

Lectures on Advanced Calculus with Applications, I

audrey terras

Math. Dept., U.C.S.D., La Jolla, CA 92093-0112

November, 2010

Part I

Introduction, Motivation, Basics About Sets, Functions, Counting

1 Preface

These notes come from various courses that I have taught at U.C.S.D. using Serge Lang's *Undergraduate Analysis* as the basic text. My lectures are an attempt to make the subject more accessible. Recently Rami Shakarchi published *Problems and Solutions for Undergraduate Analysis*, which provides solutions to all the problems in Lang's book. This caused me to collect my own exercises which are included. Exams are also to be found. The main difference between the approach of Lang and that of other similar books is the treatment of the integral which emphasizes the properties of the integral as a linear function from the set of piecewise continuous real valued functions on an interval to the real numbers. Thus the approach can be viewed as intermediate between the Riemann integral and the Lebesgue integral. Since we are interested mainly in piecewise continuous functions, we are really getting the Riemann integral.

In these lectures we include more pictures and examples than the usual texts. Moreover, we include less definitions from point set topology. Our aim is to make sense to an audience of potential high school math teachers, or economists, or engineers. We did not write these lectures for potential math. grad students. We will always try to include examples, pictures and applications. Applications will include Fourier analysis, fractals,

Warning to the reader: This course is to calculus as fixing a car is to driving a car. Moreover, sometimes the car is invisible because it is an infinitesimal car or because it is placed on the road at infinity. It is thus important to ask questions and do the exercises.

A Suggestion: You should treat any mathematics course as a language course. This means that you must be sure to memorize the definitions and practice the new vocabulary every day. Form a study group to discuss the subject. It is always a good idea to look at other books too; in particular, your old calculus book.

Another Warning: Also, beware of typos. I am a terrible proof reader.

Your calculus class was probably one that would have made sense to Newton and Leibniz in the 1600s. However, that turned out not to be sufficient to figure out complicated problems. The basic idea of the real numbers was missing as well as a real understanding of the concept of limit. This course starts with the foundations that were missing in your calculus course. You may not see why you need them at first. Don't be discouraged by that. Persevere and you will get to derivatives and integrals. We will assume that you know the basics of proofs, sets.

Other References:

Hans Sagan, *Advanced Calculus*

Tom Apostol, *Mathematical Analysis*

Dym & McKean, *Fourier Series and Integrals*

1.1 Some History

Around the early 1800's Fourier was studying heat flow in wires or metal plates. He wanted to model this mathematically and came up with the heat equation. Suppose that we have a wire stretched out on the x-axis from $x = 0$ to $x = 1$. Let

$u(x, t)$ represent the temperature of the wire at position x and time t . The **heat equation is the PDE below**, for $t > 0$ and $0 < x < 1$:

$$\frac{\partial u}{\partial t} = c \frac{\partial^2 u}{\partial x^2}.$$

Here c is a positive constant depending on the metal. If you are given an initial heat distribution $f(x)$ on the wire at time 0, then we have the **initial condition**: $u(x, 0) = f(x)$ also.

Fourier plugged in the function $u(x, t) = X(x)T(t)$ and found that to for the solution to satisfy the initial condition he needed to express $f(x)$ as a **Fourier series**:

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}. \quad (1)$$

Note that $e^{ix} = \cos x + i \sin x$, where $i = (-1)^{1/2}$ (which is not a real number). This means you can rewrite the series of complex exponentials as 2 series - one involving cosines and the other involving sines. Fourier made the claim that any function $f(x)$ has such an expression as a sum of $c_n \sin(nx)$ and $d_n \cos(nx)$. People took issue with this although they did believe in power series expressions of functions (Taylor series and Laurent expansions). But the conditions under which such series converge to the function were really unclear when Fourier first worked on the subject.

Fourier tells us that the **Fourier coefficients** are

$$a_n = \int_0^1 f(y) e^{-2\pi i n y} dy. \quad (2)$$

If you believe that it is legal to interchange sum and integral, then a bit of work will make you believe this, but unfortunately, that isn't always legal when f is a bad guy. This left mathematicians in an uproar in the early 1800's. And it took at least 50 years to bring some order to the subject.

