, we wish to calculate

$\frac{d}{dx}a^x$

One of the oldest tricks in mathematics is to break a problem down into a form that we already know we can handle. Since we have already determined the derivative of e^x, we will attempt to rewrite a^x in that form.

Using that $e^l n(c) = c$ and that $ln(a^b) = b \cdot ln(a)$, we find that

$a^x = e^{x\cdot\ln(a)}$

Thus, we simply apply the chain rule

$\frac{d}{dx}e^{x\cdot\ln(a)} = \left[\frac{d}{dx}x\cdot\ln(a)\right]e^{x\cdot\ln(a)}$

In which we can solve for the derivative and substitute back with $e^x \cdot ln(a) = a^x$ to get

Derivative of the exponential function' $\frac{d}{dx}a^x = \ln(a)\, a^x$

Derivation.

[`author=wikibooks`, file =`text_files/derivatives_logarithms`]
Closely related to the exponentiation, is the logarithm. Just as with exponents, we will derive the equation for a specific case first (the natural log, where the base is e), and then work to generalize it for any logarithm.

First let us create a variable y such that

$y = \ln(x)$

It should be noted that what we want to find is the derivative of y or $\frac{dy}{dx}$.

Next we will put both sides to the power of e in an attempt to remove the logarithm from the right hand side

$e^y = x$

Now, applying the chain rule and the property of exponents we derived earlier, we take the derivative of both sides

$\frac{dy}{dx}\cdot e^y = 1$

This leaves us with the derivative

$\frac{dy}{dx} = \frac{1}{e^y}$

Substituting back our original equation of $x = e^y$, we find that

Derivative of the Natural Logarithm' $\frac{d}{dx}\ln(x) = \frac{1}{x}$

Derivation.

[author=wikibooks, file =text_files/derivatives_logarithms]
If we wanted, we could go through that same process again for a generalized base, but it is easier just to use properties of logs and realize that

$\log_b(x) = \frac{\ln(x)}{\ln(b)}$

Since 1 / ln(b) is a constant, we can just take it outside of the derivative

$\frac{d}{dx}\log_b(x) = \frac{1}{\ln(b)} \cdot \frac{d}{dx}\ln(x)$

Which leaves us with the generalized form of

Derivative of the Logarithm' $\frac{d}{dx}\log_b(x) = \frac{1}{x\ln(b)}$

Discussion.

[author=wikibooks, file =text_files/derivatives_trig_functions]
Sine, Cosine, Tangent, Cosecant, Secant, Cotangent. These are functions that crop up continuously in mathematics and engineering and have a lot of practical applications. They also appear a lot in more advanced calculus, particularly when dealing with things such as line integrals with complex numbers and alternate representations of space such as spherical coordinates.

Derivation.

[author=wikibooks,uses=complexnumbers, file =text_files/derivatives_trig_functions]
There are two basic ways to determine the derivative of these functions. The first is to sit down with a table of trigonometric identities and work your way through using the formal equation for the derivative. This is tedious and requires either memorizing or using a table with a lot of equations on it. It is far simpler to just use Euler's Formula

Euler's Formula' $e^{i\,x} = \cos(x) + i\,\sin(x)$

Where $i = \sqrt{-1}$.

This leads us to the equations for the sine and cosine

$\sin(x) = \frac{e^{i\,x} - e^{-i\,x}}{2i}$ $\cos(x) = \frac{e^{i\,x} + e^{-i\,x}}{2}$

Using the rules discussed above for exponents, we find that

$\frac{d}{dx}\sin(x) = \frac{i\,e^{i\,x} + i\,e^{-i\,x}}{2i}$ $\frac{d}{dx}\cos(x) = \frac{i\,e^{i\,x} - i\,e^{-i\,x}}{2}$

Which when we simplify them down, leaves us with

'Derivative of Sine and Cosine' $\frac{d}{dx}\sin(x) = \cos(x)$ $\frac{d}{dx}\cos(x) = -\sin(x)$ We use the definition of the derivative, i.e., $f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$, to work these out.

Derivation.

[author=wikibooks, file =text_files/derivatives_trig_functions]
Let us find the derivative of $\sin(x)$, using the above definition.

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

$= \lim_{h\to 0} \frac{\sin(x)\cos h+\cos(x)\sin h-\sin(x)}{h}$

$= \lim_{h\to 0} \frac{\sin(x)(\cos h-1)+\cos(x)\sin h}{h}$

$= \lim_{h\to 0} \frac{\sin(x)(\cos h-1)}{h} + \frac{\cos(x)\sin h}{h}$

$= \lim_{h\to 0} \frac{\sin(x)(\cos h-1)}{h} + \lim_{h\to 0} \frac{\cos(x)\sin h}{h}$

$= 0 + \lim_{h\to 0} \frac{\cos(x)\sin h}{h}$

$= \cos(x)$

Derivation.

[author=wikibooks, file =text_files/derivatives_trig_functions]

To find the derivative of the tangent, we just remember that

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

Which is a quotient. Applying the quotient rule, we get

$\frac{d}{dx}\tan(x) = \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)} = 1 + \tan^2(x) = \frac{1}{\cos^2(x)} = \sec^2(x)$

Derivative of the Tangent $\frac{d}{dx}\tan(x) = \sec^2(x)$

Derivation.

[author=wikibooks, file =text_files/derivatives_trig_functions]
For secants, we just need to apply the chain rule to the derivations we have already determined.

$\sec(x) = \frac{1}{\cos(x)}$

So for the secant, we state the equation as

$\sec(x) = \frac{1}{u}$ $u(x) = \cos(x)$

Take the derivative of both equations, we find

$\frac{d}{dx}\sec(x) = \frac{-1}{u^2} \cdot \frac{du}{dx}$ $\frac{du}{dx} = -\sin(x)$

Leaving us with

$\frac{d}{dx}\sec(x) = \frac{\sin(x)}{\cos^2(x)}$

Simplifying, we get

Derivative of the Secant $\frac{d}{dx}\sec(x) = \sec(x)\tan(x)$

Derivation.

[author=wikibooks, file =text_files/derivatives_trig_functions]
Using the same procedure on cosecants $\csc(x) = \frac{1}{\sin(x)}$

We get

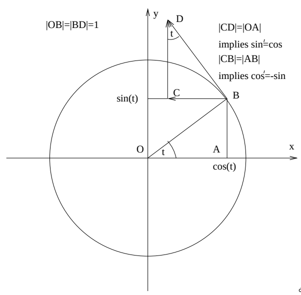
Derivative of the Cosecant $\frac{d}{dx}\csc(x) = -\csc(x)\cot(x)$

Using the same procedure for the cotangent that we used for the tangent, we get

Derivative of the Cotangent $\frac{d}{dx}\cot(x) = -\csc^2(x)$

Discussion.

[author=livshits, file =text_files/derivatives_trig_functions]
Imagine a point on the $x - y$ plane moving around the unit circle with unit speed.



deriv_sin_cos_circle

You can see from the figure that $sin(t)' = cos(t)$ and $cos(t)' = -sin(t)$.

Exercises

1. Find $\frac{d}{dx}(e^{\cos x})$
2. Find $\frac{d}{dx}(\arctan(2 - e^x))$
3. Find $\frac{d}{dx}(\sqrt{\ln\,(x-1)})$
4. Find $\frac{d}{dx}(e^{2\cos x + 5})$
5. Find $\frac{d}{dx}(\arctan(1 + \sin 2x))$
6. Find $\frac{d}{dx}\cos(e^x - x^2)$
7. Find $\frac{d}{dx}\sqrt[3]{1 - \ln\,2x}$
8. Find $\frac{d}{dx}\frac{e^x - 1}{e^x + 1}$
9. Find $\frac{d}{dx}(\sqrt{\ln\,(\frac{1}{x})})$

3.5 Product and quotient rule

Discussion.

[author=garrett, file =text_files/product_rule]
Not only will the *product rule* be of use in general and later on, but it's already helpful in perhaps unexpected ways in dealing with polynomials. Anyway, here's the general rule.

Rule 3.5.1.

[author=garrett, file =text_files/product_rule]
Product Rule

$$\frac{d}{dx}(fg) = f'g + fg'$$

Comment.

[author=garrett, file =text_files/product_rule]
While the product rule is certainly not as awful as the quotient rule just above, it is not as simple as the rule for sums, which was the good-sounding slogan that *the derivative of the sum is the sum of the derivatives.* It is *not* true that the derivative of the product is the product of the derivatives. Too bad. Still, it's not as bad as the quotient rule.

Example 3.5.1.

[author=garrett, file =text_files/product_rule]
One way that the product rule can be useful is in postponing or eliminating a lot of algebra. For example, to evaluate

$$\frac{d}{dx}\left((x^3 + x^2 + x + 1)(x^4 + x^3 + 2x + 1)\right)$$

we *could* multiply out and then take the derivative term-by-term as we did with several polynomials above. This would be at least mildly irritating because we'd have to do a bit of algebra. Rather, just apply the product rule *without* feeling compelled first to do any algebra:

$$\frac{d}{dx}\left((x^3 + x^2 + x + 1)(x^4 + x^3 + 2x + 1)\right)$$

$$= (x^3 + x^2 + x + 1)'(x^4 + x^3 + 2x + 1) + (x^3 + x^2 + x + 1)(x^4 + x^3 + 2x + 1)'$$

$$= (3x^2 + 2x + 1)(x^4 + x^3 + 2x + 1) + (x^3 + x^2 + x + 1)(4x^3 + 3x^2 + 2)$$

Now if we were somehow still obliged to multiply out, then we'd still have to do some algebra. But *we can take the derivative without multiplying out*, if we want to, by using the product rule.

Comment.

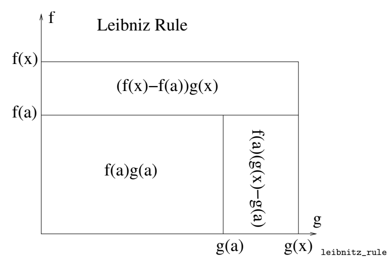
[author=garrett, file =text_files/product_rule]
For that matter, once we see that there is a *choice* about doing algebra either *before* or *after* we take the derivative, it might be possible to make a choice which minimizes our computational labor. This could matter.

Rule 3.5.2.

[author=livshits, file =text_files/product_rule]
Product or Leibniz Rule: $(fg)' = f'g + fg'$

Derivation.

[author=livshits, file =text_files/product_rule]
The product rule looks a little strange, here is the derivation of it: $(f(x)g(x) - f(a)g(a))/(x-a) = (f(x) - f(a))g(x)/(x-a) + f(a)(g(x) - g(a))/(x-a)$, both summands on the right of the $=$ sign make sense for $x = a$, the first summand becomes $f'(a)g(a)$, the second one becomes $f(a)g'(a)$.



Discussion.

[author=garrett, file =text_files/quotient_rule]
The quotient rule is one of the more irritating and goofy things in elementary calculus, but it just couldn't have been any other way.

Rule 3.5.3.

[author=garrett, file =text_files/quotient_rule]
Quotient Rule:

$$\frac{d}{dx}\left(\frac{f}{g}\right) = \frac{f'g - g'f}{g^2}$$

Comment.

[author=garrett, file =text_files/quotient_rule]
The main hazard is remembering that the numerator is as it is, rather than accidentally reversing the roles of f and g, and then being off by $\pm$, which could be fatal in real life.

Example 3.5.2.

[author=garrett, file =text_files/quotient_rule]

$$\frac{d}{dx}\left(\frac{1}{x-2}\right) = \frac{\frac{d}{dx}1\cdot(x-2) - 1\cdot\frac{d}{dx}(x-2)}{(x-2)^2} = \frac{0\cdot(x-2)-1\cdot 1}{(x-2)^2} = \frac{-1}{(x-2)^2}$$

Example 3.5.3.

[author=garrett, file =text_files/quotient_rule]

$$\frac{d}{dx}\left(\frac{x-1}{x-2}\right) = \frac{(x-1)'(x-2)-(x-1)(x-2)'}{(x-2)^2} = \frac{1\cdot(x-2)-(x-1)\cdot 1}{(x-2)^2}$$
$$= \frac{(x-2)-(x-1)}{(x-2)^2} = \frac{-1}{(x-2)^2}$$

Example 3.5.4.

[author=garrett, file =text_files/quotient_rule]

$$\frac{d}{dx}\left(\frac{5x^3+x}{2-x^7}\right) = \frac{(5x^3+x)'\cdot(2-x^7)-(5x^3+x)\cdot(2-x^7)}{(2-x^7)^2}$$
$$= \frac{(15x^2+1)\cdot(2-x^7)-(5x^3+x)\cdot(-7x^6)}{(2-x^7)^2}$$

and there's hardly any point in simplifying the last expression, unless someone gives you a good reason. In general, it's not so easy to see how much may or may not be gained in 'simplifying', and we won't make ourselves crazy over it.

Example 3.5.5.

[author=livshits, file =text_files/quotient_rule]
$(f(x)/g(x)-f(a)/g(a))/(x-a)|_{x=a} = [(f(x)/g(x)-f(x)/g(a))+(f(x)-f(a))/g(a)]/(x-a)|_{x=a} =$

$$= (f(x)/(g(x)g(a)))(g(a)-g(x))/(x-a)|_{x=a}+(f(x)-f(a))/(x-a)/g(a)|_{x=a} =$$
$$= -f(x)g'(x)/(g(x))^2 + f''(x)/g(x) = (f''(x)g(x) - f(x)g'(x))/(g(x)^2)$$

Discussion.

[author=wikibooks, file =text_files/product_quotient_rules]
When we wish to differentiate a more complicated expression such as $h(x) = (x^2+5)^5\cdot(x^3+2)^3$ our only resort (so far) is to expand and get a messy polynomial, and then differentiate the polynomial. This can get very ugly very quickly and is particularly error prone when doing such calculations by hand. It would be nice if we could just take the derivative of h(x) using just the functions $f(x) = (x^2+5)^5$

and $g(x) = (x^3 + 2)^3$ and their derivatives.

Rule 3.5.4.

[`author=wikibooks`, file =`text_files/product_quotient_rules`]
Product rule $\frac{d}{dx}[f(x) \cdot g(x)] = f'(x) \cdot g(x) + f(x) \cdot g'(x)$

Proving this rule is relatively straightforward, first let us state the equation for the derivative

$$\frac{d}{dx}[f(x) \cdot g(x)] = \lim_{h \to 0} \frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x)}{h}.$$

We will then apply one of the oldest tricks in the book – adding a term that cancels itself out to the middle

$$\frac{d}{dx}[f(x) \cdot g(x)] = \lim_{h \to 0} \frac{f(x+h) \cdot g(x+h) - \mathbf{f(x) \cdot g(x+h)} + \mathbf{f(x) \cdot g(x+h)} - f(x) \cdot g(x)}{h}.$$

Notice that those terms sum to zero, and so all we have done is add 0 to the equation.

Now we can split the equation up into forms that we already know how to solve

$$\frac{d}{dx}[f(x) \cdot g(x)] = \lim_{h \to 0} \left[\frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x+h)}{h} + \frac{f(x) \cdot g(x+h) - f(x) \cdot g(x)}{h} \right].$$

Looking at this, we see that we can separate the common terms out of the numerators to get

$$\frac{d}{dx}[f(x) \cdot g(x)] = \lim_{h \to 0} \left[g(x+h) \frac{f(x+h) - f(x)}{h} + f(x) \frac{g(x+h) - g(x)}{h} \right].$$

Which, when we take the limit, turns into

$$\frac{d}{dx}[f(x) \cdot g(x)] = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

One mnemonic for this is "one D-two plus two D-one"

This can be extended to 3 functions $D[fgh] = f(x)g(x)h'(x) + f(x)g'(x)h(x) + f'(x)g(x)h(x)$ For any number of functions, the derivative of their product is the sum, for each function, of its derivative times each other function.

Derivation.

[`author=wikibooks`, file =`text_files/product_quotient_rules`]
Quotient rule For quotients, where one function is divided by another function, the equation is more complicated but it is simply a special case of the product rule.

$$\frac{f(x)}{g(x)} = f(x) \cdot g(x)^{-1}.$$

Then we can just use the product rule and the chain rule

$$\frac{d}{dx} \frac{f(x)}{g(x)} = f'(x) \cdot g(x)^{-1} - f(x) \cdot g'(x) \cdot g(x)^{-2}.$$

We can then multiply through by 1, or more precisely $g(x)^2/g(x)^2$, which cancels out into 1, to get $\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{f'(x) \cdot g(x)}{g(x)^2} - \frac{f(x) \cdot g'(x)}{g(x)^2}$ This leads us to the so-called "quotient rule"

Rule 3.5.5.

[author=wikibooks, file =text_files/product_quotient_rules]
Quotient Rule

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)\cdot g(x) - f(x)\cdot g'(x)}{g(x)^2}.$$

Which some people remember with the mnemonic "low D-high minus high D-low over the square of what's below."

Comment.

[author=wikibooks, file =text_files/product_quotient_rules]
Remember the derivative of a product/quotient "is not' the product/quotient of the derivatives. (That is, differentiation does not distribute over multiplication or division.) However one can distribute before taking the derivative. That is $\frac{d}{dx}\left((a+b)\times(c+d)\right) \equiv \frac{d}{dx}\left(ac+ad+bc+bd\right)$

Comment.

[author=duckworth, file =text_files/product_quotient_rules]
So, we do *not* usually have $\frac{d}{dx}f(x)g(x) = f'(x)g'(x)$. If allow some curiosity into the discussion this leads to two questions: (1) When *do* we have $\frac{d}{dx}f(x)g(x) = f'(x)g'(x)$, i.e. for which functions f and g would this be true? (2) When is a product of derivatives equal to the derivative of something?

Exercises

1. Find $\frac{d}{dx}(x^3-1)(x^6+x^3+1))$
2. Find $\frac{d}{dx}(x^2+x+1)(x^4-x^2+1)$.
3. Find $\frac{d}{dx}(x^3+x^2+x+1)(x^4+x^2+1))$
4. Find $\frac{d}{dx}(x^3+x^2+x+1)(2x+\sqrt{x}))$
5. Find $\frac{d}{dx}(\frac{x-1}{x-2})$
6. Find $\frac{d}{dx}(\frac{1}{x-2})$
7. Find $\frac{d}{dx}(\frac{\sqrt{x}-1}{x^2-5})$
8. Find $\frac{d}{dx}(\frac{1-x^3}{2+\sqrt{x}})$

3.6 Chain rule

Discussion.

[author=garrett, file =text_files/chain_rule]
The *chain rule* is subtler than the previous rules, so if it *seems* trickier to you, then you're right. OK. But it is absolutely indispensable in general and later, and already is very helpful in dealing with polynomials.

The general assertion may be a little hard to fathom because it is of a different nature than the previous ones. For one thing, now we will be talking about a *composite function* instead of just adding or multiplying functions in a more ordinary way.

Rule 3.6.1.

[author=garrett, file =text_files/chain_rule]
So, for two functions f and g,

$$\frac{d}{dx}((f(g(x))) = f'(g(x)) \cdot g'(x)$$

There is also the standard notation

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

for this *composite function*, but using this notation doesn't accomplish so very much.

Comment.

[author=garrett, file =text_files/chain_rule]
A problem in successful use of the chain rule is that often it requires a little thought to recognize that some formula *is* (or can be *looked at as*) a composite function. And the very nature of the chain rule picks on weaknesses in our understanding of the notation. For example, the function

---

Example 3.6.1.

[author=garrett, file =text_files/chain_rule]

$$F(x) = (1 + x^2)^{100}$$

is really obtained by *first* using x as input to the function which *squares and adds* 1 to its input. Then the result of that is used as input to the function which *takes the* 100*th power*. It is necessary to think about it this way or we'll make a mistake. The derivative is evaluated as

$$\frac{d}{dx}(1 + x^2)^{100} = 100(1 + x^2)^{99} \cdot 2x$$

To see that this is a special case of the general formula, we need to see what corresponds to the f and g in the general formula. Specifically, let

$$f(\text{input}) = (\text{input})^{100}$$

$$g(\text{input}) = 1 + (\text{input})^2$$

The reason for writing 'input' and not 'x' for the moment is to avoid a certain kind of mistake. But we can compute that

$$f'(\text{input}) = 100(\text{input})^{99}$$

$$g'(\text{input}) = 2(\text{input})$$

The hazard here is that *the input to f is not x, but rather is $g(x)$*. So the general formula gives

$$\frac{d}{dx}(1+x^2)^{100} = f'(g(x)) \cdot g'(x) = 100g(x)^{99} \cdot 2x = 100(1+x^2)^{99} \cdot 2x$$

Examples 3.6.2.

[author=garrett, file =text_files/chain_rule]
More examples:

$$\frac{d}{dx}\sqrt{3x+2} = \frac{d}{dx}(3x+2)^{1/2} = \frac{1}{2}(3x+2)^{-1/2} \cdot 3$$

$$\frac{d}{dx}(3x^5 - x + 14)^{11} = 11(3x^5 - x + 14)^{10} \cdot (15x^4 - 1)$$

Example 3.6.3.

[author=garrett, file =text_files/chain_rule]
It is very important to recognize situations like

$$\frac{d}{dx}(ax+b)^n = n(ax+b)^{n-1} \cdot a$$

for any constants a, b, n. And, of course, this includes

$$\frac{d}{dx}\sqrt{ax+b} = \frac{1}{2}(ax+b)^{-1/2} \cdot a$$

$$\frac{d}{dx}\frac{1}{ax+b} = -(ax+b)^{-2} \cdot a = \frac{-a}{(ax+b)^2}$$

Example 3.6.4.

[author=garrett, file =text_files/chain_rule]
Of course, this idea can be combined with polynomials, quotients, and products to give enormous and excruciating things where we need to use the chain rule, the quotient rule, the product rule, etc., and possibly several times each. But this is not *hard*, merely *tedious*, since the only things we really do come in small steps. For example:

$$\frac{d}{dx}\left(\frac{1+\sqrt{x+2}}{(1+7x)^{33}}\right) = \frac{(1+\sqrt{x+2})' \cdot (1+7x)^{33} - (1+\sqrt{x+2}) \cdot ((1+7x)^{33})'}{((1+7x)^{33})^2}$$

by the quotient rule, which is then

$$\frac{(\frac{1}{2}(x+2)^{-1/2})\cdot(1+7x)^{33}-(1+\sqrt{x+2})\cdot((1+7x)^{33})'}{((1+7x)^{33})^2}$$

because our observations just above (*chain rule!*) tell us that

$$\frac{d}{dx}\sqrt{x+2}=\frac{1}{2}(x+2)^{-1/2}\cdot(x+2)'=\frac{1}{2}(x+2)^{-1/2}$$

Then we use the chain rule *again* to take the derivative of that big power of $1+7x$, so the whole thing becomes

$$\frac{(\frac{1}{2}(x+2)^{-1/2})\cdot(1+7x)^{33}-(1+\sqrt{x+2})\cdot(33(1+7x)^{32}\cdot 7)}{((1+7x)^{33})^2}$$

Although we *could* simplify a bit here, let's not. The point about having to do several things in a row to take a derivative is pretty clear without doing algebra just now.

Discussion.

[**author=wikibooks**, file =**text_files/chain_rule**]
We know how to differentiate regular polynomial functions. For example $\frac{d}{dx}(3x^3-6x^2+x)=9x^2-12x+1$ However, we've not yet explored the derivative of an unexpanded expression. If we are given the function $y=(x+5)^2$, we currently have no choice but to expand it $y=x^2+10x+25$ $f'(x)=2x+10$ However, there is a useful rule known as the "chain rule'. The function above ($y=(x+5)^2$) can be consolidated into $y=u^2$, where $u=(x+5)$. Therefore $y=f(u)=u^2$ $u=g(x)=x+5$ Therefore $y=f(g(x))$

Rule 3.6.2.

[**author=wikibooks**, file =**text_files/chain_rule**]
The chain rule states the following, in the situation described above 'Chain Rule' $\frac{dy}{dx}=\frac{dy}{du}\cdot\frac{du}{dx}$

Example 3.6.5.

[**author=wikibooks**, file =**text_files/chain_rule**]
We can now investigate the original function $\frac{dy}{dx}=2u\cdot 1$ $\frac{dy}{dx}=2(x+5)=2x+10$

Example 3.6.6.

[**author=wikibooks**, file =**text_files/chain_rule**]
This can be performed for more complicated equations. If we consider $\frac{d}{dx}\sqrt{1+x^2}$ and let $y=\sqrt{u}$ and $u=1+x^2$, so that $dy/du=1/2\sqrt{u}$ and $du/dx=2x$, then, by applying the chain rule, we find that $\frac{d}{dx}\sqrt{1+x^2}=\frac{1}{2}\sqrt{1+x^2}\cdot 2x=\frac{x}{\sqrt{1+x^2}}$

Rule 3.6.3.

[`author=livshits`, file =`text_files/chain_rule`]
Chain Rule: $(f(g(x)))' = f'(g(x))g'(x)$

In Leibniz notation it becomes $df/dx = (df/dg)(dg/dx)$, so it looks like dg just cancels out. To demonstrate the formula we notice that $f(y)-f(b) = (y-b)p(y,b)$ (because f is differentiable). By taking $y = g(x)$ and $b = g(a)$ we get $f(g(x)) - f(g(a)) = (g(x) - g(a))p(g(x), g(a))$, where $p(g(a), g(a)) = f'(g(a))$. On the other hand, $g(x) - g(a) = (x - a)r(x, a)$ where $r(a, a) = g'(a)$. Putting it all together and taking $x = a$ gives the formula we wanted.

Exercises

1. Find $\frac{d}{dx}((1 - x^2)^{100})$
2. Find $\frac{d}{dx}\sqrt{x - 3}$
3. Find $\frac{d}{dx}(x^2 - \sqrt{x^2 - 3})$
4. Find $\frac{d}{dx}(\sqrt{x^2 + x + 1})$
5. Find $\frac{d}{dx}(\sqrt[3]{x^3 + x^2 + x + 1})$
6. Find $\frac{d}{dx}((x^3 + \sqrt{x + 1})^{10})$

3.7 Hyperbolic functions

Definition 3.7.1.

[author=wikibooks, file =text_files/hyperbolics]
The hyperbolic functions are defined in analogy with the trigonometric functions

$$\begin{aligned}\sinh x &= \frac{1}{2}(e^x - e^{-x})\\ \cosh x &= \frac{1}{2}(e^x + e^{-x})\\ \tanh x &= \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{\sinh x}{\cosh x}\end{aligned}$$

The reciprocal functions cosech, sech, coth are defined from these functions.

Facts.

[author=wikibooks, file =text_files/hyperbolics]
The hyperbolic trigonometric functions satisfy identities very similar to those satisfied by the regular trigonometric functions.

$$\begin{aligned}\cosh^2 x - \sinh^2 x &= 1\\ 1 - \tanh^2 x &= \overset{2}{\operatorname{sech}} x\\ \sinh 2x &= \sinh x \cosh x\\ \cosh 2x &= \cosh^2 x + \sinh^2 x\end{aligned}$$

Rules 3.7.1.

[author=wikibooks, file =text_files/hyperbolics]
The hyperbolic trigonometric functions have very similar derivative rules as the regular trigonometric functions.

$$\begin{aligned}\frac{d}{dx}\sinh x &= \cosh x\\ \frac{d}{dx}\cosh x &= \sinh x\\ \frac{d}{dx}\tanh x &= \overset{2}{\operatorname{sech}} x\\ \frac{d}{dx}\operatorname{cosech} x &= -\operatorname{cosech} x \coth x\\ \frac{d}{dx}\operatorname{sech} x &= -\operatorname{sech} x \tanh x\\ \frac{d}{dx}\coth x &= \overset{2}{\operatorname{cosech}} x\end{aligned}$$

Definition 3.7.2.

[author=wikibooks, file =text_files/hyperbolics]
We define inverse functions for the hyperbolic functions. As with the usual trigonometric functions, we sometimes need to restrict the domain to obtain a function which is one-to-one.

- $\sinh(x)$ is one-to-one on the whole real number line, and its range is the whole real number line. Therefore, $\sinh^{-1}$ is defined on the whole real number line. The formula is given by $\sinh^{-1} = \ln(z + \sqrt{z^2+1})$.
- $\cosh(x)$ is one-to-one on the domain $[0, \infty)$. It's range is $[1, \infty)$. Therefore $\cosh^{-1}$ is defined on the interval $[1, \infty)$. The formula is given by $\cosh^{-1} z = \ln(z + \sqrt{z^2-1})$.
- $\tanh(x)$ is one-to-one on the whole real number line and it's range is the interval $(-1, 1)$. Thefore $\tanh^{-1}$ is defined on the interval $(-1, 1)$. The formula is given by $\tanh^{-1} z = \ln \sqrt{\frac{1+z}{1-z}}$.

Rules 3.7.2.

[author=wikibooks, file =text_files/hyperbolics]
Here are the derivative rules for the inverse hyperbolic trigonometric functions.

- $\frac{d}{dx} \sinh^{-1}(x) = \frac{1}{\sqrt{1+x^2}}$.
- $\frac{d}{dx} \cosh^{-1}(x) = \frac{1}{\sqrt{x^2-1}}$, $x > 1$.
- $\frac{d}{dx} \tanh^{-1}(x) = \frac{1}{\sqrt{1-x^2}}$, $-1 < x < 1$.
- $\frac{d}{dx} \operatorname{cosech}^{-1}(x) = -\frac{1}{|x|\sqrt{1+x^2}}$, $x \neq 0$.
- $\frac{d}{dx} \operatorname{sech}^{-1}(x) = -\frac{1}{x\sqrt{1-x^2}}$, $0 < x < 1$.
- $\frac{d}{dx} \coth^{-1}(x) = \frac{1}{1-x^2}$, $|x| > 1$.

3.8 Tangent and Normal Lines

Comment.

[author=garrett, file =text_files/tangent_normal_lines]
One fundamental interpretation of the derivative of a function is that it is the *slope of the tangent line* to the graph of the function. (Still, it is important to realize that this is *not the definition* of the thing, and that there are other possible and important interpretations as well).

The precise statement of this fundamental idea is as follows. Let f be a function. For each *fixed* value x_o of the input to f, the value $f'(x_o)$ of the derivative f' of f *evaluated at* x_o is the slope of the tangent line to the graph of f *at the particular point* $(x_o, f(x_o))$ *on the graph.*

Rule 3.8.1.

[**author=garrett**, file =**text_files/tangent_normal_lines**]
Recall the *point-slope form* of a line with slope m through a point (x_o, y_o):

$$y - y_o = m(x - x_o)$$

In the present context, the slope is $f'(x_o)$ and the point is $(x_o, f(x_o))$, so *the equation of the tangent line to the graph of* f *at* $(x_o, f(x_o))$ *is*

$$y - f(x_o) = f'(x_o)(x - x_o)$$

Rule 3.8.2.

[**author=garrett**, file =**text_files/tangent_normal_lines**]
The **normal line** to a curve at a particular point is the line through that point and *perpendicular* to the tangent. A person might remember from analytic geometry that the slope of any line *perpendicular to* a line with slope m is the *negative reciprocal* $-1/m$. Thus, just changing this aspect of the equation for the tangent line, we can say generally that *the equation of the normal line to the graph of* f *at* $(x_o, f(x_o))$ *is*

$$y - f(x_o) = \frac{-1}{f'(x_o)}(x - x_o)$$

The main conceptual hazard is to mistakenly name the *fixed* point 'x', as well as naming the *variable* coordinate on the tangent line 'x'. This causes a person to write down some equation which, whatever it may be, is *not* the equation of a line at all.

Another popular boo-boo is to forget the subtraction $-f(x_o)$ on the left hand side. Don't do it.

Example 3.8.1.

[**author=garrett**, file =**text_files/tangent_normal_lines**]
So, as the simplest example: let's write the equation for the tangent line to the curve $y = x^2$ at the point where $x = 3$. The derivative of the function is $y' = 2x$, which has value $2 \cdot 3 = 6$ when $x = 3$. And the value of the function is $3 \cdot 3 = 9$ when $x = 3$. Thus, the *tangent line* at that point is

$$y - 9 = 6(x - 3)$$

The *normal line* at the point where $x = 3$ is

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

So the question of finding the tangent and normal lines at various points of the graph of a function is just a combination of the two processes: computing the derivative at the point in question, and invoking the point-slope form of the equation for a straight line.

Exercises

1. Write the equation for both the *tangent line* and *normal line* to the curve $y = 3x^2 - x + 1$ at the point where $x = 1$.
2. Write the equation for both the *tangent line* and *normal line* to the curve $y = (x - 1)/(x + 1)$ at the point where $x = 0$.

3.9 End of chapter problems

Exercises

1. Derive multiplier rule from the Leibniz rule.
2. Find the formulas for $(1/f)'$ and $(g/f)'$ using Leibniz rule (Hint: differentiate the identity $(1/f)f = 1$ and solve for $(1/f)'$).
3. To make our guess a theorem we observe that every time we turn the crank to get from $(x^n)'$ to $(x^{(n+1)})'$ the pattern persists (exercise: check it).
4. Write x^8 as $((x^2)^2)^2$ and use the chain rule 2 times to get $(x^8)'$. Differentiate x^{81} using a similar approach.
5. Use the chain rule to get an easy solution for ex.1.6
6. Use the fact that $(x^{1/7})^7{=}x$ and the chain rule to get $(x^{1/7})'$.
7. Differentiate some polynomials using the differentiation rules.
8. Do the calculations (Hint: use the chain rule to get $d(x(t)^5)/dt$)
9. Redo problem 1.3 without solving for y (Hint: go implicit).
10. $sin' = cos$ (see section 2.4 for details). Compute $arcsin'$ (Hint: go implicit, starting from $sin(arcsin(x)) = x$ and use $sin^2 + cos^2 = 1$).
11. Differentiate everything that moves to get more practice.
12. For some f see how $q(x, a) = (f(x) - f(a))/(x - a)$ behaves when $x - a$ gets small.
13. As in the example **??**, More generally, the velocity at time t will be $32t$ (exercise)
14. Differentiate x^3, x^5, x^6, x^n, c (= a constant).
15. Differentiate $x^{1/3}$, $x^{1/5}$, $x^{1/7}$, $x^{1/n}$.
16. Find the slope of the tangent to the unit circle at the point $(a, \sqrt{(}1 - a^2))$. Hint: the equation of the unit circle is $x^2 + y^2 = 1$.
17. Differentiate $x^{(m/n)}$. Guess the formula for $(x^b)'$, b real.
18. Give an argument that $(f + g)' = f' + g'$ and for any constant c $(cf)' = cf'$.
19. Differentiate $(1 + x)^7$ and find a neat formula for the answer.

Differentiate the following functions

20. $x^4 + 4x^4 - 5x^3 + x + 1$ Use the constant multiplier rule, sums rule and the formula $dx^n/dx = nx^{n-1}$ $20x^3 + 15x^2 + 1$
21. $(x^2 + 3x + 2)^{10}$ use chain rule $10(x^2 + 3x + 2)^9(2x + 3)$
22. $[(x^3 + 2x + 1)^6 + (x^5 + x^3 + 2)^5]^{10}$

    use chain rule

    $10[(x^3 + 2x + 1)^6 + (x^5 + x^3 + 2)^5]^9[6(x^3 + 2x + 1)^5(3x^2 + 2) + 5(x^5 + x^3 + 2)^4(5x^4 + 3x^2)]$

23. $(3x^3 + 5x + 2)(7x^8 + 5x + 5)$

    Use the product rule

    $(9x^2 + 5)(7x^8 + 5x + 5) + (3x^3 + 5x + 2)(56x^7 + 5)$

24. $(5x^7 + 3)/(8x^9 - 3x - 1)$

    Use the quotient rule

    $(35x^6(8x^9 + 5x + 5) - (5x^7 + 3)(72x^9 + 5))/(8x^9 + 5x + 1)^2$

25. $(x^3 + 1)\sqrt{(x)}$

    Product rule

    $3x^2\sqrt{x} + (x^3 + 1)/(2\sqrt{x})$

26. Suppose that x and t satsfy the equation $x^7 + x^3 + 3t^4 + 2t + 1 = 0$ find a formula for dx/dt in terms of x and t.

    Use implicit differentiation (differentiate the equation with respect to t)

    You get $(7x^6+3x^2)(dx/dt)+12t^3+2 = 0$, so $dx/dt = -(12t^2+2)/(7x^6+3x^3)$

27. Use implicit differentiation to derive the formula for $d(x^{p/q})/dx$ where p and q are integers.

    Think about $(x^{p/q})^q = x^p$

    By differentiating the equation $(x^{p/q})^q = x^p$ we get $q(x^{p/q})^{q-1}(x^{p/q})' = px^{p-1}$ which gives us $(x^{p/q})' = (p/q)x^{p-1-(q-1)p/q} = (p/q)x^{(p/q)-1}$

28. The area of a disc of radius r is given by the formula $A(r) = \pi r^2$ Does the derivative $A'(r)$ remind you of anything? Can you see any geometric meaning of it?

29. The volume of a 3-dimensional ball of radius r is given by the formula $V(r) = 4\pi r^3/3$ Does the derivative $V'(r)$remind you of anything? Can you see any geometric meaning of it?

30. Derive a formula for $(f(x)g(x)h(x))'$

    Use the product rule twice

    $f(x)g(x)h(x) = f(x)(g(x)h(x))$, so $f(x)(g(x)h(x))' = f'(x)(g(x)h(x)) + f(x)(g(x)h(x))' = = f'(x)g(x)h(x) + f(x)g'(x)h(x) + f(x)g(x)h'(x)$

31. Try to generalize the previous problem to a product of more than 3 functions

Find the derivatives of the following functions

32. $tan(x) = sin(x)/cos(x)$

    Use the quotient rule and then some trig identities to simplify.

    $tan' = (sin/cos)' = (sin'cos - cos'sin)/cos^2 = (cos^2 + sin^2)/cos^2 = 1/cos^2$

    2. Differentiate:

33. $sin(5x)$

    Chain rule

    $5cos(5x)$

34. $cos(x^3)$

    Chair rule

    $-sin(x^3)3x^2$

35. $(sin(x - 2) + 3cos(x^2))^3$

    Chain rule

    $3(sin(x - 2) + 3cos(x^2))^2(cos(x - 2) - 3sin(x^2)2x)$

36. $ln(x^3+3)$

    Chain rule

    $(1/(x^3+3))3x^2$

37. $(1+ln(x^2+1))cos(x^3)$

    Product rule then chain rule

    $(2x/(x^2+1))cos(x^3)-(1+ln(x^2+1))sin(x^3)3x^2$

38. e^{10x}

    Chain rule

    $10e^{10x}$

39. $\exp(x^3+\sin(x))$

    Chain rule

    $\exp(x^3+\sin(x))(3x^2+\cos(x))$

40. Find a solution to the equation $y''=-y$

    Think of trig functions

    $\sin(x)$ or $\cos(x)$ or $a\sin(x)+b\cos(x)$ with any constants a and b or $Asin(x+a)$ with any constants A and a.

41. Find a solution to the equation $y''=-y$ such that $y(0)=1$ and $y'(0)=2$.

    Find a multiple of sin plus a multiple of cos which satisfy the extra conditions

    $cos(x)+2sin(x)$

    Find the following integrals.

42. $\int e^{5x}\,dx$

    U-subst

    $(e^{5x}/5)+C$

43. $\int xe^{-x^2}dx$

    U-subst

    $\int xe^{-x^2}dx=-(1/2)\int e^{-x^2}d(-x^2)=-e^{-x^2}/2\ +C$

44. $\int sin(x^2)2x\,dx$

    U-subst

    $\int sin(x^2)d(x^2)=-cos(x^2)+C$

45. $\int dx/(x+1)$

    U-subst

    $ln|1+x|+C$

46. $\int x^2dx/(x^3+3)$

    U-subst

    $(1/3)\int d(x^3+3)/(x^3+3)=(1/3)ln|x^3+3|+C$

47. $\int x^3e^xdx$

    Integration by parts, three times

    $\int x^3(e^3)'dx = x^3e^x-\int(x^3)'e^xdx = x^3e^x-3\int x^2e^xdx$, so the power of x drops by 1, integrate by parts 2 times more to get the power of x down to 0.

48. $\int e^{x+e^x} dx$

    Try $U = e^x$

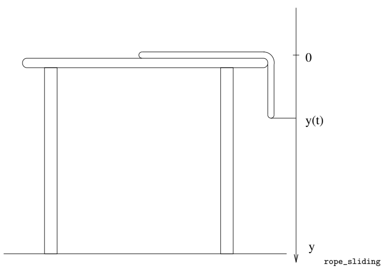
    $\int e^x e^{e^x} dx = \int e^{e^x} (e^x)' dx = \int e^{e^x} d(e^x) = e^{e^x} + C$

49. $\int x^2 cos(x) dx$

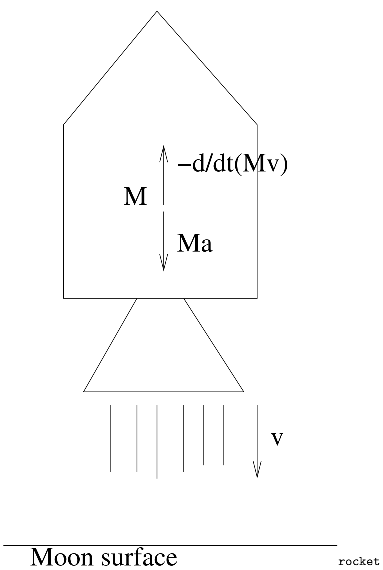
    Integrate by parts 2 times, $cos = sin'$ etc.

    $\int x^2 sin'(x) dx = x^2 sin(x) - \int (x^2)' sin(x) dx = x^2 sin(x) - 2 \int x sin(x) dx$ integrate by parts once more to get the power of $x$down to 0 ($sin = -cos'$).

h these pic-

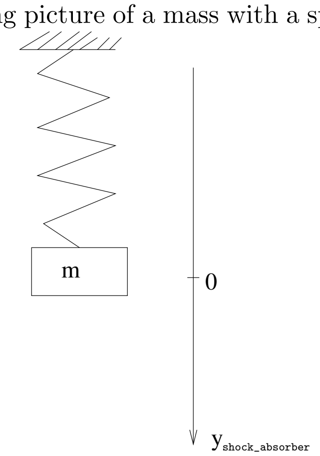
50. 

    A rope sliding off a table

51. 

    A lunar landing module

52. Consider the following picture of a mass with a spring and a shock absorber.



Chapter 4

Applications of Derivatives

Discussion.

[author=wikibooks, file =text_files/intro_to_applications_of_derivatives]
Calculus Differentiation Basic Applications

One of the most useful applications of differentiation is the determination of local extrema of a function. The derivative of a function at a local minimum or maximum is zero, as the slope changes from negative to positive or positive to negative, respectively. Specifically, you seperate the domain of the function into ranges seperated by the points where the derivative is zero, and evaluate the derivative at a point in each of the ranges, determining whether it is positive or negative. If, between any two ranges, the derivative changes from positive to negative, that point is a maximum. If it goes from negative to positive, it is a minimum. Any point where the derivative is zero is called a critical number. The local minimum can be defined as the lowest point on a graph relative to its surroundings. The local maximum can be defined as the highest point on a graph relative to its surroundings.

In physics, the derivative of a function giving position at a given point is the instataneous velocity at that point. The derivative of a function giving velocity at a given point is the instataneous acceleration at that point.

The second derivative of a function can be used to determine the concavity of a function, or, specifically, points at which a function's concavity changes (concavity refers to how the graph is shaped it is concave up if the function curves like the letter "U", and concave down if it is more like a lower case "n"). These places where concavity changes are called points of inflection, but they are only as such if concavity actually changes there. Whether or not it does can be discovered by a sign test, similar to that used for critical numbers. Additionally, the nature of critical numbers can be determined using the second derivative. If the second derivative evaluated at a critical number is positive, then the critical number is a minimum, and it is a maximum if the second derivative is negative at this point (if it is zero, then the critical number is not an extremum but a point of inflection).

4.1 Critical points, monotone increase and decrease

Definition 4.1.1.

[author=garrett, file =text_files/derivs_and_graphs]
A function is called **increasing** if it increases as the input x moves from left to right, and is called **decreasing** if it decreases as x moves from left to right.

Comment.

[author=garrett, file =text_files/derivs_and_graphs]
Of course, a function can be increasing in some places and decreasing in others: that's the complication.

We can notice that a function is increasing if the slope of its tangent is positive, and decreasing if the slope of its tangent is negative. Continuing with the idea that the slope of the tangent is the derivative: *a function is increasing where its derivative is positive, and is decreasing where its derivative is negative.*

This is a great principle, because we don't have to graph the function or otherwise list lots of values to figure out where it's increasing and decreasing. If anything, it should be a big help in graphing to know in advance where the graph goes up and where it goes down.

Definition 4.1.2.

[author=garrett, file =text_files/derivs_and_graphs]
And the points where the tangent line is horizontal, that is, where the derivative is zero, are **critical points**. The points where the graph has a *peak* or a *trough* will certainly lie among the critical points, although there are other possibilities for critical points, as well.

Rule 4.1.1.

[author=garrett, file =text_files/derivs_and_graphs]
Further, for the kind of functions we'll deal with here, there is a fairly systematic way to get all this information: to find the intervals of increase and decrease of a function f:

- Compute the derivative f' of f, and *solve* the equation $f'(x) = 0$ for x to find all the critical points, which we list in order as $x_1 < x_2 < \ldots < x_n$.
- (If there are points of discontinuity or non-differentiability, these points should be added to the list! But points of discontinuity or non-differentiability are *not* called *critical points*.)
- We need some *auxiliary points*: To the left of the leftmost critical point x_1 pick any convenient point t_o, between each pair of consecutive critical points x_i, x_{i+1} choose any convenient point t_i, and to the right of the rightmost critical point x_n choose a convenient point t_n.
- Evaluate the *derivative* f' at all the *auxiliary* points t_i.

- Conclusion: if $f'(t_{i+1}) > 0$, then f is *increasing* on (x_i, x_{i+1}), while if $f'(t_{i+1}) < 0$, then f is *decreasing* on that interval.
- Conclusion: on the 'outside' interval $(-\infty, x_o)$, the function f is *increasing* if $f'(t_o) > 0$ and is *decreasing* if $f'(t_o) < 0$. Similarly, on (x_n, ∞), the function f is *increasing* if $f'(t_n) > 0$ and is *decreasing* if $f'(t_n) < 0$.

Comment.

[author=garrett, file =text_files/derivs_and_graphs]
It is certainly true that there are many possible shortcuts to this procedure, especially for polynomials of low degree or other rather special functions. However, if you are able to quickly compute values of (derivatives of!) functions on your calculator, you may as well use this procedure as any other.

Exactly which *auxiliary points* we choose does not matter, as long as they fall in the correct intervals, since we just need a single sample on each interval to find out whether f' is positive or negative there. Usually we pick integers or some other kind of number to make computation of the derivative there as easy as possible.

It's important to realize that even if a question does not directly ask for *critical points*, and maybe does not ask about *intervals* either, still it is *implicit* that we have to find the critical points and see whether the functions is increasing or decreasing on the *intervals between critical points*. Examples:

Example 4.1.1.

[author=garrett, file =text_files/derivs_and_graphs]
Find the critical points and intervals on which $f(x) = x^2 + 2x + 9$ is increasing and decreasing: Compute $f'(x) = 2x + 2$. Solve $2x + 2 = 0$ to find only one critical point -1. To the left of -1 let's use the *auxiliary point* $t_o = -2$ and to the right use $t_1 = 0$. Then $f'(-2) = -2 < 0$, so f is *decreasing* on the interval $(-\infty, -1)$. And $f'(0) = 2 > 0$, so f is *increasing* on the interval $(-1, \infty)$.

Example 4.1.2.

[author=garrett, file =text_files/derivs_and_graphs]
Find the critical points and intervals on which $f(x) = x^3 - 12x + 3$ is increasing, decreasing. Compute $f'(x) = 3x^2 - 12$. Solve $3x^2 - 12 = 0$: this simplifies to $x^2 - 4 = 0$, so the *critical points* are ± 2. To the left of -2 choose *auxiliary point* $t_o = -3$, between -2 and $= 2$ choose auxiliary point $t_1 = 0$, and to the right of $+2$ choose $t_2 = 3$. Plugging in the auxiliary points to the derivative, we find that $f'(-3) = 27 - 12 > 0$, so f is *increasing* on $(-\infty, -2)$. Since $f'(0) = -12 < 0$, f is *decreasing* on $(-2, +2)$, and since $f'(3) = 27 - 12 > 0$, f is *increasing* on $(2, \infty)$.

Notice too that we don't really need to know the exact value of the derivative at the auxiliary points: all we care about is whether the derivative is positive or negative. The point is that sometimes some tedious computation can be avoided by stopping as soon as it becomes clear whether the derivative is positive or negative.

Exercises

1. Find the critical points and intervals on which $f(x) = x^2+2x+9$ is increasing, decreasing.
2. Find the critical points and intervals on which $f(x) = 3x^2 - 6x + 7$ is increasing, decreasing.
3. Find the critical points and intervals on which $f(x) = x^3 - 12x + 3$ is increasing, decreasing.

4.2 Minimization and Maximization

Definition 4.2.1.

[author=wikibooks, file =text_files/extreme_values]
A minimum or maximum is the function value at which a function has the lowest or highest value or values. There are two types

Absolute minima and maxima, which are on the interval $(-\infty, \infty)$. Local minima and maxima, where there exists an interval such that the value is the lowest or highest value.

Theorem 4.2.1.
[author=wikibooks, file =text_files/extreme_values]
The extreme value theorem states that for function f(x), continuous on the closed interval [a,b], f(x) must attain its maximum and minimum value each at least once. Mathematically, there exists numbers m and M such that $m \le f(x) \le M$ And there exist some c and d such that $f(c) = m$ and $f(d) = M$

Comment.

[author=wikibooks, file =text_files/extreme_values]
To formulate a proof of the extreme value theorem is quite hard, because it is so obviously true and that it almost seems a proof is unnecessary. However, various proofs are available.

Corollary 4.2.1.
[author=wikibooks, file =text_files/extreme_values]
An important result that the extreme value theorem establishes is the following Suppose that f is differentiable and that f has a local maximum or a local minimum at $x = c$. Then f'(c) = 0.

Definition 4.2.2.

[author=duckworth, file =text_files/max_mins]
Let $x = a$ be in the domain of $f(x)$.

$x = c$ is an	absolute maximum	if $f(x) \le f(c)$ for all x in the domain.
$x = c$ is a	local maximum	if $f(x) \le f(c)$ for all x near c.
		(c cannot be an endpoint)
$x = c$ is an	absolute minimum	if $f(x) \ge f(c)$ for all x in the domain.
$x = c$ is a	local minimum	if $f(x) \ge f(c)$ for all x near c.
		(c cannot be an endpoint)

Example 4.2.1.

[author=duckworth, file =text_files/max_mins]
Make up graphs showing some of each kind of thing.

Theorem 4.2.2.
[author=duckworth, file =text_files/max_mins]

If $x = c$ is a local min/max then $f'(c) = 0$ or $f'(c)$ is undefined. Look at some pictures of local min/max. If $f'(c) = 0$ or $f'(c)$ is undefined we call c a critical point. This fact justifies our approach to finding local min/max's which always starts with finding the critical points.

Rule 4.2.1.

[author=duckworth, file =text_files/max_mins]
Finding absolute max/mins. Suppose you want to find the absolute max/mins of $f(x)$ on an interval $[a, b]$

1. Find $f'(x)$, solve $f'(x) = 0$ and identify where $f'(x)$ is undefined (i.e. find the critical numbers).
2. Plug the critical numbers (which you found in step 1) and a and b into $f(x)$. This makes a list of y-values. The biggest y-value on this list is the absolute maximum. The smallest y-value on this list is the absolute mininum.

Rule 4.2.2.

[author=duckworth, file =text_files/max_mins]
Finding local min/maxs

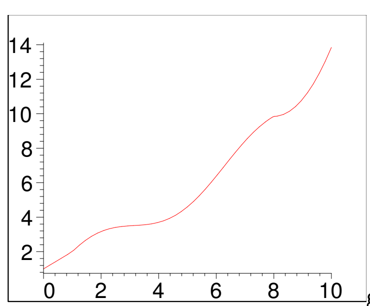
- Find $f'(x)$, solve $f'(x) = 0$ and identify where $f'(x)$ is undefined (i.e. find the critical numbers).
- Test each critical number (which you found in step 1) using the first derivative test or the second derivative test.

Comment.

[author=duckworth, file =text_files/max_mins]
So now we need to learn about the first and second derivative tests. Allthough it's not 100% necessary, we first introduce some more vocabulary.

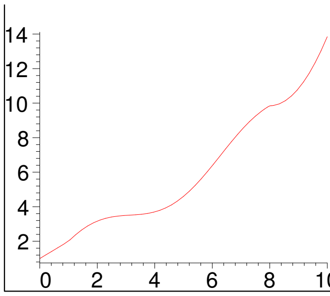
Definition 4.2.3.

[author=duckworth, file =text_files/max_mins]
If $f'(x) > 0$ we say that $f(x)$ is increasing. If $f'(x) < 0$ we say that $f(x)$ is decreasing.



generic_increasing_function

$f'(x) > 0$



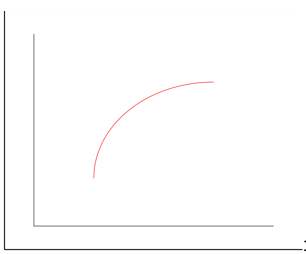
generic_increasing_function

$f'(x) < 0$

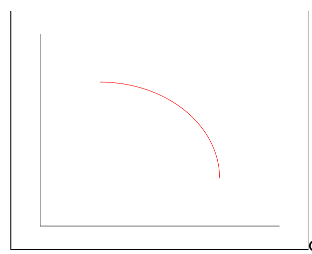
If $f''(x) > 0$ we say that $f(x)$ is concave up. This means that it is curving more upwards (it does not mean that it is increasing). If $f''(x) < 0$ we say that $f(x)$ is concave down. This means that it is curving more downwards.

Comment.

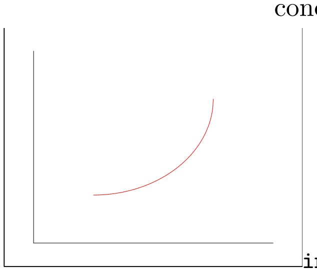
[author=duckworth, file =text_files/max_mins]
Note that concavity is *not* related to whether or not the graph is increasing or decreasing. In fact you can have any combination of concavity (up or down) with increasing or decreasing. This gives four possible pictures which you might want to keep in mind.


increasing_concave_down_graph

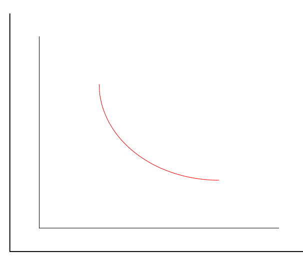
$f''(x) < 0$, $f'(x) > 0$
conc. down, incr.


decreasing_concave_down_graph

$f''(x) < 0$, $f'(x) < 0$
conc. down, decr.


increasing_concave_up_graph

$f''(x) > 0$, $f'(x) > 0$
conc. up, incr.

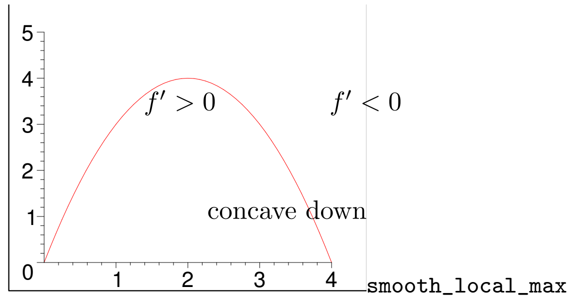

decreasing_concave_up_graph

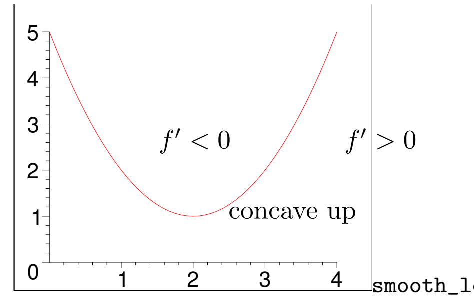
$f''(x) > 0$, $f'(x) < 0$
conc. up, decr.

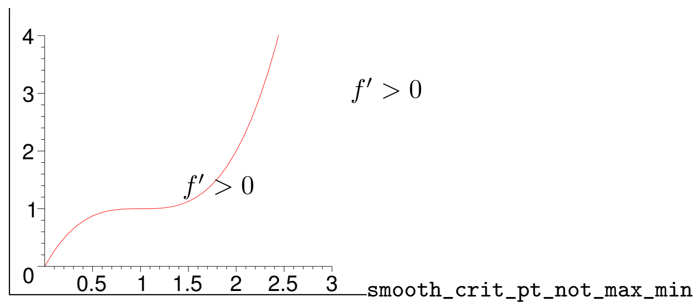
(Note, you can get these pictures from the four quadrants of a circle.)

Rule 4.2.3.

[author=duckworth, file =text_files/max_mins]
First and second derivative tests. You can see figure out what the tests should be, just by looking at pictures of max and mins, and thinking about what the first or second derivative is doing there.







concavity ?

First derivative test.

- If $f'(x)$ changes from $+$ on the left to $-$ on the right at $x = c$ then $x = c$ is a local min.
- If $f'(x)$ changes from $-$ on the left to $+$ on the right at $x = c$ then $x = c$ is a local max.
- If $f'(x)$ stays the same sign on both sides of $x = c$ then $x = c$ is neither a min or max.

Second derivative test.

- If $f''(c) < 0$ then $x = c$ is a local max.
- If $f''(c) > 0$ then $x = c$ is a local min.
- If $f''(c) = 0$ or $f''(c)$ is undefined then the second derivative test tells us nothing.

Comment.

[`author=duckworth`, file =`text_files/max_mins`]
We will almost never do both the first and the second derivative test. Only if we want to practice both of them will we do both.

Rule 4.2.4.

[author=duckworth, file =text_files/max_mins]
Finding changens of sign To use the first derivative test we need to be able to take a function $f'(x)$ and say when it is positive and when it is negative. Here's how you do this:

1. Solve for when $f'(x) = 0$ or is undefined. These are the only places when $f'(x)$ can change signs. (By the Intermediate Value Theorem! Yay! You thought you could forget about this. Not!)
2. Test a single value of x between each pair of numbers you found in step 1 (including a value to the right of all the numbers and a value to the left of all the numbers)

Example 4.2.2.

[author=duckworth, file =text_files/max_mins]
Let $f(x) = x - 2\sin(x)$.

(a) Find the critical points of $f(x)$ in the interval $0 \le x \le 4\pi$.

(b) Apply the first derivative test to each point in (a) and determine which points are local mins/max.

(c) Apply the second derivative test to each point in (a) and determine which points are local mins/max. (Usually we will *not* do both tests.)

(d) Find the absolute mins and maxs on the interval $0 \le x \le 4\pi$.

textbfSolution. The derivative is $f'(x) = 1 - 2\cos(x)$. The equation $f'(x) = 0$ has solutions on the interval $0 \le x \le 4\pi$ of $x = \pi/3,\ 5\pi/3,\ 7\pi/3,\ 11\pi/3$ which answers part (a). Testing values we find that $f'(x)$ is positive (so f is increasing) on $(\pi/3, 5\pi/5) \cup (7\pi/3, 11\pi/3)$ and $f'(x)$ is negative (so f is decreasing) on $[0, \pi/3) \cup (5\pi/3, 7\pi/3) \cup (11\pi/3, 4\pi]$. This shows that we have local max at $5\pi/3$ and $11\pi/3$ and local mins at $\pi/3$ and $7\pi/3$. This answers part (b) (actually we did more because we described the intervals where f is increasing and the intervals where f is decreasing). The second derivative is $f''(x) = 2\sin(x)$ and we have f'' is positive at $\pi/3$ and $7\pi/3$ (so these are local mins) and f'' is negative at $5\pi/3$ and $11\pi/3$ (so these are local maxs). This answers part (c). To find the absolute max and mins we compare y-values at $x = 0,\ \pi/3,\ 5\pi/3,\ 7\pi/3,\ 11\pi/3$ and 4π. One finds that $f(\pi/3) = -.6849$ is the absolute min and $f(11\pi/3) = 13.25$ is the absolute max. This answers part (d).

Example 4.2.3.

[author=duckworth, file =text_files/max_mins]
Let $f(x) = 5x^{2/3} + x^{5/3}$.

(a) Find all the critical points of $f(x)$.

(b) Apply the first derivative test to each point in (a) and determine which are local min/max.

(c) Apply the second derivative test to each point in (a) and determine which are local min/max.

Solution. The derivative $f'(x) = \frac{10}{3}x^{-1/3} + \frac{5}{3}x^{2/3}$. We see that $f'(x)$ is undefined at $x = 0$ and $f'(x) = 0$ at $x = -2$. This anwers part (a). Testing values we find that $f'(x)$ is positive on $(-\infty, -2) \cup (0, \infty)$ and $f'(x)$ is negative on $(-2, 0)$ so $x = -2$ is a local max and $x = 0$ is a local max. This answers part (b). The second derivative is $f''(x) = -\frac{10}{9}x^{-4/3} + \frac{10}{9}x^{-1/3}$. We see that $f''(-2)$ is $-$ and $f''(0)$ is $+$ so $x = -2$ is a local max and $x = 0$ is a local min. This answers part (c).

Discussion.

[`author=garrett`, file =`text_files/max_mins`]
The fundamental idea which makes calculus useful in understanding problems of maximizing and minimizing things is that at a *peak* of the graph of a function, or at the bottom of a *trough*, the tangent is *horizontal*. That is, *the derivative* $f'(x_o)$ *is* 0 *at points* x_o *at which* $f(x_o)$ *is a maximum or a minimum.*

Well, a little sharpening of this is necessary: sometimes for either natural or artificial reasons the variable x is restricted to some interval $[a, b]$. In that case, we can say that *the maximum and minimum values of* f *on the interval* $[a, b]$ *occur among the list of critical points and endpoints of the interval.*

And, if there are points where f is not differentiable, or is discontinuous, then these have to be added in, too. But let's stick with the basic idea, and just ignore some of these complications.

Rule 4.2.5.

[`author=garrett`, file =`text_files/max_mins`]
Let's describe a systematic procedure to find the minimum and maximum values of a function f on an interval $[a, b]$.

- Solve $f'(x) = 0$ to find the list of critical points of f.
- Exclude any critical points not inside the interval $[a, b]$.
- Add to the list the *endpoints* a, b of the interval (and any points of discontinuity or non-differentiability!)
- At each point on the list, evaluate the function f: the biggest number that occurs is the maximum, and the littlest number that occurs is the minimum.

Example 4.2.4.

[`author=garrett`, file =`text_files/max_mins`]
Find the minima and maxima of the function $f(x) = x^4 - 8x^2 + 5$ on the interval $[-1, 3]$. First, take the derivative and set it equal to zero to solve for critical points: this is

$$4x^3 - 16x = 0$$

or, more simply, dividing by 4, it is $x^3 - 4x = 0$. Luckily, we can see how to factor this: it is

$$x(x - 2)(x + 2)$$

So the critical points are $-2, 0, +2$. Since the interval does not include -2, we drop it from our list. And we *add* to the list the endpoints $-1, 3$. So the list of numbers to consider as potential spots for minima and maxima are $-1, 0, 2, 3$. Plugging these numbers into the function, we get (in that order) $-2, 5, -11, 14$. Therefore, the maximum is 14, which occurs at $x = 3$, and the minimum is -11, which occurs at $x = 2$.

Notice that in the previous example the maximum did not occur at a critical point, but by coincidence did occur at an endpoint.

Example 4.2.5.

[`author=garrett`, file =`text_files/max_mins`]
You have 200 feet of fencing with which you wish to enclose the largest possible rectangular garden. What is the largest garden you can have?

Let x be the length of the garden, and y the width. Then the area is simply xy. Since the perimeter is 200, we know that $2x + 2y = 200$, which we can solve to express y as a function of x: we find that $y = 100 - x$. Now we can rewrite the area as a function of x alone, which sets us up to execute our procedure:

$$area = xy = x(100 - x)$$

The derivative of this function with respect to x is $100 - 2x$. Setting this equal to 0 gives the equation

$$100 - 2x = 0$$

to solve for critical points: we find just *one*, namely $x = 50$.

Now what about endpoints? What is the interval? In this example we must look at 'physical' considerations to figure out what interval x is restricted to. Certainly a *width* must be a positive number, so $x > 0$ and $y > 0$. Since $y = 100 - x$, the inequality on y gives another inequality on x, namely that $x < 100$. So x is in $[0, 100]$.

When we plug the values $0, 50, 100$ into the function $x(100 - x)$, we get $0, 2500, 0$, in that order. Thus, the corresponding value of y is $100 - 50 = 50$, and the maximal possible area is $50 \cdot 50 = 2500$.

Definition 4.2.4.

[`author=wikibooks`, file =`text_files/max_mins`]
We say that the function f is differentiable at the point x if the derivative f'(x) exists. Since the derivative f'(x) of f at x is defined as a limit, it's quite possible that it won't exist. For example, if f is not even continuous at x then it can't be differentiable there (exercise). Continuity of f at x is not enough, though.

Example 4.2.6.

[`author=wikibooks`, file =`text_files/max_mins`]
For example, consider the function $f(x) = |x|$. The function f is differentiable at every point x other than x = 0. To see that it's differentiable at x for $x \neq 0$, you can either informally draw a graph and "see" it, or you can prove it with the epsilon-delta definition of differentiability as a limiting process. At x=0, though,

the direction of the graph changes suddenly at that point, so there is no well-defined tangent line (and so no derivative) for f at 0. We call the point 0 a "critical point" of f.

Definition 4.2.5.

[author=wikibooks, file =text_files/max_mins]
A critical number is defined, for the function f, as any number where the derivative f' is zero or undefined.

Example 4.2.7.

[author=wikibooks, file =text_files/max_mins]
For example, a critical point for the function $f(x) = x^2$ is 0, since f'(x) = 2x and f'(0) = 0. In fact, it is the only critcal number.

Comment.

[author=wikibooks, file =text_files/max_mins]
Critical numbers are significant because extrema only occur at critical numbers. However, the converse is not true. An example of this is $f(x) = x^3$, since it has one critical number $f'(x) = 3x^2$, f'(0) = 0. However, it is not an extrema.

Example 4.2.8.

[author=wikibooks, file =text_files/max_mins]
Example What are the local extrema of $f(x) = (x + 1)^2/2x$?

Find the critical numbers of f(x).

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

Set f' to zero to find the critical points.

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

Either find the zeros of the function...

$$x - 1 = 0, x = 1 \; x + 1 = 0, x = -1$$

...or do it symbolically and find $(f')^{-}1$.

$$(f')^{-1}(y) = \pm\sqrt{\frac{-1}{2y-1}} \; (f')^{-1}(0) = \pm\sqrt{\frac{-1}{2(0)-1}} = \pm 1$$

We can also see that f'(x) will be undefined at $x = 0$ (divide by zero).

Now we know that this function may have minima or maxima at -1, 0, or 1. Since f is continuous except for at 0, we can use the Intermediate Value Theorem find out whether they are minima, maxima, or nothing at all by picking intermediate values and checking them. We now pick intermediate values and test to see whether they show that the function value indicates an extreme value. Use convenient values when possible.

$x =$	-2	-1	$-1/2$	0	$1/2$	1	2
$f(x) =$	$-1/4$	0	$-1/4$	DNE	2.25	2	2.25

Since $f(-1)$ is greater than the numbers around it, 0 is a local maximum. Also, $f(1)$ is lower than the numbers around it, it is a local minimum. However, since $f(0)$ is also undefined, it is not anything.

Exercises

1. Olivia has 200 feet of fencing with which she wishes to enclose the largest possible rectangular garden. What is the largest garden she can have?

2. Find the minima and maxima of the function $f(x) = 3x^4 - 4x^3 + 5$ on the interval $[-2, 3]$.

3. The cost per hour of fuel to run a locomotive is $v^2/25$ dollars, where v is speed, and other costs are $100 per hour regardless of speed. What is the speed that minimizes cost *per mile*?

4. The product of two numbers x, y is 16. We know $x \geq 1$ and $y \geq 1$. What is the greatest possible sum of the two numbers?

5. Find both the minimum and the maximum of the function $f(x) = x^3 + 3x + 1$ on the interval $[-2, 2]$.

4.3 Local minima and maxima (First Derivative Test)

Definition 4.3.1.

