

MATH 1431 (CALCULUS I) LECTURE NOTES INVITATION-ONLY SECTION

VAUGHN CLIMENHAGA

A WORD OF CAUTION. These notes are a work in progress, and will continue to be updated the next time I teach this course. This version is from Fall 2019.

CONTENTS

I	Functions, limits, and continuity	3
1	Sets of numbers	3
2	Functions: Review of basic concepts	6
3	Examples of functions	9
4	Limits, intuitively	13
5	Computing limits	14
6	Limits, rigorously	17
7	Proving the limit laws	21
8	Theorems about limits	24
9	Continuity	26
10	Intermediate Value Theorem: Preparation	32
11	IVT: Proof and consequences	35
II	Derivatives	41
12	Derivatives	41
13	Derivative as a function	44
14	Derivatives of polynomials and exponentials	48
15	Product and quotient rules	51
16	Trigonometric functions	54
17	Differentiating exponentials	58
18	Chain rule	66
19	Implicit differentiation	70
20	Inverse functions	73
21	Rates of change in sciences	77
22	Exponential growth and decay	79
23	Related rates; linear approximation	83
24	Hyperbolic functions	87
25	The Extreme Value Theorem	90
26	Local extrema; Mean Value Theorem	92
27	Shapes of graphs	96

Date: July 15, 2020.

28	l'Hospital's rule	100
29	More on l'Hospital's rule, and basics of curve sketching	105
30	Curves, optimization, and Newton's method	108
III	Integrals	112
31	Antiderivatives	112
32	Approximating areas by sums	114
33	Lower sums, upper sums, and integrals	116
34	The Fundamental Theorem of Calculus	121
35	More about integration	124
36	Substitution rule	132
37	Finding areas between curves	135
38	Volumes	139

Part I. Functions, limits, and continuity

Lecture 1

Sets of numbers

DATE: MONDAY, AUGUST 19

The material in this lecture corresponds to Chapters 1 and 2 of Spivak's book. For full details of the construction of the real numbers – which we do not give in this course! – see Chapters 28–30 of Spivak.

We start with an informal description of five important sets of numbers.

- $\mathbb{N} = \{1, 2, 3, \dots\}$ is the set of *natural numbers*. Some people include 0 in \mathbb{N} as well.
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ is the set of *integers*. The notation stands for the German word *Zahlen* (numbers).
- $\mathbb{Q} = \{p/q : p \in \mathbb{Z}, q \in \mathbb{N}\}$ is the set of *rational numbers*. The notation stands for “quotient”. Recall that the “set builder” notation used here means “the set of all possible quotients p/q formed by choosing $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ ”.¹
- \mathbb{R} is the set of *real numbers*.
- $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$ is the set of *complex numbers*. Here i is the *imaginary* square root of -1 .

These sets are nested:

$$(1.1) \quad \mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R} \subsetneq \mathbb{C}.$$

Observe that we didn't say anything about what the set of real numbers actually is. You may have heard \mathbb{R} described as the set of all points on the number line, or perhaps even more vaguely as “all the rationals together with all the irrationals like $\sqrt{2}$ and π ”. For the moment, let's start by observing that every real number x has a decimal representation

$$(1.2) \quad x = [x] + 0.a_1a_2a_3a_4 \dots,$$

where $[x]$ means “the greatest integer that is less than or equal to x ”, and each a_i is a digit from the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Since we have rules for adding, subtracting, multiplying, and dividing decimal representations, this seems to give us the tools that we need to work with real numbers. However, there are a few ambiguities and questions that we would do well to keep in mind.

- The rules for arithmetic operations with decimal representations are defined for *finite* decimal representations, not infinite ones. For example, in the algorithm for addition, we start at the right-most decimal place and then work to the left. But if the decimal representation is infinite, there is no right-most decimal place. So what do we do?

¹The notation “ $p \in \mathbb{Z}$ ” means “ p is an element of the set of integers”, or more concisely, “ p is in the set of integers”, or even more concisely, “ p is an integer”.

- A single real number may have more than one decimal representation: for example, $1 = 0.99999\dots$. So it is not quite accurate to say that “ \mathbb{R} is the set of all decimal representations”.
- Why is \mathbb{Q} not enough? That is, is $\mathbb{R} \setminus \mathbb{Q} = \{x \in \mathbb{R} : x \notin \mathbb{Q}\}$ nonempty? We talk about “the square root of 2” (for example) as the unique real number $x > 0$ such that $x^2 = 2$, but how do we know that such a real number exists? And if we know that it exists, how do we know that it is not actually a rational number?

For the first issue raised above, it is instructive to consider how we might compute a decimal expansion for $x + y$ in the following cases.

- (1) $x = 0.123123123\dots$ (which we abbreviate as $x = 0.\overline{123}$) and $y = 0.\overline{456}$.
- (2) $x = 0.\overline{456}$ and $y = 0.\overline{789}$.
- (3) $x = 0.\overline{142857}$ and $y = 0.\overline{3}$.
- (4) $x = 0.95955955595559\dots$ and $y = 0.\overline{4}$.

In the first example, there is no issue; we simply add separately in each decimal place, there are no carries, and we obtain

$$\begin{array}{r} 0.123123123\dots \\ +0.456456456\dots \\ \hline = 0.579579579\dots \end{array}$$

In the second example, life gets a little more complicated, since the sum in every decimal place is ≥ 10 , so every decimal place has a carry:

$$\begin{array}{r} 1\ 111111111 \\ 0.456456456\dots \\ +0.789789789\dots \\ \hline = 1.246246246\dots \end{array}$$

In the third example, we have carries in some places but not in others:

$$\begin{array}{r} \overset{1}{1} \overset{1}{1} \overset{1}{1} \overset{1}{1} \\ 0.142857142857\dots \\ +0.333333333333\dots \\ \hline = 0.476190476190\dots \end{array}$$

Our implicit strategy here has been to start at the left and work to the right – which is fine because the decimal expansion *has* a leftmost position – and at each step to check whether or not there is a carry by looking at the next digit and determining whether or not the sum will be ≥ 10 . So far this has worked just fine, because we can check this by looking only one digit ahead. But the fourth example requires a little more care: when we add $9 + 4$ we clearly get a carry in the previous position, but what happens when we add $5 + 4$? In this case we need to look one digit further and see if *that* digit sum is ≥ 10 , in which case we would end up with $5 + 4 + 1 = 10$. For the values of x and y that are given, we actually end up with a carry in every position – but we sometimes need to look very far ahead to confirm this!

Another strategy that works for the first three examples is to observe that x and y have repeating decimal expansions and thus actually represent rational numbers. In the third example above, we have $x = \frac{1}{7}$ and $y = \frac{1}{3}$, so $x + y = \frac{3}{21} + \frac{7}{21} = \frac{10}{21}$, and with some more work you can see that this agrees with the decimal answer we got. This strategy,

though, does not work for the fourth example, because in that instance x cannot be represented as a fraction.

One final strategy, which works equally well for all choices of x and y , is the following: for each $n \in \mathbb{N}$, truncate x and y to the first n digits of their decimal expansions, which can be easily added by starting at the right and working left. As n gets larger and larger, so that we write down more and more digits of x and y , this should give us a better and better approximation to the true value of $x + y$. Here is what the first few steps of this procedure look like for the fourth example above:

$$\begin{array}{r} \overset{1}{0.9} \\ +0.4 \\ \hline = 1.3 \end{array} \quad \begin{array}{r} \overset{1}{0.95} \\ +0.44 \\ \hline = 1.39 \end{array} \quad \begin{array}{r} \overset{1}{0.959} \\ +0.444 \\ \hline = 1.403 \end{array} \quad \begin{array}{r} \overset{1}{0.9595} \\ +0.4444 \\ \hline = 1.4039 \end{array} \quad \begin{array}{r} \overset{1}{0.95955} \\ +0.44444 \\ \hline = 1.40399 \end{array} \quad \begin{array}{r} \overset{1}{0.959559} \\ +0.444444 \\ \hline = 1.404003 \end{array}$$

We see that as we take more and more digits in our computation, the sum is getting closer and closer to $1.4040040004\dots$, which is the value of $x + y$.

We can take away a few lessons from this example.

- (1) There may be multiple approaches to solving a particular kind of problem.
- (2) Some of those approaches may work in more cases than others.
- (3) When we work with real numbers, the idea of *approximation* – by rational numbers or otherwise – is very powerful. It may be the case that it is not clear how to do something for a particular real number x , but that we do know how to do it for some values of y that approximate x , and that we can use this to get closer and closer to the actual answer for x itself.

This language of “approximate” and “closer and closer” will be made more precise soon, when we discuss *limits*.

The remainder of this section was skipped in the classroom lecture.

For the second issue raised above (multiple decimal representations), we point out that in fact this also arises in our definition of \mathbb{Q} , since we may have $p/q = a/b$ for some $p, a \in \mathbb{Z}$ and $q, b \in \mathbb{N}$ even if $p \neq a$ and $q \neq b$. In that case we can guarantee a unique representation by requiring that p, q have no common factors except 1 (they are *relatively prime*); we can do a similar thing for real numbers by requiring that our decimal representations do not terminate in an infinite sequence of 9s.

For the third issue above, we will defer a discussion of why $x^2 = 2$ has a solution in \mathbb{R} , and present the proof that it does *not* have a solution in \mathbb{Q} .

Theorem 1.1. *There is no rational number x such that $x^2 = 2$.*

Proof. We use the technique of *proof by contradiction*: that is, we start by assuming that there *is* a rational number x with $x^2 = 2$, then show that this would imply a statement that we know to be false. This means that the original assumption must be false.

To this end, suppose that $p, q \in \mathbb{N}$ are relatively prime and that $(p/q)^2 = 2$. Then $p^2 = 2q^2$. Now we need the following fact, whose proof we leave to the reader.

Exercise 1.2. An integer $n \in \mathbb{Z}$ is even if and only if its square n^2 is even.

Since $p^2 = 2q^2$ is even, Exercise 1.2 implies that p is even as well; thus $p = 2a$ for some $a \in \mathbb{N}$, and we get

$$2q^2 = p^2 = (2a)^2 = 4a^2 \quad \Rightarrow \quad q^2 = 2a^2.$$

Thus q^2 is even, and again Exercise 1.2 implies that q is even. But then p, q are both even, contradicting our assumption that they are relatively prime. This contradiction shows that no such p, q exist, which completes the proof of the theorem. \square

The result of Theorem 1.1 is usually stated as “The square root of 2 is irrational”. In order to state it this way, though, we need to show that there *is* a square root of 2 in the real numbers, which we defer until later.

Lecture 2 Functions: Review of basic concepts

DATE: WEDNESDAY, AUGUST 21

This lecture corresponds to §1.1 of Stewart’s book and Chapter 3 of Spivak.

2.1. Functions and their graphs

A *function* f from a set X to a set Y is a rule that assigns to each element $x \in X$ an element $f(x) \in Y$. The set X is the *domain* of f , and the set Y is the *codomain* or *target space*. The *range* of f is

$$f(X) := \{f(x) : x \in X\} \subset Y,$$

where we use “ $A := B$ ” to mean that A is defined as equal to B . Note that we use the notation $A \subset B$ to mean “ A is a subset of B ”, without requiring that $A \neq B$; some authors use $A \subseteq B$ instead. If we need to specify that A is a *proper* subset of B , we will write $A \subsetneq B$, as in (1.1).

There are several ways to define a function.

- We can use a formula, such as $f(x) = x^2$. In this case the domain is usually understood to be the set of all real numbers for which the formula also gives a real number, although sometimes a smaller domain is explicitly specified.
- We can list every element of X and then say which element of Y it is mapped to by f . In this case the domain is explicitly specified.
- We can give a verbal description or an algorithm that defines the function. For example, consider the function $f: \mathbb{N} \rightarrow \mathbb{N}$ defined by taking $f(n)$ to be the n th prime number when the primes are listed in increasing order.
- We can define a function *piecewise* by partitioning the domain X into subsets and using one of the above methods to define f on each of these. For example, the “Collatz function” $f: \mathbb{N} \rightarrow \mathbb{N}$ is defined by

$$f(n) = \begin{cases} 3n + 1 & \text{if } n \text{ is odd,} \\ n/2 & \text{if } n \text{ is even.} \end{cases}$$

We are used to drawing the graph of a function from the real line to itself as a curve in the plane \mathbb{R}^2 , which consists of all the points (x, y) for which $y = f(x)$. More generally, given two sets X and Y we can form the *direct product*

$$(2.1) \quad X \times Y := \{(x, y) : x \in X, y \in Y\},$$

and then the graph of a function $f: X \rightarrow Y$ is defined to be

$$(2.2) \quad \text{graph}(f) := \{(x, f(x)) : x \in X\} \subset X \times Y.$$

Remark 2.1. The difference between the *ordered pair* (x, y) in (2.1) and the set $\{x, y\}$ is two-fold: firstly, the elements of the ordered pair are allowed to be the same (if X, Y have any elements in common), and secondly, in the ordered pair we keep track of the order in which the elements appear, so that (x, y) is distinct from (y, x) if $x \neq y$, whereas $\{x, y\} = \{y, x\}$. Notice that the same notation (x, y) can refer both to an ordered pair and to the open interval with endpoints x and y (when x, y are real numbers), so one must always be alert to the context in which it occurs to see which is meant.

**Remark 2.2.*² Formally, since set theory is taken as the foundation of mathematics, we should define ordered pairs in terms of sets: this can be done by declaring (x, y) to be the set $\{\{x\}, \{x, y\}\}$ (recall that an element of a set could be a set in its own right); then (x, x) is represented by $\{\{x\}, \{x\}\} = \{\{x\}\}$, and we see that $(x, y) = \{\{x\}, \{x, y\}\}$ and $(y, x) = \{\{y\}, \{x, y\}\}$ are distinct from each other.³ For practical purposes we do not bother with this level of detail, however, and continue to simply work with (x, y) as we always have.

Recall the vertical line test, which says that a set $\Gamma \subset \mathbb{R}^2$ is the graph of a function (on some domain in \mathbb{R}) if and only if every vertical line intersects Γ at most once. A more precise version says that Γ is the graph of a function from X to \mathbb{R} if and only if every vertical line whose x -coordinate is in X intersects Γ exactly once. The vertical line through $(x, 0)$ is the set $\{(x, y) : y \in \mathbb{R}\}$, and so we can formulate a version of the vertical line test that works for general sets X and Y .

Vertical Line Test. *Given two sets X and Y , and a subset $\Gamma \subset X \times Y$, the following are equivalent.*

- (1) *There is a function $f: X \rightarrow Y$ such that $\Gamma = \text{graph}(f)$.*
- (2) *For all $x \in X$, we have $\#(\Gamma \cap \{(x, y) : y \in Y\}) = 1$.*
- (3) *For all $x \in X$, there exists a unique $y \in Y$ such that $(x, y) \in \Gamma$.*

Remark 2.3. If A is a set, the notation $\#A$ denotes the number of elements in A . Thus the equation in the second item above is just the statement that each “vertical line” intersects Γ in exactly one point.

The terms “for all” and “there exists” that we used above are important enough that we give them their own notation: we write \forall to mean “for all”, and \exists to mean “there exists”. Thus the third item above could be written as

$$(2.3) \quad \forall x \in X \exists \text{ a unique } y \in Y \text{ s.t. } (x, y) \in \Gamma,$$

where we also use the abbreviation “s.t.” for “such that”.

²Throughout the notes, we will use a star to mark those remarks that go somewhat beyond the scope of this course and were not mentioned in lecture.

³See https://en.wikipedia.org/wiki/Ordered_pair#Kuratowski's_definition for more.

2.2. Injectivity, surjectivity, and bijectivity

Definition 2.4. Consider a function $f: X \rightarrow Y$. Given $x \in X$, the *image* of x under f is $f(x) \in Y$. If $f(x) = y$, we say that x is a *pre-image* of y . The function f is *one-to-one* (1-1) if every $y \in Y$ has at most one pre-image; such a function is also called *injective*. In this case we can define an inverse function $f^{-1}: \text{range}(f) \rightarrow X$ by the condition that $f^{-1}(y)$ is the unique $x \in X$ such that $f(x) = y$; that is, the unique pre-image of y .

Remark 2.5. Do not confuse the *inverse* f^{-1} with the *reciprocal* $1/f$. The reason for the power-type notation in the inverse is that if $Y = X$, so that f maps the set X to itself, then for each $n \in \mathbb{N}$ we can define

$$f^n = \overbrace{f \circ f \circ \cdots \circ f}^{n \text{ times}}$$

to be the n th iterate of f under *composition*, and these functions have the property that⁴

$$(2.4) \quad f^m \circ f^n = f^{m+n}.$$

The identity function $\text{Id}(x) = x$ has the property that $\text{Id} \circ f^n = f^n$, so it makes sense to write $f^0 = \text{Id}$. Then if (2.4) is to hold for negative integers as well, we should have

$$f^{-1} \circ f = f^{-1} \circ f^1 = f^{-1+1} = f^0 = \text{Id};$$

in other words, f^{-1} should be the inverse function for f .

The following are all equivalent to the definition of injectivity.

- (1) For every $y \in Y$, $\#\{x \in X : f(x) = y\} \leq 1$.
- (2) For every $y \in Y$, $\#(\text{graph}(f) \cap \{(x, y) : x \in X\}) \leq 1$.
- (3) If $x_1, x_2 \in X$ are such that $f(x_1) = f(x_2)$, then $x_1 = x_2$.

The first of these just restates the definition. The second is a mildly more complicated version of the first, which has a geometric interpretation called the *horizontal line test*: the set $\{(x, y) : x \in X\}$ represents the horizontal line with second coordinate y , so f is injective if every horizontal line meets its graph in at most one point. It is a (short) exercise to prove that the third condition is equivalent to the other two.

Definition 2.6. A function $f: X \rightarrow Y$ is *onto* (or *surjective*) if every $y \in Y$ has a pre-image in X ; in other words, for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$. If f is both 1-1 and onto (both injective and surjective), then it is called a *bijection*.

Using the “quantifiers”⁵ \forall and \exists , the definition of surjectivity can be rewritten as

$$(2.5) \quad \forall y \in Y \exists x \in X \text{ s.t. } f(x) = y.$$

Remark 2.7. Whenever you see the symbols \forall and \exists , it is helpful to interpret the statement containing them by thinking of a game between you and an adversary. Your goal is to verify the truth of the statement, and your adversary’s goal is to make the statement be false. Each \forall represents a turn taken by your adversary, in which they make a choice over which you have no control. Each \exists represents a turn that you take, in which you get to control the choice. The overall statement is true if you have a winning strategy:

⁴With this notation in mind, observe that $f^2(x)$, $f(x)^2$, and $f(x^2)$ all mean different things.

⁵In logic, \forall is called the *universal quantifier* and \exists is called the *existential quantifier*.

that is, no matter what your adversary does, you can make your choices in such a way that the innermost statement is true.

Consider the specific example given in (2.5); suppose that $X = Y = \mathbb{R}$ and that $f(x) = 2x + 1$. Then on the first turn of the game, your adversary picks a real number y , over which you have no control. You want to choose a number x such that $y = f(x) = 2x + 1$; some simple algebra suggests that you should pick $x = (y - 1)/2$. Indeed, if you make this pick, then $f(x) = 2(y - 1)/2 + 1 = y$, and thus no matter what number y your adversary picks (with the \forall), you can always make a choice of x (with the \exists) such that the innermost statement, “ $f(x) = y$ ”, is true. Thus the entire statement in (2.5) is true, and indeed, this function is onto.

On the other hand, if $f(x) = x^2$, then your adversary, after a moment’s thought, will realize that if they choose $y = -1$ on their turn, then no matter what choice of x you make, you will have $f(x) = x^2 \geq 0 > -1 = y$, and in particular, $f(x) \neq y$. Thus your adversary can force the innermost statement to be false, and thus the entire statement in (2.5) is false. And indeed, this function is not onto.

When we discuss limits in a few lectures, we will encounter more complicated expressions involving \forall and \exists .

Lecture 3

Examples of functions

DATE: FRIDAY, AUGUST 23

This lecture corresponds to §§1.2–1.5 in Stewart and Chapter 4 in Spivak.

3.1. Polynomials and rational functions

We will need to study functions whose domain and range lie in \mathbb{R} . One fundamental class of examples is the set of *polynomials*: a function $f: \mathbb{R} \rightarrow \mathbb{R}$ is a polynomial if there is a nonnegative integer n (called the *degree* of f and sometimes written $\deg f$) and real numbers a_0, a_1, \dots, a_n (the *coefficients* of f) such that

$$(3.1) \quad f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n.$$

This can also be written using summation notation:

$$f(x) = \sum_{k=0}^n a_k x^k.$$

Low-degree polynomials are given specific names:

A polynomial of degree is called ...
0	constant
1	linear
2	quadratic
3	cubic
4	quartic
5	quintic

If f is a polynomial, then a number $r \in \mathbb{R}$ such that $f(r) = 0$ is called a *root* of the polynomial. You can use polynomial long division to prove the following.

Exercise 3.1. Prove that $r \in \mathbb{R}$ is a root of a polynomial f if and only if there is a polynomial g such that $f(x) = (x - r)g(x)$ for all $x \in \mathbb{R}$. Show that in this case, $\deg g = (\deg f) - 1$.

Exercise 3.2. Use Exercise 3.1 to show that the number of roots of a non-constant polynomial f is at most $\deg f$.

Remark 3.3. When f is linear, it is easy to find the unique root. When f is quadratic, we have the *quadratic formula* that produces the roots (if they exist) in terms of the coefficients. There is a corresponding formula to find the roots of cubic polynomials, but it is rather more complicated. There is even a formula to solve quartic equations, but it is nothing short of horrific to write out in full (it would take several pages to do so). It turns out that there is no formula to solve quintic equations in general. By this we do not mean that “no formula is known”. Rather, we mean that it can be *proved* that no such formula exists, so that the fact that we do not know such a formula is not a reflection of our own ignorance, but rather a fundamental fact about the mathematical universe. The proof of this fact uses what is called *Galois theory*, which is well beyond the scope of this course.

Definition 3.4. If f, g are polynomials, then $r(x) := f(x)/g(x)$ is called a *rational function*. A polynomial $p(x)$ is a *factor* of a polynomial $f(x)$ if there is a polynomial $q(x)$ such that $f(x) = p(x)q(x)$; then polynomials f and g *have no common factor* if there is no non-constant polynomial p that is a factor of both f and g . If f, g are polynomials with no common factors, then the *degree* of the rational function f/g is $\max(\deg f, \deg g)$.

If the word “polynomial” is replaced with the word “integer” in the previous definition, then this turns into the description of the rational numbers. The rules for doing arithmetic with rational functions are completely analogous to those for doing arithmetic with rational numbers: for multiplication we simply multiply the numerators and denominators of the two functions, while for addition and subtraction we must first put everything over a common denominator. Thus for the rational functions $\frac{1}{x}$ and $\frac{1}{x+1}$ we have

$$(3.2) \quad \frac{1}{x} - \frac{1}{x+1} = \frac{x+1}{x(x+1)} - \frac{x}{x(x+1)} = \frac{1}{x(x+1)}.$$

3.2. Trigonometric functions

We define the sine and cosine functions as follows. Given $t \in \mathbb{R}$, let $P(t) \in \mathbb{R}^2$ be the point on the unit circle obtained by starting at the point $(1, 0)$ and moving counterclockwise until a total arc length of t has been reached. Then $\cos(t)$ is the x -coordinate of $P(t)$, and $\sin(t)$ is the y -coordinate of $P(t)$.

There are four more standard trigonometric functions, defined as:⁶

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

Because $P(t)$ lies on the unit circle $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, we have

$$\cos^2 t + \sin^2 t = 1 \text{ for all } t \in \mathbb{R}.$$

Two other fundamental trigonometric identities that we will need later on are the formulas for sine and cosine of sums of angles:

$$(3.3) \quad \sin(x + y) = \sin x \cos y + \cos x \sin y,$$

$$(3.4) \quad \cos(x + y) = \cos x \cos y - \sin x \sin y.$$

For the time being we omit the proofs of these identities, which can be given by elementary geometric arguments.

Remark 3.5. We said “sums of angles” even though no angles appeared in the discussion so far. In the case when $0 < t < \pi/2$, we can consider the triangle with vertices at $O = (0, 0)$, $P = (\cos t, \sin t)$, and $Q = (\cos t, 0)$; then $t = \angle(POQ)$ and $\cos t$, $\sin t$ give the lengths of the sides PQ and OP , respectively.

There is a weak point in these definitions, though. What exactly do we mean by “arc length along the circle”? To make this a little more concrete: given a point (x, y) in the first quadrant (so $x > 0$ and $y > 0$), how do we compute the arc length along the circle from $(0, 0)$ to (x, y) ? If we replace the arc between these two points with a straight line, then length is easy to compute; the length of the straight line between two points (x_1, y_1) and (x_2, y_2) is just the distance between those points, which is given by the Pythagorean distance formula $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. We can get a better approximation by picking a few points on the arc, connecting them with straight lines, and adding up the lengths of those line segments. Intuitively, it seems reasonable to say that the length of the arc is somehow given by this approximation procedure, provided we take enough points. But this will take some work to make precise, and until we do that work we need to regard our definitions of the trigonometric functions as provisional, since they rely on a notion of arc length that we have not really nailed down yet.

3.3. Exponential functions

Given $a > 0$, we would like to define an *exponential* function $f: \mathbb{R} \rightarrow \mathbb{R}$ by $f(x) = a^x$. But what does a^x mean? When x is a natural number it is clear enough: $a^1 = a$, $a^2 = a \cdot a$, $a^3 = a \cdot a \cdot a$, and so on: a^x means the product of a with itself x times.

What about negative values of x ? Or non-integer values? The first thing to observe is that when x, y are natural numbers, we clearly have

$$(3.5) \quad a^{x+y} = a^x a^y.$$

⁶Yes, we’re using a different variable (x) here than we did in the paragraph above (t). This should not bother you. The identity of a function is not affected by what we call the variable we are feeding into it. *That which we call a rose*, and so on and so forth.

We want to define a^x for more general values of x in such a way that (3.5) continues to hold. To this end, we first note that whatever a^0 is, it should have the property that $a^0 a^x = a^{0+x} = a^x$ for all $x \in \mathbb{N}$; this is only possible if $a^0 = 1$, so we define

$$(3.6) \quad a^0 := 1.$$

Now if x is a negative integer, then $x = -n$ for some $n \in \mathbb{N}$, and in order for (3.5) to hold with $y = n$, we must have

$$a^{-n} = \frac{a^{x+y}}{a^y} = \frac{a^{-n+n}}{a^n} = \frac{a^0}{a^n} = \frac{1}{a^n}.$$

Thus we define

$$(3.7) \quad a^{-n} := \frac{1}{a^n}$$

for every $n \in \mathbb{N}$. Now we have defined a^x for every $x \in \mathbb{Z}$. But what about non-integer values? Again, (3.5) seems to tell us what to do. For example, whatever $a^{1/2}$ is, using (3.5) with $x = y = 1/2$ tells us that

$$a = a^1 = a^{1/2+1/2} = a^{1/2} a^{1/2} = (a^{1/2})^2 \quad \Rightarrow \quad a^{1/2} = \sqrt{a}.$$

More generally, if p/q is any rational number, then iterating (3.5) q times gives

$$(a^{p/q})^q = a^{\frac{p}{q} \cdot q} = a^p,$$

and thus $a^{p/q}$ must be *defined* to be the q th root of a^p . We are nearly there – we have defined a^x whenever $x \in \mathbb{Q}$ – but two questions remain to be addressed.

- (1) Why does a^p always have a q th root when $p, q \in \mathbb{N}$, and why should it be unique?
- (2) What are we to make of a^x when x is an irrational number?

The first question will be addressed when we study the intermediate value theorem. For the second question, we start by observing that in order to describe an irrational number such as π , we can use a sequence of increasingly accurate rational approximations: $\pi \approx 3.14$, then $\pi \approx 3.14159$, then $\pi \approx 3.1415926535$, and so on. Since a^x was defined whenever $x \in \mathbb{Q}$, we know what is meant by $a^{3.14}$, $a^{3.14159}$, etc., and it would be reasonable to expect that these are “increasingly accurate approximations” to a^π . To make this precise requires the notion of *limit*, which we start discussing in the next lecture.

In the meantime, we observe that if $a > 1$, then the function $\mathbb{N} \rightarrow \mathbb{R}$ defined by $f(x) = a^x$ also has the property that it is *strictly increasing*: if $x, y \in \mathbb{N}$ satisfy $x < y$, then $f(x) = a^x < a^y = f(y)$. If $a < 1$, then it is *strictly decreasing*. (If $a = 1$, then $f(x) = 1$ for all x , and the function is not so exciting.) Just as we want to define $f: \mathbb{R} \rightarrow \mathbb{R}$ in such a way that (3.5) is preserved, we would also like to preserve this property of being strictly increasing (if $a > 1$) or decreasing (if $a < 1$). If and when we can do this, we will have a 1-1 function $f: \mathbb{R} \rightarrow \mathbb{R}$. We will eventually prove that the range of this function is $(0, \infty)$, and thus there is an inverse function $f^{-1}: (0, \infty) \rightarrow \mathbb{R}$, which is called the *logarithm with base a* , and denoted \log_a . Observe that given $x, y \in (0, \infty)$, if we write $s = \log_a x$ and $t = \log_a y$, then (3.5) gives

$$a^{s+t} = a^s a^t = xy,$$

and thus $s + t = \log_a(xy)$. In other words, \log_a satisfies the identity

$$(3.8) \quad \log_a(xy) = \log_a(x) + \log_a(y).$$

So far we have no reason to prefer one value of $a > 0$ over another. Eventually we will see that there is a *natural logarithmic base* $e \approx 2.71828\dots$, also called *Euler's constant*; however, the motivation for this number needs to wait until we discuss derivatives.

Lecture 4

Limits, intuitively

DATE: MONDAY, AUGUST 26

This lecture corresponds to §§2.1–2.2 in Stewart, and the beginning of Chapter 5 in Spivak.

We have now encountered several situations in which there is some quantity that we want to compute but cannot do so directly, while at the same time we can compute approximations to this quantity. We briefly describe these, together with some new ones.

- (1) *Arithmetic with decimal expansions.* If we want to add (or multiply) the decimal expansions for x and y , we can take the first n digits of each, add those using the normal rules for addition, and then let n get larger and larger to get a better and better approximation for $x + y$.
- (2) *Exponential functions.* If x is an irrational number and we want to make sense of the expression 2^x , we can approximate x with rational numbers p/q , for which $2^{p/q}$ is defined as the q th root of 2^p (which is itself 2 multiplied by itself p times).
- (3) *Arc length.* The number π is the circumference of a circle with diameter 1. Here “circumference” means the length traveled if we go once around the circle, and this can be approximated by drawing a regular polygon with a large number of sides. The perimeter of this polygon is something we can calculate using Pythagoras’ formula.
- (4) *Area.* The number π is also the area of a circle with radius 1. Here “area” is something we can make sense of for rectangles – where it is width times height – and so we can imagine covering the circle by a large number of small rectangles, then adding up their areas to get an approximate value for the area of the circle. As we use smaller and smaller rectangles in this procedure, our approximation should get better and better.
- (5) *Instantaneous velocity.* Suppose I throw a ball straight up into the air, and $f(t)$ represents its height at time t . Then the *average velocity* of the ball between time t and time $t + h$ is given by (total distance traveled) / (time elapsed), which is $\frac{f(t+h)-f(t)}{h}$. As h gets smaller and smaller, this gives a better and better approximation to the *instantaneous velocity* of the ball at time t .
- (6) *Tangent lines.* The graph of a function $y = f(x)$ gives a curve in \mathbb{R}^2 . The tangent line to this curve at a given point $(a, f(a))$ is the line through this point that “goes in the same direction as the curve”. But what does “direction of the curve” mean? If $(b, f(b))$ is a nearby point on the curve, then the *secant line* corresponding to a and b is the line passing through $(a, f(a))$ and $(b, f(b))$, which has slope given by $\frac{\text{rise}}{\text{run}} = \frac{f(b)-f(a)}{b-a}$. As b gets closer and closer to a , this secant

line gives a better and better approximation to the tangent line to f at $(a, f(a))$. For example, if $f(x) = x^2$ and $a = 2$, then the slope of the secant line is

$$\frac{b^2 - 4}{b - 2} = \frac{(b - 2)(b + 2)}{b - 2} = b + 2;$$

note that this makes sense as long as $b \neq 2$, but if $b = 2$ then the initial expression is no longer defined. Nevertheless we can use the final expression to deduce that as b gets closer and closer to 2, the slope of the secant line gets closer and closer to 4, so the slope of the tangent line is 4, and the equation of the tangent line is $y - 4 = 4(x - 2)$, which simplifies to $y = 4x - 4$.

If you have taken some calculus before, you may recognize the first two of these as instances of *continuity*, the next two as instances of *integrals*, and the last two as instances of *derivatives*. For now, we focus on the common thread between all of them, which is the idea of a *limit*.

Definition 4.1. Given a real-valued function f , a real number a in the domain of f , and a real number L , we say that L is the limit of f at a , and write $\lim_{x \rightarrow a} f(x) = L$, if we can make the values of $f(x)$ be arbitrarily close to L by taking x to be sufficiently close (but not equal) to a . In this case we also write “ $f(x) \rightarrow L$ as $x \rightarrow a$ ”, or sometimes $f(x) \xrightarrow{x \rightarrow a} L$.

Note that the value $f(a)$ does not affect the limit. We will make Definition 4.1 more precise in a couple lectures. For now we just mention that we will also have occasion to talk about the limit of a sequence: if x_1, x_2, x_3, \dots is a sequence of real numbers, we say that L is the limit of x_n (as $n \rightarrow \infty$), and write $\lim_{n \rightarrow \infty} x_n = L$, if we can make the values of x_n be arbitrarily close to L by taking n to be sufficiently large. The first three examples in the list above can be interpreted as the limit of a sequence (sequence of rational approximations, sequence of polygons, sequence of coverings by rectangles), while the last two can be interpreted as the limit of a function.

Lecture 5

Computing limits

DATE: WEDNESDAY, AUGUST 28

This lecture corresponds to §2.3 in Stewart, and the end of Chapter 5 in Spivak.

5.1. Algebraic tricks and numerical approximations

Sometimes we can use some algebraic simplifications to compute limits: for example, the tangent line computation in the previous lecture went as follows:

$$\lim_{x \rightarrow 2} \frac{x^2 - 4}{x - 2} = \lim_{x \rightarrow 2} \frac{(x - 2)(x + 2)}{x - 2} = \lim_{x \rightarrow 2} (x + 2),$$

and it is not much of a stretch to convince yourself that $x+2$ approaches 4 as x approaches 2. (We will give a more complete justification of this in the next few lectures.) What

about the limit of the reciprocal quantity? We can write

$$\lim_{x \rightarrow 2} \frac{x-2}{x^2-4} = \lim_{x \rightarrow 2} \frac{x-2}{(x-2)(x+2)} = \lim_{x \rightarrow 2} \frac{1}{x+2},$$

and it seems reasonable to expect that this approaches $\frac{1}{4}$, but how can we be sure of this? (Recall, after all, that in the definition of limit we never actually make the substitution $x = 2$, we merely choose values of x that are extremely close to 2.) We might try gathering some numerical evidence, choosing numbers like $x = 2.000001$ and seeing what happens. But we should be wary, because strange things can happen if we blindly rely on a calculator or computer. For example, suppose we want to compute

$$\lim_{t \rightarrow 0} f(t) \text{ for } f(t) = \frac{\sqrt{t^2+1}-1}{t^2}.$$

When $t = .01$, a numerical computation shows that

$$f(.01) = 0.499988\dots,$$

which suggests that the limit is $\frac{1}{2}$. (Though it's worth pausing for a moment and thinking about whether we should really always expect the limit to be a "nice" number.) When $t = .001$, a similar computation gives $f(.001) = 0.499999875\dots$, lending further credence to the conjecture. But if we keep going, then for some extremely small t , perhaps $.00001$, perhaps 10^{-10} (it depends on the details of which calculator or computer you use), the computer will start returning the answer 0. This is because it only stores some finite number of digits, and eventually the numerator of $f(t)$ gets stored as 0. On the other hand, we can use some algebra to observe that

$$(5.1) \quad \frac{\sqrt{t^2+1}-1}{t^2} = \frac{\sqrt{t^2+1}-1}{t^2} \frac{\sqrt{t^2+1}+1}{\sqrt{t^2+1}+1} = \frac{(t^2+1)-1}{t^2(\sqrt{t^2+1}+1)} = \frac{1}{\sqrt{t^2+1}+1},$$

and it once again seems very reasonable to expect that this approaches $\frac{1}{\sqrt{1+1}} = \frac{1}{2}$.

Remark 5.1. The algebraic trick we have used several times in this and the previous lecture is worth remembering: a^2-b^2 factors as $(a-b)(a+b)$, and an expression containing something of the form $a-b$ (or $a+b$) can sometimes be simplified by multiplying both numerator and denominator by the *conjugate* expression, so that for example we get $\sqrt{A}-\sqrt{B} = \frac{A-B}{\sqrt{A}+\sqrt{B}}$.

In each of the examples above, we reached a point where we concluded with "it seems reasonable to expect that" the limit is given by substituting the limiting value of t (or x , or whatever the independent variable is) into the expression. We will start justifying this shortly. First one more cautionary tale is in order.

Example 5.2. Define $f: (0, \infty) \rightarrow \mathbb{R}$ by $f(x) = \sin \frac{1}{x}$. Then for $x = \frac{1}{n\pi}$ we have $f(x) = \sin \frac{1}{x} = \sin n\pi = 0$, but it is *not* true that $f(x) \rightarrow 0$ as $x \rightarrow 0$. Indeed, if $x = \frac{2}{n\pi}$, then $f(x) = \sin \frac{n\pi}{2}$, which is equal to ± 1 whenever n is odd. Thus in this case the limit *does not exist*. A similar example with sequences (which is easier to state) is $x_n = (-1)^n$, so that x_n is 1 when n is even and -1 when n is odd.

5.2. The limit laws

Our primary tools for computing limits will be algebraic manipulations of the sort described above, together with the following set of *limit laws*.

Theorem 5.3 (Limit laws). *Let f, g be functions defined around a point $a \in \mathbb{R}$, and suppose that $\lim_{x \rightarrow a} f(x)$ and $\lim_{x \rightarrow a} g(x)$ exist. Let c be any real number. Then the following limits all exist and are given by the values shown.*

- (1) $\lim_{x \rightarrow a} (f(x) + g(x)) = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x)$
- (2) $\lim_{x \rightarrow a} (f(x) - g(x)) = \lim_{x \rightarrow a} f(x) - \lim_{x \rightarrow a} g(x)$
- (3) $\lim_{x \rightarrow a} (cf(x)) = c \lim_{x \rightarrow a} f(x)$
- (4) $\lim_{x \rightarrow a} (f(x)g(x)) = \left(\lim_{x \rightarrow a} f(x) \right) \left(\lim_{x \rightarrow a} g(x) \right)$
- (5) $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)}$ provided $\lim_{x \rightarrow a} g(x) \neq 0$
- (6) $\lim_{x \rightarrow a} \left((f(x))^n \right) = \left(\lim_{x \rightarrow a} f(x) \right)^n$
- (7) $\lim_{x \rightarrow a} c = c$
- (8) $\lim_{x \rightarrow a} x = a$
- (9) $\lim_{x \rightarrow a} x^n = a^n$ for every $n \in \mathbb{Z}$
- (10) $\lim_{x \rightarrow a} \sqrt[n]{x} = \sqrt[n]{a}$ for every odd $n \in \mathbb{Z}$ (works for even n also if $a > 0$)
- (11) $\lim_{x \rightarrow a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x \rightarrow a} f(x)}$ for every odd n (works for even n also if $\lim_{x \rightarrow a} f(x) > 0$)

Laws 1, 4, 5, 7, and 8 will be proved later, after we have given the precise definition of a limit. The remaining laws follow from these four:

- Law 3 follows from Laws 4 and 7 by putting $g(x) = c$.
- Law 2 follows from Laws 1 and 3 by writing $f - g = f + (-1)g$ and putting $c = -1$.
- Law 6 is proved by iterating Law 4.
- Law 9 follows from Laws 6 and 8.
- Law 11 follows from Law 6.
- Law 10 follows from Laws 8 and 11.

Exercise 5.4. Write down the details of the proofs of Laws 2, 3, 6, 9, 10, and 11 using Laws 1, 4, 5, 7, and 8 as suggested in the list above.

As an example of the limit laws in action, we can justify the examples from the start of this lecture. For the first two we note that

$$\begin{aligned} \lim_{x \rightarrow 2} (x + 2) &\stackrel{\text{Law 2}}{=} \lim_{x \rightarrow 2} x + \lim_{x \rightarrow 2} 2 \stackrel{\text{Laws 8 and 7}}{=} 2 + 2 = 4, \\ \lim_{x \rightarrow 2} \frac{1}{x + 2} &\stackrel{\text{Law 5 (and 7)}}{=} \frac{1}{\lim_{x \rightarrow 2} (x + 2)} = \frac{1}{4}. \end{aligned}$$

For the third, we have

$$\begin{aligned}
 \lim_{t \rightarrow 0} \frac{\sqrt{t^2 + 1} - 1}{t^2} &= \lim_{t \rightarrow 0} \frac{1}{\sqrt{t^2 + 1} + 1} && \text{algebra from (5.1)} \\
 &= \frac{\lim_{t \rightarrow 0} 1}{\lim_{t \rightarrow 0} (\sqrt{t^2 + 1} + 1)} && \text{by Law 5} \\
 &= \frac{1}{(\lim_{t \rightarrow 0} \sqrt{t^2 + 1}) + 1} && \text{by Laws 1 and 7} \\
 &= \frac{1}{\sqrt{(\lim_{t \rightarrow 0} t^2) + 1} + 1} && \text{by Laws 11, 1, and 7} \\
 &= \frac{1}{\sqrt{(\lim_{t \rightarrow 0} t)^2 + 1} + 1} && \text{by Law 9} \\
 &= \frac{1}{\sqrt{0^2 + 1} + 1} && \text{by Law 8} \\
 &= \frac{1}{2}.
 \end{aligned}$$

In practice we eventually carry out these steps without writing each of them explicitly, but when you are first encountering limits it is important to understand why the overall computation works.

We occasionally work with *one-sided limits*. For example, we say that L is the *left-hand limit* of f at a if we can make $f(x)$ arbitrarily close to L by taking x sufficiently close to a and to the left of a (that is, $x < a$). In this case we write

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

The *right-hand limit* is defined analogously and written $\lim_{x \rightarrow a^+}$.

Example 5.5. Let $f(x) = x$ for $x \leq 0$ and $f(x) = x + 1$ for $x > 0$. Then

$$\lim_{x \rightarrow 0^-} f(x) = 0 \text{ and } \lim_{x \rightarrow 0^+} f(x) = 1.$$

Finally, we mention *infinite limits*. We say that $\lim_{x \rightarrow a} f(x) = \infty$ if $f(x)$ can be made arbitrarily large by taking x sufficiently close to (but not equal to) a . We define $\lim_{x \rightarrow a} f(x) = -\infty$ analogously, and make the obvious definitions for infinite one-sided limits.

Example 5.6. Let $f(x) = \frac{1}{x}$ for $x \neq 0$. Then

$$\lim_{x \rightarrow 0^-} f(x) = -\infty \text{ and } \lim_{x \rightarrow 0^+} f(x) = \infty.$$

Definition 5.7. If any of the one-sided or two-sided limits of f at a is ∞ or $-\infty$, then we say that $x = a$ is a *vertical asymptote* of $y = f(x)$.

This lecture corresponds to §2.4 in Stewart and the middle of Chapter 5 in Spivak.

6.1. The definition at the heart of calculus

To prove the limit laws, we need the formal definition of limits, which uses the quantifier notation we introduced earlier.

Definition 6.1. Let f be a function that is defined on an open interval containing a , except possibly at a itself. Then $L \in \mathbb{R}$ is the limit of f as $x \rightarrow a$ if⁷

$$(6.1) \quad \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. if } 0 < |x - a| < \delta, \text{ then } |f(x) - L| < \varepsilon.$$

In this case we write $L = \lim_{x \rightarrow a} f(x)$.

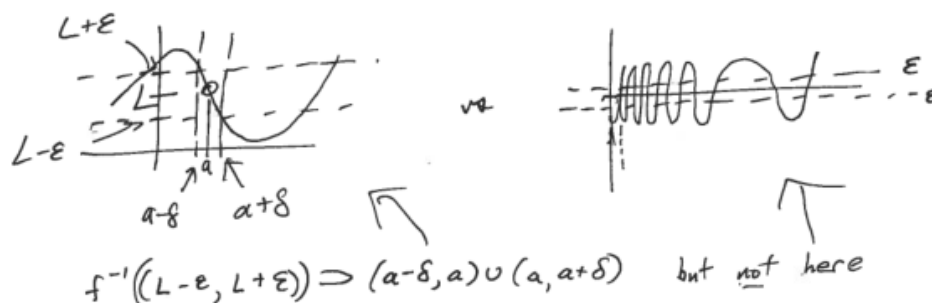


FIGURE 1. The ε - δ definition of limit

Figure 1 shows how the definition can be interpreted in terms of a graph: our adversary (recall Remark 2.7) chooses ε , which determines a horizontal strip with height 2ε , and we must choose δ such that the part of the graph lying inside the corresponding vertical strip (with width 2δ) is contained inside the horizontal strip chosen by our adversary. Observe that a smaller choice of ε will generally require a smaller choice of δ , so it is crucial that δ is allowed to depend on ε .

Exercise 6.2. Use the definition of limit to prove that $\lim_{x \rightarrow a} f(x) = \lim_{h \rightarrow 0} f(x + h)$, where by this equality we mean (as in the limit laws) that if one of the limits exists, then so does the other one, and in this case they are equal.

6.2. Some examples

Example 6.3. Earlier we claimed that $\lim_{x \rightarrow 2} (x + 2) = 4$. To prove this directly from the definition, observe that we are putting $f(x) = x + 2$, $a = 2$, and $L = 4$. Suppose our adversary chooses $\varepsilon > 0$. We want to choose x close enough to 2 that we are guaranteed to have $|(x + 2) - 4| < \varepsilon$. Observe that this ‘error term’ can be written as

$$|(x + 2) - 4| = |x - 2|,$$

⁷Here ε and δ are the Greek letters ‘epsilon’ and ‘delta’; one can imagine that they stand for “error” and “displacement”, respectively. If you do not yet know the Greek alphabet, you should learn it; mathematicians tend to run out of letters if they are restricted to one alphabet, and so it is useful to have another one handy.

and thus we have $|(x + 2) - 4| < \varepsilon$ if and only if $|x - 2| < \varepsilon$. Let $\delta = \varepsilon$. Then if x satisfies $0 < |x - 2| < \delta$, we must have $|x - 2| < \varepsilon$, and therefore $|f(x) - 4| < \varepsilon$, which proves that $\lim_{x \rightarrow 2}(x + 2) = 4$.

Example 6.4. Based on the previous example and the limit law for multiplication, we expect to find that $\lim_{x \rightarrow 2}(x + 2)^2 = 16$. So we put $f(x) = (x + 2)^2$, $a = 2$, and $L = 16$, then we check the definition. The error term $|f(x) - L|$ can be written as

$$|(x + 2)^2 - 16| = |x^2 + 4x + 4 - 16| = |x^2 + 4x - 12| = |(x - 2)(x + 6)|.$$

Thus once our adversary has chosen $\varepsilon > 0$, we have⁸

$$|f(x) - L| < \varepsilon \quad \Leftrightarrow \quad |(x - 2)(x + 6)| < \varepsilon.$$

Our goal is to choose $\delta > 0$ such that

$$(6.2) \quad \text{if } 0 < |x - 2| < \delta, \text{ then } |(x - 2)(x + 6)| < \varepsilon.$$

It is tempting to look at this and decide that we should make δ be equal to $\varepsilon/|x + 6|$ – after all, if $\delta = \varepsilon/|x + 6|$ and $|x - 2| < \delta$, then the bound we want follows immediately.

The problem with this is that δ is not allowed to depend on x . Remember the order of events in (6.1): first our adversary chooses ε , then we choose δ , and only after δ is chosen do we start checking the error estimate for various values of x .

Thus we must proceed in a different way, and address (6.2) in two steps. First we will require that $\delta \leq 1$, so that we have

$$(6.3) \quad \text{if } 0 < |x - 2| < \delta, \text{ then } 1 \leq 2 - \delta < x < 2 + \delta \leq 3, \text{ so } 7 < x + 6 < 9.$$

Using this bound, we deduce that

$$(6.4) \quad \text{if } 0 < |x - 2| < \delta, \text{ then } |(x - 2)(x + 6)| < 9\delta.$$

Thus if we also have $\delta \leq \varepsilon/9$, then we can use (6.4) to conclude that (6.2) is true. We set $\delta = \min(1, \varepsilon/9)$, and conclude that

$$0 < |x - 2| < \delta \Rightarrow |x + 6| < 9 \Rightarrow |f(x) - L| = |(x - 2)(x + 6)| < 9\delta \leq \varepsilon.$$

This proves that $\lim_{x \rightarrow 2}(x + 2)^2 = 16$.

When working through examples like these, it is very important to always keep in mind the distinction between a statement that you *are trying to prove is true*, such as (6.2), and a statement that you *have already proved is true*, such as (6.3) and (6.4). You will need to write down both sorts; remember which is which.

Example 6.5. To prove that $\lim_{x \rightarrow 2} \frac{1}{x+2} = \frac{1}{4}$, observe that we are putting $f(x) = \frac{1}{x+2}$, $a = 2$, and $L = \frac{1}{4}$. The error term $|f(x) - L|$ can be written as

$$\left| \frac{1}{x+2} - \frac{1}{4} \right| = \left| \frac{4 - (x+2)}{4(x+2)} \right| = \frac{|2-x|}{4|x+2|}.$$

Once our adversary chooses $\varepsilon > 0$, our goal is to choose $\delta > 0$ such that

$$\text{if } 0 < |x - 2| < \delta, \text{ then } \frac{|2-x|}{4|x+2|} < \varepsilon.$$

⁸The notation $P \Leftrightarrow Q$ means that statements P and Q are equivalent: if P is true then Q is true as well, and vice versa. In this case we often say that “ P is true if and only if Q is true”, and abbreviate the written form to “ P is true iff Q is true”.

To get some control on the denominator, we use the same trick as in the previous example and first assume that $\delta \leq 1$, so that if $0 < |x - 2| < \delta$, then $1 < x < 3$, and thus $3 < |x + 2| < 5$. It follows that for every such x we have

$$|f(x) - L| = \frac{|x - 2|}{4|x + 2|} \leq \frac{\delta}{12}.$$

Thus we put $\delta = \min(1, 12\varepsilon)$, and conclude that

$$0 < |x - 2| < \delta \Rightarrow \frac{1}{|x + 2|} < \frac{1}{3} \Rightarrow |f(x) - L| = \frac{|x - 2|}{4|x + 2|} < \frac{\delta}{12} \leq \varepsilon,$$

which proves that $\lim_{x \rightarrow 2} \frac{1}{x+2} = \frac{1}{4}$.

Already these examples involve enough calculations that you can imagine how much worse it would be to prove that $\lim_{t \rightarrow 0} \frac{\sqrt{t^2+1}-1}{t^2} = \frac{1}{2}$ directly from the definition. This illustrates the power of the limit laws, which we will prove soon.

6.3. Another formulation, using sequences

This section was skipped in the classroom lecture.

The formal definition of the limit of a function has a natural analogue for sequences.

Definition 6.6. Given a sequence x_n of real numbers, we say that L is the limit of x_n as $n \rightarrow \infty$, and write $L = \lim_{n \rightarrow \infty} x_n$ (or sometimes $x_n \rightarrow L$) if the following is true: for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$, we have $|x_n - L| < \varepsilon$.

Notice that the role of δ here is replaced by N , because we are interested in what happens as n becomes very large.

Exercise 6.7. Prove that $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$.

It is useful to bear in mind the following consequence of the definition of limit.

Proposition 6.8. *If $\lim_{x \rightarrow a} f(x) = L$ and x_n is a sequence converging to a such that $x_n \neq a$ for all n , then $\lim_{n \rightarrow \infty} f(x_n) = L$.*

Proof. For every $\varepsilon > 0$, the definition of limit gives $\delta > 0$ such that if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$. Similarly, the definition of limit of a sequence gives $N \in \mathbb{N}$ such that $|x_n - a| < \delta$ for all $n \geq N$. Since $x_n \neq a$ gives $0 < |x_n - a|$, we conclude that $|f(x_n) - L| < \varepsilon$ for all $n \geq N$. By the definition of limit of a sequence, this means that $\lim_{n \rightarrow \infty} f(x_n) = L$. \square

In fact the converse of this is true as well.