Part of the problem was that in the early 1800's people viewed integrals as antiderivatives. And they had no precise meaning for the convergence of a series of functions of x such as the Fourier series above. They argued a lot. They would not let Fourier publish his work until many years had passed. False formulas abounded. Confusion reigned supreme. So this course was invented. We won't have time to go into the history much, but it is fascinating. Bressoud, *A Radical Approach to Real Analysis*, says a little about the history. Another reference is Grattan-Guinness and Ravetz, *Joseph Fourier*. Still another is Lakatos, *Proofs and Refutations*.

We will end up with a precise formulation of Fourier's theorems. And we will be able to do many more things of interest in applied mathematics. In order to do all this we need to understand what the real numbers are, what we mean by the limit of a sequence of numbers or of a sequence of functions, what we mean by derivatives and integrals. You may think that you learned this in calculus, but unless you had an unusual calculus class, you just learned to compute derivatives and integrals not so much how to prove things about them.

Fourier series (and integrals) are important for all sorts of things such as analysis of time series, looking for periodicities. The finite version leads to a computer algorithm called the fast Fourier transform, which has made it possible to do things such as weather prediction in a reasonable amount of time. Matlab has a nice demo of the search for periodicities. We modified it in our book *Fourier Analysis on Finite Groups and Applications* to look for periodicities in LA yearly rainfall. The first answer I found was 12.67 years. See p. 159 of my book. Another version leads to the number 28.75 years.

2 Why Analysis? Some Motivation and a Look Forward

Almost any applied math. problem leads to an analysis question. Look at any book on mathematical methods of physics and engineering. There are also many theoretical problems in computer science that lead to analysis questions. The same can be said of economics, chemistry and biology. Here we list a few examples. We do not give all the details. The idea is to get a taste of such problems.

Example 1. Population Growth Model - The Logistic Equation.

References.

I. Stewart, *Does God Play Dice? The Mathematics of Chaos*, p. 155.

J. T. Sandefur, *Discrete Dynamical Systems*.

Define the **logistics function** $L_k(x) = kx(1 - x)$, for $x \in [0, 1]$. Here k is a fixed real number with $0 < k < 4$. Let $x_0 \in [0, 1]$ be fixed. Form a sequence

$$x_0, x_1 = L_k(x_0), x_2 = L_k(x_1), \dots, x_n = L_k(x_{n-1}), \dots$$

Question: What happens to x_n as $n \rightarrow \infty$?

The answer depends on k . For k near 0 there is a limit. For k near 4 the behavior is chaotic. Our course should give us the tools to solve this sort of problem. Similar problems come from weather forecasting, orbits of asteroids. You can put these problems on a computer to get some intuition. But you need analysis to prove that your intuition is correct (or not).

Example 2. Central Limit Theorem in Probability and Statistics.

References.

Feller, *Probability Theory*

Dym and McKean, *Fourier Series and Integrals*, p. 114

Terras, *Harmonic Analysis on Symmetric Spaces and Applications, Vol. I*

Where does the bell shaped curve originate?