[author=garrett, file =text_files/first_deriv_test]
A function f has a **local maximum** or **relative maximum** at a point x_o if the values $f(x)$ of f for x 'near' x_o are all less than $f(x_o)$. Thus, the graph of f near x_o has a *peak* at x_o. A function f has a **local minimum** or **relative minimum** at a point x_o if the values $f(x)$ of f for x 'near' x_o are all greater than $f(x_o)$. Thus, the graph of f near x_o has a *trough* at x_o. (To make the distinction clear, sometimes the 'plain' maximum and minimum are called **absolute** maximum and minimum.)

Comment.

[author=garrett, file =text_files/first_deriv_test]
Yes, in both these 'definitions' we are tolerating ambiguity about what 'near' would mean, although the peak/trough requirement on the graph could be translated into a less ambiguous definition. But in any case we'll be able to execute the procedure given below to *find* local maxima and minima without worrying over a formal definition.

This procedure is just a variant of things we've already done to analyze the intervals of increase and decrease of a function, or to find absolute maxima and minima. This procedure starts out the same way as does the analysis of intervals of increase/decrease, and also the procedure for finding ('absolute') maxima and minima of functions.

Rule 4.3.1.

[author=garrett, file =text_files/first_deriv_test]
To find the local maxima and minima of a function f on an interval $[a, b]$:

- Solve $f'(x) = 0$ to find *critical points* of f.
- Drop from the list any critical points that aren't in the interval $[a, b]$.
- Add to the list the endpoints (and any points of discontinuity or non-differentiability): we have an *ordered* list of special points in the interval:

$$a = x_o < x_1 < \ldots < x_n = b$$

- Between each pair $x_i < x_{i+1}$ of points in the list, choose an auxiliary point t_{i+1}. Evaluate the *derivative* f' at all the auxiliary points.
- For each critical point x_i, we have the auxiliary points to each side of it: $t_i < x_i < t_{i+1}$. There are four cases *best remembered by drawing a picture!*:
- if $f'(t_i) > 0$ and $f'(t_{i+1}) < 0$ (so f is *increasing* to the left of x_i and *decreasing* to the right of x_i, then f has a *local maximum* at x_o.
- if $f'(t_i) < 0$ and $f'(t_{i+1}) > 0$ (so f is *decreasing* to the left of x_i and *increasing* to the right of x_i, then f has a *local minimum* at x_o.

- if $f'(t_i) < 0$ and $f'(t_{i+1}) < 0$ (so f is *decreasing* to the left of x_i and *also decreasing* to the right of x_i, then f has *neither* a local maximum nor a local minimum at x_o.
- if $f'(t_i) > 0$ and $f'(t_{i+1}) > 0$ (so f is *increasing* to the left of x_i and *also increasing* to the right of x_i, then f has *neither* a local maximum nor a local minimum at x_o.

The endpoints require separate treatment: There is the auxiliary point t_o just to the *right* of the left endpoint a, and the auxiliary point t_n just to the *left* of the right endpoint b:

- At the *left* endpoint a, if $f'(t_o) < 0$ (so f' is *decreasing* to the right of a) then a is a *local maximum.*
- At the *left* endpoint a, if $f'(t_o) >$ (so f' is *increasing* to the right of a) then a is a *local minimum.*
- At the *right* endpoint b, if $f'(t_n) < 0$ (so f' is *decreasing* as b is approached from the left) then b is a *local minimum.*
- At the *right* endpoint b, if $f'(t_n) >$ (so f' is *increasing* as b is approached from the left) then b is a *local maximum.*

Comment.

[author=garrett, file =text_files/first_deriv_test]
The possibly bewildering list of possibilities really shouldn't be bewildering after you get used to them. We are already acquainted with evaluation of f' at auxiliary points between critical points in order to see whether the function is increasing or decreasing, and now we're just applying that information to see whether the graph *peaks, troughs, or does neither* around each critical point and endpoints. That is, *the geometric meaning of the derivative's being positive or negative is easily translated into conclusions about local maxima or minima.*

Example 4.3.1.

[author=garrett, file =text_files/first_deriv_test]
Find all the local (=relative) minima and maxima of the function $f(x) = 2x^3 - 9x^2 + 1$ on the interval $[-2, 2]$: To find critical points, solve $f'(x) = 0$: this is $6x^2 - 18x = 0$ or $x(x-3) = 0$, so there are two critical points, 0 and 3. Since 3 is not in the interval we care about, we drop it from our list. Adding the endpoints to the list, we have

$$-2 < 0 < 2$$

as our ordered list of special points. Let's use auxiliary points $-1, 1$. At -1 the derivative is $f'(-1) = 24 > 0$, so the function is increasing there. At $+1$ the derivative is $f'(1) = -12 < 0$, so the function is decreasing. Thus, since it is increasing to the left and decreasing to the right of 0, it must be that 0 is a *local maximum.* Since f is increasing to the right of the left endpoint -2, that left endpoint must give a *local minimum.* Since it is decreasing to the left of the right endpoint $+2$, the right endpoint must be a *local minimum.*

Comment.

[**author=garrett**, file =**text_files/first_deriv_test**]
Notice that although the processes of finding *absolute* maxima and minima and *local* maxima and minima have a lot in common, they have essential differences. In particular, the only relations between them are that *critical points* and *endpoints* (and points of discontinuity, etc.) play a big role in both, and that the *absolute* maximum is certainly a *local* maximum, and likewise the *absolute* minimum is certainly a *local* minimum.

For example, just plugging critical points into the function does not reliably indicate which points are *local* maxima and minima. And, on the other hand, knowing which of the critical points are *local* maxima and minima generally is only a small step toward figuring out which are *absolute*: values still have to be plugged into the funciton! *So don't confuse the two procedures!*

(By the way: while it's fairly easy to make up story-problems where the issue is to find the maximum or minimum value of some function on some interval, it's harder to think of a simple application of *local* maxima or minima).

Exercises

1. Find all the local (=relative) minima and maxima of the function $f(x) = (x+1)^3 - 3(x+1)$ on the interval $[-2, 1]$.

2. Find the local (=relative) minima and maxima on the interval $[-3, 2]$ of the function $f(x) = (x+1)^3 - 3(x+1)$.

3. Find the local (relative) minima and maxima of the function $f(x) = 1 - 12x + x^3$ on the interval $[-3, 3]$.

4. Find the local (relative) minima and maxima of the function $f(x) = 3x^4 - 8x^3 + 6x^2 + 17$ on the interval $[-3, 3]$.

4.4 An algebra trick

Rule 4.4.1.

[author=garrett, file =text_files/algebra_for_first_deriv_test]
The algebra trick here goes back at least 350 years. This is worth looking at if only as an additional review of algebra, but is actually of considerable value in a variety of hand computations as well.

The algebraic identity we use here starts with a product of factors each of which may occur with a *fractional or negative exponent.* For example, with 3 such factors:

$$f(x) = (x-a)^k\ (x-b)^\ell\ (x-c)^m$$

The derivative can be computed by using the product rule twice:

$$\begin{aligned} f'(x) &= \\ &= k(x-a)^{k-1}(x-b)^\ell(x-c)^m + (x-a)^k\ell(x-b)^{\ell-1}(x-c)^m + (x-a)^k(x-b)^\ell m(x-c)^{m-1} \end{aligned}$$

Now all three summands here have a common factor of

$$(x-a)^{k-1}(x-b)^{\ell-1}(x-c)^{m-1}$$

which we can take out, using the distributive law in reverse: we have

$$\begin{aligned} f'(x) &= \\ &= (x-a)^{k-1}(x-b)^{\ell-1}(x-c)^{m-1}[k(x-b)(x-c) + \ell(x-a)(x-c) + m(x-a)(x-b)] \end{aligned}$$

The minor miracle is that the big expression inside the square brackets is a mere quadratic polynomial in x.

Then to determine *critical points* we have to figure out the roots of the equation $f'(x) = 0$: If $k-1 > 0$ then $x = a$ is a critical point, if $k-1 \le 0$ it isn't. If $\ell - 1 > 0$ then $x = b$ is a critical point, if $\ell - 1 \le 0$ it isn't. If $m-1 > 0$ then $x = c$ is a critical point, if $m-1 \le 0$ it isn't. And, last but not least, *the two roots of the quadratic equation*

$$k(x-b)(x-c) + \ell(x-a)(x-c) + m(x-a)(x-b) = 0$$

are critical points.

There is also another issue here, about not wanting to take square roots (and so on) of negative numbers. We would exclude from the domain of the function any values of x which would make us try to take a square root of a negative number. But this might also force us to give up some critical points! Still, this is not the main point here, so we will do examples which avoid this additional worry.

Example 4.4.1.

[author=garrett, file =text_files/algebra_for_first_deriv_test]
A very simple *numerical* example: suppose we are to find the *critical points* of the function

$$f(x) = x^{5/2}(x-1)^{4/3}$$

Implicitly, we have to find the critical points first. We compute the derivative by using the product rule, the power function rule, and a tiny bit of chain rule:

$$f'(x) = \frac{5}{2}x^{3/2}(x-1)^{4/3} + x^{5/2}\frac{4}{3}(x-1)^{1/3}$$

And now *solve* this for x? It's not at all a polynomial, and it is a little ugly.

But our algebra trick transforms this issue into something as simple as *solving a linear equation*: first figure out the largest power of x that occurs in *all* the terms: it is $x^{3/2}$, since $x^{5/2}$ occurs in the first term and $x^{3/2}$ in the second. The largest power of $x-1$ that occurs in *all* the terms is $(x-1)^{1/3}$, since $(x-1)^{4/3}$ occurs in the first, and $(x-1)^{1/3}$ in the second. *Taking these common factors out* (using the distributive law 'backward'), we rearrange to

$$\begin{aligned} f'(x) &= \frac{5}{2}x^{3/2}(x-1)^{4/3} + x^{5/2}\frac{4}{3}(x-1)^{1/3} \\ &= x^{3/2}(x-1)^{1/3}\left(\frac{5}{2}(x-1) + \frac{4}{3}x\right) \\ &= x^{3/2}(x-1)^{1/3}(\frac{5}{2}x - \frac{5}{2} + \frac{4}{3}x) \\ &= x^{3/2}(x-1)^{1/3}(\frac{23}{6}x - \frac{5}{2}) \end{aligned}$$

Now to see when this is 0 is not so hard: first, since the power of x appearing in front is *positive*, $x=0$ make this expression 0. Second, since the power of $x+1$ appearing in front is *positive*, if $x-1=0$ then the whole expression is 0. Third, and perhaps *unexpectedly*, from the simplified form of the complicated factor, if $\frac{23}{6}x - \frac{5}{2} = 0$ then the whole expression is 0, as well. So, altogether, the *critical points* would appear to be

$$x = 0, \frac{15}{23}, 1$$

Many people would overlook the critical point $\frac{15}{23}$*, which is visible only after the algebra we did.*

Exercises

1. Find the critical points and intervals of increase and decrease of $f(x) = x^{10}(x-1)^{12}$.

2. Find the critical points and intervals of increase and decrease of $f(x) = x^{10}(x-2)^{11}(x+2)^3$.

3. Find the critical points and intervals of increase and decrease of $f(x) = x^{5/3}(x+1)^{6/5}$.

4. Find the critical points and intervals of increase and decrease of $f(x) = x^{1/2}(x+1)^{4/3}(x-1)^{-11/3}$.

4.5 Linear approximations: approximation by differentials

Discussion.

[author=garrett, file =text_files/linear_approximation]
The idea here in 'geometric' terms is that in some vague sense a curved line can be approximated by a straight line tangent to it. Of course, this approximation is only good at all 'near' the point of tangency, and so on. So the only formula here is secretly the formula for the tangent line to the graph of a function. There is some hassle due to the fact that there are so many different choices of symbols to *write* it.

Rule 4.5.1.

[author=duckworth, file =text_files/linear_approximation]
Linearization. Let f be a function, fix an x-value $x = a$ and let $L(x)$ be the tangent line of $f(x)$ at $x = a$. Then $f(x)$ and $L(x)$ are approximately equal for those x-values near $x = a$. In symbols:

$$\text{for } x \text{ near } a \implies f(x) \approx L(x)$$

In this context we call $L(x)$ the **linearization** of $f(x)$ at $x = a$.

Comment.

[author=garrett,author=duckworth, file =text_files/linear_approximation]
We note the following:

- One formula for $L(x)$ is given by $L(x) = f'(a)(x - a) + f(a)$. One really important thing abouth this formula is that to make it explicit, we only need to know two numbers: $f(a)$ and $f'(a)$.
- The purpose of linearization is that $L(x)$ might be easier to calculate or work with than $f(x)$. In this sense, we use $L(x)$ to tell us about $f(x)$.
- We will not spend time making precise what "near" means in this definition or how good the approximation "$\approx$" is. However we note that $f(a) = L(a)$, so that at $x = a$ the $L(x)$ is an exact approximation of $f(x)$. Also, $f'(a) = \frac{d}{dx}L(x)\Big|_{x=a}$, i.e. the derivative of f at $x = a$ equals the derivative of $L(x)$ at $x = a$. So again, at $x = a$, the slope of $L(x)$ is an exact approximation of $f(x)$.

Notation.

[author=garrett, file =text_files/linear_approximation]
The aproximation statement has many paraphrases in varying choices of symbols, and a person needs to be able to recognize all of them. For example, one of

the more traditional paraphrases, which introduces some slightly silly but oh-so-traditional notation, is the following one. We might also say that y is a function of x given by $y = f(x)$. Let

$$\Delta x = \text{ small change in } x$$

$$\Delta y = \text{ corresponding change in } y \ = f(x + \Delta x) - f(x)$$

Then the assertion is that

$$\Delta y \approx f'(x)\,\Delta x$$

Sometimes some texts introduce the following questionable (but traditionally popular!) notation:

$$dy = f'(x)\ dx = \text{ approximation to change in } y$$

$$dx = \Delta x$$

and call the dx and dy *'differentials'*. And then this whole procedure is **'approximation by differentials'**. A not particularly enlightening paraphrase, using the previous notation, is

$$dy \approx \Delta y$$

Even though you may see people writing this, don't do it.

More paraphrases, with varying symbols:

$$f(x + \Delta x) \approx f(x) + f'(x)\Delta x$$

$$f(x + \delta) \approx f(x) + f'(x)\delta$$

$$f(x + h) \approx f(x) + f'(x)h$$

$$f(x + \Delta x) - f(x) \approx f'(x)\Delta x$$

$$y + \Delta y \approx f(x) + f'(x)\Delta x$$

$$\Delta y \approx f'(x)\Delta x$$

Comment.

[`author=garrett`, file =`text_files/linear_approximation`]
A little history: Until just 20 or 30 years ago, calculators were not widely available, and especially not typically able to evaluate trigonometric, exponential, and logarithm functions. In that context, the kind of vague and unreliable 'approximation' furnished by 'differentials' was certainly worthwhile in many situations.

By contrast, now that pretty sophisticated calculators are widely available, some things that once seemed sensible are no longer. For example, a very traditional type of question is to 'approximate $\sqrt{10}$ by differentials'. A reasonable contemporary response would be to simply punch in '$1, 0, \sqrt{}$' on your calculator and get the answer immediately to 10 decimal places. But this was possible only relatively recently.

Comment.

[author=duckworth, file =text_files/linear_approximation]
Try to keep the following in mind as we do some examples. We will start with examples that are easy, or historically relevant, but do not show you, a modern reader, why linearization is a useful thing. These examples include anything where we have a formula for $f(x)$, and we are trying to approximate $f(b)$ for some number b. The examples that are useful to us today will came later.

Historically, using linearizaton to approximate $\sqrt{10}$ would have been a useful trick for most of the last 1000 years. Today, we (or our machines) can calculate $\sqrt{10}$; we will do this example just as a means of *practicing* linearization.

However, linearization is still very useful today. The following applications are incredibly important and we'll return to them later in these notes.

1. Solving an equation of the form $f(x) = 0$.

   Let $f(x) = e^x + x$ and consider solving $e^x + x = 0$. I cannot solve this equation exactly, but I can find an approximation using linearization. Let $L(x)$ be the tangent line at $x = 0$. It is easy to show that $L(x) = 2x + 1$. Instead of solving $f(x) = 0$ I solve $L(x) = 0$ to get $2x + 1 = 0$, $x = -1/2$. Since $f(x)$ and $L(x)$ are approximately the same thing, $x = -1/2$ is approximately a solution of $f(x) = 0$.

   Repeating this process will (usually) give you a more accurate approximation of $f(x) = 0$, and is called **Newton's Method**.

2. Suppose that we know only a little bit about a function. For example, suppose that we know that some moving object has position $p(t)$ satisfying $p(0) = 7$, and that the velocity is given by $v(t) = (t - 1)\cos(t)$. Can we approximate $p(.5)$, $p(1)$, etc.?

   Let $L(t)$ be the linear approximation of $p(t)$ at $t = 0$. To write down a formula for $L(t)$ we only need to know two numbers: $p(0)$ and $p'(0)$. Well, we were explicitly told that $p(0) = 7$, and we can find $p'(0) = v(0) = (0 - 1)\cos(0) = -1$. Therefore $L(t) = -t + 7$. Therefore $p(.5) \approx L(.5) = 6.5$ and $p(1) \approx L(1) = 6$.

Example 4.5.1.

[author=garrett, file =text_files/linear_approximation]
For example let's approximate $\sqrt{17}$ by differentials. For this problem to make sense at all you should imagine that you have no calculator. We take $f(x) = \sqrt{x} = x^{1/2}$. The idea here is that we can easily evaluate 'by hand' both f and f' at the point $x = 16$ which is 'near' 17. (Here $f'(x) = \frac{1}{2}x^{-1/2}$). Thus, here

$$\Delta x = 17 - 16 = 1$$

and

$$\sqrt{17} = f(17) \approx f(16) + f'(16)\Delta x = \sqrt{16} + \frac{1}{2}\frac{1}{\sqrt{16}} \cdot 1 = 4 + 18$$

Similarly, if we wanted to approximate $\sqrt{18}$ 'by differentials', we'd again take $f(x) = \sqrt{x} = x^{1/2}$. Still we imagine that we are doing this 'by hand', and then of course we can 'easily evaluate' the function f and its derivative f' at the point $x = 16$ which is 'near' 18. Thus, here

$$\Delta x = 18 - 16 = 2$$

and

$$\sqrt{18} = f(18) \approx f(16) + f'(16)\Delta x = \sqrt{16} + \frac{1}{2}\frac{1}{\sqrt{16}} \cdot 2 = 4 + \frac{1}{4}$$

Why not use the 'good' point 25 as the 'nearby' point to find $\sqrt{18}$? Well, in broad terms, the further away your 'good' point is, the worse the approximation will be. Yes, it is true that we have little idea how good or bad the approximation is *anyway*.

Comment.

[**author=garrett**, file =**text_files/linear_approximation**]
It is somewhat more sensible to *not* use this idea for numerical work, but rather to say things like

$$\sqrt{x+1} \approx \sqrt{x} + \frac{1}{2}\frac{1}{\sqrt{x}}$$

and

$$\sqrt{x+h} \approx \sqrt{x} + \frac{1}{2}\frac{1}{\sqrt{x}} \cdot h$$

This kind of assertion is more than any particular numerical example would give, because it gives a *relationship*, telling how much the *output* changes for given change in *input*, depending what *regime* (=interval) the input is generally in. In this example, we can make the *qualitative* observation that *as x increases the difference $\sqrt{x+1} - \sqrt{x}$ decreases.*

Example 4.5.2.

[**author=garrett**, file =**text_files/linear_approximation**]
Another numerical example: Approximate $\sin 31^o$ 'by differentials'. Again, the point is *not* to hit $3, 1, \sin$ on your calculator (after switching to degrees), but rather to *imagine that you have no calculator*. And we are supposed to remember from pre-calculator days the 'special angles' and the values of trig functions at them: $\sin 30^o = \frac{1}{2}$ and $\cos 30^o = \frac{\sqrt{3}}{2}$. So we'd use the function $f(x) = \sin x$, and we'd imagine that we can evaluate f and f' easily by hand at 30^o. Then

$$\Delta x = 31^o - 30^o = 1^o = 1^o \cdot \frac{2\pi \text{ radians}}{360^o} = \frac{2\pi}{360} \text{ radians}$$

We have to rewrite things in radians since we really only can compute derivatives of trig functions in radians. Yes, this is a complication in our supposed 'computation by hand'. Anyway, we have

$$\sin 31^o = f(31^o) = f(30^o) + f'(30^o)\Delta x = \sin 30^o + \cos 30^o \cdot \frac{2\pi}{360}$$

$$= \frac{1}{2} + \frac{\sqrt{3}}{2} 2\pi 360$$

Evidently we are to *also* imagine that we *know* or can easily *find* $\sqrt{3}$ (by differentials?) as well as a value of π. *Yes*, this is a lot of trouble in comparison to just punching the buttons, and from a contemporary perspective may seem senseless.

Example 4.5.3.

[author=garrett, file =text_files/linear_approximation]
Approximate $\ln(x+2)$ 'by differentials', in terms of $\ln x$ and x: This *non-numerical* question is somewhat more sensible. Take $f(x) = \ln x$, so that $f'(x) = \frac{1}{x}$. Then

$$\Delta x = (x+2) - x = 2$$

and by the formulas above

$$\ln(x+2) = f(x+2) \approx f(x) + f'(x) \cdot 2 = \ln x + \frac{2}{x}$$

Example 4.5.4.

[author=garrett, file =text_files/linear_approximation]
Approximate $\ln(e+2)$ in terms of differentials: Use $f(x) = \ln x$ again, so $f'(x) = \frac{1}{x}$. We probably have to imagine that we can 'easily evaluate' both $\ln x$ and $\frac{1}{x}$ at $x = e$. (Do we know a numerical approximation to e?). Now

$$\Delta x = (e+2) - e = 2$$

so we have

$$\ln(e+2) = f(e+2) \approx f(e) + f'(e) \cdot 2 = \ln e + \frac{2}{e} = 1 + \frac{2}{e}$$

since $\ln e = 1$.

Exercises

1. Approximate $\sqrt{101}$ 'by differentials' in terms of $\sqrt{100} = 10$.
2. Approximate $\sqrt{x+1}$ 'by differentials', in terms of $\sqrt{x}$.
3. Granting that $\frac{d}{dx} \ln x = \frac{1}{x}$, approximate $\ln(x+1)$ 'by differentials', in terms of $\ln x$ and x.
4. Granting that $\frac{d}{dx} e^x = e^x$, approximate e^{x+1} in terms of e^x.
5. Granting that $\frac{d}{dx} \cos x = -\sin x$, approximate $\cos(x+1)$ in terms of $\cos x$ and $\sin x$.

4.6 Implicit differentiation

Discussion.

[author=garrett, file =text_files/implicit_derivatives]
There is nothing 'implicit' about the differentiation we do here, it is quite 'explicit'. The difference from earlier situations is that we have a *function defined 'implicitly'*. What this means is that, instead of a clear-cut (if complicated) formula for the value of the function in terms of the input value, we only have a *relation* between the two. This is best illustrated by examples.

Example 4.6.1.

[author=garrett, file =text_files/implicit_derivatives]
For example, suppose that y is a function of x and

$$y^5 - xy + x^5 = 1$$

and we are to find some useful expression for dy/dx. Notice that it is not likely that we'd be able to *solve* this equation for y as a function of x (nor vice-versa, either), so our previous methods do not obviously do anything here! But both sides of that equality are functions of x, and are *equal*, so their derivatives are equal, surely. That is,

$$5y^4 \frac{dy}{dx} - 1 \cdot y - x\frac{dy}{dx} + 5x^4 = 0$$

Here the trick is that we can 'take the derivative' without knowing exactly what y is as a function of x, but just using the rules for differentiation.

Specifically, to take the derivative of the term y^5, we view this as a *composite* function, obtained by applying the take-the-fifth-power function after applying the (not clearly known!) function y. Then use the chain rule!

Likewise, to differentiate the term xy, we use the product rule

$$\frac{d}{dx}(x \cdot y) = \frac{dx}{dx} \cdot y + x \cdot \frac{dy}{dx} = y + x \cdot \frac{dy}{dx}$$

since, after all,

$$\frac{dy}{dx} = 1$$

And the term x^5 is easy to differentiate, obtaining the $5x^4$. The other side of the equation, the function '1', is *constant*, so its derivative is 0. (The fact that this means that the left-hand side is also constant should not be mis-used: we need to use the very non-trivial looking expression we have for that constant function, there on the left-hand side of that equation!).

Now the amazing part is that this equation can be *solved for* y', if we tolerate a formula involving not only x, but also y: first, regroup terms depending on whether they have a y' or not:

$$y'(5y^4 - x) + (-y + 5x^4) = 0$$

Then move the non-y' terms to the other side

$$y'(5y^4 - x) = y - 5x^4$$

and divide by the 'coefficient' of the y':

$$y' = \frac{y - 5x^4}{5y^4 - x}$$

Yes, this is *not* as good as if there were a formula for y' *not* needing the y. But, on the other hand, the initial situation we had did not present us with a formula for y in terms of x, so it was necessary to lower our expectations.

Yes, if we are given a value of x and told to find the corresponding y', it would be impossible without luck or some additional information. For example, in the case we just looked at, if we were asked to find y' when $x = 1$ and $y = 1$, it's easy: just plug these values into the formula for y' in terms of *both* x and y: when $x = 1$ and $y = 1$, the corresponding value of y' is

$$y' = \frac{1 - 5 \cdot 1^4}{5 \cdot 1^4 - 1} = -4/4 = -1$$

If, instead, we were asked to find y and y' when $x = 1$, not knowing in advance that $y = 1$ fits into the equation when $x = 1$, we'd have to hope for some luck. First, we'd have to try to solve the original equation for y with x replace by its value 1: solve

$$y^5 - y + 1 = 1$$

By luck indeed, there is some cancellation, and the equation becomes

$$y^5 - y = 0$$

By further luck, we can factor this 'by hand': it is

$$0 = y(y^4 - 1) = y(y^2 - 1)(y^2 + 1) = y(y - 1)(y + 1)(y^2 + 1)$$

So there are actually *three* real numbers which work as y for $x = 1$: the values $-1, 0, +1$. There is no clear way to see which is 'best'. But in any case, any one of these three values could be used as y in substituting into the formula

$$y' = \frac{y - 5x^4}{5y^4 - x}$$

we obtained above.

Yes, there are really *three solutions*, three functions, etc.

Note that we *could* have used the Intermediate Value Theorem and/or Newton's Method to *numerically* solve the equation, even without too much luck. In 'real life' a person should be prepared to do such things.

Discussion.

[`author=livshits`, file =`text_files/implicit_derivatives`]
We sometimes can calculate the derivative of a function *without knowing an explicit expression* of this function. This approach is called *implicit* differentiation. We already saw some simple examples of it in the exercises.

Example 4.6.2.

[author=livshits, file =text_files/implicit_derivatives]
Let $x(t)$ be the real root of the equation $x^5 + x = t^2 + t$ (you can sketch the curve $y = x^5 + x$ or notice that $x^5 + x$ is an increasing function of x to see that there is only one such solution, so the function $x(t)$ is well defined). It turns out that it is impossible to write an expression for $x(t)$ in terms of the familiar functions, so we are stuck. But if we differentiate our equation (with respect to t) we will get a linear equation for $x'(t)$ that is easy to solve. Doing that will give us an expression for $x'(t)$ in terms of $x(t)$ and t. Remembering that $x(0) = 0$, we can figure out that $x'(0) = 1$.

Comment.

[author=livshits, file =text_files/implicit_derivatives]
This example illustrates the following phenomenon: the equations usually simplify when we differentiate them (but at a price of the derivatives popping up in the resulting equation). As another example, you can think of the planetary motions in the solar system. They are very complicated, but if we differentiate 2 times, we get Newton's second law of dymnamics and his law of gravitation, both of which can be written in one line. We will touch upon these matters more later.

There is one subtlety here: we *assumed* that $x(t)$ is a differentiable function. This assumption has to be justified even if we could compute $x'(t)$. To illustrate what can go wrong, let us assume that there is a biggest natural number N. Then $N^2 \leq N$, but 1 is the only such natural number, therefore $N = 1$. Of course it is a joke (it's called Perron's paradox), but it shows that you can end up with the wrong thing even if you find it, if you assume the existence of a thing that doesn't exist. We will encounter less ridiculous examples of this phenomenon when we treat maxima and minima. We will return to this particular question of the existence of $x'(0)$ later.

Meanwhile, there is a comforting fact that as long as we don't have to divide by zero to carry out the implicit differentiation, the derivative that we are looking for indeed exists under some very mild assumptions about the equation. This fact is called the implicit function theorem.

Rule 4.6.1.

[author=wikibooks, file =text_files/derivatives_inverse_trig]
Arccsine, arccosine, arctangent. These are the functions that allow you to determine the angle given the sine, cosine, or tangent of that angle.

First, let us start with the arcsine such that

$y = \arcsin(x)$

To find dy/dx we first need to break this down into a form we can work with

$x = \sin(y)$

Then we can take the derivative of that

$1 = \cos(y) \cdot \frac{dy}{dx}$

...and solve for dy / dx

$\frac{dy}{dx} = \frac{1}{\cos(y)}$

At this point we need to go back to the unit triangle. Since y is the angle and the opposite side is $\sin(y)$ (which is equal to x), the adjacent side is $\cos(y)$ (which is equal to the square root of 1 minus x^2, based on the pythagorean theorem), and the hypotenuse is 1. Since we have determined the value of cos(y) based on the unit triangle, we can substitute it back in to the above equation and get

Derivative of the Arcsine $\frac{d}{dx}\arcsin(x) = \frac{1}{\sqrt{1+x^2}}$

Rule 4.6.2.

[author=wikibooks, file =text_files/derivatives_inverse_trig]
We can use an identical procedure for the arccosine and arctangent

Derivative of the Arccosine $\frac{d}{dx}\arccos(x) = \frac{-1}{\sqrt{1-x^2}}$

Derivative of the Arctangent $\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$

Exercises

1. Suppose that y is a function of x and

$$y^5 - xy + x^5 = 1$$

   Find dy/dx at the point $x = 1, y = 0$.

2. Suppose that y is a function of x and

$$y^3 - xy^2 + x^2y + x^5 = 7$$

   Find dy/dx at the point $x = 1, y = 2$. Find $\frac{d^2y}{dx^2}$ at that point.

4.7 Related rates

Discussion.

[author=garrett, file =text_files/related_rates]
In this section, most functions will be functions of a parameter t which we will think of as *time*. There is a convention coming from physics to write the derivative of any function y of t as $\dot{y} = dy/dt$, that is, with just a *dot* over the functions, rather than a *prime*.

The issues here are variants and continuations of the previous section's idea about *implicit differentiation*. Traditionally, there are other (non-calculus!) issues introduced at this point, involving both story-problem stuff as well as requirement to be able to deal with *similar triangles*, the *Pythagorean Theorem*, and to recall formulas for *volumes* of cones and such.

Example 4.7.1.

[author=garrett, file =text_files/related_rates]
Continuing with the idea of describing a function by a relation, we could have *two* unknown functions x and y of t, *related* by some formula such as

$$x^2 + y^2 = 25$$

A typical question of this genre is 'What is $\dot{y}$ when $x = 4$ and $\dot{x} = 6$?'

The fundamental rule of thumb in this kind of situation is *differentiate the relation with respect to* t: so we differentiate the relation $x^2 + y^2 = 25$ with respect to t, even though we don't know any details about those two function x and y of t:

$$2x\dot{x} + 2y\dot{y} = 0$$

using the chain rule. We can solve this for $\dot{y}$:

$$\dot{y} = -\frac{x\dot{x}}{y}$$

So *at any particular moment*, if we knew the values of $x, \dot{x}, y$ then we could find $\dot{y}$ *at that moment*.

Here it's easy to solve the original relation to find y when $x = 4$: we get $y = \pm 3$. Substituting, we get

$$\dot{y} = -\frac{4 \cdot 6}{\pm 3} = \pm 8$$

(The $\pm$ notation means that we take $+$ chosen if we take $y = -3$ and $-$ if we take $y = +3$).

Discussion.

[author=duckworth, file =text_files/related_rates]
The basic ideas of related rates are these:

- You have a formula which has more than one independant variable in it. Each variable is a letter. You let each letter represent a function of t, and then you take the derivative with respect to t. For instance, if you have

$A = fg$ where f and g are each functions of t, then the product rule says that $\frac{dA}{dt} = \frac{df}{dt}g + f\frac{dg}{dt}$.

- Now you look at the information in the problem and plug in numbers for everything in the formula except one unknown quantity, which you can solve for.
- We interpret $\frac{df}{dt}$ as the rate of change of f with respect to t. Similarly for $\frac{dA}{dt}$ and $\frac{dg}{dt}$.
- Often, the hardest part is just figuring out what formula to start with.

Example 4.7.2.

[**author=duckworth**, file =**text_files/related_rates**]
Let $A = fg$ so $\frac{dA}{dt} = \frac{df}{dt}g + f\frac{dg}{dt}$ as above.

1. $f = t^2$ and $g = \cos(t)$. Find $\frac{df}{dt}$ and $\frac{dg}{dt}$. If you plug all of this into the formula for $\frac{dA}{dt}$ just given do you get the same thing as if you found $\frac{d}{dt}t^2\cos(t)$ in one step?

2. Now suppose that instead of part (a) you know that $f(0) = 10$, $\frac{df}{dt}\Big|0 = 1$, $g(0) = 1$ and $\frac{dg}{dt}\Big|0 = 0$. What is $\frac{dA}{dt}\Big|0$?

3. Suppose now that you know $f = t + 10$ and $g = \cos(t)$. Find $\frac{dA}{dt}\Big|0$ by taking the derivative of $(t + 10)\cos(t)$ and evaluating at 0.

4. Suppose now that you know $f(1) = 7$, $\frac{df}{dt}\Big|1 = 3$, $g(1) = 5$, $\frac{dA}{dt}\Big|1 = 2$. What is $\frac{dg}{dt}\Big|1$?

Example 4.7.3.

[**author=duckworth**, file =**text_files/related_rates**]
The formula for the volume of a sphere is $V = \frac{4}{3}\pi r^3$.

1. Find the formula for $\frac{dV}{dt}$.

2. Suppose you know that the radius of the sphere is 5, and it is increasing at a rate of 10m/s. How fast is the volume increasing?

3. Suppose that you know that the radius of the sphere is 10, and that the volume is decreasing at a rate of $-3\text{m}^3/\text{s}$. How fast is the radius decreasing?

Example 4.7.4.

[author=duckworth, file =text_files/related_rates]
There are two cars, one going east and one going south.

1. Find a formula for the distance D between the cars in terms of x and y.
2. Find a formula for $\frac{dD}{dt}$. (Hint: if you can't figure out where to put $\frac{dx}{dt}$ and $\frac{dy}{dt}$ think about where the chain rule says you should put the derivative of the inside.)
3. Suppose you know that car A is travelling at 60 m/h and is 100 miles from the starting point. Suppose you know that car B is travelling 30 m/h and is 50 miles from the starting point. Find how fast the distance between the cars is increasing.
4. Suppose you know that the distance between the cars is increasing at the rate of 37 m/h. Suppose you know that car A is 75 miles from the starting point and going 30 m/h. Suppose you know that car B is 55 miles from the starting point. How fast is car B going

Example 4.7.5.

[author=duckworth, file =text_files/related_rates]
The volume of a cone is given by $V = \frac{1}{3}\pi r^2 h$ where r is the radius of the cone and h is the height.

1. Find a formula for $\frac{dV}{dt}$.
2. Suppose $\frac{dV}{dt} = 3$, $r = 2$, $\frac{dr}{dt} = 5$ and $h = 7$. Find $\frac{dh}{dt}$.
3. Suppose you know that the volume of water is 1000, and that the height is 10. Suppose also that you know that the radius is increasing at a rate of 1 and that the volume is increasing at a rate of 5. How fast is the height increasing?

Discussion.

[author=duckworth, file =text_files/related_rates]
The basic idea here is that we have a formula, and the letters in the formula stand for some function of t. We can take $\frac{d}{dt}$ of both sides of the formula and treat every letter as a function of time. Then you plug numbers into every spot except the one you're solving for. Then you solve for the unknown.

In general, I emphasize the formula first, and taking the derivative. Afterwards I go back to the problem and see how to plug the numbers in.

Example 4.7.6.

[author=duckworth, file =text_files/related_rates]

A ladder is leaning against the wall, and sliding downwards. The ladder is 10 feet long.

The equation is $x^2 + y^2 = 10$. Taking $\frac{d}{dt}$ of both sides gives $2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0$. (Note, we need $\frac{dx}{dt}$ because x is some function of t. If we knew the formula for x we could write the formula for $\frac{dx}{dt}$.)

Now suppose that you know the base is 2 feet from the wall and moving at the rate of $\frac{1}{4}$ ft/sec. How fast is the top sliding down? We plug these numbers in and we have $2 \cdot 2 \cdot \frac{1}{4} + 2y\frac{dy}{dt} = 0$. We need to get rid of y before we can solve for $\frac{dy}{dt}$. Use the Pythagorean theorem: $2^2 + y^2 = 10^2$ so $y = \sqrt{100-4} = 9.8$. Then we have $2 \cdot 2 \cdot \frac{1}{4} + 2 \cdot 9.8 \cdot \frac{dy}{dt} = 0$ whence $\frac{dy}{dt} = -.05$ ft/sec.

Notice that we took $\frac{d}{dt}$ of both sides first, and then plugged in numbers. In general, you cannot plug in any numbers before taking $\frac{d}{dt}$ unless they are numbers which cannot change in the problem.

Exercises

1. Suppose that x, y are both functions of t, and that $x^2 + y^2 = 25$. Express $\frac{dx}{dt}$ in terms of x, y, and $\frac{dy}{dt}$. When $x = 3$ and $y = 4$ and $\frac{dy}{dt} = 6$, what is $\frac{dx}{dt}$?

2. A 2-foot tall dog is walking away from a streetlight which is on a 10-foot pole. At a certain moment, the tip of the dog's shadow is moving away from the streetlight at 5 feet per second. How fast is the dog walking at that moment?

3. A ladder 13 feet long leans against a house, but is sliding down. How fast is the top of the ladder moving at a moment when the base of the ladder is 12 feet from the house and moving outward at 10 feet per second?

4.8 The intermediate value theorem and finding roots

Theorem 4.8.1.
[author=Garret, file =text_files/intermediate_value_theorem]
If a function f is continuous *on an interval $[a, b]$ and if $f(a) < 0$ and $f(b) > 0$ (or vice-versa), then there is some third point c with $a < c < b$ so that $f(c) = 0$.*

Comment.

[author=garrett, file =text_files/intermediate_value_theorem]
The assertion of the Intermediate Value Theorem is something which is probably 'intuitively obvious', and is also *provably true*.

This result has many relatively 'theoretical' uses, but for our purposes can be used to give a crude but simple way to locate the roots of functions. There is a lot of guessing, or trial-and-error, involved here, but that is fair. Again, in this situation, it is to our advantage if we are reasonably proficient in using a calculator to do simple tasks like evaluating polynomials! If this approach to estimating roots is taken to its logical conclusion, it is called the method of *interval bisection*, for a reason we'll see below. We will not pursue this method very far, because there are *better* methods to use once we have invoked *this* just to get going.

Example 4.8.1.

[author=garrett, file =text_files/intermediate_value_theorem]
For example, we probably don't know a formula to solve the cubic equation

$$x^3 - x + 1 = 0.$$

But the function $f(x) = x^3 - x + 1$ is certainly continuous, so we can invoke the Intermediate Value Theorem as much as we'd like. For example, $f(2) = 7 > 0$ and $f(-2) = -5 < 0$, so we know that there is a root in the interval $[-2, 2]$. We'd like to cut down the size of the interval, so we look at what happens at the *midpoint*, bisecting the interval $[-2, 2]$: we have $f(0) = 1 > 0$. Therefore, since $f(-2) = -5 < 0$, we can conclude that there is a root in $[-2, 0]$. Since both $f(0) > 0$ and $f(2) > 0$, we can't say anything at this point about whether or not there are roots in $[0, 2]$. Again *bisecting* the interval $[-2, 0]$ where we know there is a root, we compute $f(-1) = 1 > 0$. Thus, since $f(-2) < 0$, we know that there is a root in $[-2, -1]$ (and have no information about $[-1, 0]$).

Example 4.8.2.

[author=garrett, file =text_files/intermediate_value_theorem]
If we continue with this method, we can obtain as good an approximation as we want! But there are faster ways to get a really good approximation, as we'll see.

Unless a person has an amazing intuition for polynomials, there is really no way to anticipate what guess is better than any other in getting started.

Invoke the Intermediate Value Theorem to find an interval of length 1 or less in which there is a root of $x^3 + x + 3 = 0$: Let $f(x) = x^3 + x + 3$. Just, guessing, we compute $f(0) = 3 > 0$. Realizing that the x^3 term probably 'dominates' f when x is large positive or large negative, and since we want to find a point where f is negative, our next guess will be a 'large' negative number: how about -1? Well, $f(-1) = 1 > 0$, so evidently -1 is not negative enough. How about -2? Well, $f(-2) = -7 < 0$, so we have succeeded. Further, the failed guess -1 actually was worthwhile, since now we know that $f(-2) < 0$ and $f(-1) > 0$. Then, invoking the Intermediate Value Theorem, there is a root in the interval $[-2, -1]$.

Of course, typically polynomials have several roots, but *the number of roots of a polynomial is never more than its degree.* We can use the Intermediate Value Theorem to get an idea where *all* of them are.

Invoke the Intermediate Value Theorem to find *three different intervals* of length 1 or less in each of which there is a root of $x^3 - 4x + 1 = 0$: first, just starting anywhere, $f(0) = 1 > 0$. Next, $f(1) = -2 < 0$. So, since $f(0) > 0$ and $f(1) < 0$, there is at least one root in $[0, 1]$, by the Intermediate Value Theorem. Next, $f(2) = 1 > 0$. So, with some luck here, since $f(1) < 0$ and $f(2) > 0$, by the Intermediate Value Theorem there is a root in $[1, 2]$. Now if we somehow imagine that there is a *negative root* as well, then we try -1: $f(-1) = 4 > 0$. So we know *nothing* about roots in $[-1, 0]$. But continue: $f(-2) = 1 > 0$, and still no new conclusion. Continue: $f(-3) = -14 < 0$. Aha! So since $f(-3) < 0$ and $f(2) > 0$, by the Intermediate Value Theorem there is a *third* root in the interval $[-3, -2]$.

Notice how even the 'bad' guesses were not entirely wasted.

4.9 Newton's method

Discussion.

[`author=duckworth`, file =`text_files/newtons_method`]
Recall: The equation of the tangent line of $f(x)$ at $x = a$ is given by:

$$y = f'(a)(x - a) + f(a).$$

Definition 4.9.1.

[`author=duckworth,uses=linear_approximation`, file =`text_files/newtons_method`]
Linear approximation. Let $f(x)$ be a function and $L(x)$ the equation of it's tangent line at $x = a$. Then linear approximation states that

$$f(x) \cong L(x) \quad \text{for } x \text{ near } a.$$

Example 4.9.1.

[`author=duckworth,uses=sin,uses=linear_approximation`, file =`text_files/newtons_method`]
Let $f(x) = \sin(x)$. Then the equation of the tangent line at $x = 0$ is $L(x) = x$. Then $\sin(x) \cong x$ for x near 0. If you like, make a table of some values of $y_1 = \sin(x)$ and $y_2 = x$ for x near 0. (By the way, this explains why $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$.)

Rule 4.9.1.

[author=duckworth,uses=newtons_method, file =text_files/newtons_method]
One step Newton's method. Let $f(x)$ be a function and $L(x)$ the equation of the tangent line at $x = a$. Suppose $f(x)$ has a x-intercept near $x = a$. Then

$$\text{the solution of } f(x) = 0 \quad \text{is approximately the solution of } L(x) = 0.$$

Notice that it may be hard to solve $f(x) = 0$ but it is always easy to solve $L(x) = 0$ because this is a line.

Example 4.9.2.

[author=duckworth,uses=cos,uses=newtons_method, file =text_files/newtons_method]
Let $f(x) = x + \cos(x)$. Then $f(x) = 0$ has a solution near $x = 0$. The equation of the tangent line at $x = 0$ is $L(x) = x + 1$. I can't solve $x + \cos(x) = 0$, but I can solve $x + 1 = 0$. This gives $x = -1$. This is close to the solution of $x + \cos(x) = 0$. To make it more accurate, repeat this whole process, starting at $x = -1$. The equation of the tangent line is $L(x) = 1.841(x + 1) - .459$. Solving $1.841(x + 1) - .459 = 0$ gives $x = -.75$ which is very close to the exact solution of $f(x) = 0$.

Rule 4.9.2.

[author=duckworth,uses=newtons_method, file =text_files/newtons_method]
Multi-step Newton's method. Any tangent line at $x = a$ has equation $y = f'(a)(x - a) + f(a)$. Solving this for the x-intercept gives

$$x = a - \frac{f(a)}{f'(a)}.$$

You can iterate this process. Start with any $x = a_1$, then $a_2 = a_1 - f(a_1)/f'(a_1)$. Now that you have a_2 you can get $a_3 = a_2 - f(a_2)/f'(a_2)$. Etc.

Note that this formula can easily be adapted to being run on a computer.

Program.

[author=duckworth,uses=program, file =text_files/newtons_method]
These directions are for the `TI-83`, although similar directions would work on a variet of calculators and even computer systems.

To use the following program you must enter $y_1 = f(x)$ before running the program.

Hit PRGRM , choose `NEW` , and enter the name "NEWT" or something like that. To get `Input` you hit (while editing a program) PRGRM , choose `I/O` , then hit 1. (To get out of a menu while editing a program you can hit QUIT , which takes you back to the main screen, or CLEAR , which takes you back to the program.) To get a space after "GUESS" and "STEPS" you hit the green symbol right above the 0 button. To get `For` you hit PRGRM , then choose 4.

To get → you hit the STO→ button (right above ON). To get **End** you hit PRGRM and choose 7. To get **Disp** you hit PRGRM and choose **I/O** and then choose 3. To get y_1 you hit VARS , then choose y-vars, then choose 1. To get **nDeriv(** you hit MATH, then choose **nDeriv(** .

```
:Input ‘‘GUESS ’’,X
:Input ‘‘STEPS ’’,N
:For(I,1,N)
:X-Y1/nDeriv( (Y1,X,X)→X
:Disp X
:End
```

Example 4.9.3.

[author=duckworth,uses=e^x,uses=newtons_method, file =text_files/newtons_method]
Using the program, find an approximation of the solution of $x + e^x = 0$. Start with $x = 0$ and run five steps. , $y_2 = 1 + e^x$. Run NEWT with an initial guess of $x = 0$, and try 5 steps. You should get $-.567$. If you look at the graph of y_1 this should be very close to the x-intercept.

By the way, if you want to run it again you can just hit enter after you've run the program, but before you hit anything else.

Discussion.

[author=garrett,uses=newtons_method, file =text_files/newtons_method]
This is a method which, once you get started, quickly gives a very good approximation to a root of polynomial (and other) equations. The idea is that, if x_o is *not* a root of a polynomial equation, but is pretty close to a root, then *sliding down the tangent line at x_o to the graph of f gives a good approximation to the actual root.* The point is that this process can be repeated as much as necessary to give as good an approximation as you want.

Derivation.

[author=garrett,uses=newtons_method, file =text_files/newtons_method]
Let's derive the relevant formula: if our blind guess for a root of f is x_o, then the tangent line is

$$y - f(x_o) = f'(x_o)(x - x_o)$$

'Sliding down' the tangent line to hit the x-axis means to find the intersection of this line with the x-axis: this is where $y = 0$. Thus, we solve for x in

$$0 - f(x_o) = f'(x_o)(x - x_o)$$

to find

$$x = x_o - \frac{f(x_o)}{f'(x_o)}$$

Well, let's call this *first serious guess* x_1. Then, repeating this process, the *second serious guess* would be

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

and generally, if we have the nth guess x_n then the $n+1$th guess x_{n+1} is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

OK, that's the formula for *improving our guesses*. How do we decide when to quit? Well, it depends upon to how many decimal places we want our approximation to be good. Basically, if we want, for example, 3 decimal places accuracy, then as soon as x_n and x_{n+1} *agree* to three decimal places, we can presume that those are the *true* decimals of the true root of the equation. This will be illustrated in the examples below.

Comment.

[author=garrett, file =text_files/newtons_method]
It is important to realize that there is some uncertainty in Newton's method, both because it alone cannot assure that we have a root, and also because the idea just described for approximating roots to a given accuracy is not foolproof. But to worry about what could go wrong here is counter-productive.

Example 4.9.4.

[author=garrett,uses=newtons_method, file =text_files/newtons_method]
Approximate a root of $x^3 - x + 1 = 0$ using the intermediate value theorem to get started, and then Newton's method:

First let's see what happens if we are a little foolish here, in terms of the 'blind' guess we start with. If we ignore the advice about using the intermediate value theorem to *guarantee* a root in some known interval, we'll waste time. Let's see: The general formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

becomes

$$x_{n+1} = x_n - \frac{x^3 - x + 1}{3x^2 - 1}$$

If we take $x_1 = 1$ as our 'blind' guess, then plugging into the formula gives

$$x_2 = 0.5$$

$$x_3 = 3$$

$$x_4 = 2.0384615384615383249$$

This is discouraging, since the numbers are jumping around somewhat. But if we are stubborn and can compute quickly with a calculator (not by hand!), we'd see what happens:

$$\begin{aligned}
x_5 &= 1.3902821472167361527\\
x_6 &= 0.9116118977179270555\\
x_7 &= 0.34502849674816926662\\
x_8 &= 1.4277507040272707783\\
x_9 &= 0.94241791250948314662\\
x_{10} &= 0.40494935719938018881\\
x_{11} &= 1.7069046451828553401\\
x_{12} &= 1.1557563610748160521\\
x_{13} &= 0.69419181332954971175\\
x_{14} &= -0.74249429872066285974\\
x_{15} &= -2.7812959406781381233\\
x_{16} &= -1.9827252470441485421\\
x_{17} &= -1.5369273797584126484\\
x_{18} &= -1.3572624831877750928\\
x_{19} &= -1.3256630944288703144\\
x_{20} &= -1.324718788615257159\\
x_{21} &= -1.3247179572453899876
\end{aligned}$$

Well, after quite a few iterations of 'sliding down the tangent', the last two numbers we got, x_{20} and x_{21}, agree to 5 decimal places. This would make us think that the *true* root is *approximated to five decimal places* by -1.32471.

The stupid aspect of this little scenario was that our initial 'blind' guess was *too far from an actual root*, so that there was some wacky jumping around of the numbers before things settled down. If we had been computing by hand this would have been hopeless.

Let's try this example again using the Intermediate Value Theorem to pin down a root with some degree of accuracy: First, $f(1) = 1 > 0$. Then $f(0) = +1 > 0$ also, so we might *doubt* that there is a root in $[0, 1]$. Continue: $f(-1) = 1 > 0$ again, so we might *doubt* that there is a root in $[-1, 0]$, either. Continue: at last $f(-2) = -5 < 0$, so since $f(-1) > 0$ by the Intermediate Value Theorem we do indeed know that there is a root between -2 and -1. Now to start using *Newton's Method*, we would reasonably guess

$$x_o = -1.5$$

since this is the midpoint of the interval on which we know there is a root. Then computing by Newton's method gives:

$$\begin{aligned}
x_1 &= -1.3478260869565217295\\
x_2 &= -1.3252003989509069104\\
x_3 &= -1.324718173999053672\\
x_4 &= -1.3247179572447898011
\end{aligned}$$

so right away we have what appears to be 5 decimal places accuracy, in 4 steps rather than 21. Getting off to a good start is important.

Example 4.9.5.

[`author=garrett,uses=newtons_method`, file =`text_files/newtons_method`]
Approximate *all three* roots of $x^3-3x+1 = 0$ using the intermediate value theorem to get started, and then Newton's method. Here you have to take a little care in choice of beginning 'guess' for Newton's method:

In this case, since we are *told* that there are three roots, then we should certainly be wary about where we start: presumably we have to start in different places in order to successfully use Newton's method to find the different roots. So, starting thinking in terms of the intermediate value theorem: letting $f(x) = x^3 - 3x + 1$, we have $f(2) = 3 > 0$. Next, $f(1) = -1 < 0$, so we by the Intermediate Value Theorem we know there is a root in $[1, 2]$. Let's try to approximate it pretty well before looking for the other roots: The general formula for Newton's method becomes

$$x_{n+1} = x_n - \frac{x^3 - 3x + 1}{3x^2 - 3}$$

Our initial 'blind' guess might reasonably be the midpoint of the interval in which we know there is a root: take

$$x_o = 1.5$$

Then we can compute

$$\begin{aligned} x_1 &= 1.5333333333333333437 \\ x_2 &= 1.5320906432748537807 \\ x_3 &= 1.5320888862414665521 \\ x_4 &= 1.5320888862379560269 \\ x_5 &= 1.5320888862379560269 \\ x_6 &= 1.5320888862379560269 \end{aligned}$$

So it appears that we have quickly approximated a root in that interval! To what looks like 19 decimal places!

Continuing with this example: $f(0) = 1 > 0$, so since $f(1) < 0$ we know by the intermediate value theorem that there is a root in $[0, 1]$, since $f(1) = -1 < 0$. So as our blind gues let's use the midpoint of this interval to start Newton's Method: that is, now take $x_o = 0.5$:

$$\begin{aligned} x_1 &= 0.33333333333333337034 \\ x_2 &= 0.3472222222222222654 \\ x_3 &= 0.34729635316386797683 \\ x_4 &= 0.34729635533386071788 \\ x_5 &= 0.34729635533386060686 \\ x_6 &= 0.34729635533386071788 \\ x_7 &= 0.34729635533386060686 \\ x_8 &= 0.34729635533386071788 \end{aligned}$$

so we have a root evidently approximated to 3 decimal places after just 2 applications of Newton's method. After 8 applications, we have apparently 15 correct decimal places.

Discussion.

[author=livshits,uses=newtons_method,uses=sqrt, file =text_files/newtons_method]
We will consider first a well known method for calculating an approximate value of $\sqrt{a}$. The idea is to start with a crude guess x_1 and then to improve the approximation by taking $x_2 = (x_1 + a/x_1)/2$, then to improve it again by taking $x_3 = (x_2 + 2/x_2)/2$ and so on, in general we take

$$x_{n+1} = (x_n + a/x_n)/2.$$

Let us try to figure out how fast the approximation improves. We get: $x_{n+1}^2 - a = (x_n + a/x_n)^2/4 - a = (x_n^2 + 2a + a^2/x_n^2)/4 - a = (x_n^2 - 2a + a^2/x_n^2)/4 =$

$(x_n^2 - a)^2/(4x_n^2)$, and therefore, assuming that $x_n^2 \approx a$,

$$x_{n+1}^2 - a \approx (x_n^2 - a)^2/(4a).$$

So, roughly speaking, every iteration doubles the number of accurate decimal places in the approximation if the approximation is good enough to begin with. If the approximation is not good – then, starting with the second iteration, we will get twice closer to the solution every time we turn the crank.

This trick was already known to the Babylonians about 4000 years ago (see pp. 21-23 in Analysis by Its History). By looking at it from a more modern perspective we will arrive at the Newton's method. Here is how.

Derivation.

[`author=livshits,uses=newtons_method`, file =`text_files/newtons_method`]
Assume that we have an approximate solution x_n to the equation t

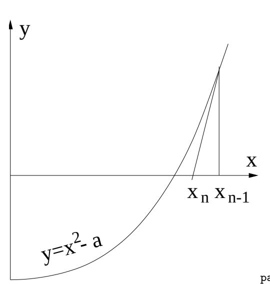
$$f(x) = 0 \tag{4.1}$$

where f is ULD. For x close to x_n $f(x)$ is well approximated by $f(x_n) + f'(x_n)(x - x_n)$, so we may hope that the solution to the approximate equation

$$f(x_n) + f'(x_n)(x - x_n) = 0 \tag{4.2}$$

will be a good approximation to the solution of our original equation. But the approximate equation is easy to solve because it is linear. Its solution is

$$x_{n+1} = x_n - f(x_n)/f'(x_n) \tag{4.3}$$

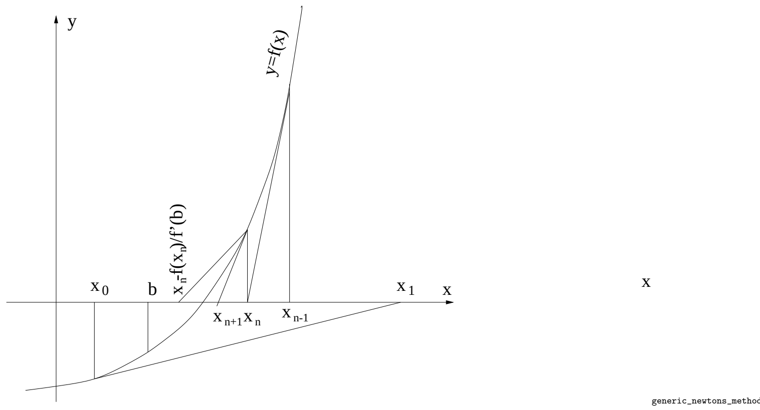


parab_newtons_method

Discussion.

[`author=livshits`, file =`text_files/newtons_method`]
Now we want to show that Newton's method always works for a ULD f that changes sign and has a positive and increasing derivative.

Assume that we start with the original guess x_0, then calculate x_1 using 4.3 with $n = 0$, then by taking $n = 1$ in 4.3 we get x_2, then x_3 by taking $n = 2$ and so on. Notice that $f(x_1) \geq 0$ no matter what x_0 is.



We can see next that for $n \geq 1$ we will have $x_{n+1} \leq x_n$, so the sequence $x_1, x_2, ..., x_n...$ will be decreasing.

On the other hand, there is b such that $f(b) < 0$ (we assumed that f changes sign), therefore, since f is increasing (because $f' > 0$), we can conclude that $b < x_n$. It follows that for any given $t > 0$ there will be m such that $x_m - x_{m+1} < t$ (otherwise $b < x_n$ will break), whence we will have $f(x_m) = (x_m - x_{m+1})f'(x_m) < tf'(x_m) \leq tf'(x_1)$, and for any $n > m$ it will be $0 \leq f(x_n) \leq f(x_m) \leq tf'(x_1)$. Now we can take t small enough for the fast convergence mentioned in exercise 3 to kick in and demonstrate that Newton's method works. Here are some details. By taking $a = x_n$ and $x = x_{n+1}$ in estimate **??** from section 2.4, and taking into account the equation 4.2 and the formula 4.3, we get

$$|f(x_{n+1})| \leq K(x_n - x_{n+1})^2 = K\left(\frac{f(x_n)}{f'(x_n)}\right)^2 < \frac{K}{f'(b)^2} f(x_n)^2 = Mf(x_n)^2,$$

where $M = K/f'(b)^2$ is a (positive) constant. Now, if $M < 10^k$ and $|f(x_n)| < 10^{-l}$, then $|f(x_{n+1})| < 10^{k-2l}$. To estimate how well x_n approximates the true solution we notice that $f(x_n - f(x_n)/f'(b)) \leq 0$, while $f(x_n) \geq 0$ (for $n > 0$), therefore the true solution will be between $x_n - f(x_n)/f'(b)$ and x_n, and will be not farther than $f(x_n)/f'(b)$ from x_n.

Comment.

[`author=livshits`, file =`text_files/newtons_method`]
A few remarks are in order here.

1. As you may have noticed, all the action took place on the segment $[b, x_1]$, so we can assume that the constant K that appeared in our finite analysis of approximation, is good only for this segment.

2. We assumed that the (there can be only one) true solution to the eqution was between $x_n - f(x_n)/f'(b)$ and x_n without justifying that assumption. It is clear that the solution can not be anywhere outside of $[x_n - f(x_n)/f'(b),\ x_n]$, but we haven't shown that it exists. To do it requires some properties of the real numbers that we will discuss later. For now we can be content that Newton's method allows us to get an approximate solution of as high quality as we want, and rather quickly at the final stage of the computation.

3. The whole argument was a bit heavy, it can be made more elegant by using convergence of sequences, we will learn later about this powerful tool.

4. While Newton's method is really good for making a good approximation to the solution much better, its perfomance may become very sluggish if the original approximation is not good.

Exercises

1. Approximate a root of $x^3 - x + 1 = 0$ using the intermediate value theorem to get started, and then Newton's method.

2. Approximate a root of $3x^4 - 16x^3 + 18x^2 + 1 = 0$ using the intermediate value theorem to get started, and then Newton's method. You might have to be sure to get sufficiently close to a root to start so that things don't 'blow up'.

3. Approximate *all three* roots of $x^3 - 3x + 1 = 0$ using the intermediate value theorem to get started, and then Newton's method. Here you have to take a little care in choice of beginning 'guess' for Newton's method.

4. Approximate the unique positive root of cos $x = x$.

5. Approximate a root of $e^x = 2x$.

6. Approximate a root of $\sin x = \ln x$. Watch out.

7. Try to prove that the algorithm given in the text gives better and better approxmiations of the squar root of a number. Try to prove it, also see what happens when $a = 0$, play with a calculator and try to understand what is going on).

8. Check that if we take $f(x) = x^2 - a$ we will arrive at the same Babylonian formula that we started with.

9. Investigate how Newton's iteration will improve the approximate solution, assuming that $f' > c > 0$ and the approximation that we start with is good enough. Do some calculations to get a feel for the performance of the method. Hint: use the inequality **??** from section 2.4 together with 4.3 to estimate $f(x_{n+1})$ and then to estimate $|x_{n+1} - x_{n+2}|$ in terms of $|x_n - x_{n+1}|$.

10. Now we want to show that Newton's method always works for a ULD f that changes sign and has a positive and increasing derivative.

    Assume that we start with the original guess x_0, then calculate x_1 using 4.3 with $n = 0$, then by taking $n = 1$ in 4.3 we get x_2, then x_3 by taking $n = 2$ and so on. Notice that $f(x_1) \geq 0$ no matter what x_0 is.

    Look at the diagram and see why, then prove it.

11. We can see next that for $n \geq 1$ we will have $x_{n+1} \leq x_n$, so the sequence $x_1, x_2, ..., x_n...$ will be decreasing. Prove it

12. 4) While Newton's method is really good for making a good approximation to the solution much better, its perfomance may become very sluggish if the original approximation is not good.

    Play with the equation $e^x = 2$ to see that.

13. See what can go wrong when different conditions on f don't hold, for example, when $f(x) = e^x$ or $f(x) = x + \sqrt{x^2 + 1}$ or $f(x) = x^2 + 1$ or $f(x) = x^{1/3}$.

4.10 L'Hospital's rule

Discussion.

[author=garrett, file =text_files/lhospitals_rule]
L'Hospital's rule is the definitive way to simplify evaluation of limits. It does not directly evaluate limits, but only *simplifies evaluation if used appropriately.*

In effect, this rule is the ultimate version of 'cancellation tricks', applicable in situations where a more down-to-earth genuine algebraic cancellation may be hidden or invisible.

Rule 4.10.1.

[author=garrett, file =text_files/lhospitals_rule]
Suppose we want to evaluate

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

where the limit a could also be $+\infty$ or $-\infty$ in addition to 'ordinary' numbers. Suppose that *either*

$$\lim_{x\to a} f(x) = 0 \text{ and } \lim_{x\to a} g(x) = 0$$

or

$$\lim_{x\to a} f(x) = \pm\infty \text{ and } \lim_{x\to a} g(x) = \pm\infty$$

(The $\pm$'s don't have to be the same sign). Then we cannot just 'plug in' to evaluate the limit, and these are traditionally called **indeterminate forms**. The unexpected trick that works often is that (amazingly) we are entitled to *take the derivative of both numerator and denominator:*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

No, this is *not the quotient rule*. No, it is not so clear why this would help, either, but we'll see in examples.

Example 4.10.1.

[author=garrett, file =text_files/lhospitals_rule]
Find $\lim_{x\to 0} (\sin x)/x$: both numerator and denominator have limit 0, so we are entitled to apply L'Hospital's rule:

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{\cos x}{1}$$

In the new expression, *neither* numerator nor denominator is 0 at $x = 0$, and we can just plug in to see that the limit is 1.

Example 4.10.2.

[author=garrett, file =text_files/lhospitals_rule]
Find $\lim_{x\to 0} x/(e^{2x} - 1)$: both numerator and denominator go to 0, so we are entitled to use L'Hospital's rule:

$$\lim_{x\to 0} \frac{x}{e^{2x} - 1} = \lim_{x\to 0} \frac{1}{2e^{2x}}$$

In the new expression, the numerator and denominator are both non-zero when $x = 0$, so we just plug in 0 to get

$$\lim_{x\to 0} \frac{x}{e^{2x}-1} = \lim_{x\to 0} \frac{1}{2e^{2x}} = \frac{1}{2e^0} = \frac{1}{2}$$

Example 4.10.3.

[author=garrett, file =text_files/lhospitals_rule]
Find $\lim_{x\to 0^+} x \ln x$: The 0^+ means that we approach 0 from the positive side, since otherwise we won't have a real-valued logarithm. This problem illustrates the *possibility* as well as *necessity* of *rearranging* a limit to make it be a *ratio* of things, in order to legitimately apply L'Hospital's rule. Here, we rearrange to

$$\lim_{x\to 0^+} x \ln x = \lim_{x\to 0} \frac{\ln x}{1/x}$$

In the new expressions the top goes to $-\infty$ and the bottom goes to $+\infty$ as x goes to 0 (from the right). Thus, we are entitled to apply L'Hospital's rule, obtaining

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

Now it is very necessary to rearrange the expression inside the last limit: we have

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

The new expression is very easy to evaluate: the limit is 0.

Comment.

[author=garrett, file =text_files/lhospitals_rule]
It is often necessary to apply L'Hospital's rule repeatedly: Let's find $\lim_{x\to +\infty} x^2/e^x$: both numerator and denominator go to ∞ as $x \to +\infty$, so we are entitled to apply L'Hospital's rule, to turn this into

$$\lim_{x\to +\infty} \frac{2x}{e^x}$$

But still both numerator and denominator go to ∞, so apply L'Hospital's rule again: the limit is

$$\lim_{x\to +\infty} \frac{2}{e^x} = 0$$

since now the numerator is fixed while the denominator goes to $+\infty$.

Example 4.10.4.

[author=garrett, file =text_files/lhospitals_rule]
Now let's illustrate more ways that things can be rewritten as ratios, thereby possibly making L'Hospital's rule applicable. Let's evaluate

$$\lim_{x\to 0} x^x$$

It is less obvious now, but we can't just plug in 0 for x: on one hand, we are taught to think that $x^0 = 1$, but also that $0^x = 0$; but then surely 0^0 can't be both at once. And this exponential expression is not a ratio.

The trick here is to *take the logarithm*:

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

The reason that we are entitled to *interchange* the logarithm and the limit is that *logarithm is a continuous function* (on its domain). Now we use the fact that $\ln(a^b) = b\ln a$, so the log of the limit is

$$\lim_{x\to 0^+} x\ln x$$

Aha! The question has been turned into one we already did! But ignoring that, and repeating ourselves, we'd first rewrite this as a ratio

$$\lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \frac{\ln x}{1/x}$$

and then apply L'Hospital's rule to obtain

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

But we have to remember that we've computed the *log* of the limit, not the limit. Therefore, the actual limit is

$$\lim_{x\to 0^+} x^x = e^{\text{ log of the limit }} = e^0 = 1$$

This trick of taking a logarithm is important to remember.

Example 4.10.5.