Proposition 6.9. *If $\lim_{x \rightarrow a} f(x) \neq L$, then there exists a sequence $x_n \rightarrow a$ such that $x_n \neq a$ for all n and $\lim_{n \rightarrow \infty} f(x_n) \neq L$.*

Proof. Since $\lim_{x \rightarrow a} f(x) \neq L$, our adversary has a winning move, and can choose $\varepsilon > 0$ such that no matter what $\delta > 0$ we choose, there is x with $0 < |x - a| < \delta$ such that $|f(x) - L| \geq \varepsilon$. In particular, given $n \in \mathbb{N}$ we can consider what happens when $\delta = \frac{1}{n}$; the previous sentence guarantees that there exists x_n with $0 < |x_n - a| < \frac{1}{n}$ such that $|f(x_n) - L| \geq \varepsilon$. Now it is a straightforward exercise to show that $x_n \rightarrow a$ and $f(x_n) \not\rightarrow L$. \square

Combining Propositions 6.8 and 6.9 gives the following useful criterion.

Corollary 6.10. $\lim_{x \rightarrow a} f(x) = L$ if and only if $\lim_{x_n \rightarrow a} f(x_n) = L$ for every sequence $x_n \rightarrow a$ with $x_n \neq a$.

6.4. A limit that does not exist

Let us briefly look at an example of how to use the definition to show that a limit does *not* exist. Let $f(x) = \sin(\frac{1}{x})$ for all $x > 0$; see the right-hand graph in Figure 1. To show that $\lim_{x \rightarrow 0^+} f(x)$ does not exist, we must show that no matter what value of $L \in \mathbb{R}$ we choose for the putative limit, our adversary can win the game by choosing a value of $\varepsilon > 0$ that makes (6.1) fail, demonstrating that $\lim_{x \rightarrow 0^+} f(x) \neq L$. Indeed, if we choose a value of L and then our adversary chooses $\varepsilon = \frac{1}{2}$, in order to win the game (which we could do if L is the limit) we would need to find $\delta > 0$ such that

$$(6.5) \quad \text{if } 0 < x < \delta, \text{ then } |f(x) - L| < \frac{1}{2}.$$

If we could do this, then for every $x, y \in (0, \delta)$, we would have

$$(6.6) \quad |f(x) - f(y)| = |(f(x) - L) + (L - f(y))| \leq |f(x) - L| + |f(y) - L| < \frac{1}{2} + \frac{1}{2} = 1.$$

Observe that for every $n \in \mathbb{N}$, the points

$$x_n := \frac{1}{(2n + \frac{1}{2})\pi} \quad \text{and} \quad y_n := \frac{1}{(2n - \frac{1}{2})\pi}$$

have the property that

$$f(x_n) = \sin(2\pi n + \frac{\pi}{2}) = \sin \frac{\pi}{2} = 1 \quad \text{and} \quad f(y_n) = \sin(2\pi n - \frac{\pi}{2}) = \sin(-\frac{\pi}{2}) = -1.$$

We want to choose n large enough that $x_n, y_n \in (0, \delta)$. Since $x_n < y_n$ it is enough to guarantee that $\frac{1}{(2n - \frac{1}{2})\pi} < \delta$, or equivalently, $2n - \frac{1}{2} > \frac{1}{\delta\pi}$. Thus by choosing $n \in \mathbb{N}$ with $n > \frac{1}{2\delta\pi} + \frac{1}{4}$, we obtain two points $x_n, y_n \in (0, \delta)$ for which

$$|f(x_n) - f(y_n)| = |1 - (-1)| = 2 > 1,$$

so that (6.6) is false. Since this happens no matter what $\delta > 0$ we choose, we conclude that we cannot win the game, and thus $\lim_{x \rightarrow 0^+} \sin(\frac{1}{x})$ does not exist.

Remark 6.11. Notice that the argument in the last part of this discussion, about choosing n large enough that $x_n, y_n \in (0, \delta)$, is exactly the proof that $\lim_{n \rightarrow \infty} y_n = 0$.

Remark 6.12. Instead of working directly with the definition of limit, we could observe that $x_n \rightarrow 0$ and $y_n \rightarrow 0$, with $x_n, y_n \neq 0$, but $\lim_{n \rightarrow \infty} f(x_n) \neq \lim_{n \rightarrow \infty} f(y_n)$, so no matter what value of L we choose, it is impossible to have both $L = \lim_{n \rightarrow \infty} f(x_n)$ and $L = \lim_{n \rightarrow \infty} f(y_n)$. By Corollary 6.10, this implies that $\lim_{x \rightarrow 0} f(x)$ does not exist.

Lecture 7

Proving the limit laws

DATE: WEDNESDAY, SEPTEMBER 4

This lecture corresponds to §2.4 and Appendix F in Stewart, and the end of Chapter 5 in Spivak.

As promised earlier, we now prove the limit laws from Theorem 5.3. It is enough to prove Laws 1, 4, 5, 7, and 8; the others are consequences of these, as explained there.

Exercise 7.1. Prove Limit Laws 7 and 8 using the definition of limit by observing that for Law 7, you can choose any $\delta > 0$ that you like,⁹ and for Law 8, you can choose $\delta = \varepsilon$.

Proposition 7.2 (Limit Law 1). *If $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$, then $\lim_{x \rightarrow a} (f(x) + g(x)) = L + M$.*

Proof. The error bound we wish to control can be estimated as follows:

$$|(f(x) + g(x)) - (L + M)| = |(f(x) - L) + (g(x) - M)| \leq \underbrace{|f(x) - L|}_{\text{I}} + \underbrace{|g(x) - M|}_{\text{II}}.$$

We can control I using the fact that $\lim_{x \rightarrow a} f(x) = L$, and II using the fact that $\lim_{x \rightarrow a} g(x) = M$. Indeed, once our adversary chooses $\varepsilon > 0$, then

- (1) we can choose $\delta_1 > 0$ such that $0 < |x - a| < \delta_1$ implies $|f(x) - L| < \frac{\varepsilon}{2}$, and
- (2) we can choose $\delta_2 > 0$ such that $0 < |x - a| < \delta_2$ implies $|g(x) - M| < \frac{\varepsilon}{2}$.

Let $\delta = \min(\delta_1, \delta_2)$. Then for every $0 < |x - a| < \delta$, we have

$$|(f(x) + g(x)) - (L + M)| \leq |f(x) - L| + |g(x) - M| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this proves that $\lim_{x \rightarrow a} (f(x) + g(x)) = L + M$, and completes the proof of Limit Law 1. \square

The remainder of this section was skipped in the classroom lecture

We go next to Law 4.

Proposition 7.3 (Limit Law 4). *If $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$, then $\lim_{x \rightarrow a} (f(x)g(x)) = LM$.*

Proof. Now the error term we must control is

$$\begin{aligned} |f(x)g(x) - LM| &= |f(x)g(x) - Lg(x) + Lg(x) - LM| \\ &\leq \underbrace{|g(x)||f(x) - L|}_{\text{I}} + \underbrace{|L||g(x) - M|}_{\text{II}}, \end{aligned}$$

where we use the trick of adding and subtracting the same expression in order to gain some control over what we are dealing with. (This trick will appear many times.) As in the previous proof, we control I and II using the limits of f and g , respectively; once our adversary has chosen $\varepsilon > 0$, we want to make each of them $< \varepsilon/2$. Start with II. Since $\lim_{x \rightarrow a} g(x) = M$, there is $\delta_1 > 0$ such that

$$(7.1) \quad \text{if } 0 < |x - a| < \delta_1, \text{ then } |g(x) - M| < \frac{\varepsilon}{2|L| + 1}.$$

This further implies that

$$|L||g(x) - M| < \frac{|L|\varepsilon}{2|L| + 1} < \frac{\varepsilon}{2},$$

⁹This is the one and only case in which δ does not depend on ε .

which controls II as desired.¹⁰ But what about I? To control this, we first use one more time the fact that $\lim_{x \rightarrow a} g(x) = M$, to choose $\delta_2 > 0$ such that

$$(7.2) \quad \text{if } 0 < |x - a| < \delta_2, \text{ then } |g(x) - M| < 1, \text{ so } |g(x)| < |M| + 1.$$

Then since $\lim_{x \rightarrow a} f(x) = L$, there is $\delta_3 > 0$ such that

$$(7.3) \quad \text{if } 0 < |x - a| < \delta_3, \text{ then } |f(x) - L| < \frac{\varepsilon}{2(|M| + 1)}.$$

Finally, we set $\delta = \min(\delta_1, \delta_2, \delta_3)$; then whenever $0 < |x - a| < \delta$, the inequalities in (7.1), (7.2), and (7.3) are all true, and we have

$$\begin{aligned} |f(x)g(x) - LM| &\leq |g(x)||f(x) - L| + |L||g(x) - M| \\ &\leq (|M| + 1)\frac{\varepsilon}{2(|M| + 1)} + |L|\frac{\varepsilon}{2|L| + 1} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this proves that $\lim_{x \rightarrow a} f(x)g(x) = LM$, as claimed. \square

Finally, we prove Law 5, starting with the special case when $f(x) = 1$.

Proposition 7.4. *If $\lim_{x \rightarrow a} g(x) = M \neq 0$, then $\lim_{x \rightarrow a} \frac{1}{g(x)} = \frac{1}{M}$.*

Proof. Suppose our adversary chooses $\varepsilon > 0$. The error term that we must control is

$$(7.4) \quad \left| \frac{1}{g(x)} - \frac{1}{M} \right| = \frac{|M - g(x)|}{|g(x)M|}.$$

The numerator becomes small when $x \approx a$, but what if the denominator also becomes small? We need to get a lower bound on $|g(x)|$, which we do by choosing $\delta_1 > 0$ small enough that

$$\text{if } 0 < |x - a| < \delta_1, \text{ then } |g(x) - M| < \frac{|M|}{2}, \text{ so } |g(x)| > |M| - \frac{|M|}{2} = \frac{|M|}{2}.$$

In this case we have $\frac{1}{|g(x)|} < \frac{2}{|M|}$, so the error term in (7.4) can be estimated as

$$\left| \frac{1}{g(x)} - \frac{1}{M} \right| = \frac{|g(x) - M|}{|M|} \frac{1}{|g(x)|} < \frac{2|g(x) - M|}{|M|^2}.$$

Once more using the fact that $\lim_{x \rightarrow a} g(x) = M$, there is $\delta_2 > 0$ such that

$$\text{if } 0 < |x - a| < \delta_2, \text{ then } |g(x) - M| < \frac{\varepsilon|M|^2}{2}.$$

Let $\delta = \min(\delta_1, \delta_2)$; then for every x with $0 < |x - a| < \delta$, both of the above estimates hold, and we have

$$\left| \frac{1}{g(x)} - \frac{1}{M} \right| < \frac{2}{|M|^2}|g(x) - M| < \frac{2}{|M|^2} \frac{\varepsilon|M|^2}{2} = \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this proves that $\lim_{x \rightarrow a} \frac{1}{g(x)} = \frac{1}{M}$. \square

¹⁰The reason we use $2|M| + 1$ in the denominator, and not $2|L|$, is that we might have $L = 0$.

The general case of Law 5 follows from this proposition together with Law 4: if $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$, then

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} f(x) \cdot \frac{1}{g(x)} = \left(\lim_{x \rightarrow a} f(x) \right) \left(\lim_{x \rightarrow a} \frac{1}{g(x)} \right) = L \cdot \frac{1}{M} = \frac{L}{M},$$

where the second equality uses Law 4 (limit of products) and the third equality uses Proposition 7.4.

Lecture 8 Theorems about limits

DATE: FRIDAY, SEPTEMBER 6

This lecture corresponds to §2.3 and Appendix F in Stewart, and parts of Chapter 5 in Spivak.

8.1. Direct substitution

Some limits can be computed via the method of “direct substitution”.

Proposition 8.1. *If f is a polynomial, then $\lim_{x \rightarrow a} f(x) = f(a)$ for every $a \in \mathbb{R}$.*

Proof. We prove this by induction in the degree of f . If $\deg f = 0$ then $f(x) = c$ is a constant, so the claim is true by Law 7. If the claim is true for polynomials of degree n , and f is a polynomial of degree $n + 1$, then $f(x) = cx^{n+1} + g(x)$, where $c \in \mathbb{R}$ and g is a polynomial of degree n , and we have

$$\begin{aligned} \lim_{x \rightarrow a} f(x) &= \lim_{x \rightarrow a} cx^{n+1} + \lim_{x \rightarrow a} g(x) && \text{by Law 1} \\ &= c \lim_{x \rightarrow a} x^{n+1} + \lim_{x \rightarrow a} g(x) && \text{by Law 3} \\ &= ca^{n+1} + g(a) && \text{by Law 9 and inductive hypothesis} \\ &= f(a). \end{aligned}$$

Thus the result holds for every n by induction. □

Proposition 8.2. *If f is a rational function and a is in the domain of f , then $\lim_{x \rightarrow a} f(x) = f(a)$.*

Proof. Let g, h be polynomials such that $f(x) = g(x)/h(x)$ for all x where $h(x) \neq 0$. By Proposition 8.1, we have

$$\lim_{x \rightarrow a} g(x) = g(a) \quad \text{and} \quad \lim_{x \rightarrow a} h(x) = h(a) \neq 0,$$

so Limit Law 5 gives

$$\lim_{x \rightarrow a} f(x) = \frac{\lim_{x \rightarrow a} g(x)}{\lim_{x \rightarrow a} h(x)} = \frac{g(a)}{h(a)} = f(a). \quad \square$$

8.2. Two-sided and one-sided limits

Now we relate two-sided and one-sided limits.

Theorem 8.3. $\lim_{x \rightarrow a} f(x) = L$ if and only if $\lim_{x \rightarrow a^-} f(x) = L = \lim_{x \rightarrow a^+} f(x)$.

Proof. (\Rightarrow):¹¹ If $\lim_{x \rightarrow a} f(x) = L$, then for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$\text{if } 0 < |x - a| < \delta, \text{ then } |f(x) - L| < \varepsilon.$$

In particular, $x \in (a, a + \delta) \Rightarrow |f(x) - L| < \varepsilon$, and since $\varepsilon > 0$ was arbitrary this implies that $\lim_{x \rightarrow a^+} f(x) = L$. Similarly, $x \in (a - \delta, a) \Rightarrow |f(x) - L| < \varepsilon$, and thus $\lim_{x \rightarrow a^-} f(x) = L$.

(\Leftarrow): If both of the one-sided limits exist and are equal to L , then for every $\varepsilon > 0$ there are $\delta_1, \delta_2 > 0$ such that

$$\text{if } x \in (a, a + \delta_1), \text{ then } |f(x) - L| < \varepsilon, \text{ and}$$

$$\text{if } x \in (a - \delta_2, a), \text{ then } |f(x) - L| < \varepsilon.$$

Taking $\delta = \min(\delta_1, \delta_2)$, we see that for every x with $0 < |x - a| < \delta$, we have $x \in (a, a + \delta_1)$ or $x \in (a - \delta_2, a)$, and in either case we get $|f(x) - L| < \varepsilon$. Thus for every $\varepsilon > 0$ we can produce the required $\delta > 0$, which shows that $\lim_{x \rightarrow a} f(x) = L$. \square

Example 8.4. Consider $\lim_{x \rightarrow 0} |x|$. For every $x > 0$ we have $|x| = x$, so

$$\lim_{x \rightarrow 0^+} |x| = \lim_{x \rightarrow 0^+} x = 0 \text{ by Limit Law 8.}$$

For every $x < 0$ we have $|x| = -x$, so

$$\lim_{x \rightarrow 0^-} |x| = \lim_{x \rightarrow 0^-} (-x) = - \lim_{x \rightarrow 0^-} x = -0 = 0 \text{ by Limit Laws 3 and 8.}$$

Since the one-sided limits exist and agree, Theorem 8.3 implies that $\lim_{x \rightarrow 0} |x| = 0$.

Example 8.5. Define $f: \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ by $f(x) = x/|x|$, so $f(x) = 1$ if $x > 0$ and -1 if $x < 0$. Then

$$\lim_{x \rightarrow 0^+} f(x) = \lim_{x \rightarrow 0^+} 1 = 1 \quad \text{and} \quad \lim_{x \rightarrow 0^-} f(x) = \lim_{x \rightarrow 0^-} -1 = -1 \quad \text{by Law 7,}$$

so by Theorem 8.3, $\lim_{x \rightarrow 0} f(x)$ does not exist.

8.3. Inequalities and limits

Finally, we prove two results demonstrating that inequalities between functions can be passed to the corresponding limits.

Theorem 8.6. If $f(x) \leq g(x)$ for every x , and if both $\lim_{x \rightarrow a} f(x)$ and $\lim_{x \rightarrow a} g(x)$ exist, then $\lim_{x \rightarrow a} f(x) \leq \lim_{x \rightarrow a} g(x)$.

Proof. Let $L = \lim_{x \rightarrow a} f(x)$ and $M = \lim_{x \rightarrow a} g(x)$. We use proof by contradiction. Suppose that $M < L$. Then by Law 2, we have

$$\lim_{x \rightarrow a} (g(x) - f(x)) = M - L < 0,$$

¹¹To prove an “if and only if” result, one often proves each direction separately. Here “(\Rightarrow)” means that we are proving that the first statement (two-sided limit) implies the second (one-sided limits), and “(\Leftarrow)” means that we are proving that the second implies the first.

and so putting $\varepsilon = L - M > 0$, the definition of limit gives $\delta_1 > 0$ such that

$$\text{if } 0 < |x - a| < \delta_1, \text{ then } 0 < |(g(x) - f(x)) - (M - L)| < \varepsilon = L - M,$$

and thus for all such x , we have

$$g(x) \leq |g(x) - f(x) - (M - L)| + f(x) + M - L < L - M + f(x) + M - L = f(x).$$

This contradicts the assumption that $f(x) < g(x)$, demonstrating that we must have $M \geq L$ after all. \square

Theorem 8.7 (Squeeze Theorem). *Suppose that f, g, h are functions such that $f(x) \leq g(x) \leq h(x)$ for every x , and that moreover $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = L$. Then we have $\lim_{x \rightarrow a} g(x) = L$.*

Proof. Given $\varepsilon > 0$, there are $\delta_1, \delta_2 > 0$ such that

$$\text{if } 0 < |x - a| < \delta_1, \text{ then } L - \varepsilon < f(x) < L + \varepsilon, \text{ and}$$

$$\text{if } 0 < |x - a| < \delta_2, \text{ then } L - \varepsilon < h(x) < L + \varepsilon.$$

Let $\delta = \min(\delta_1, \delta_2)$. Then if $0 < |x - a| < \delta$, the estimates on $f(x)$ and $h(x)$ both hold, so

$$L - \varepsilon < f(x) \leq g(x) \leq h(x) < L + \varepsilon,$$

which implies that $|g(x) - L| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, this proves that $\lim_{x \rightarrow a} g(x) = L$. \square

Remark 8.8. Both of the preceding theorems continue to hold if we replace two-sided limits by one-sided limits. Moreover, in the hypotheses of the theorems, it is enough for the inequalities to hold for every x that is sufficiently close to a ; in other words, if there is $\delta > 0$ such that the inequalities hold for all $0 < |x - a| < \delta$, then the conclusion of the theorem still holds.

Lecture 9

Continuity

DATE: MONDAY, SEPTEMBER 9

This lecture corresponds to §2.5 in Stewart and Chapter 6 in Spivak.

9.1. Definition and basic examples

The ‘direct substitution’ property from the start of the previous lecture is important enough to study at greater length, and we make the following definition.

Definition 9.1. A function f is *continuous at a point a* if $\lim_{x \rightarrow a} f(x) = f(a)$.

Note that this definition actually requires three things to be true:

- (1) $f(a)$ must be defined (a is in the domain of f);
- (2) the limit $\lim_{x \rightarrow a} f(x)$ exists;
- (3) the two values are equal.

If any of these three fails, then the function is not continuous at a ; in this case we say that f is *discontinuous at a* .

Example 9.2. Consider the piecewise constant *Heaviside function*

$$H(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0. \end{cases}$$

This function is continuous at a for every $a \neq 0$, but is discontinuous at 0, because the limit does not exist there.

The value of the Heaviside function “jumps” from 0 to 1 as x increases through 0. This type of behavior is important enough to give a name to.

Definition 9.3. If the left-hand and right-hand limits of f at a both exist, but take different values, then we say that f has a *jump discontinuity* at a .

Example 9.4. Consider the *floor function* $f(x) = \lfloor x \rfloor$, which takes x to the largest integer n such that $n \leq x$. This is continuous at every $x \in \mathbb{R} \setminus \mathbb{Z}$ and has a jump discontinuity at every $x \in \mathbb{Z}$.

Remark 9.5. Using the ϵ - δ definition of a limit, we see that f is continuous at a if and only if the following is true: for every $\epsilon > 0$ there exists $\delta > 0$ such that for every $|x - a| < \delta$, we have $|f(x) - f(a)| < \epsilon$.

Remark 9.6. Note that in the previous remark we wrote $|x - a| < \delta$ instead of the usual $0 < |x - a| < \delta$ that appears in the definition of the limit. The reason we can omit the first inequality is that when $|x - a| = 0$ we have $x = a$, and thus $f(x) = f(a)$, so $|f(x) - f(a)| = 0 < \epsilon$ automatically.

Exercise 9.7. Use the definition of limit to show that if $\lim_{x \rightarrow a} f(x)$ exists, then so does $\lim_{h \rightarrow 0} f(a + h)$, and the two limits are equal. Similarly, show that if $\lim_{h \rightarrow 0} f(a + h)$ exists, then so does $\lim_{x \rightarrow a} f(x)$, and the two limits are equal. In particular, a function f is continuous at a if and only if $\lim_{h \rightarrow 0} f(a + h) = f(a)$.

Definition 9.8. A function f is *right-continuous* at a if $\lim_{x \rightarrow a^+} f(x) = f(a)$, and *left-continuous* at a if $\lim_{x \rightarrow a^-} f(x) = f(a)$.

Recalling Theorem 8.3, we see immediately that f is continuous at a if and only if it is both left- and right-continuous at a .

Definition 9.9. A function f is *continuous on an interval I* if it is continuous at every point $a \in I$. If I contains an endpoint that is also an endpoint of the domain of f , then we interpret ‘continuous’ to mean either ‘left-continuous’ or ‘right-continuous’, as appropriate.

9.2. Polynomials, rational, and root functions are continuous

Limit Laws 9 and 10 say that the functions $f(x) = x^n$ and $g(x) = \sqrt[n]{x}$ are continuous on their entire domains. A generalization of this first fact is given by Proposition 8.1 from the last lecture, which says that if f is a polynomial, then $f(a) = \lim_{x \rightarrow a} f(x)$ for every $a \in \mathbb{R}$. This can be restated by saying that f is continuous at every $a \in \mathbb{R}$, or even more succinctly as “ f is continuous on \mathbb{R} ”. Similarly, Proposition 8.2 says that if

f is a rational function, then it is continuous on its domain (that is, the set of values of x for which the denominator is nonzero).

Remark 9.10. The domain of a rational function is always $\mathbb{R} \setminus A$, where A is a finite set. This is because if p, q are polynomials, then $f(x) = p(x)/q(x)$ is defined whenever $q(x) \neq 0$, and by Exercise 3.2, there are at most $\deg f$ values of x for which $q(x) = 0$.

Example 9.11.

- (1) $f(x) = \frac{1}{x}$ is continuous at a for every $a \neq 0$ and discontinuous at 0.
- (2) $f(x) = \frac{x^2-1}{x-1}$ is continuous at every $x \neq 1$ and discontinuous at 1, where it is undefined. Observe that the function $g(x) = x + 1$ is continuous at every x (including $x = 1$) and agrees with $f(x)$ everywhere that the latter is defined.

Definition 9.12. If a function f has the property that the left-hand and right-hand limits at a both exist and agree with each other, but disagree with $f(a)$ (which may or may not be defined), then we say that f has a *removable discontinuity* at a .

Removable discontinuities primarily occur when a function admits some algebraic simplification. For example, the function

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

is undefined at $x = 0$ (and indeed at every multiple of π) because $\sin x$ vanishes there, but the fundamental trig identity $\cos^2 x + \sin^2 x = 1$ gives $f(x) = \sin x$ at every x where f is defined, so the discontinuities of f are removable.

Remark 9.13. When we discussed the notation “ f^{-1} ” for the inverse function of f , we used the notation $f^n(x)$ to denote the result of *composing* f with itself n times, so

$$f^n(x) = \overbrace{f \circ f \circ \cdots \circ f}^{n \text{ times}}(x).$$

In particular, according to that convention, $f^2(x)$ would mean $f(f(x))$, and we would write $f(x)^2$ or $(f(x))^2$ to denote $f(x)f(x)$. However, with the trigonometric functions it is standard to write $\cos^2 x$ or $\cos^2(x)$ to mean $(\cos x)(\cos x)$, instead of $\cos(\cos x)$, and from now on we will use this ‘power’ notation to mean multiplication of trigonometric functions. The one exception is that $\cos^{-1} x$ will still mean the inverse function arccosine.

This situation of ambiguous notation, where the same way of writing things can have multiple meanings, is unfortunate but occurs from time to time in mathematics. In general you can deduce the meaning from context, but you should be aware that this possibility exists.

9.3. New continuous functions from old ones

Theorem 9.14. *If f and g are functions that are continuous at $a \in \mathbb{R}$, then the following functions are also continuous at a :*

- (1) $f + g$;
- (2) $f - g$;
- (3) cf for every $c \in \mathbb{R}$;
- (4) fg (note that this means the product $(fg)(x) = f(x)g(x)$ rather than the composition);

(5) f/g provided $g(a) \neq 0$.

Proof. These assertions follow immediately from the corresponding limit laws. \square

Theorem 9.15. *If f, g are real-valued functions such that g is continuous at a and f is continuous at $g(a)$, then $f \circ g$ is continuous at a .*

Proof. Given $\epsilon > 0$, we want to produce $\delta > 0$ such that for every $|x - a| < \delta$, we have $|f(g(x)) - f(g(a))| < \epsilon$. To accomplish this, we proceed as follows.

- (1) Use the fact that f is continuous at $g(a)$ to deduce that there exists $\delta_1 > 0$ such that for all y with $|y - g(a)| < \delta_1$, we have $|f(y) - f(g(a))| < \epsilon$.
- (2) Use continuity of g at a to deduce that there exists $\delta > 0$ such that for all x with $|x - a| < \delta$, we have $|g(x) - g(a)| < \delta_1$.

Now given any x with $|x - a| < \delta$, the second item gives $|g(x) - g(a)| < \delta_1$, and then the first gives $|f(g(x)) - f(g(a))| < \epsilon$, which proves the theorem. \square

Example 9.16. Because $x^2 + 1 \geq 0$ for all $x \in \mathbb{R}$ and the square root function is continuous on $[0, \infty)$, it follows from Theorem 9.15 that the function $f(x) = \sqrt{x^2 + 1}$ is continuous on \mathbb{R} . Note that we could also have deduced this using Limit Law 11.

9.4. Trigonometric functions are continuous

Theorem 9.17. *The sine and cosine functions are both continuous on \mathbb{R} .*

Proof. We start by proving that both of these functions are continuous at 0; then we use the sum-of-angles formulas (3.3) and (3.4) to deduce continuity at every $a \in \mathbb{R}$. For continuity at 0, we start by recalling the definition of cosine and sine, as illustrated in the left-hand side of Figure 2, which suggests that $\lim_{\theta \rightarrow 0} \cos \theta = 1$ and $\lim_{\theta \rightarrow 0} \sin \theta = 0$, so that \cos and \sin are continuous at 0.

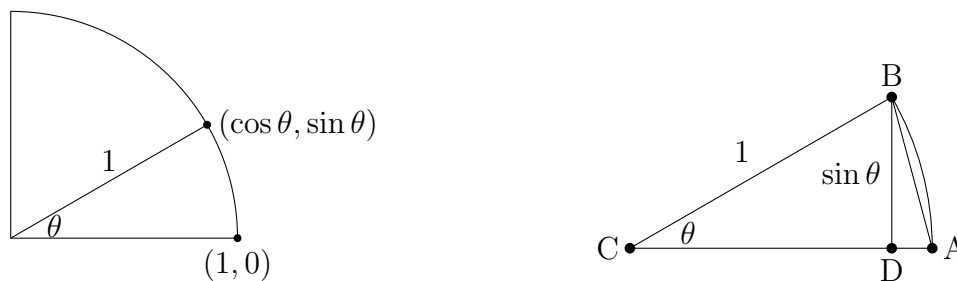


FIGURE 2. Proving continuity of \sin at 0.

To prove this rigorously, we can use the right-hand side of Figure 2 to observe that

$$\sin \theta = BD < AB \leq (\text{length of arc from } A \text{ to } B) = \theta,$$

where the first inequality is because the hypotenuse of a right triangle is longer than either leg, and the second inequality is because the length of any curve between two points is at least as large as the distance between them. Thus for $\theta \in (0, \pi/2)$ we have $0 < \sin \theta < \theta$. Since $\lim_{\theta \rightarrow 0} 0 = \lim_{\theta \rightarrow 0} \theta = 0$, the squeeze theorem proves that $\lim_{\theta \rightarrow 0^+} \sin \theta = 0$. Since $\sin(-\theta) = -\sin \theta$, we conclude that $\lim_{\theta \rightarrow 0} \sin \theta = 0$, so \sin is continuous at 0.

For the corresponding result for cosine, we use this fact together with the identity $1 - \cos^2 \theta = \sin^2 \theta$, which implies $1 - \cos \theta = \frac{\sin^2 \theta}{1 + \cos \theta}$, to get

$$\lim_{\theta \rightarrow 0} (1 - \cos \theta) = \frac{(\lim_{\theta \rightarrow 0} \sin \theta)^2}{\lim_{\theta \rightarrow 0} (1 + \cos \theta)} = \frac{0}{2} = 0,$$

so $\lim_{\theta \rightarrow 0} \cos \theta = 1$, which shows that \cos is continuous at 0.

The remainder of this proof was done in class on Wednesday, September 11.

With these two results in hand, we can use (3.3) to deduce that for any $a \in \mathbb{R}$, the sine function is continuous at a . Indeed, from Exercise 9.7 we see that it suffices to show that $\lim_{h \rightarrow 0} \sin(a + h) = \sin a$, and the sum-of-angles formula (3.3) gives

$$\begin{aligned} \lim_{h \rightarrow 0} \sin(a + h) &= \lim_{h \rightarrow 0} (\sin a \cos h + \cos a \sin h) = \sin a \left(\lim_{h \rightarrow 0} \cos h \right) + \cos a \left(\lim_{h \rightarrow 0} \sin h \right) \\ &= (\sin a) \cdot 1 + (\cos a) \cdot 0 = \sin a. \end{aligned}$$

This proves continuity of $x \mapsto \sin x$ on \mathbb{R} . For cosine, we use (3.4) to get

$$\begin{aligned} \lim_{h \rightarrow 0} \cos(a + h) &= \lim_{h \rightarrow 0} (\cos a \cos h - \sin a \sin h) = \cos a \left(\lim_{h \rightarrow 0} \cos h \right) - \sin a \left(\lim_{h \rightarrow 0} \sin h \right) \\ &= (\cos a) \cdot 1 - (\sin a) \cdot 0 = \cos a, \end{aligned}$$

and thus $x \mapsto \cos x$ is also continuous on \mathbb{R} . This completes the proof of Theorem 9.17. \square

Exercise 9.18. Use the results proved so far to show that \tan , \cot , \sec , and \csc are all continuous on their domains.

Using Theorems 9.15 and 9.17 together with the fact that rational functions are continuous, we deduce the following.

Example 9.19.

- (1) $\sin \frac{1}{x}$ is continuous at every $x \neq 0$. There is no way to extend this function to a continuous function $f: \mathbb{R} \rightarrow \mathbb{R}$ such that f is continuous at 0. That is, if $f: \mathbb{R} \rightarrow \mathbb{R}$ is any function such that $f(x) = \sin \frac{1}{x}$ for all $x \neq 0$, then f is discontinuous at 0.
- (2) The function

$$f(x) := \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

is continuous at every $x \in \mathbb{R}$. Continuity at $x \neq 0$ is a consequence of the previous part and Theorem 9.14 (product of continuous functions is continuous). For continuity at $x = 0$, we use the Squeeze Theorem, observing that $-|x| \leq x \sin \frac{1}{x} \leq |x|$ for all $x \neq 0$ and $\lim_{x \rightarrow 0} -|x| = \lim_{x \rightarrow 0} |x| = 0$.

At this point one may reasonably ask about the inverse trigonometric functions. Are they continuous? We will eventually see that they are, but our proof will require the *Intermediate Value Theorem*, which we discuss in the next lecture.

9.5. What about exponentials?

The details of this section were skipped in the classroom lecture.

Recall that given $a > 0$, we defined $f(x) = a^x$ in the following way.

- When $x \in \mathbb{N}$, it is defined iteratively (multiply a by itself x times).
- To extend to $x \in \mathbb{Z}$, we take reciprocals: $a^{-x} = 1/a^x$.
- To extend to $x \in \mathbb{Q}$, we take roots: $a^{p/q}$ is the q th root of a^p . (There's a subtlety here, though – why do q th roots exist?)
- To extend to $x \in \mathbb{R}$, we take limits: a^x is the limit of a^{r_n} when r_n is a sequence of rational numbers approaching x .

The last step here strongly suggests that the exponential function is continuous, and indeed it is. As with the sine and cosine functions, we start by proving continuity at 0.

In fact, we start by proving something about the behavior of the exponential function when the exponent gets large.

Lemma 9.20. *For every $a \geq 1$ and $b > 1$, there exists $n \in \mathbb{N}$ such that $b^n > a$.*

Proof. Let $t = b - 1$, then $b^n = (1 + t)^n \geq 1 + tn$, where the last inequality can be proved by induction or by observing that $(1 + t)^n = 1 + tn + \frac{n(n-1)}{2}t^2 + \dots + t^n \geq 1 + tn$ since all the extra terms are ≥ 0 . Thus it suffices to take $n > \frac{a-1}{t}$, since this gives $b^n \geq 1 + tn > 1 + (a - 1) = a$. \square

Now we prove right-continuity at 0 when $a \geq 1$.

Lemma 9.21. *For every $a \geq 1$, we have $\lim_{x \rightarrow 0^+} a^x = 1$.*

Proof. Given $\epsilon > 0$, use Lemma 9.20 to find $n \in \mathbb{N}$ such that $(1 + \epsilon)^n > a$, and let $\delta = 1/n$. Then taking n th roots gives $a^\delta = a^{1/n} < 1 + \epsilon$. Now for every $x \in (0, \delta)$ we have

$$1 \leq a^x \leq a^\delta < 1 + \epsilon,$$

where the second inequality uses the fact that $a^x \leq a^y$ whenever $a \geq 1$ and $x \leq y$. Because $\epsilon > 0$ was arbitrary, this completes the proof. \square

Remark 9.22. Although this proof certainly establishes the desired result, it is a little bit cryptic because it gives no indication of how we decided on that particular choice of δ . The reasoning behind this choice is as follows: “Given $\epsilon > 0$, we want to find $\delta > 0$ such that every $x \in (0, \delta)$ has $|a^x - 1| < \epsilon$. Since $a^x > 1$ for all $x > 0$, this is equivalent to showing that $a^x < 1 + \epsilon$. Moreover, since $a^x < a^\delta$ for all $x \in (0, \delta)$, this is equivalent to choosing $\delta > 0$ such that $a^\delta < 1 + \epsilon$. This last inequality is equivalent to $a < (1 + \epsilon)^{1/\delta}$.”

Now we use a similar argument to get left-continuity.

Lemma 9.23. *For every $a \geq 1$, we have $\lim_{x \rightarrow 0^-} a^x = 1$.*

Proof. Given $\epsilon > 0$, use Lemma 9.20 to find $n \in \mathbb{N}$ such that $(\frac{1}{1-\epsilon})^n > a$, and let $\delta = 1/n$. Raising both sides to the power $-1/n$ gives $1 - \epsilon < a^{-1/n}$, and thus for every $x \in (-\delta, 0)$ we have $1 - \epsilon < a^{-\delta} \leq a^{-x} \leq 1$. \square

Combining Lemmas 9.21 and 9.23 shows that $f(x) = a^x$ is continuous at 0 for every $a \geq 1$. In fact it is continuous at *every* $x \in \mathbb{R}$:

$$\lim_{h \rightarrow 0} a^{x+h} = \lim_{h \rightarrow 0} a^x a^h = a^x \lim_{h \rightarrow 0} a^h = a^x \cdot 1 = a^x.$$

Here the first equality uses the basic property of exponentials, the second equality uses Limit Law 3, and the third uses continuity at 0. Finally, we observe that for $0 < a < 1$ we have

$$\lim_{y \rightarrow x} a^y = \lim_{y \rightarrow x} (1/a)^{-y} = (1/a)^{-x} = a^x$$

for all $x \in \mathbb{R}$, where the second equality uses continuity for bases that are ≥ 1 . Putting it all together, we have proved the following.

Theorem 9.24. *For every $a > 0$, the function $f(x) = a^x$ is continuous on \mathbb{R} .*

It is also natural to ask about the logarithm function, which is the inverse function of the exponential, and just as with the inverse trig functions, our proof of continuity will need to wait until we have discussed the Intermediate Value Theorem.

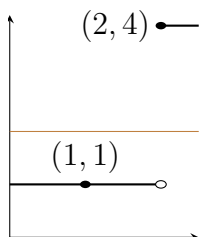
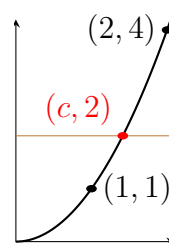
Lecture 10 Intermediate Value Theorem: Preparation

DATE: WEDNESDAY, SEPTEMBER 11

This lecture corresponds to §2.5 of Stewart and parts of Chapters 7 and 8 of Spivak; however, Stewart does not prove the theorem, and Spivak's proof differs from ours.

Now we return to a question raised in the opening lecture: Why does $\sqrt{2}$ exist? To put this a little bit more precisely, why is there a real number x with the property that $x^2 = 2$?

Informally, we might reason as follows: *The graph of the function $f(x) = x^2$ is a parabola opening upwards, as shown in the picture. The graph goes through the point $(1, 1)$, which is below the line $y = 2$, and through the point $(2, 4)$, which is above it. In order to go between these points, the graph has to cross the line $y = 2$ at some point $(c, 2)$, and then we must have $f(c) = c^2 = 2$, so $c = \sqrt{2}$.*



This reasoning is made precise by the *Intermediate Value Theorem*, which we will state in a moment. Before doing so, we observe that this reasoning must use some property of the function $f(x) = x^2$. The picture at left shows the function

$$g(x) = \begin{cases} 1 & x < 2 \\ 4 & x \geq 2 \end{cases},$$

which also has the property that its graph goes through the points $(1, 1)$ and $(2, 4)$; however, there is no value of c for which $g(c) = 2$.

This example illustrates that a function can have a discontinuity that lets it jump from one side of a line to the other without intersecting it. The Intermediate Value Theorem says that a discontinuity is the *only* way that this can happen.

Theorem 10.1 (Intermediate Value Theorem). *If $f: [a, b] \rightarrow \mathbb{R}$ is continuous and $f(a) < r < f(b)$, then there exists a real number $c \in [a, b]$ such that $f(c) = r$.*

Remark 10.2. The IVT is not true if we replace “real number” by “rational number”. Indeed, as we saw in Theorem 1.1, there is no rational number c such that $c^2 = 2$.

Remark 10.3. You may sometimes see the IVT summarized as the statement that “if you draw a curve from one side of a line to the other without lifting your pen from the paper, then you must intersect the line somewhere.” And stated this way, the theorem seems obvious; how could it be otherwise? We cannot imagine how the theorem could fail, so what is there to prove?

However, “we cannot imagine X happening” is quite different from saying that X never happens. After all, the early Pythagoreans could not imagine that there were two lengths whose ratio could not be expressed as a rational number,¹² but it turns out that the diagonal and side of a square form such a pair, since their ratio is $\sqrt{2}$. So until we produce a genuine proof, we must consider the possibility that perhaps there is a failure in our imagination.

Another way of thinking about this is the following: the statement involving “drawing a curve without lifting your pen from the paper” is meant to describe the graph of a continuous function. However, the definitions of continuity in Definition 9.1 and Remark 9.5 are quite technical; what do they have to do with “drawing without lifting your pen”? One can interpret the IVT as providing a connection between the technical definition and the intuitive one.

Before embarking on the proof itself, we make a mild simplification. The precise values of a, b, r turn out not to be particularly important to the theorem; what is important is the relationship $f(a) < r < f(b)$. Looking at the two functions shown below, one may reasonably expect that the same argument should work for both of them, even though the values are different.¹³



The function g in the right-hand picture is obtained from the function f in the left-hand picture by rescaling horizontally (the interval $[1, 4]$ gets mapped to the interval $[0, 1]$) and shifting vertically (the line $y = 2$ gets moved to the line $y = 0$). In fact, suppose we have *any* real numbers $a < b$, a function $f: [a, b] \rightarrow \mathbb{R}$, and a real number r such that $f(a) < r < f(b)$. Then the function

$$g(x) := f(a + (b - a)x) - r$$

has the following properties:

¹²A pair of such lengths are called *incommensurable*.

¹³Also note that as these pictures show, there may well be more than one value of c with $f(c) = r$. To claim that there is a *unique* $c \in [a, b]$ with this property requires some extra information about f .

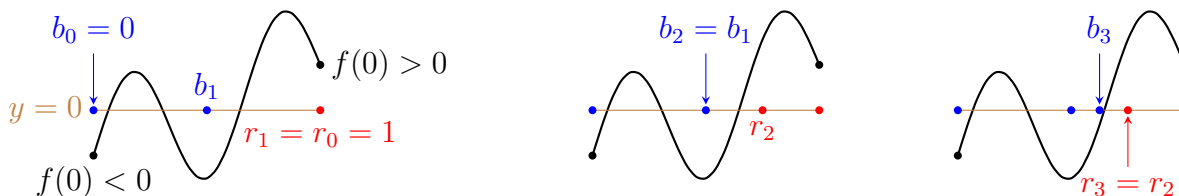
- $g: [0, 1] \rightarrow \mathbb{R}$ is continuous;
- $g(0) < 0 < g(1)$; and
- given $c \in [0, 1]$, we have $g(c) = 0$ if and only if $f(a + (b - a)c) = r$.

In light of this, we will first prove the IVT under the additional assumption that $a = 0$, $b = 1$, and $r = 0$. The procedure just described shows that if we know the theorem is true in this case, then it is true for arbitrary values of a, b, r .¹⁴

Theorem 10.4 (IVT: simple form). *If $f: [0, 1] \rightarrow \mathbb{R}$ is continuous and $f(0) < 0 < f(1)$, then there exists a real number $c \in [0, 1]$ such that $f(c) = 0$.*

Our proof of Theorem 10.4 uses *bisection sequences*.¹⁵ We build two sequences $b_n, r_n \in [0, 1]$ as shown in the pictures below (where there are blue points and red points, respectively) by the following iterative procedure:

- (1) start by putting $b_0 = 0$ and $r_0 = 1$;
- (2) if b_n, r_n have been defined, then take $m_n = \frac{1}{2}(b_n + r_n)$ to be the midpoint of $[b_n, r_n]$, and define b_{n+1}, r_{n+1} as follows:
 - if $f(m_n) = 0$, then stop (we found the desired c);
 - if $f(m_n) < 0$, then let $b_{n+1} = m_n, r_{n+1} = r_n$;
 - if $f(m_n) > 0$, then let $b_{n+1} = b_n, r_{n+1} = m_n$.



If the first case above ($f(m_n) = 0$) happens for some n , then we put $c = m_n$ and the theorem is proved. So we need to consider the situation where the first case never happens, and the sequences b_n, r_n go on forever. Then the following properties are immediate consequences of the definitions:

- $0 \leq b_0 \leq b_1 \leq b_2 \leq \dots \leq r_2 \leq r_1 \leq r_0 = 1$;
- $f(b_n) < 0 < f(r_n)$ for every $n \in \mathbb{N}$;
- $r_n - b_n = 2^{-n}$ for every $n \in \mathbb{N}$.

The pictures suggest that the root c of the equation $f(c) = 0$ should lie to the right of every b_n , and to the left of every r_n . In fact, we expect to see that $c = \lim_{n \rightarrow \infty} b_n = \lim_{n \rightarrow \infty} r_n$, where we recall that the definition of limit of a sequence was given in Definition 6.6. In the next lecture, we will complete the proof via the following steps:

- $b = \lim_{n \rightarrow \infty} b_n$ and $r = \lim_{n \rightarrow \infty} r_n$ both exist;
- $f(b) \leq 0$ and $f(r) \geq 0$;
- $b = r$;
- putting $c = b = r$ gives $f(c) = 0$.

¹⁴Mathematicians often summarize this whole discussion by saying, “Without loss of generality, we assume that $a = 0, b = 1, r = 0$ ”. This language means that we prove a specific case, which turns out to contain all the ingredients for the general case.

¹⁵Our proof follows an article by Stephen M. Walk entitled “The Intermediate Value Theorem is NOT Obvious—and I Am Going to Prove It to You”, which appeared in *The College Mathematics Journal*, vol. 42, No. 4 (Sep. 2011), pp. 254–259.

Lecture 11

IVT: Proof and consequences

DATE: FRIDAY, SEPTEMBER 13

The first parts of this lecture correspond to §2.5 of Stewart and parts of Chapters 7 and 8 of Spivak; however, Stewart does not prove the theorem, and Spivak's proof differs from ours. The end of this lecture corresponds to §2.6 of Stewart and parts of Chapter 5 of Spivak.

11.1. Completion of the proof

To finish proving the Intermediate Value Theorem, we need to justify the list of claims at the end of the previous lecture.

The first of these is a fundamental property of the real numbers: *every bounded nondecreasing sequence converges to a real number*. More precisely, the following is true.

Monotone Convergence Theorem for real numbers. *Let x_1, x_2, x_3, \dots be a sequence of real numbers with the following properties:*

- *the sequence is nondecreasing, meaning that $x_n \leq x_{n+1}$ for every $n \in \mathbb{N}$;*
- *the sequence is bounded above, meaning that there is some $B \in \mathbb{R}$ such that $x_n \leq B$ for every $n \in \mathbb{N}$.*

Then there exists some $x \in \mathbb{R}$ such that $x = \lim_{n \rightarrow \infty} x_n$.

Remark 11.1. We will not prove this theorem; rather, we will take it as a fundamental property of the real numbers that distinguishes them from smaller sets of numbers such as \mathbb{Q} . (Note that the theorem fails if we replace \mathbb{R} by \mathbb{Q} : consider the sequence 1, 1.4, 1.41, 1.414, 1.4142, \dots that converges to $\sqrt{2}$, which is not in \mathbb{Q} .) In order to prove this theorem one needs to *construct* the real numbers, which we will not do in this course.

Remark 11.2. A number B with the property stated above is called an *upper bound* for the sequence $\{x_n\}$. The real number x produced by the theorem is in fact an upper bound for the sequence (can you prove this?), and actually satisfies the stronger property of being a *least* upper bound: in other words, every number $y < x$ is *not* an upper bound for the sequence. If you are interested in exploring the foundations of the subject, it is a worthwhile exercise to prove that the monotone convergence theorem stated above is equivalent to the *least upper bound property*, which says that every *set* of real numbers that is bounded above has a least upper bound. This least upper bound property is used in Spivak's book.

Since b_n is a nondecreasing sequence, the Monotone Convergence Theorem implies that $b := \lim_{n \rightarrow \infty} b_n$ exists. For the sequence r_n , we observe that it is nonincreasing, meaning that $r_{n+1} \leq r_n$ for all n . We use the following consequence of the MCT.

Corollary 11.3. *If $x_1 \geq x_2 \geq x_3 \geq \dots$ is a nonincreasing sequence of real numbers that is bounded below, then $\lim_{n \rightarrow \infty} x_n$ exists.*

Proof. Apply the Monotone Convergence Theorem to the nondecreasing sequence $-x_1 \leq -x_2 \leq -x_3 \leq \dots$. □

This corollary shows that $r := \lim_{n \rightarrow \infty} r_n$ exists, so we have proved the first claim on our list. For the second claim, we observe that $f(b_n) < 0$ for all n , and recall Theorem 8.6, which said that if $f(x) \leq g(x)$ for all x , then $\lim_{x \rightarrow a} f(x) \leq \lim_{x \rightarrow a} g(x)$ provided both limits exist. It is easy to prove an analogous result for sequences:

$$(11.1) \quad \text{if } x_n \leq y_n \text{ for all } n, \text{ then } \lim_{n \rightarrow \infty} x_n \leq \lim_{n \rightarrow \infty} y_n \text{ provided both limits exist.}$$

Putting $x_n = f(b_n)$ and $y_n = 0$, this gives

$$\begin{aligned} f(b) &= f\left(\lim_{n \rightarrow \infty} b_n\right) && \text{by definition of } b \\ &= \lim_{n \rightarrow \infty} f(b_n) && \text{by continuity of } f \\ &\leq 0 && \text{by (11.1).} \end{aligned}$$

A similar argument with r_n and r gives $f(r) \geq 0$, so we have proved

$$(11.2) \quad f(b) \leq 0 \leq f(r).$$

Now we recall that from the definition of the sequences b_n, r_n , we have $r_n - b_n = 2^{-n}$ for all n . In particular, we have

$$r - b = \left(\lim_{n \rightarrow \infty} r_n\right) - \left(\lim_{n \rightarrow \infty} b_n\right) = \lim_{n \rightarrow \infty} (r_n - b_n) = \lim_{n \rightarrow \infty} 2^{-n},$$

where the second inequality uses Limit Law 2.

Exercise 11.4. Use Lemma 9.20 to prove that $\lim_{n \rightarrow \infty} 2^{-n} = 0$.

The exercise implies that $r - b = 0$, so $r = b$ and we can define $c = r = b$. Then (11.2) implies that $f(c) \leq 0 \leq f(c)$, which is only possible if $f(c) = 0$. This completes the proof of the IVT.

11.2. Applications of the IVT

Theorem 11.5. *If $f: [0, 1] \rightarrow [0, 1]$ is continuous, then there exists $x \in [0, 1]$ such that $f(x) = x$; such an x is called a fixed point for f .*

Proof. Because f is continuous, so is $g(x) := x - f(x)$. If $f(0) = 0$ or $f(1) = 1$, then we are done. If $f(0) > 0$ and $f(1) < 1$, then $g(0) = 0 - f(0) < 0$ and $g(1) = 1 - f(1) > 0$, so the IVT applies to g on $[0, 1]$ and gives $x \in (0, 1)$ such that $g(x) = 0$. But this means that $x - f(x) = 0$, so $f(x) = x$. \square

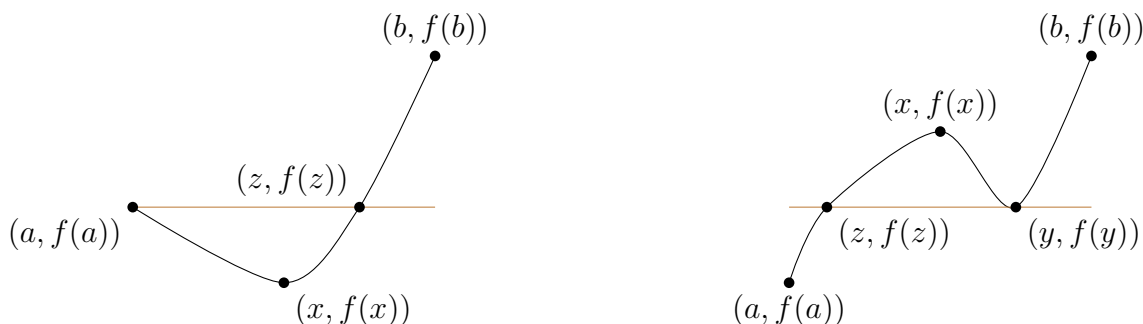
Note that the hypothesis of the preceding theorem requires that $0 \leq f(x) \leq 1$ for all $0 \leq x \leq 1$, so the theorem only applies when the image of the unit interval $[0, 1]$ lies inside the unit interval.

Theorem 11.6. *If $f: [a, b] \rightarrow \mathbb{R}$ is 1-1 and continuous, then it is either an increasing function ($x < y$ implies $f(x) < f(y)$) or a decreasing function ($x < y$ implies $f(x) > f(y)$). In either case, its range is the closed interval I whose endpoints are $f(a)$ and $f(b)$, and the inverse function $f^{-1}: I \rightarrow [a, b]$ is continuous as well.*

Proof. Since f is 1-1, we either have $f(a) < f(b)$ or $f(a) > f(b)$. We give the proof in the case $f(a) < f(b)$, when f will turn out to be increasing; the proof for $f(a) > f(b)$, when f ends up being decreasing, is completely analogous.

First we prove that for every $x \in (a, b)$, we have $f(x) > f(a)$. We prove this by contradiction; we assume that there is some $x \in (a, b)$ for which $f(x) \leq f(a)$, then we derive a contradiction from this, and conclude that our assumption must have been false.

Indeed, if there is $x \in (a, b)$ such that $f(x) \leq f(a)$, then since f is 1-1 we have $f(x) < f(a) < f(b)$, as in the left-hand picture below. By applying the IVT to f on the interval $[x, b]$, we conclude that there is $z \in [x, b]$ such that $f(z) = f(a)$. Since $a < x$, we have $z \neq x$, which contradicts the fact that f is 1-1. Thus we have proved that $f(x) > f(a)$ for all $x \in (a, b)$.