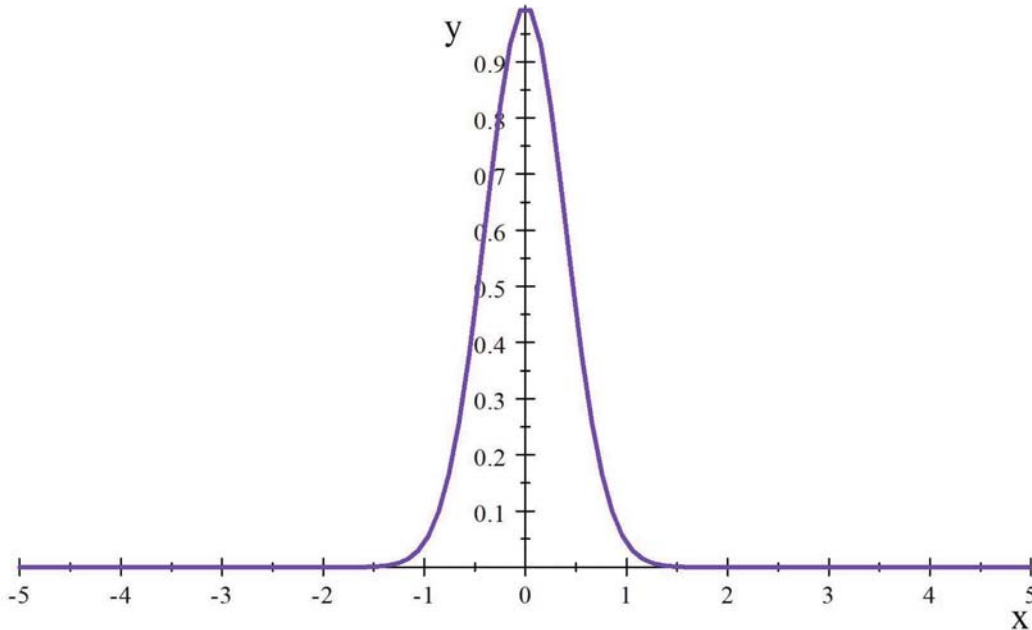


Figure 1: normal curve $e^{-\pi x^2}$, $-\infty < x < \infty$

The central limit theorem is the main theorem in probability and statistics. It is the foundation for the chi-squared test. In the language of probability, it says the following.

Central Limit Theorem I. Let X_n be a sequence of independent identically distributed random variables with density $f(x)$ normalized to have mean 0 and standard deviation 1. Then, as $n \rightarrow \infty$, the normalized sum of these variables

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} \text{the normal distribution with density } G(x), \text{ where } G(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Here " \rightarrow " means approaches. To translate this into analysis, we need a definition.

Definition 1 For integrable functions f and $g: \mathbb{R} \rightarrow \mathbb{R}$, define the convolution $f * g$ (f "splat" g) to be

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x - y)dy.$$

Then we have the analysis version of the central limit theorem.

Central Limit Theorem II. Suppose that $f : \mathbb{R} \rightarrow [0, \infty)$ is a probability density normalized to have mean 0 and standard deviation 1. This means that

$$\int_{-\infty}^{\infty} f(x)dx = 1, \int_{-\infty}^{\infty} xf(x)dx = 0, \int_{-\infty}^{\infty} x^2f(x)dx = 1.$$

Then we have the following limit as $n \rightarrow \infty$

$$\int_{a\sqrt{n}}^{b\sqrt{n}} (f \underbrace{* \dots *}_n f)(x)dx \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Example 3. Surprising Formulas.

a) **Riemann zeta function.**

References.

Lang, *Undergraduate Analysis*

Edwards, *Riemann's Zeta Function*

Terras, *Harmonic Analysis on Symmetric Spaces and Applications, Vol. I*

Definition 2 The *Riemann zeta function* $\zeta(s)$ is defined for $s > 1$ by

$$\zeta(s) = \sum_{n \geq 1} n^{-s}.$$

Euler proved $\zeta(2) = \pi^2/6$ and similar formulas for $\zeta(2n)$, $n = 1, 2, 3, \dots$.

b) **Gamma Function.**

References.

Lang, *Undergraduate Analysis*

Edwards, *Riemann's Zeta Function*

Terras, *Harmonic Analysis on Symmetric Spaces and Applications, Vol. I*

Definition 3 For $s > 0$, define the **gamma function** by $\Gamma(s) = \int_0^{\infty} e^{-y}y^{s-1}dy$.

Then we can get n factorial from gamma:

$$\Gamma(n + 1) = n! = n(n - 1)(n - 2) \dots 1.$$

Another result says

$$\Gamma(1/2) = \sqrt{\pi}.$$

c) **Theta Function.**

References.

Lang, *Undergraduate Analysis*

Terras, *Harmonic Analysis on Symmetric Spaces and Applications, Vol. I*

One of the Jacobi identities says that for $t > 0$, we have

$$\theta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi t^2} = \sqrt{\frac{\pi}{t}} \theta\left(\frac{1}{t}\right).$$

This is a rather unexpected formula - a hidden symmetry of the theta function. It implies (as Riemann showed in a paper published in 1859) that the Riemann zeta function also has a symmetry, relating $\zeta(s)$ with $\zeta(1-s)$.

d) **Famous Inequalities.**

i) **Cauchy-Schwarz Inequality**

Reference.