[author=garrett, file =text_files/lhospitals_rule]
Here is another issue of rearranging to fit into accessible form: Find

$$\lim_{x\to +\infty} \sqrt{x^2+x+1} - \sqrt{x^2+1}$$

This is not a ratio, but certainly is 'indeterminate', since it is the difference of two expressions both of which go to $+\infty$. To make it into a ratio, we take out the largest reasonable power of x:

$$\lim_{x\to +\infty} \sqrt{x^2+x+1} - \sqrt{x^2+1} = \lim_{x\to +\infty} x\cdot(\sqrt{1+\frac{1}{x}+\frac{1}{x^2}} - \sqrt{1+\frac{1}{x^2}})$$

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

The last expression here fits the requirements of the L'Hospital rule, since both numerator and denominator go to 0. Thus, by invoking L'Hospital's rule, it becomes

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

This is a large but actually tractable expression: multiply top and bottom by x^2, so that it becomes

$$= \lim_{x\to+\infty} 12 + 1x\sqrt{1+1x+1x^2} + \frac{-1}{x}\sqrt{1+\frac{1}{x^2}}$$

At this point, we *can* replace every $\frac{1}{x}$ by 0, finding that the limit is equal to

$$\frac{\frac{1}{2}+0}{\sqrt{1+0+0}} + \frac{0}{\sqrt{1+0}} = \frac{1}{2}$$

It is important to recognize that in additional to the actual application of L'Hospital's rule, it may be necessary to *experiment* a little to get things to settle out the way you want. *Trial-and-error is not only ok, it is necessary.*

Exercises

1. Find $\lim_{x\to 0} (\sin\ x)/x$
2. Find $\lim_{x\to 0} (\sin\ 5x)/x$
3. Find $\lim_{x\to 0} (\sin(x^2))/x^2$
4. Find $\lim_{x\to 0}\ x/(e^{2x}-1)$
5. Find $\lim_{x\to 0}\ x\ \ln x$
6. Find
$$\lim_{x\to 0^+} (e^x-1)\ln x$$
7. Find
$$\lim_{x\to 1} \frac{\ln x}{x-1}$$
8. Find
$$\lim_{x\to+\infty} \frac{\ln\ x}{x}$$
9. Find
$$\lim_{x\to+\infty} \frac{\ln x}{x^2}$$
10. Find $\lim_{x\to 0} (\sin x)^x$

4.11 Exponential growth and decay: a differential equation

Discussion.

[author=garrett, file =text_files/expon_growth_diffeq]
This little section is a tiny introduction to a very important subject and bunch of ideas: *solving differential equations*. We'll just look at the simplest possible example of this.

The general idea is that, instead of solving equations to find unknown *numbers*, we might solve equations to find unknown *functions*. There are many possibilities for what this might mean, but one is that we have an unknown function y of x and are given that y and its derivative y' (with respect to x) satisfy a relation

$$y' = ky$$

where k is some constant. Such a relation between an unknown function and its derivative (or *derivatives*) is what is called a **differential equation**. Many basic 'physical principles' can be written in such terms, using 'time' t as the independent variable.

Having been taking derivatives of exponential functions, a person might remember that the function $f(t) = e^{kt}$ has exactly this property:

$$\frac{d}{dt}e^{kt} = k \cdot e^{kt}$$

For that matter, any *constant multiple* of this function has the same property:

$$\frac{d}{dt}(c \cdot e^{kt}) = k \cdot c \cdot e^{kt}$$

And it turns out that these really are *all* the possible solutions to this differential equation.

There is a certain buzz-phrase which is supposed to alert a person to the occurrence of this little story: if a function f has **exponential growth** or **exponential decay** then that is taken to mean that f can be written in the form

$$f(t) = c \cdot e^{kt}$$

If the constant k is *positive* it has exponential *growth* and if k is *negative* then it has exponential *decay*.

Since we've described all the solutions to this equation, what questions remain to ask about this kind of thing? Well, the usual scenario is that some *story problem* will give you information in a way that requires you to take some trouble in order to *determine the constants* c, k. And, in case you were wondering where you get to take a derivative here, the answer is that you don't really: all the 'calculus work' was done at the point where we granted ourselves that all solutions to that differential equation are given in the form $f(t) = ce^{kt}$.

First to look at some general ideas about determining the constants before getting embroiled in story problems: One simple observation is that

$$c = f(0)$$

that is, that the constant c is the value of the function at time $t = 0$. This is true simply because

$$f(0) = ce^{k \cdot 0} = ce^0 = c \cdot 1 = c$$

from properties of the exponential function.

Example 4.11.1.

[**author=garrett**, file =**text_files/expon_growth_diffeq**]
More generally, suppose we know the values of the function at two different times:

$$y_1 = ce^{kt_1}$$

$$y_2 = ce^{kt_2}$$

Even though we certainly do have 'two equations and two unknowns', these equations involve the unknown constants in a manner we may not be used to. But it's still not so hard to solve for c, k: dividing the first equation by the second and using properties of the exponential function, the c on the right side cancels, and we get

$$\frac{y_1}{y_2} = e^{k(t_1-t_2)}$$

Taking a logarithm (base e, of course) we get

$$\ln y_1 - \ln y_2 = k(t_1 - t_2)$$

Dividing by $t_1 - t_2$, this is

$$k = \frac{\ln y_1 - \ln y_2}{t_1 - t_2}$$

Substituting back in order to find c, we first have

$$y_1 = ce^{\frac{\ln y_1 - \ln y_2}{t_1 - t_2}t_1}$$

Taking the logarithm, we have

$$\ln y_1 = \ln c + \frac{\ln y_1 - \ln y_2}{t_1 - t_2}t_1$$

Rearranging, this is

$$\ln c = \ln y_1 - \frac{\ln y_1 - \ln y_2}{t_1 - t_2}t_1 = \frac{t_1 \ln y_2 - t_2 \ln y_1}{t_1 - t_2}$$

Therefore, in summary, the two equations

$$y_1 = ce^{kt_1}$$

$$y_2 = ce^{kt_2}$$

allow us to solve for c, k, giving

$$k = \frac{\ln y_1 - \ln y_2}{t_1 - t_2}$$

$$c = e^{\frac{t_1 \ln y_2 - t_2 \ln y_1}{t_1 - t_2}}$$

A person might manage to remember such formulas, or it might be wiser to remember the way of *deriving* them.

Example 4.11.2.

[`author=garrett`, file =`text_files/expon_growth_diffeq`]
A herd of llamas has 1000 llamas in it, and the population is growing exponentially. At time $t = 4$ it has 2000 llamas. Write a formula for the number of llamas at *arbitrary* time t.

Here there is no direct mention of differential equations, but use of the buzz-phrase *'growing exponentially'* must be taken as indicator that we are talking about the situation

$$f(t) = ce^{kt}$$

where here $f(t)$ is the number of llamas at time t and c, k are constants to be determined from the information given in the problem. And the use of language should probably be taken to mean that at time $t = 0$ there are 1000 llamas, and at time $t = 4$ there are 2000. Then, either repeating the method above or plugging into the formula derived by the method, we find

$$c = \text{ value of } f \text{ at } t = 0 \ = 1000$$

$$k = \frac{\ln f(t_1) - \ln f(t_2)}{t_1 - t_2} = \frac{\ln 1000 - \ln 2000}{0 - 4}$$

$$= \ln \frac{1000}{2000} - 4 = \frac{\ln \frac{1}{2}}{-4} = (\ln 2)/4$$

Therefore,

$$f(t) = 1000\, e^{\frac{\ln 2}{4} t} = 1000 \cdot 2^{t/4}$$

This is the desired formula for the number of llamas at arbitrary time t.

Example 4.11.3.

[`author=garrett`, file =`text_files/expon_growth_diffeq`]
A colony of bacteria is growing exponentially. At time $t = 0$ it has 10 bacteria in it, and at time $t = 4$ it has 2000. At what time will it have $100,000$ bacteria?

Even though it is not explicitly demanded, we need to find the general formula for the number $f(t)$ of bacteria at time t, set this expression equal to $100,000$, and solve for t. Again, we can take a *little* shortcut here since we know that $c = f(0)$ and we are given that $f(0) = 10$. (This is easier than using the bulkier more general formula for finding c). And use the formula for k:

$$k = \frac{\ln f(t_1) - \ln f(t_2)}{t_1 - t_2} = \frac{\ln 10 - \ln 2,000}{0 - 4} = \frac{\ln \frac{10}{2,000}}{-4} = \frac{\ln 200}{4}$$

Therefore, we have

$$f(t) = 10 \cdot e^{\frac{\ln 200}{4}\ t} = 10 \cdot 200^{t/4}$$

as the general formula. Now we try to solve

$$100,000 = 10 \cdot e^{\frac{\ln 200}{4}\ t}$$

for t: divide both sides by the 10 and take logarithms, to get

$$\ln 10,000 = \frac{\ln 200}{4}\, t$$

Thus,

$$t = 4\,\frac{\ln 10,000}{\ln 200} \approx 6.953407835$$

Exercises

1. A herd of llamas is growing exponentially. At time $t = 0$ it has 1000 llamas in it, and at time $t = 4$ it has 2000 llamas. Write a formula for the number of llamas at *arbitrary* time t.

2. A herd of elephants is growing exponentially. At time $t = 2$ it has 1000 elephants in it, and at time $t = 4$ it has 2000 elephants. Write a formula for the number of elephants at *arbitrary* time t.

3. A colony of bacteria is growing exponentially. At time $t = 0$ it has 10 bacteria in it, and at time $t = 4$ it has 2000. At what time will it have $100,000$ bacteria?

4. A colony of bacteria is growing exponentially. At time $t = 2$ it has 10 bacteria in it, and at time $t = 4$ it has 2000. At what time will it have $100,000$ bacteria?

4.12 The second and higher derivatives

Definition 4.12.1.

[author=garrett, file =text_files/higher_derivs]
The **second derivative** of a function is simply *the derivative of the derivative*. The **third derivative** of a function is the derivative of the second derivative. And so on.

The second derivative of a function $y = f(x)$ is written as

$$y'' = f''(x) = \frac{d^2}{dx^2} f = \frac{d^2 f}{dx^2} = \frac{d^2 y}{dx^2}$$

The third derivative is

$$y''' = f'''(x) = \frac{d^3}{dx^3} f = \frac{d^3 f}{dx^3} = \frac{d^3 y}{dx^3}$$

And, generally, we can put on a 'prime' for each derivative taken. Or write

$$\frac{d^n}{dx^n} f = \frac{d^n f}{dx^n} = \frac{d^n y}{dx^n}$$

for the nth derivative. There is yet another notation for high order derivatives where the number of 'primes' would become unwieldy:

$$\frac{d^n f}{dx^n} = f^{(n)}(x)$$

as well.

The geometric interpretation of the higher derivatives is subtler than that of the first derivative, and we won't do much in this direction, except for the next little section.

Exercises

1. Find $f"(x)$ for $f(x) = x^3 - 5x + 1$.
2. Find $f"(x)$ for $f(x) = x^5 - 5x^2 + x - 1$.
3. Find $f"(x)$ for $f(x) = \sqrt{x^2 - x + 1}$.
4. Find $f"(x)$ for $f(x) = \sqrt{x}$.

4.13 Inflection points, concavity upward and downward

Definition 4.13.1.

[author=garrett, file =text_files/concavity_etc]
A **point of inflection** of the graph of a function f is a point where the *second* derivative f'' is 0. We have to wait a minute to clarify the geometric meaning of this.

A piece of the graph of f is **concave upward** if the curve 'bends' upward. For example, the popular parabola $y = x^2$ is concave upward in its entirety.

A piece of the graph of f is **concave downward** if the curve 'bends' downward. For example, a 'flipped' version $y = -x^2$ of the popular parabola is concave downward in its entirety.

The relation of *points of inflection* to *intervals where the curve is concave up or down* is exactly the same as the relation of *critical points* to *intervals where the function is increasing or decreasing*. That is, the points of inflection mark the boundaries of the two different sort of behavior Further, only one sample value of f'' need be taken between each pair of consecutive inflection points in order to see whether the curve bends up or down along that interval.

Rule 4.13.1.

[author=garrett, file =text_files/concavity_etc]
Expressing this as a systematic procedure: *to find the intervals along which f is concave upward and concave downward:*

- Compute the *second* derivative f'' of f, and *solve* the equation $f''(x) = 0$ for x to find all the inflection points, which we list in order as $x_1 < x_2 < \ldots < x_n$. (Any points of discontinuity, etc., should be added to the list!)
- We need some *auxiliary points*: To the left of the leftmost inflection point x_1 pick any convenient point t_o, between each pair of consecutive inflection points x_i, x_{i+1} choose any convenient point t_i, and to the right of the rightmost inflection point x_n choose a convenient point t_n.
- Evaluate the *second derivative* f'' at all the *auxiliary* points t_i.
- Conclusion: if $f''(t_{i+1}) > 0$, then f is *concave upward* on (x_i, x_{i+1}), while if $f''(t_{i+1}) < 0$, then f is *concave downward* on that interval.
- Conclusion: on the 'outside' interval $(-\infty, x_o)$, the function f is *concave upward* if $f''(t_o) > 0$ and is *concave downward* if $f''(t_o) < 0$. Similarly, on (x_n, ∞), the function f is *concave upward* if $f''(t_n) > 0$ and is *concave downward* if $f''(t_n) < 0$.

Example 4.13.1.

[author=garrett, file =text_files/concavity_etc]
Find the inflection points and intervals of concavity up and down of

$$f(x) = 3x^2 - 9x + 6$$

First, the second derivative is just $f''(x) = 6$. Since this is never zero, there are *not* points of inflection. And the value of f'' is always 6, so is always > 0, so the curve is entirely *concave upward*.

Example 4.13.2.

[author=garrett, file =text_files/concavity_etc]
Find the inflection points and intervals of concavity up and down of

$$f(x) = 2x^3 - 12x^2 + 4x - 27$$

First, the second derivative is $f''(x) = 12x - 24$. Thus, solving $12x - 24 = 0$, there is just the one inflection point, 2. Choose auxiliary points $t_o = 0$ to the left of the inflection point and $t_1 = 3$ to the right of the inflection point. Then $f''(0) = -24 < 0$, so on $(-\infty, 2)$ the curve is concave *downward*. And $f''(2) = 12 > 0$, so on $(2, \infty)$ the curve is concave *upward*.

Example 4.13.3.

[author=garrett, file =text_files/concavity_etc]
Find the inflection points and intervals of concavity up and down of

$$f(x) = x^4 - 24x^2 + 11$$

the second derivative is $f''(x) = 12x^2 - 48$. Solving the equation $12x^2 - 48 = 0$, we find inflection points ± 2. Choosing auxiliary points $-3, 0, 3$ placed between and to the left and right of the inflection points, we evaluate the second derivative: First, $f''(-3) = 12 \cdot 9 - 48 > 0$, so the curve is concave *upward* on $(-\infty, -2)$. Second, $f''(0) = -48 < 0$, so the curve is concave *downward* on $(-2, 2)$. Third, $f''(3) = 12 \cdot 9 - 48 > 0$, so the curve is concave *upward* on $(2, \infty)$.

Exercises

1. Find the inflection points and intervals of concavity up and down of $f(x) = 3x^2 - 9x + 6$.

2. Find the inflection points and intervals of concavity up and down of $f(x) = 2x^3 - 12x^2 + 4x - 27$.

3. Find the inflection points and intervals of concavity up and down of $f(x) = x^4 - 2x^2 + 11$.

4.14 Another differential equation: projectile motion

Discussion.

[author=garrett, file =text_files/projectile_motion_diffeq]
Here we encounter the fundamental idea that *if* $s = s(t)$ *is position, then* $\dot{s}$ *is velocity, and* $\ddot{s}$ *is acceleration.* This idea occurs in all basic physical science and engineering.

Derivation.

[author=garrett, file =text_files/projectile_motion_diffeq]
In particular, for a projectile near the earth's surface travelling straight up and down, ignoring air resistance, acted upon by no other forces but *gravity*, we have

$$\text{acceleration due to gravity } = -32 \text{ feet/sec }^2$$

Thus, letting $s(t)$ be position at time t, we have

$$\ddot{s}(t) = -32.$$

We take this (approximate) *physical fact* as our starting point.

From $\ddot{s} = -32$ we *integrate* (or *anti-differentiate*) once to undo one of the derivatives, getting back to *velocity*:

$$v(t) = \dot{s} = \dot{s}(t) = -32t + v_o$$

where we are calling the *constant of integration* 'v_o'. (No matter which constant v_o we might take, the derivative of $-32t + v_o$ with respect to t is -32.)

Specifically, when $t = 0$, we have

$$v(o) = v_o$$

Thus, the constant of integration v_o is **initial velocity**. And we have this formula for the velocity at *any* time in terms of *initial* velocity.

We integrate once more to undo the last derivative, getting back to the *position* function itself:

$$s = s(t) = -16t^2 + v_o t + s_o$$

where we are calling the constant of integration 's_o'. Specifically, when $t = 0$, we have

$$s(0) = s_o$$

so s_o is **initial position**. Thus, we have a formula for position at *any* time in terms of *initial position* and *initial velocity.*

Of course, in many problems the data we are given is *not* just the initial position and initial velocity, but something else, so we have to determine these constants indirectly.

Exercises

1. You drop a rock down a deep well, and it takes 10 seconds to hit the bottom. How deep is it?

2. You drop a rock down a well, and the rock is going 32 feet per second when it hits bottom. How deep is the well?

3. If I throw a ball straight up and it takes 12 seconds for it to go up and come down, how high did it go?

4.15 Graphing rational functions, asymptotes

Discussion.

[author=garrett, file =text_files/graphing_with_calculus]
This section shows another kind of function whose graphs we can understand effectively by our methods.

Definition 4.15.1.

[author=garrett, file =text_files/graphing_with_calculus]
There is one new item here, the idea of *asymptote* of the graph of a function.

A **vertical asymptote** of the graph of a function f most commonly occurs when f is defined as a *ratio* $f(x) = g(x)/h(x)$ of functions g, h continuous at a point x_o, but with the denominator going to zero at that point while the numerator doesn't. That is, $h(x_o) = 0$ but $g(x_o) \neq 0$. Then we say that f *blows up* at x_o, and that the line $x = x_o$ is a **vertical asymptote** of the graph of f.

And as we take x closer and closer to x_o, the graph of f zooms off (either up or down or both) *closely to the line* $x = x_o$.

Example 4.15.1.

[author=garrett, file =text_files/graphing_with_calculus]
A very simple example of this is $f(x) = 1/(x-1)$, whose denominator is 0 at $x = 1$, so causing a *blow-up* at that point, so that $x = 1$ is a *vertical asymptote*. And as x approaches 1 from the right, the values of the function zoom *up* to $+\infty$. When x approaches 1 from the *left*, the values zoom *down* to $-\infty$.

Definition 4.15.2.

[author=garrett, file =text_files/graphing_with_calculus]
A **horizontal asymptote** of the graph of a function f occurs if either limit

$$\lim_{x\to+\infty} f(x)$$

or

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

exists. If $R = \lim_{x\to+\infty} f(x)$, then $y = R$ is a **horizontal asymptote** of the function, and if $L = \lim_{x\to-\infty} f(x)$ exists then $y = L$ is a horizontal asymptote.

As x goes off to $+\infty$ the graph of the function gets closer and closer to the horizontal line $y = R$ if *that* limit exists. As x goes of to $-\infty$ the graph of the function gets closer and closer to the horizontal line $y = L$ if *that* limit exists.

So in rough terms *asymptotes* of a function are *straight lines* which the graph of the function approaches *at infinity*. In the case of *vertical asymptotes*, it is the y-coordinate that goes off to infinity, and in the case of *horizontal asymptotes* it is the x-coordinate which goes off to infinity.

Example 4.15.2.

[**author=garrett**, file =**text_files/graphing_with_calculus**]
Find asymptotes, critical points, intervals of increase and decrease, inflection points, and intervals of concavity up and down of $f(x) = \frac{x+3}{2x-6}$: First, let's find the asymptotes. The denominator is 0 for $x = 3$ (and this is *not* cancelled by the numerator) so the line $x = 3$ is a *vertical asymptote*. And as x goes to $\pm\infty$, the function values go to $1/2$, so the line $y = 1/2$ is a horizontal asymptote.

The derivative is

$$f'(x) = \frac{1 \cdot (2x-6) - (x+3) \cdot 2}{(2x-6)^2} = \frac{-12}{(2x-6)^2}$$

Since a ratio of polynomials can be zero only if the numerator is zero, this $f'(x)$ can *never* be zero, so there are *no critical points*. There is, however, the discontinuity at $x = 3$ which we must take into account. Choose auxiliary points 0 and 4 to the left and right of the discontinuity. Plugging in to the derivative, we have $f'(0) = -12/(-6)^2 < 0$, so the function is *decreasing* on the interval $(-\infty, 3)$. To the right, $f'(4) = -12/(8-6)^2 < 0$, so the function is also decreasing on $(3, +\infty)$.

The second derivative is $f''(x) = 48/(2x-6)^3$. This is never zero, so there are *no inflection points*. There is the discontinuity at $x = 3$, however. Again choosing auxiliary points $0, 4$ to the left and right of the discontinuity, we see $f''(0) = 48/(-6)^3 < 0$ so the curve is *concave downward* on the interval $(-\infty, 3)$. And $f''(4) = 48/(8-6)^3 > 0$, so the curve is concave *upward* on $(3, +\infty)$.

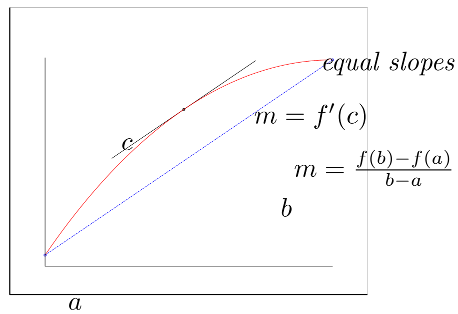
Plugging in just two or so values into the function then is enough to enable a person to make a fairly good qualitative sketch of the graph of the function.

Exercises

1. Find all asymptotes of $f(x) = \frac{x-1}{x+2}$.
2. Find all asymptotes of $f(x) = \frac{x+2}{x-1}$.
3. Find all asymptotes of $f(x) = \frac{x^2-1}{x^2-4}$.
4. Find all asymptotes of $f(x) = \frac{x^2-1}{x^2+1}$.

4.16 The Mean Value Theorem

Mean value theorem 4.16.1.
[author=duckworth, file =text_files/mean_value_theorem]
Suppose that $f(x)$ is a continuous function on the interval $[a, b]$ and differentiable on the interval (a, b). Then there is a number c, between a and b, such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.



mean_ value_ theorem

Comment.

[author=duckworth, file =text_files/mean_value_theorem]
The main use we will have for theorem is to prove the Fundamental Theorem of Calculus. However, we can do some concrete examples.

Example 4.16.1.

[author=duckworth, file =text_files/mean_value_theorem]
Let $f(x) = x + \sin(x)$. Consider $x = 0$ and $x = \pi$. Can we find a number c such that $f'(c) = \frac{\pi+\sin(\pi)-0-0}{\pi}$? The theorem tells us that we will be able to solve this (not necessarily algebraically): $f'(x) = 1 + \cos(x) = \frac{\pi}{\pi} = 1$. I.e. there is a number x, between 0 and π such that $1 + \cos(x) = 1$. In this case, it is easy to solve, namely let $x = \pi/2$. Again, the theorem tells us that even if the equation is not easy to solve, that there is some solution.

Comment.

[author=duckworth, file =text_files/mean_value_theorem]
The previous example was really stupid! Although lots of calculus books (ours included) have problems just like the previous one, that is not how the Mean Value Theorem is ever used! I mean, you could always set up the equation like in the previous example and then look at it and see if there is a solution.

So, if the MVT is *not* used to tell us that we can find a point where the derivative equals that formula using a, b, $f(a)$ and $f(b)$ then what does it do? I'll let you think for a minute. You've got an equation:

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

I've just told you that we *don't* use this to tell us about $f'(c)$. So it has to be the case that we use the equation to tell us about the right hand side! In other words, if we know something about $f'(c)$, *then* we can say something about $f(b) - f(a)$. This is incredibly important. It gives us a formula for how the derivative affects what we know about $f(x)$.

Before we had this equation, if I told you that $f'(x)$ is always ≥ 1, then all you would have been able to conclude is that $f(b)$ is always $\geq f(a)$ (since $f(x)$ is increasing). Now, I can tell you exactly how much bigger $f(b)$ has to be.

Example 4.16.2.

[author=duckworth, file =text_files/mean_value_theorem]
Suppose $a = 4$, $f(a) = 7$ and that we know $f'(x) \geq 1$ for all x. What can we say about $f(b)$ for $b \geq a$? We start with

$$\frac{f(b) - f(a)}{b - a} = f'(x) \geq 1$$

we drop the middle term "$f'(x)$", and multiply by $b - a$ to get:

$$f(b) - f(a) \geq b - a$$

whence $f(b) \geq f(a)+b-a$. Thus, for $b \geq 4$ we can say that $f(b) \geq 7+b-4 = b+3$.

Comment.

[author=duckworth, file =text_files/mean_value_theorem]
In the previous example we used the MVT to take information about $f'(x)$ and turn it into very specific, quantitive information about $f(x)$. This idea will be crucial when we prove the Fundamental Theorem of Calculus. In fact, you can alread imagine how the proof will go, in heuristic terms. In the previous example I used one piece of information about the derivative, namely that it was bigger than 1, to tell us one piece of information about $f(x)$ (when $x \geq 4$), namely that $f(x)$ was bigger than $x + 3$. Now, suppose I told you *exactly* what the derivative was at a whole bunch of points. Then you should be able to say more precisely what $f(x)$ is. If I told you what $f'(x)$ is at every point, then you should be able to say what $f(x)$ is every point.

Chapter 5

Integration

5.1 Basic integration formulas

Discussion.

[author=garrett, file =text_files/integration_basics]
The fundamental *use* of *integration* is as a *continuous version of summing*. But, paradoxically, often integrals are *computed* by viewing integration as essentially an *inverse operation to differentiation*. (That fact is the so-called *Fundamental Theorem of Calculus*.)

The notation, which we're stuck with for historical reasons, is as peculiar as the notation for derivatives: the **integral of a function** $f(x)$ **with respect to** x is written as

$$\int f(x)\,dx$$

The remark that integration is (almost) an inverse to the operation of differentiation means that if

$$\frac{d}{dx}F(x) = f(x)$$

then

$$\int f(x)\,dx = F(x) + C$$

The extra C, called the **constant of integration**, is really necessary, since after all differentiation kills off constants, which is why integration and differentiation are not *exactly* inverse operations of each other.

Rules 5.1.1.

[author=garrett, file =text_files/integration_basics]
Since integration is *almost* the inverse operation of differentiation, recollection of formulas and processes for *differentiation* already tells the most important

formulas for *integration:*

$$\begin{aligned}
\int x^n \, dx &= \frac{1}{n+1} x^{n+1} \quad \text{unless } n = -1 \\
\int e^x \, dx &= e^x \\
\int \frac{1}{x} \, dx &= \ln|x| \\
\int \sin x \, dx &= -\cos x \\
\int \cos x \, dx &= \sin x \\
\int \sec^2 x \, dx &= \tan x \\
\int \frac{1}{1+x^2} \, dx &= \arctan x
\end{aligned}$$

Rule 5.1.2.

[**author=garrett**, file =**text_files/integration_basics**]
Since the derivative of a sum is the sum of the derivatives, the *integral of a sum is the sum of the integrals:*

$$\int f(x) + g(x) \, dx = \int f(x) \, dx + \int g(x) \, dx.$$

Similarly, constants 'go through' the integral sign:

$$\int c \cdot f(x) \, dx = c \cdot \int f(x) \, dx$$

Example 5.1.1.

[**author=garrett**, file =**text_files/integration_basics**]
For example, it is easy to integrate polynomials, even including terms like $\sqrt{x}$ and more general power functions. The only thing to watch out for is terms $x^{-1} = \frac{1}{x}$, since these integrate to $\ln x$ instead of a power of x. So

$$\int 4x^5 - 3x + 11 - 17\sqrt{x} + \frac{3}{x} \, dx = \frac{4x^6}{6} - \frac{3x^2}{2} + 11x - \frac{17x^{3/2}}{3/2} + 3\ln x + C$$

Notice that we need to include just one 'constant of integration'.

Rule 5.1.3.

[**author=garrett**, file =**text_files/integration_basics**]

Other basic formulas obtained by reversing differentiation formulas:

$$\begin{aligned}\int a^x\ dx &= \frac{a^x}{\ln a}\\ \int \log_a\ x\ dx &= \frac{1}{\ln a}\cdot\frac{1}{x}\\ \int \frac{1}{\sqrt{1-x^2}}\ dx &= \arcsin x\\ \int \frac{1}{x\sqrt{x^2-1}}\ dx &= \sec^{-1}(x)\end{aligned}$$

Example 5.1.2.

[`author=garrett`, file =`text_files/integration_basics`]
Sums of constant multiples of all these functions are easy to integrate: for example,

$$\int 5\cdot 2^x - \frac{23}{x\sqrt{x^2-1}} + 5x^2\ dx = \frac{5\cdot 2^x}{\ln 2} - 23\,\mathrm{arcsec}\,x + \frac{5x^3}{3} + C$$

Discussion.

[`author=wikibooks`, file =`text_files/integration_basics`]
When we examined differentiation, we found that, graphically, the derivative of a function at a point gives us the gradient of the curve at that point. When we examine integration, we find two important uses for this finding what function yields, under differentiation, a given function finding the area under a curve

---

Discussion.

[`author=wikibooks`, file =`text_files/integration_basics`]
One example of how to interpretation of the derivative is that it gives the velocity of an object from its position. We now want to reverse this process and find the position of the object from its velocity.

Suppose v is a constant velocity and let d be position. Then $d = vt$. However, if v is *not* constant, then this formula does not work.

So we need to take a different approach. What we do is to break up the time into small chunks of time delta t, and then find the distance by summing over small chunks.

$$d = v(t_0)\Delta t + v(t_0 + \Delta t)\Delta t + v(t_0 + 2\Delta t)\Delta t.$$

Now we make the time chunks smaller and smaller until we have approximated the smooth curve.

Since we have a new function we need a new set of symbols to represent it, this looks like

---

Discussion.

[author=wikibooks, file =text_files/integration_basics]
Let $f(x)$ be a function. The **anti-derivative** of $f(x)$ is another function $F(x)$ such that the derivative of $F(x)$ equals $f(x)$.

Simple anti-derivatives can be found by guessing, or "thinking backwards".

Example 5.1.3.

[author=duckworth, file =text_files/integration_basics]
Let's find an anti-derivative by guessing, or "thinking backwards". Let $f(x) = x^2$. Can we guess what function we would take the derivative of to get x^2?

Well, at least in your head, check the derivative of a bunch of our basic functions. You should quickly decide that we won't get x^2 by taking the derivative of e^x, $\ln(x)$, $\sin(x)$, etc. We need to take the derivative of a power of x in order to get x^2.

In fact, we will have to take the derivative of something of the form x^3 in order to get x^2. So let's check: how close to the right answer is $F(x) = x^3$. Well, $\frac{d}{dx}F(x) = \frac{d}{dx}x^3 = 3x^2$. We're trying to get x^2, not $3x^2$, so we need to change our guess for $F(x)$ a little bit. We want to cancel the 3. A little more thought leads to our next guess of $F(x) = \frac{x^3}{3}$. Now it's easy to check that $\frac{d}{dx}F(x) = \frac{d}{dx}\frac{x^3}{3} = 3\frac{x^2}{3} = x^2$.

So, we've done it, $F(x) = \frac{x^3}{3}$ is an anti-derivative of x^2.

Is $F(x)$ the only solution of this problem? Well, if you go back through our thought process above you can see that no other power of x will work, and the coefficent has to be $\frac{1}{3}$. Well, you can change $F(x)$ by adding something whose derivative will be 0. Thus, $F(x) = \frac{x^3}{3} + 12$, or $F(x) = \frac{x^3}{3} - 13427$, are also anti-derivatives.

In general, every function of the form $F(x) = \frac{x^3}{3} + C$ is an anti-derivative of x^2.

Example 5.1.4.

[author=wikibooks, file =text_files/integration_basics]
Let us consider the example $f(x) = 6x^2$. How would we go about finding the integral of this function? Recall the rule from differentiation that $Dx^n = nx^{n-1}$ In our circumstance, we have $Dx^3 = 3x^2$ This is a start! We now know that the function we seek will have a power of 3 in it. How would we get the constant of 6? Well, $2Dx^3 = 2(3x^2)$ $D2x^3 = 6x^2$

Thus, we say that $2x^3$ is the integral of $6x^2$. We write it, generally, $\int 6x^2 dx$, or in terms of the differential operator, $D^{-1}(6x^2)$.

There is an important fact that needs to be kept in mind when we are integrating. Let us examine the above example, that $\int 6x^2 dx = 2x^3$ This is true in the fact that differentiating $2x^3$ yields $6x^2$, but this is not the only solution we also have $2x^3 + 1$, $2x^3 + 2$, even $2x^3 - 98999$ giving us this same solution! Constants "disappear" on differentiation– so we generally write the integral of a function with an arbitrary constant added to the end to show all the possible solutions. So we write the full equation as $\int 6x^2 dx = 2x^3 + C$

The method we have described is terribly ad-hoc, but we will be able to generalize it, and obtain the polynomial formula in the next section.

Rule 5.1.4.

[author=duckworth, file =text_files/integration_basics]
Let's find the anti-derivative of $f(x) = x^n$ where n can be any power. One way to do this is guessing. You can probably guess and check the answer right now. It's also kind of cute to figure this out by reversing the steps of differentiation. For this purpose let's write down exactly what happens for powers of x.

Derivative of power of x:
Step 1: Multiply by power of x
Step 2: Subtract 1 from power of x

Now, I'm going to reverse each of these rules, starting at the end and moving backwards (i.e. Step 1 for the anti-derivative will undo step 2 for the derivative, etc.).

Anti-derivative of power of x:
Step 1: Add 1 to power of x
Step 2: Divide by the power of x

O.k., now we can make a formula out of the verbal description we've just found. The anti-derivative of x^n is $\frac{x^{n+1}}{n+1}$. Note, this formula will not be defined if $n+1 = 0$. Putting this all together, we have the following rule:

$$\int x^n \, dx = \frac{x^{n+1}}{n+1} \text{ if } n \neq -1.$$

Comment.

[author=duckworth, file =text_files/integration_basics]
For basic problems like the ones we're learing now, it should always be very easy to chek if your formula $F(x)$ for the anti-derivative of $f(x)$ is correct: You just take the derivative of $F(x)$ and see if you get $f(x)$.

Example 5.1.5.

[author=duckworth, file =text_files/integration_basics]
Check that the anti-derivative of $\ln(x)$ is $x\ln(x) - x$. (Note: I don't expect that you could have *found* this anti-derivative by guessing. You will learn techniques in Calculus II that can help you find this anti-derivative.)

We check:

$$\frac{d}{dx} x\ln(x) - x = 1 \cdot \ln(x) + x \cdot \frac{1}{x} - 1 = \ln(x) + 1 - 1 = \ln(x).$$

Derivation.

[author=wikibooks, file =text_files/integration_basics]

In this section we will concern ourselves with determining the integrals of other functions, such as $\sin(x)$, $\cos(x)$, $\tan(x)$, and others.

Recall the following $D\sin(x) = \cos(x)$ $D\cos(x) = -\sin(x)$ $D\tan(x) = (\sec(x))^2$

and given the above rule that $Df(x) = g(x)$, $\int g(x)dx = f(x) + C$

we instantly have the integrals of $\cos(x)$, $\sin(x)$, and $(secx)^2$ $\int \cos(x)\ dx = \sin(x) + C$ $\int \sin(x)\ dx = -\cos(x) + C$ $\int (\sec(x))^2 dx = \tan(x) + C$

Derivation.

[author=wikibooks, file =text_files/integration_basics]
Recall that when we integrate, we wish to solve the equation, given the function g $Df = g$ for the function f.

When we look at the exponential function e^x, we see immediately from the above result that $\int e^x dx = e^x + C$

Discussion.

[author=livshits, file =text_files/integration_basics]
In the previous two sections we have developed (somewhat heuristically) differentiation as an operation on functions. As soon as a new operation is introduced, it is reasonable to consider an inverse operation.

In case of differentiation this operation is (naturally) called *antidifferentiation*.

More specifically, a function F is an *antiderivative* or a *primitive* of f if f is the derivative of F, i.e. $F' = f$.

When $f(x)$ is the velocity at time x, the antiderivative $F(x)$ will be the distance, when $f(x)$ is the rate of change, $F(x)$ will be the total change.

Because the derivative of any constant is zero, there are (infinitely) many antiderivatives of a given function, we can add any constant C to F, F' doesn't change because $(F + C)' = F'$ (differentiation kills constants and antidifferentiation resurrects them).

The appearance of an arbitrary additive constant C is not surprising. The velocity doesn't depend on where we measure our distance from, and whether we measure the total change from yesterday or from 100 years ago, the rate will be the same, although the total change will be not.

Later on we will prove that by adding different constants to a fixed antiderivative we can get all of them. This fact would easily follow if we knew that any function with zero derivative is a constant. It looks obviuos, but to prove it one has to take a closer look at differentiation, we will do it in section **??**.

Notation.

[author=livshits, file =text_files/integration_basics]
Meanwhile we will assume that it is true and introduce the notation

$$\int f(x)dx$$

for the set of all the antiderivatives of a given function f. This set is also called *the indefinite integral of* f. Since all the antiderivatives of f are of the form $F+C$ where F is one of them and C is a constant, we can write

$$\int f(x)dx = F(x) + C$$

C is called the **integration constant**.

Example 5.1.6.

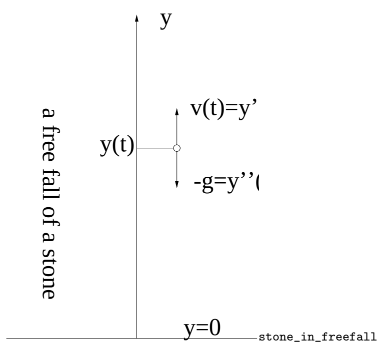
[author=livshits, file =text_files/integration_basics]
We will start with a simple problem of motion under gravity.

A stone is thrown vertically with the original velocity v_0. Find the motion of the stone, given its original position y_0.

The motion of the stone will be described by a function of time $y(t)$ that will satisfy the equation

$$y'' = -g$$

where y'' denotes the derivative of y' that is called *the second derivative* of y.



It follows from Newton's Second law: $F = ma$, where m is the mass of the stone, $a = y''$ is its acceleration and $F = -gm$ is the force of gravity. We also have two additional conditions:

$$y(0) = y_0 \text{ and } y'(0) = v_0.$$

The equation simly says that the acceleration equals to $-g$. We can find the velocity by integrating acceleration and using the initial velocity to get the integration constant. This gives us

$$v(t) = y'(t) = v_0 - gt.$$

To find the position we integrate the velocity and use the initial position to figure out the integration constant. By doing so we get

$$y(t) = y_0 + v_0 t - gt^2/2.$$

In case of zero initial velocity ($v_0 = 0$) the velocity and the position of the stone at time t will be:

$$v(t) = -gt \text{ and } y(t) = y_0 - gt^2/2.$$

In particular, it will take $T = \sqrt{2y_0/g}$ seconds for the stone to hit the ground. At that point its speed (which is the absolute value of the velocity) will be

$v(T) = gT = \sqrt{2gy_0}$. While the stone drops, it loses height, but it picks up speed. However, *the energy*

$$E = \frac{1}{2}mv^2 + mgy$$

will stay the same. The energy of the stone consists of 2 parts:

$$K = \frac{1}{2}mv^2$$

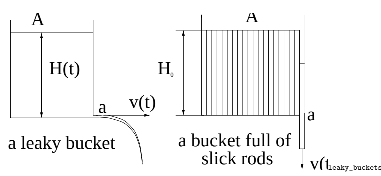
is called the *kinetic energy*, it is the energy of motion, it depends only on the speed, and

$$P = mgy$$

is called the *potential energy*, it depends only on the position of the stone. *Conservation of energy* is one of the most important principles in physics.

Example 5.1.7.

[`author=livshits`, file =`text_files/integration_basics`]
Assume there is a cylindrical bucket filled with water, and there is a small hole in the bottom. How long will it take for the bucket to get empty? The area of the horizontal cross-section of the bucket, the area of the hole and the original level of water in the bucket are given.



Let A be the area of the horizontal cross-section of the bucket, a be the area of the hole and H_0 be the original water level in the bucket. Assume that the hole in the bucket was opened up at time 0, so $H(0) = H_0$ where $H(t)$ is the water level at time t.

This problem of the detailed description of the flow of water is rather complicated, so we will add some simplifying assumptions to make things manageable. We first have to figure out how fast the wat water is squirting out of the hole, depending on the level of water in the bucket. Let us say it is squirting out at velocity v. If a small mass of water, say m, escapes through the hole, the mass of the water left in the bucket will be reduced by m, and the reduction will take place at the level H, so the potential energy of the water will drop by mgH. On the other hand, the kinetic energy of mass m of water moving at velocity v is $mv^2/2$, and the water that escapes has the potential energy zero because the hole is at the level zero. From the conservation of energy we must have

$$\frac{1}{2}mv^2 = mgH, \text{ therefore } v(t) = \sqrt{2gH(t)}. \tag{5.1}$$

In other words, the velocity $v(t)$ at which the water escapes is the same velocity that it would pick up by a free fall from level $H(t)$ to level 0 where the hole is (compate to the results from the previous problem). In deriving this formula for $v(t)$ we neglected a few things, such as the internal friction in water, the change in the flow pattern inside the bucket and the variations in velocity across the jet of water squirting out of the hole.

Now, after we get a handle on how fast the water is flowing out, it is easy to see how fast the water level will drop. Indeed, the rate of change of the volume of the water in the bucket is $-AH'(t)$, that must be equal to the rate at which the water passes through the hole, which is $av(t)$, and, using the formula 5.1 for $v(t)$, we get

$$H'(t) = -\frac{a}{A}\sqrt{2gH(t)}. \tag{5.2}$$

Dividing sides by $2\sqrt{H(t)}$ we can rewrite Equation **??** as

$$(\sqrt{H(t)})' = -\frac{a}{A}\sqrt{g/2}.$$

Taking into account that $H(0) = H_0$, we get

$$(\sqrt{H(t)}) = \sqrt{H_0} - t\frac{a}{A}\sqrt{g/2}.$$

Finally, solving the equation $H(T) = 0$ leads to

$$T = \frac{A}{a}\sqrt{2H_0/g}.$$

Let us take a closer look at this formula and see why it makes sense.

The case $a = A$ corresponds to the bottom of the bucket falling off, so all the water will be in a free fall. As we know, it will take just $\sqrt{2H_0/g}$ seconds for the water to drop the distance H_0, and that's exactly what the formula says.

The formula also says that the time it takes the bucket to empty out is proportional to the cross-section of the bucket and inversely proportional to the size of the hole, which makes sense.

Now assume that the bucket is slightly inclined and filled with a bunch of identical well lubricated metal rods, assume that each rod fits into the hole snugly, so it slides out, as soon as it gets to it (see the picture). It takes $\sqrt{2H_0/g}$ seconds for each rod to slide out, there are A/a of them that will fit into the bucket, and we arrive at the same formula for T.

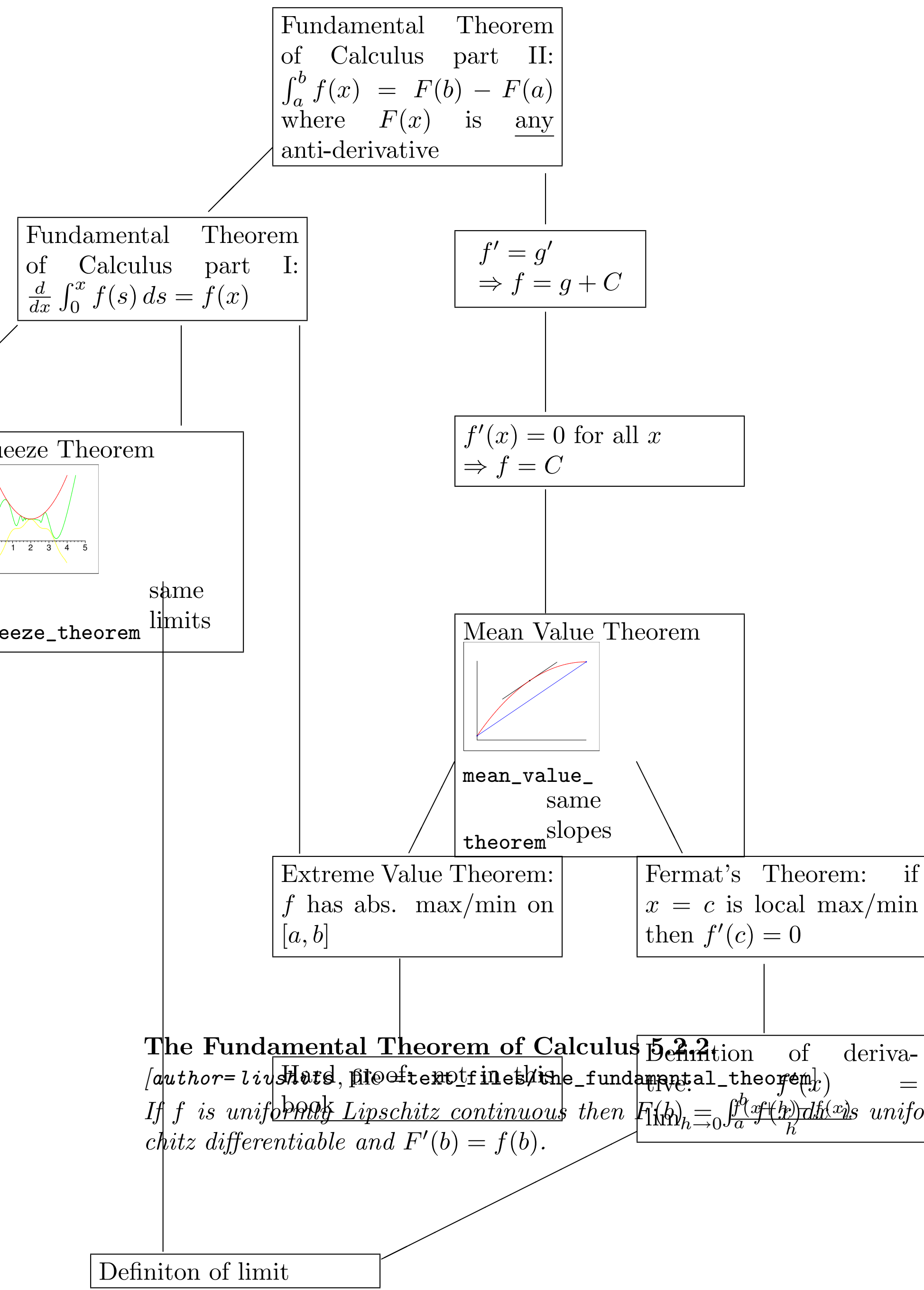
Exercises

1. $\int 4x^3 - 3\cos x + \frac{7}{x} + 2\; dx =?$
2. $\int 3x^2 + e^{2x} - 11 + \cos x\, dx =?$
3. $\int \sec^2 x\, dx =?$
4. $\int \frac{7}{1+x^2}\; dx =?$
5. $\int 16x^7 - \sqrt{x} + \frac{3}{\sqrt{x}}\; dx =?$
6. $\int 23\sin x - \frac{2}{\sqrt{1-x^2}}\; dx =?$

5.2 Introduction to the Fundamental Theorem of Calculus

Theorem 5.2.1.
[*author=wikibooks*, file =text_files/introduction_to_fundamental_theorem_calculus]
The Fundamental Theorem of Calculus states that If $\int_0^x f(t)\ dt = F(x),$ *then* $\int f(x)\ dx = F(x) + C,$ *and* $\int_a^b f(x)\ dx = F(b) - F(a)$ *for any continuous function* f.

This diagram shows the logical structure of the heart of Calculus. The main results are on top, the Fundamental Theorem of Calculus parts I and II. Each result is only true if the results below it are true, so the whole thing builds up piece by piece. It's pretty amazing that you can have something that works and that has this many logical steps, each one of which could break the whole structure! If one part of this were really false, there would probably be satellites falling out of the sky!!



The Fundamental Theorem of Calculus 5.2.2
[author=livshits, file =text_files/the_fundamental_theorem]
If f is uniformly Lipschitz continuous then $F(b) = \int_a^b f(t)\,dt$ is uniformly Lipschitz differentiable and $F'(b) = f(b)$.

Proof.
[author=livshits, file =text_files/the_fundamental_theorem]
We have to establish the inequality

$$|F(c) - F(b) - f(b)(c-b)| \leq K(c-b)^2$$

but by our integration rules the LHS can be rewritten as

$$\left|\int_b^c (f(x) - f(b))\,dx\right| \leq \int_b^c |f(x) - f(b)|\,dx \leq \int_b^c L|x-b|dx = (L/2)(b-c)^2$$

and we can take $K = L/2$ where L is the Lipschitz constant for f. □

5.3 The simplest substitutions

Comment.

[author=garrett, file =text_files/integration_simple_subst]
The simplest kind of chain rule application

$$\frac{d}{dx} f(ax+b) = a \cdot f'(x)$$

(for constants a, b) can easily be run backwards to obtain the corresponding integral formulas: some and illustrative important examples are

---

Examples 5.3.1.

[author=garrett, file =text_files/integration_simple_subst]

$$\begin{aligned}
\int \cos(ax+b)\,dx &= \frac{1}{a} \cdot \sin(ax+b) + C \\
\int e^{ax+b}\,dx &= \frac{1}{a} \cdot e^{ax+b} + C \\
\int \sqrt{ax+b}\,dx &= \frac{1}{a} \cdot \frac{(ax+b)^{3/2}}{3/2} + C \\
\int \frac{1}{ax+b}\,dx &= \frac{1}{a} \cdot \ln(ax+b) + C
\end{aligned}$$

Examples 5.3.2.

[author=garrett, file =text_files/integration_simple_subst]
Putting numbers in instead of letters, we have examples like

$$\begin{aligned}
\int \cos(3x+2)\,dx &= \frac{1}{3} \cdot \sin(3x+2) + C \\
\int e^{4x+3}\,dx &= \frac{1}{4} \cdot e^{4x+3} + C \\
\int \sqrt{-5x+1}\,dx &= \frac{1}{-5} \cdot \frac{(-5x+1)^{3/2}}{3/2} + C \\
\int \frac{1}{7x-2}\,dx &= \frac{1}{7} \cdot \ln(7x-2) + C
\end{aligned}$$

Comment.

[**author=garrett**, file =**text_files/integration_simple_subst**]
Since this kind of substitution is pretty undramatic, and a person should be able to do such things *by reflex* rather than having to think about it very much.

Rule 5.3.1.

[**author=livshits**, file =**text_files/integration_simple_subst**]

$$\int f(g(x))g'(x)dx = \int f(g)dg$$

in the right-hand side of this formula g is considered as an independent variable. The formula means that the equality holds if we plug $g = g(x)$ into the right-hand side after performing the integration.

Exercises

1. $\int e^{3x+2}\,dx = ?$
2. $\int \cos(2-5x)\,dx = ?$
3. $\int \sqrt{3x-7}\,dx = ?$
4. $\int \sec^2(2x+1)\,dx = ?$
5. $\int (5x^7 + e^{6-2x} + 23 + \frac{2}{x})\,dx = ?$
6. $\int \cos(7-11x)\,dx = ?$

5.4 Substitutions

Discussion.

[author=garrett, file =text_files/integration_subst]
The *chain rule* can also be 'run backward', and is called **change of variables** or **substitution** or sometimes **u-substitution**. Some examples of what happens are straightforward, but others are less obvious. It is at this point that the capacity to *recognize derivatives* from past experience becomes very helpful.

Examples 5.4.1.

[author=garrett, file =text_files/integration_subst]
Here are a variety of examples of simple backwards chain rules.

1. Since (by the chain rule)

$$\frac{d}{dx}e^{\sin x} = \cos x \, e^{\sin x},$$

then we can anticipate that

$$\int \cos x \, e^{\sin x} \, dx = e^{\sin x} + C$$

2. Since (by the chain rule)

$$\frac{d}{dx}\sqrt{x^5+3x} = \frac{1}{2}(x^5+3x)^{-1/2} \cdot (5x^4+3)$$

then we can anticipate that

$$\int \frac{1}{2}(5x^4+3)(x^5+3x)^{-1/2} \, dx = \sqrt{x^5+3x} + C$$

3. Since (by the chain rule)

$$\frac{d}{dx}\sqrt{5+e^x} = \frac{1}{2}(5+e^x)^{-1/2} \cdot e^x$$

then

$$\int e^x \, (5+e^x)^{-1/2} \, dx = 2\int \frac{1}{2}e^x(5+e^x)^{-1/2} \, dx = 2\sqrt{5+e^x} + C.$$

Notice how for 'bookkeeping purposes' we put the $\frac{1}{2}$ into the integral (to make the constants right there) and put a compensating 2 outside.

4. Since (by the chain rule)

$$\frac{d}{dx}\sin^7(3x+1) = 7 \cdot \sin^6(3x+1) \cdot \cos(3x+1) \cdot 3$$

then we have

$$\begin{aligned}\int \cos(3x+1) \, \sin^6(3x+1) \, dx &= \frac{1}{21}\int 7 \cdot 3 \cdot \cos(3x+1)\sin^6(3x+1) \, dx \\ &= \frac{1}{21}\sin^7(3x+1) + C\end{aligned}$$

Exercises

1. $\int \cos x \, \sin x \, dx = ?$
2. $\int 2x \, e^{x^2} \, dx = ?$
3. $\int 6x^5 \, e^{x^6} \, dx = ?$
4. $\int \frac{\cos x}{\sin x} \, dx = ?$
5. $\int \cos x e^{\sin x} \, dx = ?$
6. $\int \frac{1}{2\sqrt{x}} e^{\sqrt{x}} \, dx = ?$
7. $\int \cos x \sin^5 x \, dx = ?$
8. $\int \sec^2 x \tan^7 x \, dx = ?$
9. $\int (3 \cos x + x) \, e^{6 \sin x + x^2} \, dx = ?$
10. $\int e^x \sqrt{e^x + 1} \, dx = ?$

5.5 Area and definite integrals

Discussion.

[author=wikibooks, file =text_files/summation_notation]
Summation notation allows an expression that contains a sum to be expressed in a simple, compact manner. The greek letter sigma, Σ, is used to denote the sum of a set of numbers. A dummy variable is substituted into the expression sequentially, and the result is summed.

Examples 5.5.1.

[author=wikibooks, file =text_files/summation_notation]
It's easiest to learn summation notation by example, so we list a number of examples now.

1. $\sum_{i=1}^{5} i = 1 + 2 + 3 + 4 + 5$

   Here, the dummy variable is i, the lower limit of summation is 1, and the upper limit is 5.

2. $\sum_{j=2}^{7} j^2 = 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2$

   Here, the dummy variable is j, the lower limit of summation is 2, and the upper limit is 7.

3. The name of the dummy variable doesn't matter. For example, the following are all the same:

   $$\sum_{i=1}^{4} i = \sum_{j=1}^{4} j = \sum_{\alpha=1}^{4} \alpha = 1 + 2 + 3 + 4.$$

   This means we can change the name of the dummy variable whenever we like. Conventionally we use the letters i, j, k, m.

4. Sometimes, you will see summation signs with no dummy variable specified, e.g

   $$\sum_{1}^{4} i^3 = 100$$

   In such cases the correct dummy variable should be clear from the context.

5. You may also see cases where the limits are unspecified. Here too, they must be deduced from the context. For example, later we will always be studying infinite summations that start at 0 or 1, so

   $$\sum \frac{1}{n}$$

   would mean (in that context)

   $$\sum_{n=0}^{\infty} \frac{1}{n}$$

Examples 5.5.2.

[author=wikibooks, file =text_files/summation_notation]
Here are some common summations, together with a closed form formula for their sum (note: having a closed form formula is quite rare in general)

1. $\sum_{i=1}^{n} c = c + c + ... + c = nc$ (for any real number c)
2. $\sum_{i=1}^{n} i = 1 + 2 + 3 + ... + n = \frac{n(n+1)}{2}$
3. $\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + 3^2 + ... + n^2 = \frac{n(n+1)(2n+1)}{6}$
4. $\sum_{i=1}^{n} i^3 = 1^3 + 2^3 + 3^3 + ... + n^3 = \frac{n^2(n+1)^2}{4}$

Notation.

[author=wikibooks, file =text_files/summation_notation]
In order to avoid writing long sequences of numbers, mathemeticians use summation notation to denote sums of sequences.

$\sum_{k=1}^{n} f(k)$

denotes the sum of the values of f(k) for k=1, k=2, etc., up to k=n. For example,

$\sum_{k=1}^{4} 2k = (2 \cdot 1) + (2 \cdot 2) + (2 \cdot 3) + (2 \cdot 4) = 2 + 4 + 6 + 8 = 20$

This will become useful when defining areas under curves.

Definition 5.5.1.

[author=wikibooks, file =text_files/summation_notation]
Definition of Area The area under the graph of $f(x)$ from $x = a$ to $x = b$ is denoted by

$\int_a^b f(x)\ dx$

and is defined as

$\int_a^b f(x)\ dx = \lim_{n\to\infty} \frac{b-a}{n} \cdot \left[\sum_{k=1}^{n} f\left(\frac{k(b-a)}{n}\right)\right]$

Intuitively, this can be thought of as adding the areas of "bars" in the curve to obtain an approximation of the area, and it gets more accurate as the number of bars (n) increases.

Definition 5.5.2.

[author=garrett, file =text_files/area_defn_integrals]
The actual *definition* of 'integral' is as a limit of sums, which might easily be viewed as having to do with *area*. One of the original issues integrals were intended to address was computation of area.

First we need more notation. Suppose that we have a function f whose integral is another function F:

$$\int f(x)\,dx = F(x) + C$$

Let a, b be two numbers. Then the **definite integral** of f **with limits** a, b is

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

The left-hand side of this equality is just *notation* for the definite integral. The use of the word 'limit' here has little to do with our earlier use of the word, and means something more like 'boundary', just like it does in more ordinary English.

A similar notation is to write

$$[g(x)]_a^b = g(b) - g(a)$$

for any function g. So we could also write

$$\int_a^b f(x)\,dx = [F(x)]_a^b$$

Example 5.5.3.

[`author=garrett`, file =`text_files/area_defn_integrals`]
For example,

$$\int_0^5 x^2\,dx = [\frac{x^3}{3}]_0^5 = \frac{5^3 - 0^3}{3} = \frac{125}{3}$$

As another example,

$$\int_2^3 3x + 1\,dx = [\frac{3x^2}{2} + x]_2^3 = (\frac{3 \cdot 3^2}{2} + 3) - (\frac{3 \cdot 2^2}{2} + 2) = \frac{21}{2}$$

Comment.

[`author=garrett`, file =`text_files/area_defn_integrals`]
All the other integrals we had done previously would be called **indefinite integrals** since they didn't have 'limits' a, b. So a *definite* integral is just the difference of two values of the function given by an *indefinite* integral. That is, there is almost nothing new here except the idea of evaluating the function that we get by integrating.

But now we *can* do something new: compute *areas*:

For example, if a function f is *positive* on an interval $[a, b]$, then

$$\int_a^b f(x)\,dx = \text{ area between graph and } x\text{-axis, between } x = a \text{ and } x = b$$

It is important that the function be *positive*, or the result is false.

Example 5.5.4.

[author=garrett, file =text_files/area_defn_integrals]
For example, since $y = x^2$ is certainly always positive (or at least non-negative, which is really enough), the area 'under the curve' (and, implicitly, above the x-axis) between $x = 0$ and $x = 1$ is just

$$\int_0^1 x^2\,dx = [\frac{x^3}{3}]_0^1 = \frac{1^3 - 0^3}{3} = \frac{1}{3}$$

More generally, *the area below $y = f(x)$, above $y = g(x)$, and between $x = a$ and $x = b$ is*

$$\text{area...} \;=\int_a^b f(x) - g(x)\,dx$$

$$= \int_{\text{left limit}}^{\text{right limit}} (\text{upper curve - lower curve})\,dx$$

It is important that $f(x) \geq g(x)$ throughout the interval $[a, b]$.

For example, the area below $y = e^x$ and above $y = x$, and between $x = 0$ and $x = 2$ is

$$\int_0^2 e^x - x\,dx = [e^x - \frac{x^2}{2}]_0^2 = (e^2 - 2) - (e^0 - 0) = e^2 + 1$$

since it really is true that $e^x \geq x$ on the interval $[0, 2]$.

As a person might be wondering, in general it may be not so easy to tell whether the graph of one curve is above or below another. The procedure to examine the situation is as follows: given two functions f, g, to find the intervals where $f(x) \leq g(x)$ and vice-versa:

• Find where the graphs cross by solving $f(x) = g(x)$ for x to find the x-coordinates of the points of intersection.
• Between any two solutions x_1, x_2 of $f(x) = g(x)$ (and also to the left and right of the left-most and right-most solutions!), plug in *one* auxiliary point of your choosing to see which function is larger.

Of course, this procedure works for a similar reason that the *first derivative test* for local minima and maxima worked: we implicitly assume that the f and g are *continuous*, so if the graph of one is above the graph of the other, then the situation can't *reverse* itself without the graphs actually *crossing*.

Example 5.5.5.

[author=garrett, file =text_files/area_defn_integrals]
As an example, and as an example of a certain delicacy of wording, consider the problem to *find the area between $y = x$ and $y = x^2$ with* $0 \leq x \leq 2$. To find where $y = x$ and $y = x^2$ *cross*, solve $x = x^2$: we find solutions $x = 0, 1$. In the present problem we don't care what is happening to the left of 0. Plugging in the value 1/2 as auxiliary point between 0 and 1, we get $\frac{1}{2} \geq (\frac{1}{2})^2$, so we see that in $[0, 1]$ the curve $y = x$ is the higher. To the right of 1 we plug in the auxiliary point 2, obtaining $2^2 \geq 2$, so the curve $y = x^2$ is higher there.

Therefore, the area between the two curves has to be broken into two parts:

$$\text{area} \;= \int_0^1 (x - x^2)\,dx + \int_1^2 (x^2 - x)\,dx$$

since we must always be integrating in the form

$$\int_{\text{left}}^{\text{right}} \text{higher - lower } dx$$

In some cases the 'side' boundaries are redundant or only *implied*. For example, the question might be to *find the area between the curves* $y = 2 - x$ *and* $y = x^2$. What is implied here is that these two curves themselves enclose one or more *finite* pieces of area, without the need of any 'side' boundaries of the form $x = a$. First, we need to see where the two curves intersect, by solving $2 - x = x^2$: the solutions are $x = -2, 1$. So we *infer* that we are supposed to find the area from $x = -2$ to $x = 1$, and that the two curves *close up* around this chunk of area without any need of assistance from vertical lines $x = a$. We need to find which curve is higher: plugging in the point 0 between -2 and 1, we see that $y = 2 - x$ is higher. Thus, the desired integral is

$$\text{area...} \quad = \int_{-2}^{1} (2 - x) - x^2 \, dx$$

Definition 5.5.3.

[**author=livshits**, file =**text_files/area_defn_integrals**]
Let us say we move from time $t = a$ to time $t = b$ with velocity $v(t)$ what will be the total distance traveled? If we denote by $D_a(t)$ the total distance traveled at time t, then $D'_a(t) = v(t)$, so $D_a(t)$ is a primitive of $v(t)$. We also know that $D_a(a) = 0$. Now, if V is any other primitive of v, then $D_a(t) = V(t) - V(a)$. The total distance traveled at time $t = b$ will be $D_a(b) = V(b) - V(a)$. This expression is called the *definite integral* and is denoted by

$$\int_a^b v(t)dt = V(b) - V(a),$$

V being any primitive of v. Going back to our usual notation and using the rules of integration for indefinite integrals, we get

$$\int_a^b f = F(b) - F(a), \text{ where } F \text{ is any primitive of } f.$$

Rule 5.5.1.

[**author=livshits**, file =**text_files/area_defn_integrals**]

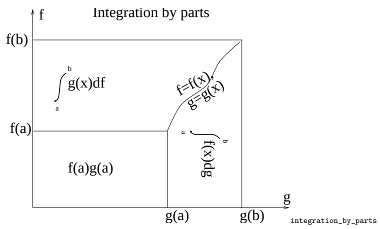
We have the following rules.

- **Sums Rule:**
$$\int_a^b (f + g) = \int_a^b f + \int_a^b g$$
- **Multiplier Rule:**
$$\int_a^b cf = c \int_a^b f, \text{ wrere } c \text{ is a constant}$$

- **Integration by Parts:**

$$\int_a^b f'g = fg|_a^b - \int_a^b fg',$$

where $fg|_a^b$ means $f(b)g(b) - f(a)g(a)$.



- **Change of Variable:**

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(g)dg$$

- There is an additional rule for definite integrals.
  Additivity:

$$\int_a^c f = \int_a^b f + \int_b^c f$$

- In section **??** we will show the following (for nice enough f):
  Positivity:

$$\int_a^b f \geq 0 \quad if \quad f \geq 0 \quad and \quad a \leq b$$

Example 5.5.6.

[**author=livshits**, file =**text_files/area_defn_integrals**]
Here is a bit sleeker way to finish the problem about a leaky bucket by using definite integrals. We can rewrite **??** as

$$\frac{dH}{dt} = -\frac{a}{A}\sqrt{2gH},$$

turning it upside down produces

$$\frac{dt}{dH} = -\frac{A}{a\sqrt{2gH}},$$

multiplying both sides by dH gives

$$dt = -\frac{A}{a}\sqrt{\frac{2}{g}}\frac{dH}{2\sqrt{H}} = -\frac{A}{a}\sqrt{\frac{2}{g}}d\sqrt{H},$$

and finally, integrating both parts yields

$$T = \int_0^T dt = -\frac{A}{a}\sqrt{\frac{2}{g}}\int_{\sqrt{H_0}}^0 d\sqrt{H} = \frac{A}{a}\sqrt{\frac{2}{g}}\int_0^{\sqrt{H_0}} d\sqrt{H} = \frac{A}{a}\sqrt{2H_0/g}$$

Discussion.

[**author=livshits**, file =**text_files/area_defn_integrals**]
As we saw in section **??** (Theorem 3.3.2), the derivative of a ULD function is ULC. It is natural to ask whether any ULC function is a derivative of some ULD function. In this section we will see that it is indeed the case. In other words, any ULC function has a ULD primitive and it makes sense to talk about definite and indefinite integrals of any ULC function. We will also take a closer look at the notion of area and prove the Newton-Leibniz theorem for ULC functions This will provide a rigorous foundation for Calculus in the realm of ULC and ULD functions.

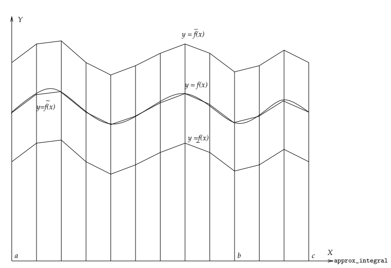
The central idea is to approximate a ULC function f from above by $\overline{f}$ and from below by $\underline{f}$ with some simple (piecewise-linear) functions that are easy to integrate. Then, using positivity of definite integral (that is equivalent to IFT) we can conclude that

$$\int_a^b \overline{f}(x)dx \leq \int_a^b \underline{f}(x)dx$$

(we assume that $a < b$), and if we want to keep positivity, we conclude that

$$\int_a^b \overline{f}(x)dx \leq \int_a^b f(x)dx \leq \int_a^b \underline{f}(x)dx \tag{5.3}$$

The assumption that f is ULC will allow us to take $\overline{f}$ and $\underline{f}$ as close to each other as we want, therefore their integrals can be made as close to each other as we want, and this will define $\int_a^b f(x)dx$ uniquely. After this construction is understood, the Newton-Leibniz theorem becomes an easy check and provides a construction for a ULD primitive of f.



Proof.
[**author=livshits**, file =**text_files/area_defn_integrals**]
So let us assume that f is defined on the segment $[a, b]$ and is ULC, i.e. $|f(x) - f(u)| \leq L|x - u|$. First we introduce a mesh of points $a = x_0 < x_1 < ... < x_{n-1} < x_n = b$ such that $x_k - x_{k-1} \leq h$. Then we put $\underline{f}(x_k) = f(x_k) - 2Lh$ and $\overline{f}(x_k) = f(x_k) + 2Lh$ for $k = 0, ..., n$ and assume $\underline{f}$ and $\overline{f}$ to be linear on each segment $[x_{k-1}, x_k]$. It is easy to check that $\underline{f}(x) \leq f(x) \leq \overline{f}(x)$ for any x in $[a, b]$. Also $\overline{f} - \underline{f} = 4Lh$, therefore

$$\int_a^b \overline{f}(x)dx - \int_a^b \underline{f}(x)dx = 4Lh(b - a) \tag{5.4}$$

Since the $h > 0$ is arbitrary, there is at most one real number I such that $\int_a^b \underline{f}(x)dx \leq I \leq \int_a^b \overline{f}(x)dx$ for any piecewise-linear $\underline{f}$ and $\overline{f}$ such that $\underline{f} \leq f \leq \overline{f}$.