A similar argument shows that $f(x) < f(b)$ for all $x \in (a, b)$, and we conclude that

$$(11.3) \quad f(a) < f(x) < f(b) \text{ for all } x \in (a, b).$$

In other words, $\text{range}(f) \subset [f(a), f(b)] =: I$. In fact, given any $y \in [f(a), f(b)]$, the IVT implies that there is $x \in [a, b]$ such that $f(x) = y$, and we conclude that $\text{range}(f) = I$, so that $f: [a, b] \rightarrow I$ is a bijection.

Now we must prove that f is increasing and that f^{-1} is continuous. To prove that it is increasing, we again proceed by contradiction. Suppose that it is *not* increasing; then there exist $x < y$ such that $f(x) > f(y)$. Because both $f(x)$ and $f(y)$ lie in I , we must have $f(a) < f(y) < f(x)$ as shown in the right-hand picture above. Applying the IVT to f on $[a, x]$ gives $z \in [a, x]$ such that $f(z) = f(y)$. Since $z \neq y$, this contradicts the assumption that f is 1-1. We conclude that f is increasing.

Finally, we show that $f^{-1}: I \rightarrow [a, b]$ is continuous. Given $c \in I$ and $\epsilon > 0$, let $x = f^{-1}(c)$ and consider the interval $(x - \epsilon, x + \epsilon)$. By the first part of the proof above, we see that

$$(11.4) \quad \{f(z) : z \in (x - \epsilon, x + \epsilon)\} = (f(x - \epsilon), f(x + \epsilon)).$$

Let $\delta_1 = c - f(x - \epsilon)$ and $\delta_2 = f(x + \epsilon) - c$; note that $\delta_1, \delta_2 > 0$ because f is increasing. Let $\delta = \min(\delta_1, \delta_2)$; then for every y with $0 < |y - c| < \delta$, we have

$$y \in (c - \delta, c + \delta) \subset (f(x - \epsilon), f(x + \epsilon)),$$

and by (11.4) we have $f^{-1}(y) \in (x - \epsilon, x + \epsilon)$. Rewriting this we get $|f^{-1}(y) - f^{-1}(c)| = |f^{-1}(y) - x| < \epsilon$, so f^{-1} is continuous at c . \square

As a consequence of Theorem 11.6, we see that inverse trigonometric functions and logarithmic functions are continuous.

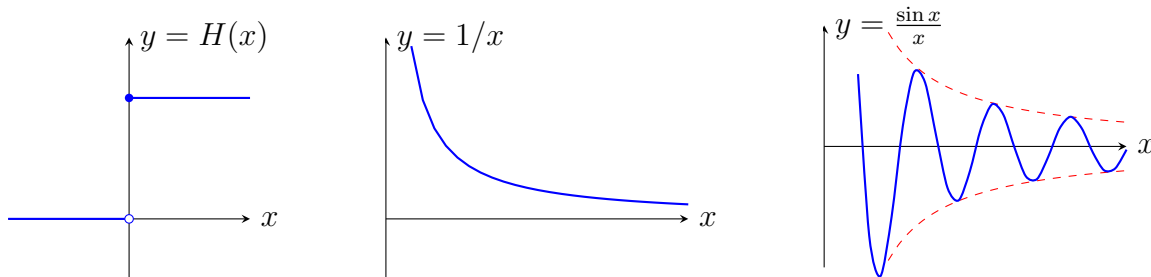
11.3. Limits at infinity

In addition to limits of a function f as x approaches a point $a \in \mathbb{R}$, we sometimes want to talk about “limits at infinity”, where x approaches ∞ or $-\infty$. Informally, we

say that $\lim_{x \rightarrow \infty} f(x) = L$ if $f(x)$ can be made arbitrarily close to L by taking x to be sufficiently large. Formally, we have the following definition, which should be compared to Definition 6.6 for the limit of a sequence.

Definition 11.7. We say that $\lim_{x \rightarrow \infty} f(x) = L$ if for every $\epsilon > 0$ there exists $R > 0$ such that for all $x > R$, we have $|f(x) - L| < \epsilon$.

There is an analogous definition for $\lim_{x \rightarrow -\infty}$, replacing $x > R$ with $x < -R$. If $\lim_{x \rightarrow \infty} f(x) = L$ or $\lim_{x \rightarrow -\infty} f(x) = L$, we say that the line $y = L$ is a *horizontal asymptote* of the graph of $y = f(x)$.



The three pictures above illustrate various ways that a function can approach a horizontal asymptote. In the first example, the Heaviside function from Example 9.2 has horizontal asymptotes at $y = 0$ and $y = 1$, and the graph eventually coincides with these asymptotes. In the second example, the graph of $y = 1/x$ has a horizontal asymptote at $y = 0$ because $\lim_{x \rightarrow \infty} 1/x = 0$, and the graph never touches the asymptote.

In the third example, the function $\frac{\sin x}{x}$ satisfies the inequalities $-\frac{1}{x} \leq \frac{\sin x}{x} \leq \frac{1}{x}$ for all $x > 0$, and since $\lim_{x \rightarrow \infty} -\frac{1}{x} = \lim_{x \rightarrow \infty} \frac{1}{x} = 0$, the Squeeze Theorem implies that $\lim_{x \rightarrow \infty} \frac{\sin x}{x} = 0$, so the graph has a horizontal asymptote at $y = 0$, which it crosses infinitely often.

Example 11.8. The inverse tangent function $y = \tan^{-1} x$ has two horizontal asymptotes, one at $y = -\frac{\pi}{2}$ and one at $y = \frac{\pi}{2}$.

The limit laws work for limits at infinity as well, although we need to be careful with Laws 8–10; these need to be rewritten as

$$(11.5) \quad \lim_{x \rightarrow \infty} x = \infty, \quad \lim_{x \rightarrow \infty} \sqrt[x]{x} = \infty, \quad \lim_{x \rightarrow \infty} x^n = \infty.$$

These use the following definition.

Definition 11.9. Given a function f that is defined for all sufficiently large x , we say that $\lim_{x \rightarrow \infty} f(x) = \infty$ if for every $Y > 0$ (no matter how large) there exists $X > 0$ such that every $x > X$ has $f(x) > Y$.

We define limits involving $-\infty$ in a similar way.

Exercise 11.10. Prove (11.5) using Definition 11.9.

Theorem 11.11. If $\lim_{x \rightarrow \infty} f(x) = \infty$, then $\lim_{x \rightarrow \infty} \frac{1}{f(x)} = 0$.

Proof. Given any $\epsilon > 0$, it follows from Definition 11.9 that there is $X > 0$ such that for all $x > X$, we have $f(x) > 1/\epsilon$. This implies that $0 < 1/f(x) < \epsilon$, and since $\epsilon > 0$ was arbitrary this shows that $\lim_{x \rightarrow \infty} \frac{1}{f(x)} = 0$. \square

Remark 11.12. It is tempting to think of Theorem 11.11 as just another version of Limit Law 5 by writing

$$\text{“} \lim_{x \rightarrow \infty} \frac{1}{f(x)} = \frac{1}{\lim_{x \rightarrow \infty} f(x)} = \frac{1}{\infty} = 0.\text{”}$$

However, this is not a correct application of Law 5, because “ $1/\infty$ ” is not a well-defined expression; ∞ is not a real number and cannot actually be divided, multiplied, added, subtracted, etc. While it is true that this informal computation gives the correct answer in this instance, it should be thought of as a way of remembering Theorem 11.11 rather than as a legitimate computation. It is worth keeping in mind the following examples where naive computations along these lines lead to trouble.

- (1) When $f(x) = \frac{\sin x}{x}$, we saw above that $\lim_{x \rightarrow \infty} f(x) = 0$, and it would be natural to write “ $\lim_{x \rightarrow \infty} \frac{1}{f(x)} = 1/\lim_{x \rightarrow \infty} f(x) = 1/0 = \infty$ ”; however, **this is incorrect**. A moment’s thought reveals that $\frac{1}{f(x)} = \frac{x}{\sin x}$ alternates between positive and negative values as x grows, changing sign at each vertical asymptote $x = n\pi$ ($n \in \mathbb{N}$), and so cannot go to ∞ .
- (2) From (11.5) we see that $\lim_{x \rightarrow \infty} x^2 = \infty$ and $\lim_{x \rightarrow \infty} x = \infty$. It is not hard to show that $\lim_{x \rightarrow \infty} (x^2 - x) = \infty$ and $\lim_{x \rightarrow \infty} (x - x^2) = -\infty$, so that any attempt to use some version of Limit Law 2 and make sense of “ $\infty - \infty$ ” is doomed to failure. (We will return to limits with such “indeterminate forms” later, when we study l’Hospitals’ rule.)

Exercise 11.13. Prove that the first example in the previous remark can be fixed by adding an extra assumption: if $\lim_{x \rightarrow \infty} f(x) = 0$ and if $f(x) > 0$ for all x , then $\lim_{x \rightarrow \infty} f(x) = \infty$.

The following exercise gives a version of Limit Laws 1, 3, and 4 for infinite limits.

Exercise 11.14. Prove that if $\lim_{x \rightarrow \infty} f(x) = \infty$ and $\lim_{x \rightarrow \infty} g(x) = \infty$, then

$$\lim_{x \rightarrow \infty} (f(x) + g(x)) = \infty \text{ and } \lim_{x \rightarrow \infty} (f(x)g(x)) = \infty.$$

Also prove that $\lim_{x \rightarrow \infty} cf(x) = \infty$ for all $c > 0$, and $\lim_{x \rightarrow \infty} cf(x) = -\infty$ for all $c < 0$.

One important consequence of Theorem 11.11 is the following.

Corollary 11.15. *For every $r > 0$, we have $\lim_{x \rightarrow \infty} \frac{1}{x^r} = 0$.*

Proof. By Theorem 11.11, it suffices to prove that $\lim_{x \rightarrow \infty} x^r = \infty$. Choosing $n \in \mathbb{N}$ sufficiently large that $\frac{1}{n} < r$, we have $x^r \geq x^{1/n} = \sqrt[n]{x}$, and $\lim_{x \rightarrow \infty} \sqrt[n]{x} = \infty$ by (11.5), so Theorem 8.6 (or rather, the equivalent theorem for limits at infinity) gives $\lim_{x \rightarrow \infty} x^r = \infty$. \square

We can use Corollary 11.15 to evaluate the limit at infinity of any rational function.

Example 11.16.

$$\lim_{x \rightarrow \infty} \frac{2x^2 + x - 5}{3x^2 - 2x + 1} = \lim_{x \rightarrow \infty} \frac{2 + \frac{1}{x} - \frac{5}{x^2}}{3 - \frac{2}{x} + \frac{1}{x^2}} \quad \text{dividing top and bottom by } x^2$$

$$\begin{aligned}
& 2 + \lim_{x \rightarrow \infty} \frac{1}{x} - 5 \lim_{x \rightarrow \infty} \frac{1}{x^2} \\
&= \frac{2 + \lim_{x \rightarrow \infty} \frac{1}{x} - 5 \lim_{x \rightarrow \infty} \frac{1}{x^2}}{3 - 2 \lim_{x \rightarrow \infty} \frac{1}{x} + \lim_{x \rightarrow \infty} \frac{1}{x^2}} \quad \text{by Limit Laws 5, 1, and 3} \\
&= \frac{2 + 0 - 5 \cdot 0}{3 - 2 \cdot 0 + 0} = \frac{2}{3} \quad \text{by Corollary 11.15.}
\end{aligned}$$

More complicated limits can sometimes be evaluated by using a little more algebraic manipulation. The following example requires our old trick of multiplying by the conjugate when we see an expression that involves adding or subtracting a square root.

Example 11.17.

$$\begin{aligned}
\lim_{x \rightarrow \infty} (\sqrt{x^2 + 1} - x) &= \lim_{x \rightarrow \infty} \frac{(\sqrt{x^2 + 1} - x)(\sqrt{x^2 + 1} + x)}{\sqrt{x^2 + 1} + x} = \lim_{x \rightarrow \infty} \frac{(x^2 + 1) - x^2}{\sqrt{x^2 + 1} + x} \\
&= \lim_{x \rightarrow \infty} \frac{1}{\sqrt{x^2 + 1} + x}.
\end{aligned}$$

From Exercise 11.14 we see that $\lim_{x \rightarrow \infty} (\sqrt{x^2 + 1} + x) = \infty$, and so Theorem 11.11 gives $\lim_{x \rightarrow \infty} (\sqrt{x^2 + 1} - x) = 0$.

Sometimes limits at finite values can be rephrased in terms of limits at infinity.

Example 11.18. Consider $\lim_{t \rightarrow 0^+} \left(\sqrt{\frac{1}{t} + 1} - \frac{1}{\sqrt{t}} \right)$. Writing $x = 1/\sqrt{t}$ so that $1/t = x^2$, we see that $x \rightarrow \infty$ as $t \rightarrow 0^+$, and thus

$$\lim_{t \rightarrow 0^+} \left(\sqrt{\frac{1}{t} + 1} - \frac{1}{\sqrt{t}} \right) = \lim_{x \rightarrow \infty} (\sqrt{x^2 + 1} - x) = 0.$$

Part II. Derivatives

Lecture 12

Derivatives

DATE: MONDAY, SEPTEMBER 16

This lecture corresponds to §2.7 in Stewart and Chapter 9 in Spivak.

Recall our discussion from Lecture 4 about finding the tangent line to the graph of a function $y = f(x)$ at the point $(a, f(a))$, for $x \approx a$ we considered the *secant line* through the points $(a, f(a))$ and $(x, f(x))$, which has slope $\frac{f(x)-f(a)}{x-a}$, and then took a limit as $x \rightarrow a$ to obtain the slope of the tangent line. Equivalently, we can write $x = a + h$ and obtain the slope of the tangent line as $\lim_{h \rightarrow 0} \frac{f(a+h)-f(a)}{h}$. This procedure is fundamental to the remainder of the course, and we formalize it with the following definition.

Definition 12.1. The *derivative* of a function f at a point a in the domain of f is

$$(12.1) \quad f'(a) := \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h},$$

if the limit exists. (The notation f' is pronounced “ f prime”.)

We call the function f *differentiable* at a if the limit in (12.1) exists, and *nondifferentiable* if it does not. We say that f is *differentiable on the interval* (a, b) if it is differentiable at every $c \in (a, b)$.

Remark 12.2. The two limits on the right-hand side of (12.1) are the same by Exercise 6.2, and either can be used as the definition of derivative.

Example 12.3. The function $f(x) = |x|$ is not differentiable at 0, because

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

so $\lim_{x \rightarrow 0^-} \frac{f(x)-f(0)}{x-0} = -1$ and $\lim_{x \rightarrow 0^+} \frac{f(x)-f(0)}{x-0} = 1$. Since the one-sided limits are different, the two-sided limit does not exist.

The derivative of a function has various interpretations in different applications.

- The tangent line to the curve $y = f(x)$ at the point $(a, f(a))$ has slope $f'(a)$. Thus the equation of this tangent line is $\frac{y-f(a)}{x-a} = f'(a)$. This can also be written as $y - f(a) = f'(a)(x - a)$, or $y = f(a) + f'(a)(x - a)$. In this case the *difference quotients* $\frac{f(x)-f(a)}{x-a}$ represent the slopes of the secant lines.
- A *linear function*¹⁶ is a function $g(x) = mx + b$, where m, b are real numbers (the slope and the y -intercept, respectively); equivalently, a linear function is a polynomial with degree 1. The function $g(x) = f(a) + f'(a)(x - a)$ is the linear function that best approximates the function f near $x = a$; one way to think of

¹⁶Technically we should call this an *affine function* and reserve the term *linear* for the case when $b = 0$ so that $g(x) = mx$, but here we will continue to use the terminology that you are probably more familiar with from previous math courses, which reflects the fact that the graph of $y = g(x)$ is a line.

this property of being the *best linear approximation* is that as we zoom in closer and closer to the point $(a, f(a))$, the graph of $y = f(x)$ looks more and more like the tangent line $y = f(a) + f'(a)(x - a)$.

- If $f(t)$ represents the position of an object at time t – for example, $f(t)$ may represent the height of a ball that is thrown straight up into the air, or the total distance from an observer to a car traveling along a straight road – then $f'(t)$ represents the *instantaneous velocity* of the object at time t . In this case the difference quotient $\frac{f(t+h)-f(t)}{h}$ represents the average velocity of the object between time t and time $t + h$.
- More generally, if $f(t)$ represents any quantity that changes with time, then $f'(t)$ represents the instantaneous rate of change at time t , and the difference quotient represents the average rate of change from t to $t + h$. For example, $f(t)$ might represent the number of bacteria in a petri dish at time t , and then $f'(t)$ represents the rate of growth of the population.¹⁷
- If $f(x)$ represents the cost of producing x units of something, then $\frac{f(x+h)-f(x)}{h}$ represents the extra cost incurred by producing h more units, assuming x units were already produced. The limit $f'(x)$ is called the *marginal cost* and represents the rate at which the cost changes per extra unit when x units are being produced.

There are many other applications and interpretations besides the ones listed above. We will return to applications later. For the time being we compute a few examples to get a feel for the process.

Example 12.4. If $f(x) = c$ is a constant function, then $f'(x) = \lim_{h \rightarrow 0} \frac{c-c}{h} = 0$ at every x . Thus *a constant function has vanishing derivative*. Later we will see that the converse is true as well provided f is differentiable everywhere.

Example 12.5. More generally, if $f(x) = mx + b$ for some constants $m, b \in \mathbb{R}$, then

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{m(x+h) + b - (mx + b)}{h} = \lim_{h \rightarrow 0} \frac{mh}{h} = m.$$

This is consistent with our interpretation of $f'(x)$ as the slope of the tangent line, since in this example the graph of f itself is a straight line with slope m .

Example 12.6. Consider the function $f(x) = x^2$. The derivative of f at a is

$$\begin{aligned} f'(a) &= \lim_{h \rightarrow 0} \frac{(a+h)^2 - a^2}{h} = \lim_{h \rightarrow 0} \frac{a^2 + 2ah + h^2 - a^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{2ah + h^2}{h} = \lim_{h \rightarrow 0} (2a + h) = 2a. \end{aligned}$$

Suppose we wish to find the tangent line to the parabola $y = x^2$ at the point $(3, 9)$. We have $f'(3) = 6$, so the tangent line has equation $y = 9 + 6(x - 3) = 9 + 6x - 18 = 6x - 9$.

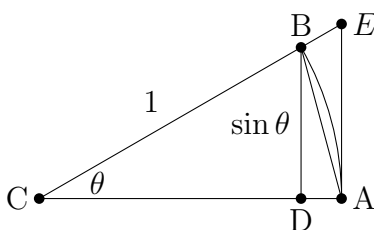
¹⁷Strictly speaking, the number of bacteria should be an integer, so $f(t)$ would need to be either constant or discontinuous, since otherwise the intermediate value theorem would imply that at some point in time it takes a non-integer value. But it is useful to pretend that the population can be an arbitrary real number, so that we can use the tools of calculus. If the population is very large, then this fiction is generally not too disruptive.

Example 12.7. The function $f(x) = \frac{1}{x}$ has derivative

$$f'(x) = \lim_{h \rightarrow 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \lim_{h \rightarrow 0} \frac{1}{h} \cdot \frac{h - (x+h)}{x(x+h)} = \lim_{h \rightarrow 0} \frac{-h}{hx(x+h)} = \lim_{h \rightarrow 0} \frac{-1}{x(x+h)} = -\frac{1}{x^2}.$$

The *normal line* to a curve $y = f(x)$ at the point $(a, f(a))$ is the line through $(a, f(a))$ that is perpendicular to the tangent line at that point. Since the tangent line has slope $f'(a)$, the normal line has slope $-\frac{1}{f'(a)}$. Suppose we wish to find the normal line to the curve $y = \frac{1}{x}$ at the point $(2, \frac{1}{2})$. The derivative is $f'(x) = -\frac{1}{x^2}$, so the normal line has slope 4, and the equation of the normal line is $y = \frac{1}{2} + 4(x - 2)$.

The following example was done in class on Wednesday, September 18



Example 12.8. The derivative of the function $f(x) = \sin x$ at the point $x = 0$, if it exists, is given by

$$f'(0) = \lim_{x \rightarrow 0} \frac{\sin x - \sin 0}{x - 0} = \lim_{x \rightarrow 0} \frac{\sin x}{x}.$$

To compute this limit, we use the figure shown above and observe that

$$BD = \sin \theta, \quad AE = \tan \theta, \quad CB = CA = 1.$$

The triangles CBA and CEA have areas $\frac{1}{2} \sin \theta$ and $\frac{1}{2} \tan \theta$, respectively, while the circular wedge with vertices C, B, A has area $\frac{1}{2} \theta$; this wedge contains the triangle CBA and is contained in the triangle CEA , so we have

$$\sin \theta < \theta < \tan \theta$$

for all $\theta \in (0, \frac{\pi}{2})$. The first inequality gives $\frac{\sin \theta}{\theta} < 1$, and the second gives $\cos \theta < \frac{\sin \theta}{\theta}$. Putting these together gives

$$\cos \theta < \frac{\sin \theta}{\theta} < 1.$$

The left-hand and right-hand functions converge to 1 as $\theta \rightarrow 0^+$, and thus by the Squeeze Theorem we have $\lim_{\theta \rightarrow 0^+} \frac{\sin \theta}{\theta} = 1$. The sine function is even, so for $x < 0$ we have $\frac{\sin x}{x} = \frac{\sin |x|}{|x|}$, and we conclude that

$$f'(0) = \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

Later we will use this to find the derivative of $f(x) = \sin x$ at any real number, not just 0.

Lecture 13**Derivative as a function**

DATE: WEDNESDAY, SEPTEMBER 18

This lecture corresponds to §2.8 in Stewart and Chapter 9 in Spivak.

13.1. Domain of the derivative

Suppose the function f is differentiable on the interval (a, b) ; that is, the derivative $f'(x) = \lim_{h \rightarrow 0} \frac{1}{h}(f(x+h) - f(x))$ exists for every $x \in (a, b)$. Then we can interpret $f': (a, b) \rightarrow \mathbb{R}$ as a function in its own right.

Remark 13.1. The domain of f' may be smaller than the domain of f . Indeed, consider the function $f(x) = \sqrt{x}$, whose domain is $[0, \infty)$. Then we have

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

and we see that the domain of f' is $(0, \infty)$.¹⁸

We identify some of the common ways that differentiability can fail.

- (1) If f has a discontinuity at a , then by Theorem 13.3 it is not differentiable there.
- (2) If the one-sided limits in the definition of derivative exist at a but are not equal, then f is not differentiable at a , and its graph has a ‘corner’ there; this is what occurs for $f(x) = |x|$. In particular, this shows that continuity does not imply differentiability.
- (3) If $\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \infty$ or $-\infty$, then the graph of f has “infinite slope” at a , and f is not differentiable there. We saw this occur with $f(x) = \sqrt{x}$.

This list is not comprehensive.

Exercise 13.2. Prove that the function $x \sin \frac{1}{x}$ from Example 9.19(2) is nondifferentiable at 0, but does not fit into any of the categories listed above.

It is worth noting that the domain of f' – that is, the set of points at which f is differentiable – must be contained in the set of points at which f is continuous.

Theorem 13.3. *If f is differentiable at a , then f is continuous at a .*

Proof. Since $f'(a)$ exists, we can use Limit Law 4 to conclude that

$$\begin{aligned} \lim_{x \rightarrow a} (f(x) - f(a)) &= \lim_{x \rightarrow a} \left(\frac{f(x) - f(a)}{x - a} (x - a) \right) \\ &= \left(\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \right) \left(\lim_{x \rightarrow a} (x - a) \right) = f'(a) \cdot 0 = 0. \end{aligned}$$

¹⁸In addition to the fact that the formula for f' only makes sense when $x > 0$, it is also standard practice to only consider the derivative as defined at points where the *two-sided* limit makes sense; when f is only defined on one side of a point a , we usually talk only about one-sided derivatives at a .

Then we have

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} (f(a) + f(x) - f(a)) = f(a) + \lim_{x \rightarrow a} (f(x) - f(a)) = f(a) + 0 = f(a),$$

which proves that f is continuous at a . \square

As the absolute value function shows, the converse of Theorem 13.3 is false. So we summarize the situation like this: **Every differentiable function is continuous, but not every continuous function is differentiable.**

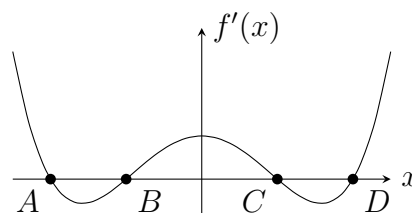
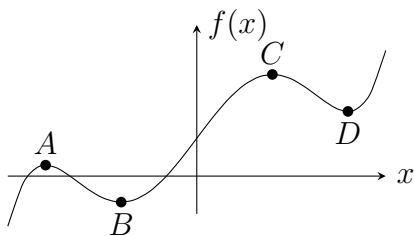
13.2. Connection between properties of f and f'

13.2.1. Monotonicity properties

It is useful to relate properties of the function f' to properties of the function f . Intuitively, since $f'(x)$ is the limit of the ratios $\frac{f(y)-f(x)}{y-x}$ as $y \rightarrow x$, we expect that a positive derivative, $f'(x) > 0$, corresponds to an increasing function, where $x < y$ implies $y - x > 0$ implies $f(y) - f(x) \approx f'(x)(y - x) > 0$, so $f(y) > f(x)$. We will make this precise when we study the Mean Value Theorem later on; for now, we merely observe that $f' > 0$ corresponds to regions where f is increasing, and $f' < 0$ corresponds to regions where f is decreasing.

Example 13.4. When $f(x) = x^2$, we saw in Example 12.6 that $f'(x) = 2x$. Thus $f' > 0$ on the interval $(0, \infty)$, where f is increasing, and $f' < 0$ on the interval $(-\infty, 0)$, where f is decreasing.

This qualitative relationship between f and f' is useful even when we do not have formulas for the functions. For example, in the pictures below we see that f' is positive between B and C , where the function f is increasing, and negative between A and B , and between C and D , where the function f is decreasing.



13.2.2. Symmetry properties

Proposition 13.5. If f is an even function ($f(-x) = f(x)$ for all x) then f' is odd ($f'(-x) = -f'(x)$ for all x where f' exists), and vice versa.

Proof. If f is even and is differentiable at x , then we have

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

where the first equality is the definition of derivative, and the second uses the fact that f is even. Writing $k = -h$ we can rewrite the last ratio as $\frac{f(x+k) - f(x)}{-k}$, and obtain

$$f'(-x) = \lim_{k \rightarrow 0} \frac{f(x+k) - f(x)}{-k} = - \lim_{k \rightarrow 0} \frac{f(x+k) - f(x)}{k} = -f'(x),$$

which proves that f' is odd. Conversely, if f is odd then we can make a similar computation and show that f' is even. \square

Example 13.6. The function $f(x) = x^2$ is even, while its derivative $f'(x) = 2x$ (recall Example 12.6) is odd.

13.3. Notation

We will usually write the derivative of f at x as $f'(x)$, but you should be aware of other notations that are in common use. One of these, which will appear in this class, is $\frac{d}{dx}f(x)$, or $\frac{df}{dx}$. If we want to stress that this derivative is evaluated at a particular point $x = a$, we may write $\frac{d}{dx}f(x)|_{x=a}$. It is important to note that although $\frac{df}{dx}$ looks like a fraction (and one might be tempted to do things like ‘cancel the ds ’), it is not a fraction, and the symbols df and dx have no independent meaning. This notation comes from the idea that $\frac{df}{dx}$ is the limit of the ratios $\frac{\Delta f}{\Delta x}$, where Δx represents the change in x , and $\Delta f = f(x + \Delta x) - f(x)$ represents the corresponding change in the value of f . These ratios are of course the difference quotients $\frac{f(x+h)-f(x)}{h}$.

Remark 13.7. Rather than thinking of $\frac{df}{dx}$ as representing a fraction, it is better to think about $\frac{d}{dx}$ as representing the *operation* of differentiation, so that $\frac{d}{dx}f$ is the function that results from applying the operation of differentiation to the function f . From this point of view, $\frac{d}{dx}$ is the *differentiation operator*, and is a function whose inputs and outputs are themselves functions. We will not dwell on this point of view in this course, but it is an important one in more advanced courses.

The notation \dot{f} is sometimes used for the derivative of f ; this is most common when f is a function of time t , so that \dot{f} represents a derivative with respect to time. In this course we will generally stick to the notation f' , though. When the variable is t instead of x , we will also use the notation $\frac{df}{dt}$ to make it clear what we are differentiating with respect to, and similarly if some other variable is used instead of x or t .

You may sometimes see a *dependent* variable y used to represent a function of an *independent* variable x , and then the notation $\frac{dy}{dx}$, y' , or \dot{y} is sometimes used for the derivative.

Finally, you may even see the notation Df for the derivative. This is a convenient notation if we want to talk about the *one-sided* derivatives $D^+f(x) = \lim_{h \rightarrow 0^+} \frac{f(x+h)-f(x)}{h}$ and $D^-f(x) = \lim_{h \rightarrow 0^-} \frac{f(x+h)-f(x)}{h}$. By Theorem 8.3, we see that f is differentiable at x if and only if $D^+f(x)$ and $D^-f(x)$ both exist and take the same value. Example 12.3 shows that the absolute value function $f(x) = |x|$ has $D^-f(0) = -1$ and $D^+f(0) = 1$.

The following section was done in the lecture on Friday, September 20 Mon, Sep 23

13.4. Higher derivatives

If f is differentiable on (a, b) , then f' is a function on (a, b) in its own right, and it is reasonable to ask whether this function is itself differentiable. If it is, then we refer to its derivative

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

as the *second derivative* of f , and denote it by $f''(x)$. This last notation is often pronounced “ f double prime”.

Another common notation for the second derivative of f with respect to x is

$$f''(x) = \frac{d^2}{dx^2} f(x) = \frac{d^2 f}{dx^2}(x).$$

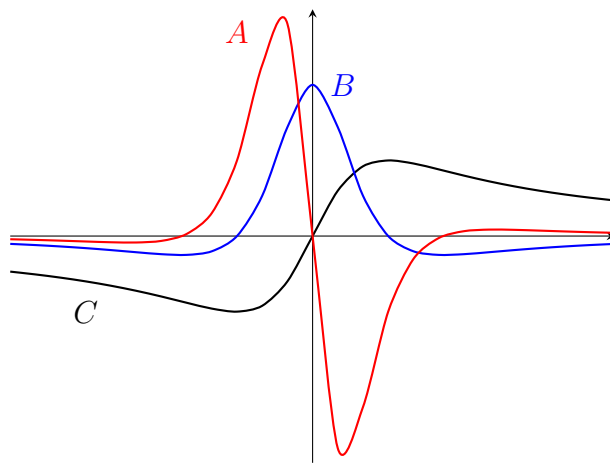
It should be noted that the appearance of a superscript “2” here is *not* used to indicate that any quantity is squared, or multiplied by itself; rather, it indicates that the operation of differentiation is done twice. One might also see the notation \ddot{f} , especially when f is a function of time, although we will generally stick to the notation f'' .

Example 13.8. If $f(t)$ represents the height of an object at time t , then $f'(t)$ represents its velocity at time t , and $f''(t)$ represents the rate at which its velocity is changing; in other words, its *acceleration*.

One can go further and define the third derivative $f'''(x)$ to be the derivative of $f''(x)$, and so on. In general we use the notation $f^{(n)}(x)$ or $\frac{d^n}{dx^n} f(x)$ or $\frac{d^n f}{dx^n}$ for the n th derivative of f at x , when it exists; this can be defined iteratively by

$$f^{(n)}(x) = \frac{d^n}{dx^n} f(x) := \frac{d}{dx} f^{(n-1)}(x).$$

We will be most interested in first- and second-order derivatives f' and f'' , but higher-order derivatives will play a role next semester when we discuss *Taylor polynomials and series*.



Example 13.9. The picture above shows the graphs of f , f' , and f'' for some function f . Which curve represents which function?

To answer this question, look at the intervals on which the curves labeled A , B , C are positive, and see if any of the other curves are increasing on these intervals. We see that the interval on which B is positive is the same as the interval on which C is increasing, so B should be the graph of the derivative of the function whose graph is C . Similarly, A is positive on the intervals where B is increasing, and negative on the intervals where B is decreasing. So we conclude that C is the graph of f , B is the graph of f' , and A is the graph of f'' .

Something to think about: is there a connection between the sign of f'' (as shown in the curve A) and the shape of the graph of f (as shown in the curve C)? We will return to this later when we discuss *convexity*.

Lecture 14 Derivatives of polynomials and exponentials

DATE: ~~FRIDAY, SEPTEMBER 20~~ MONDAY, SEPTEMBER 23

The date of this lecture was moved due to the university closure on Friday, Sept. 20.

This lecture corresponds to §3.1 in Stewart and Chapter 10 in Spivak

14.1. Power rule and polynomial functions

Now we start developing systematic rules to evaluate derivatives of broad classes of functions. When we did a similar procedure for limits, we started with polynomial functions, and we will do the same thing here.

Start by recalling that Exercises 12.4, 12.5, and 12.6 gave us

$$\frac{d}{dx}c = 0, \quad \frac{d}{dx}x = 1, \quad \frac{d}{dx}x^2 = 2x.$$

What about the derivative of x^n for other values of n ? Start with $n = 3$:

$$\begin{aligned} \frac{d}{dx}x^3 &= \lim_{h \rightarrow 0} \frac{(x+h)^3 - x^3}{h} = \lim_{h \rightarrow 0} \frac{x^3 + 3x^2h + 3xh^2 + h^3 - x^3}{h} \\ &= \lim_{h \rightarrow 0} \frac{3x^2h + 3xh^2 + h^3}{h} = \lim_{h \rightarrow 0} (3x^2 + 3xh + h^2) = 3x^2. \end{aligned}$$

A pattern is beginning to emerge, which is confirmed by the following theorem.

Theorem 14.1 (Power rule). *If $f(x) = x^n$ for some $n \in \mathbb{N}$, then $f'(x) = nx^{n-1}$.*

First proof of the power rule. Start by recalling the *Binomial Theorem*, which says that given $x, h \in \mathbb{R}$ and $n \in \mathbb{N}$, we have

$$(x+h)^n = x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 + \frac{n(n-1)(n-2)}{3 \cdot 2}x^{n-3}h^3 + \cdots + nh^{n-1} + h^n,$$

where the coefficient on the term $x^{n-k}h^k$ is given by

$$\binom{n}{k} := \frac{n(n-1)(n-2) \cdots (n-k+1)}{k(k-1) \cdots 2} = \frac{n!}{k!(n-k)!}$$

We can rewrite the expansion in the Binomial Theorem as

$$(14.1) \quad (x+h)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} h^k = x^n + nx^{n-1}h + \sum_{k=2}^n \binom{n}{k} x^{n-k} h^k,$$

where in the last expression we have separated out the first two terms for reasons that will become clear momentarily.

Exercise 14.2. Prove the Binomial Theorem (14.1) using induction on n .

Using (14.1) we can compute $f'(x)$ for $f(x) = x^n$ as follows:

$$\begin{aligned} f'(x) &= \lim_{h \rightarrow 0} \frac{1}{h} ((x+h)^n - x^n) = \lim_{h \rightarrow 0} \frac{1}{h} \left(x^n + nx^{n-1}h + \left(\sum_{k=2}^n \binom{n}{k} x^{n-k} h^k \right) - x^n \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(nx^{n-1}h + \left(\sum_{k=2}^n \binom{n}{k} x^{n-k} h^k \right) \right) = nx^{n-1} + \sum_{k=2}^n \lim_{h \rightarrow 0} \binom{n}{k} x^{n-k} h^{k-1}, \end{aligned}$$

where the last equality uses Limit Law 1 for addition. For every $k = 2, 3, \dots, n$, we have $\lim_{h \rightarrow 0} \binom{n}{k} x^{n-k} h^{k-1} = 0$ since $k-1 \geq 1$, and thus we conclude that $f'(x) = nx^{n-1}$. \square

An alternate proof of the power rule uses the following exercise, which generalizes the formula $x^2 - y^2 = (x-y)(x+y)$ for a difference of squares.

Exercise 14.3. Prove that for every $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$, we have

$$(14.2) \quad y^n - x^n = (y-x)(y^{n-1} + y^{n-2}x + y^{n-3}x^2 + \dots + yx^{n-2} + x^{n-1}) = (y-x) \sum_{k=1}^n y^{n-k} x^{k-1}.$$

Second proof of the power rule. Using (14.2), we have

$$\begin{aligned} \frac{d}{dx} x^n &= \lim_{y \rightarrow x} \frac{y^n - x^n}{y - x} = \lim_{y \rightarrow x} \frac{(y-x)(y^{n-1} + y^{n-2}x + y^{n-3}x^2 + \dots + yx^{n-2} + x^{n-1})}{y-x} \\ &= \lim_{y \rightarrow x} \sum_{k=1}^n y^{n-k} x^{k-1} = \sum_{k=1}^n \lim_{y \rightarrow x} y^{n-k} x^{k-1} = \sum_{k=1}^n x^{n-1} = nx^{n-1}. \quad \square \end{aligned}$$

By a short application of the limit laws, we can find the derivatives of cf and $f \pm g$ if f', g' are known and $c \in \mathbb{R}$.

Theorem 14.4. If $c \in \mathbb{R}$ and f is differentiable at a , then the function cf is differentiable at a and $(cf)'(a) = c \cdot f'(a)$.

Proof. Writing $g(x) = cf(x)$, Limit Law 3 gives

$$g'(x) = \lim_{h \rightarrow 0} \frac{cf(x+h) - cf(x)}{h} = c \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = cf'(x). \quad \square$$

Theorem 14.5. If f, g are differentiable at a , then so are the functions $f \pm g$, and $(f \pm g)'(a) = f'(a) \pm g'(a)$.

Proof. Writing $F(x) = f(x) + g(x)$, Limit Law 4 gives

$$\begin{aligned} F'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) + g(x+h) - f(x) - g(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} + \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} = f'(x) + g'(x). \quad \square \end{aligned}$$

Now we can differentiate any polynomial function.

Example 14.6. If $f(x) = 2 + 3x^2 + 7x^5$, then

$$f'(x) = \frac{d}{dx} 2 + \frac{d}{dx} (3x^2) + \frac{d}{dx} (7x^5) = 0 + 3 \frac{d}{dx} (x^2) + 7 \frac{d}{dx} (x^5) = 6x + 35x^4,$$

where the first equality uses Theorem 14.4, the second uses Theorem 14.5, and the third uses the power rule for differentiation.

In fact the power rule holds more generally. We saw in Example 12.7 that $\frac{d}{dx} \frac{1}{x} = -\frac{1}{x^2}$, and in Remark 13.1 that $\frac{d}{dx} \sqrt{x} = \frac{1}{2\sqrt{x}}$; these can be written as

$$\frac{d}{dx} x^{-1} = -x^{-2} \quad \text{and} \quad \frac{d}{dx} x^{\frac{1}{2}} = -\frac{1}{2} x^{-\frac{1}{2}},$$

which both have the same form as the power rule.

Theorem 14.7. *For every $\beta \in \mathbb{R}$, the function $f(x) = x^\beta$ has $f'(x) = \beta x^{\beta-1}$.*

Proof. Deferred; first we need to study derivatives of exponential functions, logarithmic functions, and the chain rule. \square

Using this, we can also differentiate functions that are not polynomials but can be written as sums of power functions.

Example 14.8. If $g(t) = \sqrt{t}(t-1)$, then we can write

$$g'(t) = \frac{d}{dt} (t^{3/2} - t^{1/2}) = \frac{d}{dt} t^{3/2} - \frac{d}{dt} t^{1/2} = \frac{3}{2} t^{1/2} - \frac{1}{2} t^{-1/2}.$$

The following section appeared in the video lecture for Friday, September 27

14.2. Exponential functions

Fix $a > 0$ and let $f(x) = a^x$. To find the derivative of f at x (if it exists) we write

$$(14.3) \quad f'(x) = \lim_{h \rightarrow 0} \frac{a^{x+h} - a^x}{h} = \lim_{h \rightarrow 0} a^x \cdot \frac{a^h - 1}{h} = a^x \lim_{h \rightarrow 0} \frac{a^h - 1}{h}.$$

But why should this last limit exist? In fact we will prove the following in Lecture 17:

For every $a > 0$, the limit $\lim_{h \rightarrow 0} \frac{a^h - 1}{h}$ exists. Writing $g(a)$ for the value of this limit, the function $g: (0, \infty) \rightarrow \mathbb{R}$ has the following properties:

- (1) g is (strictly) increasing and continuous;
- (2) $\lim_{a \rightarrow 0^+} g(a) = -\infty$, $g(1) = 0$, and $\lim_{a \rightarrow \infty} g(a) = +\infty$;
- (3) $g(ab) = g(a) + g(b)$ for all $a, b > 0$.

For the moment we defer the proof of this, and study its consequences. It follows from 1 and 2 that g is a bijection from $(0, \infty)$ to \mathbb{R} ; indeed, it is 1-1 since it is strictly increasing, and onto since for every $x \in \mathbb{R}$ we can use 2 to find $a, b \in (0, \infty)$ with $g(a) < x < g(b)$, so that the Intermediate Value Theorem guarantees existence of some $c \in (a, b)$ such that $g(c) = x$.

The function $g^{-1}: \mathbb{R} \rightarrow (0, \infty)$ is continuous by Theorem 11.6. Moreover, given $x, y \in \mathbb{R}$, writing $a = g^{-1}(x)$ and $b = g^{-1}(y)$, 3 gives

$$g(ab) = g(a) + g(b) = x + y, \quad \text{therefore} \quad g^{-1}(x + y) = ab = g^{-1}(x)g^{-1}(y).$$

Recall that we defined the exponential function $f(x) = a^x$ by the condition that $f: \mathbb{R} \rightarrow (0, \infty)$ is continuous, satisfies $f(x+y) = f(x)f(y)$ for all $x, y \in \mathbb{R}$, and has $f(1) = a$. Since f is the only function with these properties, we conclude that $g^{-1} = f$, where $a = g^{-1}(1)$. This leads us to the following definition.

Definition 14.9. The function g defined in Theorem 17.1 is called the *natural logarithm function*, and denoted $\ln a := g(a)$. Since $\ln: (0, \infty) \rightarrow \mathbb{R}$ is a bijection, we can define the number e as the unique positive real number such that $\ln e = \lim_{h \rightarrow 0} \frac{e^h - 1}{h} = 1$.

Now we can rewrite the computation in (14.3) as

$$(14.4) \quad \frac{d}{dx}a^x = (\ln a)a^x \quad \text{for all } a > 0.$$

In particular, observe that $f(x) = e^x$ has $f'(x) = \ln(e)e^x = e^x$.

Remark 14.10. The identity $\frac{d}{dx}e^x = e^x$ says that the exponential function with base e is its own derivative. By Theorem 14.4, the same thing is true of $f(x) = ce^x$ for any $c \in \mathbb{R}$. It turns out that these are the *only* functions with this property, as we will see in Lecture 22.1.

Let us observe one further consequence of (14.4). For $c \in \mathbb{R}$ the function $g(x) = e^{cx} = (e^c)^x$ has $g'(x) = (\ln e^c)e^{cx} = ce^{cx}$. We can write this as

$$(14.5) \quad \frac{d}{dx}f(cx) = cf'(cx).$$

In fact, (14.5) is always true, not just for the exponential function; indeed, if f is differentiable at cx , then we have

$$\begin{aligned} \frac{d}{dx}f(cx) &= \lim_{h \rightarrow 0} \frac{f(c(x+h)) - f(cx)}{h} = \lim_{h \rightarrow 0} \frac{f(cx+ch) - f(cx)}{h} \\ &= c \cdot \lim_{ch \rightarrow 0} \frac{f(cx+ch) - f(cx)}{ch} = cf'(cx). \end{aligned}$$

This is the first instance of a more general result, the *chain rule*, that we will come to soon. It is worth noting that while $\frac{d}{dx}f(x)$ and $f'(x)$ mean the same thing, (14.5) shows that $\frac{d}{dx}f(cx)$ and $f'(cx)$ mean *different* things. To make sense of this, observe that $x \mapsto f(cx)$ is the composition of two distinct functions; first we multiply by c , then we apply f . This can be represented by the following diagram:

$$x \xrightarrow{\cdot c} cx \xrightarrow{f} f(cx).$$

In $\frac{d}{dx}f(cx)$, we are measuring the response of the final output to a variation in the **initial input** x . In $f'(cx)$, on the other hand, we are measuring the response of the final output to a variation in the **input of the function** f , which is cx . The two do not give the same result in general, and the relationship between them is given by (14.5).

On Wednesday, September 25 we did a review session for Test 1.

Lecture 15

Product and quotient rules

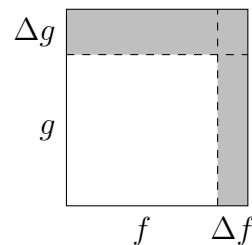
DATE: FRIDAY, SEPTEMBER 27 (VIDEO)

This lecture corresponds to §3.2 in Stewart and Chapter 10 in Spivak.

Suppose f, g are differentiable at x . What can we say about $(fg)'(x)$? Does it exist? What is its value?

The first thing to observe is that $(fg)'(x)$ is **not** given by the naive formula $f(x)g'(x)$; indeed, if $f(x) = g(x) = x$, then $f'(x) = g'(x) = 1$ and $(fg)(x) = x^2$, so $(fg)'(x) = 2x \neq 1 = f'(x)g'(x)$.

To find the correct formula, we start by imagining a rectangle whose width and height vary with time. Let $f(t)$ be the width and $g(t)$ the height at time t . Then the area at time t is $A(t) = f(t)g(t)$. If we go from time t to time $t + \Delta t$, then the rectangle at the new time has width $f + \Delta f$ and height $g + \Delta g$, as shown in the picture, and the change in the area is



$$\Delta(fg) = \Delta A = (f + \Delta f)(g + \Delta g) - fg = (\Delta f)g + f(\Delta g) + \Delta f\Delta g.$$

Observe that the three terms in this sum correspond to the three shaded rectangles in the picture. We see that

$$\frac{\Delta(fg)}{\Delta t} = \frac{\Delta f}{\Delta t}g(t) + f(t)\frac{\Delta g}{\Delta t} + \frac{\Delta f\Delta g}{\Delta t} \approx f'(t)g(t) + f(t)g'(t) + f'(t)\Delta g$$

taking Δt tending to 0 suggests that $(fg)'(t) = f'(t)g(t) + f(t)g'(t)$. Now we make this formal.

Theorem 15.1 (Product Rule). *If f and g are differentiable at x , then fg is also differentiable at x , and $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$.*

Proof. The derivative of fg at x , if it exists, is given by the limit

$$\begin{aligned} (fg)'(x) &= \lim_{y \rightarrow x} \frac{f(y)g(y) - f(x)g(x)}{y - x} \\ &= \lim_{y \rightarrow x} \frac{f(y)g(y) - f(x)g(y) + f(x)g(y) - f(x)g(x)}{y - x} \\ &= \lim_{y \rightarrow x} \underbrace{\frac{f(y) - f(x)}{y - x}}_{\text{I}} \underbrace{\lim_{y \rightarrow x} g(y)}_{\text{II}} + f(x) \underbrace{\lim_{y \rightarrow x} \frac{g(y) - g(x)}{y - x}}_{\text{III}}, \end{aligned}$$

where the first line is the definition of derivative, the second line comes by adding and subtracting $f(x)g(y)$, and the third line uses the limit laws for addition and multiplication. In order for this to be valid, we need to verify that the limits in the last line exist. Limits I and III exist and are equal to $f'(x)$ and $g'(x)$, respectively, because we assumed that f and g are differentiable at x . Because g is differentiable at x , it is also continuous at x by Theorem 13.3; thus limit II exists and is equal to $g(x)$. This completes the proof of the theorem. \square

Example 15.2. In Example 14.8 we evaluated the derivative of $f(x) = \sqrt{x}(x - 1)$ by expanding it as $x^{3/2} - x^{1/2}$ and using the power rule. We can also evaluate it without expanding by using the product rule (and then the power rule):

$$\begin{aligned} f'(x) &= \left(\frac{d}{dx} \sqrt{x} \right) (x - 1) + \sqrt{x} \frac{d}{dx} (x - 1) = \frac{1}{2\sqrt{x}} (x - 1) + \sqrt{x} \\ &= \frac{\sqrt{x}}{2} - \frac{1}{2\sqrt{x}} + \sqrt{x} = \frac{3}{2}\sqrt{x} - \frac{1}{2\sqrt{x}}, \end{aligned}$$

which agrees with our earlier answer.

The product rule lets us give a third proof of the power rule $\frac{d}{dx}x^n = nx^{n-1}$, by induction this time: the power rule for $n = 1$ is just the observation that $\frac{d}{dx}x = 1$, and if the power rule is true for a given value of n , then the product rule gives

$$\frac{d}{dx}x^{n+1} = \frac{d}{dx}(x^n x) = \left(\frac{d}{dx}x^n\right)x + x^n \frac{d}{dx}x = nx^{n-1} \cdot x + x^n \cdot 1 = (n+1)x^n,$$

so the power rule is true for $n+1$. By induction, the power rule holds for all $n \in \mathbb{N}$.

To compute the derivative of a quotient function $f(x)/g(x)$, we start by considering the case when $f(x) = 1$. Suppose that g is differentiable at x and that $g(x) \neq 0$; suppose moreover that $h(x) := 1/g(x)$ is differentiable at x ; then we have $(gh)(x) = g(x)h(x) = 1$, and differentiating both sides gives

$$0 = \frac{d}{dx}1 = \frac{d}{dx}(gh)(x) = g'(x)h(x) + g(x)h'(x).$$

Solving for h' gives

$$\frac{d}{dx} \frac{1}{g(x)} = h'(x) = -\frac{g'(x)h(x)}{g(x)} = -\frac{g'(x)}{g(x)^2},$$

where the last equality uses the definition of $h(x)$. This suggests what the formula for the derivative of a reciprocal should be. However, it does **not** prove quite the result that we want; recall Theorem 15.1, where we only needed to assume that f and g were differentiable at x , and then were able to conclude that f/g was also differentiable at x . Here we are forced to assume differentiability of $1/g$, when really this ought to be one of the conclusions of the theorem. In order to do this we need to go back to the definition of derivative rather than relying on the product rule.

Theorem 15.3 (Reciprocal rule). *If g is differentiable at x and $g(x) \neq 0$, then $1/g$ is also differentiable at x and $(1/g)'(x) = -g'(x)/g(x)^2$.*

Proof. Use the definition of derivative together with the limit laws:

$$\begin{aligned} \left(\frac{1}{g}\right)'(x) &= \lim_{y \rightarrow x} \frac{\frac{1}{g(y)} - \frac{1}{g(x)}}{y - x} = \lim_{y \rightarrow x} \frac{1}{y - x} \frac{g(x) - g(y)}{g(y)g(x)} \\ &= -\frac{1}{g(x)} \underbrace{\lim_{y \rightarrow x} \frac{1}{g(y)}}_{\text{I}} \underbrace{\lim_{y \rightarrow x} \frac{g(y) - g(x)}{y - x}}_{\text{II}}, \end{aligned}$$

provided limits I and II exist. Since g is differentiable at x , II exists and is equal to $g'(x)$; moreover, Theorem 13.3 implies that g is continuous at x , so I exists and is equal to $1/g(x)$, which proves the theorem. \square

The reciprocal rule lets us prove the power rule for negative integers:

$$\frac{d}{dx}(x^{-n}) = \frac{d}{dx} \frac{1}{x^n} = -\frac{\frac{d}{dx}x^n}{(x^n)^2} = -\frac{nx^{n-1}}{x^{2n}} = -nx^{-n-1}.$$

However, we are still awaiting a proof for the case when n is not an integer.

Putting the product rule and reciprocal rule together gives the quotient rule.

Theorem 15.4 (Quotient rule). *If f, g are differentiable at x and $g(x) \neq 0$, then f/g is also differentiable at x and*

$$(15.1) \quad \left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}.$$

Proof. The reciprocal rule implies that $1/g$ is differentiable; then the product rule implies that $f/g = f \cdot \frac{1}{g}$ is differentiable. Its derivative can be computed by combining the two formulas given above:

$$\begin{aligned} \left(\frac{f}{g}\right)'(x) &= \left(f \cdot \frac{1}{g}\right)'(x) = f'(x) \cdot \frac{1}{g(x)} + f(x) \left(\frac{1}{g}\right)'(x) \\ &= \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g(x)^2} = \frac{g(x)f'(x)}{g(x)^2} - \frac{f(x)g'(x)}{g(x)^2}. \quad \square \end{aligned}$$

Formula (15.1) is mildly clunky but is very important and should be memorized. I find the incantation “bottom times derivative of top, minus top times derivative of bottom, over bottom squared” helpful. It may also be helpful to observe the similarity between the numerator in (15.1) and the product rule; the only difference is the negative sign, and to remember where this goes you can remind yourself of the principle that differentiating something in the denominator tends to produce a negative sign, as in the reciprocal rule or as in $\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}$.