Lang, *Undergraduate Analysis*

Suppose that V is a vector space such as \mathbb{R}^n with a scalar product $\langle v, w \rangle = \sum_{i=1}^n v_i w_i$, if v_i denotes the i th coordinate of v in \mathbb{R}^n . Then the length of v is $\|v\| = \sqrt{\langle v, v \rangle}$. The **Cauchy-Schwarz inequality** says

$$|\langle v, w \rangle| \leq \|v\| \|w\|. \tag{3}$$

This inequality implies the **triangle inequality**

$$\|v + w\| \leq \|v\| + \|w\|,$$

which says that the sum of the lengths of 2 sides of a triangle is greater than or equal to the length of the third side.

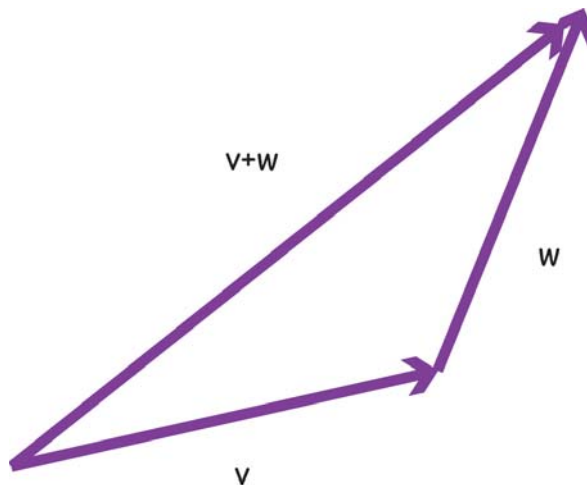


Figure 2: sum of vectors in the plane

The inequality of Cauchy-Schwarz is very general. It works for any inner product space V - even one that is infinite dimensional such as $V = C[0, 1]$, the space of continuous real-valued functions on the interval $[0, 1]$. Here the inner product for $f, g \in V$ is

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx.$$

In this case, Cauchy-Schwarz says

$$\left\{ \int_0^1 f(x)g(x)dx \right\}^2 \leq \int_0^1 f(x)^2 dx \int_0^1 g(x)^2 dx. \tag{4}$$

Amazingly the same proof works for inequality (3) as for inequality (4).

ii) **The Isoperimetric Inequality.**

Reference.

Dym and McKean, *Fourier Series and Integrals*

This inequality is related to Queen Dido's problem which is to maximize the area enclosed by a curve of fixed length. In 800 B.C., as recorded in Virgil's Aeneid, Queen Dido wanted to buy land to found the ancient city of Carthage. The locals

would only sell her the amount of land that could be enclosed with a bull's hide. She cut the hide into narrow strips and then made a long strip and used it to enclose a circle (actually a semicircle with one boundary being the Mediterranean Sea).

The isoperimetric inequality says that if A is the area enclosed by a plane curve and L is the length of the curve enclosing this area,

$$4\pi A \leq L^2.$$

Moreover, equality only holds for the circle which maximizes A for fixed L .

A =area enclosed by curve
 L =length of curve

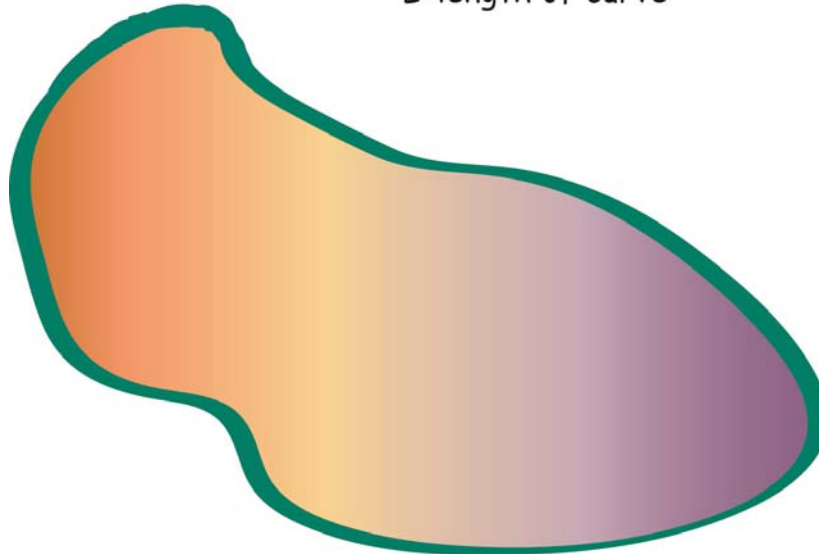


Figure 3: curve in plane of length L enclosing area A

iii) Heisenberg Inequality and the Uncertainty Principle.