And there will be such a number because $\underline{f} \leq f \leq \overline{f}$ implies $\int_a^b \underline{f}(x)dx \leq \int_a^b \overline{f}(x)dx$, so we can define $\int_a^b f(x)dx = I$. This works when $a < b$, and we can put $\int_a^a f(x)dx = 0$ and $\int_a^b f(x)dx = -\int_b^a f(x)dx$ when $b < a$.

The piecewise-linear function $\tilde{f}$ such that $\tilde{f}(x_k)$ equals $f(x_k)$ and $\tilde{f}$ is linear on every $[x_{k-1}, x_k]$ approximates f better than $\underline{f}$ or $\overline{f}$ because it sits between them together with f, so $\int_a^b \tilde{f}(x)dx$ is often used in practical calculations of $\int_a^b f(x)dx$. It is called *the trapezoid rule* because the approximating integral is the sum of the (appropriately signed) areas of a bunch of trapezoids.

In particular, we can conclude from the estimate 5.4 that

$$\left| \int_a^b \tilde{f}(x)dx - \int_a^b f(x)dx \right| \leq 4Lh|b-a|$$

and the previous exercise shows that the factor 4 in the right-hand side can be dropped.

Now, using this estimate, it is easy to see that the definite integral that we have just constructed for ULC functions possesses the positivity and additivity properties and satisfies the sums and the constant multiple rules from section **??**. It inherits these properties from the approximations, so to speak. □

Discussion.

[`author=livshits`, file =`text_files/area_defn_integrals`]
For example, to prove positivity, we can observe that from $f \geq 0$ it follows that $\tilde{f} \geq 0$ and therefore $\int_a^b \tilde{f}(x)dx \geq 0$ (we assume here that $a \leq b$ and we know that positivity holds for the piecewise-linear functions), so we can conclude that $\int_a^b f(x)dx \geq -4Lh(b-a)$, and therefore $\int_a^b f(x)dx \geq 0$ we can take $h = (b-a)/n$ (Archimedes principle again). Additivity and the sums and the constant multiple rules are demonstrated in a similar fashion (exercise).

There is an important and easy consequence of positivity of our newly constructed definite integral that will be handy soon:

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx$$

(to check it one can "integrate the inequality" $-|f| \leq f \leq |f|$).

Exercises

1. Find the area between the curves $y = x^2$ and $y = 2x + 3$.

2. Find the area of the region bounded vertically by $y = x^2$ and $y = x + 2$ and bounded horizontally by $x = -1$ and $x = 3$.

3. Find the area between the curves $y = x^2$ and $y = 8 + 6x - x^2$.

4. Find the area between the curves $y = x^2 + 5$ and $y = x + 7$.

5. It is easy to check that $\underline{f}(x) \leq f(x) \leq \overline{f}(x)$ for any x in $[a, b]$.

6. For example, to prove positivity, we can observe that from $f \geq 0$ it follows that $\tilde{f} \geq 0$ and therefore $\int_a^b \tilde{f}(x)dx \geq 0$ (we assume here that $a \leq b$ and we know that positivity holds for the piecewise-linear functions), so we can conclude that $\int_a^b f(x)dx \geq -4Lh(b-a)$, and therefore $\int_a^b f(x)dx \geq 0$ we can take $h = (b-a)/n$ (Archimedes principle again). Additivity and the sums and the constant multiple rules are demonstrated in a similar fashion (exercise).

7. It is not too difficult to see that $\underline{f}$ and $\overline{f}$ can be chosen 4 times closer together because already $\overline{f}(x_k) = f(x_k) + Lh/2$ and $\underline{f}(x_k) = f(x_k) - Lh/2$ will guarantee $\underline{f} \leq f \leq \overline{f}$.

5.6 Transcendental integration

Discussion.

[author=duckworth, file =text_files/transcendental_integration]
In this section we adopt a slightly unusual perspective. We suppose that we do not know the derivative of $\ln(x)$ or e^x. We define ln as the integral of $1/x$. Then we obtain the derivative of e^x as a consequence.

Discussion.

[author=livshits,uses=ln,establishes=deriv_of_ln, file =text_files/transcendental_integration]
Here we obtain the derivative of $\ln(x)$ by trying to find the integral of $1/x$.

As you may have noticed, the formula for integrating x^n

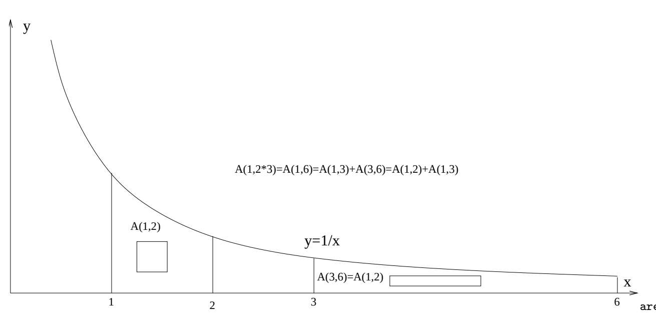
$$\int x^n dx = x^{n+1}/(n+1)$$

breaks down for $n = -1$ because we get zero in the denominator. However, if we apply this formula to calculate a definite integral from a to b where $0 < a < b$, we will get

$$\int_a^b (1/x)dx = x^0/0|_a^b = (b^0 - a^0)/0 = (1-1)/0 = 0/0,$$

and we encounter our good old friend $0/0$, so there is a glimpse of hope here.

Geometrically speaking, the definite integral above makes perfect sense and represents the area under the hyperbola $y = 1/x$ between the vertical lines $x = a$ and $x = b$. Now we have to figure out how to relate this area to something familiar. To do that, we denote by $A(a, b)$ the area under consideration and look at the picture.



area_under_1_over_x

This picture demonstrates that $A(1, 2) + A(1, 3) = A(1, 6)$. Generalizing, we get $A(1, a) + A(1, b) = A(1, ab)$ for $1 < a$ and $1 < b$ so $A(1, x)$ looks like some sort of a logarighm. It is called *the natural logarithm* and is denoted $ln(x)$. So for $1 < a \leq b$

$$\int_a^b (1/x)dx = ln(b) - ln(a) = ln(x)|_a^b$$

and for $x \geq 1$.

$$\int (1/x)dx = \ln(x) + C.$$

Notice that the formulas will hold for positive a, b, or x less than 1 if we take into account that $ln(x) = -ln(1/x)$ for $0 < x < 1$. These formulas can be extended even to the negative x as well by replacing $ln(x)$ with $ln|x|$, $ln(a)$ with $ln|a|$ and

$ln(b)$ with $ln|b|$, but should be treated with some caution since $ln|x|$ and $1/x$ blow up at 0.

Now we got yet another function that we can differentiate:

$$(\ln|x|)' = 1/x$$

Definition 5.6.1.

[`author=livshits,uses=e^x,establishes=deriv_of_e^x`, file =`text_files/transcendental_integration`]
The base of the natural logarithm is called the Euler number and denoted e, so we can write

$$\ln(e^x) = x \text{ and } e^{\ln(a)} = a \text{ for any } a > 0.$$

Sometimes e^x is written as $\exp(x)$, so

$$\ln(\exp(x)) = x \text{ for any } x \text{ and } \exp(\ln(x)) = x \text{ for } x > 0.$$

We can use implicit differentiation to figure out $\frac{d}{dx}\exp(x)$:

$$1 = x' = (\ln(\exp(x)))' = \ln'(\exp(x))\exp'(x) = (1/\exp(x))\exp'(x),$$

so

$$\exp'(x) = \exp(x).$$

5.7 End of chapter problems

Find the following integrals

Exercises

1. $\int x^7 dx$
   $dx^n/dx = nx^{n-1}$
   $x^8/8 + C$
2. $\int 5x^3 dx$
   Use the constant multiplier rule
   $(5/4)x^4 + C$
3. $\int (3x^5 + 7x^{10})dx$
   Use the constant multiplier rule and the sums rule.
   $x^6/2 + 7x^{11}/11 + C$
4. $\int (x^3 + 10)3x^2 dx$
   Use $U = x^3$.
   $u = x^3$, $du = 3x^2 dx$, so the integral becomes $\int (u+10)du = u^2/2 + 10u + C$, that is $x^6/2 + 10x^3 + C$ (after going back to the original variable x).

5. $\int (x^6 + 6x)(x^5 + x)dx$

   Expand and integrate term by term.

6. $\int (2x/(x^2 + 3)^2)dx$

   $(x^2)' = 2x$, use U-subst

   $u = x^2 = 3$, $du = 2xdx$, so the integral becomes $\int u^{-2}du = -1/u + C = -1/(u^2 + 3) + C$

7. $\int x^2\sqrt{x^3 + 2}dx$

   $(x^3)' = 3x^2$, use U-subst

   $u = x^3 + 2$, $du = 3x^2dx$, so the integral becomes $\int (1/3)u^{1/2} = (2/9)u^{3/2} + C = (2/9)(x^3 + 2)^{3/2} + C$

8. Water is pored into a conical bucket at a rate 50 cubic inches per minute. How fast is the water level in the bucket rising at the moment when the area of the water surface is 100 sqare inches?

   Differentiate the formula for the volume of a cone

   The volume of the cone of height h and base area A is $V = Ah/3$ in our problem $A = ah^2$, so $V = ah^3/3$. The time derivative $V' = ah^2h'$ and finally $h' = V'/(ah^2) = V'/A = 50/100 = 1/2$ inches per second.

9. A sperical balloon is pumped up at 5 cubic inches per second. How fast is its area growing when its radius is 10 inches?

   Differentiate the formula for the volume of a ball

   The volume of the balloon is $V = (4/3)\pi r^3$, its surface area is $A = 4\pi r^2$, so $V' = 4\pi r^2 r'$ and $A' = 8\pi r r' = 2V'/r = 10/10 = 1$

10. Conservation of energy via chain rule.

    (a) Check that the gravity force pulling the stone down is equal to $-dP/dy$ where $P(y)$ is the potential energy of the stone.

    (b) Check that Newton's Second law can be rewritten as $my'' + dP/dy = 0$.

    (c) Use the chain rule to calculate the time derivative E' of the energy and use the equation from (b) to show that $E' = 0$, which implies that E does not change with time, i.e. energy is conserved.

    For part (c) note that $(y'^2)' = 2y'y''$

11. $\int (x^5 + 3x^4 - 7)^{10}(5x^4 + 12x^3)dx$

12. $f''(x) = x^5 + x^3 + 7x^2 + 1$, $f(0) = 1$, $f(1) = 3$. Find f

13. Although about the only function that we can integrate now is x^r with $r \neq -1$, we can already solve some not totally trivial problems.

    what is the integral? why $r = -1$ is bad?

14. Check the energy conservation in case $v_0 \neq 0$.

15. Solve this differential equation,

$$H'(t) = -\frac{a}{A}\sqrt{2gH(t)}$$

    read ahead if you can't.

Chapter 6

Applications of Integration

Discussion.

[author=duckworth, file =text_files/introduction_to_applications_of_integrals]
This chapter has a bunch of applications of integration. Unfortunately we are only going to learn two of them: arc-length and surface area of a revolution. These are applications that appeal primarily to mathematicians. I wish we had time to learn the other applications too, which are used in economics, probability, physics, statistics (and therefore every emperical subject),...

6.1 Area between two curves

Rule 6.1.1.

[author=duckworth, file =text_files/area_between_two_curves]
If $f(x) \geq g(x)$ then the area between $f(x)$ and $g(x)$ is $\int_a^b f(x) - g(x)\,dx$. If $g(x)$ is sometimes on top of $f(x)$ then the area is $\int_a^b |f(x) - g(x)|\,dx$. To solve this you need to split the integral up into pieces so that you know on each piece whether f or g is on top.

6.2 Lengths of Curves

Example 6.2.1.

[author=duckworth, file =text_files/arc_length]
Find the distance travelled by a ball which has path given by $y = -x^2 + 4$.

I'll pretend I don't know how to solve this exactly and do an approximaton in 3 steps. Thus $\Delta x = 4/3$ and so I will have points at x equal to -2, $-2/3$, $2/3$, 2. The y-values corresponding to these x-values are 0, $32/9$, $32/9$, 0. Between these points I will use straight lines, thus the distance at each step will be given by the distance formula (i.e. $\sqrt{\Delta x^2 + \Delta y^2}$). So we have:

$$\begin{aligned} \text{Arc-length} &\approx \sqrt{(4/3)^2 + (32/9)^2} + \sqrt{(4/3)^2 + 0^2} + \sqrt{(4/3)^2 + (32/9)^2} \\ &= 8.928 \end{aligned}$$

Now, I want to get an exact answer. That means that I need to figure out how to replace each of those square roots by something of the form $* \cdot \Delta x$. If I can do that then I can integrate $\int_{-2}^{2} * \, dx$. So this is a trick; each of those square roots was of the form $\sqrt{\Delta x^2 + \Delta y^2}$, and if I really want something times Δx (which I do) I'll factor that out to get:

$$\sqrt{1 + \frac{\Delta y^2}{\Delta x^2}} \cdot \Delta x.$$

Thus, what we should integrate is

$$\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

In our example we have $y' = -2x$. Thus, the exact answer should be

$$\text{Arc-length} = \int_{-2}^{2} \sqrt{1 + (-2x)^2}\, dx$$

Note that the function is even, so we can integrate from 0 to 2 and multiply the result by 2. Also, $(-2x)^2$ equals $(2x)^2$, so we can find

$$2 \int_0^2 \sqrt{1 + (2x)^2}\, dx$$

Now, substitute $u = 2x$ to get

$$\frac{1}{2} \cdot 2 \int_0^4 \sqrt{1 + u^2}\, du$$

We look up this integral in the back of our book (because we've already done integrals like this in chapter 7) to get

$$\begin{aligned} &\tfrac{u}{2}\sqrt{1+u^2} + \tfrac{1}{2}\ln(u + \sqrt{1+u^2})\Big|_0^4 \\ &= 2\sqrt{17} + \tfrac{1}{2}\ln(4 + \sqrt{17}) - (0 + \ln(1+0)) \\ &= 9.29 \end{aligned}$$

Definition 6.2.1.

[`author=duckworth`, file =`text_files/arc_length`]
Based on this experience, we define arc-length as follows:

$$\text{Arc-length} = s = \int_a^b ds \text{ where } \begin{aligned} ds &= \sqrt{1 + \left(\tfrac{dy}{dx}\right)^2}\, dx \quad \left(\text{or } \sqrt{\left(\tfrac{dx}{dy}\right)^2 + 1}\, dy\right) \\ &\approx \sqrt{\Delta x^2 + \Delta y^2} \end{aligned}$$

Comment.

[author=duckworth, file =text_files/arc_length]
The problems in this section can take a long time just because there's lots of simplification and/or manipulation to get the integral into the right form. Here's some advice:

- Don't panic if it seems like the problem is getting kind of long.
- Go slowly and double check every step. If you make a mistake there's probably no way the stuff inside the square root will work out right.
- The stuff in the square root is usually rational functions (i.e. polynomials divided by polynomials). To simplify these you usually use one or more of the following tricks: (a) get common denominators, (b) foil everything out, then cancel, then factor, (c) look for perfect squares (i.e. things of the form $a^2 \pm 2ab + b^2$ which equals $(a \pm b)^2$), (d) if you don't have a perfect square then complete the square to get something of the form $\sqrt{\pm u^2 \pm a^2}$ where u equals $x \pm a$. Then try to look this integral up in the back of the book.

Rule 6.2.1.

[author=garrett, file =text_files/arc_length]
The basic point here is *a formula obtained by using the ideas of calculus*: the length of the graph of $y = f(x)$ from $x = a$ to $x = b$ is

$$\text{arc length } = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

Or, if the curve is *parametrized* in the form

$$x = f(t) \qquad y = g(t)$$

with the parameter t going from a to b, then

$$\text{arc length } = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

This formula comes from approximating the curve by straight lines connecting successive points on the curve, using the Pythagorean Theorem to compute the lengths of these segments in terms of the change in x and the change in y. In one way of writing, which also provides a good *heuristic* for remembering the formula, if a small change in x is dx and a small change in y is dy, then the length of the hypotenuse of the right triangle with base dx and altitude dy is (by the Pythagorean theorem)

$$\text{hypotenuse } = \sqrt{dx^2 + dy^2} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

Unfortunately, by the nature of this formula, most of the integrals which come up are *difficult* or *impossible* to 'do'. But if one of these really mattered, we could still estimate it by *numerical integration*.

Exercises

1. Find the length of the curve $y = \sqrt{1 - x^2}$ from $x = 0$ to $x = 1$.
2. Find the length of the curve $y = \frac{1}{4}(e^{2x} + e^{-2x})$ from $x = 0$ to $x = 1$.
3. Set up (but do not evaluate) the integral to find the length of the piece of the parabola $y = x^2$ from $x = 3$ to $x = 4$.

6.3 Numerical integration

Discussion.

[author=duckworth, file =text_files/numerical_integration]
We can approximate $\int_a^b f(x)\,dx$ using a variety of methods: the left-hand rule (LHR), the right-hand rule (RHR), and the midpoint rule (MP). In this section we will discuss which of these is better, and also get some new rules which are better still.

Example 6.3.1.

[author=duckworth, file =text_files/numerical_integration]
Consider $\int_0^1 e^{-x^2}\,dx$. Let's approximate this in four steps. So we have $n = 4$ and $\Delta x = \frac{1}{4}$. We have:

Rule	$x_1^*,\ x_2^*,\ x_3^*,\ x_4^*$	$\frac{1}{4}(e^{-(x_1^*)^2} + e^{-(x_2^*)^2} + e^{-(x_3^*)^2} + e^{-(x_4^*)^2})$	
LHR	$x_1^* = 0,\ x_2^* = \frac{1}{4},\ x_3^* = \frac{2}{4},\ x_4^* = \frac{3}{4}$	$\frac{1}{4}(e^{-0^2} + e^{-(1/4)^2} + e^{-(2/4)^2} + e^{-(3/4)^2})$	$= .821999$
RHR	$x_1^* = \frac{1}{4},\ x_2^* = \frac{2}{4},\ x_3^* = \frac{3}{4},\ x_4^* = \frac{4}{4} = 1$	$\frac{1}{4}(e^{-(1/4)^2} + e^{-(2/4)^2} + e^{-(3/4)^2} + e^{-1^2})$	$= .663969$
MP	$x_1^* = \frac{1}{8},\ x_2^* = \frac{3}{8},\ x_3^* = \frac{5}{8},\ x_4^* = \frac{7}{8}$	$\frac{1}{4}(e^{-(1/8)^2} + e^{-(3/8)^2} + e^{-(5/8)^2} + e^{-(7/8)^2})$	$= .748747$

The obvious questions at this point are: which one of these is best, and how close is it? You might think just by looking at these numbers that .821999 is too high and .663969 is too low. In this case, this is right, but the correct way to see this is to graph $f(x)$ and note that it is decreasing. This implies that the LHR is too high and the RHR is too low.

Rule 6.3.1.

[author=duckworth, file =text_files/numerical_integration]
We can summarize this for all functions:

- If $f(x)$ is increasing then LHR $> \int >$ RHR.
- If $f(x)$ is decreasing then RHR $> \int >$ LHR.

What about the MP rule? What about averaging the LHR and RHR? Let's define a new rule: TRAP $= \frac{1}{2}(\text{LHR} + \text{RHR})$. We give the outcome of this rule, together with how to calculate in terms of the x_i*:

rule	as an average	forumula	
TRAP	$= \frac{1}{2}(\text{LHR} + \text{RHR})$	$\frac{1}{2}\frac{1}{4}(f(0) + 2f(1/4) + 2f(2/4) + 2f(3/4) + f(1))$	$= .742984$

So which is better, the MP or the TRAP? To figure this out draw one "rectangle" in $f(x)$ with a quarter of a circle on top (or see the picture in the book or in lecture notes). The TRAP gives the area formed by the trapezoid connecting the right side to the left where the vertical lines hit the curve for $f(x)$. Draw the MP with a horizontal line coming half-way between the left and right sides (this is **not** the same as a horizontol line half-way between the top and the bottom of the curve). You can re-draw the MP by drawing a tangent line at the point where the MP line intersects $f(x)$. The trapezoid formed with this tangent line, has the same area as the rectangle formed with a horizontal line at the mid-point (you can see this because you just cut off one corner of the rectangle and move it to the other side to form the trapezoid). We can finally see whether MP or TRAP is better and which is too big/too small.

- If $f(x)$ is concave down, then $\mp \ > \ \int \ > \ \text{TRAP}$
- If $f(x)$ is concave up, then $\mp \ < \ \int \ < \ \text{TRAP}$
- In all cases MP is better than TRAP

The preceeding discussion justifies a new rule. We want a formula for something between MP and TRAP, which comes out a little closer to MP. This is Simpson's rule (as applied to the previous example):

rule	as an average	formula		
SIMP	$\frac{2\mp+\text{TRAP}}{3}$	$\frac{1}{3}\frac{1}{4}(f(0)+4f(1/8)+2f(2/8)+4f(3/8)$		
		$+2f(4/8)+4f(5/8)+2f(6/8)+4f(7/8)+f(1))$	$=$	$.7468261205$

Discussion.

[**author=duckworth**, file =**text_files/numerical_integration**]
This section also discusses error bounds for the rules MP, TRAP, and SIMP. To find $K_4 = \max |f^{(4)}(x)|$ you need to find the fifth derivative of f (i.e. find $f^{(5)}(x)$), set this equal to zero, solve for the critical points, then compare the y-values of $|f^{(4)}(x)|$ at the critical points and end-points. Whichever comes out biggest is the maximum. This can be a lot of work (i.e. finding 5 derivatives and setting some big formula equal to 0).

Discussion.

[**author=garrett**, file =**text_files/numerical_integration**]
As we start to see that integration 'by formulas' is a much more difficult thing than differentiation, and sometimes is impossible to do in elementary terms, it becomes reasonable to ask for *numerical approximations to definite integrals.* Since a *definite* integral is just a *number*, this is possible. By constrast, *indefinite* integrals, being *functions* rather than just numbers, are not easily described by 'numerical approximations'.

There are several related approaches, all of which use the idea that a definite integral is related to *area.* Thus, each of these approaches is really essentially a way of approximating area under a curve. Of course, this isn't exactly right, because integrals are not exactly areas, but thinking of area is a reasonable heuristic.

Of course, an approximation is not very valuable unless there is an *estimate for the error*, in other words, an idea of the *tolerance.*

Each of the approaches starts the same way: To approximate $\int_a^b f(x)\,dx$, break the interval $[a,b]$ into smaller subintervals

$$[x_0,x_1],\quad [x_1,x_2],\quad \ldots,[x_{n-2},x_{n-1}],\quad [x_{n-1},x_n]$$

each of the same length

$$\Delta x = \frac{b-a}{n}$$

and where $x_0 = a$ and $x_n = b$.

Trapezoidal rule: This rule says that

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{2}[f(x_0) + 2f(x_1) + 2f(x_2) + \ldots + 2f(x_{n-2}) + 2f(x_{n-1}) + f(x_n)]$$

Yes, all the values have a factor of '2' except the first and the last. (This method approximates the area under the curve by *trapezoids* inscribed under the curve in each subinterval).

Midpoint rule: Let $\overline{x}_i = \frac{1}{2}(x_i - x_{i-1})$ be the midpoint of the subinterval $[x_{i-1}, x_i]$. Then the **midpoint rule** says that

$$\int_a^b f(x)\,dx \approx \Delta x[f(\overline{x}_1) + \ldots + f(\overline{x}_n)]$$

(This method approximates the area under the curve by rectangles whose height is the midpoint of each subinterval).

Simpson's rule: This rule says that

$$\int_a^b f(x)\,dx \approx$$

$$\approx \frac{\Delta x}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \ldots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)]$$

Yes, the first and last coefficients are '1', while the 'inner' coefficients alternate '4' and '2'. And n has to be an *even* integer for this to make sense. (This method approximates the curve by pieces of parabolas).

In general, the smaller the Δx is, the better these approximations are. We can be more precise: the error estimates for the trapezoidal and midpoint rules depend upon the *second derivative*: suppose that $|f''(x)| \leq M$ for some constant M, for all $a \leq x \leq b$. Then

$$\text{error in trapezoidal rule } \leq \frac{M(b-a)^3}{12n^2}$$

$$\text{error in midpoint rule } \leq \frac{M(b-a)^3}{24n^2}$$

The error estimate for Simpson's rule depends on the *fourth* derivative: suppose that $|f^{(4)}(x)| \leq N$ for some constant N, for all $a \leq x \leq b$. Then

$$\text{error in Simpson's rule } \leq \frac{N(b-a)^5}{180n^4}$$

From these formulas estimating the error, it looks like the midpoint rule is always better than the trapezoidal rule. And for high accuracy, using a large number n of subintervals, it looks like Simpson's rule is the best.

6.4 Averages and Weighted Averages

Discussion.

[author=garrett, file =text_files/average_of_function]
The usual notion of *average* of a list of n numbers $x_1, \ldots, x_n$ is

$$\text{average of } x_1, x_2, \ldots, x_n = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

A *continuous* analogue of this can be obtained as an integral, using a notation which matches better:

Definition 6.4.1.

[author=garrett, file =text_files/average_of_function]
let f be a function on an interval $[a, b]$. Then

$$\text{average value of } f \text{ on the interval } [a, b] = \frac{\int_a^b f(x)\,dx}{b - a}$$

Example 6.4.1.

[author=garrett, file =text_files/average_of_function]
For example the *average* value of the function $y = x^2$ over the interval $[2, 3]$ is

$$\text{average value of } f \text{ on the interval } [a, b] = \frac{\int_2^3 x^2\,dx}{3 - 2} = \frac{[x^3/3]_2^3}{3 - 2} = \frac{3^3 - 2^3}{3 \cdot (3 - 2)} = 19/3$$

Discussion.

[author=garrett, file =text_files/average_of_function]
A **weighted average** is an average in which some of the items to be averaged are *'more important'* or *'less important'* than some of the others. The *weights* are (non-negative) numbers which measure the relative importance.

For example, the *weighted average* of a list of numbers $x_1, \ldots, x_n$ with corresponding weights $w_1, \ldots, w_n$ is

$$\frac{w_1 \cdot x_1 + w_2 \cdot x_2 + \ldots + w_n \cdot x_n}{w_1 + w_2 + \ldots + w_n}$$

Note that if the weights are all just 1, then the weighted average is just a plain average.

Definition 6.4.2.

[author=garrett, file =text_files/average_of_function]
The *continuous analogue* of a weighted average can be obtained as an integral,

using a notation which matches better: let f be a function on an interval $[a,b]$, with *weight* $w(x)$, a non-negative function on $[a,b]$. Then

$$\text{weighted average value of } f \text{ on the interval } [a,b] \text{ with weight } w = \frac{\int_a^b w(x)\cdot f(x)\,dx}{\int_a^b w(x)\,dx}$$

Notice that in the special case that the weight is just 1 all the time, then the weighted average is just a plain average.

Example 6.4.2.

[**author=garrett**, file =**text_files/average_of_function**]
For example the *average* value of the function $y=x^2$ over the interval $[2,3]$ with weight $w(x)=x$ is

$$\text{average value of } f \text{ on the interval } [a,b] \text{ with weight } x$$

$$= \frac{\int_2^3 x\cdot x^2\,dx}{\int_2^3 x\,dx} = \frac{[x^4/4]_2^3}{[x^2/2]_2^3} = \frac{\frac{1}{4}(3^4-2^4)}{\frac{1}{2}}(3^2-2^2)$$

Example 6.4.3.

[**author=duckworth**, file =**text_files/average_of_function**]
One of the best examples to think of for average value of a function is the temperature outside over one full day. It's easy to understand what the high temperature means, and what the low temperature means. Suppose you want to know the average temperature, how many times do you need to measure the temperature? 1? Not enough. 4 times? Not enough if you want the most accuracy. 24 times? In practical terms this might be enough, but in math we always want infinite precision. That leads to the following definition. The average of f on an interval $[a,b]$ is:

$$f_{\text{Avg}} = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

A rectangle with base $b-a$ and height equal to the number f_{Avg} has the same area as $\int_a^b f(x)\,dx$. This can be used to define/understand what we mean by the average.

6.5 Centers of Mass (Centroids)

Discussion.

[**author=garrett**, file =**text_files/centers_of_mass**]
For many (but certainly not all!) purposes in physics and mechanics, it is necessary or useful to be able to consider a physical object as being a mass concentrated at a single point, its *geometric center*, also called its **centroid**. *The centroid is essentially the 'average' of all the points in the object.* For simplicity, we will just consider the two-dimensional version of this, looking only at regions in the plane.

The simplest case is that of a rectangle: it is pretty clear that the centroid is the 'center' of the rectangle. That is, if the corners are $(0,0),(u,0),(0,v)$ and (u,v), then the centroid is

$$(\frac{u}{2},\frac{v}{2})$$

The formulas below are obtained by 'integrating up' this simple idea:

Definition 6.5.1.

[author=garrett, file =text_files/centers_of_mass]
For the center of mass (centroid) of the plane region described by $f(x)\le y\le g(x)$ and $a\le x\le b$, we have

$$\begin{aligned} x\text{-coordinate of the centroid} &= \text{average } x\text{-coordinate} \\ &= \frac{\int_a^b x[g(x)-f(x)]\,dx}{\int_a^b [g(x)-f(x)]\,dx} \\ = \frac{\int_{\text{left}}^{\text{right}} x[\text{upper}-\text{lower}]\,dx}{\int_{\text{left}}^{\text{right}} [\text{upper}-\text{lower}]\,dx} &= \frac{\int_{\text{left}}^{\text{right}} x[\text{upper}-\text{lower}]\,dx}{\text{area of the region}} \end{aligned}$$

And also

$$\begin{aligned} y\text{-coordinate of the centroid} &= \text{average } y\text{-coordinate} \\ &= \frac{\int_a^b \frac{1}{2}[g(x)^2-f(x)^2]\,dx}{\int_a^b [g(x)-f(x)]\,dx} \\ = \frac{\int_{\text{left}}^{\text{right}} \frac{1}{2}[\text{upper}^2-\text{lower}^2]\,dx}{\int_{\text{left}}^{\text{right}} [\text{upper}-\text{lower}]\,dx} &= \frac{\int_{\text{left}}^{\text{right}} \frac{1}{2}[\text{upper}^2-\text{lower}^2]\,dx}{\text{area of the region}} \end{aligned}$$

Comment.

[author=garrett, file =text_files/centers_of_mass]
Heuristic: For the x-coordinate: there is an amount $(g(x)-f(x))dx$ of the region at distance x from the y-axis. This is integrated, and then *averaged* dividing by the *total*, that is, dividing by the *area* of the entire region.

For the y-coordinate: in each vertical band of width dx there is amount $dx\,dy$ of the region at distance y from the x-axis. This is integrated up and then averaged by dividing by the total area.

Example 6.5.1.

[author=garrett, file =text_files/centers_of_mass]
For example, let's find the centroid of the region bounded by $x=0$, $x=1$, $y=x^2$, and $y=0$.

$$\begin{aligned} x\text{-coordinate of the centroid} &= \frac{\int_0^1 x[x^2-0]\,dx}{\int_0^1 [x^2-0]\,dx} \\ &= \frac{[x^4/4]_0^1}{[x^3/3]_0^1} = \frac{1/4-0}{1/3-0} = \frac{3}{4} \end{aligned}$$

And

$$y\text{-coordinate of the centroid } = \frac{\int_0^1 \frac{1}{2}[(x^2)^2 - 0]\,dx}{\int_0^1 [x^2 - 0]\,dx}$$

$$= \frac{\frac{1}{2}[x^5/5]_0^1}{[x^3/3]_0^1} = \frac{\frac{1}{2}(1/5 - 0)}{1/3 - 0} = \frac{3}{10}$$

Exercises

1. Find the center of mass (centroid) of the region $0 \le x \le 1$ and $0 \le y \le x^2$.
2. Find the center of mass (centroid) of the region defined by $0 \le x \le 1, 0 \le y \le 1$ and $x + y \le 1$.
3. Find the center of mass (centroid) of a homogeneous plate in the shape of an equilateral triangle.

6.6 Volumes by Cross Sections

Discussion.

[author=duckworth, file =text_files/volumes_cross_section]
The volume of a shape which has cross-sections of constant area of A is $A \cdot l$ where l is the length of the shape. But how can we find the volume of something whose cross-section is changing in area? Well, consider the following similar problem.

The area of between the x-axis and a curve with constant height h is xh where x is the width. But what is the area between the x-axis and a curve $f(x)$ whose height is changing? We break the curve into pieces, on each piece we use a constant height, i.e. we use a rectangle. Then we add all these rectangles of the form $f(x_i)\Delta x$, and then we take the limit to get $\int f(x)\,dx$.

So, suppose that we have some three dimensional shape, and we know the formlua $A(x)$ for the area of the cross section at x. We can break the shape up into pieces, on each piece use a constant cross-section area, and calculate the volume of that piece. Then we add all these pieces of the form $A(x_i)\Delta x$ together. When we take the limit we obtain the following rule.

Rule 6.6.1.

[author=duckworth, file =text_files/volumes_cross_section]
Let V be the volume of a shape between $x = a$ and $x = b$, which has cross-sectional area given by the function $A(x)$. Then V is given by

$$V = \int_a^b A(x)\,dx.$$

Comment.

[author=duckworth, file =text_files/volumes_cross_section]
When computing the volume in the previous rule, we most often have cross-sections that are squares, square, triangles, or circles, or simple combinations of these shapes. In each case you should know the how to find the area $A(x)$.

But in principle we could use the previous rule for any function $A(x)$ that we know how to integrate. For example, we could have a shape where the cross-sections are given by parabolas. In fact, we could even have a shape where we don't know how to integrate $A(x)$, but then we could approximate the volume using either our calculators or Riemann sums.

Discussion.

[author=garrett, file =text_files/volumes_cross_section]
Next to computing areas of regions in the plane, the easiest *concept* of application of the ideas of calculus is to computing volumes of solids where somehow we know a formula for the *areas of slices*, that is, *areas of cross sections*. Of course, in any particular example, the actual issue of getting the formula for the cross section, and figuring out the appropriate limits of integration, can be difficult.

Rule 6.6.2.

[`author=garrett`, file =`text_files/volumes_cross_section`]
The idea is to just 'add up slices of volume:

$$\text{volume } = \int_{\text{left limit}}^{\text{right limit}} (\text{area of cross section at } x) \; dx$$

where in whatever manner we describe the solid it extends from $x =$*left limit* to $x =$*right limit*. We must suppose that we have some reasonable *formula* for the area of the cross section.

Example 6.6.1.

[`author=garrett`, file =`text_files/volumes_cross_section`]
Find the volume of a solid ball of radius 1.

(In effect, we'll be deriving the formula for this). We can suppose that the ball is centered at the origin. Since the radius is 1, the range of x coordinates is from -1 to $+1$, so x will be integrated from -1 to $+1$. At a particular value of x, what does the cross section look like? A disk, whose radius we'll have to determine. To determine this radius, look at how the solid ball intersects the x, y-plane: it intesects in the disk $x^2 + y^2 \le 1$. For a particular value of x, the values of y are between $\pm\sqrt{1-x^2}$. This line segment, having x fixed and y in this range, is the intersection of the cross section disk with the x, y-plane, and in fact is a *diameter* of that cross section disk. Therefore, the radius of the cross section disk at x is $\sqrt{1-x^2}$. Use the formula that the area of a disk of radius r is πr^2: the area of the cross section is

$$\text{cross section at } x \; = \pi(\sqrt{1-x^2})^2 = \pi(1-x^2)$$

Then integrate this from -1 to $+1$ to get the volume:

$$\text{volume } = \int_{\text{left}}^{\text{right}} \text{area of cross-section}\, dx$$

$$= \int_{-1}^{+1} \pi(1-x^2)\, dx = \pi[x - \frac{x^3}{3}]_{-1}^{+1} = \pi[(1-\frac{1}{3}) - (-1 - \frac{(-1)^3}{3})] = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

Exercises

1. Find the volume of a circular cone of radius 10 and height 12 (not by a formula, but by cross sections).

2. Find the volume of a cone whose base is a *square* of side 5 and whose height is 6, by cross-sections.

3. A hole 3 units in radius is drilled out along a diameter of a solid sphere of radius 5 units. What is the volume of the remaining solid?

4. A solid whose base is a disc of radius 3 has vertical cross sections which are *squares*. What is the volume?

6.7 Solids of Revolution

Discussion.

[author=garrett, file =text_files/solids_revolution]
Another way of computing volumes of some special types of solid figures applies to solids obtained by *rotating plane regions* about some axis.

Rule 6.7.1.

[author=garrett, file =text_files/solids_revolution]
If we rotate the plane region described by $f(x) \leq y \leq g(x)$ and $a \leq x \leq b$ around the x-axis, then the volume of the resulting solid is

$$V = \int_a^b \pi(g(x)^2 - f(x)^2)\,dx$$

$$= \int_{\text{left limit}}^{\text{right limit}} \pi(\text{upper curve}^2 - \text{lower curve}^2)\,dx$$

It is necessary to suppose that $f(x) \geq 0$ for this to be right.

Comment.

[author=garrett, file =text_files/solids_revolution]
This formula comes from viewing the whole thing as sliced up into slices of thickness dx, so that each slice is a *disk* of radius $g(x)$ with a smaller disk of radius $f(x)$ removed from it. Then we use the formula

$$\text{area of disk } = \pi \text{ radius}^2$$

and 'add them all up'. The hypothesis that $f(x) \geq 0$ is necessary to avoid different pieces of the solid 'overlap' each other by accident, thus counting the same chunk of volume *twice*.

If we rotate the plane region described by $f(x) \leq y \leq g(x)$ and $a \leq x \leq b$ **around the y-axis** (instead of the x-axis), the volume of the resulting solid is

$$\text{volume } = \int_a^b 2\pi x(g(x) - f(x))\,dx$$

$$= \int_{\text{left}}^{\text{right}} 2\pi x(\text{ upper - lower})\,dx$$

This second formula comes from viewing the whole thing as sliced up into thin cylindrical shells of thickness dx encircling the y-axis, of radius x and of height $g(x) - f(x)$. The volume of each one is

$$(\text{area of cylinder of height } g(x) - f(x) \text{ and radius } x) \cdot dx = 2\pi x(g(x) - f(x))\,dx$$

and 'add them all up' in the integral.

Example 6.7.1.

[**author=garrett**, file =**text_files/solids_revolution**]
As an example, let's consider the region $0 \le x \le 1$ and $x^2 \le y \le x$. Note that for $0 \le x \le 1$ it really is the case that $x^2 \le y \le x$, so $y = x$ is the *upper* curve of the two, and $y = x^2$ is the *lower* curve of the two. Invoking the formula above, the volume of the solid obtained by rotating this plane region around the x-*axis* is

$$\text{volume} \ = \int_{\text{left}}^{\text{right}} \pi(\text{upper}^2 - \text{lower}^2)\, dx$$

$$= \int_0^1 \pi((x)^2 - (x^2)^2)\, dx = \pi[x^3/3 - x^5/5]_0^1 = \pi(1/3 - 1/5)$$

Example 6.7.2.

[**author=garrett**, file =**text_files/solids_revolution**]
Let's take the same function as in Example 6.7.1, and rotate it around the y-axis instead of the x-axis. Then we have

$$\text{volume} \ = \int_{\text{left}}^{\text{right}} 2\pi x(\text{upper} - \text{lower})\, dx$$

$$= \int_0^1 2\pi x(x - x^2)\, dx = \pi \int_0^1 \frac{2x^3}{3} - \frac{2x^4}{4}\, dx = [\frac{2x^3}{3} - \frac{2x^4}{4}]_0^1 = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

Discussion.

[**author=duckworth**, file =**text_files/solids_revolution**]
For some functions it's easier to slice the volume a different way. If you rotate a little bump around the y-axis, then cross-section slices aren't very good. In this case, think about a little vertical rectangle in the bump, being rotated around the y-axis and making a cylindrical shell. If we add a bunch of these shells together we'll have the whole volume.

Derivation.

[**author=duckworth**, file =**text_files/solids_revolution**]
Consider one one cylindrical shell, with height h, radius r and thickness Δr. To estimate it's volume, unwrap/unroll the shell. You'll get a rectangular piece with height h, thickness Δr and length of $2\pi r$ (from the circumference of the original cylindrical shell). Thus, the volume of this shell is $h \cdot 2\pi r \cdot \Delta r$. We translate this, and put all these pieces together as follows.

To find the volume of $f(x)$ rotated about the y-axis:

cylindrical shell	h	$2\pi r$	Δr
	$\updownarrow$	$\updownarrow$	$\updownarrow$
in terms of $f(x)$	$f(x)$	$2\pi x$	Δx
	$\updownarrow$	$\updownarrow$	$\updownarrow$
add all the shells	$\int_a^b f(x)$	$2\pi x$	dx

where x is the radius because you're rotating about the y-axis, and you need to figure out a and b (which equal the smallest and the largest radiuses) from the picture.

Sometimes, your region isn't defined by a single function $f(x)$. In this case, you draw a single shell figure out what h is.

For example, if the region is defined as being between two functions f and g you'd have

$$\begin{array}{ccc} h & 2\pi r & \Delta r \\ f(x) - g(x) & 2\pi x & \Delta x \\ \int_a^b (f(x) - g(x)) & 2\pi x \, dx & \end{array}$$

Exercises

1. Find the volume of the solid obtained by rotating the region $0 \leq x \leq 1, 0 \leq y \leq x$ around the y-axis.
2. Find the volume of the solid obtained by rotating the region $0 \leq x \leq 1, 0 \leq y \leq x$ around the x-axis.
3. Set up the integral which expresses the volume of the doughnut obtained by rotating the region $(x-2)^2 + y^2 \leq 1$ around the y-axis.

6.8 Work

Discussion.

[author=duckworth, file =text_files/work_application]
The amount of work required to move an object is:

$$W = F \cdot d$$

where F is a (positive) force acting in the opposite direction of the movement and d is the distance the object is moved. Here we assume that F is constant. Also, we change this definition slightly if the force is acting in the same direction as the movement: then we use $-F$ instead of F.

Usually we deal with problems where the force is changing or the distance is changing. In this case we figure out:

(a) the formula for doing part of the work (where "part" refers to either moving part of the object of thickness Δx or to figuring out the force over a certain distance of length Δx)

(b) and then we integrate the formula we found in part (a)

Rule 6.8.1.

[author=duckworth, file =text_files/work_application]
If an object is moving against a force of strength $F(x)$ then the work required to move the object from $x = a$ to $x = b$ is

$$W = \int_a^b F(x)\, dx$$

Example 6.8.1.

[author=duckworth, file =text_files/work_application]
if a force of $F(x)$ of strength $\sin(x)$ acts on a object at position x (for x in $[0, \pi/2]$) and the direction of $F(x)$ is towards $x = 0$, find the work required to move it from $x = 0$ to $x = 1$.

(a) Here the force is changing. Let Δx be a little distance that the object will move, at position x (for example, $\Delta x = .1$ and $x = 0$ would represent the work to move from $x = 0$ to $x = .1$). On this segment, the work will be $\sin(x)$ (from $x = 0$ to $x = .1$ we would take x in $[0, .1]$, maybe $\sin(0)$, $\sin(.1)$ or $\sin(.05)$). So the work to move a distance of Δx around position x would be

$$\text{part of work} \quad \sin(x)\Delta x$$

(b) The total work is $\int_0^1 \sin(x)\, dx$.

Rule 6.8.2.

[author=duckorth, file =text_files/work_application]
Suppose we have a substance (usually water, gravel, dirt, lengths of rope or chain) which is being moved. Suppose that the substance covers positions from $x = a$ to $x = b$. Let Δx be given, and let the phrase "the substance at position x" mean the total volume of the substance which is contained in any interval of length Δx which contains x (for example we could pick the interval $[x - \frac{1}{2}\Delta x, x + \frac{1}{2}\Delta x]$). We first approximate the amount of work required to move the substance at position x, by using a constant values for the distance the substance is moved and, if necessary, using a constant value for the force. Let $I(x)\Delta x$ be a formula for this constant approxmition of the work required to move all the substance at position x. Then the total work is given by

$$W = \int_a^b I(x)\, dx.$$

Example 6.8.2.

[author=duckworth, file =text_files/work_application]
Suppose we are pumping water out of a tank which is a cylinder of radius 2 m and height 9 m. Find the work required to empty the tank.

(a) Here, the distance being lifted is changing. Let's measure x from the top and consider a slice of the cylinder of thickness Δx at depth x. The work to lift this slice of water is

$$\begin{aligned}\text{Work of one slice} &= \text{Force} \times \text{distance}\\ &= \text{weight of slice} \times x\\ &= \text{volume of slice} \times \text{density of water} \times \text{gravity} \times x\\ &= \text{area of slice} \times \Delta x \times 1000 \times 9.8 \times x\\ &= \pi \cdot 2^2 \times \Delta x \times 1000 \times 9.8 \times x\end{aligned}$$

(b) The total work is $\int_0^9 \pi \cdot 4 \cdot 1000 \cdot 9.8 \cdot x\, dx$

6.9 Surfaces of Revolution

Discussion.

[author=duckworth, file =text_files/surface_revolution]
This section is similar in spirit and in some details to the sections on arc-length and on volumes generated by rotation. Whence terseness.

Definition 6.9.1.

[author=duckworth, file =text_files/surface_revolution]

The surface area generated by rotation a function around one of the axes is

$$SA = \int_a^b 2\pi r\, ds \text{ where } \quad ds \quad = \sqrt{1+\left(\tfrac{dy}{dx}\right)^2}\, dx \quad \left(\text{or } \sqrt{\left(\tfrac{dx}{dy}\right)^2+1}\, dy\right)$$
$$\approx \sqrt{\Delta x^2 + \Delta y^2}$$

Here r is the radius of revolution. If you're rotating around the x-axis, and your formula is given as a function of x, then you will use $r =$ function of x and $ds = \sqrt{1+\left(\frac{dy}{dx}\right)^2}\, dx$. If you're rotating around the x-axis and your formula is given in terms of y then you'll use $r = y$ and $ds = \sqrt{\left(\frac{dx}{dy}\right)^2+1}\, dy$

Comment.

[**author=duckworth**, file =**text_files/surface_revolution**]
One way to understand this formula is to think of ds as being approximately the length of a diagonal line ℓ between two points on the curve. Then $2\pi r$ times this length is the area of a rectangle with length $2\pi r$ and height ℓ. This rectangle has approximately the same area as one gets by rotating the line ℓ around a radius of r.

Definition 6.9.2.

[**author=garrett**, file =**text_files/surface_revolution**]
Here is another *formula obtained by using the ideas of calculus*: the area of the surface obtained by rotating the curve $y = f(x)$ with $a \le x \le b$ around the x-axis is

$$\text{area} \ = \int_a^b 2\pi f(x)\sqrt{1+\left(\frac{dy}{dx}\right)^2}\, dx$$

This formula comes from extending the ideas of the previous section the length of a little piece of the curve is

$$\sqrt{dx^2 + dy^2}$$

This gets rotated around the perimeter of a circle of radius $y = f(x)$, so approximately give a band of width $\sqrt{dx^2+dy^2}$ and length $2\pi f(x)$, which has area

$$2\pi f(x)\sqrt{dx^2+dy^2} = 2\pi f(x)\sqrt{1+\left(\frac{dy}{dx}\right)^2}\, dx$$

Integrating this (as if it were a sum!) gives the formula.

As with the formula for arc length, it is very easy to obtain integrals which are difficult or impossible to evaluate except numerically.

Similarly, we might rotate the curve $y = f(x)$ around the y-axis instead. The same general ideas apply to compute the area of the resulting surface. The width of each little band is still $\sqrt{dx^2+dy^2}$, but now the length is $2\pi x$ instead. So the band has area

$$\text{width} \ \times \ \text{length} \ = 2\pi x\sqrt{dx^2+dy^2}$$

Therefore, in this case the surface area is obtained by integrating this, yielding the formula

$$\text{area} \ = \int_a^b 2\pi x\sqrt{1+\left(\frac{dy}{dx}\right)^2}\, dx$$

Exercises

1. Find the area of the surface obtained by rotating the curve $y = \frac{1}{4}(e^{2x}+e^{-2x})$ with $0 \leq x \leq 1$ around the x-axis.

2. Just set up the integral for the surface obtained by rotating the curve $y = \frac{1}{4}(e^{2x} + e^{-2x})$ with $0 \leq x \leq 1$ around the y-axis.

3. Set up the integral for the area of the surface obtained by rotating the curve $y = x^2$ with $0 \leq x \leq 1$ around the x-axis.

4. Set up the integral for the area of the surface obtained by rotating the curve $y = x^2$ with $0 \leq x \leq 1$ around the y-axis.

Chapter 7

Techniques of Integration

7.1 Integration by parts

Derivation.

[`author=duckworth`, file =`text_files/integration_by_parts`]
The product rule says $(f \cdot g)' = f' \cdot g + f \cdot g'$. Taking anti-derivatives of both sides gives $f \cdot g = \int f' \cdot g + \int f \cdot g'$. Solving this for $\int f' \cdot g$ gives:

$$\text{Integration by parts} \qquad \int f' \cdot g = f \cdot g - \int f \cdot g'$$

The book writes this a different way. Let $u = f(x)$ and $v = g(x)$ so $du = f'(x)\,dx$ and $dv = g'(x)\,dx$. Then we have:

$$\text{Integration by parts} \qquad \int v\,du = u \cdot v - \int u\,dv$$

Usually you are given something to integrate that looks like a product. You have to choose which thing to call f' (or du) and which to call g (or v). The point is that $\int f \cdot g'$ should be easier for some reason than $\int f' \cdot g$.

Example 7.1.1.

[`author=duckworth`, file =`text_files/integration_by_parts`]
To find $\int xe^{3x}\,dx$ let $f' = e^{3x}$ and $g = x$. Then $f = \frac{1}{3}e^{3x}$ and $g' = 1$. So we have:

$$\int xe^{3x}\,dx \quad = \quad \frac{1}{3}xe^{3x} - \int \frac{1}{3}e^{3x}\,dx \quad = \quad \frac{x}{3}e^{3x} - \frac{1}{9}e^{3x}$$

Derivation.

[`author=garrett`, file =`text_files/integration_by_parts`]
Strangely, the subtlest standard method is just the *product rule* run backwards. This is called **integration by parts**. (This might seem strange because often people find the chain rule for differentiation harder to get a grip on than the

product rule). One way of writing the integration by parts rule is

$$\int f(x) \cdot g'(x)\ dx = f(x)g(x) - \int f'(x) \cdot g(x)\ dx$$

Sometimes this is written another way: if we use the notation that for a function u of x,

$$du = \frac{du}{dx}\ dx$$

then for two functions u, v of x the rule is

$$\int u\ dv = uv - \int v\ du$$

Yes, it is hard to see how this might be helpful, but it is. The first theme we'll see in examples is where we could do the integral except that there is a power of x 'in the way':

Example 7.1.2.

[author=garrett, file =text_files/integration_by_parts]
The simplest example is

$$\int x\ e^x\ dx = \int x\ d(e^x) = x\ e^x - \int e^x\ dx = x\ e^x - e^x + C$$

Here we have taken $u = x$ and $v = e^x$. It is important to be able to see the e^x as being the derivative of itself.

Example 7.1.3.

[author=garrett, file =text_files/integration_by_parts]
A similar example is

$$\int x\ \cos x\ dx = \int x\ d(\sin x) = x\ \sin x - \int \sin x\ dx = x\ \sin x + \cos x + C$$

Here we have taken $u = x$ and $v = \sin x$. It is important to be able to see the $\cos x$ as being the derivative of $\sin x$.

Example 7.1.4.

[author=garrett, file =text_files/integration_by_parts]
Yet another example, illustrating also the idea of *repeating* the integration by parts:

$$\int x^2\ e^x\ dx = \int x^2\ d(e^x) = x^2\ e^x - \int e^x\ d(x^2)$$

$$= x^2\ e^x - 2\int x\ e^x\ dx = x^2\ e^x - 2x\ e^x + 2\int e^x\ dx$$

$$= x^2\ e^x - 2x\ e^x + 2e^x + C$$

Here we integrate by parts twice. After the first integration by parts, the integral we come up with is $\int xe^x\,dx$, which we had dealt with in the first example.

Example 7.1.5.

[author=garrett, file =text_files/integration_by_parts]
Sometimes it is easier to integrate the *derivative* of something than to integrate the thing:

$$\int \ln x\,dx = \int \ln x\,d(x) = x\ln x - \int x\,d(\ln x)$$

$$= x\ln x - \int x\,\frac{1}{x}\,dx = x\ln x - \int 1\,dx = x\ln x - x + C$$

We took $u = \ln x$ and $v = x$.

Example 7.1.6.

[author=garrett, file =text_files/integration_by_parts]
Again in this example it is easier to integrate the derivative than the thing itself:

$$\int \arctan x\,dx = \int \arctan x\,d(x) = x\arctan x - \int x\,d(\arctan x)$$

$$= x\arctan x - \int \frac{x}{1+x^2}\,dx = x\arctan x - \frac{1}{2}\int \frac{2x}{1+x^2}\,dx$$

$$= x\arctan x - \frac{1}{2}\ln(1+x^2) + C$$

since we should recognize the

$$\frac{2x}{1+x^2}$$

as being the derivative (via the chain rule) of $\ln(1+x^2)$.

Rule 7.1.1.

[author=livshits, file =text_files/integration_by_parts]
Integration by Parts

$$\int f'g = fg - \int fg'$$

sometimes, by the use of Leibniz notation: $df = f'dx$, this rule is written as

$$\int gdf = fg - \int fdg$$

Example 7.1.7.

[author=livshits, file =text_files/integration_by_parts]

Here is a "proof" that 0 = 1 from a nice book "Mathemetical Mosaic" by Ravi Vakil. It uses integration by parts.

$$\begin{aligned}\int \frac{1}{x}\,dx &= \int x'\frac{1}{x}\,dx \\ &= x\frac{1}{x} - \int x\left(\frac{1}{x}\right)'\,dx \\ &= 1 - \int x\left(-\frac{1}{x^2}\right)\,dx \\ &= 1 + \int \frac{1}{x}\,dx.\end{aligned}$$

Therefore 0 = 1. Can you find a mistake? We will learn later how to integrate $1/x$, so the integral is a totally legitimate one, the catch is somewhere else.

Exercises

1. $\int \ln x\,dx = ?$
2. $\int xe^x\,dx = ?$
3. $\int (\ln x)^2\,dx = ?$
4. $\int xe^{2x}\,dx = ?$
5. $\int \arctan 3x\,dx = ?$
6. $\int x^3 \ln x\,dx = ?$
7. $\int \ln 3x\,dx = ?$
8. $\int x \ln x\,dx = ?$

7.2 Partial Fractions

Strategy.

[author=duckworth, file =text_files/partial_fractions]
The strategy in this section can be outlined as follows. Using basic techniques, we know how to do the following very simple rational functions:

$$\int \frac{1}{x \pm a}\,dx = \ln|x \pm a|, \qquad \int \frac{x}{x^2 \pm a}\,dx = \frac{1}{2}\ln|x^2 \pm a| \qquad \text{(both U-subst)}$$

and

$$\int \frac{1}{x^2 + a^2}\,dx = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right), \qquad \int \frac{1}{x^2 - a^2}\,dx = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right|$$

All the rest of our work is to break down more complicated problems into pieces that are polynomials, or which use the formulas just given.

Procedure.

[author=duckworth, file =text_files/partial_fractions]
We start with $\int \frac{\text{top poly}}{\text{bottom poly}}$.

- If degree top poly $\geq$ degree bottom poly, then perform polynomial division so that this is no longer the case.
- Factor the bottom poly, so that we have only linear and quadratic factors. Then do partial fractions so that we have separate fractions, each of the form $\frac{*}{x \pm a}$ or $\frac{*}{x^2+ax+b}$ (in each case $*$ should be something with a lower degree than the bottom).
- Perform completing the square on any fractions with quadratic factors so we have:
$$\frac{*}{x^2+ax+b} \to \frac{*}{u^2 \pm c^2}.$$
- This reduces the original integral as follows:
$$\int \frac{\text{poly}}{\text{bottom poly}} = \int \left(\text{poly} + \frac{*}{x \pm a} + \frac{*}{x \pm b} + \cdots + \frac{*}{u^2 \pm c^2} + \frac{*}{u^2 \pm d^2} + \ldots\right)$$
- We should be able to finish the integral using our knowledge of how to do $\int$ poly, $\int \frac{*}{x \pm a}$, $\int \frac{*}{x^2 \pm a^2}$ (again, $*$ always represents something with lower degree than the bottom).

Discussion.

[author=duckworth, file =text_files/partial_fractions]
Polynomial division. This works just like ordinary long division: you put one guy on the side, the other guy under the division sign; at each step you put a multiplier on top, multiply it by the guy on the side, subtract the result from the stuff underneath so that you kill off the leading term.

Example 7.2.1.

[`author=duckworth`, file =`text_files/partial_fractions`]

Find $\frac{123}{9}$ We rewrite this as $9\,\overline{)123}$. We will put first a 1 on top because 9 goes into 12 once:

$$9\,\overline{)123} \quad\longrightarrow\quad \begin{array}{r} 1 \\ 9\,\overline{)123} \\ -9 \\ \hline 3 \end{array} \quad\longrightarrow\quad \begin{array}{r} 13 \\ 9\,\overline{)123} \\ -9 \\ \hline 33 \\ -27 \\ \hline 6 \end{array}$$

So we have a remainder of 6. We write this as

$$\frac{123}{9} = 13 + \frac{6}{9}.$$

Example 7.2.2.

[`author=duckworth`, file =`text_files/partial_fractions`]
Find $\frac{x^4-2x^2+17x+2}{x^2+x}$. We will first put a x^2 on top, because multiplying this by x^2+x on the side will allow us to kill the x^4 underneath (note, we need to keep track of the x^3 column, so we write in $0x^3$):

$$\begin{array}{r} x^2 \\ x^2+x\,\overline{)\,x^4+0x^3-2x^2+17x+2} \\ \underline{-(x^4+\ x^3)} \\ -\ x^3-2x^2 \end{array}$$

Next, we put $-x$ on top because when we multiply this by x^2+x we can kill off the $-x^3$:

$$\begin{array}{r} x^2-\ x \\ x^2+x\,\overline{)\,x^4+0x^3-2x^2+17x+2} \\ \underline{-(x^4+\ x^3)} \\ -\ x^3-2x^2 \\ -\underline{(x^3-\ x^2)} \\ -\ x^2+17x \end{array} \quad\longrightarrow\quad \begin{array}{r} x^2-\ x-\ 1 \\ x^2+x\,\overline{)\,x^4+0x^3-2x^2+17x+2} \\ \underline{-(x^4+\ x^3)} \\ -\ x^3-2x^2 \\ -\underline{(x^3-\ x^2)} \\ -\ x^2+17x \\ -\underline{(-\ x^2-\ x)} \\ 18x+2 \end{array}$$

So the remainder is $18x+2$ and we write this all as

$$\frac{x^4-2x^2+17x+2}{x^2+x} \quad=\quad x^2-x-1+\frac{18x+2}{x^2+x}.$$

Discussion.

[`author=duckworth`, file =`text_files/partial_fractions`]
Partial Fractions. This is a way to rewrite a single fraction with factors on the

bottom as multiple fractions without fractions on the bottom.

Example 7.2.3.

[author=duckworth, file =text_files/partial_fractions]
Suppose we add $\frac{1}{x}+\frac{3}{x^2+1}$. The common denominator is $x(x^2+1)$ and we get

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

Now suppose we started with $\frac{x^2+3x+1}{x(x^2+1)}$ and didn't know that it was originally written as two fractions. We could figure out those fractions as follows. Solve for A, B and C:

$$\frac{x^2+3x+1}{x(x^2+1)} = \frac{A}{x}+\frac{Bx+C}{x^2+1}.$$

Multiplying both sides by $x(x^2+1)$ we get:

$$\begin{aligned} x^2+3x+1 &= A(x^2+1)+(Bx+C)x \\ &= Ax^2+A+Bx^2+Cx \\ &= (A+B)x^2+Cx+A \end{aligned}$$

Now, for these sides of the equation to be equal, we need the coefficients of x^2 to be the same on both sides, we need the coefficients of x to be the same on both sides, and we need the constant terms on both sides to be the same. This leads to the following equations:

$$\begin{aligned} x^2 \text{ coeff}: \quad 1 &= A+B \\ x \text{ coeff}: \quad 3 &= C \\ \text{constant}: \quad 1 &= A \end{aligned}$$

This gives us $A=1$, $B=0$ and $C=3$. Thus we have found:

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

Of course, in this example we already knew this, but the point is we figured out how to take the fraction on the left, and write it as the sum of fractions on the right.

Procedure.

[author=duckworth, file =text_files/partial_fractions]
Here's the general scheme for how to do this. You factor the bottom and look at the factors you have:

- Distinct linear factors: each gets represented once on the right hand side:

$$\frac{*}{(x+a)(x+b)\ldots} = \frac{A}{x+a}+\frac{B}{x+b}+\ldots. \qquad (a\neq b)$$

- Repeated linear factors: the ones that are repeated get represented multiple times on the right hand side:

$$\frac{*}{(x+a)^4(x+b)\ldots} = \frac{A}{x+a}+\frac{B}{(x+a)^2}+\frac{C}{(x+a)^3}+\frac{D}{(x+a)^4}+\frac{E}{x+b}+\ldots \qquad (a\neq b)$$

Hopefully the pattern is clear about what to do if you replaced $(x+a)^4$ with $(x+a)^9$.

- Distinct quadratic factors: each gets represented once on the right hand side:

$$\frac{*}{(x^2+ax+b)(x^2+cx+d)} = \frac{Ax+B}{x^2+ax+b} + \frac{Cx+D}{x^2+cx+d} \cdots$$

- Repeated quadratic factors: the ones that are repeated get represented multiple times on the right hand side:

$$\frac{*}{(x^2+ax+b)^3(x^2+cx+d)} = \frac{Ax+B}{x^2+ax+b}+\frac{Cx+B}{(x^2+ax+b)^2}+\frac{Dx+E}{(x^2+ax+b)^3}+\frac{Fx+G}{x^2+cx+d}$$

Hopefully the patter is clear about what to do if you replaced $(x+ax+b)^3$ with $(x+ax+b)^{11}$.

After you get the above equation set up, you multiply both sides by the denominator from the left, you multiply everything out on the right, you gather the x-terms, you gather the x^2-terms, the x^3-terms etc. Then you get a new system of equations by requiring that the coefficients of x be the same on both sides, the coefficients of x^2 to be the same on both sides, etc.

Discussion.

[author=duckworth, file =text_files/partial_fractions]
Completing the square. This is designed to turn something of the form x^2+ax+b into $(x+c)^2+d$. There are two ways to do this: (1) use a "recipe", (2) solve equations.

Example 7.2.4.

[author=duckworth, file =text_files/partial_fractions]
Complete the square for x^2+6x+7: Take half of the x-coefficient, square this, add and subtract this into the formula, group the first three terms and note that they look like $(x+c)^2$ and simplify the last two terms into d:

$$\begin{array}{ccc} x^2+\ \ 6x\ \ \ +7 & & x^2+6x+\ \ \ 9\ -\ 9+7 \\ \div 2 \downarrow & \longrightarrow & \div 2\downarrow \quad \nwarrow\!\nearrow \\ 3 \xrightarrow{\hat{}2} 9 & & 3 \xrightarrow{\hat{}2} 9 \end{array}$$

Note that $x^2+6x+7=(x+3)^2$, and simplify $-9+7$ to -2 to get

$$x^2+6x+7 \quad = \quad (x+3)^2-2$$

The other way to do this problem is to set:

$$x^2+6x+7=(x+a)^2+b$$

and solve for a and b. You get $x^2+6x+7=x^2+2ax+a^2+b$ so you see that $a=3$ (because we need $6x=2ax$) thus $7=a^2+b$ implies that $b=-2$.

Discussion.

[author=garrett, file =text_files/partial_fractions]
Now we return to a more special but still important technique of doing indefinite integrals. This depends on a good trick from algebra to transform complicated *rational functions* into simpler ones. Rather than try to formally describe the general fact, we'll do the two simplest families of examples.

Example 7.2.5.

[author=garrett, file =text_files/partial_fractions]
Consider the integral

$$\int \frac{1}{x(x-1)}\,dx$$

As it stands, we do not recognize this as the derivative of anything. However, we have

$$\frac{1}{x-1} - \frac{1}{x} = \frac{x-(x-1)}{x(x-1)} = \frac{1}{x(x-1)}$$

Therefore,

$$\int \frac{1}{x(x-1)}\,dx = \int \frac{1}{x-1} - \frac{1}{x}\,dx = \ln(x-1) - \ln x + C$$

That is, by separating the fraction $1/x(x-1)$ into the 'partial' fractions $1/x$ and $1/(x-1)$ we were able to do the integrals immediately by using the logarithm. How to see such identities?

Rule 7.2.1.

[author=garrett, file =text_files/partial_fractions]
Well, let's look at a situation

$$cx + d(x-a)(x-b) = \frac{A}{x-a} + \frac{B}{x-b}$$

where a, b are given numbers (not equal) and we are to *find* A, B which make this true. If we can find the A, B then we can integrate $(cx+d)/(x-a)(x-b)$ simply by using logarithms:

$$\int \frac{cx+d}{(x-a)(x-b)}\,dx = \int \frac{A}{x-a} + \frac{B}{x-b}\,dx = A\ln(x-a) + B\ln(x-b) + C$$

To find the A, B, multiply through by $(x-a)(x-b)$ to get

$$cx + d = A(x-b) + B(x-a)$$

When $x = a$ the $x - a$ factor is 0, so this equation becomes

$$c \cdot a + d = A(a-b)$$

Likewise, when $x = b$ the $x - b$ factor is 0, so we also have

$$c \cdot b + d = B(b-a)$$

That is,

$$A = \frac{c \cdot a + d}{a-b} \qquad B = \frac{c \cdot b + d}{b-a}$$

So, yes, we can find the constants to break the fraction $(cx + d)/(x - a)(x - b)$ down into simpler 'partial' fractions.

Further, if the numerator is of *bigger degree* than 1, then before executing the previous algebra trick we must first *divide the numerator by the denominator to get a remainder of smaller degree.*

Example 7.2.6.

[`author=garrett`, file =`text_files/partial_fractions`]
A simple example is

$$\frac{x^3 + 4x^2 - x + 1}{x(x-1)} = ?$$

We must recall how to divide polynomials by polynomials and get a remainder of lower degree than the divisor. Here we would divide the $x^3 + 4x^2 - x + 1$ by $x(x - 1) = x^2 - x$ to get a remainder of degree less than 2 (the degree of $x^2 - x$). We would obtain

$$\frac{x^3 + 4x^2 - x + 1}{x(x-1)} = x + 5 + \frac{4x+1}{x(x-1)}$$

since the quotient is $x + 5$ and the remainder is $4x + 1$. Thus, in this situation

$$\int \frac{x^3 + 4x^2 - x + 1}{x(x-1)}\, dx = \int x + 5 + \frac{4x+1}{x(x-1)}\, dx$$

Now we are ready to continue with the *first* algebra trick.

In this case, the first trick is applied to

$$\frac{4x+1}{x(x-1)}$$

We want constants A, B so that

$$\frac{4x+1}{x(x-1)} = \frac{A}{x} + \frac{B}{x-1}$$

As above, multiply through by $x(x - 1)$ to get

$$4x + 1 = A(x - 1) + Bx$$

and plug in the two values $0, 1$ to get

$$4 \cdot 0 + 1 = -A \qquad 4 \cdot 1 + 1 = B$$

That is, $A = -1$ and $B = 5$.

Putting this together, we have

$$\frac{x^3 + 4x^2 - x + 1}{x(x-1)} = x + 5 + \frac{-1}{x} + \frac{5}{x-1}$$

Thus,

$$\int \frac{x^3 + 4x^2 - x + 1}{x(x-1)}\, dx = \int x + 5 + \frac{-1}{x} + \frac{5}{x-1}\, dx$$
$$= \frac{x^2}{2} + 5x - \ln x + 5\ln(x-1) + C$$

Rule 7.2.2.