Example 15.5. Suppose that x and y are related by the formula $y = \frac{x^2+x+1}{x^2-1}$. To find $\frac{dy}{dx}$, we can use the quotient rule and get

$$\begin{aligned} \frac{dy}{dx} &= \frac{(x^2-1)\frac{d}{dx}(x^2+x+1) - (x^2+x+1)\frac{d}{dx}(x^2-1)}{(x^2-1)^2} \\ &= \frac{(x^2-1)(2x+1) - (x^2+x+1)(2x)}{(x^2-1)^2} = \frac{2x^3+x^2-2x-1 - (2x^3+2x^2+2x)}{(x^2-1)^2} \\ &= \frac{-x^2-4x-1}{(x^2-1)^2}. \end{aligned}$$

Example 15.6. The quotient rule leads to complicated enough computations that it is often better to use a different rule if we have the option. For example, we could differentiate the function $f(x) = (4x + 3\sqrt[3]{x})/\sqrt{x}$ using the quotient rule, but it is easier to write

$$f(x) = 4\sqrt{x} + 3x^{\frac{1}{3}-\frac{1}{2}} = 4x^{\frac{1}{2}} + 3x^{-\frac{1}{6}}$$

and then apply the power rule to each term directly.

Lecture 16

Trigonometric functions

DATE: MONDAY, SEPTEMBER 30

This lecture corresponds to §3.3 in Stewart and Chapter 15 in Spivak, although Spivak’s approach to the trigonometric functions is a little different and involves integration, which we did not discuss yet.

In Example 12.8 we proved that the sine function is differentiable at 0, with derivative 1; in other words,

$$(16.1) \quad \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

We can use this, together with some trigonometric identities, to find the derivative of \sin everywhere, similarly to the way that we used the property $e^{x+y} = e^x e^y$ of the exponential function to differentiate e^x everywhere once we understood its derivative at 0. For \sin , given any x, h , we have

$$\sin(x+h) = \sin x \cos h + \cos x \sin h,$$

and thus

$$\begin{aligned} \frac{d}{dx} \sin x &= \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin(x)}{h} = \lim_{h \rightarrow 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\ &= \lim_{h \rightarrow 0} \sin x \frac{\cos h - 1}{h} + \cos x \frac{\sin h}{h} = \sin x \underbrace{\lim_{h \rightarrow 0} \frac{\cos h - 1}{h}}_I + \cos x \underbrace{\lim_{h \rightarrow 0} \frac{\sin h}{h}}_{II}. \end{aligned}$$

The limit in II exists and is equal to 1 by (16.1). For the limit in I, we observe that

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\cos h - 1}{h} &= \lim_{h \rightarrow 0} \frac{\cos h - 1}{h} \cdot \frac{\cos h + 1}{\cos h + 1} = \lim_{h \rightarrow 0} \frac{\cos^2 h - 1}{h(\cos h + 1)} \\ &= \lim_{h \rightarrow 0} \frac{-\sin^2 h}{h(\cos h + 1)} = \left(\lim_{h \rightarrow 0} \frac{\sin h}{h} \right) \left(\lim_{h \rightarrow 0} \frac{-\sin h}{\cos h + 1} \right) = 1 \cdot \frac{-0}{2} = 0, \end{aligned}$$

and we conclude that

$$(16.2) \quad \frac{d}{dx} \sin x = \cos x.$$

Exercise 16.1. Use the formula $\cos(x+h) = \cos x \cos h - \sin x \sin h$ to prove that $\frac{d}{dx} \cos x = -\sin x$.

Remark 16.2. Euler's formula states that $e^{ix} = \cos x + i \sin x$. Differentiating this using what we just proved gives

$$(16.3) \quad \frac{d}{dx} e^{ix} = \frac{d}{dx} \cos x + i \frac{d}{dx} \sin x = -\sin x + i \cos x.$$

Recall that when we studied derivatives of exponential functions, we proved that $\frac{d}{dx} e^{cx} = ce^{cx}$. If this holds true for complex values of c as well, then we could also differentiate e^{ix} as

$$(16.4) \quad \frac{d}{dx} e^{ix} = ie^{ix} = i(\cos x + i \sin x) = i \cos x - \sin x.$$

Observe that (16.3) and (16.4) agree. However, one should observe that there's something a little bit suspect going on here; we never defined the value of the exponential function for arguments that are not real numbers, so we never actually defined e^{ix} , let alone prove that Euler's formula holds for it. In fact, one option is to use Euler's formula as the definition of e^{ix} ; then the above computations verify that differentiation still behaves as we expect, and the formula $e^{i(x+y)} = e^{ix} e^{iy}$ continues to be true as a consequence of the formulas for $\cos(x+y)$ and $\sin(x+y)$, though we omit the computation.

Now the derivatives of the other trigonometric functions can be computed by using the product and quotient rules, for example

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

Exercise 16.3. Prove that

$$\frac{d}{dx} \cot x = -\csc^2 x, \quad \frac{d}{dx} \sec x = \sec x \tan x, \quad \frac{d}{dx} \csc x = -\csc x \cot x.$$

Example 16.4. The derivative of $f(x) = \frac{\cos x}{1+\cot x}$ can be found using the quotient rule and the formulas $\frac{d}{dx} \cos x = -\sin x$, $\frac{d}{dx} \cot x = -\csc^2 x$:

$$\begin{aligned} f'(x) &= \frac{(1 + \cot x)(-\sin x) - \cos x(-\csc^2 x)}{(1 + \cot x)^2} = \frac{-\sin x - \frac{\cos x}{\sin x} \sin x + \cos x \frac{1}{\sin^2 x}}{(1 + \cot x)^2} \\ &= \frac{-\sin x + \frac{\cos x}{\sin^2 x}(1 - \sin^2 x)}{(1 + \frac{\cos x}{\sin x})^2} = \frac{-\sin x + \frac{\cos^3 x}{\sin^2 x}}{(\sin x + \cos x)^2 / \sin^2 x} = \frac{\cos^3 x - \sin^3 x}{(\cos x + \sin x)^2}. \end{aligned}$$

Remark 16.5. Since there are many trigonometric identities that allow us to convert write formulas of one trig function in terms of other trig functions, any example such as the previous one will often have multiple ways of writing the final answer. As a general rule, you should simplify as much as possible but be aware that the expression you write down might look different from what somebody else gets, even if both are correct. For example, in the previous computation we could just as easily have written the final answer as $\frac{-\sin x + \cos x \cot^2 x}{(1 + \cot x)^2}$.

Higher derivatives of sine and cosine can be easily computed since

$$\begin{aligned} \frac{d}{dx} \sin x = \cos x &\Rightarrow \frac{d^2}{dx^2} \sin x = \frac{d}{dx} \cos x = -\sin x \\ &\Rightarrow \frac{d^3}{dx^3} \sin x = \frac{d}{dx} (-\sin x) = -\frac{d}{dx} \sin x = -\cos x \\ &\Rightarrow \frac{d^4}{dx^4} \sin x = \frac{d}{dx} (-\cos x) = -\frac{d}{dx} \cos x = \sin x, \end{aligned}$$

and after this the pattern repeats with every fourth derivative. Keeping Remark 16.2 in mind, it is worth comparing this to the observation that

$$\begin{aligned} \frac{d}{dx} e^{ix} = ie^{ix} &\Rightarrow \frac{d^2}{dx^2} e^{ix} = \frac{d}{dx} ie^{ix} = i^2 e^{ix} = -e^{ix} \\ &\Rightarrow \frac{d^3}{dx^3} e^{ix} = \frac{d}{dx} (-e^{ix}) = -ie^{ix} \\ &\Rightarrow \frac{d^4}{dx^4} e^{ix} = \frac{d}{dx} (-ie^{ix}) = (-i)ie^{ix} = e^{ix}, \end{aligned}$$

which admits a geometric interpretation: multiplying a complex number by i corresponds to rotating it by 90 degrees ($\pi/2$ radians) around the origin; doing this twice corresponds to a 180 degree rotation, which is the same as multiplication by -1 , and doing it four

times returns us to where we started. For a similar geometric interpretation of the derivatives of sine, we can invoke some trigonometric identities and observe that

$$\sin\left(x + \frac{\pi}{2}\right) = \cos\frac{\pi}{2}\sin x + \sin\frac{\pi}{2}\cos x = 0 \cdot \sin x + 1 \cdot \cos x = \cos x = \frac{d}{dx}\sin x$$

$$\sin(x + \pi) = \cos\pi\sin x + \sin\pi\cos x = -1 \cdot \sin x + 0 \cdot \cos x = -\sin x = \frac{d^2}{dx^2}\sin x,$$

$$\sin\left(x + \frac{3\pi}{2}\right) = \cos\frac{3\pi}{2}\sin x + \sin\frac{3\pi}{2}\cos x = 0 \cdot \sin x + -1 \cdot \cos x = -\cos x = \frac{d^3}{dx^3}\sin x.$$

In other words, differentiating the sine function has the effect of translating the argument by $\pi/2$; doing this twice has the effect of changing the sign, since $\sin(x + \pi) = -\sin x$, and doing it four times has the same effect as doing nothing, since the function is periodic with period 2π .

Example 16.6. We can compute $\frac{d^{53}}{dx^{53}}\sin x$ by observing that $53 = 4 \cdot 13 + 1$, so the first $52 = 4 \cdot 13$ derivatives get us back to $\sin x$, and we thus have $\frac{d^{53}}{dx^{53}}\sin x = \frac{d}{dx}\sin x = \cos x$.

Of particular importance is the fact that if $f(x) = \sin x$ or $f(x) = \cos x$ (or more generally if $f(x) = a\sin x + b\cos x$ for some constants $a, b \in \mathbb{R}$), then

$$(16.5) \quad f''(x) = -f(x) \text{ for all } x \in \mathbb{R}.$$

This is a tremendously important example of a *differential equation*, which arises naturally in many different areas of science.

Example 16.7. Consider an object on a spring that at time t is displaced by a distance r from its rest position; then the force acting on the object has magnitude kr , where $k > 0$ is the *spring constant* that measures the strength of the spring, and this force is directed in the opposite direction of the displacement. Thus if m is the mass of the object, then Newton's second law gives

$$\frac{d^2r}{dt^2} = \text{acceleration} = \frac{\text{force}}{\text{mass}} = -\frac{kr}{m},$$

and if we define a function f by $f(x) = r(x/\omega)$ where $\omega := \sqrt{k/m}$, we can use (14.5) to get

$$f'(x) = \frac{d}{dt}r\left(\frac{x}{\omega}\right) = \frac{1}{\omega}r'\left(\frac{x}{\omega}\right) \quad \Rightarrow \quad \begin{cases} f''(x) = \frac{d}{dx}\left(\frac{1}{\omega}r'\left(\frac{x}{\omega}\right)\right) = \frac{1}{\omega^2}r''\left(\frac{x}{\omega}\right) \\ \quad \quad \quad = -\frac{1}{\omega^2}\frac{k}{m}r\left(\frac{x}{\omega}\right) = -r\left(\frac{x}{\omega}\right) = f(x), \end{cases}$$

which means that f must be a function that satisfies the differential equation in (16.5). One can prove (though we won't do it yet) that *every* function satisfying this differential equation can be written as $f(x) = a\sin x + b\cos x$, and we conclude that the displacement of the object is given by

$$(16.6) \quad r(t) = f(\omega t) = a\sin(\omega t) + b\cos(\omega t).$$

Exercise 16.8. Prove that if $r(t)$ is given by (16.6), then there are $A > 0$ and $\phi \in \mathbb{R}$ such that $r(t) = A\sin(\omega t + \phi)$.

The exercise demonstrates that the function r describing the position of an object on a spring can be written in terms of sine (or cosine). This system is an example of a *simple harmonic oscillator*.

Lecture 17

Differentiating exponentials

DATE: WEDNESDAY, OCTOBER 2

This lecture fills in some details that do not appear in either Stewart or Spivak, although convexity is discussed in the Appendix to Chapter 11 of Spivak, and in §4.3 of Stewart (from a different point of view).

Fix $a > 0$ and let $f(x) = a^x$. In Lecture 14.2 we computed $f'(x)$ by writing

$$(17.1) \quad f'(x) = \lim_{h \rightarrow 0} \frac{a^{x+h} - a^x}{h} = \lim_{h \rightarrow 0} a^x \cdot \frac{a^h - 1}{h} = a^x \lim_{h \rightarrow 0} \frac{a^h - 1}{h}.$$

But why should this last limit exist? In this lecture we prove the following theorem, which shows that the limit exists, and hence that f is differentiable on \mathbb{R} ; it also characterizes the natural logarithm of a as $\left. \frac{d}{dx} a^x \right|_{x=0}$.

Theorem 17.1. *For every $a > 0$, the limit $\lim_{h \rightarrow 0} \frac{a^h - 1}{h}$ exists; writing $g(a)$ for the value of this limit, the function $g: (0, \infty) \rightarrow \mathbb{R}$ has the following properties:*

- (1) g is (strictly) increasing and continuous;
- (2) $\lim_{a \rightarrow 0^+} g(a) = -\infty$, $g(1) = 0$, and $\lim_{a \rightarrow \infty} g(a) = +\infty$;
- (3) $g(ab) = g(a) + g(b)$ for all $a, b > 0$;

As we saw in Lecture 14.2, the function g defined in Theorem 17.1 is the inverse of the function $x \mapsto a^x$, where $a = g^{-1}(1)$, and we call $\ln(a) := g(a)$ the *natural logarithm* of a .

Remark 17.2. The proof of Theorem 17.1 that we give in the remainder of this lecture is fairly technical, and in the classroom lecture I omit many of the details, opting instead to highlight the main points. I do not expect that you will learn the details of the proofs in this lecture; I include them largely for completeness, in order to be consistent with the principle that I should not expect you to believe things without justification. Whether or not you follow the details of the proofs here, you should pay some attention to the notion of convexity in the next section; this will reappear later on.

17.1. Convexity of exponential functions

Before proving Theorem 17.1, we need to establish some preliminaries, which takes a bit of work but turns out to be essential. First let us recall our notion of *increasing* and *decreasing* functions, which are illustrated in Figure 3. There are multiple ways to describe the property of being increasing that is exhibited in the left-hand picture.

- The slope of every tangent line is positive.
- The derivative f' is positive everywhere.
- The slope of every secant line is positive.

- As x increases, the value of $f(x)$ increases.

The first two of these only make sense if f is differentiable, while the last two make sense for every f . (Observe that the last two are equivalent to each other.) Thus we can take either of these last two as the definition of increasing.

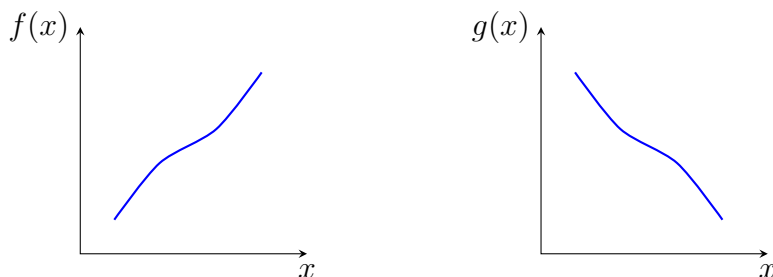


FIGURE 3. An increasing function f and a decreasing function g .

Now consider the two functions illustrated in Figure 4. The function f is an example of a *convex* function, and the function g is *concave*. But what do these words mean? How would you describe the difference between f and g to someone who could not see the pictures of their graphs?

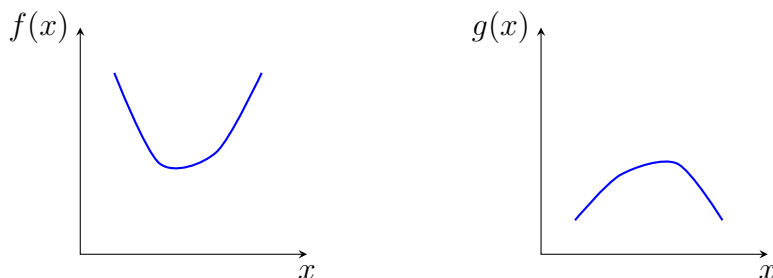


FIGURE 4. A convex function f and a concave function g .

A first attempt might be to say that the graph of f “bends upwards”, or “opens upwards”, while the graph of g “bends downwards”. One way to make this more precise would be to phrase it in terms of the tangent lines: the function f has the following two properties.

- The graph of f lies above all its tangent lines.
- As x moves to the right, the slope of the tangent line at x increases.

The second of these can also be written in terms of the derivative, since $f'(x)$ is the slope of the tangent line at x .

- As x increases, the value of $f'(x)$ increases.

If f is twice differentiable, then the derivative f' being increasing corresponds to the second derivative f'' being positive. So we might define convexity as

- the second derivative satisfies $f''(x) > 0$ everywhere.

On the other hand, we might want to consider functions that are *not* differentiable. What if we put a ‘corner’ in the graph of f , but kept the same basic shape? None of the previous descriptions would make sense, because f' would not be defined everywhere. However, if we replace ‘tangent line’ with ‘secant line’, then we get a description that works even without differentiability: f has the property that

- as either x or y moves to the right (with the other one staying in the same place), the slope of the secant line from $(x, f(x))$ to $(y, f(y))$ increases.

Suppose $x < y$. In order for the slope of the secant line to increase as y increases, we need the graph of f to lie below the secant line just to the left of y , and above the secant line just to the right of y . Similarly, the graph should lie below the secant line just to the right of x , and above it just to the left of x . So we give the following description.

- Between x and y , the secant line through those points lies above the graph.

Exercise 17.3. Using proof by contradiction, show that if the graph of f has the last property above, then at any z that is *not* between x and y , the secant line lies above the graph. Use this to prove the property about the slope of the secant line increasing.

Exercise 17.4. Determine the relationship between the six conditions listed above. Are they all equivalent when f is differentiable, or twice differentiable? Or are there some functions which satisfy certain conditions but not others?

All of the descriptions above capture various aspects of the ‘shape’ of the graph of f . For now we will take the last one as our definition of ‘convex’. To formulate it precisely, we observe that given any z between x and y , we can write $z = tx + (1 - t)y$ for some $t \in [0, 1]$ (the choice $t = \frac{z-y}{x-y}$ works), and the vertical coordinate of the secant line at this point is $tf(x) + (1 - t)f(y)$. With this in mind, we make the following definition.

Definition 17.5. Let $I \subset \mathbb{R}$ be an interval. A function $f: I \rightarrow \mathbb{R}$ is *convex* if for every $x, y \in I$ and every $t \in [0, 1]$, we have

$$(17.2) \quad f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

We call f *strictly convex* if \leq can be replaced by $<$ in (17.2) whenever $x \neq y$ and $t \in (0, 1)$.

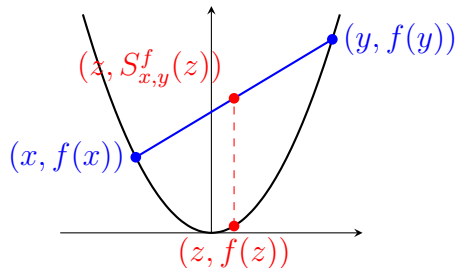
Example 17.6. The function $f(x) = x^2$ is strictly convex. To see this, we write the difference between the right- and left-hand sides of (17.2) as

$$\begin{aligned} & tf(x) + (1 - t)f(y) - f(tx + (1 - t)y) \\ &= tx^2 + (1 - t)y^2 - (t^2x^2 + 2t(1 - t)xy + (1 - t)^2y^2) \\ &= (t - t^2)x^2 - 2(t - t^2)xy + (1 - t - (1 - 2t + t^2))y^2 \\ &= (t - t^2)(x^2 - 2xy + y^2) = t(1 - t)(x - y)^2. \end{aligned}$$

This last quantity is > 0 for every $x \neq y$ and $t \in (0, 1)$, which proves (17.2) with a strict inequality. Thus f is strictly convex.

Remark 17.7. Later, in Lecture 27.2, we will see an easier way to establish convexity of $f(x) = x^2$ using second derivatives.

As described above, the definition of convexity says that on the interval between x, y , the graph of the function f lies below the secant line through the points $(x, f(x))$ and $(y, f(y))$. The picture at right illustrates this for $f(x) = x^2$, using the notation $z = tx + (1 - t)y$ and writing $S_{x,y}^f$ for the function whose graph is the secant line through $(x, f(x))$ and $(y, f(y))$.¹⁹



To find a formula for $S_{x,y}^f$, start by observing that the secant line through $(x, f(x))$ and $(y, f(y))$ is the set of all points (z, w) such that²⁰

$$\frac{w - f(x)}{z - x} = \frac{f(y) - f(x)}{y - x}.$$

This is equivalent to $w - f(x) = \frac{f(y) - f(x)}{y - x}(z - x)$, and thus we see that the secant line is the graph of the function

$$(17.3) \quad S_{x,y}^f(z) := \frac{f(y) - f(x)}{y - x}(z - x) + f(x).$$

For later use we rewrite (17.3) as

$$(17.4) \quad S_{x,y}^f(z) = f(y)\frac{z - x}{y - x} + f(x)\left(1 - \frac{z - x}{y - x}\right) = f(y)\frac{z - x}{y - x} + f(x)\frac{y - z}{y - x}.$$

Exercise 17.8. Prove that if $z = tx + (1 - t)y$ for some $t \in [0, 1]$, then $S_{x,y}^f(z) = tf(x) + (1 - t)f(y)$. Use this to deduce that a function $f: I \rightarrow \mathbb{R}$ is convex if and only if for every $x, y \in I$ and every z between x and y , we have $f(z) \leq S_{x,y}^f(z)$, and strictly convex if this inequality is strict whenever $z \neq x, y$.

Now fix $a > 0$ and consider the function $f(x) = a^x$. Our goal is to prove that this function is convex²¹ on \mathbb{R} , which amounts to showing that

$$(17.5) \quad (a^x)^t(a^y)^{1-t} \leq ta^x + (1 - t)a^y$$

for all $x \neq y \in \mathbb{R}$ and $t \in (0, 1)$. It is not immediately clear how to prove this for an arbitrary t ; however, in the specific case $t = 1/2$, our desired inequality reduces to

$$(17.6) \quad \sqrt{a^x a^y} \leq \frac{a^x + a^y}{2},$$

which seems more manageable. Can we prove that (17.6) holds for all $a > 0$ and $x \neq y$?

Our immediate response to something like (17.6) is to square both sides (perhaps after multiplying both sides by 2); since both sides are positive, we see that (17.6) is true if and only if

$$4a^x a^y \leq (a^x + a^y)^2 = a^{2x} + 2a^x a^y + a^{2y}.$$

¹⁹The subscripts x, y and superscript f are just there to remind us which points and which function we used in defining $S_{x,y}^f$.

²⁰Of course we would naturally write (x, y) for a point in \mathbb{R}^2 , but we already used these symbols, so we must pick different ones; and after all, what's in a name?

²¹In fact it turns out to be *strictly* convex, but we will settle for convex for now.

But this is equivalent to $0 \leq a^{2x} - 2a^x a^y + a^{2y}$, which we know is always true since the right-hand side can be written as $(a^x - a^y)^2$. Thus we have proved that (17.5) is true for all $x, y \in \mathbb{R}$ and $t = 1/2$; in other words, while we don't yet know that the graph of f lies below the secant line for x, y *everywhere* on the interval between x and y , it at least does so at the midpoint of this interval. We formulate this fact as a lemma, whose proof is given by the above discussion.

Lemma 17.9. *Given any $a > 0$, the exponential function $f(x) = a^x$ has the following property: for all $x, y \in \mathbb{R}$ with $x \neq y$, the point $c = \frac{x+y}{2}$ satisfies $f(c) < \frac{1}{2}(f(x) + f(y)) = S_{x,y}^f(c)$.*

In our proof of convexity, we will find it useful to have a tool that lets us translate inequalities such as the one in Lemma 17.9 into statements comparing secant lines.

Lemma 17.10. *Let I be an interval, $f: I \rightarrow \mathbb{R}$ any function, and $x, y \in I$ two numbers with $x < y$. If $c \in (x, y)$ has the property that $f(c) \leq S_{x,y}^f(c)$, then we have*

$$(17.7) \quad \text{slope}(S_{x,c}^f) \leq \text{slope}(S_{x,y}^f) \leq \text{slope}(S_{c,y}^f).$$

In particular, we have

$$(17.8) \quad \begin{aligned} S_{c,y}^f(z) &\leq S_{x,y}^f(z) \text{ for all } z \text{ between } c \text{ and } y, \\ S_{x,c}^f(z) &\leq S_{x,y}^f(z) \text{ for all } z \text{ between } x \text{ and } c. \end{aligned}$$

Proof. The inequalities in (17.8) follow directly from (17.7) by noting where the secant lines intersect each other. To prove (17.7), observe that

$$\text{slope}(S_{x,c}^f) = \frac{f(c) - f(x)}{c - x} \leq \frac{S_{x,y}^f(c) - f(x)}{c - x} = \text{slope}(S_{x,y}^f),$$

and similarly,

$$\text{slope}(S_{c,y}^f) = \frac{f(y) - f(c)}{y - c} \geq \frac{f(y) - S_{x,y}^f(c)}{y - c} = \text{slope}(S_{x,y}^f). \quad \square$$

Now the way to prove convexity of the exponential function is to iterate Lemma 17.9 using bisection sequences as in the proof of the Intermediate Value Theorem, as suggested by Figure 5. In fact we will prove a more general result, valid for any continuous function satisfying a convexity-type inequality at midpoints.

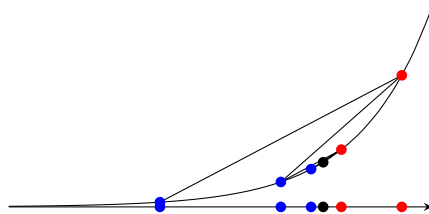


FIGURE 5. Midpoint convexity and continuity imply convexity.

Theorem 17.11. *Let $f: I \rightarrow \mathbb{R}$ be a continuous function with the property that $f(\frac{x+y}{2}) \leq \frac{1}{2}(f(x) + f(y))$ for all $x, y \in I$. Then f is convex.*

Proof. We will use the characterization of convexity from Exercise 17.8. Fix $x, y \in \mathbb{R}$; without loss of generality assume that $x < y$. Given a point $z \in (x, y)$, define a pair of bisection sequences $b_n, r_n \rightarrow z$ recursively as follows (see Figure 5):

- Start with $b_1 = x$ and $r_1 = y$.
- For every $n \geq 1$, once b_n and r_n have been determined, let $m_n = \frac{1}{2}(b_n + r_n)$ and proceed by cases:
 - (1) if $b_n \leq z \leq m_n$, then let $b_{n+1} = b_n$ and $r_{n+1} = m_n$;
 - (2) if $m_n < z \leq r_n$, then let $b_{n+1} = m_n$ and $r_{n+1} = r_n$.

This procedure should be familiar from our proof of the Intermediate Value Theorem. Note that we have $b_n \leq z \leq r_n$ for every n , and since both sequences are monotonic, their limits exist. Since $r_n - b_n \rightarrow 0$, the limits are the same, and equal to z . Finally, continuity of f and of $S_{x,y}^f$ shows that

$$(17.9) \quad f(z) = \lim_{n \rightarrow \infty} f(b_n), \quad S_{x,y}^f(z) = \lim_{n \rightarrow \infty} S_{x,y}^f(b_n).$$

(The same would be true if we replaced b_n with r_n .) At each step the number m_n is the midpoint of b_n, r_n , and by the definition of b_{n+1}, r_{n+1} , it follows from (17.8) that

$$S_{b_{n+1}, r_{n+1}}^f(w) \leq S_{b_n, r_n}^f(w) \text{ for all } w \in [b_n, r_n].$$

Since $b_n \in [b_k, r_k]$ for every $k \leq n$, it follows by induction that

$$f(b_n) = S_{b_n, r_n}^f(b_n) \leq S_{x,y}^f(b_n).$$

Taking a limit as $n \rightarrow \infty$ and using (17.9), this implies that $f(z) \leq S_{x,y}^f(z)$ by the theorem on monotonicity of limits. Since this holds for any $z \in (x, y)$, we conclude that f is convex. \square

Corollary 17.12. *For every $a > 0$, the exponential function $f(x) = a^x$ is convex.*

Remark 17.13. The proof of Theorem 17.11 can be modified to show that the exponential function is in fact *strictly* convex, using the fact that the inequality in Lemma 17.9 is strict, but we will not need this fact for now.

17.2. Exponential functions are differentiable

We will use a basic property of convex functions, which follows from Lemma 17.10.

Proposition 17.14. *Let $f: I \rightarrow \mathbb{R}$ be a convex function on an open interval I . Then the left and right derivatives of f exist everywhere in I , and for every $x < z < y$ we have*

$$\text{slope}(S_{x,z}^f) \leq D^- f(z) \leq D^+ f(z) \leq \text{slope}(S_{z,y}^f).$$

The rest of this section and the next were skipped in the classroom lecture

Proof. Given any $x_1 < x_2 < z < y$, we can apply Lemma 17.10 twice (once to $x_1 < x_2 < z$ and once to $x_2 < z < y$) to obtain

$$\text{slope}(S_{x_1,z}^f) \leq \text{slope}(S_{x_2,z}^f) \leq \text{slope}(S_{x_2,y}^f) \leq \text{slope}(S_{z,y}^f).$$

This shows that the function $x \mapsto \text{slope}(S_{x,z}^f)$ is a nondecreasing function on the interval $x < z$, and that it is bounded above by the number $\text{slope}(S_{z,y}^f)$. It follows from the

monotone convergence theorem that $\lim_{x \rightarrow z^-} \text{slope}(S_{x,z}^f) = \lim_{x \rightarrow z^-} \frac{f(x)-f(z)}{x-z}$ exists and is $\leq \text{slope}_{z,y}^f$. But this limit is the definition of the left derivative of f at z , and thus

$$\text{slope}(S_{x,z}^f) \leq D^-f(z) \leq \text{slope}(S_{z,y}^f)$$

holds for all $x < z < y$. Now a similar argument applied to the function $y \mapsto \text{slope}(S_{z,y}^f)$, with $D^-f(z)$ as the lower bound, completes the proof of the proposition. \square

Now we can prove differentiability of the exponential function. Given $b > 0$, it follows from Corollary 17.12 and Proposition 17.14 that the function $f(x) = b^x$ has left and right derivatives at every point. We must show that they agree.²² To this end, first note that for every $z \in \mathbb{R}$ we have

$$\begin{aligned} D^+f(z) &= \lim_{h \rightarrow 0^+} \frac{b^{z+h} - b^z}{h} = b^z \lim_{h \rightarrow 0^+} \frac{b^h - 1}{h} = b^z D^+f(0), \\ D^-f(z) &= \lim_{h \rightarrow 0^-} \frac{b^{z+h} - b^z}{h} = b^z \lim_{h \rightarrow 0^-} \frac{b^h - 1}{h} = b^z D^-f(0). \end{aligned}$$

For concreteness, suppose that $b > 1$. In this case, Proposition 17.14 gives

$$D^+f(0) \geq D^-f(0) \geq \text{slope}_{S_{-1,0}^f} = \frac{b^0 - b^{-1}}{0 - (-1)} = 1 - b^{-1} > 0,$$

so we can define $c := D^+f(0)/D^-f(0)$; observe that $c \geq 1$ and that $D^+f(z) = cD^-f(z)$ for all $z \in \mathbb{R}$. Thus to prove that f is differentiable everywhere, we must prove that $c = 1$. Given any $n \in \mathbb{N}$ we have

$$\begin{aligned} \frac{D^-f(1)}{D^-f(0)} &= \left(\frac{D^+f(0)}{D^-f(0)} \frac{D^-f(\frac{1}{n})}{D^+f(\frac{1}{n})} \right) \left(\frac{D^+f(\frac{1}{n})}{D^-f(\frac{1}{n})} \frac{D^-f(\frac{2}{n})}{D^+f(\frac{2}{n})} \right) \cdots \left(\frac{D^+f(\frac{n-1}{n})}{D^-f(\frac{n-1}{n})} \frac{D^-f(1)}{D^+f(\frac{n-1}{n})} \right) \\ &\geq \frac{D^+f(0)}{D^-f(0)} \frac{D^+f(\frac{1}{n})}{D^-f(\frac{1}{n})} \cdots \frac{D^+f(\frac{n-1}{n})}{D^-f(\frac{n-1}{n})} = c^n, \end{aligned}$$

where the inequality uses the fact that Proposition 17.14 gives

$$D^+f\left(\frac{k}{n}\right) \leq \text{slope}_{S_{\frac{k}{n}, \frac{k+1}{n}}^f} \leq D^-f\left(\frac{k+1}{n}\right) \text{ for all } 0 \leq k < n.$$

But this means that $1 \leq c \leq (D^-f(1)/D^-f(0))^{1/n}$ for all $n \in \mathbb{N}$, and thus $c = 1$. We conclude that f is differentiable on \mathbb{R} .

17.3. Characterization of the logarithm

Now we return to the proof of Theorem 17.1. We have shown that $f(x) = b^x$ is differentiable on \mathbb{R} for every $b > 0$, and in particular, the function

$$(17.10) \quad g(b) = \frac{d}{dx} b^x \Big|_{x=0} = \lim_{h \rightarrow 0} \frac{b^h - 1}{h}$$

is well-defined.

Lemma 17.15. *For every $a, b > 0$, we have $g(ab) = g(a) + g(b)$.*

²²I got this argument from a post by Todd Trimble at nLab.

Proof. Doing a little algebra and using the limit laws, we get

$$\begin{aligned} g(ab) &= \lim_{h \rightarrow 0} \frac{(ab)^h - 1}{h} = \lim_{h \rightarrow 0} \frac{a^h b^h - b^h + b^h - 1}{h} \\ &= \underbrace{\lim_{h \rightarrow 0} b^h}_I \underbrace{\lim_{h \rightarrow 0} \frac{a^h - 1}{h}}_{II} + \underbrace{\lim_{h \rightarrow 0} \frac{b^h - 1}{h}}_{III}. \end{aligned}$$

Limit I exists and is equal to $b^0 = 1$ because exponential functions are continuous. Limits II and III exist because exponential functions are differentiable, and the limits are equal to $g(a)$ and $g(b)$, respectively. Thus $g(ab) = 1 \cdot g(a) + g(b)$, which completes the proof of the lemma. \square

To understand the remaining properties of g , we start by getting some loose bounds. Given $c > 0$, Proposition 17.14 gives

$$g(c) = \frac{d}{dx} c^x \Big|_{x=0} \leq \text{slope}(S_{0,1}^f) = \frac{c-1}{1} = c-1,$$

and similarly

$$g(c) = \frac{d}{dx} c^x \Big|_{x=0} \geq \text{slope}(S_{-1,0}^f) = \frac{c^{-1}-1}{-1} = 1-c^{-1}.$$

Together, these imply that

$$(17.11) \quad 1 - \frac{1}{c} \leq g(c) \leq c - 1 \text{ for all } c > 0.$$

When $c = 1$ this gives $0 = 1 - \frac{1}{1} \leq g(1) \leq 1 - 1 = 0$, and thus $g(1) = 0$.

Remark 17.16. We could also deduce this last fact by using Lemma 17.15 to get $g(1) = g(1 \cdot 1) = g(1) + g(1)$, and then subtract $g(1)$ from both sides. A third option would be to work directly from the definition of g in (17.10).

The bounds in (17.11) are the key to establishing the remaining properties of g . Given $a, b > 0$, we can use Lemma 17.15 and then apply (17.11) with $c = b/a$ to get

$$g(b) = g\left(a \cdot \frac{b}{a}\right) = g(a) + g\left(\frac{b}{a}\right) \geq g(a) + 1 - \frac{a}{b},$$

and similarly,

$$g(b) = g(a) + g\left(\frac{b}{a}\right) \leq g(a) + \frac{b}{a} - 1.$$

Putting these together gives

$$(17.12) \quad g(a) + \frac{b-a}{b} \leq g(b) \leq g(a) + \frac{b-a}{a} \text{ for all } a, b > 0.$$

Lemma 17.17. *The function $g: (0, \infty) \rightarrow \mathbb{R}$ is strictly increasing.*

Proof. Given $0 < a < b$, the lower bound in (17.12) gives $g(b) \geq g(a) + (b-a)/b > g(a)$. \square

Lemma 17.18. *The function g is continuous on $(0, \infty)$.*

Proof. As $b \rightarrow a$, the lower and upper bounds for $g(b)$ given by (17.12) both approach $g(a)$, and thus $\lim_{b \rightarrow a} g(b) = g(a)$ by the Squeeze Theorem. \square

We have proved all of Theorem 17.1 except for the statements about the asymptotic behavior of g . To prove these, we observe that $g(2) > g(1) = 0$ by Lemma 17.17, and iterating Lemma 17.15 gives

$$g(2^n) = \underbrace{g(2) + \cdots + g(2)}_{n \text{ times}} = ng(2).$$

Now we claim that $\lim_{a \rightarrow \infty} g(a) = \infty$. Indeed, given any $R > 0$, let $n \in \mathbb{N}$ be such that $n \geq R/g(2)$; then for every $a > 2^n$ we have

$$g(a) > g(2^n) = ng(2) \geq R.$$

Since R was arbitrary, this proves that $\lim_{a \rightarrow \infty} g(a) = \infty$. Moreover, since $g(a^{-1}) + g(a) = g(a^{-1}a) = g(1) = 0$, we have $g(a^{-1}) = -g(a)$, and thus for all $a \in (0, 2^{-n})$ we have

$$g(a) < g(2^{-n}) = -g(2^n) = -ng(2) \leq -R.$$

This proves that $\lim_{a \rightarrow 0^+} g(a) = -\infty$, and completes the proof of Theorem 17.1.

Lecture 18

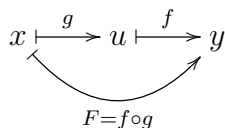
Chain rule

DATE: FRIDAY, OCTOBER 4

This lecture corresponds to §3.4 in Stewart and Chapter 10 in Spivak.

In (14.5), we saw that $\frac{d}{dx}f(cx) = cf'(cx)$. The geometric interpretation of this is that $x \mapsto f(cx)$ has a graph given by taking the graph of $f(x)$ and “squashing” it in the x -direction by a factor of $\frac{1}{c}$, which has the effect of multiply slopes by c .

Now we prove a more general version of this result. Suppose that f and g are two functions for which $f(g(a))$ is well-defined (that is, a is in the domain of g and $g(a)$ is in the domain of f). Given $x \approx a$, let $u = g(x)$ and $y = f(g(x))$, so that x, u, y are related as follows:



Writing $F = f \circ g$, we want to find a formula for F' in terms of f , g , and their derivatives. Intuitively, the idea is that $F' = \frac{dy}{dx}$ measures the ratio $\frac{\Delta y}{\Delta x}$, where Δy is the amount by which $y = F(x) = f(g(x))$ changes in response to a small change Δx in x . Using our (by now standard) trick of multiplying and dividing by the same thing, we can write

$$(18.1) \quad \frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta u} \frac{\Delta u}{\Delta x} = \lim_{\Delta u \rightarrow 0} \frac{\Delta y}{\Delta u} \cdot \lim_{\Delta x \rightarrow 0} \frac{\Delta u}{\Delta x} = \frac{dy}{du} \frac{du}{dx}.$$

Since $\frac{dy}{dx}|_{x=a} = (f \circ g)'(a)$, $\frac{dy}{du}|_{u=g(a)} = f'(g(a))$, and $\frac{du}{dx}|_{x=a} = g'(a)$, we can write (18.1) as

$$(18.2) \quad (f \circ g)'(a) = f'(g(a))g'(a),$$

an identity which is called the *chain rule*. There is one small problem, though; what if $\Delta u = 0$ in (18.1)? Then the computation is not valid, because we divided by 0. Thus

(18.1) is not quite a proof. The trick to making it a rigorous proof is to think about what quantity should replace $\frac{\Delta y}{\Delta u}$ when $\Delta u = 0$; presumably in this case we should use the derivative $\frac{dy}{du}$ itself. This reasoning leads us to the following argument.

Theorem 18.1 (Chain rule). *Let f and g be functions and $a \in \mathbb{R}$ a point such that g is differentiable at a , and f is differentiable at $g(a)$.²³ Then $f \circ g$ is differentiable at a , and the identity (18.2) holds.*

Proof. In terms of our usual notation for computing derivatives, we have $\Delta x = h$, $\Delta u = g(a + h) - g(a)$, and $\Delta y = f(g(a + h)) - f(g(a))$. We would like to write the second step in (18.1) as

$$\frac{f(g(a + h)) - f(g(a))}{h} = \frac{f(g(a + h)) - f(g(a))}{g(a + h) - g(a)} \cdot \frac{g(a + h) - g(a)}{h},$$

but the first fraction on the right-hand side may be undefined if $g(a + h) - g(a) = 0$. However, this represents a *removable* discontinuity, because f is differentiable at $g(a)$.

Lemma 18.2. *The function ϕ defined by*

$$\phi(h) = \begin{cases} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} & \text{if } g(a + h) - g(a) \neq 0, \\ f'(g(a)) & \text{if } g(a + h) - g(a) = 0, \end{cases}$$

is continuous at 0.

Proof. Because f is differentiable at $g(a)$, we have

$$\lim_{t \rightarrow 0} \frac{f(g(a) + t) - f(g(a))}{t} = f'(g(a)).$$

Thus for every $\epsilon > 0$ there exists $\delta' > 0$ such that

$$(18.3) \quad \text{if } 0 < |t| < \delta', \text{ then } \left| \frac{f(g(a) + t) - f(g(a))}{t} - f'(g(a)) \right| < \epsilon.$$

Moreover, since g is differentiable at a , it is continuous there by Theorem 13.3. Thus there is $\delta > 0$ such that

$$(18.4) \quad \text{if } |h| < \delta, \text{ then } |g(a + h) - g(a)| < \delta'.$$

Now given h with $|h| < \delta$, we can write $t = g(a + h) - g(a)$ so that $g(a + h) = g(a) + t$.

- CASE I. If $t \neq 0$, then $0 < |t| < \delta'$ by (18.4), and so

$$\phi(h) = \frac{f(g(a + h)) - f(g(a))}{g(a + h) - g(a)} = \frac{f(g(a) + t) - f(g(a))}{t}$$

satisfies $|\phi(h) - \phi(0)| = |\phi(h) - f'(g(a))| < \epsilon$ by (18.3).

- CASE II. If $t = 0$, then $\phi(h) = f'(g(a)) = \phi(0)$ by definition.

In both cases, we conclude that $|\phi(h) - \phi(0)| < \epsilon$, which proves continuity at 0. \square

²³In particular, this requires that a is in the domain of g , and $g(a)$ is in the domain of $f(a)$.

Returning to the proof of the chain rule, we observe that for every h , we have

$$\frac{f(g(a+h)) - f(g(a))}{h} = \phi(h) \cdot \frac{g(a+h) - g(a)}{h},$$

where when $g(a+h) \neq g(a)$ this follows from the definition of ϕ , and when $g(a+h) = g(a)$ it follows since both sides are equal to 0. Thus we have

$$\begin{aligned} (f \circ g)'(a) &= \lim_{h \rightarrow 0} \frac{f(g(a+h)) - f(g(a))}{h} = \lim_{h \rightarrow 0} \phi(h) \cdot \lim_{h \rightarrow 0} \frac{g(a+h) - g(a)}{h} \\ &= \phi(0) \cdot g'(a) = f'(g(a))g'(a), \end{aligned}$$

where the first equality on the second line uses Lemma 18.2, and the second uses the definition of $\phi(0)$. \square

Example 18.3. Consider the function $F(x) = \sqrt{x^2 - 1}$. We can write this as the composition of the functions $g(x) = x^2 - 1$ and $f(u) = \sqrt{u}$, whose derivatives we know to be

$$f'(u) = \frac{1}{2\sqrt{u}} \quad \text{and} \quad g'(x) = 2x.$$

By the chain rule, we have

$$F'(x) = (f \circ g)'(x) = f'(g(x))g'(x) = \frac{1}{2\sqrt{g(x)}} \cdot 2x = \frac{x}{\sqrt{x^2 - 1}}.$$

Example 18.4. The function $f(x) = \sin \frac{1}{x}$ can be written as $f = g \circ h$ where $g(u) = \sin u$ and $h(x) = \frac{1}{x}$. The chain rule gives

$$f'(x) = g'(h(x))h'(x) = \cos(h(x)) \cdot \left(-\frac{1}{x^2}\right) = -\frac{\cos \frac{1}{x}}{x^2}.$$

Example 18.5. For any differentiable function g , we can compute $\frac{d}{dx}(g(x))^n$ by recalling that $f(u) = u^n$ has $f'(u) = nu^{n-1}$ and using the chain rule to get

$$\frac{d}{dx}(g(x))^n = f'(g(x))g'(x) = n(g(x))^{n-1}g'(x).$$

Example 18.6. In (14.4) we saw that $\frac{d}{dx}a^x = (\ln a)a^x$, which can be rewritten as $\frac{d}{dx}e^{cx} = ce^{cx}$, where $c = \ln a$. In (14.5) we pointed out that since $\frac{d}{dx}e^x = e^x$, this is a special case of the more general formula $\frac{d}{dx}f(cx) = cf'(cx)$. Now we see that this in turn is a special case of the chain rule, where we put $g(x) = cx$ and get

$$\frac{d}{dx}f(g(x)) = g'(x)f'(g(x)) = cf'(cx).$$

Remark 18.7. There is a subtlety here that is worth emphasizing. We are used to writing $f'(x)$ and $\frac{d}{dx}f(x)$ to mean the same thing – the derivative of f with respect to x . However, $f'(cx)$ and $\frac{d}{dx}f(cx)$ do *not* mean the same thing. The latter – $\frac{d}{dx}f(cx)$ – means the derivative of $x \mapsto f(cx)$ with respect to x , which gives the sensitivity of $f(cx)$ to a small change in x . The former – $f'(cx)$ – means the derivative with respect to the input of f , which in this case is cx , and so it gives the sensitivity of $f(cx)$ to a small change in cx . You should keep this in mind when using and reading the notations $\frac{d}{dx}$ and f' : the first of these always means a rate of change with respect to x , while the second means a rate of change with respect to the input of f , whatever that input is called.

Example 18.8. Using the chain rule and the power rule, we get

$$\begin{aligned}\frac{d}{dx} \frac{1}{\sqrt{x^2+1}} &= \frac{d}{dx} (x^2+1)^{-1/2} \\ &= -\frac{1}{2}(x^2+1)^{-3/2} \frac{d}{dx} (x^2+1) = \frac{-2x}{2(x^2+1)^{3/2}} = \frac{-x}{(x^2+1)^{3/2}}.\end{aligned}$$

Note that although we did not write the functions out explicitly, we took $g(x) = x^2 + 1$ and $f(u) = u^{-1/2}$, so that $f \circ g(x) = f(x^2 + 1) = (x^2 + 1)^{-1/2}$; then the chain rule is used to get the first equality on the second line.

Example 18.9. The rational function $F(t) = \left(\frac{t+1}{t-1}\right)^2$ can be differentiated in several ways. We could write it as $\frac{(t+1)^2}{(t-1)^2}$ and use the quotient rule directly. Or we can use the chain rule with $g(t) = \frac{t+1}{t-1}$ and $f(u) = u^2$, then apply the quotient rule to g (which is simpler than F) and obtain

$$\begin{aligned}F'(t) &= \frac{d}{dt} \left(\left(\frac{t+1}{t-1} \right)^2 \right) = 2 \cdot \frac{t+1}{t-1} \cdot \frac{d}{dt} \left(\frac{t+1}{t-1} \right) \\ &= 2 \cdot \frac{t+1}{t-1} \left(\frac{t-1 - (t+1)}{(t-1)^2} \right) = \frac{-4(t+1)}{(t-1)^3}.\end{aligned}$$

Yet another way to obtain $F'(t)$ is to rewrite the quotient as $\frac{t+1}{t-1} = \frac{t-1+2}{t-1} = 1 + \frac{2}{t-1}$, so that

$$F(t) = \left(\frac{t+1}{t-1} \right)^2 = \left(1 + \frac{2}{t-1} \right)^2 = 1 + \frac{4}{t-1} + \frac{4}{(t-1)^2}.$$

The power rule and chain rule together imply that $\frac{d}{dt}(t-1)^n = n(t-1)^{n-1}$ for all n , so we get

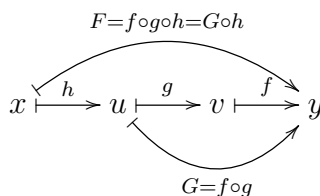
$$F'(t) = \frac{-4}{(t-1)^2} - 2 \cdot \frac{4}{(t-1)^3}.$$

A little more algebra shows that this is equal to the expression we got earlier. This example illustrates that there can be more than one correct way to compute a given derivative, and that the appearance of the answer may depend on the method chosen, even though all answers will be equivalent.

Example 18.10.

$$\frac{d}{dx} e^{\cos x} = e^{\cos x} \frac{d}{dx} \cos x = -\sin x e^{\cos x}.$$

The chain rule can be used to differentiate compositions of more than two functions; indeed, this motivates its name, since we can think of this as a ‘chain’ of functions composed with each other. For example, given three functions $f, g, h: \mathbb{R} \rightarrow \mathbb{R}$, we can visualize $F = f \circ g \circ h$ as follows:



Writing $G = f \circ g$, we can use the chain rule to write G' in terms of f' and g' , then again to write F' in terms of G' and h' , obtaining

$$F'(x) = (G \circ h)'(x) = G'(h(x))h'(x) = (f \circ g)'(h(x))h'(x) = f'(g(h(x))) \cdot g'(h(x)) \cdot h'(x).$$

It is perhaps easiest to view this using the alternate notation

$$(18.5) \quad \frac{dy}{dx} = \frac{dy}{dv} \frac{dv}{du} \frac{du}{dx},$$

where $u = h(x)$, $v = g(u) = g(h(x))$, and $y = f(v) = f(g(h(x)))$. We reiterate that although this notation makes it very tempting to view the chain rule as just a matter of ‘canceling the dv and the du from the numerator and denominator’, this is not an actual proof of anything, because the quantities $\frac{dy}{dv}$, $\frac{dv}{du}$, and $\frac{du}{dx}$ are derivatives, not fractions, and the symbols du and dv have no independent meaning.

Example 18.11. The function $F(x) = \sin(e^{3x})$ can be differentiated by writing $u = 3x$, $v = e^u = e^{3x}$, and $y = \sin v = \sin e^{3x}$. We get

$$F'(x) = \frac{dy}{dx} = \frac{dy}{dv} \frac{dv}{du} \frac{du}{dx} = (\cos v)(e^u)(3) = 3e^{3x} \cos e^{3x}.$$

Note that although applying (18.5) gave us an expression in terms of u, v, x , this was not yet our final answer; it was necessary to go one step further and write u and v in terms of the original variable x . You will encounter this phenomenon in other places as well.

Lecture 19

Implicit differentiation

DATE: MONDAY, OCTOBER 7

This lecture corresponds to §3.5 in Stewart and Chapter 12 in Spivak.

19.1. Implicitly defined functions

Suppose (x, y) is a point on the unit circle $x^2 + y^2 = 1$. What is the slope of the tangent line to the circle at (x, y) ?

One approach is to write y as a function of x , and then differentiate. If $y > 0$, then this means that $y = \sqrt{1 - x^2}$, and we can use the chain rule with $g(x) = 1 - x^2$ and $f(z) = z^{1/2}$ to deduce that

$$\text{slope} = \frac{dy}{dx} = \frac{d}{dx} \underbrace{(1 - x^2)^{1/2}}_{f \circ g(x)} = \frac{1}{2} \underbrace{(1 - x^2)^{-1/2}}_{f'(g(x))} \underbrace{(-2x)}_{g'(x)} = \frac{-2x}{2\sqrt{1 - x^2}} = -\frac{x}{y}.$$

A similar computation gives the same result if $y < 0$. Note that if $y = 0$ then the tangent line is vertical so the slope is undefined.

This approach works fine in this case, although the computation with the chain rule is a little messy. But what if we want to find the tangent line to a more complicated curve, such as the curve defined by the equation $x^3 + y^3 = xy$, which is shown in Figure 6? (Note that this curve was drawn by appealing to a computer program; we do not yet

have the tools to plot such curves by hand.) In this case it is not so easy to solve the equation and write y as a function of x . Fortunately, there is another approach available to us.

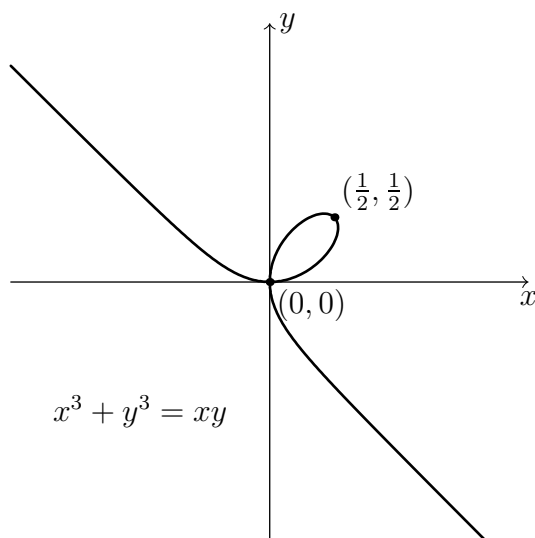


FIGURE 6. An implicitly defined function.