References.

Dym and McKean, *Fourier Series and Integrals*

Terras, *Harmonic Analysis on Symmetric Spaces and Applications, Vol. I, p. 20*

Quantum mechanics says that you cannot measure position and momentum to arbitrary precision at the same time. The analyst's interpretation goes as follows. Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ and define the Fourier transform of f to be

$$\widehat{f}(w) = \int_{-\infty}^{\infty} f(t)e^{-2\pi itw} dt.$$

Here $i = \sqrt{-1}$ and $e^{i\theta} = \cos \theta + i \sin \theta$.

Now, suppose that we have the following facts

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = 1, \int_{-\infty}^{\infty} t |f(t)|^2 dt = 0, \int_{-\infty}^{\infty} w |\widehat{f}(w)|^2 dw = 0.$$

Then we have **the uncertainty inequality**

$$\int_{-\infty}^{\infty} t^2 |f(t)|^2 dt \int_{-\infty}^{\infty} w^2 |\widehat{f}(w)|^2 dw \geq \left(\frac{1}{4\pi}\right)^2.$$

The integral over t measures the square of the time duration of the signal $f(t)$ and the integral over w measures the square of the frequency spread of the signal. The uncertainty inequality can be shown to be equivalent to the following inequality involving the derivative of f rather than the Fourier transform of f :

$$\int_{-\infty}^{\infty} t^2 |f(t)|^2 dt \int_{-\infty}^{\infty} |f'(u)|^2 du \geq \frac{1}{4}.$$

This completes our quick introduction to some famous problems of analysis. Hopefully we will manage to investigate most of them in more detail. But next to set theory.

3 Set Theory and Functions

G. Cantor (1845-1918) developed the theory of infinite sets. It was controversial. There are paradoxes for those who throw caution to the winds and consider sets whose elements are sets. For example, consider **Russell's paradox**. It was stated by B. Russell (1872-1970). We use the notation: $x \in S$ to mean that x is an element of the set S ; $x \notin S$ meaning x is not an element of the set S . The notation $\{x|x \text{ has property } P\}$ is read as the set of x such that x has property P . Consider the set X defined by

$$X = \{\text{sets } S | S \notin S\}.$$

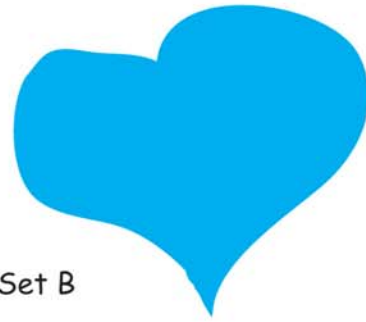
Then $X \in X$ implies $X \notin X$ and $X \notin X$ implies $X \in X$. This is a paradox. The set X can neither be a member of itself nor not a member of itself. There are similar paradoxes that sound less abstract. Consider the barber who must shave every man in town who does not shave himself. Does the barber shave himself? A mystery was written inspired by the paradox: *The Library Paradox* by Catharine Shaw. There is also a comic book about Russell, *Logicomix* by A. Doxiadis and C. Papadimitriou.

We will hopefully avoid paradoxes by restricting consideration to sets of numbers, vectors, functions. This would not be enough for "constructionists" such as E. Bishop, once at U.C.S.D. Anyway for applied math., one can hope that paradoxical sets and barbers do not appear.

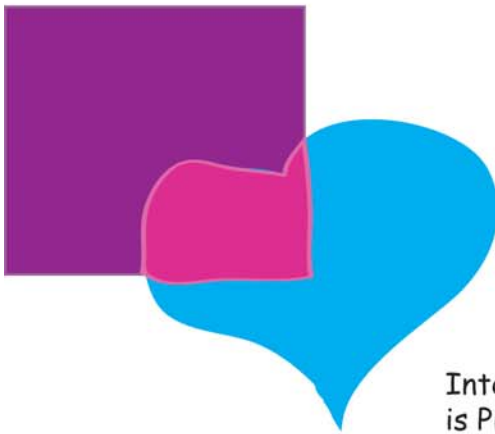
Most books on calculus do a little set theory. We assume you are familiar with the notation. Let's do pictures in the plane. We write $A \subset B$ if A is a subset of B ; i.e., $x \in A$ implies $x \in B$. If $A \subset B$, the **complement** of A in B is $B - A = \{x \in B | x \notin A\}$. The empty set is denoted \emptyset . It has no elements. The **intersection** of sets A and B is $A \cap B = \{x | x \in A \text{ and } x \in B\}$. The **union** of sets A and B is $A \cup B = \{x | x \in A \text{ or } x \in B\}$. Here or means either or both. See Figure 4.