[author=garrett, file =text_files/partial_fractions]
In a slightly different direction: we can do any integral of the form

$$\int \frac{ax+b}{1+x^2}\,dx$$

because we know two different sorts of integrals with that same denominator:

$$\int \frac{1}{1+x^2}\,dx = \arctan x + C \qquad \int \frac{2x}{1+x^2}\,dx = \ln(1+x^2) + C$$

where in the second one we use a substitution. Thus, we have to break the given integral into two parts to do it:

$$\begin{aligned}\int \frac{ax+b}{1+x^2}\,dx &= \frac{a}{2}\int \frac{2x}{1+x^2}\,dx + b\int \frac{1}{1+x^2}\,dx \\ &= \frac{a}{2}\ln(1+x^2) + b\arctan x + C\end{aligned}$$

Example 7.2.7.

[author=garrett, file =text_files/partial_fractions]
And, as in the first example, if we are given a numerator of degree 2 or larger, then we *divide* first, to get a remainder of lower degree. For example, in the case of

$$\int \frac{x^4+2x^3+x^2+3x+1}{1+x^2}\,dx$$

we divide the numerator by the denominator, to allow us to write

$$\frac{x^4+2x^3+x^2+3x+1}{1+x^2} = x^2+2x+\frac{x+1}{1+x^2}$$

since the quotient is x^2+2x and the remainder is $x+1$. Then

$$\begin{aligned}\int \frac{x^4+2x^3+x^2+3x+1}{1+x^2}\,dx &= \int x^2+2x+\frac{x+1}{1+x^2} \\ &= \frac{x^3}{3}+x^2+\frac{1}{2}\ln(1+x^2)+\arctan x + C\end{aligned}$$

These two examples are just the simplest, but illustrate the idea of using algebra to simplify rational functions.

Example 7.2.8.

[author=wikibooks, file =text_files/partial_fractions]
First, an example. $\frac{1}{x^3+x^2+x+1} = \frac{1}{x} - \frac{x}{1+x^2}$ so

$$\begin{aligned}\int \frac{dx}{x^3+x^2+x+1} &= \int \frac{dx}{x} - \int \frac{x\,dx}{1+x^2} \\ &= \ln x - \frac{1}{2}\ln(1+x^2) \\ &= \ln \frac{x}{\sqrt{1+x^2}}\end{aligned}$$

Rewriting the integrand as a sum of simpler fractions has allowed us to reduce the initial, more complex, integral to a sum of simpler integrals.

Rule 7.2.3.

[author=wikibooks, file =text_files/partial_fractions]
More generally, if we have a $Q(x)$ which is the product of p factors of the form $(x-a_i)^{n_i}$ and q factors of the form $\left((x-b_i)^2-c_i\right)^{n_i}$ then we can write any P/Q as a sum of simpler terms, each with a power of only one factor in the denominator:

$$\begin{aligned}\frac{P(x)}{Q(x)} = & \frac{d_{1,1}}{x-a_1} + \cdots + \frac{d_{p,n_1}}{(x-a_p)^{n_p}} + \cdots \\ & + \frac{f_{1,1}+g_{1,1}x}{(x-b_1)^2-c_1} + \cdots + \frac{f_{q,n_q}+g_{q,n_q}x}{((x-b_q)^2-c_q)^{n_q}}\end{aligned}$$

then solve for the new constants. If we were using complex numbers none of the factors of Q would be quadratic.

Example 7.2.9.

[author=wikibooks, file =text_files/partial_fractions]
We will consider a few more examples, to see how the procedure goes. Consider

$$1/P(x) = 1+x^2 \text{ and } Q(x) = (x+3)(x+5)(x+7).$$

We first write $\frac{1+x^2}{(x+3)(x+5)(x+7)} = \frac{a}{x+3} + \frac{b}{x+5} + \frac{c}{x+7}$ Multiply both sides by the denominator $1+x^2 = a(x+5)(x+7) + b(x+3)(x+7) + c) + c(x+3)(x+7)$ Substitute in three values of x to get three equations for the unknown constants,

$x=-3 \quad 1+3^2 = 2\cdot 4a$
$x=-5 \quad 1+5^2 = -2\cdot 2b$ so a=5/4, b=-13/2, c=25/2, and $\frac{1+x^2}{(x+3)(x+5)(x+7)} =$
$x=-7 \quad 1+7^2 = (-4)\cdot(-2)c$

$\frac{5}{4x+12} - \frac{13}{2x+10} + \frac{25}{2x+14}$ We can now integrate the left hand side. $\int \frac{1+x^2\,dx}{(x+3)(x+5)(x+7)} = \ln \frac{(x+3)^{\frac{5}{4}}(x+7)^{\frac{25}{2}}}{(x+5)^{\frac{13}{2}}}$

Example 7.2.10.

[author=wikibooks, file =text_files/partial_fractions]
$2/P(x) = 1$, $Q(x) = (x+1)(x+2)^2$ We first write $\frac{1}{(x+1)(x+2)^2} = \frac{a}{x+1} + \frac{b}{x+2} + \frac{c}{(x+2)^2}$ Multiply both sides by the denominator $1 = a(x+2)^2 + b(x+1)(x+2) + c(x+1)$ Substitute in three values of x to get three equations for the unknown constants,

$x=0 \quad 1 = 2^2a + 2b + c$
$x=-1 \quad 1 = a$ so a=1, b=-1, c=-1, and $\frac{1}{(x+1)(x+2)^2} = \frac{1}{x+1} - \frac{1}{x+2} -$
$x=-2 \quad 1 = -c$

$\frac{1}{(x+2)^2}$ We can now integrate the left hand side. $\int \frac{dx}{(x+1)(x+2)^2} = \ln \frac{x+1}{x+2} + \frac{1}{x+2}$

Exercises

1. $\int \frac{1}{x(x-1)}\,dx = ?$

2. $\int \frac{1+x}{1+x^2}\,dx = ?$

3. $\int \frac{2x^3+4}{x(x+1)}\, dx =?$

4. $\int \frac{2+2x+x^2}{1+x^2}\, dx =?$

5. $\int \frac{2x^3+4}{x^2-1}\, dx =?$

6. $\int \frac{2+3x}{1+x^2}\, dx =?$

7. $\int \frac{x^3+1}{(x-1)(x-2)}\, dx =?$

8. $\int \frac{x^3+1}{x^2+1}\, dx =?$

7.3 Trigonometric Integrals

Discussion.

[author=duckworth, file =text_files/trigonometric_integrals]
This section gives tricks for solving integrals of the form $\int \sin^n \cos^m$ and $\int \tan^n \sec^m$.

Procedure.

[author=duckworth, file =text_files/trigonometric_integrals]
For $\int \sin^n(x) \cos^m(x)\, dx$ use:

- if n is odd get rid of all but 1 power of sin using $\sin^2 = 1 - \cos^2$, then use $u = \cos$ and $du = -\sin\, dx$.
- if m is odd get rid of all but 1 power of cos using $\cos^2 = 1 - \sin^2$, then use $u = \sin$ and $du = \cos\, dx$.
- if n and m are even, use $\sin^2(x) = \frac{1}{2}(1 - \cos(2x))$ and $\cos^2(x) = \frac{1}{2}(1 + \cos(2x))$ (may have to repeat this step) to get everything in terms of $\cos(2x)$, $\cos(4x)$ etc.

Procedure.

[author=duckworth, file =text_files/trigonometric_integrals]
For $\int \tan^n(x) \sec^m(x)\, dx$ use:

- if n is odd get rid of all but 1 power of tan using $\tan^2 = \sec^2 - 1$, force one power of sec out next to tan, and use $u = \sec$, $du = \sec\tan$.
- if m is even get rid of all but 2 powers of sec using $\sec^2 = \tan^2 + 1$, use $u = \tan$ and $du = \sec^2$.
- if n is even and m is odd get rid of all powers of tan using $\tan^2 = \sec^2 - 1$. Now we have only powers of sec, use integration by parts and $\int \sec(x)\, dx = \ln|\sec(x) + \tan(x)|$.

Example 7.3.1.

[author=duckworth, file =text_files/trigonometric_integrals]
$\int \sin^7(x) \cos^2(x)\, dx$. We get rid of $\sin^6(x)$ by rewriting it as $(1 - \cos^2(x))^3$. Then we have:

$$\int \sin^7(x) \cos^2(x)\, dx = \int \sin(x)(1 - \cos^2(x))^3 \cos^2(x)\, dx = -\int (1 - u^2)^3 u^2\, du$$

which you can solve by multiplying out.

Example 7.3.2.

[author=duckworth, file =text_files/trigonometric_integrals]
$\int \tan^2(x)\sec(x)\,dx$. We get rid of $\tan^2(x)$ by rewriting it as $\sec^2(x) - 1$. Then we have:

$$\int \tan^2(x)\sec(x)\,dx = \int (\sec^2(x) - 1)\sec(x)\,dx = \int \sec^3(x) - \sec(x)\,dx$$

The book does $\int \sec^3(x)\,dx$ (the trick for this is integration by parts once, $\tan^2 = \sec^2 - 1$, and solving an equation for $\int \sec^3(x)\,dx$) and we stated $\int \sec(x)\,dx$ above.

Discussion.

[author=garrett, file =text_files/trigonometric_integrals]
Here we'll just have a *sample* of how to use trig identities to do some more complicated integrals involving trigonometric functions. This is 'just the tip of the iceberg'. We don't do more for at least two reasons: first, hardly anyone remembers all these tricks anyway, and, second, in real life you can look these things up in tables of integrals. Perhaps even more important, in 'real life' there are more sophisticated viewpoints which even make the whole issue a little silly, somewhat like evaluating $\sqrt{26}$ 'by differentials' without your calculator seems silly.

The only identities we'll need in our examples are

$$\begin{array}{ll} \cos^2(x) + \sin^2(x) = 1 & \text{Pythagorean identity} \\ \sin(x) = \sqrt{\frac{1-\cos(2x)}{2}} & \text{half-angle formula} \\ \cos(x) = \sqrt{\frac{1+\cos(2x)}{2}} & \text{half-angle formula} \end{array}$$

Example 7.3.3.

[author=garrett, file =text_files/trigonometric_integrals]
The first example is

$$\int \sin^3 x\ dx$$

If we ignore all trig identities, there is no easy way to do this integral. But if we use the Pythagorean identity to rewrite it, then things improve:

$$\int \sin^3 x\ dx = \int (1 - \cos^2 x)\sin x\ dx = -\int (1 - \cos^2 x)(-\sin x)\ dx$$

In the latter expression, we can view the $-\sin x$ as the derivative of $\cos x$, so with the substitution $u = \cos x$ this integral is

$$-\int (1 - u^2)\ du = -u + \frac{u^3}{3} + C = -\cos x + \frac{\cos^3 x}{3} + C$$

Example 7.3.4.

[author=garrett, file =text_files/trigonometric_integrals]
This idea can be applied, more generally, to integrals

$$\int \sin^m x\ \cos^n x\ dx$$

where *at least one of m, n is odd*. For example, if n is odd, then use

$$\cos^n x = \cos^{n-1} x\ \cos x = (1 - \sin^2 x)^{\frac{n-1}{2}}\ \cos x$$

to write the whole thing as

$$\int \sin^m x\ \cos^n x\ dx = \int \sin^m x\ (1 - \sin^2 x)^{\frac{n-1}{2}}\ \cos x\ dx$$

The point is that we have obtained something of the form

$$\int\ (\text{polynomial in } \sin x)\ \cos x\ dx$$

Letting $u = \sin x$, we have $\cos x\ dx = du$, and the integral becomes

$$(\text{polynomial in } u)\ du$$

which we can do.

Example 7.3.5.

[author=garrett, file =text_files/trigonometric_integrals]
But this Pythagorean identity trick does not help us on the relatively simple-looking integral

$$\int \sin^2(x)\, dx$$

since there is no odd exponent anywhere. In effect, we 'divide the exponent by two', thereby getting an odd exponent, by using the *half-angle formula*:

$$\int \sin^2 x\ dx = \int \frac{1 - \cos\ 2x}{2}\ dx = \frac{x}{2} - \frac{\sin 2x}{2 \cdot 2} + C$$

Example 7.3.6.

[author=garrett, file =text_files/trigonometric_integrals]
A bigger version of this application of the half-angle formula is

$$\int \sin^6 x\ dx = \int (\frac{1 - \cos\ 2x}{2})^3\ dx = \int \frac{1}{8} - 38 \cos 2x + \frac{3}{8} \cos^2 2x - \frac{1}{8} \cos^3 2x\ dx$$

Of the four terms in the integrand in the last expression, we can do the first two directly:

$$\int \frac{1}{8}\ dx = \frac{x}{8} + C \qquad \int -38 \cos 2x\ dx = \frac{-3}{16} \sin 2x + C$$

But the last two terms require further work: using a half-angle formula *again*, we have

$$\int \frac{3}{8} \cos^2 2x\ dx = \int \frac{3}{16}(1 + \cos 4x)\ dx = \frac{3x}{16} + \frac{3}{64} \sin 4x + C$$

And the $\cos^3 2x$ needs the Pythagorean identity trick:

$$\int \frac{1}{8}\cos^3 2x\, dx = \frac{1}{8}\int (1-\sin^2 2x)\cos 2x\, dx = \frac{1}{8}[\sin 2x - \frac{\sin^3 2x}{3}] + C$$

Putting it all together, we have

$$\int \sin^6 x\, dx = \frac{x}{8} + \frac{-3}{16}\sin 2x + \frac{3x}{16} + \frac{3}{64}\sin 4x + \frac{1}{8}[\sin 2x - \frac{\sin^3 2x}{3}] + C$$

This last example is typical of the kind of repeated application of all the tricks necessary in order to treat all the possibilities.

Example 7.3.7.

[author=garrett, file =text_files/trigonometric_integrals]
In a slightly different vein, there is the horrible

$$\int \sec x\, dx$$

There is no decent way to do this at all from a first-year calculus viewpoint. A sort of rationalized-in-hindsight way of explaining the answer is:

$$\int \sec x\, dx = \int \frac{\sec x(\sec x + \tan x)}{\sec x + \tan x}\, dx$$

All we did was multiply and divide by $\sec x + \tan x$. Of course, we don't pretend to answer the question of how a person would get the idea to do this. But then (another miracle?) we 'notice' that the numerator is the derivative of the denominator, so

$$\int \sec x\, dx = \ln(\sec x + \tan x) + C$$

There is something distasteful about this rationalization, but at this level of technique we're stuck with it.

Comment.

[author=garrett, file =text_files/trigonometric_integrals]
Maybe this is enough of a sample. There are several other tricks that one would have to know in order to claim to be an 'expert' at this, but it's not really sensible to *want* to be 'expert' at these games, *because there are smarter alternatives.*

Discussion.

[author=wikibooks, file =text_files/trigonometric_integrals]
We're going to find formulas for integrals of the form $\int \cos^m \sin^n$, but we start with an example.

Example 7.3.8.

[author=wikibooks, file =text_files/trigonometric_integrals]

Let $I = \int (\cos(x))^3 (\sin(x))^2\,dx$. Making the substitution $u = \sin(x)$, $du = \cos(x)dx$ and using the fact $\cos(x)^2 = 1 - \sin(x)^2$ we obtain $I = \int (1-u^2)u^2\,du$ which we can solve easily to obtain $I = \int u^2\,du - \int u^4\,du = 1/3u^3 + 1/5u^5 + C$ $= 1/3(\sin(x))^3 - 1/5(\sin(x))^5 + C$

Rule 7.3.1.

[author=wikibooks, file =text_files/trigonometric_integrals]
In general we have, for $\int (\cos(x))^m (\sin(x))^n\,dx$

- for m odd substitute $u = \sin x$ and use the fact that $(\cos x)^2 = 1 - (\sin x)^2$
- for m even substitute $u = \cos x$ and use the fact that $(\sin x)^2 = 1 - (\cos x)^2$
- for m and n both even, use the fact that $(\sin x)^2 = 1/2(1 - \cos 2x)$ and $(\cos x)^2 = 1/2(1 + \cos 2x)$

Example 7.3.9.

[author=wikibooks, file =text_files/trigonometric_integrals]
For example, for m and n even, say $I = \int (\sin x)^2 (\cos x)^4\,dx$ making the substitutions gives $I = \int \left(\frac{1}{2}(1-\cos 2x)\right)\left(\frac{1}{2}(1+\cos 2x)\right)^2\,dx$

Expanding this out $I = \frac{1}{8}\left(\int 1 - \cos^2 2x + \cos 2x - \cos^3 2x\,dx\right)$

Using the multiple angle identities

$$\begin{aligned} I &= \frac{1}{8}\left(\int 1\,dx - \int \cos^2 2x\,dx + \int \cos 2x\,dx - \int \cos^3 2x\,dx\right) \\ &= \frac{1}{8}\left(x - \frac{1}{2}\int (1+\cos 4x)\,dx + \frac{1}{2}\sin 2x - \int \cos^2 2x \cos 2x\,dx\right) \\ &= \frac{1}{16}\left(x + \sin 2x + \int \cos 4x\,dx - 2\int (1-\sin^2 2x)\cos 2x\,dx\right) \end{aligned}$$

then we obtain on evaluating

$$I = \frac{x}{16} - \frac{\sin 4x}{64} + \frac{\sin^3 2x}{48} + C.$$

Discussion.

[author=wikibooks, file =text_files/trigonometric_integrals]
Another useful change of variables is $t = \tan(x/2)$. With this transformation, using the double-angle trig identities, $\sin x = \frac{2t}{1+t^2}$, $\cos x = \frac{1-t^2}{1+t^2}$, $\tan x = \frac{2t}{1-t^2}$ and $dx = \frac{2dt}{1+t^2}$. This transforms a trigonometric integral into a algebraic integral, which may be easier to integrate.

Example 7.3.10.

[author=wikibooks, file =text_files/trigonometric_integrals]

For example, if the integrand is $1/(1+\sin x)$ then

$$\begin{aligned} \int_0^{\pi/2} \frac{dx}{1+\sin x} &= \int_0^1 \frac{2dt}{(1+t)^2} \\ &= \left[-\frac{2}{1+t}\right]_0^1 \\ &= 1 \end{aligned}$$

This method can be used to further simplify trigonometric integrals produced by the changes of variables described earlier.

Example 7.3.11.

[author=wikibooks, file =text_files/trigonometric_integrals]
For example, if we are considering the integral

$$I = \int_{-1}^{1} \frac{\sqrt{1-x^2}}{1+x^2}\, dx$$

we can first use the substition $x = \sin\theta$, which gives

$$I = \int_{-\pi/2}^{\pi/2} \frac{\cos^2\theta}{1+\sin^2\theta} d\theta$$

then use the tan-half-angle substition to obtain

$$I = \int_{-1}^{1} \frac{(1-t^2)^2}{1+6t^2+t^4} \frac{2dt}{1+t^2}.$$

In effect, we've removed the square root from the original integrand. We could do this with a single change of variables, but doing it in two steps gives us the opportunity of doing the trigonometric integral another way.

Having done this, we can split the new integrand into partial fractions, and integrate.

$$\begin{aligned} I &= \int_{-1}^{1} \frac{2-\sqrt{2}}{t^2+3-\sqrt{8}} dt + \int_{-1}^{1} \frac{2+\sqrt{2}}{t^2+3+\sqrt{8}} dt - \int_{-1}^{1} \frac{2}{1+t^2} dt \\ &= \frac{4-\sqrt{8}}{\sqrt{3-\sqrt{8}}} \tan^{-1}(\sqrt{3+\sqrt{8}}) + \frac{4+\sqrt{8}}{\sqrt{3+\sqrt{8}}} \tan^{-1}(\sqrt{3-\sqrt{8}}) - \pi \end{aligned}$$

This result can be further simplified by use of the identities

$$3 \pm \sqrt{8} = \left(\sqrt{2} \pm 1\right)^2 \quad \tan\left(\sqrt{2} \pm 1\right) = \left(\frac{1}{4} \pm \frac{1}{8}\right)\pi$$

ultimately leading to

$$I = (\sqrt{2}-1)\pi$$

In principle, this approach will work with any integrand which is the square root of a quadratic multiplied by the ratio of two polynomials. However, it should not be applied automatically.

Example 7.3.12.

[author=wikibooks, file =text_files/trigonometric_integrals]
E.g, in this last example, once we deduced

$$I = \int_{-\pi/2}^{\pi/2} \frac{\cos^2\theta}{1+\sin^2\theta} d\theta$$

we could have used the double angle formulae, since this contains only even powers of cos and sin. Doing that gives

$$I = \int_{-\pi/2}^{\pi/2} \frac{1+\cos 2\theta}{3-\cos 2\theta} d\theta = \frac{1}{2}\int_{-\pi}^{\pi} \frac{1+\cos\phi}{3-\cos\phi} d\phi$$

Using tan-half-angle on this new, simpler, integrand gives

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \tfrac{1}{1+2t^2} \tfrac{dt}{1+t^2} \\ &= \int_{-\infty}^{\infty} \tfrac{2dt}{1+2t^2} - \int_{-\infty}^{\infty} \tfrac{dt}{1+t^2} \end{aligned}$$

This can be integrated on sight to give

$$I = \frac{4}{\sqrt{2}}\frac{\pi}{2} - 2\frac{\pi}{2} = (\sqrt{2}-1)\pi$$

This is the same result as before, but obtained with less algebra, which shows why it is best to look for the most straightforward methods at every stage.

Rule 7.3.2.

[**author=wikibooks**, file =**text_files/trigonometric_integrals**]
For the integrals $\int \sin nx \cos mx\, dx$, $\int \sin nx \sin mx\, dx$, $\int \cos nx \cos mx\, dx$ use the following identities $2\sin a \cos b = (\sin(a+b) + \sin(a-b))$, $2\sin a \sin b = (\cos(a-b) - \cos(a+b))$, $2\cos a \cos b = (\cos(a-b) + \cos(a+b))$

Example 7.3.13.

[**author=wikibooks**, file =**text_files/trigonometric_integrals**]
Find the integral $\int \sin 3x \cos 5x\, dx$.

We use the fact that $\sin(a)\cos(b) = \frac{1}{2}(\sin(a+b) + \sin(a-b))$, so $\sin 3x \cos 5x = a\frac{1}{2}(\sin(7x) + \sin(-2x)) = frac12(\sin(7x) - \sin(2x))/2$, where we have used the fact that $\sin(x)$ is an odd function. And now we can integrate

$$\begin{aligned} \int \sin(3x)\cos(5x)\, dx &= \tfrac{1}{2}\int \sin(7x) - \sin(2x) dx \\ &= \tfrac{1}{2}(-\tfrac{1}{7}\cos(7x) + \tfrac{1}{2}\cos(2x)) \end{aligned}$$

Example 7.3.14.

[**author=wikibooks**, file =**text_files/trigonometric_integrals**]
Find the integral $\int \sin(x)\sin(2x)\, dx$.

Use $\sin x \sin 2x = \frac{1}{2}(\cos(-x) - \cos(3x)) = \frac{1}{2}(\cos x - \cos 3x)$. Then $\int \sin(x)\sin 2x\, dx = \frac{1}{2}\int($ $= \frac{1}{2}($

Rule 7.3.3.

[**author=wikibooks**, file =**text_files/trigonometric_integrals**]
A reduction formula is one that enables us to solve an integral problem by reducing it to a problem of solving an easier integral problem, and then reducing that to

the problem of solving an easier problem, and so on.

Example 7.3.15.

[author=wikibooks, file =text_files/trigonometric_integrals]
For example, if we let $I_n = \int x^n e^x \, dx$ Integration by parts allows us to simplify this to $I_n = x^n e^x - n \int x^{n-1} e^x \, dx = I_n = x^n e^x - nI_{n-1}$ which is our desired reduction formula. Note that we stop at $I_0 = e^x$.

Similarly, if we let

$I_n = \int_0^\alpha \sec^n \theta \, d\theta$

then integration by parts lets us simplify this to

$I_n = \sec^{n-2} \alpha \tan \alpha - (n-2) \int_0^\alpha \sec^{n-2} \theta \tan^2 \theta \, d\theta$

Using the trigonometric identity, $tan^2 = sec^2 - 1$, we can now write

$$\begin{aligned} I_n &= \sec^{n-2} \alpha \tan \alpha \quad +(n-2) \left(\int_0^\alpha \sec^{n-2} \theta \, d\theta - \int_0^\alpha \sec^n \theta \, d\theta \right) \\ &= \sec^{n-2} \alpha \tan \alpha \quad +(n-2) \left(I_{n-2} - I_n \right) \end{aligned}$$

Rearranging, we get

$I_n = \frac{1}{n-1} \sec^{n-2} \alpha \tan \alpha + \frac{n-2}{n-1} I_{n-2}$

Note that we stop at n=1 or 2 if n is odd or even respectively.

As in these two examples, integrating by parts when the integrand contains a power often results in a reduction formula.

Exercises

1. $\int \cos^2 x \, dx = ?$
2. $\int \cos x \sin^2 x \, dx = ?$
3. $\int \cos^3 x \, dx = ?$
4. $\int \sin^2 5x \, dx = ?$
5. $\int \sec(3x + 7) \, dx$
6. $\int \sin^2 (2x + 1) \, dx = ?$
7. $\int \sin^3 (1 - x) \, dx = ?$

7.4 Trigonometric Substitutions

Discussion.

[author=duckworth, file =text_files/trigonometric_subst]
The basic idea here is that we reverse the usual role of u-substitution. Usually, we set u equal to some function of x because this "covers up" some complicated function. But here, we're going to set x equal to a more complicated function (of θ) because of the special properties of trig functions.

Procedure.

[author=duckworth, file =text_files/trigonometric_subst]

If the integral involves	use
$\sqrt{a^2-x^2}$	$x = a\sin(\theta)$
$\sqrt{a^2+x^2}$	$x = a\tan(\theta)$
$\sqrt{x^2-a^2}$	$x = a\sec(\theta)$

Example 7.4.1.

[author=duckworth, file =text_files/trigonometric_subst]

(a) Find the area under a circle with radius 1, from $x = 0$ to $x = 1/2$. This is $\int_0^{1/2}\sqrt{1-x^2}\,dx$. The hard part is coming up with the definite integral. Let $x = \sin(\theta)$, then $dx = \cos(\theta)\,d\theta$. Note that $\sqrt{1-x^2} = \sqrt{1-\sin^2(\theta)} = \cos(\theta)$. We also translate the endpoints of the integral. When $x = 0$ we have $\sin(\theta) = 0$ so $\theta = 0$. When $x = 1/2$ we have $\sin(\theta) = 1/2$ so $\theta = \pi/6$. So we have

$$\int_0^{1/2}\sqrt{1-x^2}\,dx = \int_0^{\pi/6}\cos(\theta)\cdot\cos(\theta)\,d\theta = \int_0^{\pi/6}\cos^2(\theta)\,d\theta.$$

We look up this integral from section 7.1 or 7.2 as $\frac{1}{2}\theta + \frac{1}{2}\sin(\theta)\cos(\theta)$ so the final answer is found by plugging in $\theta = \pi/6$ and $\theta = 0$.

(b) Find the indefinite integral in part (a) (i.e. the anti-derivative $\int\sqrt{1-x^2}\,dx$). Well, we know this is $\frac{1}{2}\theta + \frac{1}{2}\sin(\theta)\cos(\theta)$, so we just need to translate from θ back to x. By the definition of our substitution we have $x = \sin(\theta)$. To find $\cos(\theta)$ in terms of x you can draw a right triangle, label an angle as θ, the opposite side as x, the hypotenuse as 1 (this is because $\sin(\theta) = x$) and solve for the missing side. You should find that $\cos(\theta) = \sqrt{1-x^2}$ (by the way, it always works out this way; the missing side is the $\sqrt{\ }$ that you started with in the integral). Finally, $\theta = \sin^{-1}(x)$ (because $\sin(\theta) = x$). Thus,

$$\int\sqrt{1-x^2}\,dx \quad = \quad \frac{1}{2}\theta + \frac{1}{2}\sin(\theta)\cos(\theta) \quad = \quad \frac{1}{2}\sin^{-1}(x) + \frac{1}{2}x\sqrt{1-x^2}.$$

(If you want, you can get the same answer as in (a) by plugging in $x = 1/2$ and $x = 0$ to evaluate this definite integral, i.e. to find the area under the curve.)

Discussion.

[author=garrett, file =text_files/trigonometric_subst]
This section continues development of relatively special tricks to do special kinds of integrals. Even though the application of such things is limited, it's nice to be *aware* of the possibilities, at least a little bit.

The key idea here is to use trig functions to be able to 'take the square root' in certain integrals. There are just three prototypes for the kind of thing we can deal with:

$$\sqrt{1-x^2}, \quad \sqrt{1+x^2}, \sqrt{x^2-1}$$

Examples will illustrate the point.

In rough terms, the idea is that in an integral where the 'worst' part is $\sqrt{1-x^2}$, replacing x by $\sin u$ (and, correspondingly, dx by $\cos u\, du$), *we will be able to take the square root*, and then obtain an integral in the variable u which is one of the *trigonometric integrals* which in principle we now know how to do. The point is that then

$$\sqrt{1-x^2} = \sqrt{1-\sin^2 x} = \sqrt{\cos^2 x} = \cos x$$

We have 'taken the square root'.

Example 7.4.2.

[author=garrett, file =text_files/trigonometric_subst]
For example, in

$$\int \sqrt{1-x^2}\ dx$$

we replace x by $\sin u$ and dx by $\cos u\, du$ to obtain

$$\int \sqrt{1-x^2}\ dx = \int \sqrt{1-\sin^2 u}\ \cos u\ du = \int \sqrt{\cos^2 u}\ \cos u\ du =$$

$$= \int \cos u\ \cos u\ du = \int \cos^2 u\ du$$

Now we have an integral we know how to integrate: using the half-angle formula, this is

$$\int \cos^2 u\ du = \int \frac{1+\cos 2u}{2}\ du = \frac{u}{2} + \frac{\sin 2u}{4} + C$$

And there still remains the issue of *substituting back* to obtain an expression in terms of x rather than u. Since $x = \sin u$, it's just the definition of *inverse function* that

$$u = \arcsin x$$

To express $\sin 2u$ in terms of x is more aggravating. We use another *half-angle formula*

$$\sin 2u = 2 \sin u \cos u$$

Then

$$\frac{1}{4}\sin 2u = \frac{1}{4} \cdot 2 \sin u \cos u = \frac{1}{4} x \cdot \sqrt{1-x^2}$$

where 'of course' we used the Pythagorean identity to give us

$$\cos u = \sqrt{1-\sin^2 u} = \sqrt{1-x^2}$$

Whew.

Rule 7.4.1.

[author=garrett, file =text_files/trigonometric_subst]
The next type of integral we can 'improve' is one containing an expression

$$\sqrt{1+x^2}$$

In this case, we use another Pythagorean identity

$$1+\tan^2 u = \sec^2 u$$

(which we can get from the usual one $\cos^2 u + \sin^2 u = 1$ by dividing by $\cos^2 u$). So we'd let

$$x = \tan u \, dx = \sec^2 u \, du$$

(mustn't forget the dx and du business!).

Example 7.4.3.

[author=garrett, file =text_files/trigonometric_subst]
For example, in

$$\int \frac{\sqrt{1+x^2}}{x}\, dx$$

we use

$$x = \tan u \qquad dx = \sec^2 u \, du$$

and turn the integral into

$$\int \frac{\sqrt{1+x^2}}{x}\, dx = \int \frac{\sqrt{1+\tan^2 u}}{\tan u} \sec^2 u \, du =$$

$$= \int \frac{\sqrt{\sec^2 u}}{\tan u} \sec^2 u \, du = \int \frac{\sec u}{\tan u} \sec^2 u \, du = \int \frac{1}{\sin u \cos^2 u}\, du$$

by rewriting everything in terms of $\cos u$ and $\sin u$.

Rule 7.4.2.

[author=garrett, file =text_files/trigonometric_subst]
For integrals containing $\sqrt{x^2-1}$, use $x = \sec u$ in order to invoke the Pythagorean identity

$$\sec^2 u - 1 = \tan^2 u$$

so as to be able to 'take the square root'. Let's not execute any examples of this, since nothing new really happens.

Discussion.

[author=garrett, file =text_files/trigonometric_subst]
Let's examine some *purely algebraic variants* of these trigonometric substitutions, where we can get some mileage out of *completing the square*.

Example 7.4.4.

[author=garrett, file =text_files/trigonometric_subst]
For example, consider

$$\int \sqrt{-2x - x^2}\, dx$$

The quadratic polynomial inside the square-root is *not* one of the three simple types we've looked at. But, by completing the square, we'll be able to rewrite it in essentially such forms:

$$-2x - x^2 = -(2x + x^2) = -(-1 + 1 + 2x + x^2) = -(-1 + (1 + x)^2) = 1 - (1 + x)^2$$

Note that always when completing the square we 'take out' the coefficient in front of x^2 in order to see what's going on, and then put it back at the end.

So, in this case, we'd let

$$\sin u = 1 + x, \quad \cos u\, du = dx$$

Example 7.4.5.

[author=garrett, file =text_files/trigonometric_subst]
In another example, we might have

$$\int \sqrt{8x - 4x^2}\, dx$$

Completing the square again, we have

$$8x - 4x^2 = -4(-2 + x^2) = -4(-1 + 1 - 2x + x^2) = -4(-1 + (x - 1)^2)$$

Rather than put the whole '-4' back, we only keep track of the $\pm$, and take a '$+4$' outside the square root entirely:

$$\int \sqrt{8x - 4x^2}\, dx = \int \sqrt{-4(-1 + (x - 1)^2)}\, dx$$

$$= 2\int \sqrt{-(-1 + (x - 1)^2)}\, dx = 2\int \sqrt{1 - (x - 1)^2)}\, dx$$

Then we're back to a familiar situation.

Rule 7.4.3.

[author=wikibooks, file =text_files/trigonometric_subst]
If the integrand contains a factor of this form we can use the substitution

$$x = a\sin(\theta)\, dx = a\cos(\theta)\, d\theta$$

This will transform the integrand to a trigonometic function. If the new integrand can't be integrated on sight then the tan-half-angle substitution described below will generally transform it into a more tractable algebraic integrand.

Example 7.4.6.

[author=wikibooks, file =text_files/trigonometric_subst]
Find the integral of $\sqrt(1-x^2)$,

$$\begin{aligned}
\int_0^1 \sqrt{1-x^2}dx &= \int_0^{\pi/2} \sqrt{1-\sin^2\theta}\cos\theta\, d\theta \\
&= \int_0^{\pi/2} \cos^2\theta\, d\theta \\
&= \frac{1}{2}\int_0^{\pi/2} 1+\cos 2\theta\, d\theta \\
&= \frac{\pi}{4}
\end{aligned}$$

Example 7.4.7.

[author=wikibooks, file =text_files/trigonometric_subst]
Find the integral of $\sqrt(1+x)/\sqrt(1-x)$. We first rewrite this as

$$\sqrt{\frac{1+x}{1-x}} = \sqrt{\frac{1+x}{1+x}\frac{1+x}{1-x}} = \frac{1+x}{\sqrt{1-x^2}}$$

Then we can make the substitution

$$\begin{aligned}
\int_0^a \frac{1+x}{\sqrt{1-x^2}}dx &= \int_0^\alpha \frac{1+\sin\theta}{\cos\theta}\cos\theta\, d\theta & 0<a<1 \\
&= \int_0^\alpha 1+\sin\theta, d\theta & \alpha=\sin^{-1} a \\
&= \alpha + [-\cos\theta]_0^\alpha \\
&= \alpha + 1 - \cos\alpha \\
&= 1+\sin^{-1} a - \sqrt{1-a^2}
\end{aligned}$$

Rule 7.4.4.

[author=wikibooks, file =text_files/trigonometric_subst]
If the integrand contains a factor of the form $\sqrt{x^2-a^2}$ we use the substitution

$$x = a\sec\theta \quad dx = a\sec\theta\tan\theta d\theta \quad \sqrt{x^2-a^2} = \tan\theta$$

This will transform the integrand to a trigonometic function. If the new integrand can't be integrated on sight then another substitution may transform it to a more tractable algebraic integrand.

Example 7.4.8.

[author=wikibooks, file =text_files/trigonometric_subst]
Find the integral of $\sqrt(x^2-1)/x$.

We use substitution:

$$\begin{aligned}
\int_1^z \frac{\sqrt{x^2-1}}{x}dx &= \int_1^\alpha \frac{\tan\theta}{\sec\theta}\sec\theta\tan\theta\, d\theta & z>1 \\
&= \int_0^\alpha \tan^2\theta\, d\theta & \alpha=\sec^{-1} z \\
&= [\tan\theta-\theta]_0^\alpha & \tan\alpha=\sqrt{\sec^2\alpha-1} \\
&= \tan\alpha-\alpha & \tan\alpha=\sqrt{z^2-1} \\
&= \sqrt{z^2-1}-\sec^{-1} z
\end{aligned}$$

Since the integrand is approximately 1 for large x we should expect the integral at large z to be z plus a constant. It is actually $z - \pi/2$, as we expected. We can use this line of reasoning to check our calculations.

Example 7.4.9.

[**author=wikibooks**, file =**text_files/trigonometric_subst**]
Find the integral of $\sqrt{(}x^2 - 1)/x^2$.

Note that the integrand is approximately $1/x$ for large x, so the antiderivative should be approximately $\ln x$. Using the substitution we find

$$\begin{array}{rcll} \int_1^z \frac{\sqrt{x^2-1}}{x^2}\,dx & = & \int_1^\alpha \frac{\tan\theta}{\sec^2\theta}\sec\theta\tan\theta\,d\theta & z > 1 \\ & = & \int_0^\alpha \frac{\sin^2\theta}{\cos\theta}d\theta & \alpha = \sec^{-1} z \end{array}$$

We can now integrate by parts

$$\begin{array}{rcl} \int_1^z \frac{\sqrt{x^2-1}}{x^2}\,dx & = & -\left[\tan\theta\cos\theta\right]_0^\alpha + \int_0^\alpha \sec\theta\,d\theta \\ & = & -\sin\alpha + \left[\ln(\sec\theta + \tan\theta)\right]_0^\alpha \\ & = & \ln(\sec\alpha + \tan\alpha) - \sin\alpha \\ & = & \ln(z + \sqrt{z^2-1}) - \frac{\sqrt{z^2-1}}{z} \end{array}$$

which for large z behaves like $\ln z + \ln 2 - 1$, just as expected.

Rule 7.4.5.

[**author=wikibooks**, file =**text_files/trigonometric_subst**]
When the integrand contains a factor of this form $\sqrt{a^2 + x^2}$ we can use the substitution

$$x = a\tan\theta \quad \sqrt{x^2 + a^2} = a\sec\theta \quad dx = a\sec^2\theta d\theta$$

Example 7.4.10.

[**author=wikibooks**, file =**text_files/trigonometric_subst**]
Find the integral of $(x^2 + a^2)^-3/2$.

We make the substitution:

$$\begin{array}{rcll} \int_0^z \left(x^2 + a^2\right)^{-\frac{3}{2}} dx & = & a^{-2}\int_0^\alpha \cos\theta\,d\theta & z > 0 \\ & = & a^{-2}\left[\sin\theta\right]_0^\alpha & \alpha = \tan^{-1}(z/a) \\ & = & a^{-2}\sin\alpha & \\ & = & a^{-2}\frac{z/a}{\sqrt{1+z^2/a^2}} & = \frac{1}{a^2}\frac{z}{\sqrt{a^2+z^2}} \end{array}$$

If the integral is

$$I = \int_0^z \sqrt{x^2 + a^2} \quad z > 0$$

then on making this substitution we find

$$\begin{array}{rcllll} I & = & a^2\int_0^\alpha \sec^3\theta\,d\theta & & & \alpha = \tan^{-1}(z/a) \\ & = & a^2\int_0^\alpha \sec\theta\,d\tan\theta & & & \\ & = & a^2\left[\sec\theta\tan\theta\right]_0^\alpha & - & a^2\int_0^\alpha \sec\theta\tan^2\theta\,d\theta & \\ & = & a^2\sec\alpha\tan\alpha & - & a^2\int_0^\alpha \sec^3\theta\,d\theta & +a^2\int_0^\alpha \sec\theta\,d\theta \\ & = & a^2\sec\alpha\tan\alpha & - & I & +a^2\int_0^\alpha \sec\theta\,d\theta \end{array}$$

After integrating by parts, and using trigonometric identities, we've ended up with an expression involving the original integral. In cases like this we must now rearrange the equation so that the original integral is on one side only

$$\begin{array}{rcccc} I & = & \frac{1}{2}a^2\sec\alpha\tan\alpha & + & \frac{1}{2}a^2\int_0^\alpha \sec\theta\,d\theta \\ & = & \frac{1}{2}a^2\sec\alpha\tan\alpha & + & \frac{1}{2}a^2\left[\ln\left(\sec\theta+\tan\theta\right)\right]_0^\alpha \\ & = & \frac{1}{2}a^2\sec\alpha\tan\alpha & + & \frac{1}{2}a^2\ln\left(\sec\alpha+\tan\alpha\right) \\ & = & \frac{1}{2}a^2\left(\sqrt{1+\frac{z^2}{a^2}}\right)\frac{z}{a} & + & \frac{1}{2}a^2\ln\left(\sqrt{1+\frac{z^2}{a^2}}+\frac{z}{a}\right) \\ & = & \frac{1}{2}z\sqrt{z^2+a^2} & + & \frac{1}{2}a^2\ln\left(\frac{z}{a}+\sqrt{1+\frac{z^2}{a^2}}\right) \end{array}$$

As we would expect from the integrand, this is approximately $z^2/2$ for large z.

Example 7.4.11.

[author=wikibooks, file =text_files/trigonometric_subst]
Consider the problem

$$\int \frac{1}{x^2+a^2}\,dx$$

with the substitution $x = a\tan(\theta)$, we have $dx = asec^2\theta d\theta$, so that

$$\int \frac{1}{x^2+a^2}\,dx = \frac{\arctan(x/a)}{a}$$

Exercises

1. Tell what trig substitution to use for $\int x^8\sqrt{x^2-1}\,dx$
2. Tell what trig substitution to use for $\int \sqrt{25+16x^2}\,dx$
3. Tell what trig substitution to use for $\int \sqrt{1-x^2}\,dx$
4. Tell what trig substitution to use for $\int \sqrt{9+4x^2}\,dx$
5. Tell what trig substitution to use for $\int x^9\sqrt{x^2+1}\,dx$
6. Tell what trig substitution to use for $\int x^8\sqrt{x^2-1}\,dx$

7.5 Overview of Integration

Strategy.

[`author=duckworth`, file =`text_files/integration_strategy`]
This is just an outline of the techniques we have developed, and what order to try them in:

- Familiarize yourself with a list of basic anti-derivatives, like the one in given in class or elsewhere in these notes. This does mean memorizing part of the list. The part of the list that you don't memorize you should at least recognize.
- Simplify the integral.
- Try u-substitution.
- Classify the integral according to type:
  - Trigonometric functions.
  - Rational functions.
  - Integration by parts.
  - Radicals ($\sqrt{\pm x^2 \pm a^2}$ is in 7.3, and $\sqrt[n]{ax+b}$ often reduces to a rational function and 7.4 via $u = \sqrt[n]{ax+b}$).

7.6 Improper Integrals

Rule 7.6.1.

[`author=duckworth`, file =`text_files/improper_integrals`]
The word "improper" here just means that $\int_a^b f(x)\,dx$ has one (or more) of the following: $a = \infty$, $b = -\infty$, $f(x)$ has a vertical asymptote (VA) in the interval $[a, b]$ (i.e. we have y-values approaching $\pm\infty$). To handle any of these you use limits.

- $\int_{-\infty}^b f(x),\ dx = \lim_{a\to-\infty} \int_a^b f(x)\,dx$. If we can find $F(x)$ then this equals $\lim_{a\to-\infty} F(x)\Big|_a^b$.
- $\int_a^\infty f(x),\ dx = \lim_{b\to\infty} \int_a^b f(x)\,dx$. If we can find $F(x)$ then this equals $\lim_{b\to\infty} F(x)\Big|_a^b$.
- $\int_{-\infty}^\infty = \int_{-\infty}^0 + \int_0^\infty$ where both of the integrals on the right hand side have to exist.
- If $x = c$ is a VA then $\int_a^c f(x)\,dx = \lim_{t\to c} \int_a^t f(x)\,dx$. If we can find $F(x)$ this equals $\lim_{t\to c} F(x)\Big|_a^t$.

- If $x = c$ is a VA then $\int_c^b f(x)\,dx = \lim_{t\to c} \int_t^b f(x)\,dx$. If we can find $F(x)$ then this equals $\lim_{t\to c} F(x)\Big|_t^b$.
- If $x = c$ is a VA and c is in (a, b) then $\int_a^b = \int_a^c + \int_c^b$ and both of these integrals have to exist.

Chapter 8

Taylor polynomials and series

8.1 Historical and theoretical comments: Mean Value Theorem

Discussion.

[author=garrett, file =text_files/taylor_background]
For several reasons, the traditional way that *Taylor polynomials* are taught gives the impression that the ideas are inextricably linked with issues about *infinite series*. This is not so, but every calculus book I know takes that approach. The reasons for this systematic mistake are complicated. Anyway, we will *not* make that mistake here, although we may talk about infinite series later.

Instead of following the tradition, we will immediately talk about Taylor polynomials, *without* first tiring ourselves over infinite series, and *without* fooling anyone into thinking that Taylor polynomials have the infinite series stuff as prerequisite!

The theoretical underpinning for these facts about Taylor polynomials is *The Mean Value Theorem*, which itself depends upon some fairly subtle properties of the real numbers. It asserts that, *for a function f differentiable on an interval $[a, b]$, there is a point c in the interior (a, b) of this interval so that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Note that the latter expression is the formula for the slope of the 'chord' or 'secant' line connecting the two points $(a, f(a))$ and $(b, f(b))$ on the graph of f. And the $f'(c)$ can be interpreted as the slope of the *tangent* line to the curve at the point $(c, f(c))$.

In many traditional scenarios a person is expected to commit the statement of the Mean Value Theorem to memory. And be able to respond to issues like 'Find a point c in the interval $[0, 1]$ satisfying the conclusion of the Mean Value Theorem for the function $f(x) = x^2$.' This is pointless and we won't do it.

Discussion.

[**author=duckworth**, file =**text_files/taylor_background**]
We start by looking at approximating a function using polynomials. To make the approximation more accurate we usually have to use more and more terms in the polynomial. This leads to "infinite" polynomials (which we always approximate with finite ones). We need tests to measure how accurately the approximation holds, and which numbers it even makes sense to plug in. For any of this to make sense, you should do about a hundred examples.

8.2 Taylor polynomials: formulas

Discussion.

[**author=garrett**, file =**text_files/taylor_poly_formula**]
Before attempting to illustrate what these funny formulas can be used for, we just write them out. First, some reminders:

The notation $f^{(k)}$ means the kth derivative of f. The notation $k!$ means k-*factorial*, which by definition is

$$k! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots \cdot (k-1) \cdot k$$

Taylor's Formula with Remainder 8.2.1.

[**author=garrett**, file =**text_files/taylor_poly_formula**]
First somewhat verbal version: Let f be a reasonable function, and fix a positive integer n. Then we have

$$\begin{aligned} rcl f(\text{input}) &= f(\text{basepoint}) + \frac{f'(\text{basepoint})}{1!}(\text{input} - \text{basepoint}) \\ &+ \frac{f''(\text{basepoint})}{2!}(\text{input} - \text{basepoint})^2 \\ &+ \frac{f'''(\text{basepoint})}{3!}(\text{input} - \text{basepoint})^3 \\ &+ \ldots \\ &+ \frac{f^{(n)}(\text{basepoint})}{n!}(\text{input} - \text{basepoint})^n \\ &+ \frac{f^{(n+1)}(c)}{(n+1)!}(\text{input} - \text{basepoint})^{n+1} \end{aligned}$$

for some c between *basepoint* and *input*.

That is, the value of the function f for some *input* presumably 'near' the *basepoint* is expressible in terms of the values of f and its derivatives *evaluated at the basepoint*, with the only mystery being the precise nature of that c between *input* and *basepoint*.

Taylor's Formula with Remainder Term 8.2.2.

[`author=garrett`, file =`text_files/taylor_poly_formula`]
Second somewhat verbal version: Let f be a reasonable function, and fix a positive integer n.

$$\begin{aligned} f(\text{basepoint} + \text{increment}) &= f(\text{basepoint}) + \frac{f'(\text{basepoint})}{1!}(\text{increment}) \\ &+ \frac{f''(\text{basepoint})}{2!}(\text{increment})^2 \\ &+ \frac{f'''(\text{basepoint})}{3!}(\text{increment})^3 \\ &+ \dots \\ &+ \frac{f^{(n)}(\text{basepoint})}{n!}(\text{increment})^n \\ &+ \frac{f^{(n+1)}(c)}{(n+1)!}(\text{increment})^{n+1} \end{aligned}$$

for some c between *basepoint* and *basepoint* + *increment*.

This version is really the same as the previous, but with a different emphasis: here we still have a *basepoint*, but are thinking in terms of moving a little bit away from it, by the amount *increment*.

Taylors Formula with remainder 8.2.3.

[`author=garrett`, file =`text_files/taylor_poly_formula`]
And to get a more compact formula, we can be more symbolic: let's repeat these two versions:

Let f be a reasonable function, fix an input value x_o, and fix a positive integer n. Then for input x we have

$$\begin{aligned} f(x) = f(x_o) + \frac{f'(x_o)}{1!}(x - x_o) + \frac{f''(x_o)}{2!}(x - x_o)^2 + \frac{f'''(x_o)}{3!}(x - x_o)^3 + \dots \\ + \frac{f^{(n)}(x_o)}{n!}(x - x_o)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_o)^{n+1} \end{aligned}$$

for some c between x_o and x.

Comment.

[`author=garrett`, file =`text_files/taylor_poly_formula`]
Note that in every version, in the very last term where all the indices are $n + 1$, the input into $f^{(n+1)}$ is *not* the basepoint x_o but is, instead, that mysterious c about which we truly know nothing but that it lies between x_o and x. The part of this formula *without* the error term is the **degree-n Taylor polynomial for f at x_o**, and that last term is the **error term** or **remainder term**. The Taylor series is said to be **expanded at** or **expanded about** or **centered at** or simply **at** the basepoint x_o.

Comment.

[`author=garrett`, file =`text_files/taylor_poly_formula`]
There are many other possible forms for the error/remainder term. The one here

was chosen partly because it resembles the other terms in the main part of the expansion.

Linear Taylor Polynomial with Remainder Term 8.2.4.

[`author=garrett`, file =`text_files/taylor_poly_formula`]
Let f be a reasonable function, fix an input value x_o. For any (reasonable) input value x we have

$$f(x) = f(x_o) + \frac{f'(x_o)}{1!}(x - x_o) + \frac{f''(c)}{2!}(x - x_o)^2$$

for some c between x_o and x.

Comment.

[`author=garrett`, file =`text_files/taylor_poly_formula`]
The previous formula is of course a very special case of the first, more general, formula. The reason to include the 'linear' case is that *without* the error term it is the old *approximation by differentials* formula, which had the fundamental flaw of having no way to estimate the error. Now we *have* the error estimate.

Comment.

[`author=garrett`, file =`text_files/taylor_poly_formula`]
The general idea here is to approximate 'fancy' functions by polynomials, especially if we restrict ourselves to a fairly small interval around some given point. (That 'approximation by differentials' circus was a very crude version of this idea).

It is at this point that it becomes relatively easy to 'beat' a calculator, in the sense that the methods here can be used to give whatever precision is desired. So at the very least this methodology is not as silly and obsolete as some earlier traditional examples.

But even so, there is more to this than getting numbers out: it ought to be of some intrinsic interest that pretty arbitrary functions can be approximated as well as desired by polynomials, which are so readily computable (by hand *or* by machine)!

One element under our control is choice of *how high degree polynomial to use*. Typically, the higher the degree (meaning more terms), the better the approximation will be. (There is nothing comparable to this in the 'approximation by differentials').

Of course, for all this to really be worth anything either in theory or in practice, we do need a tangible *error estimate*, so that we can be sure that we are within whatever tolerance/error is required. (There is nothing comparable to this in the 'approximation by differentials', either).

And at this point it is not at all clear what exactly can be done with such formulas. For one thing, there are choices.

Notation.

[author=duckworth, file =text_files/taylor_poly_formula]
Recall the notation $f^{(k)}$ means the kth derivative of f. Recall the definition of $n!$: $0! = 1$, $1! = 1$, $2! = 2$, $3! = 3 \cdot 2$, $4! = 4 \cdot 3 \cdot 2$ and in general $k! = k(k-1) \cdots 3 \cdot 2$.

Theorem 8.2.1.
[author=duckworth, file =text_files/taylor_poly_formula]
If $f(x)$ is a nice function near $x = 0$, then $f(x)$ may be approximated by the following degree n polynomial

$$f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n$$

In other words, the coefficient of x^k is $\frac{f^{(k)}(0)}{k!}$.

Comment.

[author=duckworth, file =text_files/taylor_poly_formula]
The polynomial in Theorem 8.2.1 is called the Maclaurin or Taylor polynomial.

Comment.

[author=duckworth, file =text_files/taylor_poly_formula]
You might want to think about the following questions when you look at this theorem (we will persue these questions later):

1. What does “nice” mean?
2. Can we replace “near $x = 0$” with some other number?
3. How good an approximation is it?
4. How do we make the approximation better?
5. Does it make sense to use infinitely many terms in the polynomial?
6. Can we prove that our answers to any of the questions are correct?

Comment.

[author=duckworth, file =text_files/taylor_poly_formula]
What does “nice” mean in Theorem 8.2.1? It means that f has as many derivatives as we want, all continuous, on some open interval containing $x = 0$.

Example 8.2.1.

[author=duckworth, file =text_files/taylor_poly_formula]
Let’s find the Maclaurin polynomial for $f(x) = \sin(x)$. For the above recipe we need to calculate $f^{(k)}(x)$, i.e. a bunch of derivatives, and we need to calculate

$f^{(k)}(0)$, i.e. evaluate these derivatives at $x = 0$. We calculate:

$$\begin{array}{rclrcl} f(x) & = & \sin(x) & f(0) & = & 0 \\ f'(x) & = & \cos(x) & f'(0) & = & 1 \\ f''(x) & = & -\sin(x) & f''(0) & = & 0 \\ f'''(x) & = & -\cos(x) & f'''(0) & = & -1 \\ & & \text{(After this it repeats)} & & & \\ f^{(4)} & = & \sin(x) & f^{(4)} & = & 0 \\ f^{(5)} & = & \cos(x) & f^{(5)} & = & 1 \\ \vdots & & & \vdots & & \end{array}$$

Thus we have

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \dots$$

We'll worry about how to write the "last" term later, and we'll worry about Σ notation later. Note that only odd terms remain in this polynomial. That's because $\sin(x)$ is an odd function.

Just to see how good this approximation is, let's take a look at some graphs. Let's graph $y_1 = \sin(x)$, $y_2 = x - \frac{1}{3!}x^3$, $y_3 = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5$ and $y_4 = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7$.

Comment.

[**author=duckworth**, file =**text_files/taylor_poly_formula**]
What about the "last" term in Example 8.2.1? Judging from the above pattern we know it will be odd. We can write any odd number as $2n+1$, so the last term will be of the form $\pm\frac{1}{(2n+1)!}x^{2n+1}$. But that's not very satisfying, is it "+" or is it "−"? Well, that alternates. The first term is positive, the next is negative, the next positive, etc. So we want a formula that alternates like this between positive and negative. The most common formula for this is $(-1)^n$. Thus, including the last term we have:

$$\text{Maclaurin poly for } \sin(x) \text{ is} \quad x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \dots + \frac{(-1)^n}{(2n+1)!}x^{2n+1}$$

Example 8.2.2.

[**author=duckworth**, file =**text_files/taylor_poly_formula**]
Let's find the Maclaurin polynomial for $f(x) = \cos(x)$. We calculate:

$$\begin{array}{rclrcl} f(x) & = & \cos(x) & f(0) & = & 1 \\ f'(x) & = & -\sin(x) & f'(0) & = & 0 \\ f''(x) & = & -\cos(x) & f''(0) & = & -1 \\ f'''(x) & = & \sin(x) & f'''(0) & = & 0 \\ & & \text{(After this it repeats)} & & & \\ f^{(4)} & = & \cos(x) & f^{(4)} & = & 1 \\ f^{(5)} & = & -\sin(x) & f^{(5)} & = & 0 \\ \vdots & & & \vdots & & \end{array}$$

Thus we have

$$\cos(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \dots$$

Notice that only even terms appear in this polynomial. That's because $\cos(x)$ is an even function.

Example 8.2.3.

[author=duckworth, file =text_files/taylor_poly_formula]
Let's find the Maclaurin polynomial for $f(x) = e^x$. We calculate:

$$\begin{array}{rclrcl} f(x) & = & e^x & f(0) & = & 1 \\ f'(x) & = & e^x & f'(0) & = & 1 \\ f''(x) & = & e^x & f''(0) & = & 1 \\ & & \text{(After this it repeats)} & & & \\ & \vdots & & & \vdots & \end{array}$$

Thus we have

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots$$

Discussion.

[author=duckworth, file =text_files/taylor_poly_formula]
For the next example we need to change $x = 0$ to $x = 1$ (that's because the next function is not defined at $x = 0$). In general we can replace $x = 0$ with $x = a$, but of course we need to change the recipe in Theorem 8.2.1.

Theorem 8.2.2.
[author=duckworth, file =text_files/taylor_poly_formula]
If $f(x)$ is a nice function near $x = a$, then $f(x)$ may be approximated by the following polynomial:

$$f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

In other words the coefficient of $(x-a)^k$ is $\frac{f^{(k)}(a)}{k!}$.

Comment.

[author=duckworth, file =text_files/taylor_poly_formula]
The polynomial in Theorem **??** is called the **Taylor polynomial of** $f(x)$ **at** $x = a$. People also say that the polynomial is **defined at** $x = a$ or **centered at** $x = a$ or that a is the **center** of the polynomial.

Note that for $a = 0$ this formula is identical to the formula for the Maclaurin polynomial. In other words, the Maclaurin polynomial is just a special case of the Taylor polynomial. However, this "special" case is the one which we will see most often.

Example 8.2.4.

[author=duckworth, file =text_files/taylor_poly_formula]

Find the Taylor polynomial at $x = 1$ for $f(x) = 1/x$ (I chose $x = 1$ because 1 is halfway between 0 and ∞). Since we'll be taking lots of derivatives let's rewrite $1/x$ as x^{-1}.

$$\begin{array}{rclrcl} f(x) & = & x^{-1} & f(1) & = & 1 \\ f'(x) & = & -x^{-2} & f(1) & = & -1 \\ f''(x) & = & 2x^{-3} & f(1) & = & 2 \\ f'''(x) & = & -6x^{-4} = -3!\,x^{-4} & f(1) & = & -3! \\ f^{(4)}(x) & = & 24x^{-5} = 4!\,x^{-5} & f(1) & = & 4! \end{array}$$

Thus we have

$$1/x \approx 1 - (x-1) + (x-1)^2 - (x-1)^3 + (x-1)^4 - \ldots$$

Exercises

1. Write the first three terms of the Taylor series *at 0* of $f(x) = 1/(1+x)$.
2. Write the first three terms of the Taylor series *at 2* of $f(x) = 1/(1-x)$.
3. Write the first three terms of the Taylor series *at 0* of $f(x) = e^{\cos x}$.

8.3 Classic examples of Taylor polynomials

Examples 8.3.1.

[author=garrett, file =text_files/taylor_examples]
Some of the most famous (and important) examples are the expansions of $\frac{1}{1-x}$, e^x, $\cos x$, $\sin x$, and $\log(1+x)$ at 0: right from the formula, although simplifying a little, we get

1. $\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + \ldots$
2. $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots$
3. $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \ldots$
4. $\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$
5. $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \frac{x^6}{6} + \ldots$

where here the *dots* mean to *continue to whatever term you want, then stop, and stick on the appropriate remainder term.*

It is entirely reasonable if you can't really see that these are what you'd get, but in any case you should do the computations to verify that these are right. It's not so hard.

Note that the expansion for cosine has no *odd* powers of x (meaning that the coefficients are *zero*), while the expansion for sine has no *even* powers of x (meaning that the coefficients are *zero*).

Comment.

[author=garrett, file =text_files/taylor_examples]
At this point it is worth repeating that we are *not* talking about *infinite* sums (series) at all here, although we do allow arbitrarily large *finite* sums. Rather than worry over an infinite sum that we can never truly evaluate, we use the *error* or *remainder* term instead. Thus, while in other contexts the dots *would* mean 'infinite sum', that's not our concern here.

The first of these formulas you might recognize as being a *geometric series*, or at least a part of one. The other three patterns might be new to you. A person would want to be learn to recognize these on sight, as if by reflex!

8.4 Computational tricks regarding Taylor polynomials

Discussion.

[author=garrett, file =text_files/taylor_calculation_tricks]
The obvious question to ask about Taylor polynomials is 'What are the first so-many terms in the Taylor polynomial of some function expanded at some point?'.

The most straightforward way to deal with this is just to do what is indicated by the formula: take however high order derivatives you need and plug in. However, very often this is not at all the most efficient.

Especially in a situation where we are interested in a composite function of the form $f(x^n)$ or, more generally, $f(\text{polynomial in } x)$ with a 'familiar' function f, there are alternatives.

Example 8.4.1.

[**author=garrett**, file =**text_files/taylor_calculation_tricks**]
For example, looking at $f(x) = e^{x^3}$, if we start taking derivatives to expand this at 0, there will be a big mess pretty fast. On the other hand, we might start with the 'familiar' expansion for e^x

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{e^c}{4!}x^4$$

with some c between 0 and x, where our choice to cut it off after that many terms was simply a whim. But then replacing x by x^3 gives

$$e^{x^3} = 1 + x^3 + \frac{x^6}{2!} + \frac{x^9}{3!} + \frac{e^c}{4!}x^{12}$$

with some c between 0 and x^3. Yes, we need to keep track of c in relation to the *new* x.

So we get a polynomial plus that funny term with the 'c' in it, for the remainder. Yes, this gives us a different-looking error term, but that's fine.

So we obtain, with relative ease, the expansion of degree *eleven* of this function, which would have been horrible to obtain by repeated differentiation and direct application of the general formula. Why 'eleven'?: well, the error term has the x^{12} in it, which means that the polynomial itself stopped with a x^{11} term. Why didn't we see that term? Well, evidently the coefficients of x^{11}, and of x^{10} (not to mention $x, x^2, x^4, x^5, x^7, x^8$!) are *zero*.

Example 8.4.2.

[**author=garrett**, file =**text_files/taylor_calculation_tricks**]
As another example, let's get the degree-eight expansion of $\cos x^2$ at 0. Of course, it makes sense to use

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{-\sin c}{5!}x^5$$

with c between 0 and x, where we note that $-\sin x$ is the fifth derivative of $\cos x$. Replacing x by x^2, this becomes

$$\cos x^2 = 1 - \frac{x^4}{2!} + \frac{x^8}{4!} + \frac{-\sin c}{5!}x^{10}$$

where now we say that c is between 0 and x^2.

Exercises

1. Use a shortcut to compute the Taylor expansion at 0 of $\cos(x^5)$.
2. Use a shortcut to compute the Taylor expansion at 0 of $e^{(x^2+x)}$.
3. Use a shortcut to compute the Taylor expansion at 0 of $\log(\frac{1}{1-x})$.

8.5 Getting new Taylor polynomials from old

Discussion.

[author=duckworth, file =text_files/new_taylor_series_from_old]
For our next example, we need to know how to take an old example and get a new one. In other words, there are two ways to figure out a Taylor polynomial: (1) take lots of derivatives and use the recipe given above (2) take a Taylor polynomial for some other function, and change it to make a new function. We make this idea more precise in the following theorem.

Theorem 8.5.1.
[author=duckworth, file =text_files/new_taylor_series_from_old]
Suppose $f(x) \approx c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \dots$. Then we may find polynomial approximations for a function $g(x)$ as follows:

1. *If $g(x) = f(x^2)$ (or $f(2x)$, or $f(-x^2)$ or ...) the polynomial for $g(x)$ is found by substituting x^2 (or $2x$, or $-x^2$ or ...) in place of x in the polynomial for $f(x)$.*
2. *If $g(x)$ is the anti-derivative of $f(x)$ (or the derivative) then the polynomial for $g(x)$ is found by taking the anti-derivative (or the derivative) of the polynomial for $f(x)$.*
3. *If $g(x)$ equals $f(x)$ times x (or $\sin(x)$, or e^x, or ..., or divided by one of these) then the polynomial for $g(x)$ is found by multipliying by x (or by the polynomial for $\sin(x)$, or the polynomial for $\sin(x)$ or by dividing by one of these).*

Example 8.5.1.

[author=duckworth, file =text_files/new_taylor_series_from_old]
Find the Maclaurin polynomial for $\sin(x^2)$.

Note, if we tried to do this in the same way as our other examples it would be difficult. For the first derivative we'd need the chain rule, after that we'd need the product rule and we'd get two terms. After that we'd need the product rule again and we'd need three terms, after that it just keeps getting worse.

Let's start with $\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$. Note that if we plug in 289 we have $\sin(289) = 289 - \frac{1}{3!}289^3 + \frac{1}{5!}289^5 - \dots$. But, if you happen to notice that $289 = 17^2$ then you could write $\sin(17^2) = 17^2 - \frac{1}{3!}(17^2)^3 + \frac{1}{5!}(17^2)^5 - \dots$. Replacing 17 with x and you see that

$$\sin(x^2) = x^2 - \frac{1}{3!}x^6 + \frac{1}{5!}x^5 - \dots$$

Example 8.5.2.

[author=duckworth, file =text_files/new_taylor_series_from_old]
Find the Taylor polynomial at $x = 1$ for $\ln(x)$.

This problem we could do by taking lots of derivatives, but it's easier to do it by starting with an example we already know. Let's start with $1/x$ and take the anti-derivative.

$$\begin{aligned}\ln(x) &= \int \tfrac{1}{x}\,dx + C\\ &= \int 1-(x-1)+(x-1)^2-(x-1)^3+\dots\,dx + C\\ &= x - \tfrac{(x-1)^2}{2} + \tfrac{(x-1)^3}{3} - \tfrac{(x-1)^4}{4} + \cdots + C\end{aligned}$$

But what's C? Well, we know that we should have $\ln(1) = 0$. Plugging this in we get

$$1 = 1 - 0 + 0 - 0 + \cdots + C$$

so $C = -1$ and we can write

$$\ln(x) = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \dots$$

Example 8.5.3.

[`author=duckworth`, file =`text_files/new_taylor_series_from_old`]
Find the Maclaurin polynomial for $\tan^{-1}(x)$.

Again, I want to start with an example we already know. If I think about derivatives and anti-derivatives, I see that $\tan^{-1}$ is the anti-derivative of $\frac{1}{1+x^2}$. So, we'd have a plan if we knew the polynomial for $\frac{1}{1+x^2}$. Well, $\frac{1}{1+x^2}$ looks like $1/x$ where we've replaced x with $1+x^2$. So my plan is this: take the polynomial for $1/x$, substitute $1+x^2$ into this, then take the antiderivative:

$$\begin{aligned}\tfrac{1}{x} &= 1-(x-1)+(x-1)^2-(x-1)^3+(x-1)^4-\dots\\ \downarrow &\\ \tfrac{1}{1+x^2} &= 1-(1+x^2-1)+(1+x^2-1)^2-(1+x^2-1)^3+(1+x^2-1)^4-\dots\\ &= 1-x^2+x^4-x^6+x^8\\ \downarrow &\\ \tan^{-1}(x) &= \int 1-x^2+x^4-x^6+x^8-\cdots+C\\ &= x-\tfrac{x^3}{3}+\tfrac{x^5}{5}-\tfrac{x^7}{7}-\cdots+C\end{aligned}$$

Again you can find C by plugging in $\tan^{-1}(0) = 0$. In this case you find that $C = 0$, thus:

$$\tan^{-1}(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} - \dots$$

Example 8.5.4.