First let us return to the circle. The idea is that if $y = y(x)$ is a function that gives the dependence of y on x , then this function must satisfy the relationship $x^2 + (y(x))^2 = 1$. Both sides of this equation are functions of x , which we can differentiate using the rules we have discovered so far:

$$\frac{d}{dx}(x^2 + (y(x))^2) = 2x + 2y(x)\frac{dy}{dx},$$

$$\frac{d}{dx}1 = 0.$$

In the first equation we use the chain rule to differentiate $y(x)^2$. Because the two functions $x^2 + y(x)^2$ and 1 are equal, their derivatives are as well, and we conclude that

$$2x + 2y\frac{dy}{dx} = 0.$$

Solving for $\frac{dy}{dx}$ gives $\frac{dy}{dx} = -\frac{x}{y}$, just as before.

Remark 19.1. Note that this approach only works if we know beforehand that the point (x, y) is actually on the circle! If you use the formula $\frac{dy}{dx} = -\frac{x}{y}$ for a point that is not on the circle, you will get a meaningless answer.

The benefit of this second approach – *implicit differentiation* – is seen by considering the second example, $x^3 + y^3 = xy$. Here we differentiate both sides and get

$$(19.1) \quad 3x^2 + 3y^2\frac{dy}{dx} = x\frac{dy}{dx} + y \quad \Rightarrow \quad \frac{dy}{dx} = \frac{y - 3x^2}{3y^2 - x}.$$

Note that in both examples, we obtain an expression for $\frac{dy}{dx}$ that involves both x and y ; this is typical of solutions obtained using implicit differentiation.

Example 19.2. Let us find the points at which the curve $x^3 + y^3 = xy$ has a horizontal tangent line. From the computation in (19.1) we have $\frac{dy}{dx} = 0$ if and only if $y = 3x^2$ and $3y^2 \neq x$. If the point (a, b) lies on the curve and satisfies $b = 3a^2$, then we have

$$0 = a^3 + b^3 - ab = a^3 + (3a^2)^3 - a(3a^2) = 27a^6 - 2a^3 = a^3(27a^3 - 2),$$

so $a = 0$ or $a = \sqrt[3]{2/27} = \frac{1}{3}\sqrt[3]{2}$. In the first case we have $b = 3a^2 = 0$, in the second we have $b = 3a^2 = 3 \cdot \frac{1}{9} \cdot 2^{2/3} = \frac{1}{3}2^{2/3}$. At $(a, b) = (0, 0)$ the denominator in the formula for $\frac{dy}{dx}$ vanishes, while at $(a, b) = (\frac{1}{3}2^{1/3}, \frac{1}{3}2^{2/3})$ we have $3b^2 - a = \frac{3}{9}2^{4/3} - \frac{1}{9}2^{1/3} = \frac{5}{9}2^{1/3} \neq 0$, so at this point we have $\frac{dy}{dx} = 0$ and the curve has a horizontal tangent line.

Remark 19.3. The problem with the point $(0, 0)$ in the previous example has to do with the fact that the curve crosses itself at this point, as shown in Figure 6, so it does not have a uniquely defined tangent line.

Example 19.4. We can find the points with a vertical tangent line by reversing the roles of x and y , so that x is a function of y . Mimicking (19.1), we can differentiate both sides of $x^3 + y^3 = xy$ with respect to y and get

$$3x^2 \frac{dx}{dy} + 3y^2 = x + y \frac{dx}{dy} \quad \Rightarrow \quad \frac{dx}{dy} = \frac{x - 3y^2}{3x^2 - y}.$$

A similar computation to the one in the previous example shows that there is a vertical tangent line at $(a, b) = (\frac{1}{3}2^{2/3}, \frac{1}{3}2^{1/3})$.

Example 19.5. We can iterate this process to find higher derivatives. Consider the circle $x^2 + y^2 = 1$. Differentiating this gave $2x + 2yy' = 0$, which we solved to get $y' = -\frac{x}{y}$. We can differentiate this to get

$$y'' = -\frac{y \cdot 1 - xy'}{y^2} = -\frac{1}{y} + \frac{x}{y^2} \left(-\frac{x}{y} \right) = -\frac{1}{y} - \frac{x^2}{y^3}.$$

Observe that such computations will usually yield an expression involving y' , which needs to be substituted in order to simplify and obtain the final answer. We could also have proceeded by differentiating both sides of $x + yy' = 0$ to get

$$1 + (y')^2 + yy'' = 0 \quad \Rightarrow \quad y'' = \frac{-1 - (y')^2}{y} = -\frac{1}{y} - \frac{x^2}{y^3}.$$

Our discussion so far has sidestepped an important issue. The graphs associated to the equations $x^2 + y^2 = 1$ and $x^3 + y^3 = xy$ do not satisfy the vertical line test, so they are not graphs of functions; so what do we mean when we write $y = y(x)$?

The resolution of this is that if (a, b) is a point on the graph of one of these equations, then in most cases there is a small piece of the graph near (a, b) that *is* the graph of a function. For the circle, if $b > 0$ then for $x \approx a$ we can write $y = \sqrt{1 - x^2}$, while if $b < 0$ we can write $y = -\sqrt{1 - x^2}$. The exception comes at the points $(\pm 1, 0)$, for which $b = 0$; there is no neighborhood of these points on which the circle is the graph of a function $y = y(x)$. Whenever we use implicit differentiation, we must restrict ourselves to points near which y can be written as a function of x ; we also need to know that $y(x)$ is differentiable at these points. The theoretical tool for determining which points satisfy these criteria is the following theorem; this theorem really belongs

to multivariable calculus, so our description here is just to provide some intuition rather than to give a complete justification. In particular, this theorem highlights one of the recurring themes of calculus, which is the power of linear approximations.

Theorem 19.6 (Implicit Function Theorem). *Let $F(x, y)$ be a continuously differentiable²⁴ real-valued function of two variables, and suppose that $a, b, c \in \mathbb{R}$ are such that $F(a, b) = c$. Suppose that the function $g(y) := F(a, y)$ has the property that $g'(b) \neq 0$. Then there exists $\epsilon > 0$ and a differentiable function $f: (a - \epsilon, a + \epsilon) \rightarrow \mathbb{R}$ such that given $x \in (a - \epsilon, a + \epsilon)$, $y = f(x)$ is the only solution of $F(x, y) = c$ that is near b .*

Idea of proof. The idea is to use the fact that near (a, b) , the continuously differentiable function F has the linear approximation $F(x, y) \approx F(a, b) + \ell(x - a) + m(y - b)$, where $\ell, m \in \mathbb{R}$ are the *partial derivatives* that represent how sensitive $F(x, y)$ is to changes in x and y , respectively. In particular, $m = g'(b) \neq 0$, so the equation $F(x, y) = c$ is very close to $F(a, b) + \ell(x - a) + m(y - b) = c$. Since $F(a, b) = c$, the solutions of the latter equation form a line $y - b = -\frac{\ell}{m}(x - a)$; this is where we use the assumption that $m \neq 0$. Then one chooses $\epsilon > 0$ small enough that the linear approximation is accurate enough to give a function $f(x) \approx b - \frac{\ell}{m}(x - a)$ with the desired property. \square

Example 19.7. With $F(x, y) = x^2 + y^2$ and $c = 1$, at a point (a, b) on the unit circle we have $g(y) = a^2 + y^2$, so $g'(y) = 2y$. Thus we have $g'(b) = 0$ if and only if $b = 0$. The points on the circle with $b = 0$ are $(\pm 1, 0)$, which are exactly the points near which the circle *cannot* be represented as the graph of a function $y = f(x)$. Near every other point on the circle, the theorem applies.

Example 19.8. With $F(x, y) = x^3 + y^3 - xy$, for a given a we have $g(y) = a^3 + y^3 - ay$, so $g'(y) = 3y^2 - a$.

Lecture 20

Inverse functions

DATE: WEDNESDAY, OCTOBER 9

This lecture corresponds to §§3.5–3.6 in Stewart and Chapter 12 in Spivak.

We have encountered several functions (e^x , $\sin x$, $\cos x$) whose inverses ($\ln x$, $\arcsin x$, $\arccos x$) are important functions in their own right. In particular, it would be helpful to be able to compute derivatives of the inverse functions in terms of the original functions. Implicit differentiation suggests a solution: if f^{-1} is differentiable at $b = f(a)$ and f is differentiable at a , then differentiating both sides of $y = f(f^{-1}(y))$ with respect to y (using the chain rule) and evaluating at $y = b$ gives

$$1 = (f^{-1})'(b)f'(a) \quad \Rightarrow \quad (f^{-1})'(b) = \frac{1}{f'(a)},$$

²⁴A precise definition of “continuously differentiable” for functions of two variables involves more technicalities than are appropriate for this course. Basically it means that near (a, b) , there is a good linear approximation $F(x, y) \approx F(a, b) + \ell(x - a) + m(y - b)$ for some $\ell, m \in \mathbb{R}$. Most ‘nice’ functions you encounter will have this property.

which can be rewritten as

$$(20.1) \quad (f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}.$$

There is one caveat. As with our previous applications of implicit differentiation, in order for this to be valid we need to know that f^{-1} is differentiable at $b = f(a)$. This can be guaranteed by the Implicit Function Theorem: we want $x = f^{-1}(y)$ to satisfy $f(x) - y = 0$, so we take $F(x, y) = f(x) - y$. Now the Implicit Function Theorem says that x is a differentiable function of y near the values $y = b$, $x = a$ as long as the derivative $\frac{d}{dx}(f(x) - y)|_{x=a}$ is nonzero.²⁵ This derivative is $f'(a)$, and so we see that f^{-1} is differentiable at $b = f(a)$ whenever $f'(a) \neq 0$, and in this case $(f^{-1})'(b) = 1/f'(a)$.

This formula is easy to remember if we write it in the form

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}.$$

Once again, though, we emphasize that $\frac{dy}{dx}$ and $\frac{dx}{dy}$ are not fractions, so this relationship is a theorem that makes our notation reasonable, rather than a simple consequence of how fractions work.

20.1. Inverse trigonometric functions

Since we proved that $\frac{d}{dx} \sin x = \cos x$, it follows from the formula for derivatives of inverse functions that

$$\frac{d}{dx} \sin^{-1} x = \frac{1}{\cos(\sin^{-1} x)}.$$

This can be simplified further by observing that if $t = \sin^{-1} x$, then $-\frac{\pi}{2} \leq t \leq \frac{\pi}{2}$ and $\sin t = x$, so $\cos t = \sqrt{1 - \sin^2 t} = \sqrt{1 - x^2}$, and we deduce that

$$(20.2) \quad \frac{d}{dx} \sin^{-1} x = \frac{1}{\sqrt{1 - x^2}}.$$

Exercise 20.1. Mimic the above computation to prove that $\frac{d}{dx} \cos^{-1} x = -1/\sqrt{1 - x^2}$.

We can also use implicit differentiation directly: for example, if $y = \tan^{-1} x$, then $x = \tan y(x)$, and differentiating both sides with respect to x gives

$$1 = (\sec^2 y) \frac{dy}{dx} \quad \Rightarrow \quad \frac{dy}{dx} = \frac{1}{\sec^2 y} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2}.$$

20.2. Logarithms

Given $a > 0$, the function $y = \log_a x$ is the inverse of $x = a^y$, and differentiating gives

$$1 = (\ln a) a^y \frac{dy}{dx} = x \ln a \frac{dy}{dx},$$

²⁵The hypothesis in the Implicit Function Theorem placed a requirement on the derivative with respect to y . But that was because the theorem was set up to find y as a function of x . Here we are doing the reverse, and the general rule is that we need to check that the derivative *with respect to the desired dependent variable* is nonzero.

so we conclude that

$$(20.3) \quad \frac{d}{dx} \log_a x = \frac{1}{x \ln a},$$

and in particular,

$$(20.4) \quad \frac{d}{dx} \ln x = \frac{1}{x}.$$

This has many important consequences. For example, we can now prove the power rule stated in Theorem 14.7: given any $\beta \in \mathbb{R}$, the function $f: (0, \infty) \rightarrow \mathbb{R}$ defined by $f(x) = x^\beta$ can be rewritten as $f(x) = e^{\beta \ln x}$, and then the chain rule together with the rules for differentiating exponentials and logarithms gives

$$f'(x) = \left(\frac{d}{dx} (\beta \ln x) \right) e^{\beta \ln x} = \frac{\beta}{x} e^{\beta \ln x} = \frac{\beta}{x} x^\beta = \beta x^{\beta-1}.$$

Example 20.2. If g is a differentiable positive function, then the chain rule gives

$$(20.5) \quad \frac{d}{dx} \ln g(x) = \frac{g'(x)}{g(x)}.$$

The expression on the right-hand side of (20.5) is called the *logarithmic derivative* of g .

Example 20.3. If $y = \ln(x^2 + 3x + 4)$, then

$$\frac{dy}{dx} = \frac{\frac{d}{dx}(x^2 + 3x + 4)}{x^2 + 3x + 4} = \frac{2x + 3}{x^2 + 3x + 4}.$$

Example 20.4.

$$\frac{d}{dx} \ln(\cos x) = \frac{-\sin x}{\cos x} = -\tan x.$$

Example 20.5. The function $g: (-\infty, 0) \rightarrow (0, \infty)$ defined by $g(x) = -x = |x|$ has $g'(x) = -1$, so

$$\frac{d}{dx} \ln g(x) = \frac{g'(x)}{g(x)} = \frac{-1}{-x} = \frac{1}{x},$$

which is the same formula as for $\frac{d}{dx} \ln x$ when $x > 0$, and we conclude that

$$(20.6) \quad \frac{d}{dx} \ln |x| = \frac{1}{x} \text{ for all } x \neq 0.$$

As with previous methods we introduced, we now have multiple options for how to evaluate certain derivatives.

Example 20.6. Logarithmic differentiation and the quotient rule give

$$\frac{d}{dx} \ln \left(\frac{x+1}{x-1} \right) = \frac{1}{\frac{x+1}{x-1}} \cdot \left(\frac{(x-1) - (x+1)}{(x-1)^2} \right) = \frac{-2}{(x+1)(x-1)} = \frac{2}{1-x^2}.$$

We could also compute this by using properties of the logarithm before taking the derivative:

$$\frac{d}{dx} \ln \left(\frac{x+1}{x-1} \right) = \frac{d}{dx} (\ln(x+1) - \ln(x-1)) = \frac{1}{x+1} - \frac{1}{x-1} = \frac{(x-1) - (x+1)}{(x+1)(x-1)} = \frac{2}{1-x^2}.$$

Even when the original problem does not include a logarithm, it is sometimes easier to evaluate the derivative by using logarithms instead of the potentially messier product and quotient rules.

Example 20.7. If $y = x^{2/3}\sqrt{x^2 + 2}/(2x + 1)^3$, then we can evaluate $\frac{dy}{dx}$ by observing that

$$\ln y = \frac{2}{3} \ln x + \frac{1}{2} \ln(x^2 + 1) - 3 \ln(2x + 1),$$

and differentiating gives

$$\frac{1}{y} \frac{dy}{dx} = \frac{2}{3x} + \frac{x}{x^2 + 1} - \frac{6}{2x + 1},$$

which we solve to get

$$\frac{dy}{dx} = \frac{x^{2/3}\sqrt{x^2 + 2}}{(2x + 1)^3} \left(\frac{2}{3x} + \frac{x}{x^2 + 1} - \frac{6}{2x + 1} \right).$$

We can also use logarithmic differentiation to address expressions such as $F(x) = f(x)^{g(x)}$, where both the base and the power are functions of x , so neither the power rule nor exponential differentiation are sufficient on their own. By taking logarithms we get

$$\ln F(x) = g(x) \ln f(x) \quad \Rightarrow \quad \frac{F'(x)}{F(x)} = g'(x) \ln f(x) + g(x) \frac{f'(x)}{f(x)},$$

and thus

$$(20.7) \quad F'(x) = g'(x) f(x)^{g(x)} \ln f(x) + g(x) f'(x) f(x)^{g(x)-1}.$$

Note that the first term behaves like an exponential derivative (the original expression is still there, multiplied by the rate of change in the exponent and by the natural logarithm of the base), while the second behaves like the power rule (the power gets decreased by 1 and we multiply by the original power, as well as by a factor that comes from the chain rule). Indeed, if $g(x) = \beta$ in (20.7) is constant, then the equation reduces to the rule

$$\frac{d}{dx} f(x)^\beta = \beta f'(x) f(x)^{\beta-1},$$

and if $f(x) = a$ is constant, then we get

$$\frac{d}{dx} a^{g(x)} = (\ln a) g'(x) a^{g(x)}.$$

We can also use the formula for $\frac{d}{dx} \ln x$ to recover an alternate expression for the natural logarithmic base e . On the one hand, $f(x) = \ln x$ has $f'(1) = \frac{1}{1} = 1$. On the other hand,

$$f'(1) = \lim_{x \rightarrow 0} \frac{f(1+x) - f(1)}{x} = \lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = \lim_{x \rightarrow 0} \ln((1+x)^{1/x}).$$

Since the exponential function is continuous, we get

$$e = e^1 = e^{f'(1)} = \lim_{x \rightarrow 0} e^{\ln((1+x)^{1/x})} = \lim_{x \rightarrow 0} (1+x)^{1/x}.$$

In particular, since $\frac{1}{n} \rightarrow 0$ as $n \rightarrow \infty$, this proves that

$$(20.8) \quad e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} \right)^n.$$

Lecture 21**Rates of change in sciences**

DATE: FRIDAY, OCTOBER 11

This lecture corresponds to §3.7 in Stewart

In this lecture we briefly discuss various places in physics, chemistry, and biology where rates of change, and hence derivatives, naturally appear. Given more time, one could extend this discussion further to include other sciences and many more applications.

21.1. Physics21.1.1. *Velocity and acceleration*

Consider an object moving along a straight line. If $s = s(t)$ represents the *position* of the object at time t , then the rate of change of the object's position is its *velocity* $v = \frac{ds}{dt} = s'$. This is also sometimes denoted \dot{s} . As we discussed when we first defined the derivative, the average velocity from time t_1 to time t_2 is $\frac{s(t_2) - s(t_1)}{t_2 - t_1}$, and the (instantaneous) velocity v is the limit of this quantity as $t_2 \rightarrow t_1$.

Differentiating again gives the rate at which velocity is changing, which is the *acceleration* $a = \frac{dv}{dt} = \frac{d^2s}{dt^2} = v' = s'' = \dot{v} = \ddot{s}$.

21.1.2. *Density*

Consider a straight rod whose density may vary from point to point. Given a point on the rod, let x denote the distance from the beginning of the rod to this point, and let $m(x)$ denote the mass of that part of the rod. Then the mass of the part of the rod lying between points x_1 and x_2 is $m(x_2) - m(x_1)$, and the *average density* of this part of the rod is $\frac{m(x_2) - m(x_1)}{x_2 - x_1}$. Taking a limit as $x_2 \rightarrow x_1$ gives the *linear density* $\rho = \rho(x) := \frac{dm}{dx} = m'(x)$.

21.1.3. *Electricity*

Consider an electric current passing through a wire. Fix a cross-section of the wire, and let $Q = Q(t)$ be the total amount of electrical charge that has passed through that cross-section by time t . Then $\frac{\Delta Q}{\Delta t} = \frac{Q(t_2) - Q(t_1)}{t_2 - t_1}$ gives the *average current* between time t_1 and time t_2 , and $I = I(t) = \frac{dQ}{dt}$ gives the *current* at time t .

21.2. Chemistry21.2.1. *Rate of reaction*

Suppose we have a chemical reaction in which two substances A and B combine to form a third substance C . Let $[A]$, $[B]$, $[C]$ denote the *concentrations* of these substances at a given time t ; this is measured in moles (a unit for number of molecules) per liter. The *average rate of reaction of C* is $\frac{\Delta[C]}{\Delta t} = \frac{[C](t_2) - [C](t_1)}{t_2 - t_1}$. Taking the limit as $\Delta t \rightarrow 0$ gives the *instantaneous rate of reaction of C*, $\frac{d[C]}{dt}$; this is the rate at which C is being produced. The rates of reaction for A and B are defined similarly.

Suppose that 2 molecules of A and 1 molecules of B are required to produce each molecules of C . Then we write the reaction as $2A + B \rightarrow C$, and we have $\frac{d[A]}{dt} = -2\frac{d[C]}{dt}$ and $\frac{d[B]}{dt} = -\frac{d[C]}{dt}$.

21.2.2. Compressibility

Let V denote the volume of a certain quantity of gas, and P the pressure at which the gas is held. Then V is determined by P , and in particular we can consider the rate of change of V with respect to P , which is given by the derivative $\frac{dV}{dP}$. It is reasonable to expect that increasing the pressure will cause the volume to decrease by an amount proportional to V ; if a given pressure increase causes 10 liters of gas to compress by 1 liter, then we expect that 20 liters would compress by 2 liters under the same pressure increase. Thus the relevant quantity is the rate of change of V per unit V , and with this in mind we define the *isothermal*²⁶ *compressibility* of the gas to be $\beta := -\frac{1}{V} \frac{dV}{dP}$. Note that this is equal to $-\frac{d}{dP} \log V$. The negative sign appears because we measure ‘compressibility’, which is the rate at which the volume compresses (decreases) as pressure increases.

21.3. Biology

21.3.1. Blood flow

Under certain assumptions,²⁷ a fluid flowing through a cylindrical tube, such as blood flowing through a vein, flows according to *Poiseuille’s law of laminar flow*, which says that if R is the radius of the tube, η the viscosity of the fluid, P the difference in pressure between the two ends of the tube, and ℓ the length of the tube, then a particle of fluid at a distance r from the center of the tube flows with velocity

$$(21.1) \quad v = \frac{P}{4\eta\ell}(R^2 - r^2).$$

To use this equation to determine the overall rate at which fluid is flowing through a given cross-section, we would need to use the theory of *integration*, which we will develop near the end of the course. For the time being we point out that the sensitivity of velocity to changes in the distance from the center is given by the *velocity gradient* $\frac{dv}{dr} = -\frac{Pr}{2\eta\ell}$.

21.3.2. Population growth

Consider a population that changes over time, with $n = n(t)$ giving the size of the population at time t . Technically n should only take integer values, since the number of individuals in a population is always a whole number; however, when the population is large enough, it is useful to approximate the true population with a differentiable function of time, to which the tools of calculus can be applied. The average rate of growth $\frac{\Delta n}{\Delta t}$ makes sense in either case, but if we allow n to take non-integer values then it is reasonable to also consider the instantaneous rate of growth $\frac{dn}{dt}$.

²⁶‘Isothermal’ refers to the fact that we are considering a gas held at a constant temperature.

²⁷The fluid should be incompressible and Newtonian, the flow should be laminar rather than turbulent, and the tube should be substantially longer than it is wide.

As an important example, suppose we have a colony of bacteria that reproduces at such a rate that its population doubles every hour. Writing n_0 for the initial population at time 0, and $n(t)$ for the population at time t , we see that

$$n(1) = 2n_0, \quad n(2) = 2^2n_0, \quad n(3) = 2^3n_0, \quad \dots \quad n(k) = 2^kn_0.$$

If we use the same formula for times t that are not whole multiples of an hour, we see that the population formula $n(t) = 2^tn_0$ satisfies the given growth condition. With this formula, the rate of growth of the population is

$$\frac{dn}{dt} = (\ln 2)2^tn_0 = (\ln 2)n(t).$$

This situation, where the rate of growth of a quantity is proportional to the quantity itself, leads to exponential growth or decay, and we will examine this phenomenon in more detail shortly. For the moment we observe that if this were to really happen, and if we were to begin with just a single bacteria with mass $\approx 10^{-12}$ g, then after 50 hours (a little over 2 days) the total population would be $2^{50} \approx (10^3)^5 = 10^{15}$ (here we use the approximation $2^{10} = 1024 \approx 10^3$) and thus the total mass would be $10^{15} \cdot 10^{-12}$ g = 10^3 g = 1 kg. After 80 more hours (130 hours total), the mass would be $2^{80} = (2^{10})^8 \approx (10^3)^8 = 10^{24}$ kg. The mass of the earth is $\approx 6 \times 10^{24}$ kg, so after 133 hours of exponential growth the total mass of the bacteria ($\approx 8 \times 10^{24}$ kg) would exceed that of the earth. Clearly this does not happen in practice, which illustrates the limitations of the exponential growth model; it is valid only as long as there are sufficient resources to permit unconstrained growth. Next semester when we study differential equations we will examine some more realistic models.

Lecture 22

Exponential growth and decay

DATE: MONDAY, OCTOBER 14

This lecture corresponds to §3.8 in Stewart (related rates are in §3.9)

22.1. Solving a differential equation

As we saw in the previous lecture, the simplest model for population growth leads to the *differential equation*

$$(22.1) \quad \frac{dy}{dt} = ky,$$

where $y = y(t)$ is the population at time t , and $k > 0$ is a constant parameter that determines the rate of growth. We know that $y = e^{kt}$ is a solution, and so is $y = Ce^{kt}$ for every $C \geq 0$, since

$$\frac{d}{dt}(Ce^{kt}) = C \frac{d}{dt}e^{kt} = Cke^{kt}.$$

Note that $y(0) = Ce^{k \cdot 0} = Ce^0 = C$, so, we can write this as

$$(22.2) \quad y(t) = y(0)e^{kt}.$$

There are two natural questions to ask.

- (1) Suppose we didn't know in advance that (22.2) is a solution of (22.1). How could we find the solution?
- (2) Is this the only solution? Or could there be another function $y(t)$ that also satisfies (22.1) but is not given by (22.2)?

To address the first question, we can divide both sides of (22.1) by y and obtain $y'/y = k$; since the left-hand side is the logarithmic derivative of y , we get

$$\frac{d}{dt} \ln y = k.$$

This is easier to solve: indeed, we know that given any $\ell \in \mathbb{R}$, the function $f(t) = kt + \ell$ has the property that $f'(t) = k$ for every t . Thus $\ln y = kt + \ell$ gives a solution of (22.1), and exponentiating both sides gives $y(t) = e^{kt} e^\ell$. But is it the only solution? Note that if $f'(t) = k$ for every t , then $g(t) := f(t) - kt$ has the property that $g'(t) = f'(t) - k = 0$ for every t . Our intuition strongly suggests that in order to have zero derivative at every point, g must be a constant function. And indeed, this is true.

Theorem 22.1. *If $g: (a, b) \rightarrow \mathbb{R}$ is differentiable and $g'(x) = 0$ for every $x \in (a, b)$, then g is constant on (a, b) .*

Remark 22.2. It is easiest to prove Theorem 22.1 as a corollary of the Mean Value Theorem, which we will prove later on. It is possible to prove it now using bisection sequences, and we include this proof here for completeness (this proof was omitted in the classroom lecture).

Proof of Theorem 22.1. We start by proving that

$$(22.3) \quad \text{for every } \varepsilon > 0, \text{ we have } |g(x) - g(y)| \leq \varepsilon|x - y| \text{ for every } x, y \in (a, b).$$

This implies the conclusion of the theorem because if $x, y \in (a, b)$ were such that $g(x) \neq g(y)$, then taking $\varepsilon = \frac{1}{2} \frac{|g(x) - g(y)|}{|x - y|}$ would give $|g(x) - g(y)| = 2\varepsilon|x - y|$, contradicting (22.3). To prove (22.3), fix $\varepsilon > 0$ and $a < x < y < b$. (The case $y > x$ is similar, and $y = x$ is automatic.) Aiming for a contradiction, suppose that $|g(x) - g(y)| > \varepsilon|x - y|$. Let m be the midpoint of x and y and observe that $|g(x) - g(y)| \leq |g(x) - g(m)| + |g(m) - g(y)|$; moreover, $|x - m| = |m - y| = \frac{|x - y|}{2}$, so

$$\varepsilon < \frac{|g(x) - g(y)|}{|x - y|} \leq \frac{1}{2} \left(\frac{|g(x) - g(m)|}{|x - m|} + \frac{|g(m) - g(y)|}{|m - y|} \right).$$

It follows that at least one of $\frac{|g(x) - g(m)|}{|x - m|}$ and $\frac{|g(m) - g(y)|}{|m - y|}$ must exceed ε . Replacing the interval (x, y) with whichever of the intervals (x, m) or (m, y) has this property, we can iterate this argument to produce two bisection sequences; more precisely, we write $b_0 = x$, $r_0 = y$, and observe that if $|g(r_n) - g(b_n)| > \varepsilon|r_n - b_n|$, then we can choose $b_{n+1}, r_{n+1} \in [b_n, r_n]$ such that

- one of b_{n+1}, r_{n+1} is the midpoint of $[b_n, r_n]$, and the other is at an endpoint of that interval;
- $|g(r_{n+1}) - g(b_{n+1})| > \varepsilon|r_{n+1} - b_{n+1}|$.

As in our previous proofs with bisection sequences, the sequences b_n and r_n converge to a common limit, which we denote by

$$c = \lim_{n \rightarrow \infty} b_n = \lim_{n \rightarrow \infty} r_n.$$

For every n , we have

$$|g(b_n) - g(r_n)| \leq |g(b_n) - g(c)| + |g(c) - g(r_n)|$$

and

$$|b_n - r_n| \geq |b_n - c|, \quad |b_n - r_n| \geq |c - r_n|;$$

dividing gives

$$\varepsilon < \frac{|g(b_n) - g(r_n)|}{|b_n - r_n|} \leq \frac{|g(b_n) - g(c)|}{|b_n - c|} + \frac{|g(c) - g(r_n)|}{|c - r_n|},$$

and thus there is $x_n \in \{b_n, r_n\}$ such that $|g(x_n) - g(c)|/|x_n - c| > \varepsilon/2$. Sending $n \rightarrow \infty$ we get

$$g'(c) = \lim_{n \rightarrow \infty} \frac{|g(x_n) - g(c)|}{|x_n - c|} \geq \frac{\varepsilon}{2} > 0,$$

contradicting the assumption that g' vanishes on the entire interval. This proves (22.3), and by the argument given at the start of the proof, this is enough to prove Theorem 22.1. \square

Returning to the question of solutions of (22.1), we see that every solution $y = y(t)$ has the property that $g(t) = \ln(y(t)) - kt$ has zero derivative at all t . By Theorem 22.1, this implies that $\ln(y(t)) - kt = g(t) = g(0) = \ln(y(0))$ for all t , and exponentiating gives $y(t)e^{-kt} = y(0)$, demonstrating that (22.2) is the only solution of (22.1).

22.2. Carbon dating

The differential equation (22.1) and its solution (22.2) also arise in the process of *radioactive decay*. A carbon atom has 6 protons and 6 electrons, and can have either 6, 7, or 8 neutrons. Roughly 99% of carbon atoms have 6 neutrons, giving “carbon-12”, and nearly all remaining carbon atoms have 7 neutrons, giving “carbon-13”. Roughly one in every trillion carbon atoms has 8 neutrons, giving “carbon-14”. The ratio is so low because this is an unstable isotope; given a sample of carbon with N_0 atoms of carbon-14, after $\sim 5,730$ years about half of these atoms will have undergone radioactive decay and been transformed into nitrogen, so that the number of carbon-14 atoms remaining is $\frac{1}{2}N_0$. Extrapolating, we see that the number $N(t)$ of carbon-14 atoms after time t is

$$(22.4) \quad N(t) = \left(\frac{1}{2}\right)^{\frac{t}{5730}} N_0 = e^{(\frac{-\ln 2}{5730})t} N_0,$$

so that $N(t)$ obeys (22.1) and (22.2) with $k = -(\ln 2)/5730$. This is the calculation that leads to the process of “carbon dating”: the proportion of carbon-14 in the atmosphere is relatively stable, because the exponential decay is balanced out by the production of carbon-14 in the upper atmosphere via cosmic rays, and thus as long as a plant or animal is alive, it exchanges carbon with its environment so that the proportion of carbon-14 in its body is stable. Once it dies, however, it no longer takes in new carbon-14 atoms, and the decay process begins. Thus if we wish to find the age of a particular sample of organic matter, we can measure its current carbon-14 content to determine $N(t)$, and

then since N_0 is known (via the background carbon-14 level in the atmosphere), we can take logs in (22.4) and solve for t , obtaining

$$\ln N(t) = kt + \ln N_0 \quad \Rightarrow \quad t = \frac{1}{k}(\ln N(t) - \ln N_0).$$

22.3. Newton's law of cooling

Suppose that an object has temperature $T(t)$ at time t , and that its surroundings have constant temperature T_s . Newton's law of cooling states that

$$\frac{dT}{dt} = k(T - T_s),$$

where k is a constant that depends on physical properties of the object and its surroundings. Note that $k < 0$ since we expect the object's temperature to decrease when $T > T_s$. Writing $y = T - T_s$, we get $y' = \frac{dT}{dt} = ky$, so y is a solution of (22.1), and thus (22.2) gives

$$T(t) = T_s + y(t) = T_s + y(0)e^{kt} = T_s + (T(0) - T_s)e^{kt}.$$

Note that this discussion makes some simplifying assumptions.

- We assume that the object is always at a single temperature throughout; of course in reality a spatially extended object can vary in temperature from point to point, and if this variation becomes significant then a more refined model is needed.
- We assume that the object's surroundings remain at a constant temperature even as they absorb the heat from the object (or give heat to the object). This is of course not true in the strictest sense, but as long as the surroundings are sufficiently large relative to the object, it is a reasonable approximation to make. This assumption that the environment functions as a "heat bath" is a common one in the study of thermodynamics.

22.4. Compound interest

Suppose I have a bank account in the amount $A(t)$ at time t . Suppose also that I make no withdrawals or deposits, but that the money grows with annual interest rate r (here $r = 0.03$ would correspond to 3% interest). If interest is applied once per year, then after 1 year the amount is $A(1) = (1 + r)A(0)$, after 2 years it is $A(2) = (1 + r)^2A(0)$, and in general after t years it is

$$A(t) = (1 + r)^t A(0).$$

If interest is applied twice per year, then every 6 months the amount of money in the account is multiplied by $(1 + \frac{r}{2})$, and we get

$$A(t) = \left(1 + \frac{r}{2}\right)^{2t} A(0).$$

In general, if interest is applied n times per year, then the amount is multiplied by $(1 + \frac{r}{n})$ every time interest is applied, and after t years, interest has been applied nt times, so

$$A(t) = \left(1 + \frac{r}{n}\right)^{nt} A(0).$$

Recall from a homework assignment that $\lim_{n \rightarrow \infty} (1 + \frac{r}{n})^n = e^r$; thus in the limit as $n \rightarrow \infty$, we get

$$A(t) = e^{rt} A_0;$$

this is the case of *continuously compounded* interest.

Remark 22.3. The first three equations above have a small problem; they are only valid when nt is an integer, since otherwise the expression $(1 + \frac{r}{n})^{nt}$ includes a pro-rated share of the next term's interest, which has not yet been applied.

22.5. Related rates

It is often the case that we are interested in determining the rate of change of one quantity, while the information we are given is in terms of a different quantity; in this case we need to write down the function that relates the two, and apply the chain rule. For example, suppose we fill a balloon with air at the constant rate of $100 \text{ cm}^3/\text{s}$, and we want to determine the rate at which the radius is increasing when the diameter is 50 cm. Writing $V(t)$ for the volume at time t and $r(t)$ for the radius at time t , we have the following.

$$\text{Given quantity: } \frac{dV}{dt} \quad \text{Desired quantity: } \frac{dr}{dt} \quad \text{Relationship: } V = \frac{4}{3}\pi r^3$$

The last equation, which expresses volume as a function of radius, is valid as long as we assume that the balloon is spherical, and will be proved later on after we have studied integration. For now we simply take it as a given fact from geometry. Applying the chain rule and the power rule, we get

$$\frac{dV}{dt} = \frac{dV}{dr} \frac{dr}{dt} = (4\pi r^2) \frac{dr}{dt},$$

and solving for $\frac{dr}{dt}$ gives

$$\frac{dr}{dt} = \frac{1}{4\pi r^2} \frac{dV}{dt}.$$

When the diameter is 50 cm, the radius is 25 cm, so

$$\left. \frac{dr}{dt} \right|_{r=25} = \frac{1}{4\pi \cdot 625 \text{ cm}^2} \cdot 100 \text{ cm}^3/\text{s} = \frac{1}{25\pi} \text{ cm/s} \approx 0.0127 \text{ cm/s}.$$

Lecture 23

Related rates; linear approximation

DATE: WEDNESDAY, OCTOBER 16

This lecture corresponds to §§3.9–3.10 in Stewart

23.1. More related rates examples

Example 23.1. Suppose I have two rings that can slide along a straight track, which are connected by a string of length 20 m. I pull the middle of the string in a direction perpendicular to the track with a constant velocity v . How fast is the distance between the two rings decreasing when they are 10 m apart?

Solution: Let $x(t)$ denote the distance between the two rings, and let $y(t)$ denote the distance that I have pulled the middle of the string away from the track. Then we have

$$\text{Given quantity: } \frac{dy}{dt} \quad \text{Desired quantity: } \frac{dx}{dt} \quad \text{Relationship: } \left(\frac{x}{2}\right)^2 + y^2 = 10^2$$

where the last equation comes by looking at the right triangle with vertices at my hand, the midpoint of the rings, and one of the two rings. Using implicit differentiation and the chain rule we get

$$0 = \frac{d}{dt}10^2 = \frac{d}{dt}\left(\frac{x^2}{4} + y^2\right) = \frac{x}{2}\frac{dx}{dt} + 2y\frac{dy}{dt},$$

and solving for $\frac{dx}{dt}$ gives

$$\frac{dx}{dt} = -\frac{4y}{x}\frac{dy}{dt}.$$

When the rings are 10 m apart, we have $x = 10$ and $y = \sqrt{10^2 - 5^2} = \sqrt{75} = 5\sqrt{3}$, and we are given that $\frac{dy}{dt}$ takes the constant value v , so we conclude that at this moment we have

$$\frac{dx}{dt} = -\frac{4 \cdot 5\sqrt{3}}{10}v = -2\sqrt{3}v.$$

Thus the distance between the rings is decreasing at an instantaneous rate of $2\sqrt{3}v$ at the moment when they are 10 m apart.

Example 23.2. Consider a conical tank whose top is a circle of radius 2 m and whose height is 5 m. Suppose that water drains from the bottom of the tank at a rate of 10 L/min. How fast is the height of the water decreasing when the tank is half empty?

Step 1: Identify the quantities involved and assign notation. Let V be the volume of water remaining at time t , h the height of the top of the water (relative to the lowest point of the cone), and r the radius of the circle formed by the top of the water.

Step 2: Identify the given information and the desired information in terms of derivatives. We are given that the water is draining at a rate of 10 L/min, so $\frac{dV}{dt} = -10$ L/min. The rate at which the height of the water is changing is $\frac{dh}{dt}$. We want the rate at which it is *decreasing*, so our desired quantity is $-\frac{dh}{dt}$. (If the derivative is negative, then the height is decreasing and this number will be positive.)

Step 3: Write down formulas relating the various quantities. The water is in the shape of a cone with height h whose base is a circle of radius r : recalling a formula from geometry,²⁸ the volume of this cone is $V = \frac{1}{3}\pi r^2 h$. We also need to relate r and h ; their ratio is the same as the ratio of the radius of the tank to its height, so $\frac{r}{h} = \frac{2}{5}$. Writing this as $r = \frac{2}{5}h$, we can relate V and h by

$$(23.1) \quad V = \frac{4}{75}\pi h^3.$$

Step 4: Differentiate using the chain rule. Differentiating (23.1) gives

$$\frac{dV}{dt} = \frac{4}{25}\pi h^2 \frac{dh}{dt}.$$

²⁸When we study integrals later, we will see how to obtain this formula directly.

Step 5: Substitute the given information to obtain the solution. We are given that $\frac{dV}{dt} = -10$ L/min, so

$$(23.2) \quad \frac{dh}{dt} = \frac{25}{4\pi h^2} \frac{dV}{dt} \text{ L/min} = -\frac{250}{4\pi h^2} \text{ L/min}.$$

To obtain the final solution we need to know the value of h at the desired instant. When half the water is gone, the volume of the water is half the volume of the tank, so

$$\underbrace{\frac{4}{75}\pi h^3}_{\text{volume of water}} = \frac{1}{2} \cdot \underbrace{\frac{4}{75}\pi 5^3}_{\text{volume of tank}} \Rightarrow h^3 = \frac{5^3}{2} \Rightarrow h = \frac{5}{\sqrt[3]{2}} \text{ m}.$$

Together with (23.2) this gives

$$\frac{dh}{dt} = -\frac{250}{4\pi} \cdot \frac{2^{2/3}}{5^2} \text{ L/min/m}^2 = -\frac{5}{\pi\sqrt[3]{2}} \text{ L/min/m}^2.$$

We are not quite done, because we need to put our answer in the correct units. The volume was given in liters, and we have $1 \text{ L} = 1000 \text{ cm}^3$. Note that $1 \text{ m}^3 = 100^3 \text{ cm}^3 = 10^6 \text{ cm}^3$, so $1 \text{ L} = 10^{-3} \text{ m}^3$. We conclude that at the moment the tank is half full, we have

$$\frac{dh}{dt} = -\frac{5}{\pi\sqrt[3]{2}} \cdot 10^{-3} \text{ m}^3/\text{min/m}^2 = -\frac{1}{200\pi\sqrt[3]{2}} \text{ m/min} = -\frac{1}{2\pi\sqrt[3]{2}} \text{ cm/min}.$$

Thus the water level is falling at a rate of $\frac{1}{2\pi\sqrt[3]{2}} \approx 0.126$ cm/min.

23.2. Linear approximation and differentiating inverse functions

We have already mentioned several times that the derivative of a function gives a linear approximation to that function via the tangent line. This is made precise by the following.

Proposition 23.3. *Consider a function $f: I \rightarrow \mathbb{R}$, where I is an interval. Given a point a in the interior of I , a real number $m \in \mathbb{R}$ is the derivative of f at a if and only if there is a function $r: I \rightarrow \mathbb{R}$ such that*

- $f(x) = f(a) + m(x - a) + r(x)$ for all $x \in I$, and
- $\lim_{x \rightarrow a} \frac{r(x)}{x - a} = 0$.

Proof. If f is differentiable at a , then writing $m := f'(a)$, the function $r(x) := f(x) - (f(a) + f'(a)(x - a))$ satisfies the first condition by definition, and moreover

$$\lim_{x \rightarrow a} \frac{r(x)}{x - a} = \lim_{x \rightarrow a} \frac{f(x) - (f(a) + f'(a)(x - a))}{x - a} = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} - f'(a) = 0$$

by the definition of derivative. Conversely, if there is a function r satisfying the two conditions given, then

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lim_{x \rightarrow a} \frac{m(x - a) + r(x)}{x - a} = \lim_{x \rightarrow a} m + \frac{r(x)}{x - a} = m,$$

and thus $m = f'(a)$. □

Remark 23.4. Later, in multivariable calculus, the characterization of derivative in Proposition 23.3 turns out to be the best way to define derivative for functions of multiple variables. For a function of n variables, $x - a$ is a vector with n components, not a real number, and the derivative is an $n \times n$ matrix.

Remark 23.5. The function $r(x)$ can be thought of as a ‘remainder term’ that captures the difference between $f(x)$ and the linear approximation given by the tangent line at a . We will encounter generalizations of this later on when we study Taylor series. A simple first step in this direction would be to ask for the *quadratic* function that gives the “best fit” to $f(x)$ at the point a . The best linear approximation was the function $L(x) = c_0 + c_1x$ such that $L(a) = f(a)$ and $L'(a) = f'(a)$, so similarly the best quadratic approximation should be the function $Q(x) = c_0 + c_1x + c_2x^2$ such that $Q(a) = f(a)$, $Q'(a) = f'(a)$, and $Q''(a) = f''(a)$. Rewriting as $Q(x) = A + B(x - a) + C(x - a)^2$ we see that

$$Q(a) = A, \quad Q'(x) = B + 2C(x - a) \Rightarrow Q'(a) = B, \quad Q''(x) = 2C.$$

Thus $A = Q(a)$, $B = Q'(a)$, and $C = \frac{1}{2}Q''(a)$, so the best quadratic approximation to $f(x)$ near a is

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

This is the *quadratic Taylor polynomial*.

We can use Proposition 23.3 to study derivatives of inverse functions; in particular, to show that they exist (and thus are given by (20.1)) *without* relying on the Implicit Function Theorem from Lecture 19 (which after all we did not really prove). Suppose that $f: I \rightarrow \mathbb{R}$ is injective, continuous, and that $f'(a)$ exists and is not equal to 0. By Theorem 11.6, f^{-1} is continuous. Given $y \approx b$, let $x = f^{-1}(y)$. By Proposition 23.3, we have

$$y = f(x) = f(a) + f'(a)(x - a) + r(x),$$

and solving for x gives

$$y - f(a) - r(x) = f'(a)(x - a) \quad \Rightarrow \quad f^{-1}(y) = x = a + \frac{y - f(a) - r(x)}{f'(a)}.$$

Since $a = f^{-1}(b)$ and $x = f^{-1}(y)$, we can rewrite this as

$$f^{-1}(y) = f^{-1}(b) + \frac{1}{f'(a)}(y - b) - \underbrace{\frac{r(f^{-1}(y))}{f'(a)}}_{R(y)}.$$

We want to apply Proposition 23.3 by proving that $R(y)/(y - b) \rightarrow 0$ as $y \rightarrow b$. Indeed, since f^{-1} is continuous at b , we have $x \rightarrow a$ as $y \rightarrow b$, and thus

$$\begin{aligned} \lim_{y \rightarrow b} \frac{R(y)}{y - b} &= \lim_{x \rightarrow a} \frac{r(x)}{f'(a)(f(x) - f(a))} = \lim_{x \rightarrow a} \frac{r(x)}{f'(a)(f'(a)(x - a) + r(x))} \\ &= \lim_{x \rightarrow a} \frac{r(x)/(x - a)}{f'(a)^2 + f'(a)r(x)/(x - a)} = \frac{0}{f'(a)^2 + 0} = 0, \end{aligned}$$

where we have used the fact that $f'(a) \neq 0$. This proves that f^{-1} is differentiable at $b = f(a)$, and that

$$(23.3) \quad (f^{-1})'(b) = \frac{1}{f'(a)} = \frac{1}{f'(f^{-1}(b))}.$$

23.3. Using linear approximations

Linear approximations are useful when we only need a rough estimate for a function that is perhaps difficult to compute exactly. For example, if we want a reasonable estimate for $\sqrt{4.1}$, we might observe that $f(x) = \sqrt{x}$ has $f(4) = 2$ and $f'(x) = \frac{1}{2\sqrt{x}}$, so $f'(4) = \frac{1}{4}$; thus the linear approximation to f near 4 is

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

With $x = 4.1$ we get

$$\sqrt{4.1} \approx 2 + \frac{1}{4}(.1) = 2.025.$$

In fact the first few digits of the true value are 2.02485..., so this is a reasonable approximation.

Note that we must be careful in how we use linear approximations. For $x = 4.1$ we got a reasonable approximation. But if we go too far, things break down. For example, at $x = 9$ we have $\sqrt{9} = 3$, but the linear approximation gives $2 + \frac{1}{4}(9 - 4) = 2 + \frac{5}{4} = 3.25$, which is not particularly close. It is a very important problem to understand how the error term between the linear approximation and the true value of the function can be controlled, and we will return to this near the end of next semester when we study Taylor polynomials.

Lecture 24

Hyperbolic functions

DATE: FRIDAY, OCTOBER 18

This lecture corresponds to §3.11 in Stewart

The standard trigonometric functions are defined by considering the unit circle $x^2 + y^2 = 1$, and writing $\cos t$ and $\sin t$ for the x - and y -coordinates of the point on the circle obtained by moving a distance t counterclockwise from $(1, 0)$. Another useful way of describing this point is that if $\gamma(s)$ is a parametrization of the unit circle that moves counterclockwise and satisfies $\gamma(0) = 1$, and if $L(s)$ is the length of γ from 0 to s , then $(\cos t, \sin t) = \gamma(L^{-1}(t))$.

To define the hyperbolic functions we start by observing that $\cos t$ and $\sin t$ can also be characterized in terms of area. If we write $A(s)$ for the area swept out by the line from $(0, 0)$ to $\gamma(s)$, as the parameter moves from 0 to s , then $(\cos t, \sin t) = \gamma(A^{-1}(t/2))$. In other words, $(\cos t, \sin t)$ is the point at which this line has swept out an area of $t/2$. Now replace the circle with the hyperbola $x^2 - y^2 = 1$ as shown in Figure 7. Let γ be a parametrization of this hyperbola such that $\gamma(0) = (1, 0)$ and $\gamma(s)$ moves into the first quadrant as s increases. Again writing $A(s)$ for the area swept out by the line from

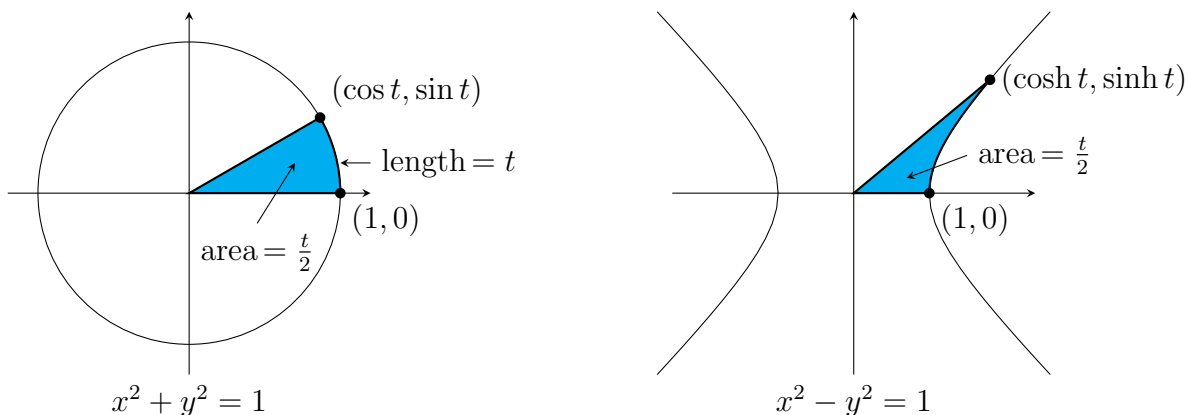


FIGURE 7. The circle and the hyperbola.

$(0, 0)$ to γ as the parameter moves from 0 to s , we define $(\cosh t, \sinh t) = \gamma(A^{-1}(t/2))$, so that $(\cosh t, \sinh t)$ is the point on the hyperbola at which this line has swept out an area of $t/2$.

When we first discussed arc length, we observed that it needs to be defined using some kind of limiting procedure. A similar statement is true for area; we know what is meant by “area of a rectangle” – length times height – but what is meant by “area of the region swept out by such-and-such a line”? To make this properly precise requires integration, and so we defer it until later; for the time being we merely observe that it is once again a limiting procedure, in which the region whose area we wish to determine is approximated by simpler regions (rectangles) whose area we know. Once we have developed the theory of integration, we will see that \cosh and \sinh admit the following formulas:

$$(24.1) \quad \cosh(t) = \frac{e^t + e^{-t}}{2}, \quad \sinh(t) = \frac{e^t - e^{-t}}{2}.$$

Compare these to the formulas using $e^{\pm it}$ for \cos and \sin .

Because $(\cosh t, \sinh t)$ lies on the hyperbola $x^2 - y^2 = 1$, we obtain the fundamental identity

$$(24.2) \quad \cosh^2 t - \sinh^2 t = 1 \text{ for all } t \in \mathbb{R},$$

which is analogous to the identity $\cos^2 t + \sin^2 t = 1$. Note that this identity can also be deduced directly from (24.1). We also see from (24.1), or from the geometric description, that \sinh is an odd function, while \cosh is even.

Example 24.1. Hyperbolic functions arise naturally in various applications. For example, if a cable is supported at its endpoints and hangs between them under its own weight (such as a power line, a suspension bridge, etc.), then it takes the shape of a *catenary*, which is a curve given as the graph of the function $y = c + a \cosh(x/a)$, where a, c are parameters determined by the physical characteristics of the situation. (We will prove this next semester.) Another example comes if we consider a wave of wavelength L propagating in water of depth d ; the velocity of the wave is $\sqrt{\frac{gL}{2\pi} \tanh(\frac{2\pi d}{L})}$, where $g = 9.8 \text{ m/s}^2$ is the force of gravity, and $\tanh t = \sinh t / \cosh t$.