Set A



Set B



Intersection of A and B
is Pink



Union of A and B is
Purple

Figure 4: intersection and union

Definition 4 If A and B are sets, the **Cartesian product** of A and B is the set of ordered pairs (a, b) with $a \in A$ and $b \in B$; i.e.,

$$A \times B = \{(a, b) | a \in A, b \in B\}.$$

Example 1. Suppose A and B are both equal to the set of all real numbers; $A = B = \mathbb{R}$. Then $A \times B = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. That is the Cartesian product of the real line with itself is the set of points in the plane.

Example 2. Suppose C is the interval $[0, 1]$ and D is the set consisting of the point $\{2\}$. Then $C \times D$ is the line segment of length 1 at height 2 in the plane. See Figure 5 below.

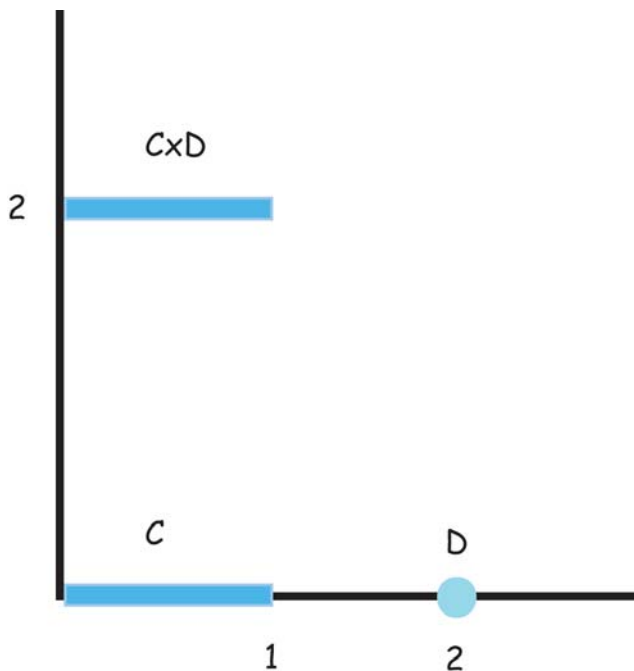


Figure 5: The Cartesian product $[0, 1] \times \{2\}$.

Example 3. $[0, 1] \times [0, 1] \times [0, 1] = [0, 1]^3$ is the unit cube in 3-space. See Figure 6.

Example 4. $[0, 1] \times [0, 1] \times [0, 1] \times [0, 1] = [0, 1]^4$ is the 4-dimensional cube or tesseract. Draw it by "pulling out" the 3-dimensional cube. See T. Banchoff, *Beyond the Third Dimension*. Figure 7 below shows the edges and vertices of the 4-cube (actually more of a 4-rectangular solid) as drawn by Mathematica.