[`author=duckworth`, file =`text_files/new_taylor_series_from_old`]
Find the Maclaurin series for $e^x\sin(x)$. Actually, just find the first few terms. The idea here is just that you muliply the polynomials for e^x and $\sin(x)$. So we have

$$e^x\sin(x) = (1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots)(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots)$$

Not everyone knows how to multiply things like this together. If you apply the distributive law over and over again the result is this: pick a term on the left, multiply it by each term on the right, then move to the next term on the left. Thus we get:

$$\begin{aligned} e^x \sin(x) = \quad & 1 \cdot x - 1 \cdot \tfrac{x^3}{3!} + 1 \tfrac{x^5}{5!} - \ldots && \text{(1 times the polynomial on the right)} \\ & + x \cdot x - x \cdot \tfrac{x^3}{3!} + x \cdot \tfrac{x^5}{5!} - \ldots && \text{(x times the polynomial on the right)} \\ & + \tfrac{x^2}{2} \cdot x - \tfrac{x^2}{2} \cdot \tfrac{x^3}{3!} + \tfrac{x^2}{2} \cdot \tfrac{x^5}{5!} - \ldots && \text{($\tfrac{x^2}{2}$ times the polynomial on the right)} \\ & + \tfrac{x^3}{3!} \cdot x - \tfrac{x^3}{3!} \cdot \tfrac{x^3}{3!} + \tfrac{x^3}{3!} \cdot \tfrac{x^5}{5!} - \ldots && \text{($\tfrac{x^3}{3!}$ times the polynomial on the right)} \end{aligned}$$

Now one simplifies this

$$\begin{aligned} = \quad & x - \tfrac{x^3}{3!} + \tfrac{x^5}{5!} - \ldots \\ & + x^2 - \tfrac{x^4}{3!} + \tfrac{x^6}{5!} - \ldots \\ & + \tfrac{x^3}{2} - \tfrac{x^5}{2 \cdot 3!} + \tfrac{x^7}{2 \cdot 5!} - \ldots \\ & + \tfrac{x^4}{3!} - \tfrac{x^6}{3! \cdot 3!} + \tfrac{x^8}{3! \cdot 5!} - \ldots \end{aligned}$$

Now one collects the constant terms in front, then all the x terms, then all the x^2 terms, etc.

$$= x + x^2 + \left(-\frac{1}{3!} + \frac{1}{2} \right) x^3 + \left(\frac{2}{3!} \right) x^4 + \left(\frac{1}{5!} - \frac{1}{2 \cdot 3!} \right) x^5 + \ldots$$

Note that here there is not a clear pattern as to what the next term would look like.

8.6 Prototypes: More serious questions about Taylor polynomials

Discussion.

[`author=garrett`, file =`text_files/taylor_questions`]
Beyond just writing out Taylor expansions, we could actually use them to approximate things in a more serious way. There are roughly three different sorts of *serious* questions that one can ask in this context. They all use similar words, so a careful reading of such questions is necessary to be sure of answering the question asked.

(The word 'tolerance' is a synonym for 'error estimate', meaning that we know that the error is *no worse* than such-and-such)

Here are the big questions:

1. Given a Taylor polynomial approximation to a function, expanded at some given point, and given an interval around that given point, *within what tolerance* does the Taylor polynomial approximate the function on that interval?

2. Given a Taylor polynomial approximation to a function, expanded at some given point, and given a required tolerance, *on how large an interval* around the given point does the Taylor polynomial achieve that tolerance?
3. Given a function, given a fixed point, given an interval around that fixed point, and given a required tolerance, find *how many terms* must be used in the Taylor expansion to approximate the function to within the required tolerance on the given interval.

As a special case of the last question, we can consider the question of *approximating* $f(x)$ *to within a given tolerance/error in terms of* $f(x_o), f'(x_o), f''(x_o)$ *and higher derivatives of* f *evaluated at a given point* x_o.

In 'real life' this last question is not really so important as the third of the questions listed above, since evaluation at just one point can often be achieved more simply by some other means. Having a polynomial approximation that works *all along an interval* is a much more substantive thing than evaluation at a single point.

It must be noted that there are also *other* ways to approach the issue of *best approximation by a polynomial on an interval*. And beyond worry over approximating the *values* of the function, we might also want the values of one or more of the *derivatives* to be close, as well. The theory of **splines** is one approach to approximation which is very important in practical applications.

Discussion.

[`author=duckworth`, file =`text_files/taylor_questions`]
Question: How good is the approximation? What happens as we use more terms? We saw some of this answer in the graphs of $\sin(x)$ and its Maclaurin polynomial; here we make it more precise. We start by giving an exact meaning to this question.

Definition 8.6.1.

[`author=duckworth`, file =`text_files/taylor_questions`]
Let $f(x)$ be a nice function and let $c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$ be it's degree n Maclaurin polynomial. The degree n **remainder**, or **error**, is defined as

$$R_n(x) = f(x) - (c_0 + c_1x + c_2x^2 + \ldots c_nx^n).$$

In other words, $R_n(x)$ is the gap between the original function f and the polynomial.

Comment.

[`author=ducworth`, file =`text_files/taylor_questions`]
Now we need a formula which tells us something about $R_n(x)$. Ideally, we can can use this formula to say how big $R_n(x)$ is, and maybe even show that $R_n(x)$ goes to 0 as we use more and more terms in the Maclaurin polynomial (i.e. $\lim_{n\to\infty} R_n(x) = 0$).

Maclaurin remainder theorem 8.6.1.
[author=duckworth, file =`text_files/taylor_questions`]
Let $R_n(x)$ be the degree n remainder of the Maclaurin approximation of $f(x)$. Let $[a, b]$ be some interval containing 0 and let M be some number such that $|f^{(n+1)}(x)| \leq M$ on the interval $[a, b]$. Then

$$|R_n(x)| \leq \frac{M}{(n+1)!}|x|^{n+1}$$

for all x in the interval $[-a, a]$.

Comments.

[`author=duckworth`, file =`text_files/taylor_questions`]

1. Note, usually we will find M by finding the absolute max and min of $f^{(n+1)}(x)$ on the interval $[a, b]$. Sometimes, however, we can find a value for M without calculating absolute max's and min's. For example, if $f(x)$ equals $\sin(x)$, then we can always take $M = 1$.
2. Note that this theorem gives some impression of why Maclaurin approximations get better by using more terms. As n gets bigger, the fraction $\frac{M}{(n+1)!}$ will almost always get smaller. Why? Because $(n+1)!$ get's really big. O.K., so does $\frac{M}{(n+1)!}$ always get smaller? Well, be to be rigorous, M might change with n. Off the top of my head, I can't think of a function where M would change enough to prevent $\frac{M}{(n+1)!}$ from getting smaller, but I believe such a function exists.
3. Note, we are often interested in bounding $R_n(x)$ on an interval; in such a case we replace $|x|^{n+1}$ by its absolute max on the interval. In other words, if the interval is $[a, b]$, we'll replace $|x|^{n+1}$ by $|a|^{n+1}$ or b^{n+1}, whichever is bigger.

Example 8.6.1.

[`author=duckworth`, file =`text_files/taylor_questions`]
Consider the Maclaurin polynomial for $\sin(x)$.

(a) Find an upper bound for the error of approximating $\sin(.5)$ using the degree three Maclaurin polynomial.

(b) Find n so that the error would be at most .00001.

Solution: (a) By the work above the degree three polynomial is $x - \frac{1}{3!}x^3$. Thus the error is

$$R_3(.5) \leq \frac{M}{4!}(.5)^4$$

where M is an upper bound on the fourth derivative of $\sin(x)$. You should always remember, for $\sin(x)$ and $\cos(x)$, you can always take 1 as an upper bound on any derivative. So, let $M = 1$. Then we have

$$R_3(.5) \leq \frac{1}{4!}(.5)^4 = .0026$$

To make this more concrete, let's calculate the approximation we're discussing (note that so far in this problem, we've calculated the error without ever knowing what the approximation is; this is kind of strange). The approximation is

$$\sin(.5) \approx .5 - \frac{(.5)^3}{3!} = .4792$$

and our calculation above says that the "real" number is within .0026 of this.

(b) Now we don't know n, but we use the same value for M. So we want:

$$\frac{1}{(n+1)!}(.5)^{n+1} \leq .00001$$

Now, the truth is, I don't know how to solve this for n. So, I'll just guess and check. Note, I'll just guess odd values for n since there are no even terms in $\sin(x)$.

$$n = 5 \Rightarrow \quad \tfrac{1}{6!}(.5)^6 = .000022$$

$$n = 7 \Rightarrow \quad \tfrac{1}{8!}(.5)^8 = .968 \times 10^{-7}$$

Thus $n = 7$ gives an error which is quite small.

Comment.

[author=duckworth, file =text_files/taylor_questions]
By the way, Maclaurin polynomials (or possibly a refined version of them) are how your calculator really finds different values of $\sin(x)$. It doesn't have a big table of all possible values of $\sin(x)$. Instead, it first reduces an angle x, to a value between 0 and $\pi/2$, and then uses a polynomial formula.

Actually, it's probably even smarter than this, and it's an interesting topic of how to use these formulas in the most efficient way possible. Why do people care about efficiency? Well, suppose you're graphing $\sin(x^2 + x)$. This involves calculating a y-value for each pixel on the calculator (or computer) screen. If each pixel required using 30 terms in the Maclaurin polynomial, that would take a long time to graph. So, if you can reduce any angle to one near 0, you don't need very many terms at all to approximate sine of that angle. For example, if I wanted to calculate $\sin(25.23274123)$ that might take a lot of terms, but I know that $\sin(x) = \sin(x - 2\pi)$, so I can subtract 2π first from 25.23274123. Well, $25.23274123 - 2\pi = 18.94955592$, and I can subtract 2π again, and again, etc. Note that $25.23274123 - 8\pi \approx .1$ and that $\sin(.1)$ can be approximated with very few terms. My computer says that $\sin(.1) \approx 0.09983341665$. You can get $\sin(.1) \approx 0.09983341666$ by using the first three nonzero terms of the Maclaurin polynomial (i.e. up to degree 5).

Well, the trick with subtracting 2π works well for angles that end up near $x = 0$, but what if you start with an angle like $x = 3.241592654$. This is near π and subtracting 2π won't help. Well, here you use the fact that $\sin(x) = -\sin(x - \pi)$, and this means $\sin(3.241592654) = -\sin(.1)$, and of course we already know how to calculate $\sin(.1)$.

The tricks I've just described would reduce calculations of $\sin(x)$, for all values of x to calculations involving only x between $-\pi/2$ and $\pi/2$. But maybe this is still too big of a range. Maybe calculation of some angle like $x = 1.47$ would take a long time. My computer says that $\sin(1.47) \approx 0.9949243498$. Suppose I need to get 10 places of accuracy (which is actually a little less than what your calculators have). Then I would need the first 8 nonzero terms of the sine polynomial (i.e. up to degree 15). That's a lot more calculation than before, especially since this would been raising something to the 15th power. So is there another shortcut?

Sure, I can think of another approache, and allthought I'm sure that the calculator uses something vaguely like this, it probably is much more sophisticated (i.e. efficient, but complicated) than what I'm presenting here. The goal is to reduce 1.47 to something closer to zero. One trick would be to use the identity $\sin(x) = \cos(x - \pi/2)$. Then $\sin(1.47) = \cos(-0.100796327)$ and now I would use the cosine polynomial, probably with only a few terms since what I'm plugging in is close to zero. If I use the first four nonzero terms of the Maclaurin polynomial I find that $\cos(-0.100796327) \approx 0.9949243497$.

Using all of the above tricks would reduce calculations of $\sin(x)$, and $\cos(x)$, for all values of x to calculations involving only x between $-\pi/4$ and $\pi/4$.

What if this still isn't good enough? What if calculating $\sin(.78)$ takes too long? Note that .78 is near $\pi/4$ and can't be made any closer to zero by subtracting π, or 2π, or $\pi/2$ etc. Well, then you could use other identities. Remember, in trigonometry, there are a million identities! So, you could use the half angle identity: $\sin(x) = 2\sin(x/2)\cos(x/2)$. So, you could calculate $\sin(.78/2)$ and $\cos(.78/2)$ using Maclaurin series, and then multiply these together and multiply by 2.

Well, you get the idea. If time matters (which it usually does) and if calculations take time (which they always do) and if you're doing lots of calculations (which is probably the case in most interesting problems) then it's worth your time to optimize the process using whatever tricks you can. The tricks I've shown you here are "naive" in the sense that they didn't use anything more than basic trigonometry. In real life, there are whole books and classes full tricks to speed calculations. This topic is part of *numerical analysis* and *numerical recipes*.

8.7 Determining Tolerance/Error

Discussion.

[`author=garrett`, file =`text_files/taylor_error`]
This section treats a simple example of the second kind of question mentioned above: 'Given a Taylor polynomial approximation to a function, expanded at some given point, and given an interval around that given point, *within what tolerance* does the Taylor polynomial approximate the function on that interval?'

Example 8.7.1.

[`author=garrett`, file =`text_files/taylor_error`]
Let's look at the approximation $1 - \frac{x^2}{2} + \frac{x^4}{4!}$ to $f(x) = \cos^x$ on the interval $[-\frac{1}{2}, \frac{1}{2}]$. We might ask *'Within what tolerance does this polynomial approximate* $\cos x$ *on that interval?'*

To answer this, we first recall that the error term we have after those first (oh-so-familiar) terms of the expansion of cosine is

$$\frac{-\sin c}{5!}x^5$$

For x in the indicated interval, we want to know the *worst-case scenario* for the size of this thing. A sloppy but good and simple *estimate* on $\sin c$ is that $|\sin c| \leq 1$, regardless of what c is. This is a very happy kind of estimate because it's not so bad and because it doesn't depend at all upon x. And the biggest that x^5 can be is $(\frac{1}{2})^5 \approx 0.03$. Then the *error is estimated as*

$$|\frac{-\sin c}{5!}x^5| \leq \frac{1}{2^5 \cdot 5!} \leq 0.0003$$

This is not so bad at all!

We could have been a little clever here, taking advantage of the fact that a lot of the terms in the Taylor expansion of cosine at 0 are already zero. In particular, we could *choose* to view the original polynomial $1 - \frac{x^2}{2} + \frac{x^4}{4!}$ as *including* the *fifth-degree* term of the Taylor expansion as well, which simply happens to be zero, so is invisible. Thus, instead of using the remainder term with the '5' in it, we are actually entitled to use the remainder term with a '6'. This typically will give a better outcome.

That is, instead of the remainder we had must above, we would have an error term

$$\frac{-\cos c}{6!}x^6$$

Again, in the *worst-case scenario* $|-\cos c| \leq 1$. And still $|x| \leq \frac{1}{2}$, so we have the *error estimate*

$$|\frac{-\cos c}{6!}x^6| \leq \frac{1}{2^6 \cdot 6!} \leq 0.000022$$

This is less than a tenth as much as in the first version.

But what happened here? Are there two different answers to the question of how well that polynomial approximates the cosine function on that interval? Of course not. Rather, there were two *approaches* taken by us to *estimate* how well it approximates cosine. In fact, we still do not know the *exact* error!

The point is that the second estimate (being a little wiser) is *closer* to the truth than the first. The first estimate is *true*, but is a *weaker* assertion than we are able to make if we try a little harder.

This already illustrates the point that 'in real life' there is often no single 'right' or 'best' estimate of an error, in the sense that the estimates that we can obtain by practical procedures may not be perfect, but represent a trade-off between time, effort, cost, and other priorities.

Exercises

1. How well (meaning 'within what tolerance') does $1 - x^2/2 + x^4/24 - x^6/720$ approximate $\cos x$ on the interval $[-0.1, 0.1]$?

2. How well (meaning 'within what tolerance') does $1 - x^2/2 + x^4/24 - x^6/720$ approximate $\cos x$ on the interval $[-1, 1]$?

3. How well (meaning 'within what tolerance') does $1 - x^2/2 + x^4/24 - x^6/720$ approximate $\cos x$ on the interval $[\frac{-\pi}{2}, \frac{\pi}{2}]$?

8.8 How large an interval with given tolerance?

Discussion.

[author=garrett, file =text_files/taylor_interval_size]
This section treats a simple example of the first kind of question mentioned above: 'Given a Taylor polynomial approximation to a function, expanded at some given point, and given a required tolerance, *on how large an interval* around the given point does the Taylor polynomial achieve that tolerance?'

Example 8.8.1.

[author=garrett, file =text_files/taylor_interval_size]
The specific example we'll get to here is *'For what range of* $x \ge 25$ *does* $5 + \frac{1}{10}(x-25)$ *approximate* $\sqrt{x}$ *to within* $.001$*?'*

Again, with the degree-one Taylor polynomial and corresponding remainder term, for reasonable functions f we have

$$f(x) = f(x_o) + f'(x_o)(x - x_o) + \frac{f''(c)}{2!}(x - x_o)^2$$

for some c between x_o and x. The **remainder** term is

$$\text{remainder term } = \frac{f''(c)}{2!}(x - x_o)^2$$

The notation 2! means '2-factorial', which is just 2, but which we write to be 'forward compatible' with other things later.

Again: no, we do not know what c *is, except that it is between* x_o *and* x. But this is entirely reasonable, since if we really knew it exactly then we'd be able to evaluate $f(x)$ exactly and we are evidently presuming that this isn't possible (or we wouldn't be doing all this!). That is, we have *limited information* about what c is, which we could view as the limitation on how precisely we can know the value $f(x)$.

To give an example of how to use this limited information, consider $f(x) = \sqrt{x}$ (yet again!). Taking $x_o = 25$, we have

$$\sqrt{x} = f(x) = f(x_o) + f'(x_o)(x - x_o) + \frac{f''(c)}{2!}(x - x_o)^2 =$$

$$= \sqrt{25} + \frac{1}{2}\frac{1}{\sqrt{25}}(x - 25) - \frac{1}{2!}\frac{1}{4}\frac{1}{(c)^{3/2}}(x - 25)^2 =$$

$$= 5 + \frac{1}{10}(x - 25) - \frac{1}{8}\frac{1}{c^{3/2}}(x - 25)^2$$

where all we know about c is that it is between 25 and x. What can we expect to get from this?

Well, we have to make a choice or two to get started: let's suppose that $x \ge 25$ (rather than smaller). Then we can write

$$25 \le c \le x$$

From this, because the three-halves-power function is *increasing*, we have

$$25^{3/2} \le c^{3/2} \le x^{3/2}$$

Taking inverses (with positive numbers) reverses the inequalities: we have

$$25^{-3/2} \geq c^{-3/2} \geq x^{-3/2}$$

So, *in the worst-case scenario*, the value of $c^{-3/2}$ is at most $25^{-3/2} = 1/125$.

And we can rearrange the equation:

$$\sqrt{x} - [5 + \frac{1}{10}(x - 25)] = -\frac{1}{8}\frac{1}{c^{3/2}}(x - 25)^2$$

Taking absolute values *in order to talk about error*, this is

$$|\sqrt{x} - [5 + \frac{1}{10}(x - 25)]| = |\frac{1}{8}\frac{1}{c^{3/2}}(x - 25)^2|$$

Now let's use our **estimate** $|\frac{1}{c^{3/2}}| \leq 1/125$ to write

$$|\sqrt{x} - [5 + \frac{1}{10}(x - 25)]| \leq |\frac{1}{8}\frac{1}{125}(x - 25)^2|$$

OK, having done this simplification, *now* we can answer questions like *For what range of $x \geq 25$ does $5 + \frac{1}{10}(x-25)$ approximate $\sqrt{x}$ to within .001?* We cannot hope to tell *exactly*, but only to give a range of values of x for which we *can* be sure *based upon our estimate*. So the question becomes: solve the inequality

$$|\frac{1}{8}\frac{1}{125}(x - 25)^2| \leq .001$$

(with $x \geq 25$). Multiplying out by the denominator of $8 \cdot 125$ gives (by coincidence?)

$$|x - 25|^2 \leq 1$$

so the solution is $25 \leq x \leq 26$.

So we can conclude that $\sqrt{x}$ is approximated to within .001 for all x in the range $25 \leq x \leq 26$. This is a worthwhile kind of thing to be able to find out.

Exercises

1. For what range of values of x is $x - \frac{x^3}{6}$ within 0.01 of $\sin x$?
2. Only consider $-1 \leq x \leq 1$. For what range of values of x *inside this interval* is the polynomial $1 + x + x^2/2$ within .01 of e^x?
3. On how large an interval around 0 is $1 - x$ within 0.01 of $1/(1+x)$?
4. On how large an interval around 100 is $10 + \frac{x-100}{20}$ within 0.01 of $\sqrt{x}$?

8.9 Achieving desired tolerance on desired interval

Discussion.

[author=garrett, file =text_files/taylor_adjusting_degree]
We saw before two questions about the accuracy of the Taylor polynomial (they were ??). Now we look at the most difficult question about accuracy:

'Given a function, given a fixed point, given an interval around that fixed point, and given a required tolerance, find how many terms *must be used in the Taylor expansion to approximate the function to within the required tolerance on the given interval.*

Example 8.9.1.

[author=garrett, file =text_files/taylor_adjusting_degree]
For example, let's get a Taylor polynomial approximation to e^x which is within 0.001 on the interval $[-\frac{1}{2}, +\frac{1}{2}]$. We use

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \frac{e^c}{(n+1)!}x^{n+1}$$

for some c between 0 and x, and where we do not yet know what we want n to be. It is very convenient here that the nth derivative of e^x is still just e^x! We are wanting to *choose n large-enough to guarantee that*

$$|\frac{e^c}{(n+1)!}x^{n+1}| \leq 0.001$$

for all x in that interval (without knowing anything too detailed about what the corresponding c's are!).

The error term is estimated as follows, by thinking about the *worst-case scenario* for the sizes of the parts of that term: we know that the exponential function is increasing along the whole real line, so in any event c lies in $[-\frac{1}{2}, +\frac{1}{2}]$ and

$$|e^c| \leq e^{1/2} \leq 2$$

(where we've not been too fussy about being accurate about how big the square root of e is!). And for x in that interval we know that

$$|x^{n+1}| \leq (\frac{1}{2})^{n+1}$$

So we are wanting to *choose n large-enough to guarantee that*

$$|\frac{e^c}{(n+1)!}(\frac{1}{2})^{n+1}| \leq 0.001$$

Since

$$|\frac{e^c}{(n+1)!}(\frac{1}{2})^{n+1}| \leq \frac{2}{(n+1)!}(\frac{1}{2})^{n+1}$$

we can be confident of the desired inequality if we can be sure that

$$\frac{2}{(n+1)!}(\frac{1}{2})^{n+1} \leq 0.001$$

That is, we want to 'solve' for n in the inequality

$$\frac{2}{(n+1)!}(\frac{1}{2})^{n+1} \leq 0.001$$

There is no genuine formulaic way to 'solve' for n to accomplish this. Rather, we just evaluate the left-hand side of the desired inequality for larger and larger values of n until (hopefully!) we get something smaller than 0.001. So, trying $n = 3$, the expression is

$$\frac{2}{(3+1)!}(\frac{1}{2})^{3+1} = \frac{1}{12 \cdot 16}$$

which is more like 0.01 than 0.001. So just try $n = 4$:

$$\frac{2}{(4+1)!}(\frac{1}{2})^{4+1} = \frac{1}{60 \cdot 32} \leq 0.00052$$

which is better than we need.

The conclusion is that we needed to take the Taylor polynomial of degree $n = 4$ to achieve the desired tolerance along the whole interval indicated. Thus, the polynomial

$$1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4}$$

approximates e^x to within 0.00052 for x in the interval $[-\frac{1}{2}, \frac{1}{2}]$.

Yes, such questions can easily become very difficult. And, as a reminder, there is no real or genuine claim that this kind of approach to polynomial approximation is 'the best'.

Exercises

1. Determine how many terms are needed in order to have the corresponding Taylor polynomial approximate e^x to within 0.001 on the interval $[-1, +1]$.

2. Determine how many terms are needed in order to have the corresponding Taylor polynomial approximate $\cos x$ to within 0.001 on the interval $[-1, +1]$.

3. Determine how many terms are needed in order to have the corresponding Taylor polynomial approximate $\cos x$ to within 0.001 on the interval $[\frac{-\pi}{2}, \frac{\pi}{2}]$.

4. Determine how many terms are needed in order to have the corresponding Taylor polynomial approximate $\cos x$ to within 0.001 on the interval $[-0.1, +0.1]$.

5. Approximate $e^{1/2} = \sqrt{e}$ to within .01 by using a Taylor polynomial with remainder term, expanded at 0. *(Do NOT add up the finite sum you get!)*

6. Approximate $\sqrt{101} = (101)^{1/2}$ to within 10^{-15} using a Taylor polynomial with remainder term. *(Do NOT add up the finite sum you get! One point here is that most hand calculators do not easily give 15 decimal places. Hah!)*

8.10 Integrating Taylor polynomials: first example

Discussion.

[author=garrett, file =text_files/taylor_integration]
Thinking simultaneously about the difficulty (or impossibility) of 'direct' symbolic integration of complicated expressions, by contrast to the ease of integration of *polynomials*, we might hope to get some mileage out of *integrating Taylor polynomials*.

Example 8.10.1.

[author=garrett, file =text_files/taylor_integration]
As a promising example: on one hand, it's not too hard to compute that

$$\int_0^T \frac{dx}{1-x}\,dx = [-\log(1-x)]_0^T = -\log(1-T)$$

On the other hand, if we write out

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots$$

then we could obtain

$$\begin{aligned}\int_0^T (1 + x + x^2 + x^3 + x^4 + \ldots)\,dx = [x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots]_0^T \\ = T + \frac{T^2}{2} + \frac{T^3}{3} + \frac{T^4}{4} + \ldots\end{aligned}$$

Putting these two together (and changing the variable back to 'x') gives

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots$$

(For the moment let's not worry about what happens to the error term for the Taylor polynomial).

This little computation has several useful interpretations. First, we obtained a Taylor polynomial for $-\log(1-T)$ from that of a geometric series, without going to the trouble of recomputing derivatives. Second, from a different perspective, we have an expression for the integral

$$\int_0^T \frac{dx}{1-x}\,dx$$

without necessarily mentioning the logarithm: that is, with some suitable interpretation of the trailing dots,

$$\int_0^T \frac{dx}{1-x}\,dx = T + \frac{T^2}{2} + \frac{T^3}{3} + \frac{T^4}{4} + \ldots$$

8.11 Integrating the error term: example

Example 8.11.1.

[author=garrett, file =text_files/taylor_integrating_error]
Being a little more careful, let's keep track of the error term in the example we've been doing: we have

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots + x^n + \frac{1}{(n+1)} \frac{1}{(1-c)^{n+1}} x^{n+1}$$

for some c between 0 and x, and also depending upon x and n. One way to avoid having the $\frac{1}{(1-c)^{n+1}}$ 'blow up' on us, is to keep x itself in the range $[0, 1)$ so that c is in the range $[0, x)$ which is inside $[0, 1)$, keeping c away from 1. To do this we might demand that $0 \le T < 1$.

For simplicity, and to illustrate the point, let's just take $0 \le T \le \frac{1}{2}$. Then in the *worst-case scenario*

$$|\frac{1}{(1-c)^{n+1}}| \le \frac{1}{(1-\frac{1}{2})^{n+1}} = 2^{n+1}$$

Thus, *integrating the error term,* we have

$$|\int_0^T \frac{1}{n+1} \frac{1}{(1-c)^{n+1}} x^{n+1}\, dx| \le \int \frac{1}{n+1} 2^{n+1} x^{n+1}\, dx = \frac{2^{n+1}}{n+1} \int_0^T x^{n+1}\, dx$$

$$= 2^{n+1} n + 1 [\frac{x^{n+2}}{n+2}]_0^T = \frac{2^{n+1} T^{n+2}}{(n+1)(n+2)}$$

Since we have cleverly required $0 \le T \le \frac{1}{2}$, we actually have

$$|\int_0^T \frac{1}{n+1} \frac{1}{(1-c)^{n+1}} x^{n+1}\, dx| \le \frac{2^{n+1} T^{n+2}}{(n+1)(n+2)} \le$$

$$\le \frac{2^{n+1} (\frac{1}{2})^{n+2}}{(n+1)(n+2)} = \frac{1}{2(n+1)(n+2)}$$

That is, we have

$$| - \log(1-T) - [T + \frac{T^2}{2} + \ldots + \frac{T^n}{n}]| \le \frac{1}{2(n+1)(n+2)}$$

for all T in the interval $[0, \frac{1}{2}]$. Actually, we had obtained

$$| - \log(1-T) - [T + \frac{T^2}{2} + \ldots + \frac{T^n}{n}]| \le \frac{2^{n+1} T^{n+2}}{2(n+1)(n+2)}$$

and the latter expression shrinks rapidly as T approaches 0.

8.12 Applications of Taylor series

Comment.

[author=duckworth, file =text_files/applications_of_taylor_series]
Finally, I want to show you an application of this stuff. The first application is a little artificial, since we have other ways to do it. But it's a good application none-the-less.

Example 8.12.1.

[author=duckworth, file =text_files/applications_of_taylor_series]
Use Maclaurin polynomials to find an approximation of the integral $\int_0^1 e^{-x^2}\,dx$.

We start with the polynomial for e^x: namely $1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \ldots$. Replacing x with $-x^2$ we obtain

$$e^{-x^2} = 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \frac{x^8}{4!} - \ldots$$

Now we intgrate this polynomial:

$$\begin{aligned}\int_0^1 e^{-x^2}\,dx &= \int_0^1 1 - x^2 + \tfrac{x^4}{2!} - \tfrac{x^6}{3!} + \tfrac{x^8}{4!} - \ldots \\ &= x - \tfrac{x^3}{3} + \tfrac{x^5}{5\cdot 2!} - \tfrac{x^7}{7\cdot 3!} + \tfrac{x^9}{9\cdot 4!} - \ldots \Big|_0^1 \\ &= 1 - \tfrac{1}{3} + \tfrac{1}{5\cdot 2} - \tfrac{1}{7\cdot 3!} + \tfrac{1}{9\cdot 4!} - \ldots\end{aligned}$$

If we add the first three terms here we get .7666. As a rough idea of how accurate this is, supppose we added the next term. This would change the result to .7429. This isn't much of a change. If we added one more term, this would change it even less.

Discussion.

[author=duckworth, file =text_files/binomial_series]
Recall that

$$\begin{aligned}(a+b)^2 &= a^2 + 2ab + b^2 \\ (a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3 \\ (a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4\end{aligned}$$

To get these coefficients we can look at Pascal's triangle. In this triangle, the numbers on row n are the coefficients used in $(a+b)^n$. You get a coefficient by adding the two numbers above it.

1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

This triangle is great, but what if we want to find $(a+b)^{27}$? Do we really want to write down 27 rows of this triangle? I think not. Then, is there a closed formula for the coefficients? Yes.

$$\text{Define: } \binom{n}{k} := \frac{\overbrace{n(n-1)\ldots}^{k \text{ factors}}}{k!}$$

Then we have

$$(a+b)^n = a^n + na^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \binom{n}{3}a^{n-3}b^3 + \cdots + nab^{n-1} + b^n$$

How does this relate to polynomials? Newton realized first that we could replace whole numbers for n by any real numbers, and secondly, we could replace b by x. (A critic of Newton once said that "any clever school boy could have thought of this"!). The following theorem is due to Newton.

Theorem 8.12.1.
[author=duckworth, file =text_files/binomial_series]
For any real number n, and for $|x| < 1$, we have

$$(1+x)^n = 1 + nx + \binom{n}{2}x^2 + \binom{n}{3}x^3 + \dots$$

Proof.
[author=duckworth, file =text_files/binomial_series]
To prove that the binomial series is correct one just applies the Maclaurin series to $(1+x)^n$. To use the binomial series for something like $(a+b)^n$ you factor out the larger number. So suppose $a \geq b$, then we write $(a+b)^n = a^n\left(1+\frac{b}{a}\right)^n$. Also, $(1-x)^n$ we treat as $(1+u)^n$ and substitute $-x$ in for u. This will give an alternating series. □

Example 8.12.2.

[author=duckworth, file =text_files/binomial_series]
You should double check the following yourself.

$$(1+x)^{-1/2} = 1 - \frac{1}{2}x + \frac{1\cdot 3}{2!\cdot 2^2}x^2 - \frac{1\cdot 3\cdot 5}{3!\cdot 2^3}x^3 + \frac{1\cdot 3\cdot 5\cdot 7}{4!\cdot 2^4}x^4 + \cdots + (-1)^n\frac{\overbrace{1\cdot 3\cdot 5\cdot \ldots}^{n \text{ factors}}}{n!\cdot 2^n}x^n + \ldots$$

To get the series for $\frac{\sin(x)}{\sqrt{1+x/4}}$ one would substitute $x/4$ into the series for $(1+x)^{-1/2}$, then multiply the result by the series for $\sin(x)$.

Chapter 9

Infinite Series

Definition 9.0.1.

[author=wikibooks, file =text_files/introduction_to_series]
An arithmetic series is the sum of a sequence of terms. For example, an interesting series which appears in many practical problems in science, engineering, and mathematics is the geometric series $r + r^2 + r^3 + r^4 + ...$ where the ... indicates that the series continues indefinetly. A common way to study a particular series (following Cauchy) is to define a sequence consisting of the sum of the first n terms. For example, to study the geometric series we can consider the sequence which adds together the first n terms $S_n(r) = \sum_{i=1}^{n} r^i$. Generally by studying the sequence of partial sums we can understand the behavior of the entire infinite series.

Two of the most important questions about a series are

Does it converge? If so, what does it converge to?

9.1 Convergence

Definition 9.1.1.

[author=duckworth, file =text_files/introduction_to_series_convergence]
If we are given a sequence of numbers $a_0, a_1, a_2, a_3, \ldots$ we say $\lim_{n\to\infty} a_n$ exists, and equals L, if the values of a_n get closer and closer to L as n gets bigger and bigger. We say $\sum_{i=0}^{\infty} c_i$ exists if $\lim_{n\to\infty} a_n$ exists for the following sequence of numbers:

$$\begin{aligned} a_0 &= c_0 \\ a_1 &= c_0 + c_1 \\ a_2 &= c_0 + c_1 + c_2 \\ &\vdots \end{aligned}$$

Example 9.1.1.

[**author=duckworth**, file =**text_files/introduction_to_series_convergence**]
Let r be a real number. Then $\sum_{i=0}^{\infty} ar^i$ equals $\frac{a}{1-r}$ if $|r| < 1$, and does not exist otherwise.

Note: this is proven in an *ad hoc* manner, meaning, the proof is made up just for this series and does not follow a general strategy (essentially you multiply the partial sum $r^0 + r^1 + \cdots + r^n$ by $r - 1$ and simplify). If you need to find $\sum_{i=a}^{\infty}$ you use the equation

$$\underbrace{\sum_{i=0}^{\infty}}_{\text{use formula}} = \underbrace{\sum_{i=0}^{a-1}}_{\text{finite}} + \underbrace{\sum_{i=a}^{\infty}}_{\text{solve for this}} .$$

You should always think of the example of Zeno's paradox. (i.e. to get half way across the room, then half of the remaining distance, then half of the remaining distance etc. So you have $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1.$)

Example 9.1.2.

[**author=wikibooks**, file =**text_files/introduction_to_series_convergence**]
For example, it is fairly easy to see that for $r > 1$, the geometric series $S_n(r)$ will not converge to a finite number (i.e., it will diverge to infinity). To see this, note that each time we increase the number of terms in the series $S_n(r)$ increases.

Example 9.1.3.

[**author=wikibooks**, file =**text_files/introduction_to_series_convergence**]
Perhaps a more surprising and interesting fact is that for $|r| < 1$, $S_n(r)$ will converge to a finite value. Specifically, it is possible to show that $\lim_{n\to\infty} S_n(r) = \frac{r}{1-r}$. Indeed, consider the quantity $(1-r)S_n(r) = (1-r)\sum_{i=1}^{n} r^n = \sum_{i=1}^{n} r^n - \sum_{i=2}^{n+1} r^n = r - r^{n+1}$ Since $r^{n+1} \to 0$ as $n \to \infty$ for $|r| < 1$, this shows that $(1-r)S_n(r) \to r$ as $n \to \infty$. The quantity $1 - r$ is non-zero and doesn't depend on n so we can divide by it and arrive at the formula we want.

We'd like to be able to draw similar conclusions about any series.

Unfortunately, there is no simple way to sum a series. The most we will be able to do is determine if it converges.

Example 9.1.4.

[**author=wikibooks**, file =**text_files/introduction_to_series_convergence**]
It is obvious that for a series to converge, the a_n must tend to zero, but this is not sufficient.

Consider the harmonic series, the sum of 1/n, and group terms

$$\begin{aligned} \sum_1^{2^m} \frac{1}{n} &= 1 + \frac{1}{2} + \quad \frac{1}{3} + \frac{1}{4} + \\ &> \frac{3}{2} + \quad \frac{1}{4}2 + \\ &= \frac{3}{2} + \quad \frac{1}{2} + \end{aligned}$$

As m tends to infinity, so does this final sum, hence the series diverges.

We can also deduce something about how quickly it diverges. Using the same grouping of terms, we can get an upper limit on the sum of the first so many terms, the partial sums.

$1 + \frac{m}{2} < \sum_1^{2^m} \frac{1}{n} < 1 + m$ or $1 + \frac{\ln_2 m}{m} < \sum_1^{m} \frac{1}{n} < 1 + \ln_2 m$ and the partial sums increase like log m, very slowly.

Notice that to discover this, we compared the terms of the harmonic series with a series we knew diverged.

Test.

[author=wikibooks, file =text_files/introduction_to_series_convergence]
Comparison test This is a convergence test (also known as the direct comparison test) we can apply to any pair of series. If b_n converges and $|a_n| \leq |b_n|$ then a_n converges. If b_n diverges and $|a_n| \geq |b_n|$ then a_n diverges.

There are many such tests, the most important of which we'll describe in this chapter.

Definition 9.1.2.

[author=duckworth, file =text_files/absolute_convergence_of_series]
We say $\sum_{i=0}^{\infty} a_i$ is *absolutely convergent* if the series $\sum_{i=0}^{\infty} |a_i|$ converges. (Note: in general, it is easier for a series to converge if some of the terms are negative. For example, see the Alternating series test.) We say the series is *conditionally convergent* if it converges but is not absolutely convergent. Any series which is absolutely convergent is also convergent without absolute values.

Theorem 9.1.1.
[author=wikibooks, file =text_files/absolute_convergence_of_series]
If the series of ***absolute*** *values,* $\sum_{n=1}^{\infty} |a_n|$*, converges, then so does the series* $\sum_{n=1}^{\infty} a_n$

Comment.

[author=wikibooks, file =text_files/absolute_convergence_of_series]
We say such a series converges absolutely.

The converse does not hold. The series 1-1/2+1/3-1/4 ... converges, even though the series of its absolute values diverges.

A series like this that converges, but not absolutely, is said to converge conditionally.

Comment.

[author=wikibooks, file =text_files/absolute_convergence_of_series]
If a series converges absolutely, we can add terms in any order we like. The limit will still be the same.

If a series converges conditionally, rearranging the terms changes the limit. In fact, we can make the series converge to any limit we like by choosing a suitable rearrangement.

E.g, in the series 1-1/2+1/3-1/4 ..., we can add only positive terms until the partial sum exceeds 100, subtract 1/2, add only positive terms until the partial sum exceeds 100, subtract 1/4, and so on, getting a sequence with the same terms that converges to 100.

This makes absolutely convergent series easier to work with. Thus, all but one of convergence tests in this chapter will be for series all of whose terms are postive, which must be absolutely convergent or divergent series. Other series will be studied by considering the corresponding series of absolute values.

9.2 Various tests for convergence

Rule 9.2.1.

[author=duckworth, file =text_files/ratio_test]
Consider the series $\sum_{i=0}^{\infty} a_i$. Let $\lim_{i\to\infty} \frac{a_{i+1}}{a_i} = L$. If $L < 1$ then the series is absolutely convergent. If $L > 1$ (or $L = \infty$) then the series diverges. If $L = 1$ then the test tells you nothing.

Comment.

[author=duckworth, file =text_files/ratio_test]
Note: when we say that the ratio test tells us nothing about the case $L = 1$, this means that there are convergent series with $L = 1$ and there are divergent series with $L = 1$. Note, the test is easy to remember because for convergence we need (for positive numbers) that the terms decrease; if the terms decrease this means that a_{i+1} should be smaller than a_i, and if this is the case then $\frac{a_{i+1}}{a_i} < 1$. Note, we have learned how to find $\lim_{i\to\infty}$ of many fractions.

Example 9.2.1.

[author=wikibooks, file =text_files/ratio_test]
E.g, suppose $a_n = \frac{n!n!}{(2n)!}$ then $\frac{a_{n+1}}{a_n} = \frac{(n+1)^2}{(2n+1)(2n+2)} = \frac{n+1}{4n+2} \to \frac{1}{4}$ so this series converges.

Rule 9.2.2.

[author=duckworth, file =text_files/root_test_for_series]
The Root Test. Consider the series $\sum_{i=0}^{\infty} a_i$. Let $L = \lim_{i\to\infty} \sqrt[i]{|a_i|}$. If $L < 1$ then the series is absolutely convergent. If $L > 1$ (or $L = \infty$) then the series diverges. If $L = 1$ then the test tells us nothing abouth the convergence of the series.

Comment.

[author=duckworth, file =text_files/root_test_for_series]
Note: the statement that the series tells us nothing when $L = 1$ means that there are convergent series with $L = 1$ and there are divergent series with $L = 1$. Note: it is often easier to apply the ratio test than the root test. So the root test is best

to apply when we have i^{th} powers in a_i which we are trying to cancel.

Rule 9.2.3.

[`author=duckworth`, file =`text_files/integral_test`]
Integral Test If $c_i = f(i)$ where $f(x)$ is some function defined on the interval $[1, \infty)$ then $\sum_{i=1}^{\infty} c_i$ exists $\iff$ $\int_1^{\infty} f(x)\,dx$ exists. Of course, this is only useful if we know how to evaluate the integral. (Note: " $\iff$ " means the things on either side are equivalent.) Let $R_n = \sum_{i=0}^{\infty} c_i - \sum_{i=0}^{n} c_i$ be the error. Then we have: $\int_{n+1}^{\infty} f(x)\,dx \leq R_n \leq \int_n^{\infty} f(x)\,dx$. Furthermore, the total sum $\int_{i=0}^{\infty} c_i$ may be estimated via $s_n + \int_{n+1}^{\infty} f(x)\,dx \leq \sum_{i=0}^{\infty} c_i \leq s_n + \int_n^{\infty} f(x)\,dx$.

We can prove this test works by writing the integral as $\int_1^{\infty} f(x)dx = \sum_{n=1}^{\infty} \int_n^{n+1} f(x)\,dx$ and comparing each of the integrals with rectangles, giving the inequalities $f(n) \geq \int_n^{n+1} f(x)dx \geq f(n+1)$ Applying these to the sum then shows convergence.

Rule 9.2.4.

[`author=duckworth`, file =`text_files/integral_test`]
p-series Test The series $\sum_{i=1}^{\infty} \frac{1}{i^p}$ converges $\iff$ $p > 1$.

If $p = 1$ then $\int_1^{\infty} \frac{1}{x}\,dx = \lim_{b\to\infty} \ln(x)\Big|_1^b$. Since $\lim_{b\to\infty} \ln(b) = \infty$, the integral and the series diverge.

If $p \neq 1$ then

$$\int_1^{\infty} \frac{1}{x^p}\,dx = \int_1^{\infty} x^{-p}\,dx = \lim_{b\to\infty} \frac{x^{-p+1}}{-p+1}.$$

If $-p+1 > 0$, then this last fraction has more x's on top and therefore $\lim_{b\to\infty} \frac{x^{-p+1}}{=}\infty$ and the series diverges. If $-p+1 < 0$, then this last fraction has x's on the bottom and therefore $\lim_{b\to\infty} \frac{x^{-p+1}}{=}0$.

Rule 9.2.5.

[`author=duckworth`, file =`text_files/comparison_test_for_series`]
(a) if $a_n \geq b_n \geq 0$ and $\sum_{i=0}^{\infty} a_n$ exists then so does $\sum_{i=0}^{\infty} b_n$.

(b) If $\lim_{n\to\infty} \frac{a_n}{b_n}$ equals a non-zero number, then $\sum_{i=0}^{\infty} a_n$ exists $\iff$ $\sum_{i=0}^{\infty} b_n$ exists.

Example 9.2.2.

[`author=duckworth`, file =`text_files/comparison_test_for_series`]
The comparison theorem part (b) shows that $\sum_{i=1}^{\infty} \frac{i^2+i+1}{i^3-100}$ does not exist by comparing it to $\sum_{i=1}^{\infty} \frac{1}{i}$.

The comparison theorem part (b) shows that $\sum_{i=1}^{\infty} \frac{i^2+i+1}{i^4-100}$ does exist by comparing it to $\sum_{i=1}^{\infty} \frac{1}{i^2}$.

Rule 9.2.6.

[author=wikibooks, file =text_files/limit_comparison_test]
If b_n converges, and $\lim \frac{|a_n|}{b_n} < \infty$ then a_n converges.

If c_n diverges, and $\lim \frac{|a_n|}{c_n} > 0$ then a_n converges

---

Example 9.2.3.

[author=wikibooks, file =text_files/limit_comparison_test]
Let $a_n = n^{-\frac{n+1}{n}}$

For large n, the terms of this series are similar to, but smaller than, those of the harmonic series. We compare the limits.

$$\lim \frac{|a_n|}{c_n} = \lim \frac{n}{n^{\frac{n+1}{n}}} = \lim \frac{1}{n^{\frac{1}{n}}} = 1 > 0$$

so this series diverges.

Definition 9.2.1.

[author=wikibooks, file =text_files/alternating_series_test]
If the signs of the a_n alternate, $a_n = (-1)^n|a_n|$ and they are decreasing, then we call this an **alternating** series.

Theorem 9.2.1.
[author=wikibooks, file =text_files/alternating_series_test]
The series sum converges provided that $\lim_{n\to\infty} a_n = 0$.

The error in a partial sum of an alternating series is smaller than the first omitted term. $|\sum_{n=1}^{\infty} a_n - \sum_{n=1}^{m} a_n| < |a_{m+1}|$

Comment.

[author=wikibooks, file =text_files/alternating_series_test]
There are other tests that can be used, but these tests are sufficient for all commonly encountered series.

Theorem 9.2.2.
[author=duckworth, file =text_files/alternating_series_test]
If $b_i \geq b_{i+1}$ *(for all* i*) and* $\lim_{i\to\infty} b_i = 0$, *then* $\sum_{i=0}^{\infty}(-1)^i b_i$ *converges. Furthermore, if* $R_n = \sum_{i=0}^{\infty} b_i - \sum_{i=0}^{n} b_i$ *is the error, then* $|R_n| \leq b_{n+1}$.

Comment.

[author=duckworth, file =text_files/alternating_series_test]
Note: the error estimate in the alternating series test is often the best error estimate we will get.

9.3 Power series

Discussion.

[author=duckworth, file =text_files/power_series]
First of all, a power series is different than some of the other series in these notes. Many of the other series had only fixed numbers in the terms; a power series has x's in it which represent an input that we plug different numbers into.

Definition 9.3.1.

[author=duckworth, file =text_files/power_series]
A power series is one of the form: $\sum_{i=0}^{\infty} c_i(x-a)^i$ where a is some constant and the c_i are coefficients. We call a the center of the series.

Theorem 9.3.1.
[author=duckworth, file =text_files/power_series]
Given a power series $\sum_{i=0}^{\infty} c_i(x-a)^i$, one of the following situations holds:

(i) The series only converges when $x = a$.

(ii) The series converges for all x.

(iii) The series converges for those x in the interval $(a-R, a+R)$ and diverges for those x that are $> a+R$ and those x that are $< a-R$.

Comment.

[author=duckworth, file =text_files/power_series]
Let R be as in the previous theorem part (c), or let $R = 0$ in part (a), or let $R = \infty$ in part (b). In each case we call R the **radius of convergence** of the power series.

Comment.

[author=duckworth, file =text_files/power_series]
Note: This statement does not tell us what happens for $x = a \pm R$, though sometimes we can figure this out by using the another test. In general, we need to use the root or ratio test to find R.

Example 9.3.1.

[author=duckworth, file =text_files/power_series]

Series	R
$\sum_{i=0}^{\infty} x^i$	$R = 1$
$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$	$R = \infty$
$\ln(x) = \sum_{i=0}^{\infty} (-1)^{n+1} \frac{(x-1)^n}{n}$	$R = 1$

.

Discussion.

[author=wikibooks, file =text_files/power_series]
The study of power series concerns ourselves with looking at series that can approximate some function over some interval.

Recall from elementary calculus that we can obtain a line that touches a curve at one point by using differentiation. So in a sense we are getting an approximation to a curve at one point. This does not help us very much however.

Let's look at the case of $y = \cos(x)$, about the point $x = 0$. We have a first approximation using differentiation by the line $y = 1$. Observe that $\cos(x)$ looks like a parabola upside-down at $x = 0$. So naturally we think "what parabola could approximate $\cos(x)$ at this point?" The parabola $1 - x^2/2$ will do. In fact, it is the best estimate using polynomials of degree 2. But how do we know this is so? This is the study of power series finding optimal approximations to functions using polynomials.

Definition 9.3.2.

[author=wikibooks, file =text_files/power_series]
A power series is a series of the form $a_0x^0 + a_1x^1 + ... + a_nx^n = \sum_{j=0}^{n} a_jx^j$

Theorem 9.3.2.
[author=wikibooks, file =text_files/power_series]
Radius of convergence *We can only use the equation $f(x) = \sum_{j=0}^{n} a_jx^j$ to study $f(x)$ when the power series converges. This may happen for a finite range, or for all real numbers.*

If the series converges only for x is some interval, then the ***radius of convergence*** *is half of the length of this interval.*

Example 9.3.2.

[author=wikibooks, file =text_files/power_series]
Consider the series $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$ (a geometric series) this converges when —x—¡1, so the radius of convergence is 1. $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ Using the [[Calculusratio test—ratio test]], this series converges when the ratio of successive terms is less than one, $\lim_{n\to\infty} \left| \frac{x^{n+1}}{(n+1)!} \frac{n!}{x^n} \right| < 1$ or $\lim_{n\to\infty} \left| \frac{x}{n} \right| < 1$ which is always true, so this power series has an infinite radius of convergence.

If we use the ratio test on an arbitary power series, we find it converges when $\lim \frac{|a_{n+1}x|}{|a_n|} < 1$ and diverges when $\lim \frac{|a_{n+1}x|}{|a_n|} > 1$ The radius of convergence is therefore $r = \lim \frac{|a_n|}{|a_{n+1}|}$ If this limit diverges to infinity, the series has an infinite radius of convergence.

Fact.

[author=wikibooks, file =text_files/power_series]
Differentiation and integration Within its radius of convergence a power series can be differentiated and integrated term by term. $\frac{d}{dx} \sum_{j=0}^{\infty} a_jx^j = \sum_{j=0}^{\infty} (j+1)a_{j+1}x^j$

$\int \sum_{j=0}^{\infty} a_j z^j dz = \sum_{j=1}^{\infty} \frac{a_{j-1}}{j} x^j$

Both the differential and the integral have the same radius of convergence as the original series.

Example 9.3.3.

[**author=wikibooks**, file =**text_files/power_series**]
This allows us to sum exactly suitable power series. E.g $\frac{1}{1+x} = 1-x+x^2-x^3+\ldots$. This is a geometric series, which converges for $|x| < 1$. Integrate both sides, and we get $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3}\ldots$ which will also converge for $|x| < 1$. When $x = -1$ this is the harmonic series, which diverges. When $x = 1$ this is an alternating series with diminishing terms, which converges to $\ln(2)$.

It also lets us write power series for integrals we cannot do exactly. E.g $e^{-x^2} = \sum (-1)^n \frac{x^{2n}}{n!}$ The left hand side can not be integrated exactly, but the right hand side can be. $\int_0^z e^{-x^2} dx = \sum \frac{(-1)^n z^{2n+1}}{(2n+1)n!}$ This gives us a power series for the sum, which has an infinite radius of convergence, letting us approximate the integral as closely as we like.

Definition 9.3.3.

[**author=wikibooks**, file =**text_files/taylor_series_in_context_of_power_series**]
The Taylor series of an infinitely often differentiable real (or complex) function f defined on an interval $(a-r, a+r)$ is the power series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Here, $n!$ is the factorial of n and $f^(n)(a)$ denotes the nth derivative of f at the point a. If this series converges for every x in the interval $(a-r, a+r)$ and the sum is equal to $f(x)$, then the function $f(x)$ is called analytic. If $a = 0$, the series is also called a Maclaurin series.

Comment.

[**author=wikibooks**, file =**text_files/taylor_series_in_context_of_power_series**]
To check whether the series converges towards $f(x)$, one normally uses estimates for the remainder term of Taylor's theorem. A function is analytic if and only if it can be represented as a power series the coefficients in that power series are then necessarily the ones given in the above Taylor series formula.

Comment.

[**author=wikibooks**, file =**text_files/taylor_series_in_context_of_power_series**]
The importance of such a power series representation is threefold. First, differentiation and integration of power series can be performed term by term and is hence particularly easy. Second, an analytic function can be uniquely extended to a holomorphic function defined on an open disk in the complex number—complex plane, which makes the whole machinery of complex analysis available. Third, the

(truncated) series can be used to compute function values approximately.

Example 9.3.4.

[author=wikibooks, file =text_files/taylor_series_in_context_of_power_series]
The function e^-1/x^2 is not analytic the Taylor series is 0, although the function is not. Note that there are examples of infinitely often differentiable functions $f(x)$ whose Taylor series converge, but are not equal to $f(x)$. For instance, for the function defined piecewise by saying that $f(x) = e^{-1/x^2}$ if $x \neq 0$ and $f(0) = 0$, all the derivatives are zero at $x = 0$, so the Taylor series of $f(x)$ is zero, and its radius of convergence is infinite, even though the function most definitely is not zero. This particular pathology does not afflict complex-valued functions of a complex variable. Notice that e^{-1/z^2} does not approach 0 as z approaches 0 along the imaginary axis.

Comment.

[author=wikibooks, file =text_files/taylor_series_in_context_of_power_series]
Some functions cannot be written as Taylor series because they have a singularity. In these cases, one can often still achieve a series expansion if one allows also negative powers of the variable x see Laurent series. For example, $f(x) = e^{-1/x^2}$ can be written as a Laurent series.

Examples 9.3.5.

[author=wikibooks, file =text_files/taylor_series_in_context_of_power_series]
Several important Taylor series expansions follow. All these expansions are also valid for complex arguments x.

$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$	for all x
$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n$	for $\lvert x\rvert < 1$
$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$	for $\lvert x\rvert < 1$
$(1+x)^\alpha = \sum_{n=0}^{\infty} C(\alpha,n) x^n$	for all $\lvert x\rvert < 1$, and all complex α, and the $C(\alpha,n)$ are the Binomial coefficients which are defined somewhere else, or which can be calculated on a case-by-case basis
$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$	for all x
$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}$	for all x
$\tan x = \sum_{n=1}^{\infty} \frac{B_{2n}(-4)^n(1-4^n)}{(2n)!} x^{2n-1}$	for $\lvert x\rvert < \frac{\pi}{2}$ and the B_{2n} are the Bernoulli numbers which are defined somewhere else, or which can be calculated on a case-by-case basis
$\sec x = \sum_{n=0}^{\infty} \frac{(-1)^n E_{2n}}{(2n)!} x^{2n}$	for $\lvert x\rvert < \frac{\pi}{2}$ and the E_{2n} are the Euler numbers which are defined somewhere else, or which can be calculated on a case-by-case basis
$\arcsin x = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n(n!)^2(2n+1)} x^{2n+1}$	for $\lvert x\rvert < 1$
$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}$	for $\lvert x\rvert < 1$
$\sinh x = \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} x^{2n+1}$	for all x
$\cosh x = \sum_{n=0}^{\infty} \frac{1}{(2n)!} x^{2n}$	x
$\tanh x = \sum_{n=1}^{\infty} \frac{B_{2n}4^n(4^n-1)}{(2n)!} x^{2n-1}$	for $\lvert x\rvert < \frac{\pi}{2}$ and the B_{2n} are the Bernoulli numbers which are defined somewhere else, or which can be calculated on a case-by-case basis
$\sinh^{-1} x = \sum_{n=0}^{\infty} \frac{(-1)^n(2n)!}{4^n(n!)^2(2n+1)} x^{2n+1}$	for $\lvert x\rvert < 1$
$\tanh^{-1} x = \sum_{n=0}^{\infty} \frac{1}{2n+1} x^{2n+1}$	for $\lvert x\rvert < 1$
$W_0(x) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} x^n$	for $\lvert x\rvert < \frac{1}{e}$

Comment.

[author=wikibooks, file =text_files/taylor_series_in_context_of_power_series]
The Taylor series may be generalised to functions of more than one variable with the formula

$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_d=0}^{\infty} \frac{\partial^{n_1}}{\partial x^{n_1}} \cdots \frac{\partial^{n_d}}{\partial x^{n_d}} \frac{f(a_1,\cdots,a_d)}{n_1!\cdots n_d!}(x_1-a_1)^{n_1}\cdots(x_d-a_d)^{n_d}$$

Of course, to use this formula one must know how to take derivatives in more than one dimension! In fact, one way to *define* derivatives in any dimension, is to say that they are the functions which give you the correct coefficients for a Taylor polynomial to work!

9.4 Formal Convergence

Definition 9.4.1.

[author=wikibooks, file =text_files/convergent_sequences_and_series]

A sequence is an infinite list of real numbers a_1, a_2, a_3, ..., where a_1 is the firts number on the list, a_2 is the second number on the list etc.

Comment.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
We can also think of a sequence as a function from the natural numbers to the real numbers: this just means that given n there's some description of what a_n is. In fact, most of the sequences we work with are explicitly given by a function, like "$a_n = 1/2^n$".

Since we can list the integers 1,2,3,... we likewise list a sequence f(1), f(2), f(3), f(4), ... We shall denote a sequence by an italic capital letter, the set of real values that function takes by the same non-italic capital letter , and the elements of that set with the corresponding lower case letter, and subscripts. For example sequence S takes values in the set S with elements $s_1, s_2, s_3 \ldots$.

S is a set of reals, S is a function from the integers to the reals, two different concepts. While we are being rigorous we must be careful not to confuse the two, but in general usage the concepts are interchangable.

We can also denote sequences by their function. For example if we say the function S is 3k, then the sequence consists of 3,6,9,....

In particular we will be interested in special types of sequences that converge. We first introduce three definitions.

Definition 9.4.2.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
A cauchy sequence is a sequence S where for every $\epsilon > 0$, there exists an integer $n(\epsilon)$ such that for all $k > n(\epsilon)$, $|s_k - s_{n(\epsilon)}| < \epsilon$.

Definition 9.4.3.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
A sequence, a_n is bounded above if there exists some number M such that $a_n \leq M$ for all n. We define bounded below similarly.

Definition 9.4.4.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
A sequence, S, converges if there exists a number, s, such that for all $\epsilon > 0$, there exists an integer $n(\epsilon)$ such that for all $k > n(\epsilon)$, $|s - s_{n(\epsilon)}| < \epsilon$. If the series is convergent we call the number, s, the limit of the sequence S. We write, $s = \lim_{n\to\infty} s_n$

Theorem 9.4.1.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
If there exists a number, s, such that for all $\epsilon > 0$, there exists an integer $n(\epsilon)$

such that for all $k > n(\epsilon)$, $|s - s_{n(\epsilon)}| < f(\epsilon)$ where f is such that δ smaller than or equal to some $\delta(\epsilon)$ implies $f(\delta) \leq \epsilon$, $f(x)$ is positive for all positive x, and f(0)=0 then S converges.

Proof.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
For any ϵ consider $n(\delta(\epsilon))$.

If $n > n(\delta(\epsilon))$ then

From given conditions, $|s - s_{n(\delta(\epsilon))}| < f(\delta(\epsilon)) \leq \epsilon$

So, S meets the conditions in Definition 3. □

Comment.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
This theorem means it is sufficient to prove that the difference between a term and the limit is less than some continuous positive function of ϵ.

Discussion.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
Were going to prove a sequence is Cauchy if and only if it is convergent. To do this we need some preliminary theorems.

Theorem 9.4.2.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
Every Cauchy sequence is bounded above and below.

Proof.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
We prove only that the sequence is bounded above. By definition 1 with $\epsilon = 1$, $\exists N$ such that $\forall n > N \quad |s_n - s_N| < 1$. Define $r = 2 + \sup\{s_1, s_2, ...s_N\}$ Then, by definition $s_n < r \quad \forall k \leq N$. If n¿N $s_n \leq 1 + s_N < r$. Therefore the sequence meets definition 2 with $s_+ = r$ The cauchy sequence is bounded above. □

Definition 9.4.5.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
A sequence is monotonically increasing if for all $n, mn \geq m$ implies $a_n \geq a_m$ a sequence is monotonically decreasing if for all $n, mn \geq m$ implies $a_n \leq a_m$.

Theorem 9.4.3.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
If S is bounded above and monotonically increasing, S converges to sup S b. If S is bounded below and monotonically decreasing, S converges to inf S

Proof.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
For a monotonically increasing sequence bounded aboveS, and all $\epsilon > 0$, we must have $s_N > \sup S - \epsilon$ for some N, else $\sup S - \epsilon$ is an upper bound of S, contradicting definition of sup for all $n > N$, $s_N \leq s_n$, by definition combine with first inequality to get $\sup S > s_n > supS - \epsilon$ rearrange $|s_n - \sup S| < \epsilon$ for all n larger than some N Hence $\sup S$ is the limit of S 3b) is proved similarly □

Theorem 9.4.4.
[author=wikibooks , file =text_files/convergent_sequences_and_series]
(The sandwich theorem) Given three sequences, R, S , T, If R and T both converge, lim R=lim T and $\exists N \, \forall n > N \quad r_n \leq s_n \leq t_n$ *Sequence S converges to the same limit*

Proof.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
Let $s = \lim R = \lim T$. For any $\epsilon > 0$, by definition of convergence, there exist M, N such that $\forall n > M \quad |r_n - s| < \epsilon \; \forall n > N \quad |t_n - s| < \epsilon$ Combing these two inequalities with the conditions on R and T gives $s - \epsilon < r_n \leq s_n \leq t_n < s + \epsilon$ for all n greater than the maximum of M and N on rearrangement, S satisfies the definition of convergence, with limit s, and $n(\epsilon) = \max\{M, N\}$. ☐

Theorem 9.4.5.
[author=wikibooks , file =text_files/convergent_sequences_and_series]
If R, and S are both convergent series $\forall n r_n \leq s_n \Rightarrow \lim r_n \leq \lim s_n$

Theorem 9.4.6.
[author=wikibooks , file =text_files/convergent_sequences_and_series]
A sequence S is convergent if and only if it is cauchy.

Proof.
[author=wikibooks, file =text_files/convergent_sequences_and_series]
Convergence implies cauchy. Assume S is convergent, with limit s For a $\epsilon > 0$, choose n such that $\forall k > n \quad |s_k - s| < \epsilon/2$ (always possible by defintion of convegence) Via triangle inequality, $|s_k - s_j| < |s_k - s| + |s - s_j| \; |s_k - s_j| < \epsilon/2 + \epsilon/2 = \epsilon$, for $j, k > n$. This is defintion of cauchy.

Cauchy implies convergence. ¡ Let S be a cauchy sequence Define two sequences R and T by $r_n = \inf\{s_m m \geq n\} \quad t_n = \sup\{s_m m \geq n\} \; r_n = \min(s_n, r_{n+1}) \leq r_{n+1}$ R is monotonically increasing. Similarly T is monotonically decreasing. $\forall m > n \quad r_1 \leq r_n \leq s_m \leq t_n \leq t_1$ so R and T are bounded above and below respectively. Being bounded and monotonic, they converge to their supremum and infimum repsectively. $r = \lim_{n\to\infty} r_n = \sup \inf\{s_m\} \quad t = \lim_{n\to\infty} t_n = \inf \sup\{s_m\}$ By theorem 5 since $r_n \leq t_n$ for all $n, r \leq t$. If, for some N, all s_n with $n > N$ are greater than r, r is a lower bound to the s_n but it is also an upper bound. For r to be both the s_n must be constant, making the series trivially convergent. Similarly for t So, for all N, there must be n, m larger than N, with $s_n \leq r \quad s_m \geq t$ $\forall N \, \exists n, m > N \quad |s_n - s_m| \geq |r - t|$ If $r \neq t$ this contradicts the definition of Cauchy, so r=t But S is bounded betweeen R and T, so by the sandwhich theorem, S is convergent ☐

Comment.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
We can now use cauchy and convergent interchangeably, as convenient. We will often prove a sequence is convergent by proving it to be cauchy.

Discussion.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
If can add, multiply, and divide sequences in the obvious way, then the limit of a sum/product/ratio of sequences will be the sum/product/ratio of their limits.

Fact.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
We define $(S + T)$ by $(s + t)_n = s_n + t_n$.

Addition on sequences inherits the group properties of the reals.

If S and T both converge, to s and t respectively, then for all $\epsilon > 0$, $\exists N \forall n > N$ $|s_n - s| < \epsilon/2 \quad |t_n - t| < \epsilon/2$ (definition of limit) $|s_n - s| + |t_n - t| > |s_n + t_n - s - t|$ hence $|s_n + t_n - (s + t)| < \epsilon$ So, by definition of limit, S+T converges to s+t

---

Fact.

[author=wikibooks, file =text_files/convergent_sequences_and_series]
We define (ST) by $(st)_n = s_n t_n$.

Mulitplication inherits commutativity and associativity from the reals.

If S and T both converge, to s and t respectively, then, then for all $\epsilon > 0$, $\exists N \forall n > N$ $|s_n - s| < \epsilon \quad |t_n - t| < \epsilon$ (definition of limit) $|s_n t_n - st| = |(s_n - s)(t_n - t) + s(t_n - t) + t(s_n - s)|$ $|s_n t_n - st| < |s_n - s||t_n - t| + |s||t_n - t| + |t||s_n - s|$ $|(st)_n - st| < \epsilon^2 + 2\epsilon(|s| + |t|)$

The right handside is a monotonic increasing function of ϵ, therefore it can be replaced by ϵ, and hence, by the definition of limit, ST converges to st.

Chapter 10

Ordinary Differential Equations

Discussion.

[author=wikibooks, file =text_files/introduction_to_ordinary_diffeqs]
Ordinary differential equations involve equations containing variables functions their derivatives and their solutions.

In studying integration, you already have considered solutions to very simple differential equations. For example, when you look to solving $\int f(x)\,dx = g(x)$ for g(x), you are really solving the differential equation $D\ g(x) = f(x)$

Discussion.

[author=duckworth, file =text_files/introduction_to_ordinary_diffeqs]
A differential equation is an equation involving x, y, y', possibly y'' etc, that we are trying to solve for y (which is a function). Here are all of the types of equations we will solve in this course:

- You will be given an equation and told what form the solution y should take. For example:
  1. Show that $y = x - x^{-1}$ is a solution of $xy' + y = 2x$.
  2. Show that $y = \sin(x)\cos(x) - \cos(x)$ is a solution of $y' + \tan(x)y = \cos^2(x)$ such that $y(0) = -1$.
  3. Find r such that $y = e^{rx}$ is a solution of $y'' - y' - 2y = 0$.
- We can separate the equation as $f(x)dx = g(y)dy$. Then we integrate both sides. For example:
  1. $\frac{dy}{dx} = \frac{y}{x}$ which separates as $\frac{1}{y}dy = \frac{1}{x}dx$.
  2. $(x^2+1)y' = xy$ which separates as $\frac{1}{y}dy = \frac{x}{x^2+1}dx$.
- Exponential growth and decay. All the problems in this section are variations on the following: $y' = cy$, or $\frac{dP}{dt} = rP$, or "the rate of change of the population is proportional to the population". Allthough this type of equation is very usefull, it's kind of stupid that the book waited until the fourth

section to introduce it: we know already how to solve these, they're all of the form Ce^{rt}!

10.1 Simple differential equations

Notation.

[author=wikibooks, file =text_files/ordinary_diffeqs]
The notations we use for solving differential equations will be crucial in the ease of solubility for these equations.

This document will be using three notations primarily f' to denote the derivative of f D f to denote the derivative of f $\frac{df}{dx}$ to denote the derivative of f (for separable equations).

Definition 10.1.1.

[author=duckworth, file =text_files/ordinary_diffeqs]
The highest derivative which appears an a differential equation is called the **order** of the differential equation.

Example 10.1.1.

[author=wikibooks, file =text_files/ordinary_diffeqs]
Consider the differential equation $3f''(x)+5xf(x) = 11$ Since the equation's highest derivative is 2, we say that the differential equation is of order 2.