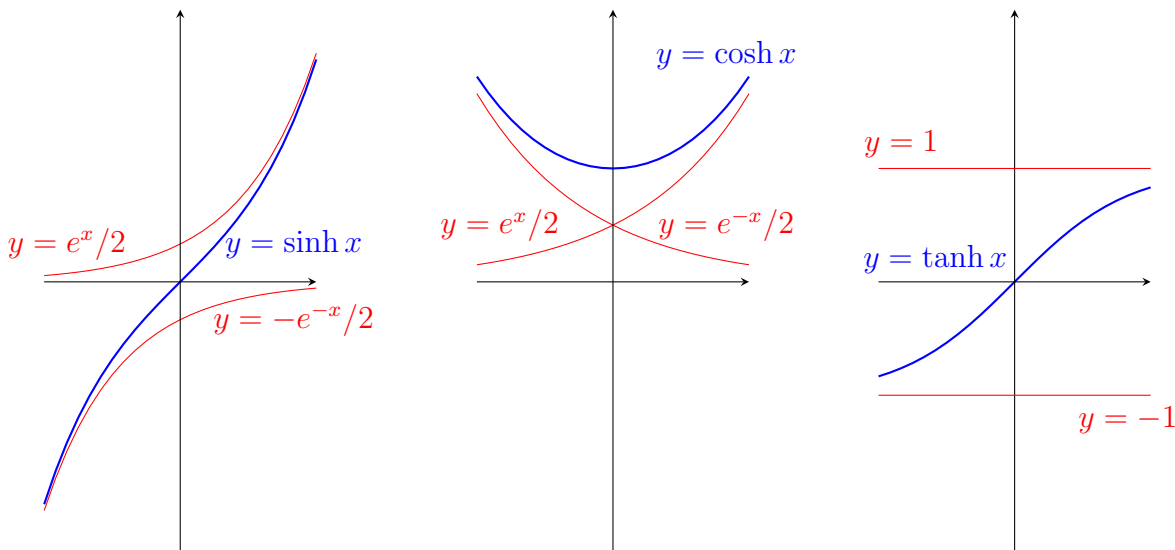


FIGURE 8. Hyperbolic functions.

The graphs of the functions \cosh , \sinh , and \tanh are shown in Figure 8; the shapes of these graphs are relatively easy to deduce from (24.1). Note that the slope of \sinh is positive at all points, and is equal to 1 as it goes through the origin; this is because

$$\frac{d}{dt} \sinh(t) = \frac{d}{dt} \frac{e^t - e^{-t}}{2} = \frac{e^t + e^{-t}}{2} = \cosh(t),$$

and $\cosh(0) = 1$. A similar observation applies to $\tanh(t)$, which we can justify either by differentiating $\tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}}$, or more cleanly by observing that the linear approximations to \sinh and \cosh near $t = 0$ are given by

$$\sinh(t) \approx t, \quad \cosh(t) \approx 1,$$

and thus $\tanh(t) \approx t$ for $t \approx 0$. Note also that

$$\frac{d}{dt} \cosh(t) = \frac{d}{dt} \frac{e^t + e^{-t}}{2} = \frac{e^t - e^{-t}}{2} = \sinh(t),$$

so that once again we recover equations very similar to the ones we are familiar with from trigonometric functions, but with some interchange of positive and negative signs.

One difference between trigonometric functions and hyperbolic functions is that the latter can be explicitly inverted. Indeed, if $y = \cosh^{-1}(x)$, where we use the branch $y \geq 0$ – note that \cosh is invertible on $[0, \infty)$ and on $(-\infty, 0]$ – then we have

$$x = \cosh y = \frac{e^y + e^{-y}}{2} \Rightarrow e^y - 2x + e^{-y} = 0 \Rightarrow e^{2y} - 2xe^y + 1 = 0,$$

and the quadratic formula gives

$$e^y = x \pm \sqrt{x^2 - 1}.$$

For $y \geq 0$ we have $e^y \geq 1$, so we use the sum instead of the difference, and get

$$(24.3) \quad \cosh^{-1} x = y = \ln(x + \sqrt{x^2 - 1}).$$

Differentiating gives

$$\frac{d}{dx} \cosh^{-1} x = \frac{1}{x + \sqrt{x^2 - 1}} \left(1 + \frac{2x}{2\sqrt{x^2 - 1}} \right) = \frac{1}{x + \sqrt{x^2 - 1}} \frac{\sqrt{x^2 - 1} + x}{\sqrt{x^2 - 1}} = \frac{1}{\sqrt{x^2 - 1}}.$$

Alternately we could obtain this via implicit differentiation:

$$y = \cosh^{-1} x \Rightarrow \cosh y = x \Rightarrow (\sinh y) \frac{dy}{dx} = 1 \Rightarrow \frac{d}{dx} \cosh^{-1} x = \frac{1}{\sinh x} = \frac{1}{\sqrt{x^2 - 1}}.$$

The class on Monday, October 21 will be a review session for Test 2.

Lecture 25

The Extreme Value Theorem

DATE: WEDNESDAY, OCTOBER 23

Stewart §4.1, Spivak Chapter 11

Now we start to look at applications of differentiation.

Definition 25.1. A function $f: D \rightarrow \mathbb{R}$ has an *absolute maximum* (or *global maximum*) at $c \in D$ if $f(c) \geq f(x)$ for all $x \in D$. The value $f(c)$ is called the *maximum value* of f .

Similarly, if $c \in D$ is such that $f(c) \leq f(x)$ for all $x \in D$, then f has an *absolute minimum* (or *global minimum*) at c , and $f(c)$ is called the *minimum value* of f .

The maximum and minimum values of a function are called its *extreme values*.

Many problems in science and engineering can be formulated as the search for the extreme points and/or values of some function; these are sometimes referred to as *optimization problems*.

First observe that global maxima and minima may or may not exist, and may or may not be unique.

Example 25.2. If $f(x) = \sin x$, then $-1 \leq f(x) \leq 1$ for all $x \in \mathbb{R}$. Moreover $f(x) = 1$ for all $x = \frac{\pi}{2} + 2n\pi$, $n \in \mathbb{Z}$, and so each of these points is a global maximum, and 1 is the maximum value. Similarly $f(x) = -1$ for all $x = -\frac{\pi}{2} + 2n\pi$, so each of these is a global minimum, and -1 is the minimum value.

Example 25.3. $f(x) = -x^2$ has a global maximum at $x = 0$ (the maximum value is 0), but no global minimum.

Example 25.4. $f(x) = x^3$ has no global maximum or minimum.

Example 25.5. $f(x) = \frac{1}{x}$ has no global maximum or minimum.

Example 25.6. $f(x) = \frac{1}{x^2+1}$ has a global maximum at 0 (the maximum value is 1), but no global minimum.

Note that the last two examples are *bounded below* in the sense that there is $M \in \mathbb{R}$ such that $f(x) \geq M$ for every x – such an M is called a *lower bound* for f – but because of their asymptotic behavior they never achieve a lower bound and so do not have global minima. This sort of behavior can be avoided by considering continuous functions on closed and bounded intervals.

Theorem 25.7 (Extreme Value Theorem). *Let $f: [a, b] \rightarrow \mathbb{R}$ be a continuous function. Then f has a global maximum and minimum.*

Before proving the theorem we make a few observations.

- The maximum and minimum need not be unique; for example, if f is constant then every point is both a global maximum and a global minimum.
- The theorem doesn't give any information about how to actually find the maximum and minimum.
- The conclusion can fail if either hypothesis is violated; if $f(x) = x$ for $x \in (0, 1)$, then f has no global maximum or minimum on $(0, 1)$ – here the domain of f is not a closed bounded interval. If we define f like this on $(0, 1)$ and then put $f(0) = f(1) = \frac{1}{2}$, we get a discontinuous function on $[0, 1]$ that has no global maximum or minimum.

Proof of the EVT. It suffices to prove existence of a global maximum; then the result for a global minimum follows by finding a global maximum for $-f$. We break the proof into two halves, both of which we prove using bisection sequences.

- (1) If $f: [a, b] \rightarrow \mathbb{R}$ is continuous, then it is bounded above.
- (2) If $f: [a, b] \rightarrow \mathbb{R}$ is continuous and bounded above, then it has a global maximum.

For the first half, we argue by contradiction. Suppose that f is not bounded above on $[a, b]$, and construct a pair of bisection sequences using the question: “Is f bounded above on the interval $[a, m]$?” Here m is the midpoint of the current points in each sequence. Note that the answer is ‘yes’ for $m = a$ and ‘no’ for $m = b$. Thus we iteratively construct sequences

$$a = r_1 \leq r_2 \leq r_3 \leq \cdots \leq b_3 \leq b_2 \leq b_1 = b$$

with the following properties:

- f is bounded above on each interval $[a, r_n]$;
- f is not bounded above on $[a, b_n]$ for any n ;
- the sequences r_n and b_n converge to a common limit $c \in [a, b]$.

By continuity there is $\delta > 0$ such that for every $x \in [a, b]$ with $|x - c| < \delta$, we have $|f(x) - f(c)| < 1$. In particular, for every $x \in (c - \delta, c + \delta)$, we have $f(x) \leq f(c) + 1$. Now we complete the proof of the first half as follows.

- Since $r_n \rightarrow c$, there is $n \in \mathbb{N}$ such that $r_n \in (c - \delta, c + \delta)$.
- By the definition of the sequences, f is bounded above on $[a, r_n]$, so there is $M \in \mathbb{R}$ such that $f(x) \leq M$ for all $x \in [a, r_n]$.
- For every $x \in [a, c + \delta)$, we either have $x \in [a, r_n]$, in which case $f(x) \leq M$, or $x \in (r_n, c + \delta) \subset (c - \delta, c + \delta)$, in which case $f(x) \leq f(c) + 1$. Thus f is bounded above by $\max(M, f(c) + 1)$ on $[a, c + \delta)$.
- Since $b_n \rightarrow c$, there is $n \in \mathbb{N}$ such that $b_n \in (c - \delta, c + \delta)$, and the previous step shows that f is bounded above on $[a, b_n]$.
- This contradicts the definition of the sequences, and we conclude that f is bounded above on $[a, b]$.

Now we carry out the second part of the proof. Let $M \in \mathbb{R}$ be an upper bound for f . Choose any $m \in \text{range}(f)$. Clearly $m \leq M$. If $m = M$ then there is $c \in [a, b]$ such that $f(c) = m$ (since m is part of the range), and moreover $f(x) \leq f(c)$ for all $x \in [a, b]$

(since $f(c) = M$ is an upper bound). So we consider the case $m < M$. Our goal is to find a value in $[m, M]$ that is simultaneously (1) in the range of f and (2) an upper bound for f .

To this end, construct a pair of bisection sequences

$$m = r_1 \leq r_2 \leq r_3 \leq \cdots \leq b_3 \leq b_2 \leq b_1 = M$$

via the rule “color the midpoint red if it is in the range of f , and blue otherwise”, so that we have

- $r_n \in \text{range}(f)$ for every n ;
- $b_n \notin \text{range}(f)$ for every n ;
- the sequences r_n and b_n converge to a common limit $L \in [m, M]$.

We prove that L is an upper bound for f and that it is in the range of f .

Upper bound: Suppose L is not an upper bound for f . Then there is $x \in [a, b]$ such that $f(x) > L$. Since $b_n \searrow L$, there is n such that $b_n \in (L, f(x))$. But then $m < b_n < f(x)$, and since m is in the range of f , the Intermediate Value Theorem implies that b_n is also in the range of f , contradicting the definition of the sequences. Thus L is an upper bound for f .

In the range: Suppose L is not in the range of f . Then $g(x) := \frac{1}{L-f(x)}$ is a continuous function on $[a, b]$, so by the first half of the EVT it is bounded above by some $K \in \mathbb{R}$. But $g(x) \leq K$ is equivalent to $L - f(x) \geq \frac{1}{K}$, which is equivalent to $f(x) \leq L - \frac{1}{K}$. Since $r_n \nearrow L$, there is n such that $r_n \in (K - \frac{1}{L}, K)$, but since $r_n \in \text{range}(f)$ this means that there is $x \in [a, b]$ such that $f(x) = r_n > L - \frac{1}{K}$, contradicting boundedness of g . This contradiction implies that L is in the range of f , so there is $c \in [a, b]$ such that $f(c) = L$. Since L is an upper bound for f , this means that c is a global maximum. \square

Lecture 26

Local extrema; Mean Value Theorem

DATE: FRIDAY, OCTOBER 25

Stewart §4.2, Spivak Chapter 11

26.1. Local extrema and Fermat's Theorem

Definition 26.1. A function $f: D \rightarrow \mathbb{R}$ has a *local maximum* (or *relative maximum*) at $c \in D$ if there exists $\delta > 0$ such that for all $x \in D \cap (c - \delta, c + \delta)$, we have $f(x) \leq f(c)$. Similarly, f has a *local minimum* (or *relative minimum*) at $c \in D$ if there exists $\delta > 0$ such that for all $x \in D \cap (c - \delta, c + \delta)$, we have $f(x) \geq f(c)$.

Remark 26.2. Stewart's definition of local extreme points differs slightly from the one above; he requires that a local maximum or minimum have the property that $(c - \delta, c + \delta) \subset D$ for some $\delta > 0$. In other words, he does not allow endpoints of intervals to be local maxima or minima. We follow Spivak's definitions instead.

Clearly a global maximum is a local maximum, and similarly for minima. The converse is not true.

Exercise 26.3. Prove that $f(x) = (x^2 - 1)^2$ has a local maximum at 0, but that this is not a global maximum.

Theorem 26.4 (Fermat's theorem). *If $f: (a, b) \rightarrow \mathbb{R}$ has a local maximum or minimum at $c \in (a, b)$, and if the derivative $f'(c)$ exists, then $f'(c) = 0$.*

Proof. We give the proof for a local maximum. The other case is similar (or we can just consider $-f$). Since $f'(c)$ exists, the left and right derivatives of f at c exist and agree. Let $\delta > 0$ be such that every $x \in (c - \delta, c + \delta)$ satisfies $x \in (a, b)$ and $f(x) \leq f(c)$. Then for every $h \in (0, \delta)$ we have $f(c + h) \leq f(c)$, and thus $\frac{f(c+h)-f(c)}{h} \leq 0$. The monotonicity property of limits gives

$$f'(c) = D^+ f(c) = \lim_{h \rightarrow 0^+} \frac{f(c+h) - f(c)}{h} \leq 0.$$

Similarly, for every $h \in (-\delta, 0)$, we have $\frac{f(c+h)-f(c)}{h} \geq 0$ since the numerator is ≤ 0 and the denominator is < 0 . Thus

$$f'(c) = D^- f(c) = \lim_{h \rightarrow 0^-} \frac{f(c+h) - f(c)}{h} \geq 0.$$

Since $f'(c)$ is simultaneously ≤ 0 and ≥ 0 , we must have $f'(c) = 0$. □

Definition 26.5. We say that c is a *critical point* for f if $f'(c) = 0$ or if $f'(c)$ does not exist. If c is a critical point, then $f(c)$ is called a *critical value*.

Fermat's theorem can be reformulated in the following way.

Corollary 26.6. *If $f: [a, b] \rightarrow \mathbb{R}$ has a local maximum or minimum at $c \in [a, b]$, then either c is an endpoint of the interval, or c is a critical point for f .*

It is important to stress that the converse of this result is not true; critical points (and endpoints) need not be local maxima or minima. For example, $f(x) = x^3$ has a critical point at $x = 0$, but this is neither a local maximum nor a local minimum. Also, the case of nondifferentiability really does need to be included in the definition of critical point; $f(x) = |x|$ has $f'(x) = \pm 1$ wherever the derivative exists, and a local (and global) minimum at $x = 0$, where it is not differentiable.

Corollary 26.6 gives us a method for finding the global maxima and minima of a continuous function $f: [a, b] \rightarrow \mathbb{R}$:

- (1) Find the critical points of f on (a, b) .
- (2) Compare the values of f at its critical points and at the endpoints a and b . This will often be a finite set of points, and the largest and smallest values from this list give the extreme values of f on $[a, b]$.

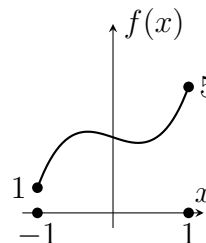
Example 26.7. Consider the function $f(x) = 3x^3 - x + 3$ on $[-1, 1]$. Because the interval is closed and the function is continuous, the method above will work to find the maxima and minima. To find the critical points, we compute

$$0 = f'(x) = 9x^2 - 1 \quad \Leftrightarrow \quad x^2 = \frac{1}{9} \quad \Leftrightarrow \quad x = \pm \frac{1}{3}.$$

Thus there are four points to check: the endpoints ± 1 and the critical points $\pm \frac{1}{3}$. We see that

- $f(-1) = 3(-1)^3 - (-1) + 3 = 1$;
- $f(1) = 3 - 1 + 3 = 5$;
- $f(-\frac{1}{3}) = 3(-\frac{1}{27}) - (-\frac{1}{3}) + 3 = -\frac{1}{9} + \frac{1}{3} + 3 = 3 + \frac{2}{9} = \frac{29}{9}$;
- $f(\frac{1}{3}) = 3(\frac{1}{27}) - \frac{1}{3} + 3 = 3 - \frac{2}{9} = \frac{25}{9}$.

Thus $f(-1) < f(\frac{1}{3}) < f(-\frac{1}{3}) < f(1)$, and we see that the global maximum is at $x = 1$ and the global minimum is at $x = -1$. The picture at right shows the shape of the graph; it appears that f has a local maximum at $-\frac{1}{3}$ and a local minimum at $\frac{1}{3}$. This can be verified by observing that on $[-1, 0]$, the only critical point is $-\frac{1}{3}$, and that $f(-\frac{1}{3}) > \max(f(-1), f(0))$; that $x = \frac{1}{3}$ is a local minimum can be proved similarly.



26.2. The Mean Value Theorem

Given a function $f: [a, b] \rightarrow \mathbb{R}$, the average rate of change of f over the entire interval is $\frac{f(b)-f(a)}{b-a}$. It is often useful to know that when f is differentiable, there is some point $c \in (a, b)$ at which the *instantaneous* rate of change $f'(c)$ is equal to the average rate of change; this is the content of the *Mean Value Theorem*. Before proving this general case of the theorem, we start with the case when $f(b) = f(a)$.

Theorem 26.8 (Rolle's Theorem). *Let f be continuous on $[a, b]$ and differentiable on (a, b) . If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. The following is almost a proof: “By the extreme value theorem, f has a global maximum at some point $x = c$. This global maximum is also a local maximum, so by Fermat's theorem it is a critical point. Since f is differentiable, we have $f'(c) = 0$.”

The only thing wrong with this “proof” is that a global maximum might occur at an endpoint. But this can be dealt with as follows. If $c \in (a, b)$ then the above argument indeed gives $f'(c) = 0$. So suppose we have $c = a$ or $c = b$. We have $f(x) \leq f(c)$ for all $x \in (a, b)$ by the definition of global maximum. If $f(x) = f(c)$ for all $x \in (a, b)$, then $f'(x) = 0$ for all $x \in (a, b)$ because constant functions have zero derivatives. On the other hand, if there is $x \in (a, b)$ such that $f(x) < f(c)$, then taking p to be the global minimum of f , we see that $p \in (a, b)$ since $f(a) = f(b) = f(c)$, and thus p is a critical point, with $f'(p) = 0$. \square

Rolle's theorem can be used to prove that certain equations have a *unique* solution, even without finding an exact value.

Example 26.9. Consider the function $f(x) = x^5 + x^3 + x - 1$. We claim that there is exactly one real number r such that $f(r) = 0$. First note that $f(0) = -1$ and $f(1) = 2$, and that f is continuous, so by the Intermediate Value Theorem there exists $r \in [0, 1]$ such that $f(r) = 0$. Now suppose there exists $s \in \mathbb{R}$ such that $f(s) = 0$ and $s \neq r$. Then by Rolle's theorem, there exists c between r and s such that $f'(c) = 0$. However, we have $f'(x) = 5x^4 + 3x^2 + 1 \geq 1$ for all $x \in \mathbb{R}$, so no such s can exist.

The rest of this section was done on Monday, October 28

Theorem 26.10 (Mean Value Theorem). *Let f be continuous on $[a, b]$ and differentiable on (a, b) . Then there exists $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$, or equivalently, $f(b) - f(a) = f'(c)(b - a)$.*

Proof. The line between $(a, f(a))$ and $(b, f(b))$ has equation $y = f(a) + \frac{f(b)-f(a)}{b-a}(x - a)$. Let $h(x)$ denote the difference between $f(x)$ and this line, so

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

Observe that h is continuous on $[a, b]$ and differentiable on (a, b) ; moreover, $h(a) = h(b) = 0$, so Rolle's theorem gives $c \in (a, b)$ such that $h'(c) = 0$. Since

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a},$$

this implies that $f'(c) = \frac{f(b)-f(a)}{b-a}$, and completes the proof. \square

If we are given bounds on f' , we can use the Mean Value Theorem to bound f .

Example 26.11. If a differentiable function f satisfies $f(0) = 1$ and $f'(x) \leq 2$ for all x , then we can get an upper bound for $f(3)$ by applying the MVT to $[0, 3]$ to get some $c \in (0, 3)$ with

$$\frac{f(3) - f(0)}{3 - 0} = f'(c) \leq 2.$$

Multiplying by 3 gives $f(3) - f(0) \leq 6$, so $f(3) \leq f(0) + 6 = 7$.

We can use the MVT to give a short proof of Theorem 22.1, which said that if $f: (a, b) \rightarrow \mathbb{R}$ is differentiable and $f'(x) = 0$ for every $x \in (a, b)$, then f is constant on (a, b) . Indeed, given any $x < y$ in (a, b) , the MVT gives $c \in (x, y)$ such that

$$f(y) - f(x) = f'(c)(y - x) = 0(y - x) = 0 \quad \Rightarrow \quad f(y) = f(x).$$

This fact has the following important consequence.

Corollary 26.12. *If f and g are differentiable on (a, b) and have $f'(x) = g'(x)$ for every $x \in (a, b)$, then there exists $C \in \mathbb{R}$ such that $f(x) = g(x) + C$ for all $x \in (a, b)$.*

Proof. Let $F(x) = f(x) - g(x)$; then $F'(x) = f'(x) - g'(x) = 0$, so F is constant on (a, b) by Theorem 22.1. \square

Example 26.13. It is crucial that the set of x on which we apply these results is a single open interval. The Heaviside function $H(x)$ defined by $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x \geq 0$ has the property that $H'(x) = 0$ for all $x \neq 0$, but it is not constant on any interval that contains both positive and negative numbers.

Example 26.14. We saw in Lecture 22 that $\frac{dy}{dt} = ky$ has $y(t) = y(0)e^{kt}$ as a solution. In fact it is the *only* solution; any solution $y(t)$ must have

$$\frac{d}{dt}(\ln y(t) - kt) = \frac{y'}{y} - k = 0,$$

and thus $\ln y(t) - kt = \ln y(0)$ for all t , which gives the formula above.

Example 26.15. We can use Theorem 22.1 to prove that $\tan^{-1}x + \cot^{-1}x = \frac{\pi}{2}$ for all x ; differentiating the left-hand side gives $\frac{1}{1+x^2} - \frac{1}{1+x^2} = 0$, and at $x = 1$ we have $\tan^{-1}1 = \cot^{-1}1 = \frac{\pi}{4}$. (One could also give a direct proof of this identity.)

Lecture 27

Shapes of graphs

DATE: MONDAY, OCTOBER 28

Stewart §4.3, Spivak Chapter 11

27.1. Monotonicity and first derivatives

Back in Lecture 13.2, we observed that a function f with positive derivative f' should be increasing and a function with negative derivative should be decreasing. One can prove this directly from the definition of derivative, though we did not do so; instead we opted to wait until now, since with the MVT in hand we can give a quick proof.

Theorem 27.1. *Let f be continuous on $[a, b]$ and differentiable on (a, b) . If $f'(x) > 0$ for every $x \in (a, b)$, then f is strictly increasing on $[a, b]$ (this means that $f(x_2) > f(x_1)$ whenever $x_2 > x_1$ for $x_1, x_2 \in [a, b]$). Similarly, if $f'(x) < 0$ for every $x \in (a, b)$, then f is strictly decreasing on $[a, b]$. If we replace $>$ and $<$ with \geq and \leq , we get a similar result with “strictly increasing” replaced by “nondecreasing”, and “strictly decreasing” replaced by “nonincreasing”.*

Proof. Suppose $f'(x) > 0$ for every $x \in (a, b)$. Then given any $x_1 < x_2$ in $[a, b]$, the MVT gives $c \in (x_1, x_2)$ such that

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) > 0.$$

Thus f is increasing. The proof for the case $f' < 0$ is the same, simply reverse the last inequality. \square

Remark 27.2. This theorem is not a dichotomy; there are plenty of functions that do not satisfy either of the derivative conditions on a given interval, and that are neither increasing or decreasing. (Though they may be increasing or decreasing on a smaller interval.)

Remark 27.3. The converse of the theorem fails; a function can be strictly increasing even if its derivative vanishes somewhere. For example, $f(x) = x^3$ is strictly increasing but $f'(0) = 0$.

Example 27.4. Consider the function $f(x) = 3x^5 - 5x^3 + 2$; the graphs of f and f' are shown in Figure 9. Observe that

$$f'(x) = 15x^4 - 15x^2 = 15x^2(x^2 - 1) = 15x^2(x - 1)(x + 1).$$

We identify the intervals on which f' is positive and negative by checking the sign of each of its factors between its zeros.

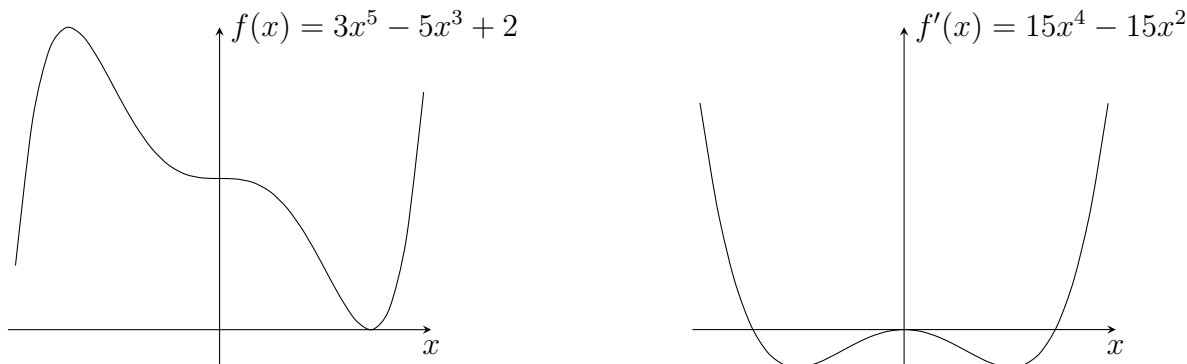


FIGURE 9. Connection between monotonicity of f and sign of f' .

	$15x^2$	$x - 1$	$x + 1$	$f'(x)$
$x < -1$	+	-	-	+
$-1 < x < 0$	+	-	+	-
$0 < x < 1$	+	-	+	-
$x > 1$	+	+	+	+

Using this table and Theorem 27.1, we conclude that f is increasing on $(-\infty, 1]$ and $[1, \infty)$, while it is decreasing on $[-1, 0]$ and $[0, 1]$ (and thus in fact it is decreasing on $[-1, 1]$).

We see that -1 is a local maximum for f , because for $x < -1$ we have $f(x) < f(-1)$ since f is increasing on this interval, while for $x \in (-1, 0)$ we have $f(x) < f(-1)$ since f is decreasing on this interval. Similar reasoning shows that 1 is a local minimum. Observe that 0 is a critical point that is neither a local maximum nor a local minimum.

The reasoning in the last paragraph of the example is worth codifying.

Theorem 27.5 (First Derivative Test). *Let f be differentiable on an interval (a, b) that contains a critical point c .*

- (1) *If f' changes from positive to negative at c ($f' > 0$ just to the left of c , and $f' < 0$ just to the right of c), then f has a local maximum at c .*
- (2) *If f' changes from negative to positive at c ($f' < 0$ just to the left of c , and $f' > 0$ just to the right of c), then f has a local minimum at c .*
- (3) *If f' does not change sign at c (either $f'(x) > 0$ for all $x \approx c$, or $f'(x) < 0$ for all $x \approx c$), then f has neither a local maximum nor a local minimum at c .*

27.2. Convexity and second derivative

We also saw in Lecture 17.1 that the sign of f'' should be related to *convexity* properties of f . Now we make this precise. Recall from Definition 17.5 that a function f is *convex* on an interval I if for every $x, y \in I$ and every $t \in [0, 1]$, we have

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

Given $t \in [0, 1]$, let $z = tx + (1 - t)y$ be the point at which f is evaluated to get the left-hand side. Note that when $t = 0$ we have $z = y$, and when $t = 1$ we have $z = x$, so z slides from y to x as t goes from 0 to 1. The right-hand side is equal to $f(y)$ when $t = 0$

and $f(x)$ when $t = 1$, and represents the point above z on the line joining $(x, f(x))$ to $(y, f(y))$. Thus a function is convex if its graph lies on or below this line between x and y , as illustrated in the left half of Figure 10

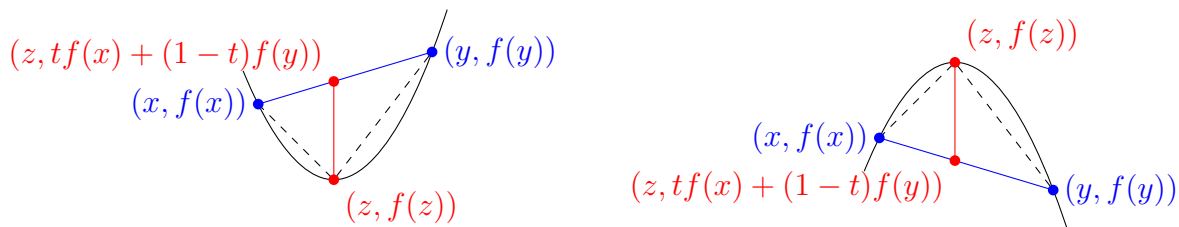


FIGURE 10. Convex (left) and concave (right).

Definition 27.6. A function f is *concave* on an interval I if for every $x, y \in I$ and every $t \in [0, 1]$, we have

$$f(tx + (1 - t)y) \geq tf(x) + (1 - t)f(y).$$

A concave function is shown in the right half of Figure 10; it has the property that the graph of f lies above its secant line between x and y .

Remark 27.7. Some authors call convex functions *concave upward*, and concave functions *concave downward*.

The definition of convexity we gave above works whether or not f is differentiable. If f' exists then we can use it to give alternate characterizations of convexity.

Theorem 27.8. Suppose f is differentiable on an interval I . Then the following are equivalent.

- (1) f is convex on I .
- (2) $\frac{f(z)-f(x)}{z-x} \leq f'(z) \leq \frac{f(y)-f(z)}{y-z}$ for all $x < z < y$ in I .
- (3) f' is a nondecreasing function on I .
- (4) The graph of f lies on or above all its tangent lines on I .

The corresponding result for ‘concave’ holds, where we reverse all the inequalities, replace ‘nondecreasing’ by ‘nonincreasing’, and replace ‘above’ by ‘below’.

The proof of this theorem was skipped in lecture

Proof. We prove that (1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1), and that (2) \Leftrightarrow (4).

(1) \Rightarrow (2). Suppose that f is convex and consider $x < z < y$. Comparing the slopes of the two secant lines associated to x and z and to z and y (see the left half of Figure 10), we see that

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(z)}{y - z}.$$

Taking a limit of the first quantity as $x \rightarrow z^-$ gives the second inequality in (2). Taking a limit of the second quantity as $y \rightarrow z^+$ gives the first inequality in (2).

(2) \Rightarrow (3). Given any $a < b$ in I , the second inequality in (2) (with $z = a$ and $y = b$) gives $f'(a) \leq \frac{f(b)-f(a)}{b-a}$, while the first inequality in (2) (with $x = a$ and $z = b$) gives $\frac{f(b)-f(a)}{b-a} \leq f'(b)$. Combining these gives $f'(a) \leq f'(b)$, which proves (3).

(3) \Rightarrow (1). Suppose that f is *not* convex. Then there exists $x < z < y$ such that $(z, f(z))$ lies above the line through $(x, f(x))$ and $(y, f(y))$ (see the right half of Figure 10), which in turn implies that

$$\frac{f(z) - f(x)}{z - x} > \frac{f(y) - f(z)}{y - z}.$$

By the MVT there are points $a \in (x, z)$ and $b \in (z, y)$ such that $f'(a) = \frac{f(z) - f(x)}{z - x}$ and $f'(b) = \frac{f(y) - f(z)}{y - z}$. But then we have $a < b$ and $f'(a) > f'(b)$, so that f' is *not* a nondecreasing function on I .

(2) \Leftrightarrow (4). The tangent line at a has equation $g(x) = f(a) + f'(a)(x - a)$. For $x < a$, we see that the following are equivalent:

- $(x, f(x))$ lies above this tangent line;
- $f(x) \geq f(a) + f'(a)(x - a)$;
- $f(a) - f(x) \leq f'(a)(a - x)$;
- $\frac{f(a) - f(x)}{a - x} \leq f'(a)$.

For $y > a$, the following are equivalent.

- $(y, f(y))$ lies above the tangent line at a ;
- $f(y) \geq f(a) + f'(a)(y - a)$;
- $f(y) - f(a) \geq f'(a)(y - a)$;
- $f'(a) \leq \frac{f(y) - f(a)}{y - a}$.

We conclude that the graph of f lies above its tangent lines if and only if for every $x < a < y$ we have $\frac{f(a) - f(x)}{a - x} \leq f'(a) \leq \frac{f(y) - f(a)}{y - a}$, which is (2). \square

Combining Theorems 27.1 and 27.8, we can give a criterion for convexity using the second derivative.

Theorem 27.9. *Let f be twice differentiable on an interval I .*

- (1) *If $f''(x) > 0$ for all $x \in I$, then f is convex on I .*
- (2) *If $f''(x) < 0$ for all $x \in I$, then f is concave on I .*

Proof. If $f''(x) > 0$ for all $x \in I$, then Theorem 27.1 implies that f' is nondecreasing on I , and then Theorem 27.8 implies that f is convex on I . The proof for the second half is similar. \square

The rest of this section was done in lecture on Wednesday, October 30

Definition 27.10. If f'' changes sign at x (goes from positive to negative or vice versa) then we say that x is an *inflection point* of f . Thus at an inflection point, f goes from being convex to being concave, or vice versa.

We can also use the second derivative to determine whether a critical point is a local maximum or minimum.

Theorem 27.11 (Second Derivative Test). *Let f be twice differentiable on an interval (a, b) that contains a critical point c .*

- (1) *If $f''(c) < 0$, then f has a local maximum at c .*
- (2) *If $f''(c) > 0$, then f has a local minimum at c .*

(3) If $f''(c) = 0$, then c could be either a local maximum, a local minimum, or neither.

Proof. If $f''(c) > 0$ then we have

$$0 < f''(c) = \lim_{h \rightarrow 0} \frac{f'(c+h) - f'(c)}{h} = \lim_{h \rightarrow 0} \frac{f'(c+h)}{h}.$$

Thus there exists $\delta > 0$ such that $\frac{f'(c+h)}{h} > 0$ whenever $0 < |h| < \delta$. In particular, for $0 < h < \delta$ we have $f'(c+h) > 0$, and $f'(c-h) < 0$. By the first derivative test, this implies that c is a local minimum. The proof when $f''(c) < 0$ is similar. \square

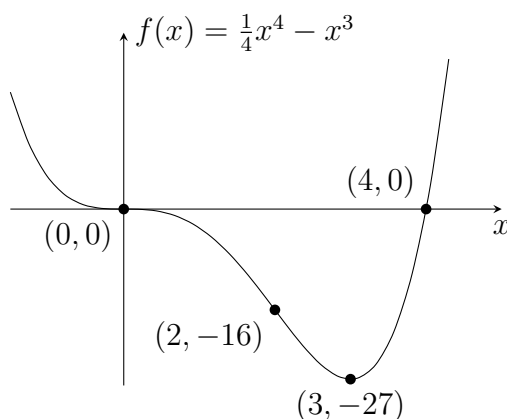


FIGURE 11. Roots, critical points, and inflection points.

Example 27.12. Consider the function $f(x) = \frac{1}{4}x^4 - x^3$, whose graph is shown in Figure 11. This function has roots at $x = 0$ and $x = 4$, is negative between the roots, and positive elsewhere. Differentiating gives

$$f'(x) = x^3 - 3x^2 = x^2(x - 3) \quad \text{and} \quad f''(x) = 3x^2 - 6x = 3x(x - 2).$$

Thus f has critical points at 0 and 3, and inflection points at 0 and 2. Because $f''(3) = 36 > 0$, the critical point at 3 is a local minimum by the Second Derivative Test. This text provides no information about the critical point at 0, and indeed checking the sign of f reveals that this is neither a local maximum nor a local minimum.

By considering the sign of f' , we see that f is decreasing on $(-\infty, 3)$ and increasing on $(3, \infty)$. Considering the sign of f'' , we see that f is convex on $(-\infty, 0)$ and $(2, \infty)$, and concave on $(0, 2)$.

Lecture 28

l'Hospital's rule

DATE: WEDNESDAY, OCTOBER 30

Stewart §4.4, Spivak Chapter 11

28.1. Indeterminate forms

The limit laws tell us how to compute $\lim \frac{f}{g}$ provided (1) $\lim f$ and $\lim g$ exist, and (2) $\lim g \neq 0$. But what if $\lim g = 0$? If $\lim f \neq 0$ then either $\lim \frac{f}{g} = \pm\infty$ (if g is consistently positive or consistently negative) or $\frac{f}{g}$ oscillates between $\pm\infty$ (if g oscillates between positive and negative values in the limit). If $\lim f = 0$, on the other hand, then the limit has the *indeterminate form* $\frac{0}{0}$, and many different outcomes are possible.

Example 28.1. We have seen the following examples in previous lectures.

- (1) $\lim_{x \rightarrow 1} \frac{x-1}{x^2-1} = \lim_{x \rightarrow 1} \frac{1}{x+1} = \frac{1}{2}$ by algebraic simplifications (difference of squares).
- (2) $\lim_{x \rightarrow 1} \frac{\sqrt{x}-1}{x-1} = \lim_{x \rightarrow 1} \frac{\sqrt{x}-1}{x-1} \cdot \frac{\sqrt{x}+1}{\sqrt{x}+1} = \lim_{x \rightarrow 1} \frac{1}{\sqrt{x}+1} = \frac{1}{2}$ by multiplying by a conjugate and simplifying.
- (3) $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ by a geometric argument that we gave in Lecture 8.
- (4) $\lim_{x \rightarrow 0} \frac{\sin(2x)}{\sin x} = \lim_{x \rightarrow 0} 2 \frac{\sin(2x)}{2x} \frac{x}{\sin x} = 2$ using the previous example.
- (5) $\lim_{x \rightarrow \infty} \frac{x^2}{e^x} = 0$ by an argument from Lecture 24 using properties of exponentials.
- (6) $\lim_{x \rightarrow \infty} \frac{\ln x}{x} = 0$ by a similar argument, or as a consequence of the previous example.

The computation above for $\lim_{x \rightarrow 0} \frac{\sin 2x}{\sin x}$ suggests the following general approach:

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \frac{x - a}{g(x) - g(a)} = \frac{f'(a)}{g'(a)},$$

where the first equality works provided $f(a) = g(a) = 0$, and the second works provided f, g are both differentiable at a . Another way of looking at this is as follows: if f, g are differentiable at a and vanish at a , then for $x \approx a$ we have the linear approximations $f(x) = f'(a)(x-a) + r(x)$ and $g(x) = g'(a)(x-a) + s(x)$, where r, s are error functions with the property that $\lim_{x \rightarrow a} \frac{r(x)}{x-a} = \lim_{x \rightarrow a} \frac{s(x)}{x-a} = 0$. Thus

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(a)(x-a) + r(x)}{g'(a)(x-a) + s(x)} = \lim_{x \rightarrow a} \frac{f'(a) + \frac{r(x)}{x-a}}{g'(a) + \frac{s(x)}{x-a}} = \frac{f'(a)}{g'(a)}.$$

This gives the intuition behind l'Hospital's rule,²⁹ but there is a shortcoming inherent in this approach; these arguments only work if f, g are differentiable at a , and if $g'(a) \neq 0$.

Example 28.2. The limits $\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2}$ and $\lim_{x \rightarrow 0^+} \frac{\sqrt{x}}{e^{-1/x}}$ have indeterminate form $\frac{0}{0}$, but *cannot* be computed via the above argument. In the first case, we have $f'(0) = g'(0) = 0$, so $f'(0)/g'(0)$ is undefined, and in the second case neither f nor g is differentiable at 0; indeed, g is not even defined at 0.

²⁹We follow Stewart's book in our spelling; Spivak's book uses the alternate spelling l'Hôpital.

In light of Example 28.2, we might try to replace $f'(a)/g'(a)$ with $f'(x)/g'(x)$ for $x \approx a$, since this quantity is defined for both examples given there, and then take a limit as $x \rightarrow a$. To relate f'/g' to f/g , we could try to use the MVT. Indeed, if $x > a$ is such that f, g are continuous on $[a, x]$ and differentiable on (a, x) , then the MVT gives $y, z \in (a, x)$ such that

$$f'(y) = \frac{f(x) - f(a)}{x - a} = \frac{f(x)}{x - a} \text{ and } g'(z) = \frac{g(x) - g(a)}{x - a} = \frac{g(x)}{x - a} \Rightarrow \frac{f'(y)}{g'(z)} = \frac{f(x)}{g(x)}.$$

Since $y, z \rightarrow a$ as $x \rightarrow a$, this *almost* does what we want. But not quite. We don't have any way to guarantee that y, z are the same point, and it is not clear why $f'(y)/g'(z)$ should tell us anything about $f'(y)/g'(y)$ or $f'(z)/g'(z)$.

28.2. Cauchy's MVT

The solution is to use a slightly stronger version of the Mean Value Theorem.

Theorem 28.3 (Cauchy's MVT). *If f, g are continuous on $[a, b]$ and differentiable on (a, b) , and if $g'(x) \neq 0$ for all $x \in (a, b)$, then there exists $c \in (a, b)$ such that*

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Note that if we choose $g(x) = x$, then this reduces to the standard MVT. And indeed, we can prove Cauchy's MVT from Rolle's Theorem by mimicking the proof of the MVT.

Proof of Cauchy's MVT. Consider the function

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

and observe that $h(a) = h(b) = 0$. Moreover, h is continuous on $[a, b]$ and differentiable on (a, b) , so by Rolle's Theorem there is $c \in (a, b)$ such that $h'(c) = 0$. Note that

$$h'(x) = f'(x) - g'(x) \frac{f(b) - f(a)}{g(b) - g(a)},$$

so $h'(c) = 0$ implies $f'(c) = g'(c) \frac{f(b) - f(a)}{g(b) - g(a)}$. Since $g'(c) \neq 0$, this completes the proof. \square

Exercise 28.4. Prove that under the conditions of Cauchy's MVT, we have $g(b) \neq g(a)$, so that the definition of $h(x)$ in the proof is valid.

28.3. l'Hospital's rule for limits of the form 0/0

Theorem 28.5 (l'Hospital's rule, right-handed limits of the form 0/0).

Suppose we are given functions f, g and $a < b$ such that the following are true:

- (1) f, g are differentiable on (a, b) and $g'(x) \neq 0$ for all $x \in (a, b)$;
- (2) $\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0$;
- (3) $\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$ exists.

Then $\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)}$ exists and is equal to $\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$.

Proof. Since none of the hypotheses or conclusions are affected by the value of f and g at a (or indeed by whether or not f and g are even defined at a), we may redefine f, g at this single point if necessary and assume that $f(a) = g(a) = 0$, so that f, g are continuous on $[a, b)$. Now let $L = \lim_{x \rightarrow a^+} \frac{f(x)}{g(x)}$. Then for every $\varepsilon > 0$, there is $\delta > 0$ such that

$$(28.1) \quad \left| \frac{f'(c)}{g'(c)} - L \right| < \varepsilon \text{ for all } c \in (a, a + \delta).$$

Given any $x \in (a, a + \delta)$ we can apply Cauchy's Mean Value Theorem to F, G on the interval $[a, x]$ and obtain $c \in (a, x)$ such that

$$(28.2) \quad \frac{f'(c)}{g'(c)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x)}{g(x)}.$$

Combining (28.1) and (28.2) gives $\left| \frac{f(x)}{g(x)} - L \right| < \varepsilon$ for all $x \in (a, a + \delta)$. Since $\varepsilon > 0$ was arbitrary, this proves that $L = \lim_{x \rightarrow a^+} \frac{f(x)}{g(x)}$. \square

Exercise 28.6. Prove the following versions of l'Hospital's rule for limits of the form $0/0$.

- (1) *Left-handed limits:* Theorem 28.5 remains true if we replace $\lim_{x \rightarrow a^+}$ by $\lim_{x \rightarrow b^-}$.
- (2) *Two-sided limits:* Theorem 28.5 remains true if we replace $\lim_{x \rightarrow a^+}$ by $\lim_{x \rightarrow a^-}$, provided there is an open interval I containing a such that f, g are differentiable on $I \setminus \{a\}$ and g' does not vanish on $I \setminus \{a\}$.

Exercise 28.7. Formulate and prove versions of l'Hospital's rule for right-handed, left-handed, and two-sided limits where we replace "lim f'/g' exists" with "lim $f'/g' = \infty$ " or "lim $f'/g' = -\infty$ ".

Remark 28.8. We emphasize that l'Hospital's rule does not require f and g to be differentiable, continuous, or even defined at a itself.

The remainder of this section was done on Friday, November 1

Corollary 28.9 (l'Hospital's rule, limits at infinity of the form $0/0$).

Suppose we are given functions f, g and $r \in \mathbb{R}$ such that the following are true:

- (1) f, g are differentiable on (r, ∞) and $g'(x) \neq 0$ for all $x > r$;
- (2) $\lim_{x \rightarrow \infty} f(x) = \lim_{x \rightarrow \infty} g(x) = 0$;
- (3) $\lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}$ exists.

Then $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$ exists and is equal to $\lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}$.

Proof. Define functions $F, G: (0, 1/r) \rightarrow \mathbb{R}$ by

$$F(x) = f(1/x) \text{ and } G(x) = g(1/x).$$

We claim that F, G satisfy the three conditions of Theorem 28.5. The chain rule gives

$$F'(x) = -x^{-2}f'(1/x) \text{ and } G'(x) = -x^{-2}g'(1/x),$$

so F, G are differentiable and G' is nonvanishing on $(0, 1/r)$, which verifies the first condition. The second condition follows from observing that $\lim_{x \rightarrow 0^+} F(x) = \lim_{x \rightarrow 0^+} f(1/x) = \lim_{t \rightarrow \infty} f(t) = 0$, and similarly for G . The third condition follows since

$$(28.3) \quad \lim_{x \rightarrow 0^+} \frac{F'(x)}{G'(x)} = \lim_{x \rightarrow 0^+} \frac{-x^{-2}f'(1/x)}{-x^{-2}g'(1/x)} = \lim_{x \rightarrow 0^+} \frac{f'(1/x)}{g'(1/x)} = \lim_{t \rightarrow \infty} \frac{f'(t)}{g'(t)}.$$

Thus Theorem 28.5 applies to F, G for the right-hand limit at 0, and we deduce that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow \infty} \frac{F(1/x)}{G(1/x)} = \lim_{t \rightarrow 0^+} \frac{F(t)}{G(t)} = \lim_{t \rightarrow 0^+} \frac{F'(t)}{G'(t)} = \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)},$$

where the last equality uses (28.3). □

Now we can revisit the limits in Examples 28.1 and 28.2 and compute them using l'Hospital's rule.

- (1) $\lim_{x \rightarrow 1} \frac{x-1}{x^2-1} = \lim_{x \rightarrow 1} \frac{1}{2x} = \frac{1}{2}$.
- (2) $\lim_{x \rightarrow 1} \frac{\sqrt{x}-1}{x-1} = \lim_{x \rightarrow 1} \frac{\frac{1}{2}x^{-1/2}}{1} = \frac{1}{2}$.
- (3) $\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{\cos x}{1} = 1$. (Note, however, that the proof that $\frac{d}{dx} \sin x = \cos x$ required us to first compute this limit!)
- (4) $\lim_{x \rightarrow 0} \frac{\sin(2x)}{\sin x} = \lim_{x \rightarrow 0} \frac{2 \cos(2x)}{\cos x} = 2$.
- (5) $\lim_{x \rightarrow \infty} \frac{x^2}{e^x} = \lim_{x \rightarrow \infty} \frac{2x}{e^x} = \lim_{x \rightarrow \infty} \frac{2}{e^x} = 0$, where we have used l'Hospital's rule *twice*.
- (6) $\lim_{x \rightarrow \infty} \frac{\ln x}{x} = \lim_{x \rightarrow \infty} \frac{1/x}{1} = 0$.
- (7) $\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \lim_{x \rightarrow 0} \frac{\sin x}{2x} = \frac{1}{2}$.
- (8) $\lim_{x \rightarrow 0^+} \frac{\sqrt{x}}{e^{-1/x}} = \lim_{x \rightarrow 0^+} \frac{\frac{1}{2}x^{-1/2}}{x^{-2}e^{-1/x}} = \frac{1}{2} \lim_{x \rightarrow 0^+} \frac{x^{3/2}}{e^{-1/x}} = ???$.

In this last example, it is not clear what to do next; we could try applying l'Hospital's rule again, and hope that the expression starts to simplify, but if you carry out the computation you will find that in fact the numerator is now $x^{5/2}$, and we are no closer to a solution. On the other hand, we might start over and go in a different direction by writing the original expression as $\lim_{x \rightarrow 0^+} e^{1/x}/x^{-1/2}$; then numerator and denominator both go to ∞ in the limit, so this is a limit with the *indeterminate form* ∞/∞ . If we had a version of l'Hospital's rule for such limits, then we could write

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

In the next lecture we will prove that l'Hospital's rule works for limits of this form.

Remark 28.10. We conclude with a warning. In order to apply l'Hospital's rule, it is crucial that the limit has indeterminate form. If we blindly differentiate top and bottom without first checking this condition, we can get ourselves into trouble. For example,

with $f(x) = \sin x$ and $g(x) = \cos x$ we have

$$\lim_{x \rightarrow 0^+} \frac{\sin x}{\cos x} = 0 \quad \text{but} \quad \lim_{x \rightarrow 0^+} \frac{-\cos x}{\sin x} = -\infty.$$

This does not violate l'Hospital's rule because $\lim g \neq 0$, so the conditions of Theorem 28.5 are not met.

Lecture 29 **More on l'Hospital's rule, and basics of curve sketching**

DATE: FRIDAY, NOVEMBER 1

Stewart §4.4 and §4.5, Spivak Chapter 11

29.1. Limits with indeterminate forms ∞/∞

As suggested last time, we start by proving another version of l'Hospital's rule.

Theorem 29.1 (l'Hospital's rule, limits at infinity of the form ∞/∞).

Suppose we are given functions f, g and $r \in \mathbb{R}$ such that the following are true:

- (1) f, g are differentiable on (r, ∞) and $g'(x) \neq 0$ for all $x > r$;
- (2) $\lim_{x \rightarrow \infty} f(x) = \lim_{x \rightarrow \infty} g(x) = \infty$;
- (3) $\lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}$ exists.

Then $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$ exists and is equal to $\lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}$.

Proof. Let r be as in the hypothesis, and let $L = \lim_{x \rightarrow \infty} f'(x)/g'(x)$. Given $\varepsilon > 0$, let $a \geq r$ be such that $|\frac{f'(x)}{g'(x)} - L| < \varepsilon/2$ for all $x > a$. Now given any $x > a$, first note that since $f, g \rightarrow \infty$ as $x \rightarrow \infty$, there is $b \in \mathbb{R}$ such that $f(x) > f(a)$, $g(x) > g(a)$, and $g(x) > 0$ for all $x > b$. For all such x , by Cauchy's MVT there is $c \in (a, x)$ such that

$$(29.1) \quad \frac{f'(c)}{g'(c)} = \frac{f(x) - f(a)}{g(x) - g(a)}.$$

Thus we have

$$(29.2) \quad \begin{aligned} \left| \frac{f(x)}{g(x)} - L \right| &\leq \left| \frac{f(x)}{g(x)} - \frac{f'(c)}{g'(c)} \right| + \left| \frac{f'(c)}{g'(c)} - L \right| \leq \left| \frac{f'(c)}{g'(c)} \right| \cdot \left| \frac{f(x)/g(x)}{f'(c)/g'(c)} - 1 \right| + \frac{\varepsilon}{2} \\ &\leq (|L| + \varepsilon) \underbrace{\left| \frac{f(x)(g(x) - g(a))}{g(x)(f(x) - f(a))} - 1 \right|}_{1} + \frac{\varepsilon}{2}. \end{aligned}$$

The limit laws give

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} \frac{g(x) - g(a)}{f(x) - f(a)} = \lim_{x \rightarrow \infty} \frac{1 - \frac{g(a)}{g(x)}}{1 - \frac{f(a)}{f(x)}} = 1$$

using the fact that $\lim f = \lim g = \infty$, so there exists $s \in \mathbb{R}$ such that for all $x > s$, we have $1 < \frac{\varepsilon}{2(|L|+\varepsilon)}$. For every such x , (29.2) gives $|\frac{f(x)}{g(x)} - L| < \varepsilon$, which completes the proof of Theorem 29.1. \square

Exercise 29.2. Formulate and prove versions of l'Hospital's rule for one- and two-sided limits of the form ∞/∞ , as well as a version that applies when $\lim f'/g' = \pm\infty$.