Discussion.

[author=wikibooks, file =text_files/ordinary_diffeqs]
A key idea in solving differential equations will be that of Integration.

Let us consider the second order differential equation $f = 2$

How would we go about solving this?. It tells us that on differentiating twice, we obtain the constant 2 so, if we integrate twice, we should obtain our result.

Integrating once first of all

$\int f\,dx = \int 2\,dx\;\; f' = 2x + C_1$

We have transformed the apparently difficult second order differential equation into a rather simpler one, viz.

$f' = 2x + C_1$

This equation tells us that if we differentiate a function once, we get $2x + C_1$. If we integrate once more, we should find the solution.

$\int f'\,dx = \int 2x + C_1\,dx\;\; f = x^2 + C_1x + C_2$

This is the solution to the differential equation. We will get $f = 2$ for all values of C_1 and C_2.

The values C_1 and C_2 are known as initial conditions.

Discussion.

[author=wikibooks, file =text_files/ordinary_diffeqs]
Why are initial conditions useful? ODEs (ordinary differential equations) are useful in modelling physical conditions. We may wish to model a certain physical system which is initially at rest (so one initial condition may be zero), or wound up to some point (so an initial condition may be nonzero and be say 5 for instance) and we may wish to see how the system reacts under such an initial condition.

When we solve a system with given initial conditions, we substitute them during our process of integration. Without initial conditions, the answer we obtain is the most general solution.

Example 10.1.2.

[author=duckworth, file =text_files/ODEs_with_solution_of_known_form]
One type of differential equation method involves being told what form a solution should have where the "form" has some unknown constant, then plugging that form into the differential equation to solve for the unknown constant.

This is similar to how we verify that a formula satisfies a certain differential equation, so let's start with an example like that.

Example 10.1.3.

[author=duckworth, file =text_files/ODEs_with_solution_of_known_form]
should take the function we've been given, find its derivative, plug that into the equation and see if it works. We've been given $y = x - x^{-1}$, its derivative is $y' = 1 + x^{-2}$. The equation is $xy' + y = 2x$. Plugging in we get

$$x(1 + x^{-2}) + (x - x^{-1}) \stackrel{?}{=} 2x$$

where I've written $\stackrel{?}{=}$ because I am pretending that I don't know if the equation will work or not. Simplifying the left hand side (I always choose to work on just one side when I can, otherwise, if you cancel stuff from both sidse, you end up getting an equation which says "0 = 0", which is true, but sometimes a little confusing):

$$x + x^{-1} + x - x^{-1}$$

which equals $2x$, as we wanted to show.

10.2 Basic Ordinary Differential Equations

Discussion.

[author=wikibooks, file =text_files/basic_ordinary_diffeqs]
In this section we will consider four main types of differential equations separable homogeneous linear exact

There are many other forms of differential equation, however, and these will be dealt with in the next section

Derivation.

[author=wikibooks, file =text_files/separable_ordinary_differential_equations]
A separable equation is one of the form

$$\frac{dy}{dx} = f(x)/g(y).$$

In this context people always use dy/dx notation. Previously we have only dealt with simple differential equations with $g(y) = 1$. How do we solve such a separable equation as above?

We group x and dx terms together, and y and dy terms together as well. This gives

$$g(y)\,dy = f(x)\,dx.$$

Integrating both sides (on the left we integrate with respect to y, and on the right with respect to x) we get

$$\int g(y)\,dy = \int f(x)\,dx + C.$$

The resulting equation gives an implicit solution for $y(x)$. In practice, it is often possible to solve this equation for y.

Example 10.2.1.

[author=duckworth, file =text_files/separable_ordinary_differential_equations]
Starting with $\frac{dy}{dx} = \frac{y}{x}$ we divide by y to get $\frac{1}{y}\frac{dy}{dx} = \frac{1}{x}$ and we multiply by dx to get $\frac{1}{y}dy = \frac{1}{x}dx$. Note, you should always do the separation steps using multiplication and division. You should never wind up with something like "$f(y) + dy$"; this is meaningless nonsense!

O.k., so now we've got it separated, we integrate both sides:

$$\begin{aligned} \int \frac{1}{y}dy &= \int \frac{1}{x}dx \\ \ln|y| &= \ln|x| + C \\ |y| &= C|x| \text{ (new } C!) \end{aligned}$$

Example 10.2.2.

[author=wikibooks, file =text_files/separable_ordinary_differential_equations]
Here is a worked example illustrating the process.

We are asked to solve $\frac{dy}{dx} = 3x^2 y$

Separating $\frac{dy}{y} = (3x^2)\,dx$ Integrating $\int \frac{dy}{y} = \int 3x^2\,dx$ $\ln y = x^3 + C$ $y = e^{x^3+C}$ Letting $k = e^C$ where k is a constant we obtain $y = ke^{x^3}$ which is the general solution.

Example 10.2.3.

[author=wikibooks, file =text_files/separable_ordinary_differential_equations] Just for practice, let's verify that our answer in Example 10.2.2 really was a solution of the given differential equation. Note, this step is only for practice, it is *not* usually part of finding a solution.

We obtained $y = ke^{x^3}$ as the solution to $\frac{dy}{dx} = 3x^2y$

Differentiating the solution, $\frac{dy}{dx} = 3kx^2e^{x^3}$

Since $y = ke^{x^3}$, we can write $dydx = 3x^2y$ We see that we obtain our original differential equation, so we can confirm our working as being correct.

Discussion.

[author=duckworth, file =text_files/separable_ordinary_differential_equations] There's one kind of problem of this type which deservers special mention: the mixing problem! Some people seem to hate these, but that's only because they hate translating words into equations. Concentrate on this step and you'll be fine. In a mixing problem y represents some substance (often salt) which is mixed into something else (usually water). The basic form of the differential equation is:

$$\frac{dy}{dx} = \text{rate in} - \text{rate out}.$$

See, that's not so hard. The part about translating is making "rate in" and "rate out" into formulas. Usually, one of these you've been given directly in the problem, e.g. "salty water with .5kg of salt per liter flows into the tank at rate of 7 liters per minute". In this case you would have rate in $= .5 \times 7$. The other rate is usually found by multiplying the concentration (which is like density) by the amount of flow. E.g. "thoroughly mixed water flows out of the tank at the same rate as water flows in". In this case, we have:

$$\begin{aligned} \text{Concentration} &= \frac{\text{total amount salt}}{\text{total amount of water}} \\ &= \frac{y}{\text{total amount of water}} \end{aligned}$$

What's the total amount of water? I haven't told you yet. Suppose the problem started with the phrase "A hundred liter tank has salty water flowing in." Then we would have concentration $= y/100$. Finally, using the fact that the flow out equals the flow in, equals 7 liters per minute, this would give us

$$\begin{aligned} \text{rate out} &= \text{concentration} \times 7 \\ &= \frac{y}{100} \times 7 \end{aligned}$$

Thus, the final differential equation would be

$$\frac{dy}{dx} = .5 \times 7 - \frac{y}{100} \times 7$$

Example 10.2.4.

[author=duckworth, file =text_files/separable_ordinary_differential_equations]
A hundred liter tank has salty water flowing in. Salty water with .5 kg/L flows into the tank at rate of 7 L/min. Thoroughly mixed water flows out of the tank at the same rate as water flows in. Find an equation for the amount of salt in the tank.

Translating these words into equations we have that the rate of salt in is $.5 \times 7$ and the rate of salt out is the concentration times 7, which becomes $\frac{5}{100} \times 7$. Thus, the differential equation would is

$$\frac{dy}{dx} = .5 \times 7 - \frac{y}{100} \times 7.$$

We separate this (with multiplication and division) as

$$\frac{1}{.5 \times 7 - 7y/100} dy = dx$$

Let's move those constants around. Multiply both sides by 7, mulitply the top and bottom of the fraction by 100.

$$\begin{aligned}
\frac{100}{50-y} dy &= 7dx \\
\int \frac{100}{50-y} dx &= \int 7dx \\
100 \ln|50-y| &= 7x + C \\
\ln|50-y| &= \frac{1}{100}(7x+C) \\
\ln|50-y| &= \frac{7}{100}x + C \text{ (new } C\text{!)} \\
|50-y| &= Ce^{.007x} \text{ (new } C\text{!)} \\
y &= 50 \pm Ce^{.007x}
\end{aligned}$$

Example 10.2.5.

[author=duckworth, file =text_files/ODE_lengthy_example]
In this example we're going to take a differential equation and analyze it every way we can. (This is often done in real life problems where you don't stop as soon as you get a solution. You graph it, you find max/mins, you analyze it every way you can.)

Suppose that $y(t)$ is a solution of $\frac{dy}{dt} = y^4 - 6y^3 + 5y^2$. (a) Which values of y give contstant solutions? (b) Which values of y give increasing solutions? (c) Which values of y give decreasing solutions?

Recall that there are infinitely many solutions to a differential equation; in our case, we will see below that these correspond to the different values of C one can have when finding an anti-derivative.

The main idea is to translate these questions into ones involving derivatives. So, (a) is equivalent to: which values of y make $\frac{dy}{dt} = 0$? (b) is equivalent to: which values of y make $\frac{dy}{dt}$ positive? (c) is equivalent to: which values of y make $\frac{dy}{dt}$ negative? Since we have an equation for $\frac{dy}{dt}$, these questions are easily solved:

$$\begin{array}{lll} \frac{dy}{dt} = 0 & \text{is equivalent to} & 0 = y^4 - 6y^3 + 5y^2 \\ & \text{is equivalent to} & 0 = y^2(y-5)(y-1) \\ & \text{is equivalent to} & y = 0, 5, 1 \end{array}$$

So these are the solutions of part (a). The constant functions, $y = 0$, $y = 5$, $y = 1$ are all solutions of the differential equation.

For parts (b) and (c) we use a standard procedure: if you want to know when a function is positive or negative (in this case $\frac{dy}{dt}$), you find when it equals zero, and then between each pair of points where it's zero, it will stay positive or stay negative: you can figure out which by testing a single point (or by looking at the factors, if you have them). In our case we consider the intervals: $y < 0$, $0 < y < 1$, $1 < y < 5$, $5 < y$. We can see that $y^2(y-5)(y-1)$ will be positive for y values bigger than 5, between 0 and 1, and for values less than 0. We can see that $y^2(y-5)(y-1)$ will be negative for y values between 1 and 5. In other words, we have horizontal asymptotes at $y = 0$, $y = 1$ and $y = 5$. On one side of an asymptote you can draw a nice curve that's increasing, or decreasing, as we just figured out, so that the graph curves as it gets closer and closer to the asymptote, never quite touching the asymptote. Recall again that we can have infinitely many solutions of this differential equation. Putting all this information together, we can make a bunch of graphs of possible solutions, some of which I show in Figure 10.1.

So far we've lookd at this problem without actually solving it. But, if we know the method of separation we can solve this differential equation.

We divide both sides of $\frac{dy}{dt} = y^2(y-5)(y-1)$ to get

$$\frac{1}{y^2(y-5)(y-1)}dy = dt$$

To integrate the left hand side we need partial fractions (yay! I'm so glad we learned that already!). We set that up as follows:

$$\frac{1}{y^2(y-5)(y-1)} = \frac{A}{y} + \frac{B}{y^2} + \frac{C}{y-5} + \frac{D}{y-1}$$

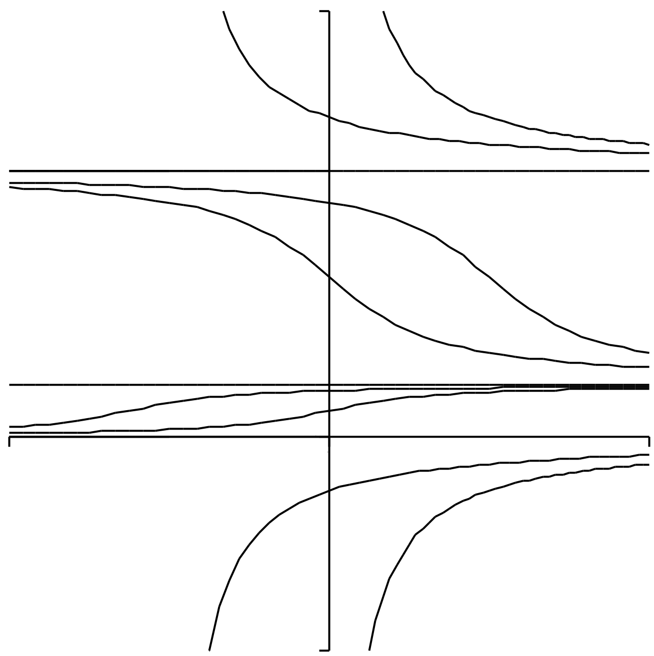
This is a relatively simple example to solve. Multiply both sides by $y^2(y-5)(y-1)$ to get

$$1 = Ay(y-5)(y-1) + B(y-5)(y-1) + Cy^2(y-1) + Dy^2(y-5)$$

If set $y = 0$ you get $B = \frac{1}{5}$. If you set $y = 1$ you get $D = -\frac{1}{4}$. If you set $y = 5$ you get $C = \frac{1}{100}$. This leaves only A to solve for. Plug in the values we've found for B, C and D, as well as any value for y that you like (besides 0, 1 or 5) and solve for $A = \frac{6}{25}$. Now we integrate both sides:

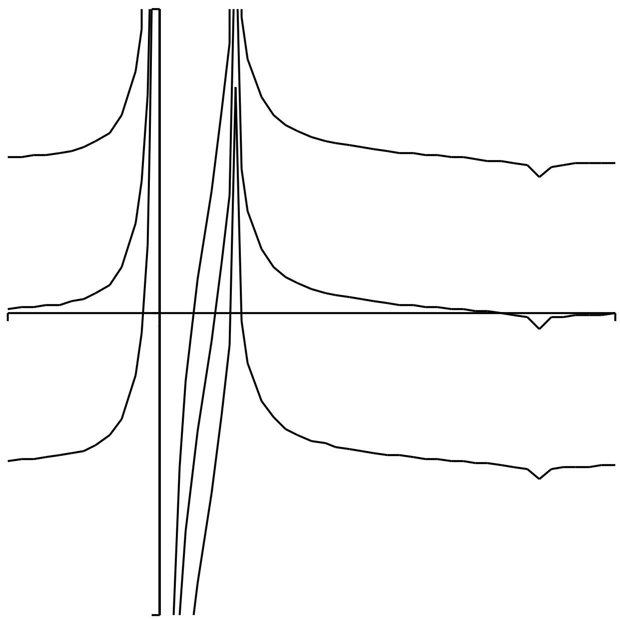
$$\begin{aligned} \int \frac{6}{25}\frac{1}{y} + \frac{1}{5}\frac{1}{y^2} + \frac{1}{100}\frac{1}{y-5} - \frac{1}{4}\frac{1}{y-1}\,dy &= \int dt \\ \frac{6}{25}\ln|y| - \frac{1}{5}\frac{1}{y} + \frac{1}{100}\ln|y-5| - \frac{1}{4}\ln|y-1| &= t + C \end{aligned}$$

Figure 10.1: A rough sketch of some solutions



ODE_lengthy_example_first_graph

Figure 10.2: The whole real graph: $-2 \leq x \leq 7$



ODE_lengthy_example_whole_graph

Now, I want to graph this solution directly, and compare it to the graph we sort of made up in Figure 10.1. But how can we graph it? I can't solve this equation of or y as a function of t. The trick is to graph t as a function of y; this is like graphing the inverse of the function that we really want. Thus, the picture we get will be like the one in Figure 10.1, except that we've switched the x and y axes (you can think about this as reflection of the graph across the line $y = x$, or you can think about this as "flipping" the graph so that the x-axis goes where the y-axis was, and you're looking at the graph through the back of the picture). So the graphs below put t on the vertical axis and y on the horizontal axis (this is like entering on our calculators $t \leftrightarrow y_1$ and $y \leftrightarrow x$). Note that the different values of C cause the graph to shift up and down. You can see here again why there should be infinitely many solutions of this differential equation.

Figure 10.2 shows the whole graph for three different values of C.

It's kind of hard to see the behavior of the graph when you look at the whole picture. This is one way that hand-drawn graphs are better than real ones: I could make each feature pretty clear in Figure 10.1, even though (actually *because*) it was not perfectly accurate.

We can work around the limitations of the real graph by looking closely at each part. In Figures 10.3–10.8 I zoom in on various parts of the graph. In each case you get (roughly) the shape that I drew in Figure 10.1.

Well, I think we've analyzed this problem from every way we can. The point was just to do the problem in two different ways; we combined tricks from integration (partial fractions) and used the method of separation. I guess we also got

Figure 10.3: Real graph: $-2 \leq x \leq 0$

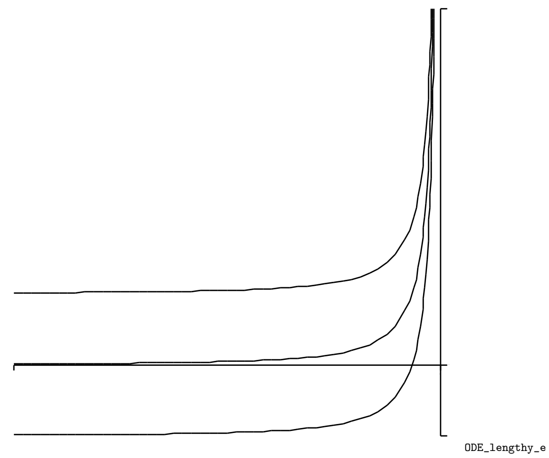


Figure 10.4: Real graph: $0 \leq x \leq 1$

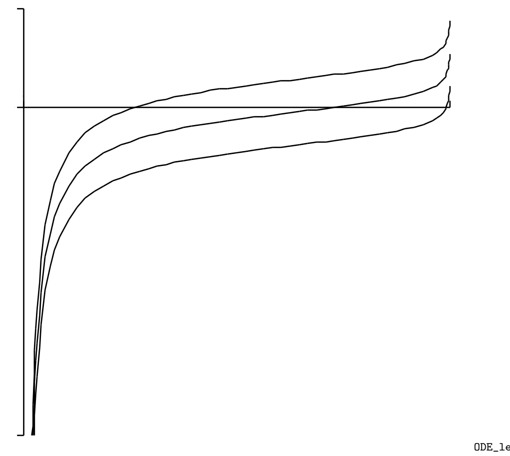


Figure 10.5: Real graph: $1 \leq x \leq 5$

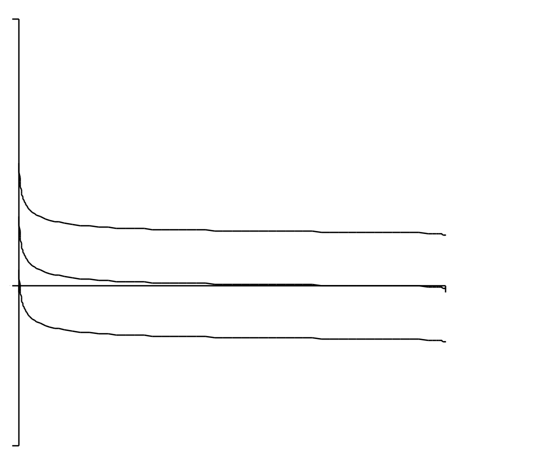


Figure 10.6: Real graph: $1 \leq x \leq 1.01$

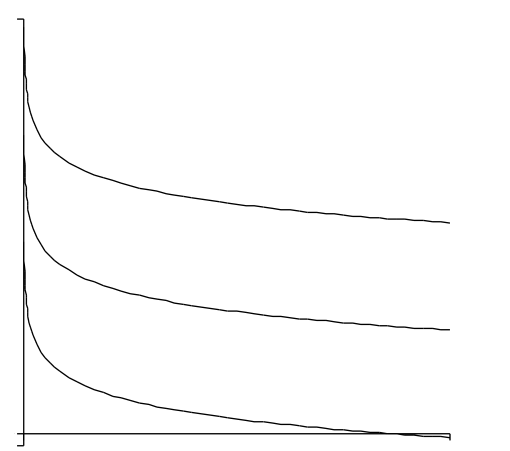


Figure 10.7: Real graph: $2 \leq x \leq 5$

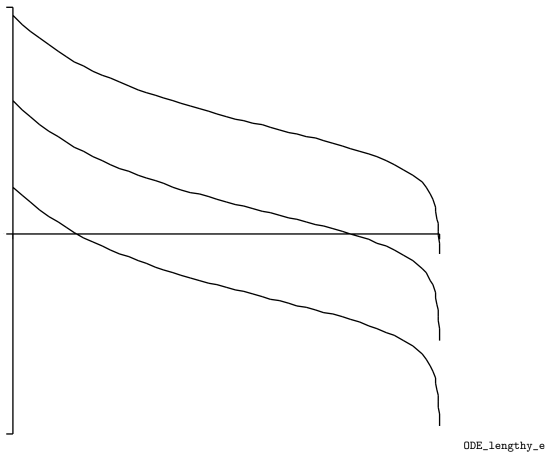
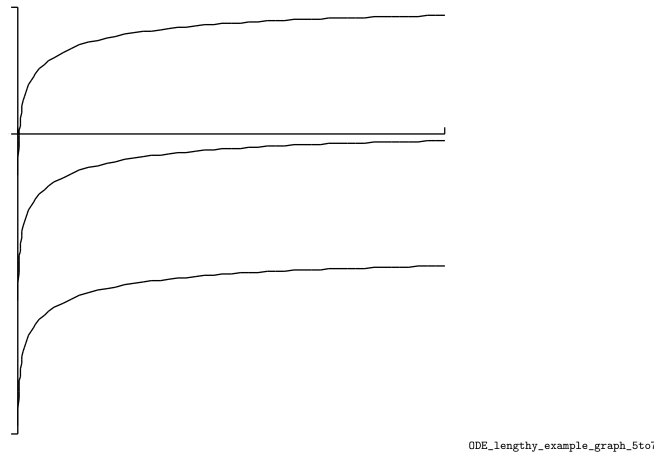


Figure 10.8: Real graph: $5 \leq x \leq 7$



practice in looking at a graph, and even a little bit of review of inverse functions (i.e. reversing the roles of x and y). Thanks for reading through this.

Definition 10.2.1.

[**author=wikibooks**, file =**text_files/homogeneous_ordinary_differential_equations**]
A homogeneous equation is in the form $\frac{dy}{dx} = f(y/x)$

This looks difficult as it stands, however we can utilize a substitution y=xv and use the product rule.

The equation above then becomes, using the product rule $\frac{dy}{dx} = v + x\frac{dv}{dx}$.

Then $v + x\frac{dv}{dx} = f(v)$ $x\frac{dv}{dx} = f(v) - v$ $\frac{dv}{dx} = \frac{f(v)-v}{x}$ which is a seperable equation and can be solved as above.

However let's look at a worked equation to see how homogeneous equations are solved.

Example 10.2.6.

[**author=wikibooks**, file =**text_files/homogeneous_ordinary_differential_equations**]
We have the equation $\frac{dy}{dx} = \frac{y^2+x^2}{yx}$

This does not appear to be immediately seperable, but let us expand to get $\frac{dy}{dx} = \frac{y^2}{yx} + \frac{x^2}{yx}$ $\frac{dy}{dx} = \frac{x}{y} + \frac{y}{x}$

Substituting y=xv which is the same as substituting v=y/x $\frac{dy}{dx} = 1/v + v$

Now $v + x\frac{dv}{dx} = 1/v + v$ Cancelling v from both sides $x\frac{dv}{dx} = 1/v$ Seperating $v\,dv = dx/x$ Integrating both sides $v^2 + C = \ln(x)$ $\left(\frac{y}{x}\right)^2 = \ln(x) - C$ $y^2 = x^2\ln(x) - Cx^2$ $y = x\sqrt{\ln(x) - C}$

which is our desired solution.

Definition 10.2.2.

[author=wikibooks, file =text_files/linear_ordinary_differential_equations]
A linear first order differential equation is a differential equation in the form $a(x)\frac{dy}{dx} + b(x)y = c(x)$

Rule 10.2.1.

[author=wikibooks, file =text_files/linear_ordinary_differential_equations]
Multiplying or dividing a linear first order differential equation by any non-zero function of x makes no difference to its solutions so we could always divide by $a(x)$ to make the coefficient of the differential 1, but writing the equation in this more general form may offer insights.

At first glance, it is not possible to integrate the left hand side, but there is one special case. If b happens to be the differential of a then we can write

$$a(x)\frac{dy}{dx} + b(x)y = a(x)\frac{dy}{dx} + y\frac{da}{dx} = ddxa(x)y$$

and integration is now straightforward.

Since we can freely multiply by any function, lets see if we can use this freedom to write the left hand side in this special form.

We multiply the entire equation by an arbitary $I(x)$ getting

$$aI\frac{dy}{dx} + bIy = cI$$

then impose the condition

$$\frac{d}{dx}aI = bI.$$

If this is satisified the new left hand side will have the special form. Note that multiplying I by any constant will leave this condition still satisfied.

Rearranging this condition gives

$$\frac{1}{I}\frac{dI}{dx} = \frac{b - \frac{da}{dx}}{a}$$

We can integrate this to get

$$\ln I(x) = \int \frac{b(z)}{a(z)}dz - \ln a(x) + c \quad I(x) = \frac{k}{a(x)}e^{\int \frac{b(z)}{a(z)}dz}.$$

We can set the constant k to be 1, since this makes no difference. Next we use I on the original differential equation, getting

$$e^{\int \frac{b(z)}{a(z)}dz}\frac{dy}{dx} + e^{\int \frac{b(z)}{a(z)}dz}\frac{b(x)}{a(x)}y = e^{\int \frac{b(z)}{a(z)}dz}\frac{c(x)}{a(x)}.$$

Because we've chosen I to put the left hand side in the special form we can rewrite this as

$$\frac{d}{dx}(ye^{\int \frac{b(z)}{a(z)}dz}) = e^{\int \frac{b(z)}{a(z)}dz}\frac{c(x)}{a(x)}.$$

Integrating both sides and dividing by I we obtain the final result

$$y = e^{-\int \frac{b(z)}{a(z)} dz} \left(\int e^{\int \frac{b(z)}{a(z)} dz} \frac{c(x)}{a(x)} dx + C \right).$$

We call I an integrating factor. Similar techniques can be used on some other calclulus problems.

Example 10.2.7.

[author=wikibooks, file =text_files/linear_ordinary_differential_equations]
Consider

$$\frac{dy}{dx} + y \tan x = 1 \quad y(0) = 0.$$

First we calculate the integrating factor.

$$I = e^{\int \tan z dz} = e^{\ln \sec x} = \sec x.$$

Multiplying the equation by this gives

$$\sec x \frac{dy}{dx} + y \sec x \tan x = \sec x$$

or

$$\frac{d}{dx} y \sec x = \sec x$$

We can now integrate

$$y = \cos x \int_0^x \sec z \, dz = \cos x \ln(\sec x + \tan x)$$

Definition 10.2.3.

[author=wikibooks, file =text_files/exact_ordinary_differential_equations]
An **exact equation** is in the form $f(x,y)dx + g(x,y)dy = 0$ and, has the property that $D_x f = D_y g$

Rule 10.2.2.

[author=wikibooks, file =text_files/exact_ordinary_differential_equations]
If we have an exact equation then there exists a function h(x, y) such that $D_y h = fandD_x h = g$

So then the solutions are in the form $h(x,y) = c$ by using total differentials. We can find the function $h(x,y)$ by integration.

Example 10.2.8.

[author=wikibooks, file =text_files/exact_ordinary_differential_equations]
Consider the differential equation $(3x^2 + 6y^2)dx + ((3x^2 + 6y^2 + 4y)dy$

It is exact since $D_x(3x^2 + 6y^2) = 6xD_y(3x^2 + 6y^2 + 4y) = 6x$

Now, there exists a function h such that 1)$D_x h = f = (3x^2 + 6y^2)$ 2) $D_y h = g = (3x^2 + 6y^2 + 4y)$

Integrate$D_x h$, with treating y as a constant $h(x, y) = 2y^3 + 3x^2 y + r(y)$ (We have the function r(y) because on differentiating with respect to x of the above expression, r(y) disappears - this is the similar procedure of adding an arbitrary constant)

So now, $D_y h = 3x^2 + 6y^2 + r'(x)$ Comparing with (2), we see $r'(y) = 4y$, so $r(y) = 2y^2 + C$

So substituting above, we get $h(x, y)2y^3 + 3x^2 y + 2y^2 + C = C_1$ where C_1 is a constant, and our most general solution is then $2y^3 + 3x^2 y + 2y^2 = k$ and we have simply moved the two constants to the one side of the expression and made this one constant

10.3 Higher order differential equations

Discussion.

[`author=wikibooks`, file =`text_files/introduction_to_higher_order_diffeqs`]
The generic solution of a $n^t h$ order ODE will contain n constants of integration. To calculate them we need n more equations. Most often, we have either boundary conditions, the values of y and its derivatives take for two different values of x or initial conditions, the values of y and its first n-1 derivatives take for one particular value of x.

Derivation.

[`author=wikibooks`, file =`text_files/reducible_higher_order_ODEs`]
If the independent variable x does not occur in the differential equation then its order can be lowered by one. This will reduce a second order ODE to first order.

Consider the equation

$$F\left(y, \frac{dy}{dx}, \frac{d^2y}{dx^2}\right) = 0.$$

Define $u = \frac{dy}{dx}$. Then

$$\frac{d^2y}{dx^2} = \frac{du}{dx} = \frac{du}{dy} \cdot \frac{dy}{dx} = \frac{du}{dy} \cdot u.$$

Substitute these two expression into the equationa and we get

$$F\left(y, u, \frac{du}{dy} \cdot u\right) = 0$$

which is a first order ODE.

Example 10.3.1.

[`author=wikibooks`, file =`text_files/reducible_higher_order_ODEs`]
Solve $1 + 2y^2\,\mathrm{D}^2\,y = 0$ if at x=0, y=Dy=1

First, we make the substitution, getting $1 + 2y^2 u \frac{du}{dy} = 0$ This is a first order ODE. By rearranging terms we can separate the variables $udu = -\frac{dy}{2y^2}$ Integrating this gives $u^2/2 = c + 1/2y$ We know the values of y and u when x=0 so we can find c $c = u^2/2 - 1/2y = 1^2/2 - 1/(2 \cdot 1) = 1/2 - 1/2 = 0$ Next, we reverse the substitution $\frac{dy}{dx}^2 = u^2 = \frac{1}{y}$ and take the square root $\frac{dy}{dx} = \pm \frac{1}{\sqrt{y}}$ To find out which sign of the square root to keep, we use the initial condition, Dy=1 at x=0, again, and rule out the negative square root. We now have another separable first order ODE, $\frac{dy}{dx} = \frac{1}{\sqrt{y}}$ Its solution is $\frac{2}{3} y^{\frac{3}{2}} = x + d$ Since y=1 when x=0, d=2/3, and $y = \left(1 + \frac{3x}{2}\right)^{\frac{2}{3}}$

Definition 10.3.1.

[**author=wikibooks**, file =**text_files/linear_higher_order_ODEs**]
An ODE of the form $\frac{d^n y}{dx^n} + a_1(x) \frac{d^{n-1} y}{dx^{n-1}} + ... + a_n y = F(x)$ is called linear. Such equations are much simpler to solve than typical non-linear ODE's. Though only a few special cases can be solved exactly in terms of elementary functions, there is much that can be said about the solution of a generic linear ODE. A full account would be beyond the scope of this book If $F(x) = 0$ for all x the ODE is called homogeneous.

Fact.

[**author=wikibooks**, file =**text_files/linear_higher_order_ODEs**]
Two useful properties of generic linear equations are Any linear combination of solutions of an homogenous linear equation is also a solution. If we have a solution of a nonhomogenous linear equation and we add any solution of the corresponding homogenous linear equation we get another solution of the nonhomogenous linear equation

Rule 10.3.1.

[**author=wikibooks**, file =**text_files/linear_higher_order_ODEs**]
Variation of constants Suppose we have a linear ODE, $\frac{d^n y}{dx^n} + a_1(x) \frac{d^{n-1} y}{dx^{n-1}} + ... + a_n y = 0$ and we know one solution, $y = w(x)$.

The other solutions can always be written as $y = wz$. This substitution in the ODE will give us terms involving every differential of z upto the $n^t h$, no higher, so we'll end up with an $n^t h$ order linear ODE for z.

We know that z is constant is one solution, so the ODE for z must not contain a z term, which means it will effectively be an $n - 1$th order linear ODE. We will have reduced the order by one.

Lets see how this works in practice.

Example 10.3.2.

[**author=wikibooks**, file =**text_files/linear_higher_order_ODEs**]
Consider $\frac{d^2 y}{dx^2} + \frac{2}{x} \frac{dy}{dx} - \frac{6}{x^2} y = 0$

One solution of this is $y = x^2$, so substitute $y = zx^2$ into this equation.

$\left(x^2 \frac{d^2z}{dx^2} + 2x\frac{dz}{dx} + 2z\right) + \frac{2}{x}\left(x^2\frac{dz}{dx} + 2xz\right) - \frac{6}{x^2}x^2 z = 0$

Rearrange and simplify. $x^2D^2z + 6xDz = 0$ This is first order for Dz. We can solve it to get $z = Ax^{-5} \quad y = Ax^{-3}$

Since the equation is linear we can add this to any multiple of the other solution to get the general solution,

$y = Ax^{-3} + Bx^2$

Rule 10.3.2.

[`author=wikibooks`, file =`text_files/linear_higher_order_ODEs`]
Linear homogenous ODE's with constant coefficents Suppose we have a ODE $(D^n + a_1D^{n-1} + ... + a_{n-1}D + a_0)y = 0$ we can take an inspired guess at a solution (motivate this) $y = e^{px}$ For this function $D^ny = p^ny$ so the ODE becomes $(p^n + a_1p^{n-1} + ... + a_{n-1}p + a_0)y = 0$

y=0 is a trivial solution of the ODE so we can discard it. We are then left with the equation $p^n + a_1p^{n-1} + ... + a_{n-1}p + a_0) = 0$ This is called the characteristic equation of the ODE.

It can have up to n roots, p_1, $p_2 \ldots p_n$, each root giving us a different solution of the ODE.

Because the ODE is linear, we can add all those solution together in any linear combination to get a general solution $y = A_1e^{p_1x} + A_2e^{p_2x} + ... + A_ne^{p_nx}$

To see how this works in practice we will look at the second order case. Solving equations like this of higher order uses the exact same principles only the algrebra is more complex.

Rule 10.3.3.

[`author=wikibooks`, file =`text_files/linear_higher_order_ODEs`]
Second order If the ODE is second order, $D^2y + bDy + cy = 0$ then the characteristic equation is a quadratic, $p^2 + bp + c = 0$ with roots $p_\pm = \frac{-b\pm\sqrt{b^2-4c}}{2}$

What these roots are like depends on the sign of $b^2 - 4c$, so we have three cases to consider.

$1/b^2 > 4c$

In this case we have two different real roots, so we can write down the solution straight away. $y = A_+e^{p_+} + A_-e^{p_-}$

$2/b^2 < 4c$

In this case, both roots are imaginary. We could just put them directly in the formula, but if we are interested in real solutions it is more useful to write them another way.

Defining $k^2 = 4c - b^2$, then the solution is $y = A_+e^{ikx-\frac{bx}{2}} + A_-e^{-ikx-\frac{bx}{2}}$

For this to be real, the A¡nowiki¿'¡/nowiki¿s must be complex conjugates $A_\pm = Ae^{\pm ia}$

Make this substitution and we can write, $y = Ae^{-bx/2}\cos(kx + a)$

If b is positive, this is a damped oscillation.

$3/b^2 = 4c$

In this case the characteristic equation only gives us one root, p=-b/2. We must use another method to find the other solution.

We'll use the method of variation of constants. The ODE we need to solve is, $D^2y - 2pDy + p^2y = 0$ rewriting b and cin terms of the root. From the characteristic equation we know one solution is $y = e^{px}$ so we make the substitution $y = ze^{px}$, giving $(e^{px}D^2z + 2pe^{px}Dz + p^2e^{px}z) - 2p(e^{px}Dz + pe^{px}z) + p^2e^{px}z = 0$ This simplifies to $D^2z = 0$, which is easily solved. We get $z = Ax + B \quad y = (Ax + B)e^{px}$ so the second solution is the first multiplied by x.

Higher order linear constant coefficent ODE's behave similarly an exponential for every real root of the characteristic and a exponent multiplied by a trig factor for every complex conjugate pair, both being multiplied by a polynomial if the root is repeated.

E.g, if the characteristic equation factors to $(p - 1)^4(p - 3)(p^2 + 1)^2 = 0$ the general solution of the ODE will be $y = (A + Bx + Cx^2 + Dx^3)e^x + Ee^{3x} + F\cos(x + a) + Gx\cos(x + b)$

The most difficult part is finding the roots of the characteristic equation.

Rule 10.3.4.

[author=wikibooks, file =text_files/linear_higher_order_ODEs]
Linear nonhomogenous ODE's with constant coefficents First, let's consider the ODE $Dy - y = x$ a nonhomogenous first order ODE which we know how to solve.

Using the integrating factor e^-x we find $y = ce^{-x} + 1 - x$

This is the sum of a solution of the coresponding homogenous equation, and a polynomial.

Nonhomogeneous ODE's of higher order behave similarly.

If we have a single solution, y_p of the nonhomogeneous ODE, called a particular solution, $(D^n + a_1D^{n-1} + \cdots + a_n)y = F(x)$ then the general solution is $y = y_p + y_h$, where y_h is the general solution of the homogeneous ODE.

Find y_p For an arbitary F(x) requires methods beyond the scope of this chapter, but there are some special cases where finding y_p is straightforward.

Remember that in the first order problem y_p for a polynomial F(x) was itself a polynomial of the same order. We can extend this to higher orders.

Example $D^2y + y = x^3 - x + 1$ Consider a particular solution $y_p = b_0 + b_1x + b_2x^2 + x^3$ Substitute for y and collect coefficients $x^3 + b_2x^2 + (6 + b_1)x + (2b_2 + b_0) = x^3 - x + 1$ So $b_2 = 0$, $b_1 = -7$, $b_0 = 1$, and the general solution is $y = a\sin x + b\cos x + 1 - 7x + x^3$

This works because all the derivatives of a polynomial are themselves polynomials.

Two other special cases are $F(x) = P_ne^{kx} \quad y_p(x) = Q_ne^{kx}$ $F(x) = A_n\sin kx + B_n\cos kx \quad y_p(x) = P_n\sin kx + Q_n\cos kx$ where P_n, Q_n, A_n, and B_n are all polynomials of degree n.

Making these substitutions will give a set of simultaneous linear equations for the coefficients of the polynomials.

Rule 10.3.5.

[author=wikibooks, file =text_files/linear_higher_order_ODEs]
Non-Linear ODE's If the ODE is not linear, first check if it is reducible. If it is neither linear nor reducible there is no generic method of solution. You may, with sufficient ingenuity and algebraic skill, be able to transform it into a linear ODE.

If that is not possible, solving the ODE is beyond the scope of this book.

Chapter 11

Vectors

Discussion.

[author=wikibooks, file =text_files/introduction_to_vectors]
In most mathematics courses up until this point, we deal with scalars. These are quantities which only need one number to express. For instance, the amount of gasoline used to drive to the grocery store is a scalar quantity because it only needs one number 2 gallons.

In this unit, we deal with vectors. A vector is a directed line segment – that is, a line segment that points one direction or the other. As such, it has an initial point and a terminal point. The vector starts at the initial point and ends at the terminal point, and the vector points towards the terminal point. A vector is drawn as a line segment with an arrow at the terminal point

The same vector can be placed anywhere on the coordinate plane and still be the same vector – the only two bits of information a vector represents are the magnitude and the direction. The magnitude is simply the length of the vector, and the direction is the angle at which it points. Since neither of these specify a starting or ending location, the same vector can be placed anywhere. To illustrate, all of the line segments below can be defined as the vector with magnitude $\sqrt{32}$ and angle 45 degrees

Multiple locations for the same vector.

It is customary, however, to place the vector with the initial point at the origin as indicated by the blue vector. This is called the standard position.

11.1 Basic vector arithmetic

Discussion.

[author=wikibooks, file =text_files/vector_operations]
In most mathematics courses up until this point, we deal with scalars. These are quantities which only need one number to express. For instance, the amount of gasoline used to drive to the grocery store is a scalar quantity because it only needs one number 2 gallons.

In this unit, we deal with vectors.

Definition 11.1.1.

[author=wikibooks, file =text_files/vector_operations]
A vector is a directed line segment – that is, a line segment that points one direction or the other. As such, it has an initial point and a terminal point. The vector starts at the initial point and ends at the terminal point, and the vector points towards the terminal point. A vector is drawn as a line segment with an arrow at the terminal point

The same vector can be placed anywhere on the coordinate plane and still be the same vector – the only two bits of information a vector represents are the magnitude and the direction. The magnitude is simply the length of the vector, and the direction is the angle at which it points. Since neither of these specify a starting or ending location, the same vector can be placed anywhere. To illustrate, all of the line segments below can be defined as the vector with magnitude $\sqrt{32}$ and angle 45 degrees

It is customary, however, to place the vector with the initial point at the origin as indicated by the blue vector. This is called the standard position.

Comment.

[author=wikibooks, file =text_files/vector_operations]
In standard practice, we don't express vectors by listing the length and the direction. We instead use component form, which lists the height (rise) and width (run) of the vectors. It is written as follows

From the diagram we can now see the benefits of the standard position the two numbers for the terminal point's coordinates are the same numbers for the vector's rise and run. Note that we named this vector u. Just as you can assign numbers to variables in algebra (usually x, y, and z), you can assign vectors to variables in calculus. The letters u, v, and w are usually used, and either boldface or an arrow over the letter is used to identify it as a vector.

When expressing a vector in component form, it is no longer obvious what the magnitude and direction are. Therefore, we have to perform some calculations to find the magnitude and direction.

Definition 11.1.2.

[author=wikibooks, file =text_files/vector_operations]
The **magnitude** of a vector is defined as

$$|\vec{u}| = \sqrt{u_x^2 + u_y^2}$$

where u_x is the width, or run, of the vector u_y is the height, or rise, of the vector. You should recognize this formula as simply the distance formula between two points. It is – the magnitude is the distance between the initial point and the terminal point.

Definition 11.1.3.

[author=wikibooks, file =text_files/vector_operations]
The **direction** of a vector is defined as,

$$\tan\theta = \frac{u_y}{u_x}$$

where θ is the direction of the vector. This formula is simply the tangent formula for right triangles.

Comment.

[author=duckworth, file =text_files/vector_operations]
Note that the definition of direction of a vector assumes that you have fixed x and y axes in the $\mathbb{R}^2$ plane. In more general settings "direction" of a vector is too vague, instead, one would refer more speceficially to "the angle between two vectors."

Definition 11.1.4.

[author=duckworth, file =text_files/vector_operations]
Let $\vec{u} = \begin{bmatrix} u_x \\ u_y \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}$ be any vectors. We define $\vec{u} + \vec{v}$ to be the vector given by

$$\begin{bmatrix} u_x + v_x \\ u_y + v_y \end{bmatrix}.$$

Comment.

[author=wikibooks, file =text_files/vector_operations]
Graphically, adding two vectors together places one vector at the end of the other. This is called tip-to-tail addition The resultant vector, or solution, is the vector drawn from the initial point of the first vector to the terminal point of the second vector when they are drawn tip-to-tail

---

Example 11.1.1.

[author=wikibooks, file =text_files/vector_operations]
For example,

$$\begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Definition 11.1.5.

[author=wikibooks, file =text_files/vector_operations]
Let c be a real number and $\vec{u}$ any vector. We define the **scalar product** $c\vec{u}$ as

the vector:

$$c\vec{u} = \begin{bmatrix} cu_x \\ cu_y \end{bmatrix}$$

Comment.

[author=wikibooks, file =text_files/vector_operations]
Graphically, multiplying a vector by a scalar changes only the magnitude of the vector by that same scalar. That is, multiplying a vector by 2 will "stretch" the vector to twice its original magnitude, keeping the direction the same.

Example 11.1.2.

[author=duckworth, file =text_files/vector_operations]
Note that the length of $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ is $\sqrt{9+25} = \sqrt{34}$. Nowe we calculate $2\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$. Note that the length of $\begin{bmatrix} 6 \\ 10 \end{bmatrix}$ is $\sqrt{36+100} = \sqrt{136} = 2\sqrt{34}$.

Fact.

[author=wikibooks, file =text_files/vector_operations]
Since multiplying a vector by a constant results in a vector in the same direction, we can reason that two vectors are parallel if one is a constant multiple of the other – that is, that $\vec{u}$ is parllel to $\vec{v}$ if $\vec{u} = c\vec{v}$ for some constant c.

Definition 11.1.6.

[author=duckworth, file =text_files/vector_operations]
Let $\vec{u} = \begin{bmatrix} u_x \\ u_y \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}$ be any vectors. We define the **dot product** $\vec{u} \cdot \vec{v}$ to be the real number given by

$$u_x \cdot v_x + u_y \cdot v_y.$$

Comment.

[author=duckworth, file =text_files/vector_operations]
Note, we have used the notation "·" both for multiplying vectors and for multiplying real numbers. We rely on the reader to whether the things being multiplied are vectors or real numbers.

Definition 11.1.7.

[author=wikibooks, file =text_files/vector_operations]

The angle θ between two vectors $\vec{u}$ and $\vec{v}$ is (implicitly) by the equation

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos\theta$$

where θ is the angle difference between the two vectors.

Fact.

[author=duckworth, file =text_files/vector_operations]
Two vectors $\vec{u}$ and $\vec{v}$ are perpendicular to each other if and only if $\vec{u} \cdot \vec{v} = 0$.

Definition 11.1.8.

[author=wikibooks, file =text_files/vector_operations]
A **unit vector** is a vector with a magnitude of 1. The **unit vector of** u is a vector in the same direction as $\vec{u}$, but with a magnitude of 1. In other words, the unit vector of u is given by the formula $\frac{1}{|\vec{u}|}\vec{u}$. The process of finding the unit vector of u is called **normalization**.

Definition 11.1.9.

[author=duckworth, file =text_files/vector_operations]
We define the **standard basis** or **standard unit vectors**. Define the vector $\hat{\imath}$ as $\begin{bmatrix}1\\0\end{bmatrix}$. Thus $\hat{\imath}$ points from the origin directly to the right with a length of 1. Define the vector $\hat{\jmath}$ as $\begin{bmatrix}0\\1\end{bmatrix}$. Thus $\hat{\jmath}$ points from the origin directly up with a length of 1. It may not be obvious to the student why it's even worth giving these vectors names; these vectors are occasionally convenient when writing formulas.

Comment.

[author=duckworth, file =text_files/vector_operations]
Using the standard unit vectors we can write an arbitrary vector $\vec{u}$ this way

$$u = u_x\hat{\imath} + u_y\hat{\jmath}$$

where u_x and u_y are the x and y-components of u, respectively.

Discussion.

[author=wikibooks, file =text_files/polar_coordinates]
Polar coordinates are an alternative two-dimensional coordinate system, which is often useful when rotations are important. Instead of specifying the position along the x and y axes, we specify the distance from the origin, r, and the direction, an angle θ .

Looking at this diagram, we can see that the values of x and y are related to those of r and θ by the equations

$$\begin{array}{ll} x = r\cos\theta & r = \sqrt{x^2+y^2} \\ y = r\sin\theta & \tan\theta = \frac{y}{x} \end{array}$$

Because $\tan^{-1}$ is multivalued, care must be taken to select the right value.

Just as for Cartesian coordinates the unit vectors that point in the x and y directions special, so in polar coordinates the unit vectors that point in the r and θ directions are special.

We will call these vectors $\hat{r}$ and $\hat{\theta}$, pronounced r-hat and theta-hat. Putting a circumflex over a vector this way is often used to mean the unit vector in that direction.

Again, on looking at the diagram we see,

$$\begin{array}{ll} i = \hat{r}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta & \hat{r} = \frac{x}{r}i + \frac{y}{r}j \\ j = \hat{r}\sin\theta + \hat{\boldsymbol{\theta}}\cos\theta & \hat{\boldsymbol{\theta}} = -\frac{y}{r}i + \frac{x}{r}j \end{array}$$

Discussion.

[`author=wikibooks`, file =`text_files/three_dimensional_vectors`]
Two-dimensional Cartesian coordinates as we've discussed so far can be easily extended to three-dimensions by adding one more value z. If the standard (x, y) coordinate axes are drawn on a sheet of paper, the z axis would extend upwards off of the paper.

Similar to the two coordinate axes in two-dimensional coordinates, there are three coordinate planes in space. These are the xy-plane, the yz-plane, and the xz-place. Each plane is the "sheet of paper" that contains both axes the name mentions. For instance, the yz-plane contains both the y and z axes and is perpendicular to the x axis.

Therefore, vectors can be extended to three dimensions by simply adding the z value. For example:

$$\vec{u} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

To faciliate standard form notation, we add another standard unit vector

$$\vec{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Again, both forms (component and standard) are equivalent. For example,

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1\vec{i} + 2\vec{j} + 3\vec{k}$$

Magnitude in three dimensions is the same as in two dimensions, with the addition of a z term in the square root:

$$|\vec{u}| = \sqrt{u_x^2 + u_y^2 + u_z^2}$$

Definition 11.1.10.

[`author=wikibooks`, file =`text_files/three_dimensional_vectors`]
The cross product of two vectors is defined as the following determinant:

$$\vec{u} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix}$$

and is vector.

The cross product of two vectors is at right angles to both vectors. The magnitude of the cross product is the product of the magnitude of the vectors and $\sin(\theta)$ where θ is the angle between the two vectors:

$$|\vec{u} \times \vec{v}| = |\vec{u}||\vec{v}| \sin(\theta).$$

This magnitude is the area of the parallelogram defined by the two vectors.

Fact.

[`author=wikibooks`, file =`text_files/three_dimensional_vectors`]
The cross product is linear and anticommutative. In other words, for any numbers a and b, and any vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$, we have

$$\vec{u} \times (a\vec{v} + b\vec{w}) = a\vec{u} \times \vec{v} + b\vec{u} \times \vec{w}$$

and

$$\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$$

If both vectors point in the same direction, their cross product is zero.

Facts.

[`author=wikibooks`, file =`text_files/three_dimensional_vectors`]
If we have three vectors we can combine them in two ways, a triple scalar product,

$$\vec{u} \cdot (\vec{v} \times \vec{w})$$

and a triple vector product

$$\vec{u} \times (\vec{v} \times \vec{w})$$

The triple scalar product is a determinant

$$\vec{u} \cdot (\vec{v} \times \vec{w}) = \begin{vmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{vmatrix}$$

If the three vectors are listed clockwise, looking from the origin, the sign of this product is positive. If they are listed anticlockwise the sign is negative.

The order of the cross and dot products doesn't matter:

$$\vec{u} \cdot (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \cdot \vec{w}$$

Either way, the absolute value of this product is the volume of the paralllelpiped defined by the three vectors, u, v, and w

The triple vector product can be simplified:

$$\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \cdot \vec{w})\vec{v} - (\vec{u} \cdot \vec{v})\vec{w}$$

This form is easier to do calculations with.

The triple vector product is not associative.

$$\vec{u} \times (\vec{v} \times \vec{w}) \neq (\vec{u} \times \vec{v}) \times \vec{w}.$$

There are special cases where the two sides are equal, but in general the brackets matter and must not be omitted.

Discussion.

[author=wikibooks, file =text_files/three_dimensional_vectors]
We will use r to denote the position of a point.

The multiples of a vector, a all lie on a line through the origin. Adding a constant vector b will shift the line, but leave it straight, so the equation of a line is, $\vec{r} = \vec{a}s + \vec{b}$

This is a parametric equation. The position is specified in terms of the parameter s.

Any linear combination of two vectors, a and b lies on a single plane through the origin, provided the two vectors are not colinear. We can shift this plane by a constant vector again and write $\vec{r} = \vec{a}s + \vec{b}t + \vec{c}$

If we choose a and b to be orthonormal vectors in the plane (i.e unit vectors at right angles) then s and t are cartesian coordiantes for points in the plane.

These parametric equations can be extended to higher dimensions.

Instead of giving parametic equations for the line and plane, we could use constraints. E.g, for any point in the $x-y$-plane $z = 0$.

For a plane through the origin, the single vector normal to the plane, n, is at right angle with every vector in the plane, by defintion, so $\vec{r} \cdot \vec{n} = 0$ is a plane through the origin, normal to n.

For planes not through the origin we get $(\vec{r} - \vec{a}) \cdot \vec{n} = 0 \quad \vec{r} \cdot \vec{n} = a$

A line lies on the intersection of two planes, so it must obey the constraint for both planes, i.e $\vec{r} \cdot \vec{n} = a \quad \vec{r} \cdot \vec{m} = b$

These constraint equations con also be extended to higher dimensions.

Discussion.

[author=wikibooks, file =text_files/three_dimensional_vectors]
For any curve given by vector function of t, $f(t)$, we can define a unit tangent vector t,

$$\vec{t} = \frac{1}{|d\vec{f}/dt|}\frac{d\vec{f}}{dt},$$

where t depends only on the geometry of the curve, not on the parameterisation.

Now, for any unit vector v we have

$$\begin{aligned} 1 &= \vec{v}\cdot\vec{v} \\ 1 &= v_x v_x + v_y v_y + v_z v_z \\ 0 &= 2v_x \dot{v_x} + 2v_y \dot{v_y} + 2v_x \dot{v_y} \\ 0 &= \vec{v}\cdot\dot{\vec{v}} \end{aligned}$$

so v and its derivative are always at right angles.

This lets us define a second unit vector, at right angles to the tangent, which also depends only on the geometry of the curve.

$$\vec{n} = \frac{1}{|d\vec{t}/dt|}\frac{d\vec{t}}{dt},$$

n is called the normal to the curve. The curve lies in its $n-t$-plane near any point. This plane is called the osculating plane.

Since we've got two perpendicular unit vectors we can define a third.

$$\vec{b} = \vec{t}\times\vec{n}$$

This vector is called the binormal. All three of these vectors depend only on the geometry of the curve, which makes them useful when studying that curve.

We can, for example, use them to define curvature.

Discussion.

[`author=wikibooks`, file =`text_files/three_dimensional_vectors`]
Suppose $x = (x(t), y(t), z(t))$. We can use Pythagoras to calculate the length of an infintesimal segment of the curve.

$$\begin{aligned} ds &= \sqrt{dx^2 + dy^2 + dz^2} \\ &= dt\sqrt{v_x^2 + v_y^2 + v_z^2} \end{aligned}$$

where s is the length measured along the curve and v is the derivative of x with respect to t, analogous to velocity.

Integrating this, we get

$$s = \int \sqrt{v_x^2 + v_y^2 + v_z^2}\, dt \quad \frac{ds}{dt} = |\vec{v}|$$

For a circle, $x = (a\cos(t), a\sin(t), 0)$, this gives $\frac{d}{dt} = a$ and the circumference of the circle as $2\pi a$ just as expected.

The curvature of a curve $\vec{x}$ is defined to be

$$\kappa = |\left|\frac{d\vec{x}}{ds}\right|$$

For circles, this is the reciprocal of the radius. E.g

$$\begin{aligned} \kappa &= \left|\tfrac{d}{ds}(\cos t, \sin t)\right| \\ &= \tfrac{dt}{ds}\left|\tfrac{d}{dt}(\cos t, \sin t)\right| \\ &= \tfrac{1}{a}\left|(-\sin t, \cos t)\right| \\ &= \tfrac{1}{a} \end{aligned}$$

We can get the general expression for κ by writing v and a in terms of t and n

$$\begin{aligned}
\vec{v} &= \vec{t}\frac{ds}{dt} \\
\vec{a} &= \frac{d}{dt}\left(\vec{t}\frac{ds}{dt}\right) \\
&= \frac{d^2s}{dt^2}\vec{t} + \frac{ds}{dt}\frac{d\vec{t}}{dt} \\
&= \frac{d^2s}{dt^2}\vec{t} + \left(\frac{ds}{dt}\right)^2\frac{d\vec{t}}{ds} \\
&= \frac{d^2s}{dt^2}\vec{t} + \left(\frac{ds}{dt}\right)^2\kappa\vec{n}
\end{aligned}$$

where the last line follows from the definitions of n and κ.

We can now take the cross product of velocity and acceleration to get

$$\vec{v} \times \vec{a} = \kappa\left(\frac{ds}{dt}\right)^3\vec{b}$$

but b is a unit vector and $|ds/dt| = |v|$ so

$$\kappa = \frac{|\vec{v} \times \vec{a}|}{|\vec{v}|^3}$$

For a two-dimensional curve there is an alternative interpretation of κ. Since t and n are both unit vectors they must be of the form

$$\vec{t} = (\cos\theta, \sin\theta) \quad \vec{t} = (-\sin\theta, \cos\theta)$$

Differentiating these vectors gives

$$\frac{d}{ds}\vec{t} = (-\sin\theta, \cos\theta)\frac{d\theta}{ds} \quad \frac{d}{ds}\vec{t} = (-\cos\theta, -\sin\theta)\frac{d\theta}{ds}.$$

Comparing this with the previous definitions we see that

$$\kappa = \frac{d\theta}{ds} \quad \frac{d}{ds}\vec{t} = \kappa\vec{n} \quad \frac{d}{ds}\vec{n} = -\kappa\vec{t}$$

So for a two-dimensional curve, the curvature is the rate at which the tangent and normal vectors rotate.

A similar expression can be deduced for three dimensional curves.

11.2 Limits and Continuity in Vector calculus

Discussion.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
In your previous study of calculus, we have looked at functions and their behaviour. Most of these functions we have examined have been all in the form $f : \mathbb{R} \to \mathbb{R}$, and only occasional examination of functions of two variables. However, the study of functions of several variables is quite rich in itself, and has applications in several fields.

We write functions of vectors - many variables - as follows $fR^m \to R^n$ and f(x) for the function that maps a vector in R^m to a vector in R^n.

Before we can do calculus in R^n, we must familiarise ourselves with the structure of R^n. We need to know which properties of R can be extended to R^n

Discussion.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Topology in R^n We are already familar with the nature of the regular real number line, which is the set R, and the two-dimensional plane, R^2. This examination of topology in R^n attempts to look at a generalization of the nature of n-dimensional spaces; R, or R^{23}, or R^n.

Discussion.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Lengths and distances If we have a vector in R^2, we can calculate its length using the Pythagorean theorem. For instance, the length of the vector (2, 3) is $\sqrt{2^2+3^2}=\sqrt{13}$

Definition 11.2.1.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
We can generalize this to R^n. We define a vector's length, written —x—, as the square root of the squares of each of its components. That is, if we have a vector $x=(x_1,...,x_n)$, $|\mathbf{x}|=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$

Now that we have established some concept of length, we can establish the distance between two vectors. We define this distance to be the length of the two vectors' difference. We write this distance d(x, y), and it is $d(\mathbf{x},\mathbf{y})=|\mathbf{x}-\mathbf{y}|=\sqrt{\sum(x_i-y_i)^2}$

This distance function is sometimes referred to as a metric. Other metrics arise in different circumstances. The metric we have just defined is known as the Euclidean metric.

Definition 11.2.2.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Open and closed balls In R, we have the concept of an interval, in that we choose a certain number of other points about some central point. For example, the interval [-1, 1] is centered about the point 0, and includes points to the left and right of zero.

In R^2 and up, the idea is a little more difficult to carry on. For R^2, we need to consider points to the left, right, above, and below a certain point. This may be fine, but for R^3 we need to include points in more directions.

We generalize the idea of the interval by considering all the points that are a given, fixed distance from a certain point - now we know how to calculate distances in R^n, we can make our generalization as follows, by introducing the concept of an open ball and a closed ball respectively, which are analogous to the open and closed interval respectively. an open ball $B(\mathbf{a},r)$ is a set in the form $\{x\in R^n|d(x,a)<r\}$ an closed ball $\overline{B}(\mathbf{a},r)$ is a set in the form $\{x\in R^n|d(x,a)\leq r\}$

In R, we have seen that the open ball is simply an open interval centered about the point x=a. In R^2 this is a circle with no boundary, and in R^3 it is a sphere with no outer surface. (What would the closed ball be?)

Neighbourhoods A neighbourhood is an important concept to determine whether a set later, is open or closed. A set N in R^n is called a neighbourhood (usually just abbreviated to nhd) of a in R^n such that a is contained in N, and that for some r, an open ball of radius r about a is a subset of N.

More symbolically, for every $r > 0, x \in N$ when $d(x, a) < r$.

Simply put, all points sufficiently close to a, are also in N. We have some terminology with certain points and their neighbourhoods - a point in a set with a neighbourhood lying completely in that set is known as an interior point of that set. The set of all interior points of a set S is known as the interior of the set S and is written S^o.

Definition 11.2.3.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
Open and closed sets With these ideas now, we can formulate the concept of an open set and a closed set.

We say that a set is open if every point in that set is an interior point of that set, which means that we can construct a neighbourhood of every point in that set. Symbolically, for all $a \in S$, there is a $r > 0$, so all x satisfying $d(x, a) < r$ is in S.

We have the fact that open balls are open sets. With the idea of the complement of a set S being all the points that are not in S, written S^c or S', a closed set is a set with its complement being open.

Example 11.2.1.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
It is important to note that most sets are not open or closed. Think about a box in R^2 with its top and bottom included, and it's left and right sides open - this set is $\{(x, y)||x| < 1and|y| \leq 1\}$.

Definition 11.2.4.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
Limit points A limit point of some set S is a point where, if we construct a neighbourhood about that point, that neighbourhood always contains some other point in S.

Example 11.2.2.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
Here's an example. If $S = \{1/n|n \in Z^+\}$, and we pick the point 0, we can always construct a neighbourhood about 0 which includes some other point of S. This brings up the important point that a limit point need not be in that set. Note that 0 is clearly not in S - but is a limit point of that set.

Definition 11.2.5.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
If we include all the limit points of a set including that set, we call that set the closure of S, and we write it $\overline{S}$.

Comment.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Limit points allow us to also characterize whether a set is open or closed - a set is closed if it contains all its limit points.

Definition 11.2.6.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Boundary points If we have some area, say a field, then the common sense notion of the boundary is the points next to both the inside and outside of the field. For any set S we can define this rigorously by saying the boundary of the set contains all those points such that we can find points both inside and outside the set. We call the set of such points ∂S.

Typically, when it exists the dimension of ∂S is one lower than the dimension of S. e.g the boundary of a volume is a surface and the boundary of a surface is a curve.

This isn't always true but it is true of all the sets we will be using.

Example 11.2.3.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
The boundary of a closed ball in R^2 is the circle surrounding the interior of that ball. In symbols this means that $\partial\overline{B}((0,0),1) = \{(x,y)|x^2+y^2=1\}$

Definition 11.2.7.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Bounded sets A set S is **bounded** if it is contained in some ball centered at 0.

Definition 11.2.8.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Curves and parametrizations If we have a function $f:\mathbb{R}\to\mathbb{R}^n$, we say that the image of f (i.e. the set $\{f(t)|t\in\mathbb{R}\}$) is a curve in $\mathbb{R}^n$ and that f is its parametrization.

Parametrizations are not necessarily unique - for example, $f(t)=(\cos t,\sin t)$ such that $t\in[0,2\pi)$ is one parametrization of the unit circle, and $g(t)=(\cos 7t,\sin 7t)$ such that $t\in[0,2\pi/7)$ is another parameterization.

Definition 11.2.9.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Collision and intersection points Say we have two different curves. It may be important to consider when the two curves cross each other - where they intersect when the two curves hit each other at the same time - where they collide.

Definition 11.2.10.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Intersection points Firstly, we have two parametrizations f(t) and g(t), and we want to find out when they intersect, this means that we want to know when the function values of each parametrization are the same. This means that we need to solve f(t) = g(s) because were seeking the function values independent of the times they intersect.

Example 11.2.4.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
For example, if we have f(t) = (t, 3t) and $g(t) = (t, t^2)$, and we want to find intersection points $f(t) = g(s) = (t, 3t) = (s, s^2)$, t = s and $3t = s^2$ with solutions (t, s) = (0, 0) and (3, 3)

So, the two curves intersect at the points (0, 0) and (3, 9).

However, if we want to know when the points "collide", with f(t) and g(t), we need to know when both the function values and the times are the same, so we need to solve instead f(t) = g(t)

For example, using the same functions as before, f(t) = (t, 3t) and $g(t) = (t, t^2)$, and we want to find collision points $f(t) = g(t)(t, 3t) = (t, t^2)$, t = t and $3t = t^2$ which gives solutions t = 0, 3 So the collision points are (0, 0) and (3, 9).

We may want to do this to actually model physical problems, such as in ballistics.

Definition 11.2.11.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Continuity and differentiability If we have a parametrization $f : \mathbb{R} \to \mathbb{R}^n$, which is built up out of component functions in the form $f(t) = (f_1(t), ..., f_n(t))$, then we say that f is continuous if and only if each component function is also.

In this case the derivative of $f(t)$ is $a_i = (f_1(t), ..., f_n(t))$. This is actually a specific consequence of a more general fact we will see later.

Definition 11.2.12.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Tangent vectors Recall in single-variable calculus that on a curve, at a certain point, we can draw a line that is tangent to that curve at exactly at that point. This line is called a tangent. In the several variable case, we can do something

similar.

We can expect the tangent vector to depend on $f'(t)$ and we know that a line is its own tangent, so looking at a parametrised line will show us precisely how to define the tangent vector for a curve.

An arbitary line is $f(t) = at + b$, with $f_i(t) = a_i t + b_i$, so $f_i(t) = a_i$ and $f(t) = a$, which is the direction of the line, its tangent vector.

Similarly, for any curve, the tangent vector is $f'(t)$.

The gradient of the line $f(t)$ in the one-variable case is $f'(t)$, likewise, the tangent vector to a curve in the several variable case is the vector $f'(t)$ (this vector must not be 0).

Definition 11.2.13.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
Angle between curves We can then formulate the concept of the angle between two curves by considering the angle between the two tangent vectors. If two curves, parametrized by f_1 and f_2 intersect at some point, which means that $f_1(s) = f_2(t) = c$, the angle between these two curves at c is the angle between the tangent vectors $f_1(s)$ and $f_2(t)$ is given by $\arccos \frac{\mathbf{f}_1'(s)-\mathbf{f}_2'(s)}{|\mathbf{f}_1'(s)||\mathbf{f}_2'(t)|}$

Definition 11.2.14.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
Tangent lines With the concept of the tangent vector as being analogous to being the gradient of the line in the one variable case, we can form the idea of the tangent line. Recall that we need a point on the line and its direction.

If we want to form the tangent line to a point on the curve, say p, we have the direction of the line $f'(p)$, so we can form the tangent line $x(t) = p + tf'(p)$

Definition 11.2.15.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
Different parametrizations One such parametrization of a curve is not necessarily unique. Curves can have several different parametrizations. For example, we already saw that the unit circle can be parametrized by $g(t) = (cos(at), \sin(at))$ such that $t \in [0, 2\pi/a)$.

Generally, if f is one parametrization of a curve, and g is another, with $f(t_0) = g(s_0)$ there is a function u(t) such that $u(t_0) = s_0$, and $g(u(t)) = f(t)$ near t_0.

This means, in a sense, the function u(t) "speeds up" the curve, but keeps the curves shape.

Definition 11.2.16.

[`author=wikibooks`, file =`text_files/formal_issues_of_vector_calculus`]
Surfaces A surface in space can be described by the image of a function $f : \mathbb{R}^2 \to \mathbb{R}^n$. We call f the parametrization of that surface.

Example 11.2.5.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
For example, consider the function $f(\alpha, \beta) = \alpha(2, 1, 3) + \beta(-1, 2, 0)$ This describes an infinite plane in R^3. If we restrict α and β to some domain, we get a parallelogram-shaped surface in R^3.

Comment.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Surfaces can also be described explicitly, as the graph of a function $z = f(x, y)$ which has a standard parametrization as $f(x, y) = (x, y, f(x, y))$, or implictly, in the form $f(x, y, z) = c$.

Definition 11.2.17.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Level sets The concept of the level set (or contour) is an important one. If you have a function f(x, y, z), a level set in R^3 is a set of the form $\{(x, y, z) \mid f(x, y, z) = c\}$. Each of these level sets is a surface.

Level sets can be similarly defined in any R^n.

Level sets in two dimensions may be familiar from maps, or weather charts. Each line represents a level set. For example, on a map, each contour represents all the points where the height is the same. On a weather chart, the contours represent all the points where the air pressure is the same.

Discussion.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Intersections of surfaces Different surfaces can intersect and produce curves as well. How can these be found? If the surfaces are simple, we can try and solve the two equations of the surfaces simultaneously.

Discussion.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Limits and continuity Before we can look at derivatives of multivariate functions, we need to look at how limits work with functions of several variables first, just like in the single variable case.

Definition 11.2.18.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
If we have a function $f : \mathbb{R} \to \mathbb{R}^n$, we write

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

if for all positive ϵ, there is a corresponding positive number δ such that $|f(x)-b| < \epsilon$ whenever $|x-a| < \delta$, with $x \neq a$.

Comment.

[author=duckworth, file =text_files/formal_issues_of_vector_calculus]
Definition 11.2.18 means that by making difference between x and a smaller, we can make the difference between $f(x)$ and b as small as we want.

For grammatical convenience we sometimes describe the situation in Definition 11.2.18 in different ways.

We read this definition as "the limit of $f(x)$, as x approaches a, equals b." We also write "$f(x) \to b$ as $x \to a$". We also will write "$\lim_{x\to a} f = b$" (where we leave out the "x" in "$f(x)$"), or even $\lim f = b$ (where we leave out the "$x \to a$"). These abbreviated forms are not used just out of laziness; it's sometimes better to simplify notation by leaving out unnecessary details.

Fact.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Since this is an almost identical formulation of limits in the single variable case, many of the limit rules in the one variable case are the same as in the multivariate case.

Let f and g be functions mapping $\mathbb{R}^m$ to $\mathbb{R}^n$, and $h(x)$ a scalar function mapping $\mathbb{R}^m$ to $\mathbb{R}$. Suppose $\lim_{x\to a} f(x) = b$, $\lim_{x\to a} g(x) = c$, and $\lim_{x\to a} h(x) = H$. Then the following hold:

- $\lim_{x\to a}(f+g) = b+c$,
- $\lim_{x\to a}(h(x)f(x)) = Hb$,
- $\lim_{x\to a}(f\cdot g) = b\cdot c$,
- $\lim_{x\to a}(f\times g) = b\times c$.
- if $n = 1$ and $c \neq 0$, then $\lim_{x\to a} \frac{f}{g} = \frac{b}{c}$.
- If $H \neq 0$ then $\lim_{x\to a} \frac{f}{h} = \frac{b}{H}$

Discussion.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Continuity Again, we can use a similar definition to the one variable case to formulate a definition of continuity for multiple variables.

Definition 11.2.19.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
If $f : \mathbb{R}^m \to \mathbb{R}^n$, then f is continuous at a point a in $\mathbb{R}^m$ if $f(a)$ is defined and

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

Comment.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Just as for functions of one dimension, if f, g are both continuous at $x = a$, then $f + g$, λf (for a scalar λ), $f \cdot g$, and $f \times g$ are also continuous at $x = a$. If $\phi : \mathbb{R}^m \to \mathbb{R}$ is continus at $x = a$, and $\phi(a) \neq 0$, then $\phi(x)f(x)$, f/ϕ are also continuous at $x = a$.

Comment.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
From these facts we also have that if A is some matrix which is $n \times m$ in size, with x in $\mathbb{R}^m$, a function $f(x) = Ax$ is continuous in that the function can be expanded in the form $x_1 a_1 + ... + x_m a_m$, which can be easily verified from the points above.

Fact.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Let $f : \mathbb{R}^m \to \mathbb{R}^n$ and write $f(x)$ in the form $f(x) = (f_1(x), ..., f_n(x)$. Then f is continuous if and only if each f_i is continuous.

Fact.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Finally, if f is continuous at $x = a$, and g is continuos at $f(a)$, then $g(f(x))$ is continuous at $x = a$.

Comment.

[author=wikibooks, file =text_files/formal_issues_of_vector_calculus]
Special note about limits It is important to note that 1-variable functions can have multiple limits too since we are looking at limits of functions of more than one variable, we must note that we can approach a point in more than one direction, and thus, the direction that we approach that point counts in our evaluation of the limit. It may be the case that a limit may exist moving in one direction, but not in another. ex

11.3 Derivatives in vector calculus

Discussion.

[author=wikibooks, file =text_files/derivatives_in_vector_calculus]

Before we define the derivative in higher dimensions, it's worth looking again at the definition of derivative in one variable.

For one variable the definition of the derivative at a point p, is $\lim_{x\to p} \frac{f(x)-f(p)}{x-p} = f(p)$

We cant divide by vectors, so this defintion cant be immediately extended to the multiple variable case. However, we can divide by the absolute value of a vector, so lets rewrite this definition in terms of absolute values

Still needs probably a little more explanation here $\lim_{x\to p} \frac{|f(x)-f(p)-f(p)(x-p)|}{|x-p|} = 0$ after pulling f(p) inside and putting it over a common denominator.

So, how can we use this for the several-variable case?

If we switch all the variables over to vectors and replace the constant,(which performs a linear map in one dimension) with a matrix (which is also a linear map), we have $\lim_{\mathbf{x}\to\mathbf{p}} \frac{|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{p})-A(\mathbf{x}-\mathbf{p})|}{|\mathbf{x}-\mathbf{p}|} = 0$ If this limit exists for some $fR^m \to R^n$, and there is a matrix A which is $m \times n$, we refer to this matrix as being the derivative and we write it as $D_p f$.

A point on terminology - in referring to the action of taking the derivative, we write $D_p f$, but in referring to this matrix itself, it is known as the Jacobian matrix and is also written $J_p f$. More on the Jacobian later.

Discussion.

[author=wikibooks, file =text_files/derivatives_in_vector_calculus]
Affine approximations We say that f is differentiable at p if we have, for x close to p, that $|f(x)-(f(p)+A(x-p)|$ is small compared to $|x-p|$. If this holds then $f(x)$ is approximately equal to $f(p)+A(x-p)$.

We call an expression of the form $g(x)+c$ affine, when $g(x)$ is linear and c is a constant. $f(p)+A(x-p)$ is an affine approximation to $f(x)$.

Discussion.

[author=wikibooks, file =text_files/derivatives_in_vector_calculus]
Jacobian matrix and partial derivatives The Jacobian matrix of a function is in the form $(J_\mathbf{p}\mathbf{f})_{ij} = \left.\frac{\partial f_i}{\partial x_j}\right|_\mathbf{p}$ for a $fR^m \to R^n$, $J_p f$ is a $m \times n$ matrix.

The consequence of this is that if f is differentiable at p, all the partial derivatives of f exist at p.

However, it is possible that all the partial derivatives of a function exist at some point yet that function is not differentiable there.

Discussion.

[author=wikibooks, file =text_files/derivatives_in_vector_calculus]
Continuity and differentiability

Furthermore, if all the partial derivatives exist, and are continuous in some neighbourhood of a point p, then f is differentiable at p. This has the consequence that functions which have their component functions built from continuous functions (such as rational functions, differentiable functions or otherwise), f is differentiable everywhere f is defined.

We use the terminology continuously differentiable for a function differentiable at p which has all its partial derivatives existing and are continuous in some neighbourhood at p.

Discussion.

[`author=wikibooks`, file =`text_files/derivatives_in_vector_calculus`]
Rules of taking Jacobians If $fR^m \to R^n$, and $h(x)R^m \to R$ are differentiable at p $J_{\mathbf{p}}(\mathbf{f}+\mathbf{g}) = J_{\mathbf{p}}\mathbf{f} + J_{\mathbf{p}}\mathbf{g}$ $J_{\mathbf{p}}(h\mathbf{f}) = hJ_{\mathbf{p}}\mathbf{f} + \mathbf{f}(\mathbf{p})J_{\mathbf{p}}h$ $J_{\mathbf{p}}(\mathbf{f}\cdot\mathbf{g}) = \mathbf{g}^T J_{\mathbf{p}}\mathbf{f} + \mathbf{f}^T J_{\mathbf{p}}\mathbf{g}$ Important make sure the order is right - matrix multiplication is not commutative!

Discussion.

[`author=wikibooks`, file =`text_files/derivatives_in_vector_calculus`]
Chain rule The chain rule for functions of several variables is as follows. For $fR^m \to R^n$ and $gR^n \to R^p$, and $g \circ f$ differentiable at p, then the Jacobian is given by $\left(J_{\mathbf{f}(\mathbf{p})}\mathbf{g}\right)\left(J_{\mathbf{p}}\mathbf{f}\right)$ Again, we have matrix multiplication, so one must preserve this exact order. Compositions in one order may be defined, but not necessarily in the other way.

come back to this Higher derivatives If one wishes to take higher-order partial derivatives, we can proceed in two ways if the order is small, calculate the first derivative, and then calculate the derivative of that and so forth, or nest using the chain rule as follows

Discussion.

[`author=wikibooks`, file =`text_files/derivatives_in_vector_calculus`]
Alternate notations For simplicity, we will often use various standard abbreviations, so we can write most of the formulae on one line. This can make it easier to see the important details.

Notation.

[`author=wikibooks`, file =`text_files/derivatives_in_vector_calculus`]
We can abbreviate partial differentials with a subscript, e.g, $\partial_x h(x,y) = \frac{\partial h}{\partial x} \quad \partial_x\partial_y h = \partial_y\partial_x h$ When we are using a subscript this way we will generally use the Heaviside D rather than ∂, $D_x h(x,y) = \frac{\partial h}{\partial x} \quad D_x D_y h = D_y D_x h$ Mostly, to make the formulae even more compact, we will put the subscript on the function itself. $D_x h = h_x \quad h_{xy} = h_{yx}$

If we are using subscripts to label the axes, $x_1, x_2, \ldots$, then, rather than having two layers of subscripts, we will use the number as the subscript.

$h_1 = D_1 h = \partial_1 h = \partial_{x_1} h = \frac{\partial h}{\partial x_1}$

We can also use subscripts for the components of a vector function, $u = (u_x, u_y, u_y)$ or $u = (u_1, u_2 \ldots u_n)$.

If we are using subscripts for both the components of a vector and for partial derivatives we will separate them with a comma.

$u_{x,y} = \frac{\partial u_x}{\partial y}$

The most widely used notation is h_x. Both h_1 and $\partial_1 h$ are also quite widely used whenever the axes are numbered. The notation $\partial_x h$ is used least frequently.

We will use whichever notation best suits the equation we are working with.

Discussion.

[author=wikibooks, file =text_files/derivatives_in_vector_calculus]
Directional derivatives Normally, a partial derivative of a function with respect to one of its variables, say, x_j, takes the derivative of that "slice" of that function parallel to the x_jth axis. needs pic

More precisely, we can think of cutting a function $f(x_1, ..., x_n)$ in space along the x_jth axis, with keeping everything but the x_j variable constant.

From the definition, we have the partial derivative at a point p of the function along this slice as $\frac{\partial \mathbf{f}}{\partial x_j} = \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{p}+t\mathbf{e}_j)-\mathbf{f}(\mathbf{p})}{t}$

provided this limit exists.

Discussion.

[author=wikibooks, file =text_files/derivatives_in_vector_calculus]
Instead of the basis vector, which corresponds to taking the derivative along that axis, we can pick a vector in any direction (which we usually take as being a unit vector), and we take the directional derivative of a function as $\frac{\partial \mathbf{f}}{\partial \mathbf{d}} = \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{p}+t\mathbf{d})-\mathbf{f}(\mathbf{p})}{t}$ where d is the direction vector.

If we want to calculate directional derivatives, calculating them from the limit definition is rather painful, but, we have the following if $fR \to R$ is differentiable at a point p, —p—=1, $\frac{\partial \mathbf{f}}{\partial \mathbf{d}} = D_{\mathbf{p}}\mathbf{f}(\mathbf{d})$

There is a closely related formulation which well look at in the next section.

11.4 Div, Grad, Curl, and other operators

Definition 11.4.1.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Gradient vectors The partial derivatives of a scalar tell us how much it changes if we move along one of the axes. What if we move in a different direction?

Well call the scalar f, and consider what happens if we move an infintesimal direction dr=(dx,dy,dz), using the chain rule. $\mathbf{df} = dx\frac{\partial f}{\partial x} + dy\frac{\partial f}{\partial y} + dz\frac{\partial f}{\partial z}$

This is the dot product of dr with a vector whose components are the partial derivatives of f, called the gradient of f

$\operatorname{grad} \mathbf{f} = \nabla \mathbf{f} = \left(\frac{\partial \mathbf{f}(\mathbf{p})}{\partial x_1}, \cdots, \frac{\partial \mathbf{f}(\mathbf{p})}{\partial x_n} \right)$

We can form directional derivatives at a point p, in the direction d then by taking the dot product of the gradient with d $\frac{\partial \mathbf{f}(\mathbf{p})}{\partial \mathbf{d}} = \mathbf{d} \cdot \nabla \mathbf{f}(\mathbf{p})$.

Notice that grad f looks like a vector multiplied by a scalar. This particular combination of partial derivatives is commonplace, so we abbreviate it to $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$

We can write the action of taking the gradient vector by writing this as an operator. Recall that in the one-variable case we can write d/dx for the action of taking the derivative with respect to x. This case is similar, but ∇ acts like a vector.

We can also write the action of taking the gradient vector as $\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots \frac{\partial}{\partial x_n} \right)$

Comment.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Properties of the gradient vector Let f be a vector function and p a point in the domain of f. Then $\nabla f(p)$ is a vector pointing in the direction of steepest slope of f. Therefore $|\nabla f(p)|$ is the rate of change of that slope at that point.

Example 11.4.1.

[author=wikibooks, file =text_files/gradients_divergence_curl]
For example, if we consider $h(x, y) = x^2 + y^2$. The level sets of h are concentric circles, centred on the origin, and $\nabla h = (h_x, h_y) = 2(x, y) = 2\mathbf{r}$ grad h points directly away from the origin, at right angles to the contours.

Along a level set, $(\nabla f)(p)$ is perpendicular to the level set $\{x | f(x) = f(p) \text{ at } x = p\}$.

If dr points along the contours of f, where the function is constant, then df will be zero. Since df is a dot product, that means that the two vectors, df and grad f, must be at right angles, i.e the gradient is at right angles to the contours.

Fact.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Algebraic properties Like d/dx, ∇ is linear. For any pair of constants, a and b, and any pair of scalar functions, f and g $\frac{d}{dx}(af + bg) = a\frac{d}{dx}f + b\frac{d}{dx}g \quad \nabla(af + bg) = a\nabla f + b\nabla g$

Since its a vector, we can try taking its dot and cross product with other vectors, and with itself.

Definition 11.4.2.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Divergence If the vector function u maps $\mathbb{R}^n$ to itself, then we can take the dot

product of u and ∇ . This dot product is called the **divergence**. In symbols $\operatorname{diverge} u = \nabla u \cdot u = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \cdots \frac{\partial u_n}{\partial x_n}$

Comment.

[author=wikibooks, file =text_files/gradients_divergence_curl]
$\operatorname{diverge} v$ tells us how much u is converging or diverging. It is positive when the vector is diverging from some point, and negative when the vector is converging on that point.

Example 11.4.2.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Define the vector function $v = (1 + x^2, xy)$. Then $\operatorname{diverge} v = 3x$, which is positive to the right of the origin, where v is diverging, and negative to the left of the origin, where v is diverging.

Fact.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Like $\frac{d}{dx}$ and ∇, diverge is linear. In symbols: if u and v are vector functions and a and b are scalars, then $\nabla \cdot (au + bv) = a\nabla u \cdot u + b\nabla v \cdot v$

Comment.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Later in this chapter we will see how the divergence of a vector function can be integrated to tell us more about the behaviour of that function.

Comment.

[author=wikibooks, file =text_files/gradients_divergence_curl]
To find the divergence we took the dot product of ∇ and a vector with ∇ on the left. If we reverse the order we get $u \cdot \nabla = u_x D_x + u_y D_y + u_z D_z$

To see what this means consider $i \cdot \nabla$ This is D_x, the partial differential in the i direction. Similarly, $u \cdot \nabla$ is the the partial differential in the u direction, multiplied by —u—

Definition 11.4.3.

[author=wikibooks, file =text_files/gradients_divergence_curl]
If u is a three-dimensional vector function on R^3 then we can take its cross product with ∇u. This cross product is called the **curl**. In symbols:

$$\operatorname{curl} u = \nabla u \times u = \begin{vmatrix} i & j & k \\ D_x & D_y & D_z \\ u_x & u_y & u_z \end{vmatrix}$$

The curl of u tells us if the vector u is rotating around a point. The direction of $\operatorname{curl} u$ is the axis of rotation.

We can treat vectors in two dimensions as a special case of three dimensions, with $u_z = 0$ and $D_z u = 0$. We can then extend the definition of curl u to two-dimensional vectors and obtain $\operatorname{curl} u = D_y u_x - D_x u_y$. This two dimensional curl is a scalar. In four, or more, dimensions there is no vector equivalent to the curl.

Example 11.4.3.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Consider the function u defined by $u = (-y, x)$. These vectors are tangent to circles centred on the origin, so appear to be rotating around it anticlockwise. It is easy to calculate $\operatorname{curl} u = D_y(-y) - D_x x = -2$.

Example 11.4.4.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Consider the function u defined by $u = (-y, x - z, y)$, which is similar to the previous example. An easy calculation shows that

$$\operatorname{curl} u = \begin{vmatrix} i & j & k \\ D_x & D_y & D_z \\ -y & x-z & y \end{vmatrix} = 2i + 2k.$$

This u is rotating round the axis $i + k$.

Comment.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Later in this chapter we will see how the curl of a vector function can be integrated to tell us more about the behaviour of that function.

Rules 11.4.1.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Product and chain rules Just as with ordinary differentiation, there are product rules for ∇, diverge and curl.

Let g be a scalar and v a vector function.

The divergence of gv is $\nabla(gv) \cdot gv = g\nabla \cdot v + (v \cdot \nabla)g$.

The curl of gv is $\nabla \times (gv) = g(\nabla \times v) + (\nabla g) \times v$.

Let u and v be two vector functions.

The gradient of $u \cdot v$ is $\nabla(u \cdot v) = u \times (\nabla \times v) + v \times (\nabla \times u) + (u \cdot \nabla)v + (v \cdot \nabla)u$.

The divergence of $u \times v$ is $\nabla \cdot (u \times v) = v \cdot (\nabla \times u) - u \cdot (\nabla \times v)$.

The curl of $u \times v$ is $\nabla \times (u \times v) = (v \cdot \nabla)u - (u \cdot \nabla)v + u(\nabla \cdot v) - v(\nabla \cdot u)$.

Rules 11.4.2.

[author=wikibooks, file =text_files/gradients_divergence_curl]
We can also write chain rules. In the general case, when both functions are vectors and the composition is defined, we can use the Jacobian defined earlier. $\nabla u(v)|_r = J_v \ \nabla v|_r$ where J_u is the Jacobian of u at the point v.

Normally J is a matrix but if either the range or the domain of u is R^1 then it becomes a vector. In these special cases we can compactly write the chain rule using only vector notation.

If g is a scalar function of a vector and h is a scalar function of g then $\nabla h(g) = \frac{dh}{dg}\nabla g$ If g is a scalar function of a vector then $\nabla = (\nabla g)\frac{d}{dg}$ This substitution can be made in any of the equations containing ∇.

Definition 11.4.4.

[author=wikibooks, file =text_files/gradients_divergence_curl]
Second order differentials We can also consider dot and cross products of ∇ with itself, whenever they can be defined. Once we know how to simplify products of two ∇'s well know out to simplify products with three or more.

The divergence of the gradient of a scalar f is $\nabla^2 f(x_1, x_2, \ldots x_n) = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \ldots + \frac{\partial^2 f}{\partial x_n^2}$

This combination of derivatives is the **Laplacian** of f. It is commmonplace in physics and multidimensional calculus because of its simplicity and symmetry.

Discussion.

[author=wikibooks, file =text_files/gradients_divergence_curl]
We can also take the Laplacian of a vector, $\nabla^2 u(x_1, x_2, \ldots x_n) = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \ldots + \frac{\partial^2 u}{\partial x_n^2}$

The Laplacian of a vector is not the same as the divergence of its gradient $\nabla(\nabla \cdot u) - \nabla^2 u = \nabla \times (\nabla \times u)$

Both the curl of the gradient and the divergence of the curl are always zero. $\nabla \times \nabla f = 0 \quad \nabla \cdot (\nabla \times u) = 0$

This pair of rules will prove useful.

11.5 Integration in vector calculus

Discussion.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
We have already considered differentiation of functions of more than one variable, which leads us to consider how we can meaningfully look at integration.

In the single variable case, we interpret the definite integral of a function to mean the area under the function. There is a similar interpretation in the multiple variable case for example, if we have a paraboloid in R^3, we may want to look at the integral of that paraboloid over some region of the xy plane, which will be the volume under that curve and inside that region.

Definition 11.5.1.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Riemann sums When looking at these forms of integrals, we look at the Riemann sum. Recall in the one-variable case we divide the interval we are integrating over into rectangles and summing the areas of these rectangles as their widths get smaller and smaller. For the multiple-variable case, we need to do something similar, but the problem arises how to split up R^2, or R^3, for instance.

To do this, we extend the concept of the interval, and consider what we call a n-interval. An n-interval is a set of points in some rectangular region with sides of some fixed with in each dimension, that is, a set in the form $\{x \in R^n \mid a_i \leq x_i \leq b_i \text{ with } i = 0, ..., n\}$, and its area/size/volume (which we simply call its measure to avoid confusin) is the product of the lengths of all its sides.

So, an n-interval in R^2 could be some rectangular partition of the plane, such as $\{(x, y) \mid x \in [0, 1] and y \in [0, 2]|\}$. Its measure is 2.

If we are to consider the Riemann sum now in terms of sub-n-intervals of a region Ω, it is $\sum_{iS_i \subset \Omega} f(x_i^*) m(S_i)$ where $m(S_i)$ is the measure of the division of Ω into k sub-n-intervals S_i, and x_i^* is a point in S_i. The index is important - we only perform the sum where S_i falls completely within Ω - any S_i that is not completely contained in Ω we ignore.

As we take the limit as k goes to infinity, that is, we divide up Ω into finer and finer sub-n-intervals, and this sum is the same no matter how we divide up Ω , we get the integral of f over Ω which we write $\int_\Omega f$ For two dimensions, we may write $\int \int_\Omega f$ and likewise for n dimensions.

Iterated integrals Thankfully, we need not always work with Riemann sums every time we want to calculate an integral in more than one variable. There are some results that make life a bit easier for us.

For R^2, if we have some region bounded between two functions of the other variable (so two functions in the form f(x) = y, or f(y) = x), between a constant boundary (so, between x = a and x =b or y = a and y = b), we have $\int_a^b \int_{f(x)}^{g(x)} h(x, y)\, dx$

An important theorem (called Fubinis theorem) assures us that this integral is the same as $\int \int_\Omega f$.

Definition 11.5.2.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Parametric integrals If we have a vector function, u, of a scalar parameter, s, we can integrate with respect to s simply by integrating each component of u seperately.

$v(s) = \int u(s)\, ds \Rightarrow v_i(s) = \int u_i(s)\, ds$

Similarly, if u is given a function of vector of parameters, s, lying in R^n, integration with respect to the parameters reduces to a multiple integral of each component.

Definition 11.5.3.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Line integrals In one dimension, saying we are integrating from a to b uniquely specifies the integral.

In higher dimensions, saying we are integrating from a to b is not sufficient. In general, we must also specify the path taken between a and b.

We can then write the integrand as a function of the arclength along the curve, and integrate by components.

Example 11.5.1.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Given a scalar function $h(r)$ we write

$$\int_C h(r)\,dr = \int_C h(r)\frac{dr}{ds}\,ds = \int_C h(r(s))t(s)\,ds$$

where C is the curve being integrated along, and t is the unit vector tangent to the curve.

Rule 11.5.1.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
There are some particularly natural ways to integrate a vector function, u, along a curve,

$$\int_C u\,ds \quad \int_C u\cdot dr \quad \int_C u\times dr \quad \int_C u\cdot nds$$

where the third possibility only applies in 3 dimensions.

Again, these integrals can all be written as integrals with respect to the arclength, s.

$$\int_C u\cdot dr = \int_C u\cdot t\,ds \text{ or } \int_C u\times dr = \int_C u\times t\,ds$$

If the curve is planar and u a vector lieing in the same plane, the second integral can be usefully rewritten. Say, $u = u_t t + u_n n + u_b b$ where t, n, and b are the tangent, normal, and binormal vectors uniquely defined by the curve.

Then $u \times t = -bu_n + nu_b$

For the 2-d curves specified b is the constant unit vector normal to their plane, and u_b is always zero.

Therefore, for such curves, $\int_C u\times dr = \int_C u\cdot nds$

Discussion.

[author=wikibooks, file =text_files/integration_in_vector_calculus]

Inverting differentials We can use line integrals to calculate functions with specified divergence, gradient, or curl.

If grad V = u $V(p) = \int_{p_0}^{p} u \cdot dr + h(p)$ where h is any function of zero gradient and curl u must be zero.

If div u = V $u(p) = \int_{p_0}^{p} V\, dr + w(p)$ where w is any function of zero divergence.

If curl u = v $u(p) = \frac{1}{2}\int_{p_0}^{p} v \times dr + w(p)$ where w is any function of zero curl.

Example 11.5.2.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
For example, if $V = r^2$ then $\nabla V = 2(x, y, z) = 2r$ and

$$\begin{aligned}\int_0^r 2u \cdot du &= \int_0^r 2\,(udu + vdv + wdw) \\ &= u^2\Big|_0^r + v^2\Big|_0^r + w^2\Big|_0^r \\ &= x^2 + y^2 + z^2 = r^2\end{aligned}$$

so this line integral of the gradient gives the original function.

Example 11.5.3.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Similarly, if $v = k$ then $u(p) = \int_{p_0}^{p} k \times dr$

Consider any curve from 0 to $p = (x, y, z)$, given by $r = r(s)$ with $r(0) = 0$ and $r(S) = p$ for some S, and do the above integral along that curve.

$$\begin{aligned}u(p) &= \int_0^S k \times \frac{dr}{ds}\, ds \\ &= \int_0^S \left(\frac{dr_x}{ds} j - \frac{dr_y}{ds} i\right) ds \\ &= j\int_0^S \frac{dr_x}{ds}\, ds - i\int_0^S \frac{dr_y}{ds}\, ds \\ &= jr_x(s)\Big|_0^S - ir_y(s)\Big|_0^S \\ &= p_x j - p_y i = xj - yi\end{aligned}$$

The curl of u is

$$\frac{1}{2}\begin{vmatrix} i & j & k \\ D_x & D_y & D_z \\ -y & x & 0 \end{vmatrix} = k = v$$

as expected.

Comment.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
We will soon see that these three integrals do not depend on the path, apart from a constant.

Discussion.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Surface and Volume Integrals Just as with curves, it is possible to parameterise surfaces then integrate over those parameters without regard to geometry of the surface.

That is, to integrate a scalar function V over a surface A parameterised by r and s we calculate

$$\int_A V(x,y,z)\,dS = \int\int_A V(r,s)\det J\,drds$$

where J is the Jacobian of the tranformation to the parameters.

To integrate a vector this way, we integrate each component seperately.

However, in three dimensions, every surface has an associated normal vector n, which can be used in integration. We write

$$dS = ndS.$$

For a scalar function V and a vector function v this gives us the integrals

$$\int_A V\,dS, \quad \int_A v\cdot dS, \quad \int_A v\times dS$$

These integrals can be reduced to parametric integrals but, written this way, it is clear that they reflect more of the geometry of the surface.

When working in three dimensions, dV is a scalar, so there is only one option for integrals over volumes.

Discussion.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Gausss divergence theorem We know that, in one dimension, $\int_a^b Df dx = f|_a^b$ Integration is the inverse of differentiation, so integrating the differential of a function returns the original function.

This can be extended to two or more dimensions in a natural way, drawing on the analogies between single variable and multivariable calculus.

The analog of D is ∇ , so we should consider cases where the integrand is a divergence.

Instead of integrating over a one-dimensional interval, we need to integrate over a n-dimensional volume.

In one dimension, the integral depends on the values at the edges of the interval, so we expect the result to be connected with values on the boundary.

This suggests the following theorem.

Theorem 11.5.1.
[author=wikibooks, file =text_files/integration_in_vector_calculus]
$\int_V \nabla\cdot u\,dV = \int_{\partial V} n\cdot udS$

Comment.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
This is indeed true, for vector fields in any number of dimensions.

This is called Gausss theorem.

Theorem 11.5.2.
[author=wikibooks, file =text_files/integration_in_vector_calculus]
There are two other, closely related, theorems for grad and curl

$$\int_V \nabla u \, dV = \int_{\partial V} u n dS,$$

and

$$\int_V \nabla \times u \, dV = \int_{\partial V} n \times u dS,$$

with the last theorem only being valid where curl is defined.

Discussion.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
Stokes curl theorem These theorems also hold in two dimensions, where they relate surface and line integrals. Gausss divergence theorem becomes

---

Theorem 11.5.3.
[author=wikibooks, file =text_files/integration_in_vector_calculus]

$$\int_S \nabla \cdot u \, dS = \oint_{\partial S} n \cdot u ds$$

where s is arclength along the boundary curvw and the vector n is the unit normal to the curve that lies in the surface S, i.e in the tangent plane of the surface at its boundary, which is not necessarily the same as the unit normal associated with the boundary curve itself.

Theorem 11.5.4.
[author=wikibooks, file =text_files/integration_in_vector_calculus]
Similarly, we get

$$\int_s \nabla \times u \, ds = \int_C n \times u ds \quad (1),$$

where C is the boundary of S.

Comment.

[author=wikibooks, file =text_files/integration_in_vector_calculus]
In the last theorem the integral does not depend on the surface S.

To see this, suppose we have different surfaces, S_1 and S_2, spanning the same curve C, then by switching the direction of the normal on one of the surfaces we can write

$$\int_{S_1+S_2} \nabla \times u \, dS = \int_S \nabla \times u \, dS - \int_S \nabla \times u \, dS \quad (2).$$

The left hand side is an integral over a closed surface bounding some volume V so we can use Gausss divergence theorem.

$\int_{S_1+S_2} \nabla \times u \, dS = \int_V \nabla \cdot \nabla \times u \, dV$

but we know this integrand is always zero so the right hand side of (2) must always be zero, i.e the integral is independant of the surface.

This means we can choose the surface so that the normal to the curve lieing in the surface is the same as the curves intrinsic normal

Then, if u itself lies in the surface, we can write

$u = (u \cdot n) \, n + (u \cdot t) \, t$

just as we did for line integrals in the plane earlier, and substitute this into (1) to get the following.

Stokes's Curl Theorem 11.5.5.
[author=wikibooks, file =`text_files/integration_in_vector_calculus`]

$$\int_S \nabla \times u \, dS = \int_C u \cdot dr$$

Chapter 12

Partial Differential Equations

Discussion.

[author=wikibooks, file =text_files/introduction_partial_diffeqs]
Any partial differential equation of the form $h_1 \frac{\partial u}{\partial x_1} + h_2 \frac{\partial u}{\partial x_2} \cdots + h_n \frac{\partial u}{\partial x_n} = b$ where h_1, h_2, $\ldots h_n$, and b are all functions of both u and R^n can be reduced to a set of ordinary differential equations.

To see how to do this, we will first consider some simpler problems.

12.1 Some simple partial differential equations

Discussion.

[author=wikibooks, file =text_files/partial_diffeqs]
We will start with the simple PDE $u_z(x, y, z) = u(x, y, z)$ (1) Because u is only differentiated with respect to z, for any fixed x and y we can treat this like the ODE, du/dz=u. The solution of that ODE is ce^z, where c is the value of u when z=0, for the fixed x and y

Therefore, the solution of the PDE is $u(x, y, z) = u(x, y, 0)e^z$

Instead of just having a constant of integration, we have an arbitary function. This will be true for any PDE.

Notice the shape of the solution, an arbitary function of points in the xy, plane, which is normal to the z axis, and the solution of an ODE in the z direction.

Discussion.

[author=wikibooks, file =text_files/partial_diffeqs]
Now consider the slightly more complex PDE $a_x u_x + a_y u_y + a_z u_z = h(u)$ (2) where h can be any function, and each a is a real constant.

We recognise the left hand side as being $a \cdot \nabla$, so this equation says that the differential of u in the a direction is $h(u)$. Comparing this with the first equation suggests that the solution can be written as an arbitary function on the plane normal to a combined with the solution of an ODE.

Remembering from earlier that any vector r can be split up into components parallel and perpendicular to a, $r = r_\perp + r_\parallel = \left(r - \frac{(r\cdot a)a}{|a|^2}\right) + \frac{(r\cdot a)a}{|a|^2}$ we will use this to split the components of r in a way suggested by the analogy with (1).

Lets write $r = (x, y, z) = r_\perp + sa \quad s = \frac{r\cdot a}{a\cdot a}$ and substitute this into (2), using the chain rule. Because we are only differentiating in the a direction, adding any function of the perpendicular vector to s will make no difference.

First we calculate grad s, for use in the chain rule, $\nabla s = \frac{a}{a^2}\frac{d}{ds}$

On making the substitution into (2), we get, $h(u) = a\cdot\nabla s\frac{d}{ds}u(s) = \frac{a\cdot a}{a\cdot a}\frac{d}{ds}u(s) = \frac{du}{ds}$ which is an ordinary differential equation with the solution $s = c(r_\perp) + \int^u \frac{dt}{h(t)}$

The constant c can depend on the perpendicular components, but not upon the parallel coordinate. Replacing s with a monotonic scalar function of s multiplies the ODE by a function of s, which doesn't affect the solution.

Example 12.1.1.

[**author=wikibooks**, file =**text_files/partial_diffeqs**]
Consider the equation: $u(x,t)_x = u(x,t)_t$

For this equation, a is $(1, -1)$, $s = x - t$, and the perpendicular vector is $(x + t)(1, 1)$. The reduced ODE is $du/ds = 0$ so the solution is $u = f(x + t)$.

To find f we need initial conditions on u. Are there any constraints on what initial conditions are suitable?

Consider, if we are given $u(x, 0)$, this is exactly $f(x)$, $u(3t, t)$, this is $f(4t)$ and $f(t)$ follows immediately $u(t^3 + 2t, t)$, this is $f(t^3 + 3t)$ and $f(t)$ follows, on solving the cubic. Consider $u(-t, t)$, then this is $f(0)$, so if the given function isn't constant we have a inconsistancy, and if it is the solution isn't specified off the initial line.

Similarly, if we are given u on any curve which the lines $x + t = c$ intersect only once, and to which they are not tangent, we can deduce f.

Derivation.

[**author=wikibooks**, file =**text_files/partial_diffeqs**]
For any first order PDE with constant coefficients, the same will be true. We will have a set of lines, parallel to $r = at$, along which the solution is gained by integrating an ODE with initial conditions specified on some surface to which the lines arent tangent.

If we look at how this works, well see we havent actually used the constancy of a, so lets drop that assumption and look for a similar solution.

The important point was that the solution was of the form $u = f(x(s), y(s))$, where $(x(s), y(s))$ is the curve we integrated along – a straight line in the previous case. We can add constant functions of integration to s without changing this form.

Consider a PDE, $a(x,y)u_x + b(x,y)u_y = c(x,y,u)$ For the suggested solution, $u = f(x(s), y(s))$, the chain rule gives $\frac{du}{ds} = \frac{dx}{ds}u_x + \frac{dy}{ds}u_y$ Comparing coefficients then gives $\frac{dx}{ds} = a(x,y) \quad \frac{dy}{ds} = b(x,y) \quad \frac{du}{ds} = c(x,y,u)$ so weve reduced our original PDE to a set of simultaneous ODEs. This procedure can be reversed.

The curves $(x(s), y(s))$ are called characteristics of the equation.

Example 12.1.2.

[author=wikibooks, file =text_files/partial_diffeqs]
Solve $yu_x = xu_y$ given $u = f(x)$ for $x \geq 0$. The ODEs are $\frac{dx}{ds} = y \quad \frac{dy}{ds} = -x \quad \frac{du}{ds} = 0$ subject to the initial conditions at $s = 0$, $x(0) = r \quad y(0) = 0 \quad u(0) = f(r) \quad r \geq 0$ This ODE is easily solved, giving $x(s) = r\cos s \quad y(s) = \sin s \quad u(s) = f(r)$ so the characteristics are concentric circles round the origin, and in polar coordinates $u(r, \theta) = f(r)$.

Considering the logic of this method, we see that the independance of a and b from u has not been used either, so that assumption too can be dropped, giving the general method for equations of this quasilinear form.

12.2 Quasilinear partial differential equations

Discussion.

[author=wikibooks, file =text_files/quasilinear_partial_diffeqs]
Summarising the conclusions of the last section, to solve a PDE $a_1(u, \mathbf{x})\frac{\partial u}{\partial x_1} + a_2(u, \mathbf{x})\frac{\partial u}{\partial x_2} \cdots + a_n(u, \mathbf{x})\frac{\partial u}{\partial x_n} = b(u, \mathbf{x})$ subject to the initial condition that on the surface, $(x_1(r_1, \ldots, r_n - 1, \ldots x_n(r_1, \ldots, r_n - 1)$, $u = f(r_1, \ldots, r_n - 1)$ –this being an arbitary paremetrisation of the initial surface–

We transform the equation to the equivalant set of ODEs, $\frac{dx_1}{ds} = a_1 \ \ldots \ \frac{dx_n}{ds} = a_n \quad \frac{du}{ds} = b$ subject to the initial conditions $x_i(0) = f(r_1, \ldots, r_{n-1}) \quad u = f(r_1, r_2, \ldots r_{n-1})$ Solve the ODEs, giving x_i as a function of s and the r_i. Invert this to get s and the r_i as functions of the x_i. Substitute these inverse functions into the expression for u as a function of s and the r_i obtained in the second step.

Both the second and third steps may be troublesome.

The set of ODEs is generally non-linear and without analytical solution. It may even be easier to work with the PDE than with the ODEs.

In the third step, the r_i together with s form a coordinate system adapted for the PDE. We can only make the inversion at all if the Jacobian of the transformation to Cartesian coordinates is not zero, $\begin{vmatrix} \frac{\partial x_1}{\partial r_1} & \cdots & \frac{\partial x_1}{\partial r_{n-1}} & a_1 \\ \vdots & \ddots & & \vdots \\ \frac{\partial x_n}{\partial r_1} & \cdots & \frac{\partial x_n}{\partial r_{n-1}} & a_n \end{vmatrix} \neq 0$ This is equivalent to saying that the vector $(a_1, \ldots, a_n)$ is never in the tangent plane to a surface of constant s.

If this condition is not false when s=0 it may become so as the equations are integrated. We will soon consider ways of dealing with the problems this can cause.

Even when it is technically possible to invert the algebraic equations it is obvious inconvenient to do so.

Example 12.2.1.

[author=wikibooks, file =text_files/quasilinear_partial_diffeqs]
To see how this works in practice, we will a/ consider the PDE, $uu_x + u_y + u_t = 0$ with generic initial condition, $u = f(x, y)$ on $t = 0$

Naming variables for future convenience, the corresponding ODEs are $\frac{dx}{d\tau} = u \quad \frac{dy}{d\tau} = 1 \quad \frac{dz}{d\tau} = 1 \quad \frac{du}{d\tau} = 0$ subject to the initial conditions at $\tau = 0$, $x = r \quad y = s \quad t = 0 \quad u = f(r, s)$

These ODEs are easily solved to give $x = r + f(r, s)\tau \quad y = s + \tau \quad t = \tau \quad u = f(r, s)$

These are the parametric equations of a set of straight lines, the characteristics.

The determinant of the Jacobian of this coordinate transformation is $\begin{vmatrix} 1 + \tau\frac{\partial f}{\partial r} & \tau\frac{\partial f}{\partial s} & f \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = 1 + \tau\frac{\partial f}{\partial r}$

This determinant is 1 when t=0, but if f_r is anywhere negative this determinant will eventually be zero, and this solution fails.

In this case, the failure is because the surface $sf_r = -1$ is an envelope of the characteristics.

For arbitary f we can invert the transformation and obtain an implicit expression for u $u = f(x - tu, y - x)$ If f is given this can be solved for u.

Example 12.2.2.

[author=wikibooks, file =text_files/quasilinear_partial_diffeqs]
Consider the form of equation $ax = f(x, y)$. The implicit solution is $u = a(x - tu) \Rightarrow u = \frac{ax}{1+at}$ This is a line in the u-x plane, rotating clockwise as t increases. If a is negative, this line eventually become vertical. If a is positive, this line tends towards u=0, and the solution is valid for all t.

Example 12.2.3.

[author=wikibooks, file =text_files/quasilinear_partial_diffeqs]
Consider the form of equation $f(x, y) = x^2$. The implicit solution is $u = (x - tu)^2 \Rightarrow u = \frac{1+2tx-\sqrt{1+4tx}}{2t^2}$ which looks like [[ImageQuasilinearPDEexample1.png—equation solution]] This solution clearly fails when $1 + 4tx < 0$, which is just when $sf_r = -1$. For any t¿0 this happens somewhere. As t increases this point of failure moves toward the origin.

Notice that the point where u=0 stays fixed. This is true for any solution of this equation, whatever f is.

We will see later that we can find a solution after this time, if we consider discontinuous solutions. We can think of this as a shockwave.

Example 12.2.4.

[author=wikibooks, file =text_files/quasilinear_partial_diffeqs]
Consider the form of equation $f(x, y) = \sin(xy)$. The implicit solution is $u(x, y, t) = \sin((x - tu)(y - x))$ and we can not solve this explitly for u. The best we can manage is a numerical solution of this equation.

Example 12.2.5.

[author=wikibooks, file =text_files/quasilinear_partial_diffeqs]
/We can also consider the closely related PDE $uu_x + u_y + u_t = y$ The corresponding ODEs are $\frac{dx}{d\tau} = u \quad \frac{dy}{d\tau} = 1 \quad \frac{dz}{d\tau} = 1 \quad \frac{du}{d\tau} = y$ subject to the initial conditions at $\tau = 0$, $x = r \quad y = s \quad t = 0 \quad u = f(r, s)$

These ODEs are easily solved to give $x = r + \tau f + \frac{1}{2}s\tau^2 + \frac{1}{6}\tau^3 \quad y = s + \tau \quad t = \tau \quad u = f + s\tau + \frac{1}{2}\tau^2$ Writing f in terms of u, s, and τ , then substituting into the equation for x gives an implicit solution $u(x, y, t) = f(x - ut + \frac{1}{2}yt^2 - \frac{1}{6}t^3, y - t) + yt - \frac{1}{2}t^2$

It is possible to solve this for u in some special cases, but in general we can only solve this equation numerically. However, we can learn much about the global properties of the solution from further analysis

12.3 Initial value problems

Discussion.

[author=wikibooks, file =text_files/intial_value_partial_diffeqs_with_discontin_sols]
So far, weve only considered smooth solutions of the PDE, but this is too restrictive. We may encounter initial conditions which arent smooth, e.g.

$$u_t = cu_x \quad u(x, 0) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

If we were to simply use the general solution of this equation for smooth initial conditions,

$$u(x, t) = u(x + ct, 0)$$

we would get

$$u(x, t) = \begin{cases} 1, & x + ct \geq 0 \\ 0, & x + ct < 0 \end{cases}$$

which appears to be a solution to the original equation. However, since the partial differentials are undefined on the characteristic x+ct=0, so it becomes unclear what it means to say that the equation is true at that point.

We need to investigate further, starting by considering the possible types of discontinuities.

If we look at the derivations above, we see weve never use any second or higher order derivatives so it doesnt matter if they arent continuous, the results above will still apply.

The next simplest case is when the function is continuous, but the first derivative is not, e.g $|x|$. Well initially restrict ourselves to the two-dimensional case, $u(x,t)$ for the generic equation.

$$a(x,t)u_x + b(x,t)u_t = c(u,x,t) \quad (1)$$

Typically, the discontinuity is not confined to a single point, but is shared by all points on some curve, $(x_0(s), t_0(s))$

Then we have $\begin{array}{ll} x > x_0 & \lim_{x\to x_0} = u_+ \\ x < x_0 & \lim_{x\to x_0} = u_- \end{array}$

We can then compare u and its derivatives on both sides of this curve.

It will prove useful to name the jumps across the discontinuity. We say

$[u] = u_+ - u_- \quad [u_x] = u_{x+} - u_{x-} \quad [u_t] = u_{t+} - u_{t-}$

Now, since the equation (1) is true on both sides of the discontinuity, we can see that both u_+ and u_-, being the limits of solutions, must themselves satisfy the equation. That is,

$\begin{array}{rcl} a(x,t)u_{+x} + b(x,t)u_{+t} & = & c(u_+, x, t) \\ a(x,t)u_{-x} + b(x,t)u_{-t} & = & c(u_-, x, t) \end{array}$ where $\begin{array}{c} x = x_0(s) \\ t = t_0(s) \end{array}$

Subtracting then gives us an equation for the jumps in the differentials

$a(x,t)[u_x] + b(x,t)[u_t] = 0$

We are considering the case where u itself is continuous so we know that [u]=0. Differentiating this with respect to s will give us a second equation in the differential jumps.

$\frac{dx_0}{ds}[u_x] + \frac{dx_0}{dt}[u_t] = 0$

The last two equations can only be both true if one is a multiple of the other, but multiplying s by a constant also multiplies the second equation by that same constant while leaving the curve of discontinuity unchanged, hence we can without loss of generality define s to be such that

$\frac{dx_0}{ds} = a \quad \frac{dt_0}{ds} = b$

But these are the equations for a characteristic, i.e discontinuities propagate along characteristics. We could use this property as an alternative definition of characteristics.

We can deal similarly with discontinuous functions by first writing the equation in conservation form, so called because conservation laws can always be written this way.

$(au)_x + (bu)_t = a_x u + b_t u + c \quad (1)$

Notice that the left hand side can be regarded as the divergence of (au, bu). Writing the equation this way allows us to use the theorems of vector calculus.

Consider a narrow strip with sides parallel to the discontinuity and width h [[imagePDEsJumpsSetup.PNG]] could be improved

We can integrate both sides of (1) over R, giving

$\int_R (au)_x + (bu)_t \, dxdt = \int_R (a_x + b_t)u + c \, dxdt$

Next we use Greens theorem to convert the left hand side into a line integral.

$\oint_{\partial R} audt - budx = \int_R (a_x + b_t)u + c \, dxdt$

Now we let the width of the strip fall to zero. The right hand side also tends to zero but the left hand side reduces to the difference between two integrals along the part of the boundary of R parallel to the curve.

$\int au_+ dt - bu_+ dx - \int au_- dt - bu_- dx = 0$

The integrals along the opposite sides of R have different signs because they are in opposite directions.

For the last equation to always be true, the integrand must always be zero, i.e

$\left(a\frac{dt_0}{ds} - b\frac{dx_0}{ds}\right)[u] = 0$

Since, by assumption [u] isnt zero, the other factor must be, which immediately implies the curve of discontinuity is a characteristic.

Once again, discontinuities propagate along characteristics.

Discussion.

[**author=wikibooks**, file =**text_files/intial_value_partial_diffeqs_with_discontin_sols**]
Above, we only considered functions of two variables, but it is straightforward to extend this to functions of n variables.

The initial condition is given on an n-1 dimensional surface, which evolves along the characteristics. Typical discontinuities in the initial condition will lie on a n-2 dimensional surface embedded within the initial surface. This surface of discontinuity will propagate along the characteristics that pass through the initial discontinuity. diagram needed

The jumps themselves obey ordinary differential equations, much as u itself does on a characteristic. In the two dimensional case, for u continuous but not smooth, a little algebra shows that

$\frac{d[u_x]}{ds} = [u_x]\left(\frac{\partial c}{\partial u} + a\frac{b_x}{b} - a_x\right)$

while u obeys the same equation as before,

$\frac{du}{ds} = c$

We can integrate these equations to see how the discontinuity evolves as we move along the characteristic.

We may find that, for some future s, $[u_t, x_t]$ passes through zero. At such points, the discontinuity has vanished, and we can treat the function as smooth at that characteristic from then on.

Conversely, we can expect that smooth functions may, under the righr circumstances, become discontinuous.

To see how all this works in practice well consider the solutions of the equation

$u_t + uu_x = 0 \quad u(x, 0) = f(x)$

for three different initial conditions.

The general solution, using the techniques outlined earlier, is

$u = f(x - tu)$

u is constant on the characteristics, which are straight lines with slope dependent on u.

First consider f such that

$$f(x) = \begin{cases} 1 & x > a \\ \frac{x}{a} & a \geq x > 0 \quad a > 0 \\ 0 & x \leq 0 \end{cases}$$

While u is continuous its derivative is discontinuous at x=0, where u=0, and at x=a, where u=1. The characteristics through these points divide the solution into three regions.

[[ImagePDEexcont1.png]]

All the characteristics to the right of the characteristic through x=a, t=0 intersect the x-axis to the right of x=1, where u=1 so u is 1 on all those characteristics, i.e whenever x-t¿a.

Similarly the characteristic through the origin is the line x=0, to the left of which u remains zero.

We could find the value of u at a point in between those two characteristics either by finding which intermediate characteristic it lies on and tracing it back to the initial line, or via the general solution.

Either way, we get

$$f(x) = \begin{cases} 1 & x - t > a \\ \frac{x}{a+t} & a + t \geq x > 0 \\ 0 & x \leq 0 \end{cases}$$

At larger t the solution u is more spread out than at t=0 but still the same shape.

We can also consider what happens when a tends to 0, so that u itself is discontinuous at x=0.

If we write the PDE in conservation form then use Greens theorem, as we did above for the linear case, we get

$[u]\frac{dx_0}{ds} = \frac{1}{2}[u^2]\frac{dt_0}{ds}$

$[u^2]$ is the difference of two squares, so if we take s=t we get

$\frac{dx_0}{dt} = \frac{1}{2}(u_- + u_+)$

In this case the discontinuity behaves as if the value of u on it were the average of the limiting values on either side.

However, there is a caveat.

Since the limiting value to the left is u_- the discontinuity must lie on that characteristic, and similarly for u_+ i.e the jump discontinuity must be on an intersection of characteristics, at a point where u would otherwise be multivalued.

For this PDE the characteristic can only intersect on the discontinuity if

$u_- > u_+$

If this is not true the discontinuity can not propogate. Something else must happen.

The limit a=0 is an example of a jump discontinuity for which this condition is false, so we can see what happens in such cases by studying it.

Taking the limit of the solution derived above gives

$$f(x) = \begin{cases} 1 & x > t \\ \frac{x}{t} & t \geq x > 0 \\ 0 & x \leq 0 \end{cases}$$

If we had taken the limit of any other sequence of initials conditions tending to the same limit we would have obtained a trivially equivalent result.

Looking at the characteristics of this solution, we see that at the jump discontinuity characteristics on which u takes every value betweeen 0 and 1 all intersect.

At later times, there are two slope discontinuities, at x=0 and x=t, but no jump discontinuity.

This behaviour is typical in such cases. The jump discontinuity becomes a pair of slope discontinuities between which the solution takes all appropriate values.

Example 12.3.1.

[`author=wikibooks`, file =`text_files/intial_value_partial_diffeqs_with_discontin_sols`]
Now, lets consider the same equation with the initial condition

$$f(x) = \begin{cases} 1 & x \leq 0 \\ 1 - \frac{x}{a} & a \geq x > 0 \qquad a > 0 \\ 0 & x > a \end{cases}$$

This has slope discontinuities at x=0 and x=a, dividing the solution into three regions.

The boundaries between these regions are given by the characteristics through these initial points, namely the two lines

$$x = t \quad x = a$$

These characteristics intersect at t=a, so the nature of the solution must change then.

In between these two discontinuities, the characteristic through x=b at t=0 is clearly

$$x = \left(1 - \frac{b}{a}\right) t + b \quad 0 \leq b \leq a$$

All these characteristics intersect at the same point, (x,t)=(a,a).

We can use these characteristics, or the general solution, to write u for t¡a

$$u(x,t) = \begin{cases} 1 & x \leq t \\ \frac{a-x}{a-t} & a \geq x > t \qquad a > t \geq 0 \\ 0 & x > a \end{cases}$$

As t tends to a, this becomes a step function. Since u is greater to the left than the right of the discontinuity, it meeets the condition for propogation deduced above, so for t¿a u is a step function moving at the average speed of the two sides.

$$u(x,t) = \begin{cases} 1 & x \leq \frac{a+t}{2} \\ 0 & x > \frac{a+t}{2} \end{cases} \quad t \geq a \geq 0$$

This is the reverse of what we saw for the initial condition previously considered, two slope discontinuities merging into a step discontinuity rather than vice versa. Which actually happens depends entirely on the initial conditions. Indeed, examples could be given for which both processes happen.

In the two examples above, we started with a discontinuity and investigated how it evolved. It is also possible for solutions which are initially smooth to become discontinuous.

For example, we saw earlier for this particular PDE that the solution with the initial condition $u = x^2$ breaks down when 2xt+1=0. At these points the solution becomes discontinuous.

Typically, discontinuities in the solution of any partial differential equation, not merely ones of first order, arise when solutions break down in this way and progogate similarly, merging and splitting in the same fashion.

12.4 Non linear PDE's

Discussion.

[author=wikibooks, file =text_files/nonlinear_partial_diffeqs]
It is possible to extend the approach of the previous sections to reduce any equation of the form $F(x_1, x_2, \dots, x_n, u, u_{x_1}, u_{x_2}, \dots, u_{x_n}) = 0$ to a set of ODEs, for any function, F.

We will not prove this here, but the corresponding ODEs are $\frac{dx_i}{d\tau} = \frac{\partial F}{\partial u_i} \quad \frac{du_i}{d\tau} = -\left(\frac{\partial F}{\partial x_i} + u_i \frac{\partial F}{\partial u}\right) \quad \frac{du}{d\tau} = \sum_{i=1}^{n} u_i \frac{\partial F}{\partial u_i}$

If u is given on a surface parameterised by $r_1 \dots r_n$ then we have, as before, n initial conditions on the n, x_i, $\tau = 0 \quad x_i = f_i(r_1, r_2, \dots, r_{n-1})$ given by the parameterisation and one initial condition on u itself, $\tau = 0 \quad u = f(r_1, r_2, \dots, r_{n-1})$ but, because we have an extra n ODEs for the u_i's, we need an extra n initial conditions.

These are, n-1 consistency conditions, $\tau = 0 \quad \frac{\partial f}{\partial r_i} = \sum_{j=1}^{n-1} u_i \frac{\partial f_i}{\partial r_j}$ which state that the u_i's are the partial derivatives of u on the initial surface, and one initial condition $\tau = 0 \quad F(x_1, x_2, \dots, x_n, u, u_1, u_2, \dots, u_n) = 0$ stating that the PDE itself holds on the initial surface.

These n initial conditions for the u_i will be a set of algebraic equations, which may have multiple solutions. Each solution will give a different solution of the PDE.

Example 12.4.1.

[author=wikibooks, file =text_files/nonlinear_partial_diffeqs]
Consider $u_t = u_x^2 + u_y^2, \quad u(x,y,0) = x^2 + y^2$ The initial conditions at $\tau = 0$ are $\begin{matrix} x = r & y = s & t = 0 & u = r^2 + s^2 \\ u_x = 2r & u_y = 2s & u_t = 4(r^2+s^2) \end{matrix}$ and the ODEs are $\begin{matrix} \frac{dx}{d\tau} = -2u_x & \frac{dy}{d\tau} = -2u_y \\ \frac{du_x}{d\tau} = 0 & \frac{du_y}{d\tau} = 0 \end{matrix}$

Note that the partial derivatives are constant on the characteristics. This always happen when the PDE contains only partial derivatives, simplifying the procedure.

These equations are readily solved to give $x = r(1-4\tau) \quad y = s(1-4\tau) \quad t = \tau \quad u = (r^2+s^2)(1-4\tau)$

On eliminating the parameters we get the solution, $u = \frac{x^2+y^2}{1-4t}$ which can easily be checked.

12.5 Higher order PDE's

Derivation.

[author=wikibooks, file =text_files/second_order_partial_diffeqs]
Suppose we are given a second order linear PDE to solve

$$a(x,y)u_{xx}+b(x,y)u_{xy}+c(x,y)u_{yy}=d(x,y)u_x+e(x,y)u_y+p(x,y)u+q(x,y) \quad (1)$$

The natural approach, after our experience with ordinary differential equations and with simple algebraic equations, is attempt a factorisation. Lets see how far this takes us.

We would expect factoring the left hand of (1) to give us an equivalent equation of the form

$$a(x,y)(D_x+\alpha_+(x,y)D_y)(D_x+\alpha_-(x,y)D_y)u$$

and we can immediately divide through by a. This suggests that those particular combinations of first order derivatives will play a special role.

Now, when studying first order PDEs we saw that such combinations were equivalent to the derivatives along characteristic curves. Effectively, we changed to a coordinate system defined by the characteristic curve and the initial curve.

Here, we have two combinations of first order derivatives each of which may define a different characteristic curve. If so, the two sets of characteristics will define a natural coordinate system for the problem, much as in the first order case.

In the new coordinates we will have

$$D_x+\alpha_+(x,y)D_y=D_r \quad D_x+\alpha_-(x,y)D_y=D_s$$

with each of the factors having become a differentiation along its respective characteristic curve, and the left hand side will become simply u_r's giving us an equation of the form

$$u_{rs}=A(r,s)u_r+B(r,s)u_s+C(r,s)u+D(r,s).$$

If A, B, and C all happen to be zero, the solution is obvious. If not, we can hope that the simpler form of the left hand side will enable us to make progress.

However, before we can do all this, we must see if (1) can actually be factorised.

Multiplying out the factors gives

$$u_{xx}+\frac{b(x,y)}{a(x,y)}u_{xy}+\frac{c(x,y)}{a(x,y)}c(x,y)u_{yy}=u_{xx}+(\alpha_++\alpha_-)u_{xy}+\alpha_+\alpha_-u_{yy}$$

On comparing coefficients, and solving for the α s we see that they are the roots of

$$a(x,y)\alpha^2+b(x,y)\alpha+c(x,y)=0$$

Since we are discussing real functions, we are only interested in real roots, so the existence of the desired factorisation will depend on the discriminant of this quadratic equation.

If $b(x,y)^2 > 4a(x,y)c(x,y)$ then we have two factors, and can follow the procedure outlined above. Equations like this are called hyperbolic

If $b(x,y)^2 = 4a(x,y)c(x,y)$ then we have only factor, giving us a single characteristic curve. It will be natural to use distance along these curves as one coordinate, but the second must be determined by other considerations. The same line of argument as before shows that use the characteristic curve this way gives a second order term of the form $u_{r,r}$, where weve only taken the second derivative with respect to one of the two coordinates. Equations like this are called parabolic

If $b(x,y)^2 < 4a(x,y)c(x,y)$ then we have no real factors. In this case the best we can do is reduce the second order terms to the simplest possible form satisfying this inequality, i.e $u_{r,r} + u_{s,s}$ It can be shown that this reduction is always possible. Equations like this are called elliptic

It can be shown that, just as for first order PDEs, discontinuities propagate along characteristics. Since elliptic equations have no real characteristics, this implies that any discontinuities they may have will be restricted to isolated points i.e, that the solution is almost everywhere smooth.

This is not true for hyperbolic equations. Their behaviour is largely controlled by the shape of their characteristic curves.

These differences mean different methods are required to study the three types of second equation. Fortunately, changing variables as indicated by the factorisation above lets us reduce any second order PDE to one in which the coefficients of the second order terms are constant, which means it is sufficient to consider only three standard equations.

$$u_{xx} + u_{yy} = 0 \quad u_{xx} - u_{yy} = 0 \quad u_{xx} - u_y = 0$$

We could also consider the cases where the right hand side of these equations is a given function, or proportional to u or to one of its first order derivatives, but all the essential properties of hyperbolic, parabolic, and elliptic equations are demonstrated by these three standard forms.

Derivation.

[author=wikibooks, file =text_files/second_order_partial_diffeqs]
While weve only demonstrated the reduction in two dimensions, a similar reduction applies in higher dimensions, leading to a similar classification. We get, as the reduced form of the second order terms,

$$a_1 \frac{\partial^2 u}{\partial x_1^2} + a_2 \frac{\partial^2 u}{\partial x_2^2} + \cdots + a_n \frac{\partial^2 u}{\partial x_n^2}$$

where each of the a_i's is equal to either 0, +1, or -1.

If all the a_i's have the same sign the equation is elliptic

If any of the a_i's are zero the equation is parabolic

If exactly one of the a_i's has the opposite sign to the rest the equation is hyperbolic

In 2 or 3 dimensions these are the only possibilities, but in 4 or more dimensions there is a fourth possibility at least two of the a_i's are positive, and at least two of the a_i's are negative.

Such equations are called ultrahyperbolic. They are less commonly encountered than the other three types, so will not be studied here.

When the coefficients are not constant, an equation can be hyperbolic in some regions of the xy plane, and elliptic in others. If so, different methods must be used for the solutions in the two regions.

Derivation.

[author=wikibooks, file =text_files/second_order_partial_diffeqs]
The canonical parabolic equation is the diffusion equation

$$\nabla^2 h = h_t$$

Here, we will consider some simple solutions of the one-dimensional case.

The properties of this equation are in many respects intermediate between those of hyperbolic and elliptic equation.

As with hyperbolic equations but not elliptic, the solution is well behaved if the value is given on the initial surface t=0.

However, the characteristic surfaces of this equation are the surfaces of constant t, thus there is no way for discontinuities to propagate to positive t.

Therefore, as with elliptic equations but not hyberbolic, the solutions are typically smooth, even when the initial conditions arent.

Furthermore, at a local maximum of h, its Laplacian is negative, so h is decreasing with t, while at local minima, where the Laplacian will be positive, h will increase with t. Thus, initial variations in h will be smoothed out as t increases.

In one dimension, we can learn more by integrating both sides,

$$\begin{array}{rcl} \int_{-a}^{b} h_t dt & = & \int_{-a}^{b} h_{xx} dx \\ \frac{d}{dt}\int_{-a}^{b} h\, dt & = & [h_x]_{-a}^{b} \end{array}$$

Provided that h_x tends to zero for large x, we can take the limit as a and b tend to infinity, deducing

$$\frac{d}{dt}\int_{-\infty}^{\infty} h\, dt$$

so the integral of h over all space is constant.

This means this PDE can be thought of as describing some conserved quantity, initially concentrated but spreading out, or diffusing, over time.

This last result can be extended to two or more dimensions, using the theorems of vector calculus.

We can also differentiate any solution with respect to any coordinate to obtain another solution. E.g if h is a solution then

$$\nabla^2 h_x = \partial_x \nabla^2 h = \partial_x \partial_t h = \partial_t h_x$$

so h_x also satisfies the diffusion equation.

Derivation.

[author=wikibooks, file =text_files/second_order_partial_diffeqs]
Looking at this equation, we might notice that if we make the change of variables

$$r = \alpha x \quad \tau = \alpha^2$$

then the equation retains the same form. This suggests that the combination of variables x^2/t, which is unaffected by this variable change, may be significant.

We therefore assume this equation to have a solution of the special form

$$h(x,t) = f(\eta) \text{ where } \eta = \frac{x}{t^{1/2}}$$

then

$$h_x = \eta_x f_\eta = t^{-1/2} f_\eta \quad h_t = \eta_t f_\eta = -\frac{\eta}{2t} f_\eta$$

and substituting into the diffusion equation eventually gives

$$f_{\eta\eta} + \frac{\eta}{2} f_\eta = 0$$

which is an ordinary differential equation.

Integrating once gives

$$f_\eta = Ae^{-\frac{\eta^2}{4}}$$

Reverting to h, we find

$$\begin{array}{rcl} h_x & = & \frac{A}{\sqrt{t}} e^{-\frac{\eta^2}{4}} \\ h & = & \frac{A}{\sqrt{t}} \int_{-\infty}^{x} e^{-s^2/4t} ds + B \\ & = & A \int_{-\infty}^{x/2\sqrt{t}} e^{-z^2} dz + B \end{array}$$

This last integral can not be written in terms of elementary functions, but its values are well known.

In particular the limiting values of h at infinity are

$$h(-\infty, t) = B \quad h(\infty, t) = B + A\sqrt{\pi},$$

taking the limit as t tends to zero gives

$$h = \begin{cases} B & x < 0 \\ B + A\sqrt{\pi} & x > 0 \end{cases}$$

and the entire solution looks like We see that the initial discontinuity is immediately smoothed out. The solution at later times retains the same shape, but is more stretched out.

The derivative of this solution with respect to x

$$h_x = \frac{A}{\sqrt{t}} e^{-x^2/4t}$$

is itself a solution, with h spreading out from its initial peak, and plays a significant role in the further analysis of this equation.

The same similiarity method can also be applied to some non-linear equations.

Derivation.

[author=wikibooks, file =text_files/second_order_partial_diffeqs]
We can also obtain some solutions of this equation by separating variables.

$$h(x,t) = X(x)T(t) \Rightarrow X''T = X\dot{T}$$

giving us the two ordinary differential equations

$$\frac{d^2X}{dx^2} + k^2X = 0 \quad \frac{dT}{dt} = -kT$$

and solutions of the general form

$$h(x,t) = Ae^{-kt}\sin(kx + \alpha).$$

12.6 Systems of partial differential equations

Discussion.

[author=wikibooks, file =text_files/systems_of_partial_diffeqs]
We have already examined cases where we have a single differential equation and found several methods to aid us in finding solutions to these equations. But what happens if we have two or more differential equations, that depend on each other? For example, consider the case where $D_tx(t) = 3y(t)^2 + x(t)t$ and $D_ty(t) = x(t) + y(t)$ Such a set of differential equations are said to be coupled. Systems of ordinary differential equations such as these are what we will look into in this section.

First order systems A general system of differential equations can be written in the form $D_t\mathbf{x} = \mathbf{F}(\mathbf{x}, t)$

Instead of writing the set of equations in a vector, we can write out each equation explicitly, in the form $D_tx_1 = F_1(x_1, \ldots, x_n, t) \vdots D_tx_i = F_i(x_1, \ldots, x_n, t)$

If we have the system at the very beginning, we can write it as $D_t\mathbf{x} = \mathbf{G}(\mathbf{x}, t)$ where $\mathbf{x} = (x(t), y(t)) = (x, y)$ and $\mathbf{G}(\mathbf{x}, t) = (3y^2 + xt, x + y)$ or, writing each equation out as shown above.

Why are these forms important? Often, this arises as a single, higher order differential equation that is changed into a simpler form in a system. For example, with the same example, $D_tx(t) = 3y(t)^2 + x(t)t \; D_ty(t) = x(t) + y(t)$

we can write this as a higher order differential equation by simple substitution. $D_ty(t) - y(t) = x(t)$ then $D_tx(t) = 3y(t)^2 + (D_ty(t) - y(t))t \; D_tx(t) = 3y(t)^2 + tD_ty(t) - ty(t)$

Notice now that the vector form of the system is dependent on t since $\mathbf{G}(\mathbf{x}, t) = (3y^2 + xt, x + y)$ the first component is dependent on t. However, if instead we had $\mathbf{H}(\mathbf{x}) = (3y^2 + x, x + y)$ notice the vector field is no longer dependent on t. We call such systems autonomous. They appear in the form $D_t\mathbf{x} = \mathbf{H}(\mathbf{x})$ We can convert between an autonomous system and a non-autonomous one by simply making a substitution that involves t, such as y=(x, t), to get a system $D_t\mathbf{y} = (\mathbf{F}(\mathbf{y}), 1) = (\mathbf{F}(\mathbf{x}, t), 1)$

In vector form, we may be able to seperate f in a linear fashion to get something that looks like $\mathbf{F}(\mathbf{x}, t) = A(t)\mathbf{x} + \mathbf{b}(t)$ where A(t) is a matrix and b is a vector. The matrix could contain functions or constants, clearly, depending on whether the matrix depends on t or not.

Appendix: the Gnu Free Documentation License

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The **"Document"**, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as **"you"**. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A **"Modified Version"** of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A **"Secondary Section"** is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The **"Invariant Sections"** are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The **"Cover Texts"** are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A **"Transparent"** copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called **"Opaque"**.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The **"Title Page"** means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section **"Entitled XYZ"** means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as **"Acknowledgements"**, **"Dedications"**, **"Endorsements"**, or **"History"**.) To **"Preserve the Title"** of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright ©YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.