29.2. Other indeterminate forms

Limits with other indeterminate forms, including $0 \cdot \infty$, $\infty - \infty$, 0^0 , ∞^0 , and 1^∞ , can often be evaluated using l'Hospital's rule by first relating them to a limit of the form $0/0$ or ∞/∞ .

Example 29.3. The limit $\lim_{x \rightarrow 0^+} x \ln x$ has the indeterminate form $0 \cdot \infty$ because $x \rightarrow 0$ and $\ln x \rightarrow -\infty$. We can write it as a limit of form ∞/∞ by writing $x \ln x = \frac{\ln x}{1/x}$; then both numerator and denominator go to $\pm\infty$, and l'Hospital's rule gives

$$\lim_{x \rightarrow 0^+} x \ln x = \lim_{x \rightarrow 0^+} \frac{\ln x}{1/x} = \lim_{x \rightarrow 0^+} \frac{1/x}{-1/x^2} = \lim_{x \rightarrow 0^+} (-x) = 0.$$

Note that if $f \rightarrow 0$ and $g \rightarrow \infty$ then we can write $fg = \frac{f}{1/g}$ to get a limit of the form $0/0$, or $fg = \frac{g}{1/f}$ to get a limit of the form ∞/∞ . We may need to choose wisely which we do: in the example above, making the other choice would give $x \ln x = \frac{x}{1/\ln x}$ and using l'Hospital's rule would require us to evaluate

$$\lim_{x \rightarrow 0^+} \frac{1}{-\frac{1}{x}(\ln x)^2} = \lim_{x \rightarrow 0^+} \frac{-x}{(\ln x)^2},$$

which is no easier to handle, and the situation would not improve upon further differentiations.

Example 29.4. $\lim_{x \rightarrow \frac{\pi}{2}^-} (\sec x - \tan x)$ has the indeterminate form $\infty - \infty$, but can be evaluated using l'Hospital's rule by first transforming it into a limit with the indeterminate form $0/0$:

$$\lim_{x \rightarrow \frac{\pi}{2}^-} (\sec x - \tan x) = \lim_{x \rightarrow \frac{\pi}{2}^-} \frac{1 - \sin x}{\cos x} = \lim_{x \rightarrow \frac{\pi}{2}^-} \frac{-\cos x}{-\sin x} = 0.$$

Example 29.5. $\lim_{x \rightarrow 0^+} x^x$ has the indeterminate form 0^0 , but can be evaluated using l'Hospital's rule by transforming it into the exponential of a limit that we already evaluated:

$$\lim_{x \rightarrow 0^+} x^x = \lim_{x \rightarrow 0^+} e^{x \ln x} = e^{\lim_{x \rightarrow 0^+} x \ln x} = e^0 = 1.$$

Here the second equality uses continuity of the exponential function.

Similarly, writing $f^g = e^{g \ln f}$ lets us evaluate limits of the forms 0^0 , ∞^0 , and 1^∞ in terms of limits of the form $0 \cdot \infty$, which can then be dealt with as above.

29.3. Curve sketching

We can assemble the tools developed so far into a fairly robust procedure for qualitative curve sketching. If we are given a function f in terms of a formula $f(x)$, then to sketch its graph we should carry out the following steps, which we illustrate in Figure 12 for the example $f(x) = \frac{3x^2}{x^2-4}$.

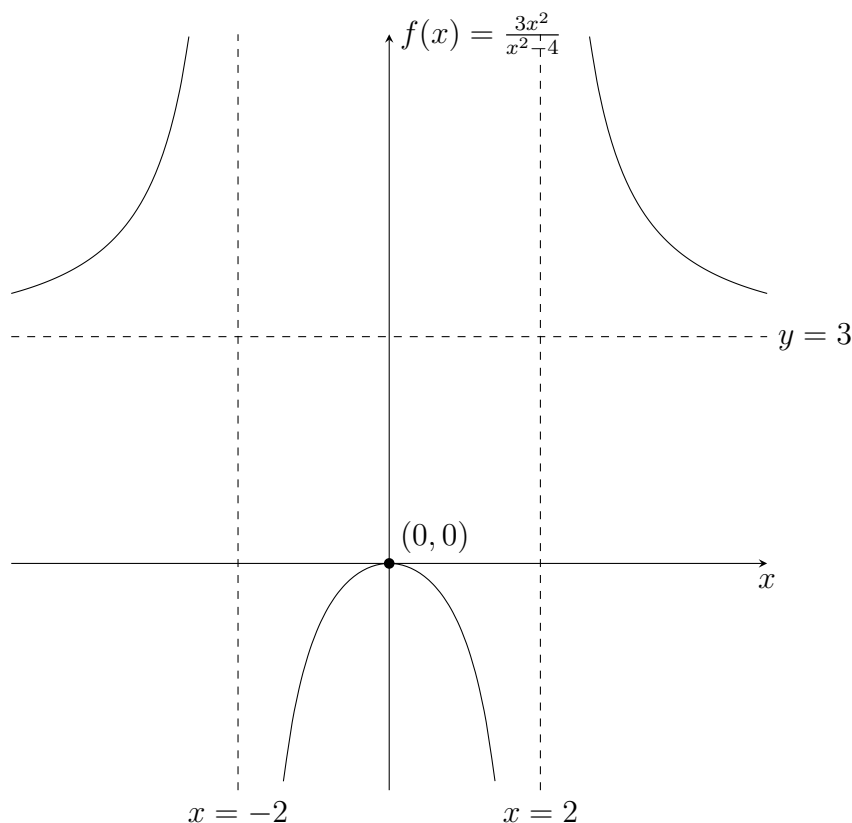


FIGURE 12. Roots, critical points, and inflection points.

- (1) Determine the domain. If the domain is not explicitly stated then we take it to be the set of x for which the formula is well defined. In the example, f is a rational function so the domain is the set of x where the denominator is nonzero: $D = \mathbb{R} \setminus \{-2, 2\} = (-\infty, -2) \cup (-2, 2) \cup (2, \infty)$.
- (2) Determine whether the function has any symmetry that should be taken into account: is it even, odd, or periodic? The example is an even function, so the graph on $(-\infty, 0]$ will be the mirror image of the graph on $[0, \infty)$.
- (3) Find the x - and y -intercepts and plot these points on the graph. In the example the only intercept occurs at the origin.
- (4) Find the asymptotes and draw them as dashed lines; determine the corresponding limits and draw these “arms” of the graph. In the example there are vertical asymptotes at ± 2 since the denominator vanishes and the numerator does not,

and we have

$$\lim_{x \rightarrow -2^+} \frac{3x^2}{x^2 - 4} = \lim_{x \rightarrow 2^-} \frac{3x^2}{x^2 - 4} = -\infty \quad \text{and} \quad \lim_{x \rightarrow -2^-} \frac{3x^2}{x^2 - 4} = \lim_{x \rightarrow 2^+} \frac{3x^2}{x^2 - 4} = +\infty,$$

as shown; note also that factoring the denominator reveals that $f(x)$ is positive on $(-\infty, -2) \cup (2, \infty)$, and negative on $(-2, 2)$. There is a horizontal asymptote at $y = 3$ because

$$\lim_{x \rightarrow \infty} \frac{3x^2}{x^2 - 4} = \lim_{x \rightarrow \infty} \frac{3}{1 - \frac{4}{x^2}} = 3,$$

and similarly for $\lim_{x \rightarrow -\infty} f(x)$. Note that the function approaches the asymptote from above because $3x^2 > 3(x^2 - 4)$ and thus $\frac{3x^2}{x^2 - 4} > 3$ whenever $x^2 - 4 > 0$.

- (5) Use f' to find the critical points of f and determine on which intervals f is increasing or decreasing. In the example, we have

$$f'(x) = \frac{6x(x^2 - 4) - 3x^2(2x)}{(x^2 - 4)^2} = \frac{-24x}{(x^2 - 4)^2},$$

so f is increasing on $(-\infty, -2)$ and $(-2, 0)$ – note that since -2 is not part of the domain, we cannot say “ f is increasing on $(-\infty, 0)$ ” – and decreasing on $(0, 2)$ and $(2, \infty)$. The only critical point is $x = 0$.

- (6) Compute f'' at the critical points (or use some other technique) to find the local maxima, local minima, and inflection points. In the example we have

$$f''(x) = \frac{(x^2 - 4)^2(-24) - (-24x) \cdot (2x)2(x^2 - 4)}{(x^2 - 4)^4} = 24 \frac{4x^2 - (x^2 - 4)}{(x^2 - 4)^3} = 24 \frac{3x^2 + 4}{(x^2 - 4)^3}$$

and thus in particular $f''(0) = 24 \cdot 4/(-4)^3 = -\frac{3}{2} < 0$, so 0 is a local maximum. Alternately we could observe that f is increasing on $(-2, 0)$ and decreasing on $(0, 2)$, which is enough to conclude that 0 is a local maximum without computing f'' .

- (7) Use f'' to determine convexity and concavity. The above computation shows that f'' takes the same sign as $x^2 - 4$, and thus f is convex on $(-\infty, -2)$, concave on $(-2, 2)$, and convex on $(2, \infty)$.
- (8) Use the information about local extrema, inflection points, monotonicity, and convexity to connect the dots and arms from the initial sketch.

Lecture 30 Curves, optimization, and Newton's method

DATE: MONDAY, NOVEMBER 4

Stewart §4.7 and §4.8

30.1. A curve sketching example

We mention one more possibility that should be looked for when sketching the graph of a function. Consider the example $f(x) = \frac{x^2}{x+1}$. The domain is $(-\infty, -1) \cup (-1, \infty)$ and the function has no symmetry; the only intercept is at the origin, and there is a

vertical asymptote at $x = -1$. There is no horizontal asymptote, but it turns out that there *is* a “slant asymptote”, or “oblique asymptote”. To explain this, we first observe that $f(x)$ is written as a sort of “improper fraction”, where the numerator has larger degree than the denominator. We can use polynomial long division to rewrite $f(x)$ as a polynomial plus a rational function in “proper fraction” form, where the numerator has smaller degree than the denominator, as follows.

$$\begin{array}{r}
 x - 1 \\
 x + 1 \overline{) x^2} \\
 \underline{-x^2 - x} \\
 -x \\
 \underline{x + 1} \\
 1
 \end{array}$$

From this we conclude that $\frac{x^2}{x+1} = x - 1 + \frac{1}{x+1}$, and in particular

$$\lim_{x \rightarrow \pm\infty} (f(x) - (x - 1)) = \lim_{x \rightarrow \pm\infty} \frac{1}{x + 1} = 0.$$

Thus $f(x)$ is asymptotic to the line $y = x - 1$. More generally, we say that $y = mx + b$ is a slant asymptote, or oblique asymptote, for a function f if $f(x) - (mx + b)$ goes to 0 at ∞ or $-\infty$.

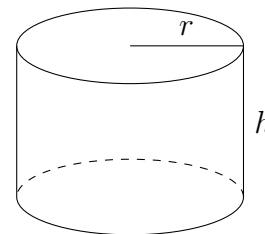
30.2. Optimization problems

Fermat’s theorem, which says that local maxima and minima of a function on a closed interval must occur at critical points or at endpoints, provides a good tool for finding extreme points. To apply this in examples, we must first translate the problem into a question of maximizing or minimizing a function on an interval.

Example 30.1. I want to make a five-gallon bucket that is cylindrical and has an open top. How much material do I need, assuming the thickness of the material is predetermined?

Step 1 is to understand the problem and identify the various quantities in play: which quantities are given, and which are unknown? In the example, the two most important quantities are the volume of the bucket and the surface area of the bucket (which determines the amount of material I use).

Step 2 (which can be done in conjunction with Step 1) is to draw a diagram illustrating the situation. In the example, when we draw the cylinder, we see that both volume and surface area are determined by the height of the cylinder and the radius of the circle that forms its base, so these two quantities will also enter our analysis.



Step 3 is to set up notation for the various quantities in play, and to identify which quantity we need to maximize or minimize in order to solve the problem. In the example, we can write V for volume, A for surface area, h for height, and r for radius; then our goal is to minimize A , because we are trying to construct the bucket using the smallest amount of material possible.

Step 4 is to write the quantity to be optimized as a function of the other variables. In the example we see that $A = \pi r^2 + 2\pi r h$, where the first term represents the area of the base, and the second term is the area of the sides of the cylinder.

Step 5 is to use the relationship between the other variables to write the quantity to be optimized as a function of a *single* variable. This step is necessary because usually there are multiple other variables that are related by certain constraints that must be taken into account. In the example, the volume is fixed, so r and h are related by $\pi r^2 h = V$. Solving this for h gives $h = \frac{V}{\pi r^2}$, and thus

$$A = \pi r^2 + 2\pi r \frac{V}{\pi r^2} = \pi r^2 + \frac{2V}{r}.$$

Step 6 is to use our tools for finding extreme points to solve the problem. In the example we observe that

$$\frac{dA}{dr} = 2\pi r - \frac{2V}{r^2} = 0 \quad \Leftrightarrow \quad \pi r = \frac{V}{r^2} \quad \Leftrightarrow \quad r^3 = \frac{V}{\pi}$$

and thus there is a single critical point at $r = \sqrt[3]{V/\pi}$. Moreover, we have $\frac{dA}{dr} < 0$ to the left of this critical point, and $\frac{dA}{dr} > 0$ to its right, so this critical point is a global minimum. At this critical point we have

$$A = \pi r^2 + \frac{2V}{r} = \pi \left(\frac{V}{\pi}\right)^{2/3} + 2V \left(\frac{\pi}{V}\right)^{1/3} = \pi^{1/3} V^{2/3} + 2\pi^{1/3} V^{2/3} = 3\pi^{1/3} V^{2/3}.$$

Note that Step 5 required us to solve for h in terms of r . We could also have solved for r in terms of h , but this would have led to $r = \sqrt{\frac{V}{\pi h}}$ and thus

$$A = \pi \frac{V}{\pi h} + 2\pi \sqrt{\frac{V}{\pi h}} h = \frac{V}{h} + 2\sqrt{\pi V h},$$

which makes the following computations a little messier because of the square root (though we could certainly carry them out). An alternate approach is to use implicit differentiation: treating h as an (unknown) function of r and differentiating the formulas $A = \pi r^2 + 2\pi r h$ and $V = \pi r^2 h$ gives

$$\frac{dA}{dr} = 2\pi r + 2\pi h + 2\pi r \frac{dh}{dr} \quad \text{and} \quad 0 = \frac{dV}{dr} = 2\pi r h + \pi r^2 \frac{dh}{dr}.$$

Solving the second of these gives $\frac{dh}{dr} = -2h/r$. Using this in the first equation gives

$$\frac{dA}{dr} = 2\pi \left(r + h - r \cdot \frac{2h}{r} \right) = 2\pi(r + h - 2h) = 2\pi(r - h).$$

Thus the critical point occurs when $r = h$, and some simple algebra gives the answer.

30.3. Newton's method

Consider the equation $x^5 + x + 1 = 0$. We do not have an analogue of the quadratic formula available to find explicitly the solution(s) of this equation. On the other hand, we can consider the function $f(x) = x^5 + x + 1$ and observe that f is continuous and $f(-1) = -1 < 0 < 1 = f(0)$, so by the IVT there is at least one solution of $f(x) = 0$ in the interval $[-1, 0]$. Moreover, $f'(x) = 5x^4 + 1 > 0$ for all $x \in \mathbb{R}$, so f is increasing and thus 1-1, which means that this is the *only* solution of $f(x) = 0$.

How can we approximate this solution numerically? One approach would be to mimic the proof of the IVT: build a pair of bisection sequences whose mutual limit is the unique root, and then take some element of one of those sequences, which gives a good approximation. Here, though, we outline another approach: *Newton's method*.

Start by making a guess at the solution, and call it x_0 . For example, we might take $x_0 = -\frac{1}{2}$ since the desired value $f(x) = 0$ is midway between $f(-1) = -1$ and $f(0) = 1$. Then we consider the linear approximation to f at x_0 , which is given by

$$g_0(x) = f(x_0) + f'(x_0)(x - x_0).$$

While solving the equation $f(x) = 0$ is hard, solving $g_0(x) = 0$ is quite easy, and if g_0 is a good enough approximation to f , we may hope that the solutions of the two equations are close together. Thus we let x_1 be defined by

$$0 = g_0(x_1) = f(x_0) + f'(x_0)(x_1 - x_0) \quad \Rightarrow \quad x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

If we are lucky, then x_1 is a better approximation to the solution than x_0 is. If we want to keep improving, we can repeat the process: let $g_1(x) = f(x_1) + f'(x_1)(x - x_1)$ be the linear approximation to f at x_1 , let x_2 be the solution to $g_1(x_2) = 0$, and so on. Thus in general, we define linear functions $g_n(x)$ and approximate solutions x_n by

$$g_n(x) = f(x_n) + f'(x_n)(x - x_n) \quad \text{and} \quad g_n(x_{n+1}) = 0.$$

More succinctly, we define a sequence x_n by

$$f(x_n) + f'(x_n)(x_{n+1} - x_n) = 0 \quad \Rightarrow \quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

If we carry this out for the example $f(x) = x^5 + x + 1$ with the starting guess $x_0 = -\frac{1}{2} = -0.5$, then using the fact that $f'(x) = 5x^4 + 1$, we get

$$\begin{aligned} x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} = x_0 - \frac{x_0^5 + x_0 + 1}{5x_0^4 + 1} = \frac{5x_0^5 + x_0 - (x_0^5 + x_0 + 1)}{5x_0^4 + 1} \\ &= \frac{4x_0^5 - 1}{5x_0^4 + 1} = \frac{4(-2)^{-5} - 1}{5(-2)^{-4} + 1} = \frac{-9/8}{21/16} = -\frac{6}{7} \approx -0.85714\dots \\ x_2 &= \frac{4x_1^5 - 1}{5x_1^4 + 1} \approx \frac{4(-0.85714)^5 - 1}{5(-0.85714)^4 + 1} \approx \frac{-2.8506}{3.6989} \approx -0.77066\dots \\ x_3 &= \frac{4x_2^5 - 1}{5x_2^4 + 1} \approx \frac{-2.0874}{2.7637} \approx -0.75529\dots \\ x_4 &= \frac{4x_3^5 - 1}{5x_3^4 + 1} \approx \frac{-1.9832}{2.6271} \approx -0.75490\dots \\ x_5 &= \frac{4x_4^5 - 1}{5x_4^4 + 1} \approx \frac{-1.9806}{2.6238} \approx -0.75486\dots, \end{aligned}$$

and so on, where the fact that successive iterates do not change by much suggests that we are approaching the true solution.

Part III. Integrals

Lecture 31

Antiderivatives

DATE: WEDNESDAY, NOVEMBER 6

Stewart §4.9

Suppose we are interested in some function F , about which all we know is that its derivative is equal to some other function f . For example, this situation arises if we want to reconstruct a car's position over time based only on the information displayed on the speedometer; we want to know $s(t)$, the position at time t , and all we know at first is $v(t) = s'(t)$, the velocity at time t . This motivates the following definition.

Definition 31.1. A function F is an *antiderivative* of f on an interval $I \subset \mathbb{R}$ if $F'(x)$ exists and is equal to $f(x)$ for all $x \in I$.

Example 31.2. If $f(x) = 3x^2$, then $F(x) = x^3$ is an antiderivative of f on \mathbb{R} . So is $G(x) = x^3 + 10$. Indeed, for *any* $C \in \mathbb{R}$, the function $x \mapsto x^3 + C$ is an antiderivative of f .

This principle is quite general: if F is an antiderivative of f , then so is $x \mapsto F(x) + C$ for any constant C . In fact, this gives *all* the antiderivatives of f .

Theorem 31.3. *If F, G are antiderivatives of f on an interval I , then there is a constant C such that $G(x) = F(x) + C$ for every $x \in I$.*

Proof. By the definition of antiderivative, $(G - F)'(x) = G'(x) - F'(x) = f(x) - f(x) = 0$ for all $x \in I$, so by Theorem 22.1 (which follows quickly from the MVT), $G - F$ is constant on I . \square

Geometrically, any two antiderivatives F and G have graphs with the property that one is a vertical translate of the other.

We stress that Theorem 31.3 only works on an interval. For example, $F(x) = \ln|x|$ is an antiderivative of $f(x)$, and so is

$$G(x) = \begin{cases} 1 + \ln x & x > 0, \\ \ln|x| & x < 0, \end{cases}$$

but $G - F$ is not a constant. The problem here is that f, F, G are not defined at 0, so we are not working on a single interval, but on the union of two disjoint intervals $(-\infty, 0)$ and $(0, \infty)$. On each of these intervals, any two antiderivatives must differ by a constant, but if we jump to a different interval then we need to allow for a new constant.

In order to recover a specific antiderivative F , we need to know enough information about F to determine the constant. For example, if F is an antiderivative of $f(x) = 3x^2$ with the property that $F(0) = 5$, then from Example 31.2 and Theorem 31.3 we see that $F(x) = 3x^2 + C$ for some constant C , and evaluating $F(0) = 3 \cdot 0^2 + C = C$ we see that $C = 5$, so $F(x) = 3x^2 + 5$.

In some situations we may need to antidifferentiate more than once. For example, if we want to know the vertical position $s(t)$ of an object moving under the influence of gravity, then we are given not the velocity $v(t) = s'(t)$, but the acceleration $a(t) = v'(t) = s''(t)$. With a constant gravitational field imparting a downward acceleration $g > 0$, we have $a(t) = -g$. Antidifferentiating once gives $v(t) = -gt + C$ for some constant C , and another antidifferentiation gives $s(t) = -\frac{g}{2}t^2 + Ct + D$ for some constant D . Thus in order to determine $s(t)$, we need to determine both C and D , which requires *two* pieces of information. For example, it would suffice to know the position at two different moments in time, or both the position and the velocity at a single moment.

It is natural at this point to ask whether every function has an antiderivative. We may reasonably start by listing the various formulas we have so far for derivatives of common functions, and obtain the following table; note that in each case we only write a single antiderivative, all the others can be obtained by adding a constant.

function $f(x)$	antiderivative $F(x)$
x^n	$\frac{1}{n+1}x^{n+1}$
$\frac{1}{x}$	$\ln x $
e^x	e^x
$\cos x$	$\sin x$
$\sin x$	$-\cos x$
$\sec^2 x$	$\tan x$
$\sec x \tan x$	$\sec x$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x$
$\frac{1}{1+x^2}$	$\arctan x$

We could keep going and include the derivatives of the other trigonometric, inverse trigonometric, and hyperbolic functions, but it should quickly become apparent that this approach will only get us so far. For example, how are we to antidifferentiate something like $\ln x$? Does it even have an antiderivative?

First observe that since $(cF)' = c(F')$ for $c \in \mathbb{R}$, and $(F + G)' = F' + G'$, we can find an antiderivative of cf and $f + g$ provided we can antidifferentiate f and g individually.

Example 31.4. To find an antiderivative F of $f(x) = 2 \sin x + \frac{2x^3 - \sqrt{x}}{x}$, we can rewrite the function as $f(x) = 2 \sin x + 2x^2 - x^{-1/2}$ and find antiderivatives of each term individually, obtaining

$$F(x) = -2 \cos x + \frac{2}{3}x^3 - 2\sqrt{x} + C.$$

This lets us assemble the functions from the table into a broader class of functions with antiderivatives, but still does not address all the possibilities. At this point it is useful to stop thinking about the *formulas* that define functions, and start thinking about other ways of representing functions. In particular, we consider the graph of f , which is a subset of \mathbb{R}^2 . For concreteness, let $f(x) = x$, so that we know $F(x) = \frac{1}{2}x^2$

is an antiderivative. The graph of f is the line through the origin with slope 1, and we observe that for $x > 0$, the triangle with vertices at the origin, $(x, 0)$, and (x, x) has area $\frac{1}{2}x^2$. This triangle can be described in terms of the graph as the region that lies underneath the graph of f , above the x -axis, to the right of the y -axis, and to the left of the vertical line through $(x, 0)$. So for this function at least, we have interpreted an antiderivative as the area of a specific region.

More generally, suppose that f is a positive function, and define a function F by letting $F(t)$ be the area of the region that is bounded above by the graph of f , to the left by the y -axis, below by the x -axis, and to the right by the line $x = t$. If we compare $F(t)$ and $F(t + h)$, we see that they are the areas of two regions which differ only in whether the strip from t to $t + h$ is included. This strip is close to being a rectangle with width h and height $f(t)$, so its area is approximately $h \cdot f(t)$. Thus we have the rough estimate

$$(31.1) \quad \frac{F(t+h) - F(t)}{h} = \frac{\text{area of the strip from } t \text{ to } t+h}{h} \approx \frac{h \cdot f(t)}{h} = f(t).$$

This suggests that as $h \rightarrow 0$, we may reasonably expect the difference quotient $\frac{F(t+h)-F(t)}{h}$ to converge to $f(t)$, which would show that $F'(t) = f(t)$.

There are a number of steps in this procedure that need to be made more precise.

- (1) What exactly do we mean by “area”? How do we compute the area of a region that is irregular in shape?
- (2) What does it mean to say that “this strip is close to being a rectangle”? Presumably we want some condition that guarantees that the approximation in (31.1) gets better and better as $h \rightarrow 0$; how do we guarantee this?

Obtaining precise and rigorous answers to these questions requires us to develop the theory of integration, which we do over the next few lectures. In a nutshell, the answers are as follows.

- (1) We can introduce a precise definition of *definite integral* that plays the role of area for the regions we are concerned with, provided f is “nice enough”. This leads to the notion of an *integrable function*.
- (2) The approximation in (31.1) leads to correct conclusions whenever f is continuous. In particular, continuous functions are integrable, and every continuous function has an antiderivative that is produced via the procedure described above. This last fact is known as the *Fundamental Theorem of Calculus*.

Lecture 32

Approximating areas by sums

DATE: FRIDAY, NOVEMBER 8

Stewart §5.1, Spivak Chapter 13

In the next few lectures we develop the basic results of the theory of integration, including the definition of integrals and of integrable functions, the observation that continuous functions are integrable, and the Fundamental Theorem of Calculus that

relates integration and differentiation. The treatment here is more theoretical and complete than that given in Stewart's book, but not as comprehensive as the one in Spivak's book. The order of the results, and the method of certain proofs, has been chosen to take the most efficient route that covers everything I want to say but does not bombard you with too much extra theory. In a number of places I have also borrowed ideas and proofs from a book by Pete L. Clark,³⁰ which is closer to Spivak than to Stewart in its style but also has quite a few differences from Spivak.

32.1. Area and Riemann sums

Given a positive function f , last time we had the idea of producing an antiderivative F for f by letting $F(t)$ be the “area” of the region in \mathbb{R}^2 bounded by the x -axis, the graph of f , and the vertical lines $x = 0$ and $x = t$. But what do we mean by “area” of a region in \mathbb{R}^2 ? We start with three axioms that any reasonable notion of area should obey:

- (1) The area of a rectangle is the product of its width and its height.
- (2) If two regions X, Y do not overlap, or if their overlap has zero area, such as a straight line, then $\text{area}(X \cup Y) = \text{area}(X) + \text{area}(Y)$.
- (3) If a region X is contained inside a region Y , then $\text{area}(X) \leq \text{area}(Y)$.

Let us write $\Gamma(f, a, b)$ for the region in \mathbb{R}^2 bounded by the x -axis, the graph of f , and the vertical lines $x = a$ and $x = b$. Let $\int_a^b f = \text{area}(\Gamma(f, a, b))$, for some reasonable notion of area. Then the three axioms above imply the following.

- (1) If there is $C \in \mathbb{R}$ such that $f(x) = C$ for every $x \in [a, b]$, then $\int_a^b f = C(b - a)$.
- (2) If $a < c < b$, then $\int_a^b f = \int_a^c f + \int_c^b f$.
- (3) If $f(x) \leq g(x)$ for every $x \in [a, b]$, then $\int_a^b f \leq \int_a^b g$.

These axioms motivate the remainder of our discussion: we want to associate to every “reasonable” function $f: [a, b] \rightarrow \mathbb{R}$ a number $\int_a^b f$ in such a way that the properties above are satisfied. We will further clarify the meaning of “reasonable” later on.

Before making any formal definitions, we describe the general principle, which is to approximate $\Gamma(f, a, b)$ with a union of rectangles: for some large $n \in \mathbb{N}$ we choose points $a = x_0 < x_1 < \cdots < x_n = b$, and over each interval $[x_{i-1}, x_i]$ we draw a rectangle whose height is roughly equal to the value of the function on that interval. Adding up the areas of the rectangles gives us a number that we hope is a good approximation to the area of $\Gamma(f, a, b)$.

The set of points $\{x_i\}_{i=0}^n$ is said to *partition* the interval $[a, b]$ into smaller subintervals, and we often write $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ to denote a specific partition. One way to choose the heights of the rectangles used in the approximation is to pick a point t_i inside each subinterval $[x_{i-1}, x_i]$, and use $f(t_i)$ as the height of the corresponding rectangle. We write τ to denote the set of points (t_1, t_2, \dots, t_n) , and refer to the pair (P, τ) as a *tagged partition*. Each tagged partition corresponds to an approximation of $\Gamma(f, a, b)$ by rectangles, and the first two axioms tell us that the area of this

³⁰“Honors Calculus” by Pete L. Clark, Univ. of Georgia, <http://math.uga.edu/~pete/2400full.pdf>

approximation is

$$(32.1) \quad R(f, P, \tau) := \sum_{i=1}^n f(t_i)(x_i - x_{i-1}).$$

The notation $R(f, P, \tau)$ should be read as “the *Riemann sum* of f associated to the partition P with tags τ ”. The idea that drives the *Riemann integral* is that as we choose tagged partitions for which the subinterval lengths $x_i - x_{i-1}$ get smaller and smaller, the corresponding Riemann sums should converge to a limit, which we will denote $\int_a^b f$. Making this precise, and giving conditions under which the limit actually exists, will require some nontrivial effort, and first we give an example that illustrates the idea.

Example 32.1. Consider $\int_0^b f$ when $f(x) = x$. Write P for the partition of $[0, b]$ into n subintervals of equal length $\frac{b}{n}$, and τ_{right} for tags chosen to be the right-hand endpoint of each subinterval. Thus $x_i = t_i = \frac{ib}{n}$, and the corresponding Riemann sum is

$$R(f, P, \tau_{\text{right}}) = \sum_{i=1}^n \frac{ib}{n} \cdot \frac{b}{n} = \frac{b^2}{n^2} \sum_{i=1}^n i = \frac{b^2}{n^2} \frac{n(n+1)}{2} = \frac{b^2}{2} \left(1 + \frac{1}{n}\right),$$

where we use the formula $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. As $n \rightarrow \infty$, the length of the subintervals in the partition P goes to 0, and the Riemann sum converges to $\frac{b^2}{2}$, which we know from elementary geometry to be the area of the right triangle $\Gamma(f, 0, b)$. But what if we had chosen a different choice of tags? Would we still get the same conclusion? For example, suppose we let τ_{left} denote tags chosen at the left endpoint of each subinterval, so $t_i = x_{i-1} = \frac{(i-1)b}{n}$. Then we have

$$R(f, P, \tau_{\text{left}}) = \sum_{i=1}^n \frac{(i-1)b}{n} \cdot \frac{b}{n} = \frac{b^2}{n^2} \sum_{i=1}^n (i-1) = \frac{b^2}{n^2} \frac{(n-1)n}{2} = \frac{b^2}{2} \left(1 - \frac{1}{n}\right),$$

and once again we see that the Riemann sums converge to the appropriate limit. Of course, there are many other choices of tags we could make as well, and we could also have chosen different partitions, in which the subintervals did not have equal length. But it turns out that all of them lead to the same limit. Let us explain why the choice of tags does not affect the limit; we will postpone a discussion of unequal subintervals until later. Given the equal-length partition P into n subintervals, since f is increasing we have $f(x_{i-1}) \leq f(x) \leq f(x_i)$ for every $x \in [x_{i-1}, x_i]$. It follows that for *any* choice of tags τ , we will have

$$(32.2) \quad \frac{b^2}{2} \left(1 - \frac{1}{n}\right) \leq R(f, P, \tau_{\text{left}}) \leq R(f, P, \tau) \leq R(f, P, \tau_{\text{right}}) \leq \frac{b^2}{2} \left(1 + \frac{1}{n}\right).$$

By the squeeze theorem, we get the same limit no matter which tags we choose.

Lecture 33 Lower sums, upper sums, and integrals

DATE: MONDAY, NOVEMBER 11

33.1. Lower and upper sums

In general we need to deal with functions that are not necessarily increasing, but we can still obtain bounds similar to those in (32.2) as follows. Suppose that P is a partition of $[a, b]$ with endpoints x_i , and that $m_i, M_i \in \mathbb{R}$ have the property that $m_i \leq f(x) \leq M_i$ for all $x \in [x_{i-1}, x_i]$. Then for every choice of tags τ , we have $m_i \leq f(t_i) \leq M_i$, and thus

$$(33.1) \quad \sum_{i=1}^n m_i(x_i - x_{i-1}) \leq \sum_{i=1}^n f(t_i)(x_i - x_{i-1}) = R(f, P, \tau) \leq \sum_{i=1}^n M_i(x_i - x_{i-1}).$$

The first and last sums in (33.1) are not necessarily Riemann sums themselves, because there may be no choice of tags that gives $f(t_i) = m_i$ or $f(t_i) = M_i$. Nevertheless, these expressions turn out to be extremely useful for studying integrals and for developing the formal theory, because they allow us to invoke the third area axiom, which we have so far neglected, and observe that for each $1 \leq i \leq n$, we have

$$m_i(x_i - x_{i-1}) = \int_{x_{i-1}}^{x_i} m_i \leq \int_{x_{i-1}}^{x_i} f \leq \int_{x_{i-1}}^{x_i} M_i = M_i(x_i - x_{i-1}).$$

Here the two equalities use the first area axiom, and the two inequalities use the third axiom. Summing over i and using the second axiom gives

$$(33.2) \quad \sum_{i=1}^n m_i(x_i - x_{i-1}) \leq \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f = \int_a^b f \leq \sum_{i=1}^n M_i(x_i - x_{i-1}).$$

By comparing (33.1) and (33.2), we see that $\int_a^b f$ and $R(f, P, \tau)$ are both contained in the interval between the lower sum $\sum m_i(x_i - x_{i-1})$ and the upper sum $\sum M_i(x_i - x_{i-1})$. Intuitively, we would like to say that if we can make these two sums be arbitrarily close together by choosing x_i, m_i, M_i appropriately, then this gives a definition of the integral $\int_a^b f$ and shows that it is the limit of the Riemann sums.

33.2. The least upper bound property

To make this discussion precise, we need the following general notions.

Definition 33.1. An *upper bound* for a set $A \subset \mathbb{R}$ is a number $M \in \mathbb{R}$ such that $x \leq M$ for every $x \in A$. We say that M is the *least upper bound* for A if every $M' < M$ is *not* an upper bound for A . In this case we also call M the *supremum* of A , and write $M = \sup A$. Equivalently, $M = \sup A$ if and only if the following are both true:

- (1) $x \leq M$ for every $x \in A$;
- (2) for every $M' < M$, there is $x \in A$ such that $x > M'$.

If we reverse the direction of all inequalities in this definition, we get the definition of *lower bound* and *greatest lower bound*, also called *infimum* and denoted $\inf A$.

Example 33.2. $\sup[0, 1] = 1$ and $\inf[0, 1] = 0$, which suggests that \sup and \inf should be thought of as \max and \min , respectively. The difference is that the maximum of a set A must lie in A , while the supremum is not required to. Thus the open interval $(0, 1)$ has no \max or \min , while $\sup(0, 1) = 1$ and $\inf(0, 1) = 0$ still exist.

Recall that in an earlier lecture, we said that the defining characteristic of the real numbers (as compared to \mathbb{Q}) is the fact that every increasing sequence that is bounded above has a limit. This implies the following property.

Theorem 33.3 (Least Upper Bound property). *If $A \subset \mathbb{R}$ is bounded above, then it admits a least upper bound in \mathbb{R} .*

Proof. Let b be an upper bound for A , and choose any $a \in A$. Define a pair of bisection sequences $a = a_1 \leq a_2 \leq a_3 \leq \cdots \leq b_3 \leq b_2 \leq b_1 = b$ by assigning each midpoint to b_n if it is an upper bound for A , and a_n if it is not. Thus every b_n is an upper bound for A , and no a_n is an upper bound for A . As usual, $c = \lim a_n = \lim b_n$ exists.

Given any $x \in A$, we have $x \leq b_n$ for every n , and thus $c = \lim b_n \geq x$. Thus c is an upper bound for A . Moreover, given any $c' < c$ there is $a_n \in (c', c)$, and since a_n is not an upper bound for A , neither is c' . Thus c is the least upper bound for A . \square

This property fails in \mathbb{Q} : the set $A = \{p/q \in \mathbb{Q} : (p/q)^2 < 2\}$ is bounded above, but has no least upper bound in \mathbb{Q} . In \mathbb{R} , on the other hand, $\sqrt{2}$ is the least upper bound.

Exercise 33.4. Suppose we know that \mathbb{R} has the least upper bound property, but do not yet know whether every bounded increasing sequence has a limit. Prove that if x_n is a bounded increasing sequence, then $\lim x_n$ exists and is equal to $\sup\{x_n : n \in \mathbb{N}\}$. This says that our description of \mathbb{R} in terms of monotone convergence is equivalent to the description in terms of the least upper bound property, which is what Spivak uses.

Exercise 33.5. Use the least upper bound property to deduce that every set that is bounded below admits a greatest lower bound.

33.3. Lower and upper integrals

Returning to our discussion of $\int_a^b f$, which so far we have studied but not defined, we begin to make some formal definitions. In what follows, we require f to be bounded, since if a function can take arbitrarily large values then there is no reason to expect the associated area to be finite. However, we do not require f to be positive. If f is a negative function, then we should think of $\int_a^b f$ as a negative number whose absolute value is the area lying between the graph of f and the x -axis. If f takes both signs, then $\int_a^b f$ represents the difference between the area above the x -axis and the area below it.

Definition 33.6. As in the previous lecture, a *partition* of $[a, b]$ is a finite set $P \subset [a, b]$ that contains a and b . We will always write the elements of P in increasing order, so $P = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$.

Given a bounded function $f: [a, b] \rightarrow \mathbb{R}$ and a partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a, b]$, consider for each $1 \leq i \leq n$ the quantities

$$m_i(f, P) = \inf\{f(x) : x \in [x_{i-1}, x_i]\} \quad \text{and} \quad M_i(f, P) = \sup\{f(x) : x \in [x_{i-1}, x_i]\}.$$

Note that the inf and sup exist because f is bounded above and below. (They may not be achieved, though; there may not be any $x \in [x_{i-1}, x_i]$ satisfying $f(x) = m_i(f, P)$ or $f(x) = M_i(f, P)$.) Define the *lower and upper sums* of f for P by

$$L(f, P) = \sum_{i=1}^n m_i(f, P)(x_i - x_{i-1}) \quad \text{and} \quad U(f, P) = \sum_{i=1}^n M_i(f, P)(x_i - x_{i-1}).$$

Recalling (33.2), we see that if we are ever to successfully define $\int_a^b f$ so that the three desired axioms are satisfied, then we must have

$$(33.3) \quad L(f, P) \leq \int_a^b f \leq U(f, P)$$

for every partition P of $[a, b]$. In other words $\int_a^b f$ should be simultaneously an upper bound for the set of all lower sums $L(f, P)$, and a lower bound for the set of all upper sums $U(f, P)$. This implies that

$$\underbrace{\sup\{L(f, P) : P \text{ is a partition of } [a, b]\}}_{\int_a^b f} \leq \int_a^b f \leq \underbrace{\inf\{U(f, P) : P \text{ is a partition of } [a, b]\}}_{\overline{\int}_a^b f}.$$

We will refer to the quantities $\int_a^b f$ and $\overline{\int}_a^b f$ as the *lower and upper integrals* of f , respectively. The lower integral can be thought of as the area of $\Gamma(f, a, b)$ that we can detect by approximating it from inside by rectangles; the upper integral is the area that we can detect by approximating it from outside by rectangles. Now we come to the crucial definition.

Definition 33.7. A function $f: [a, b] \rightarrow \mathbb{R}$ is *integrable* if it is bounded and if $\int_a^b f = \overline{\int}_a^b f$. In this case the common value is called the *integral* of f on $[a, b]$ and is denoted $\int_a^b f$.

Remark 33.8. The quantity $\int_a^b f$ is often referred to as the *definite integral* to emphasize that it is a single number associated to a definite interval $[a, b]$. It is often denoted as $\int_a^b f(x) dx$, especially when we write the function f explicitly; for example, if $f(x) = x^2$, then we would write $\int_a^b f = \int_a^b x^2 dx$. This has the same meaning as $\int_a^b t^2 dt$, and as $\int_a^b y^2 dy$, and so on. The symbol “ \int ” is called an *integral sign* and can be thought of as an elongated “S” (since integration is related to summation). The function f is called the *integrand*, and the values a, b are called the *limits of integration*.³¹

Example 33.9. With Example 32.1 in mind, let us show that $f(x) = x^2$ is integrable on $[0, b]$, and compute $\int_0^b x^2 dx$. Once again, we consider the partition $P_n = \{0 < \frac{b}{n} < \frac{2b}{n} < \dots < \frac{(n-1)b}{n} < b\}$, and observe that since f is increasing, we have $m_i = f(x_{i-1}) = (\frac{(i-1)b}{n})^2$ and $M_i = f(x_i) = (\frac{ib}{n})^2$. Thus

$$L(f, P_n) = \sum_{i=1}^n \left(\frac{(i-1)b}{n}\right)^2 \frac{b}{n} = \frac{b^3}{n^3} (0^2 + 1^2 + 2^2 + 3^2 + \dots + (n-1)^2),$$

$$U(f, P_n) = \sum_{i=1}^n \left(\frac{ib}{n}\right)^2 \frac{b}{n} = \frac{b^3}{n^3} (1^2 + 2^2 + 3^2 + 4^2 + \dots + n^2).$$

It can be proved by induction that $1^2 + 2^2 + \dots + n^2 = \frac{1}{6}n(n+1)(2n+1)$ – this is a good exercise to do if you haven’t before – and thus

$$L(f, P_n) = \frac{b^3}{n^3} \cdot \frac{1}{6}(n-1)n(2n-1), \quad U(f, P_n) = \frac{b^3}{n^3} \cdot \frac{1}{6}n(n+1)(2n+1).$$

³¹Note that this has nothing to do with our usual use of the word “limit”.

In particular, as $n \rightarrow \infty$, both $L(f, P_n)$ and $U(f, P_n)$ converge to $\frac{b^3}{3}$, and we conclude that $\int_0^b x^2 dx = \frac{b^3}{3}$.

It is worth pointing out that not every bounded function is integrable.

Exercise 33.10. Define a function $f: [0, 1] \rightarrow \mathbb{R}$ by $f(x) = 0$ if x is irrational, and $f(x) = 1$ if x is rational. Prove that $\int_0^1 f = 0$ and $\overline{\int}_0^1 f = 1$, and thus f is not integrable.

We end this lecture by observing that lower and upper integrals both satisfy the three desired axioms. Two of these are easy and we leave them as exercises; the remaining axiom takes a short proof.

Exercise 33.11. Prove that $\int_a^b C = \overline{\int}_a^b C = C(b - a)$ for every $a < b$ and $C \in \mathbb{R}$.

Exercise 33.12. Prove that if $f, g: [a, b] \rightarrow \mathbb{R}$ are bounded functions satisfying $f(x) \leq g(x)$ for all $x \in [a, b]$, then $\int_a^b f \leq \int_a^b g$ and $\overline{\int}_a^b f \leq \overline{\int}_a^b g$.

Proposition 33.13. *If $f: [a, b] \rightarrow \mathbb{R}$ is bounded and $c \in [a, b]$, then $\int_a^b f = \int_a^c f + \int_c^b f$ and $\overline{\int}_a^b f = \overline{\int}_a^c f + \overline{\int}_c^b f$.*

The following proof was omitted in the classroom lecture

Proof. We give the proof for \int and leave the other half as an exercise. First observe that if Q and R are any partitions of $[a, c]$ and $[c, b]$, respectively, then $P := Q \cup R$ is a partition of $[a, b]$, and

$$L(f, P) = L(f, Q) + L(f, R).$$

Every partition P of $[a, b]$ that contains c arises in this way, and so taking a supremum over all Q and R gives

$$(33.4) \quad \underbrace{\sup\{L(f, P) : P \text{ is a partition of } [a, b] \text{ and } c \in P\}}_I = \int_a^c f + \int_c^b f.$$

The quantity I is clearly $\leq \int_a^b f$. To conclude the proof we need to consider partitions of $[a, b]$ that do *not* contain c , and use the following lemma to see what happens when the point c is added to the partition.

Lemma 33.14. *If $f: [a, b] \rightarrow \mathbb{R}$ is bounded and P is a partition of $[a, b]$, then for every $c \in [a, b]$ we have $L(f, P) \leq L(f, P \cup \{c\})$ and $U(f, P) \geq U(f, P \cup \{c\})$.*

Proof. We prove the inequality for L ; the upper sums are left as an exercise. Let $P = \{a = x_0 < x_1 < \dots < x_n = b\}$ and choose $k \in \{1, \dots, n\}$ such that $c \in [x_{k-1}, x_k]$. Let

$$m_c^\ell := \inf\{f(x) : x \in [x_{k-1}, c]\} \quad \text{and} \quad m_c^r := \inf\{f(x) : x \in [c, x_k]\}.$$

Then we have

$$L(f, P \cup \{c\}) = \left(\sum_{i \neq k} m_i(f, P)(x_i - x_{i-1}) \right) + m_c^r(x_k - c) + m_c^\ell(c - x_{k-1})$$

$$\begin{aligned} &\geq \left(\sum_{i \neq k} m_i(f, P)(x_i - x_{i-1}) \right) + m_k(f, P)((x_k - c) + (c - x_{k-1})) \\ &= \sum_{i=1}^n m_i(f, P)(x_i - x_{i-1}) = L(f, P), \end{aligned}$$

which proves the lemma. \square

Returning to the proof of Proposition 33.13, we use Lemma 33.14 to deduce that

$$\begin{aligned} \int_a^b f &= \sup\{L(f, P) : P \text{ is a partition of } [a, b]\} \\ &\leq \underbrace{\sup\{L(f, P \cup \{c\}) : P \text{ is a partition of } [a, b]\}}_I \leq \int_a^b f \end{aligned}$$

and thus $I = \int_a^b f$; together with (33.4), this proves the proposition. \square

Lecture 34 The Fundamental Theorem of Calculus

DATE: WEDNESDAY, NOVEMBER 13

Stewart §5.3 and §5.4, Spivak Chapters 13 and 14

34.1. The Fundamental Theorem of Calculus

Now we address two important questions.

- Which functions are integrable?
- Does integration in fact give us an antiderivative, as we originally hoped?

We start with the following result: it is a little bit clunky because it works with \int and $\bar{\int}$, but it quickly implies three very important results that have cleaner formulations.

Theorem 34.1 (Fundamental Theorem of Calculus, Part 0). *Let $f: [a, b] \rightarrow \mathbb{R}$ be a bounded function, and define $L, U: [a, b] \rightarrow \mathbb{R}$ by $L(x) := \int_a^x f$ and $U(x) := \bar{\int}_a^x f$. If f is continuous at a point $c \in (a, b)$, then L and U are both differentiable at c , and satisfy $L'(c) = U'(c) = f(c)$.*

Proof. We prove the statement regarding $L(x)$; the proof for $U(x)$ is identical, since all we use are the properties of \int and $\bar{\int}$ from Proposition 33.13. Fix $c \in (a, b)$ such that f is continuous at c . Given $h > 0$, consider the quantities

$$m_h := \inf\{f(t) : c \leq t \leq c + h\} \quad \text{and} \quad M_h := \sup\{f(t) : c \leq t \leq c + h\}.$$

It follows from the first two statements in Proposition 33.13 that

$$m_h \cdot h = \int_c^{c+h} m_h \leq \int_c^{c+h} f \leq \int_c^{c+h} M_h = M_h \cdot h.$$

The third statement in Proposition 33.13 gives

$$L(c+h) = \int_{\underline{a}}^{c+h} f = \int_{\underline{a}}^c f + \int_{\underline{c}}^{c+h} f = L(c) + \int_{\underline{c}}^{c+h} f.$$

Combining these, we get

$$m_h \leq \frac{L(c+h) - L(c)}{h} \leq M_h,$$

and a similar argument shows that it remains true when $h < 0$. Because f is continuous, we have $\lim_{h \rightarrow 0} m_h = \lim_{h \rightarrow 0} M_h = f(c)$, and so the squeeze theorem implies that $\lim_{h \rightarrow 0} \frac{1}{h}(L(c+h) - L(c))$ exists and is equal to $f(c)$, which proves the theorem. \square

Remark 34.2. The only properties of \int and $\bar{\int}$ that we used in Theorem 34.1 were the three axioms; in particular, we did not need any knowledge of how these quantities are computed via partitions.

Theorem 34.1 has several tremendously important consequences.

Corollary 34.3. *If $f: [a, b] \rightarrow \mathbb{R}$ is continuous, then it is integrable.*³²

Proof. By Theorem 34.1, $L(x) := \int_{\underline{a}}^x f$ and $U(x) := \bar{\int}_{\underline{a}}^x f$ both define antiderivatives for f on $[a, b]$. By Theorem 22.1, this implies that $L - U$ is constant. But $L(a) = 0 = U(a)$, so we must have $L(x) = U(x)$ for all $x \in [a, b]$. In particular, $\int_{\underline{a}}^b f = L(b) = U(b) = \bar{\int}_{\underline{a}}^b f$, so f is integrable. \square

Corollary 34.4 (Fundamental Theorem of Calculus, Part 1). *If $f: [a, b] \rightarrow \mathbb{R}$ is integrable and is continuous at a point $x \in (a, b)$, then $F(x) := \int_{\underline{a}}^x f$ is differentiable at x and satisfies $F'(x) = f(x)$.*

In particular, to our original question about antiderivatives, we can now answer that every continuous function $f: [a, b] \rightarrow \mathbb{R}$ has an antiderivative, given by $F(x) = \int_{\underline{a}}^x f$. Moreover, the correspondence goes both ways: if we happen to know an antiderivative of f , we can use it to compute the definite integral.

Example 34.5. We can use Corollary 34.4 together with the chain rule to differentiate functions where the upper limit of integration depends on x in a more complicated way:

$$\begin{aligned} \frac{d}{dx} \int_1^{x^2} \sqrt{1+t} dt &= \frac{d}{dx} \int_1^u \sqrt{1+t} dt && (u = x^2) \\ &= \frac{du}{dx} \frac{d}{du} \int_1^u \sqrt{1+t} dt = (2x)\sqrt{1+u} = 2x\sqrt{1+x^2}. \end{aligned}$$

Corollary 34.6 (Fundamental Theorem of Calculus, Part 2). *If $f: [a, b] \rightarrow \mathbb{R}$ is continuous and $F: [a, b] \rightarrow \mathbb{R}$ is an antiderivative of f , then $\int_a^b f = F(b) - F(a)$.*

³²This theorem is proved in multiple ways in Spivak's book and in Clark's. To my mind, the proof here, which follows pages 292–293 in Spivak, is the simplest, but it does require setting up the FTC in a slightly non-standard way at first – usually the FTC is not stated for lower and upper integrals, but only for the integral itself – and there are certainly other ways of proving this result.

Proof. By the previous corollary, $G(x) := \int_a^x f$ is an antiderivative of f , so $G - F$ is constant. Thus $G(b) - F(b) = G(a) - F(a) = -F(a)$, since $G(a) = \int_a^a f = 0$, which gives $G(b) = F(b) - F(a)$ and proves the corollary. \square

Example 34.7. Corollary 34.6 gives us a faster way to evaluate the integrals $\int_0^b x \, dx$ and $\int_0^b x^2 \, dx$ from Examples 32.1 and 33.9:

$$\int_0^b x \, dx = \frac{1}{2}x^2 \Big|_0^b = \frac{1}{2}b^2 - \frac{1}{2}0^2 = \frac{b^2}{2},$$

$$\int_0^b x^2 \, dx = \frac{1}{3}x^3 \Big|_0^b = \frac{1}{3}b^3 - \frac{1}{3}0^3 = \frac{b^3}{3}.$$

Here the notation $F(x)|_a^b$ is shorthand for $F(b) - F(a)$, and is especially convenient to use when the expression for $F(x)$ is quite complicated.

Remark 34.8. Corollary 34.6 actually remains true under the weaker assumption that f is integrable, but this requires a different proof using the MVT; see Theorem 14.2 in Spivak's book for details.

Remark 34.9. With a little more effort one can show that if $f: [a, b] \rightarrow \mathbb{R}$ is bounded and is continuous everywhere except at a finite set of points, then it is integrable; see §8.3.2 in Clark's book.

Remark 34.10. With a lot more effort one can give a complete characterization of which bounded functions are integrable: see §8.5 in Clark's book.

Example 34.11. Before applying Corollary 34.6, one must check that the conditions are satisfied. A too-naive application of the FTC gives

$$\int_{-2}^1 \frac{1}{x^2} \, dx = -\frac{1}{x} \Big|_{-2}^1 = -\frac{1}{1} - \frac{1}{-2} = -\frac{1}{2},$$

despite the fact that $f \geq 0$ and so we expect to get a nonnegative integral. The issue is that we cannot apply Corollary 34.6 since $1/x^2$ is not continuous on the interval $[-2, 1]$; in fact, it is not even bounded, so it is not integrable.

34.2. Basic properties of integrals

The following properties of definite integrals are immediate consequences of the corresponding properties in Exercise 33.11, Exercise 33.12, and Proposition 33.13.

Theorem 34.12. *Let $f, g: [a, b] \rightarrow \mathbb{R}$ be integrable functions. Then for every $c \in (a, b)$ and every $C \in \mathbb{R}$ we have*

- (1) $\int_a^b C = C(b - a)$.
- (2) $\int_a^b f = \int_a^c f + \int_c^b f$.
- (3) If $f \leq g$ everywhere on $[a, b]$, then $\int_a^b f \leq \int_a^b g$.

Exercise 34.13. Prove that Theorem 34.12 continues to hold without any assumption on the ordering of a, b, c if we define $\int_b^a f = -\int_a^b f$ and $\int_a^a f = 0$.

Remark 34.14. The following two consequences of Theorem 34.12 come up so often that they are worth mentioning explicitly. Here $f: [a, b] \rightarrow \mathbb{R}$ is integrable and $m, M \in \mathbb{R}$.

$$(34.1) \quad f \geq 0 \quad \Rightarrow \quad \int_a^b f \geq 0,$$

$$(34.2) \quad m \leq f \leq M \quad \Rightarrow \quad m(b-a) \leq \int_a^b f \leq M(b-a).$$

Using Theorem 34.12, we can use the FTC to differentiate integrals where the lower limit of integration is variable: If f is integrable on $[a, b]$ then for $x \in (a, b)$ we have

$$(34.3) \quad \frac{d}{dx} \int_x^b f = \frac{d}{dx} \left(\int_a^b f - \int_a^x f \right) = -\frac{d}{dx} \int_a^x f = -f(x).$$

We can go one step further and differentiate integrals where both limits of integration are variable.

Example 34.15.

$$\frac{d}{dx} \int_{x^2}^{x^3} \sin t \, dt = \frac{d}{dx} \int_{x^2}^a \sin t \, dt + \frac{d}{dx} \int_a^{x^3} \sin t \, dt = (-2x) \sin(x^2) + (3x^2) \sin(x^3).$$

What happens to the integral $\int_a^b f$ if we multiply f by a scalar, or if we add two integrable functions together? First we consider this question in the case when f and g are continuous.

Theorem 34.16. *If $f, g: [a, b] \rightarrow \mathbb{R}$ are continuous and $c \in \mathbb{R}$, then*

$$\int_a^b (cf) = c \int_a^b f \quad \text{and} \quad \int_a^b (f+g) = \int_a^b f + \int_a^b g.$$

Proof. By FTC1, f and g have antiderivatives $F, G: [a, b] \rightarrow \mathbb{R}$. By linearity of differentiation, we have $(cF)'(x) = c(F'(x)) = cf(x)$ and $(F+G)'(x) = F'(x) + G'(x) = f(x) + g(x) = (f+g)(x)$ for all $x \in (a, b)$, so cF and $F+G$ are antiderivatives of cf and $f+g$, respectively. By FTC2, this implies that

$$\begin{aligned} \int_a^b (cf) &= (cF)(b) - (cF)(a) = c(F(b) - F(a)) = c \int_a^b f, \\ \int_a^b (f+g) &= (F+G)(b) - (F+G)(a) = F(b) - F(a) + G(b) - G(a) = \int_a^b f + \int_a^b g, \end{aligned}$$

which proves the theorem. \square

Remark 34.17. In fact, a similar result holds for all integrable functions, not just continuous ones, but the proof is harder; see Theorem 35.16 below.

Lecture 35

More about integration

DATE: FRIDAY, NOVEMBER 15

Spivak Chapter 13

35.1. Definitions via integrals

Using integrals, we can give alternate definitions of various numbers and functions that we have encountered so far. For example, a circle with radius 1 has area π , so half of the circle has area $\pi/2$. In particular, this is the area of the region between the curve $y = \sqrt{1 - x^2}$ and the x -axis over the interval $[-1, 1]$, and we conclude that

$$(35.1) \quad \pi = 2 \int_{-1}^1 \sqrt{1 - x^2} dx.$$

This gives one way of defining the number π .

Similarly, certain functions can be defined quite simply using integrals. Remember the amount of work we had to go through to define the exponential and logarithmic functions in earlier lectures. Since $\frac{d}{dx} \ln x = \frac{1}{x}$ and $\ln 1 = 0$, the FTC gives

$$\int_1^x \frac{1}{t} dt = \ln t \Big|_1^x = \ln x - \ln 1 = \ln x,$$

and thus we could just as well *define* the natural logarithm by $\ln x := \int_1^x \frac{1}{t} dt$. Indeed, this is exactly the approach taken in Spivak's book. Once \ln has been defined, the function $x \mapsto e^x$ is then defined to be its inverse.

A similar approach works for trigonometric functions. Earlier we demonstrated that $\frac{d}{dx} \sin^{-1} x = \frac{1}{\sqrt{1-x^2}}$, so instead of the geometric definition, we could give an analytic definition of arcsine as

$$\arcsin x := \int_0^x \frac{1}{\sqrt{1-t^2}} dt.$$

Then sine can be defined as the inverse function of this, and extended periodically, and we recover all the trigonometric functions from the usual identities.

35.2. Indefinite integrals

Thanks to the Fundamental Theorem of Calculus, we have now seen that an antiderivative F of a continuous function f can be produced by defining $F(x) = \int_a^x f$, and the previous section showed that some important functions can be defined this way. Since such functions are obtained by integration, it is common to use the following terminology.

Definition 35.1. An antiderivative F of a continuous function f is also called an *indefinite integral* of f , and denoted $F(x) = \int f(x) dx$.

Remark 35.2. The word “integral” is used to refer both to definite and indefinite integrals. Recall that given a function f ,

- the *definite integral* of f is a *number* associated to a specific interval $[a, b]$, so we must always specify the limits of integration and write $\int_a^b f(x) dx$; while
- an *indefinite integral* of f is a *function* that is not associated to any specific interval, and thus the notation $\int f(x) dx$ does not include any limits of integration.

We also stress that because a function has many antiderivatives on any given interval, indefinite integrals are only determined up to a constant, and indeed it is most appropriate to think of the indefinite integral of f as being a whole *family* of functions.

Example 35.3. The indefinite integral of x^2 is any of the family of functions $\frac{1}{3}x^3 + C$, where C is a constant, and we write

$$\int x^2 dx = \frac{1}{3}x^3 + C.$$

The constant C is often called a *constant of integration*. We can look up many indefinite integrals directly using our table of antiderivatives from an earlier lecture. We can also use linearity to compute many indefinite integrals:

$$\int (5x^3 + 2 \sec^2 x) dx = \frac{5}{4}x^4 + 2 \tan x + C.$$

Note that even though we combine two functions here, only a single constant of integration is needed.

Sometimes a little bit of manipulation is needed in order to put a function in a form from which we can easily evaluate the indefinite integral:

$$\int \frac{\sin x}{\cos^2 x} dx = \int \frac{1}{\cos x} \frac{\sin x}{\cos x} dx = \int \sec x \tan x dx = \sec x + C.$$

Using the Fundamental Theorem of Calculus, we can write definite integrals in terms of indefinite integrals:

$$\int_0^1 \frac{1}{x^2 + 1} dx = \left[\int \frac{1}{x^2 + 1} dx \right]_0^1 = \left[\arctan x \right]_0^1 = \arctan(1) - \arctan(0) = \frac{\pi}{4}.$$

Technically the indefinite integral above should be “ $\arctan x + C$ ” to include the constant of integration. However, because it does not matter *which* antiderivative we use in the FTC, the constant of integration can be omitted when computing a definite integral using indefinite integrals.

The FTC2 can be reformulated as follows.

Theorem 35.4 (Net Change Theorem). *If $F: [a, b] \rightarrow \mathbb{R}$ is differentiable on (a, b) , then $\int_a^b F'(x) dx = F(b) - F(a)$.*

The Net Change Theorem says that the net change in a quantity F is the integral of its rate of change. In applications of this theorem, we often (but not always) interpret F as a function of time t .

Example 35.5. If $V(t)$ represents the volume of water in a container at time t , then the Net Change Theorem says that $\int_{t_1}^{t_2} V'(t) dt = V(t_2) - V(t_1)$: the integral of the rate at which water enters or leaves the container is equal to the net change in the volume of water in the container.

Example 35.6. If $s(t)$ denotes the position of an object at time t , then $v(t) = s'(t)$ is its velocity, and the net displacement from time t_1 to time t_2 is $s(t_2) - s(t_1) = \int_{t_1}^{t_2} v(t) dt$. For example, if v is positive from t_1 to t' , negative from t' to t'' , and then positive from t'' to t_2 , and if A_1, A_2, A_3 denote the areas of the regions between the curve and the x -axis in these three intervals, then the net displacement is $s(t_2) - s(t_1) = A_1 - A_2 + A_3$. The negative sign on A_2 comes because between times t' and t'' the object is moving in the negative direction.

If we want to compute the total distance traveled, then we need to integrate not v , but $|v|$: the total distance traveled is $\int_{t_1}^{t_2} |v(t)| dt$, which in the situation above is equal to $A_1 + A_2 + A_3$.

Example 35.7. To make the previous example concrete, suppose that the velocity at time t is $v(t) = t^2 - t - 6$ and that t ranges from 1 to 4. Then the net displacement is

$$\begin{aligned} \int_1^4 v(t) dt &= \int_1^4 (t^2 - t - 6) dt = \left[\frac{1}{3}t^3 - \frac{1}{2}t^2 - 6t \right]_1^4 \\ &= \left(\frac{64}{3} - 8 - 24 \right) - \left(\frac{1}{3} - \frac{1}{2} - 6 \right) = -\frac{9}{2}. \end{aligned}$$

Note that $v(t) = (t - 3)(t + 2)$ is negative on $[1, 3)$ and positive on $(3, 4]$, so the total distance traveled is

$$\int_1^4 |v(t)| dt = \int_1^3 -v(t) dt + \int_3^4 v(t) dt = \left[\frac{1}{3}t^3 - \frac{1}{2}t^2 - 6t \right]_1^3 + \left[\frac{1}{3}t^3 - \frac{1}{2}t^2 - 6t \right]_3^4$$

which is equal to $\frac{61}{6}$ (after a little computation).

The following sections were discussed very briefly in the lecture, with all proofs skipped

35.3. Approximations by lower and upper sums

Lemma 33.14 said that if we refine a partition by adding another point (thus dividing one of the subintervals into two pieces), then the corresponding lower sum does not get smaller, and the corresponding upper sum does not get bigger. Iterating this procedure gives the following result.

Lemma 35.8. *If $f: [a, b] \rightarrow \mathbb{R}$ is bounded and P, Q are partitions of $[a, b]$ with $P \subset Q$, then $L(f, P) \leq L(f, Q)$ and $U(f, P) \geq U(f, Q)$.*

Proof. Choose partitions $P = P_0 \subset P_1 \subset P_2 \subset \cdots \subset P_m = Q$ such that each P_i is obtained from P_{i-1} by adding a single point. Then by Lemma 33.14,

$$L(f, P) = L(f, P_0) \leq L(f, P_1) \leq L(f, P_2) \leq \cdots \leq L(f, P_m) = L(f, Q).$$

The proof for U follows the same argument. □

Remark 35.9. The definition of lower and upper sums immediately gives $L(f, P) \leq U(f, P)$ for every partition P , but did not immediately tell us that $L(f, P) \leq U(f, Q)$ when P, Q are two different partitions. In particular, so far we did not even prove that $\int_a^b f \leq \overline{\int}_a^b f$ in general! Now we can do this, using Lemma 35.8. Indeed, given two partitions P, Q of $[a, b]$, Lemma 35.8 gives

$$L(f, P) \leq L(f, P \cup Q) \leq U(f, P \cup Q) \leq U(f, Q).$$

Fixing Q and taking a supremum over all P gives $\underline{\int}_a^b f \leq U(f, Q)$, and then taking an infimum over all Q gives $\underline{\int}_a^b f \leq \overline{\int}_a^b f$. Putting it all together, we have shown that for any partitions P, Q and any bounded f , we have

$$(35.2) \quad L(f, P) \leq L(f, P \cup Q) \leq \underline{\int}_a^b f \leq \overline{\int}_a^b f \leq U(f, P \cup Q) \leq U(f, Q).$$

Lemma 35.10. *Let $f: [a, b] \rightarrow \mathbb{R}$ be bounded. Then the following are equivalent:*

- (1) f is integrable;
- (2) for every $\varepsilon > 0$, there exists a partition P of $[a, b]$ such that $U(f, P) - L(f, P) < \varepsilon$.

Proof. If f is integrable, then by the definition of $\underline{\int}$ and $\overline{\int}$ there are partitions Q, R such that

$$\left(\underline{\int}_a^b f\right) - \frac{\varepsilon}{2} < L(f, Q) \leq \underline{\int}_a^b f \leq \overline{\int}_a^b f \leq U(f, R) < \left(\overline{\int}_a^b f\right) + \frac{\varepsilon}{2},$$

and by (35.2) the partition $P = Q \cup R$ satisfies

$$U(f, P) - L(f, P) \leq U(f, R) - L(f, Q) < \varepsilon.$$

Conversely, for every partition P we have

$$\overline{\int}_a^b f - \underline{\int}_a^b f \leq U(f, P) - L(f, P),$$

so if the RHS can be made arbitrarily small by choosing P appropriately, then the LHS, which does not depend on P , must be equal to 0, so f is integrable. \square

35.4. Riemann sums

In Examples 32.1 and 33.9, we saw that in order to compute the integrals $\int_0^b x \, dx$ and $\int_0^b x^2 \, dx$, we did not really need to consider all partitions, or consider the lower and upper sums. It was enough to consider a single sequence of partitions for which the lengths of the subintervals became arbitrarily small, and to compute the Riemann sums associated to some convenient choice of tags. In fact, this always works, provided $f: [a, b] \rightarrow \mathbb{R}$ is integrable.

First we observe that as shown in (33.1), for every partition P and every choice of tags τ , we have

$$L(f, P) \leq R(f, P, \tau) \leq U(f, P),$$

where L, U are the lower and upper sums used in the definition of the (Darboux) integral and R is the Riemann sum associated to the tagged partition (P, τ) . Moreover, by Lemma 35.10, if f is integrable then for every $\varepsilon > 0$ there is a partition P such that $U(f, P) - L(f, P) < \varepsilon$. In this case, every choice of tags τ will give a Riemann sum with the property that

$$\left| R(f, P, \tau) - \int_a^b f \right| < \varepsilon,$$

Thus the real question is to determine how we can find a partition with this ‘good approximation’ property. Intuitively we might expect that making the partition’s subintervals small enough guarantees that the approximation is good, and indeed this turns out to be the case. To give a precise statement, we make the following definition

Definition 35.11. The *mesh* of a partition $P\{a = x_0 < x_1 < \cdots < x_n = b\}$ is

$$\text{mesh}(P) := \max\{x_i - x_{i-1} : 1 \leq i \leq n\}.$$

Theorem 35.12. *Let $f: [a, b] \rightarrow \mathbb{R}$ be integrable. For every $\varepsilon > 0$, there exists $\delta > 0$ such that if P is any partition with $\text{mesh}(P) < \delta$, then $U(f, P) - L(f, P) < \varepsilon$. In particular, for every tagged partition (P, τ) satisfying $\text{mesh}(P) < \delta$, we have $|R(f, P, \tau) - \int_a^b f| < \varepsilon$.*

Proof. By Lemma 35.10, since f is integrable, there is a partition Q such that $U(f, Q) - L(f, Q) < \frac{\varepsilon}{2}$. Let N be the number of subintervals in Q , so $Q = \{a = y_0 < y_1 < \cdots < y_N = b\}$. Since f is bounded, there is $K > 0$ such that $|f(x)| \leq K$ for all $x \in [a, b]$. Choose $\delta > 0$ small enough that $2KN\delta < \frac{\varepsilon}{4}$.

Now given any partition P with $\text{mesh}(P) < \delta$, it follows from (35.2) that the partition $R = P \cup Q$ has $U(f, R) - L(f, R) < \frac{\varepsilon}{2}$. We claim that

$$(35.3) \quad U(f, P) < U(f, R) + \frac{\varepsilon}{4} \quad \text{and} \quad L(f, P) > L(f, R) - \frac{\varepsilon}{4},$$

which will imply that $U(f, P) - L(f, P) < U(f, R) - L(f, R) + \frac{\varepsilon}{2} < \varepsilon$, and complete the proof. So our goal is to prove (35.3).

We prove the second half of (35.3); the first half is similar. Writing $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$, we have

$$L(f, P) = \sum_{i=1}^n m_i(f, P)(x_i - x_{i-1}).$$

For each $i \in \{1, \dots, n\}$, there are two possibilities.

- Case 1: there are no elements of R between x_{i-1} and x_i . In this case $[x_{i-1}, x_i]$ is also a subinterval in the partition R , and makes exactly the same contribution to the sum that defines $L(f, R)$.
- Case 2: there is at least one element of R between x_{i-1} and x_i . In this case, $R \cap [x_{i-1}, x_i]$ gives a partition of $[x_{i-1}, x_i]$ that we can write as $\{x_{i-1} = z_0 < z_1 < \cdots < z_\ell = x_i\}$, and the corresponding lower sum is

$$\begin{aligned} \sum_{j=1}^{\ell} \inf\{f(x) : z_{j-1} \leq x \leq z_j\}(z_j - z_{j-1}) &\leq \sum_{j=1}^{\ell} (m_i(f, P) + 2K)(z_j - z_{j-1}) \\ &= (m_i(f, P) + 2K)(x_i - x_{i-1}) < m_i(f, P)(x_i - x_{i-1}) + 2K \text{mesh}(P). \end{aligned}$$

Now we sum over all n values of i to get an estimate on $L(f, R)$. Because R contains at most N elements that are not one of the x_i 's, Case 2 can happen at most N times, and we conclude that

$$L(f, R) \leq \left(\sum_{i=1}^n m_i(f, P)(x_i - x_{i-1}) \right) + 2KN \text{mesh}(P) = L(f, P) + 2KN \text{mesh}(P).$$

When $\text{mesh}(P) < \delta$, this gives

$$L(f, R) \leq L(f, P) + 2KN\delta < L(f, P) + \frac{\varepsilon}{4},$$

which proves the second half of (35.3). The first half is similar, and as explained above, this completes the proof of Theorem 35.12. \square

Remark 35.13. In particular, Theorem 35.12 shows that if (P_n, τ_n) is any sequence of tagged partitions of $[a, b]$ satisfying $\lim_{n \rightarrow \infty} \text{mesh}(P_n) = 0$, then $\lim_{n \rightarrow \infty} R(f, P_n, \tau_n) =$

$\int_a^b f$ for every integrable $f: [a, b] \rightarrow \mathbb{R}$. This includes the case when P_n is the partition into n subintervals of equal length, but is not restricted to that choice. Indeed, there are cases in which it is more convenient to use partitions into subintervals of unequal length, such as when f is known only approximately via data that has been collected for certain specific values of x that are not evenly spaced.

Remark 35.14. A function satisfying the conclusion of Theorem 35.12 is often called *Riemann integrable*, while a function satisfying $\int_a^b f = \overline{\int}_a^b f$ is called *Darboux integrable*. Theorem 35.12 says that Darboux integrable functions are Riemann integrable, and the converse direction is straightforward (we leave it as Exercise 35.15 below), so in fact Riemann and Darboux integrability are equivalent, which justifies our cavalier use of the word “integrable” without specifying which notion is meant.

Exercise 35.15. Suppose that $f: [a, b] \rightarrow \mathbb{R}$ is a bounded function and that there is $I \in \mathbb{R}$ such that for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every tagged partition (P, τ) satisfying $\text{mesh}(P) < \delta$, we have $|R(f, P, \tau) - I| < \varepsilon$. Prove that f is (Darboux) integrable and that $\int_a^b f = I$.

The following two sections were omitted from the in-class lecture

35.5. Linearity of integrals

As mentioned in Remark 34.17, the linearity properties in Theorem 34.16 continue to hold for all integrable functions.

Theorem 35.16. *If $f, g: [a, b] \rightarrow \mathbb{R}$ are integrable and $c \in \mathbb{R}$, then*

- (1) $cf: [a, b] \rightarrow \mathbb{R}$ is integrable, and $\int_a^b (cf) = c \int_a^b f$;
- (2) $f + g: [a, b] \rightarrow \mathbb{R}$ is integrable, and $\int_a^b (f + g) = \int_a^b f + \int_a^b g$.

Before proving the theorem we point out the following example demonstrating the need for integrability.

Exercise 35.17. Let $f: [0, 1] \rightarrow \mathbb{R}$ be the function from Exercise 33.10 and let $g = 1 - f$. Prove that $\int_0^1 (-f) \neq -\int_0^1 f$, and $\int_0^1 (f + g) \neq \int_0^1 f + \int_0^1 g$ even though all these lower integrals are defined.

Proof of Theorem 35.16. We give the proof for $f + g$ and leave cf as an exercise. Fix $\varepsilon > 0$. By Theorem 35.12, there is $\delta > 0$ such that for every tagged partition (P, τ) of $[a, b]$ satisfying $\text{mesh}(P) < \delta$, we have $|R(f, P, \tau) - \int_a^b f| < \frac{\varepsilon}{2}$ and $|R(g, P, \tau) - \int_a^b g| < \frac{\varepsilon}{2}$. Moreover, observe that for every tagged point t_i , we have $(f + g)(t_i) = f(t_i) + g(t_i)$, and thus $R(f + g, P, \tau) = R(f, P, \tau) + R(g, P, \tau)$. It follows that

$$\left| R(f + g, P, \tau) - \left(\int_a^b f + \int_a^b g \right) \right| \leq \left| R(f, P, \tau) - \int_a^b f \right| + \left| R(g, P, \tau) - \int_a^b g \right| < \varepsilon.$$

By Exercise 35.15, this implies that $f + g$ is integrable and that $\int_a^b (f + g) = \int_a^b f + \int_a^b g$. \square

Exercise 35.18. Adapt the above argument to prove the assertion in Theorem 35.16 regarding cf .

35.6. Alternate proof of linearity

We conclude by giving an alternate proof of Theorem 35.16 that works directly with the lower and upper sums and does not rely on Theorem 35.12. We start with the claim regarding cf , and consider the case $c \geq 0$. Given a partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$, for each $1 \leq i \leq n$ we have

$$m_i(cf, P) = \inf\{cf(x) : x \in [x_{i-1}, x_i]\} = c \inf\{f(x) : x \in [x_{i-1}, x_i]\} = cm_i(f, P),$$

and thus

$$L(cf, P) = \sum_{i=1}^n m_i(cf, P)(x_i - x_{i-1}) = \sum_{i=1}^n cm_i(f, P)(x_i - x_{i-1}) = cL(f, P).$$

Taking a supremum over all partitions P gives $\int_a^b cf = c \int_a^b f$, and since f is integrable this gives $\int_a^b cf = c \int_a^b f$. When $c < 0$, we have

$$m_i(cf, P) = \inf\{cf(x) : x \in [x_{i-1}, x_i]\} = c \sup\{f(x) : x \in [x_{i-1}, x_i]\} = cM_i(f, P),$$

and thus $L(cf, P) = cU(f, P)$. This time, taking a supremum over all partitions P gives

$$\int_a^b cf = \sup_P L(cf, P) = \sup_P cU(f, P) = c \inf_P U(f, P) = c \int_a^b f.$$

Since f is integrable this gives $\int_a^b cf = c \int_a^b f$.

The proof of the claim in Theorem 35.16 regarding $f + g$ requires a bit more work. Let $f, g: [a, b] \rightarrow \mathbb{R}$ be integrable, and consider $f + g$. We start by observing that for a given partition P , if we write $m_i^f = \inf\{f(x) : x \in [x_{i-1}, x_i]\}$ and define m_i^g, m_i^{f+g} similarly, then for every $x \in [x_{i-1}, x_i]$ we have $f(x) \geq m_i^f$ and $g(x) \geq m_i^g$, so $(f+g)(x) \geq m_i^f + m_i^g$, and consequently

$$m_i^{f+g} \geq m_i^f + m_i^g.$$

Recalling the definition of the lower sums, we get

$$L(f+g, P) = \sum_{i=1}^n m_i^{f+g}(x_i - x_{i-1}) \geq \sum_{i=1}^n (m_i^f + m_i^g)(x_i - x_{i-1}) = L(f, P) + L(g, P).$$

A similar argument gives $M_i^{f+g} \leq M_i^f + M_i^g$ and thus $U(f+g, P) \leq U(f, P) + U(g, P)$. Putting it all together, we get

$$(35.4) \quad L(f, P) + L(g, P) \leq L(f+g, P) \leq U(f+g, P) \leq U(f, P) + U(g, P)$$

for every partition P . Since f, g are integrable, for every $\varepsilon > 0$, Lemma 35.10 lets us choose partitions Q, R such that

$$(35.5) \quad \begin{aligned} \left(\int_a^b f\right) - \frac{\varepsilon}{4} &< L(f, Q) < \int_a^b f < U(f, Q) < \left(\int_a^b f\right) + \frac{\varepsilon}{4}, \\ \left(\int_a^b g\right) - \frac{\varepsilon}{4} &< L(g, R) < \int_a^b g < U(g, R) < \left(\int_a^b g\right) + \frac{\varepsilon}{4}. \end{aligned}$$

Now let $P = Q \cup R$ and use Lemma 33.14 together with (35.4) to get

$$\begin{aligned} \left(\int_a^b f + \int_a^b g \right) - \frac{\varepsilon}{2} &< L(f, Q) + L(g, R) \leq L(f, P) + L(g, P) \\ &\leq L(f + g, P) \leq \int_a^b (f + g) \leq \overline{\int}_a^b (f + g) \leq U(f + g, P) \\ &\leq U(f, P) + U(g, P) \leq U(f, Q) + U(g, R) < \left(\int_a^b f + \int_a^b g \right) + \frac{\varepsilon}{2}. \end{aligned}$$

This proves that $\overline{\int}_a^b (f + g) - \underline{\int}_a^b (f + g) < \varepsilon$, since both quantities are contained in the interval of length ε centered at $\int_a^b f + \int_a^b g$. Since this is true for any $\varepsilon > 0$, and the quantities $\overline{\int}_a^b f$, $\underline{\int}_a^b f$ do not depend on ε , it must be the case that $\overline{\int}_a^b (f + g) = \underline{\int}_a^b (f + g) = \int_a^b f + \int_a^b g$, which proves Theorem 35.16.

Lecture 36

Substitution rule

DATE: MONDAY, NOVEMBER 18

Stewart §5.5, Spivak Chapter 19

Now that we have a theory of integration, it is time to start developing the tools necessary to put it into practice. So far the only functions whose indefinite integrals we can find explicitly are the ones listed in the table in Lecture 31. The first step in extending our abilities is the *substitution rule*. Suppose we want to compute the indefinite integral

$$\int \frac{x}{\sqrt{1+x^2}} dx.$$

If we happen to guess that $\sqrt{1+x^2}$ might be an antiderivative, then it is easy to check that indeed it is, because we can write it as the composition of the functions $x \mapsto u = 1 + x^2$ and $u \mapsto \sqrt{u}$, and then use the chain rule to obtain

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

But how would we come up with this guess in the first place? Again, the chain rule provides the clue. Looking at the integrand $\frac{x}{\sqrt{1+x^2}}$, we might decide that we would get a simpler expression if we made a change of variables and wrote $u = 1 + x^2$. Then $\frac{du}{dx} = 2x$, and if we treat the notation as though it was genuinely a fraction (which it is not!) we might write $x dx = \frac{1}{2} du$ and obtain the formal, unjustified computation

$$\int \frac{x dx}{\sqrt{1+x^2}} = \int \frac{1}{2} u^{-1/2} du = u^{1/2} + C = \sqrt{1+x^2} + C.$$

The fact that this works provides some justification for why the notation is set up the way it is. This procedure is formalized by the following theorem.

Theorem 36.1 (Substitution rule). *Let f be a continuous function on an interval I and g be a differentiable function whose range is contained in I . Write $u = g(x)$; then*

$$\int f(g(x))g'(x) dx = \int f(u) du.$$

Equivalently, if $F: I \rightarrow \mathbb{R}$ is an antiderivative of f , then $F \circ g$ is an antiderivative of $(f \circ g) \cdot (g')$.

Proof. This is a direct consequence of the chain rule:

$$(F \circ g)'(x) = F'(g(x))g'(x) = f(g(x))g'(x),$$

where the first equality is the chain rule and the second uses the fact that F is an antiderivative of f . \square

Example 36.2. In the integral $\int x^2 \sin(x^3 + 1) dx$, we can write $u = x^3 + 1$ and obtain $du = 3x^2 dx$, so

$$\int x^2 \sin(x^3 + 1) dx = \int \frac{1}{3} \sin u du = -\frac{1}{3} \cos u + C = -\frac{1}{3} \cos(x^3 + 1) + C.$$

The last step in the example is important; when we are computing an indefinite integral, we always need to find a final expression that is given in terms of the original variable, and not any intermediate variables such as u that we introduced along the way.

The hardest part of the procedure is choosing which function u to use. This is often a trial and error procedure, but there are some guidelines that are helpful to keep in mind.

- If some part of the integrand represents a function whose derivative also appears in the integrand, it may be worth setting u to be this part and seeing what happens.
- If there is some complicated expression that appears inside a square root, trigonometric function, logarithm, exponential, etc., then we might make progress by setting u to be this expression.

Example 36.3. In $\int \sqrt{3x+2} dx$, we can use the expression under the square root: $u = 3x + 2$, so $du = 3 dx$, and we get

$$\int \sqrt{3x+2} dx = \int \frac{1}{3} \sqrt{u} du = \frac{1}{3} \left(\frac{1}{3/2} u^{3/2} \right) + C = \frac{2}{9} u^{3/2} + C = \frac{2}{9} (3x+2)^{3/2} + C.$$

An alternate approach would be to take $u = \sqrt{3x+2}$ so that $u^2 = 3x+2$ and $2u du = 3 dx$, giving

$$\int \sqrt{3x+2} dx = \int \frac{2}{3} u^2 du = \frac{2}{9} u^3 du + C = \frac{2}{9} (3x+2)^{3/2} + C.$$

Example 36.4. In $\int x^3 \sqrt{1+x^2} dx$, we again use the expression under the square root: $u = 1 + x^2$ gives $du = 2x dx$ and

$$\begin{aligned} \int x^3 \sqrt{1+x^2} dx &= \int (x^2 \sqrt{1+x^2}) x dx = \int (u-1) \sqrt{u} \frac{du}{2} = \frac{1}{2} \int (u^{3/2} - u^{1/2}) du \\ &= \frac{1}{2} \left(\frac{2}{5} u^{5/2} - \frac{2}{3} u^{3/2} \right) + C = \frac{1}{5} u^{5/2} - \frac{1}{3} u^{3/2} + C \end{aligned}$$

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

Example 36.5. To find the integral of $\tan x$, we can write it as $\frac{\sin x}{\cos x}$ and notice that the derivative of $\cos x$ appears in the numerator (up to a negative sign), so putting $u = \cos x$ gives $du = -\sin x dx$ and

$$\begin{aligned} \int \tan x dx &= \int \frac{\sin x}{\cos x} dx = \int \frac{-du}{u} = -\ln |u| + C = -\ln |\cos x| + C = \ln |1/\cos x| + C \\ &= \ln |\sec x| + C. \end{aligned}$$

This is an important enough example that it is worth remembering for future reference.

To compute *definite* integrals using the substitution rule, one option is to use the above procedure to evaluate the indefinite integral, and then apply the FTC. For example, we can use Example 36.3 to compute

$$(36.1) \quad \int_0^1 \sqrt{3x+2} dx = \frac{2}{9}(3x+2)^{3/2} \Big|_0^1 = \frac{2}{9}(5^{3/2} - 2^{3/2}).$$

An alternate approach is to apply the substitution rule direction to the definite integral via the following result, which uses the above procedure and the FTC in its proof.

Theorem 36.6 (Substitution rule for definite integrals). *Suppose $[a, b]$ and I are intervals in \mathbb{R} , and that we are given functions $g: [a, b] \rightarrow I$ and $f: I \rightarrow \mathbb{R}$ such that g' exists and is continuous on (a, b) , and f is continuous on I . Then*

$$\int_a^b f(g(x))g'(x) dx = \int_{g(a)}^{g(b)} f(u) du.$$

Proof. Let $F: I \rightarrow \mathbb{R}$ be an antiderivative of f ; then $\frac{d}{dx}F(g(x)) = F'(g(x))g'(x)$ so the FTC2 gives

$$\int_a^b f(g(x))g'(x) dx = F(g(x)) \Big|_a^b = F(g(b)) - F(g(a)) = F(u) \Big|_{g(a)}^{g(b)} = \int_{g(a)}^{g(b)} f(u) du.$$

□

With this approach, we can evaluate the integral in (36.1) by using $u = g(x) = 3x + 2$ to get $du = 3 dx$ and $g(0) = 2$, $g(1) = 5$, so

$$\int_0^1 \sqrt{3x+2} dx = \int_2^5 \frac{1}{3} \sqrt{u} du = \frac{1}{3} \cdot \frac{2}{3} u^{3/2} \Big|_2^5 = \frac{2}{9}(5^{3/2} - 2^{3/2}).$$

Observe that we need to change the limits of integration so that they are given in terms of the new variable. This example says that the area under the graph of $\sqrt{3x+2}$ between 0 and 1 is the same as the area under the graph of $\sqrt{u}/3$ between 2 and 5.

Example 36.7. To compute $\int_1^2 (1-2x)^{-2} dx$, we can write $u = 1-2x$ so that $du = -2 dx$ and the new integral goes from $u = -1$ to $u = -3$:

$$\int_1^2 \frac{dx}{(1-2x)^2} = -\frac{1}{2} \int_{-1}^{-3} u^{-2} du = \frac{1}{2u} \Big|_{-1}^{-3} = \frac{1}{2(-3)} - \frac{1}{2(-1)} = -\frac{1}{6} + \frac{1}{2} = \frac{1}{3}.$$

The substitution rule lets us deduce certain properties of integrals of symmetric functions.

Theorem 36.8. *Let $f: [-a, a] \rightarrow \mathbb{R}$ be continuous.*

- (1) *If f is even, then $\int_{-a}^a f = 2 \int_0^a f$.*
- (2) *If f is odd, then $\int_{-a}^a f = 0$.*

Proof. From properties of integrals, we have

$$(36.2) \quad \int_{-a}^a f = \int_{-a}^0 f + \int_0^a f = - \int_0^{-a} f + \int_0^a f.$$

The substitution $u = -x$ has $du = -dx$ and gives

$$- \int_0^{-a} f(x) dx = - \int_0^a f(-u)(-du) = \int_0^a f(-u) du.$$

When f is even, we have $f(-u) = f(u)$, so this is equal to $\int_0^a f$. When f is odd, we have $f(-u) = -f(u)$, so this is equal to $-\int_0^a f$. Using these in (36.2) proves the theorem. \square

The class on Wednesday, November 20 will be a review for Test 3

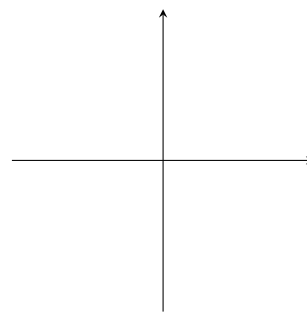
Lecture 37

Finding areas between curves

DATE: FRIDAY, NOVEMBER 22

Stewart §6.1

Suppose we want to find the area of a region such as the one shown in the picture at right, which is bounded on the left and right by the vertical lines $x = a$ and $x = b$, below by the graph of $y = g(x)$, and above by the graph of $y = f(x)$, where $f(x) \geq g(x)$ for all $x \in [a, b]$. If we partition the interval $[a, b]$ into n subintervals $[x_{i-1}, x_i]$ for $i = 1, \dots, n$, inside each of which we pick a ‘tag’ point t_i , then for each i we can consider the rectangle that ranges horizontally from x_{i-1} to x_i and vertically from $f(t_i)$ to $g(t_i)$, so its area is $(f(t_i) - g(t_i))(x_i - x_{i-1})$. The union of all of these rectangles has area $\sum_{i=1}^n (f(t_i) - g(t_i))(x_i - x_{i-1})$, which is a Riemann sum for the function $f - g$ on the interval $[a, b]$. Sending the mesh of the partition to 0, this sum converges to $\int_a^b (f - g)$, and we conclude that



$$(37.1) \quad (\text{area bounded by } f \text{ and } g \text{ between } x = a \text{ and } x = b) = \int_a^b (f(x) - g(x)) dx.$$

Lec 37

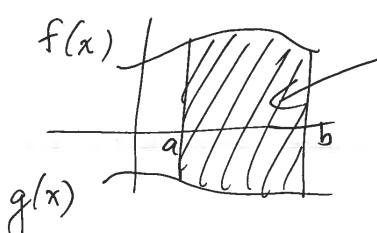
(11/21)

§6.1 | Area between curves

• Registered for 1432

• last week of lectures online

86



$$\text{area} = \lim_{n \rightarrow \infty} \sum_{i=1}^n [f(x_i^*) - g(x_i^*)] \Delta x$$

$$= \int_a^b [f(x) - g(x)] dx$$

eg Find area enclosed by $y = x^2$

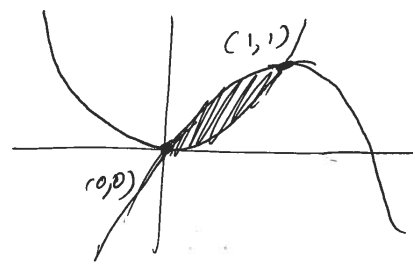
$$\& y = 2x - x^2 = 1 - (x-1)^2$$

Need to find intersection pts:

$$y = x^2 = 2x - x^2$$

$$\Rightarrow 2x^2 = 2x \Rightarrow x = 0 \text{ or } 1$$

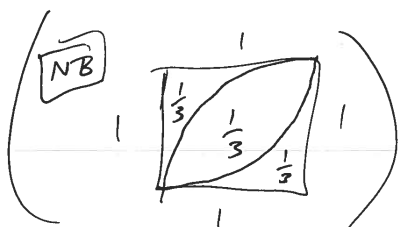
$$\Rightarrow (0, 0) \text{ \& } (1, 1)$$



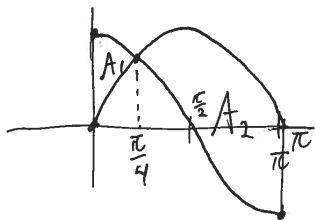
$$\text{Area} = \int_0^1 (2x - x^2) - x^2 dx$$

$$= \int_0^1 2x - 2x^2 dx$$

$$= \left[x^2 - \frac{2}{3} x^3 \right]_0^1 = 1 - \frac{2}{3} = \boxed{\frac{1}{3}}$$



If curves cross, use abs value:

eg What is area between $\sin x$ & $\cos x$ from 0 to $\frac{\pi}{2}$?

$$A_1 = \int_0^{\pi/4} \cos x - \sin x dx$$

$$= \left[\sin x + \cos x \right]_0^{\pi/4} = \sin \frac{\pi}{4} + \cos \frac{\pi}{4} - \cos 0 = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} - 1 = \sqrt{2} - 1$$

$$A_2 = \int_{\pi/4}^{\pi/2} \sin x - \cos x dx = \left[-\cos x - \sin x \right]_{\pi/4}^{\pi/2}$$

$$= -\cos \pi - \sin \pi + \cos \frac{\pi}{4} + \sin \frac{\pi}{4} = 1 + \sqrt{2}$$

(87)

So total area is

$$\int_0^{\pi} |\sin x - \cos x| dx = \int_0^{\frac{\pi}{4}} + \int_{\frac{\pi}{4}}^{\pi} = A_1 + A_2$$

$$= (\sqrt{2} - 1) + (\sqrt{2}) + 1 = 2\sqrt{2}$$

If v_1, v_2 are velocity fns for 2 objects,

then $v_2 - v_1 =$ speed at which v_2 is moving away from v_1 (assume they start together),

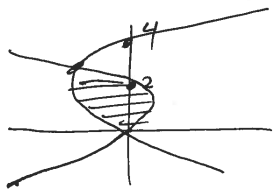
so $\int_0^T (v_2(t) - v_1(t)) dt =$ distance by which first is ahead.

Sometimes it is better to treat x as fn of y :

Q9 Area enclosed by $x = y^2 - 4y$ & $x = 2y - y^2$. Intersect at

$$2y - y^2 = y^2 - 4y \Leftrightarrow 6y = 2y^2 \Leftrightarrow y = 0 \text{ or } 3$$

$$\Leftrightarrow (0, 0) \text{ \& } (-3, 3)$$

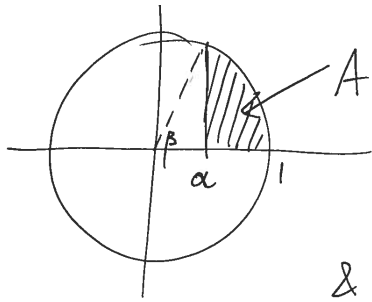


$$\text{Area} = \int_0^3 f(y) - g(y) dy$$

$$= \int_0^3 (2y - y^2) - (y^2 - 4y) dy$$

$$= \int_0^3 6y - 2y^2 dy = \left[3y^2 - \frac{2}{3}y^3 \right]_0^3$$

$$= 3 \cdot 9 - \frac{2}{3} \cdot 27 = 27 - 18 = \boxed{9}$$



$$A = ? \quad \int_a^1 \sqrt{1-x^2} dx$$

(88)

From geometry, area of sector is $\frac{\beta}{2}$
 where $\beta = \cos^{-1} a$.
 & triangle is $\frac{1}{2} a \sqrt{1-a^2}$

$$\therefore A = \frac{1}{2} (\cos^{-1} a - a \sqrt{1-a^2})$$

Can we compute integral directly?

$$\text{Try: } u = 1-x^2 \quad du = -2x dx \Rightarrow \int \sqrt{1-x^2} dx = \int \sqrt{u} \left(\frac{-du}{2x} \right)$$

$$x^2 = 1-u \quad x = \sqrt{1-u} \quad = \frac{-1}{2} \int \frac{\sqrt{u}}{\sqrt{1-u}} du = ???$$

$$\text{Try: } y = \sqrt{1-x^2} \quad y^2 = 1-x^2 \quad y dy = -x dx$$

$$\int \sqrt{1-x^2} dx = \int y \left(\frac{y dy}{-x} \right) = - \int \frac{y^2}{\sqrt{1-y^2}} dy = ???$$

$$\text{Try: } x = \cos \theta \quad (\theta = \cos^{-1} x)$$

$$\Rightarrow dx = -\sin \theta d\theta, \quad \sqrt{1-x^2} = \sqrt{1-\cos^2 \theta} = \sin \theta$$

$$\& \int_a^1 \sqrt{1-x^2} dx = \int_{\beta}^0 \sin \theta (-\sin \theta d\theta) = \int_0^{\beta} \sin^2 \theta d\theta$$

$$\text{Also, } \cos 2\theta = \cos^2 \theta - \sin^2 \theta = 1 - 2\sin^2 \theta$$

$$\Rightarrow \sin^2 \theta = \frac{1 - \cos 2\theta}{2}$$

$$\therefore \int_0^{\beta} \frac{1 - \cos 2\theta}{2} d\theta = \frac{1}{2} \left[\theta - \frac{1}{2} \sin 2\theta \right]_0^{\beta}$$

$$= \frac{1}{2} \beta - \frac{1}{4} \sin 2\beta = \frac{1}{2} \beta - \frac{1}{2} \sin \beta \cos \beta$$

$$= \frac{1}{2} (\cos^{-1} a - a \sqrt{1-a^2})$$

Lecture 38

Volumes

DATE: MONDAY, NOVEMBER 25

Stewart §§6.2-6.3

Lec 38

(1/28)

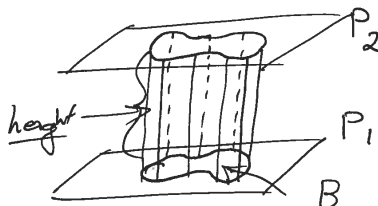
§6.2

Volumes

89

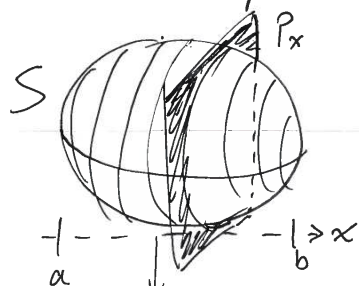
Def Given a region B in a plane P_1 & a parallel plane P_2 , the cylinder above B between P_1 & P_2 is the set of all pts in line segments starting in B , running \perp to P_1 , & ending in P_2 .

This solid region has
volume = $\text{area}(B) \cdot \text{height}$



What about volumes of other regions?

- approx by unions of cylinders
- same procedure as for rectangles in \mathbb{R}^2 .



close to being a cylinder with height Δx & base area $A(x_i^*)$

$$V = \lim_{n \rightarrow \infty} \sum_{i=1}^n A(x_i^*) \Delta x = \int_a^b A(x) dx$$

for a solid between $x=a$ & $x=b$
s.t. ~~the~~ cross-sectional area in P_x is $A(x)$.

(NB) if S is a cylinder then $A(x)$ is constant
 $\therefore V = \int_a^b A dx = A(b-a)$

(90)

Q) $S =$ sphere w/ radius r centered at O .

P_x intersects S in disc radius $\sqrt{r^2 - x^2}$

$$\therefore A(x) = \pi (r^2 - x^2)$$

$$\begin{aligned} \therefore V &= \int_{-r}^r \pi (r^2 - x^2) dx = 2\pi \int_0^r (r^2 - x^2) dx \\ &= 2\pi \left[r^2 x - \frac{1}{3} x^3 \right]_0^r = 2\pi \left(r^3 - \frac{1}{3} r^3 \right) = \frac{4\pi r^3}{3} \end{aligned}$$

Q) Cone with height h & radius r at base.

slice horizontally:

$$V = \int_0^h A(z) dz$$

$S \cap P_z$ is disc w/ radius $\frac{(h-z)r}{h}$

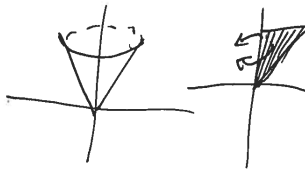
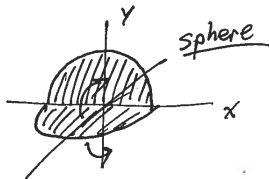
$$\therefore A(z) = \pi \left(\frac{h-z}{h} r \right)^2$$

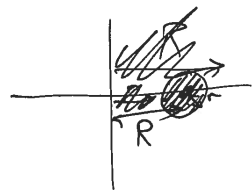
$$\begin{aligned} \therefore V &= \int_0^h \pi \left(\frac{h-z}{h} r \right)^2 dz = \frac{\pi r^2}{h^2} \int_0^h (h-z)^2 dz \\ &= \frac{\pi r^2}{h^2} \left[-\frac{1}{3} (h-z)^3 \right]_0^h = \frac{1}{3} \frac{\pi r^2}{h^2} h^3 = \frac{1}{3} \pi r^2 h \end{aligned}$$

\rightarrow could make integral easier by flipping over, so $A = \pi \left(\frac{zr}{h} \right)^2$

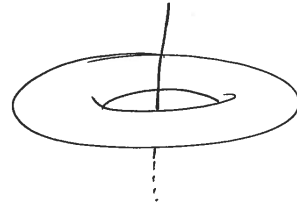


These were both solids of revolution:

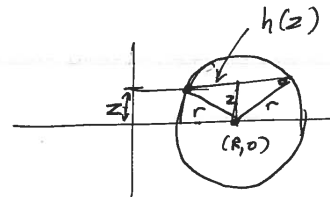




gives torus



w/ cross-section



$$A(z) = \pi (f(z)^2 - g(z)^2)$$

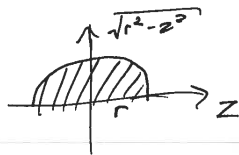
$$h(z)^2 = r^2 - z^2$$

$$f(z) = R + h(z) \quad g(z) = R - h(z)$$

$$A(z) = \pi (f+g)(f-g) = \pi \cdot 2R \cdot 2h(z) = 4\pi R h(z)$$

$$= 4\pi R \sqrt{r^2 - z^2}$$

$$\therefore V = \int_{-r}^r 4\pi R \sqrt{r^2 - z^2} dz = 4\pi R \int_{-r}^r \sqrt{r^2 - z^2} dz$$



$$\int_{-r}^r \sqrt{r^2 - z^2} dz = \frac{1}{2} \pi r^2$$

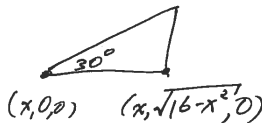
$$\therefore \text{volume of torus} = 4\pi R \left(\frac{1}{2} \pi r^2 \right) = 2\pi^2 R r^2$$

Lec 39

(11/30)

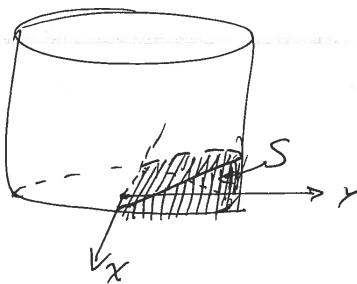
④ Circular cylinder radius 4 - cut out wedge by
2 planes: one \perp to axis of cylinder, the second
intersecting first at 30° along diam of cylinder.
Volume (wedge) = ?

SNP_x = triangle:



height = $y \tan 30^\circ$

$$= y \frac{1/\sqrt{3}}{1/2} = \frac{y}{\sqrt{3}} = \sqrt{\frac{16-x^2}{3}}$$



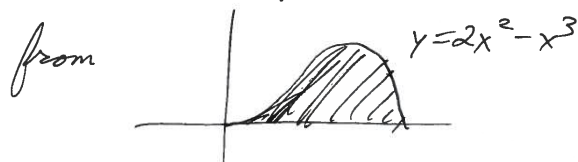
$$A(x) = \frac{1}{2} \sqrt{16-x^2} \cdot \frac{1}{\sqrt{3}} \sqrt{16-x^2} = \frac{16-x^2}{2\sqrt{3}}$$

$$V = \int_{-4}^4 A(x) dx = \frac{1}{\sqrt{3}} \int_0^4 16-x^2 dx = \frac{1}{\sqrt{3}} \left[16x - \frac{1}{3}x^3 \right]_0^4$$

$$= \frac{1}{\sqrt{3}} \left[64 - \frac{1}{3}64 \right] = \frac{128}{3\sqrt{3}}$$

§ 6.3 Volumes by cylindrical shells

Consider solid of revolution around y -axis



SNP_y annulus where inner & outer radii are
sols of $y = 2x^2 - x^3$... no nice formula

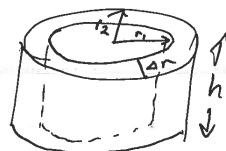
(93)

So instead, integrate out from y -axis:



each rectangle rotates into a
cylindrical shell =

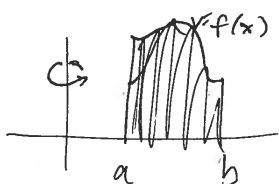
cylinder over an annulus



~~$$r = \frac{1}{2}(r_1 + r_2)$$~~

$$A(\text{annulus}) = \pi r_2^2 - \pi r_1^2 = \pi(r_2^2 - r_1^2) = \pi(r_2 - r_1)(r_2 + r_1) \\ = 2\pi r \Delta r$$

$$V(\text{cylindrical shell}) = \underbrace{2\pi r}_{\text{circumference}} \cdot \underbrace{h}_{\text{height}} \cdot \underbrace{\Delta r}_{\text{thickness}}$$



Divide $[a, b]$ into subintervals $\Delta x = \frac{b-a}{n}$ wide,
 \bar{x}_i = midpt of i^{th} , then rectangles rotate to shells,

$$\& \text{ volume } (i^{\text{th}} \text{ shell}) = 2\pi \bar{x}_i f(\bar{x}_i) \Delta x$$

$$\therefore \text{Volume}(S) = \lim_{n \rightarrow \infty} \sum_{i=1}^n 2\pi \bar{x}_i f(\bar{x}_i) \Delta x = \int_a^b 2\pi x f(x) dx$$

ex $y = 2x^2 - x^3$ on $[0, 2]$:

$$V = \int_0^2 (2\pi x)(2x^2 - x^3) dx = 2\pi \int_0^2 (2x^3 - x^4) dx \\ = 2\pi \left[\frac{1}{2}x^4 - \frac{1}{5}x^5 \right]_0^2 = 2\pi \left(8 - \frac{32}{5} \right) = \frac{16}{5} \pi$$