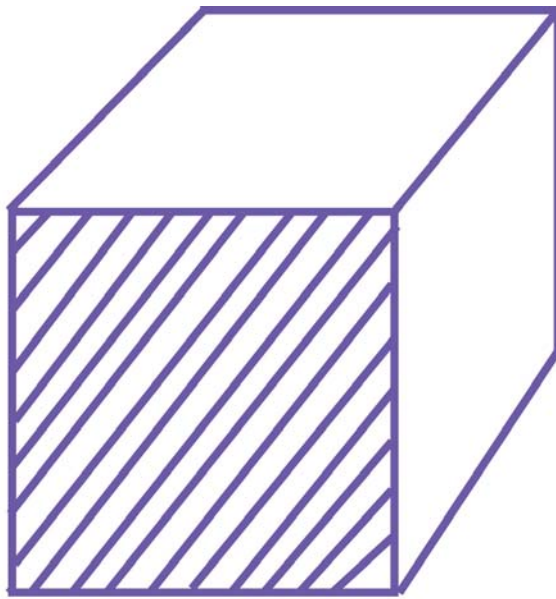


Figure 6: $[0, 1]^3$

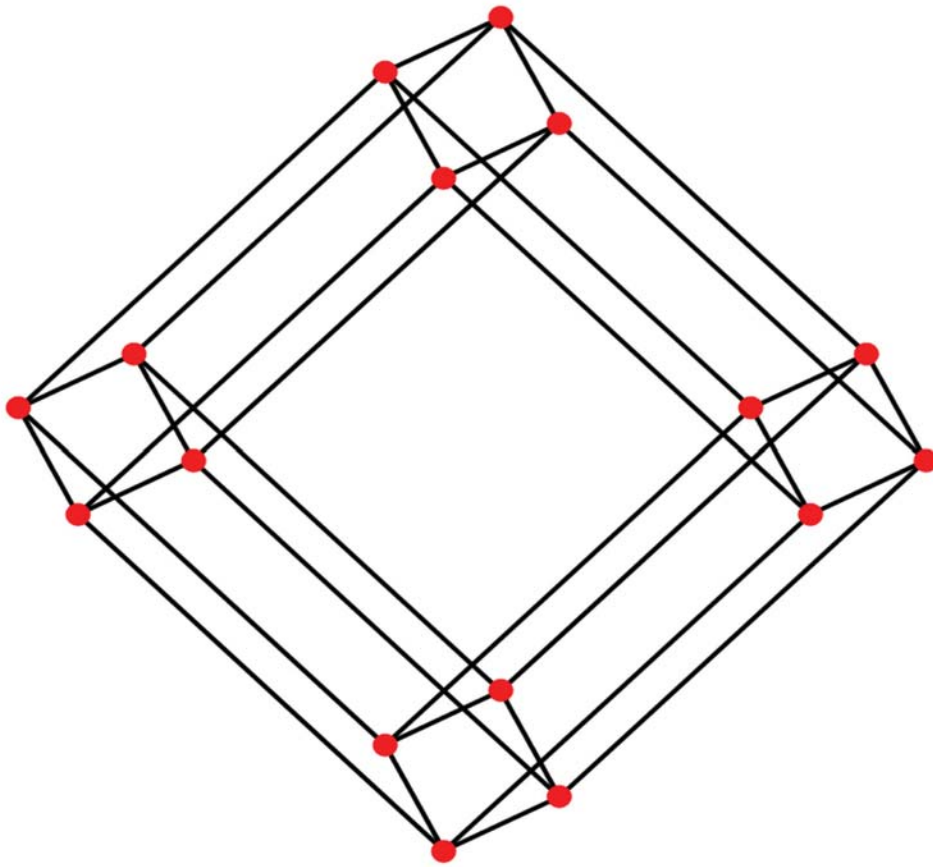


Figure 7: $[0, 1]^4$

Next we recall the definitions of functions which will be extremely important for the rest of these notes.

Definition 5 A **function (or mapping or map)** f maps set A into set B means that for every $a \in A$ there is a unique element $f(a) \in B$. The notation is $f : A \rightarrow B$.

Definition 6 The function $f : A \rightarrow B$ is **one-to-one (1-1 or injective)** if and only if $f(a) = f(a')$ implies $a = a'$.

Definition 7 The function $f : A \rightarrow B$ is **onto (or surjective)** if and only if for every $b \in B$, there exists $a \in A$ such that $f(a) = b$.

A function that is 1-1 and onto is also called **bijective** or a **bijection**. Given a function $f : A \rightarrow B$, we can draw a **graph** consisting of points $(x, f(x))$, for all $x \in A$. It is a subset of the Cartesian product $A \times B$. Equivalently a function $f : A \rightarrow B$ can be viewed as a subset F of $A \times B$ such that (a, b) and (a, c) in F implies $b = c$.

Notation. From now on I will use the abbreviations:

| | |
|---------------|---|
| iff | if and only if |
| \forall | for every |
| \exists | there exists |
| s.t. | such that |
| \rightarrow | approaches (in the limit to be defined later in gory detail) |
| \doteq | the thing on the left of \doteq is defined to be the thing on the right of \doteq |

Example 1. Suppose $A = B = [0, 1]$. A subset which is not a function is the square wave pictured below. This is a bad function for calculus since there are infinitely many values of the function at 2 points in the interval. We can make the square wave into a function by collapsing the 2 vertical lines to points.

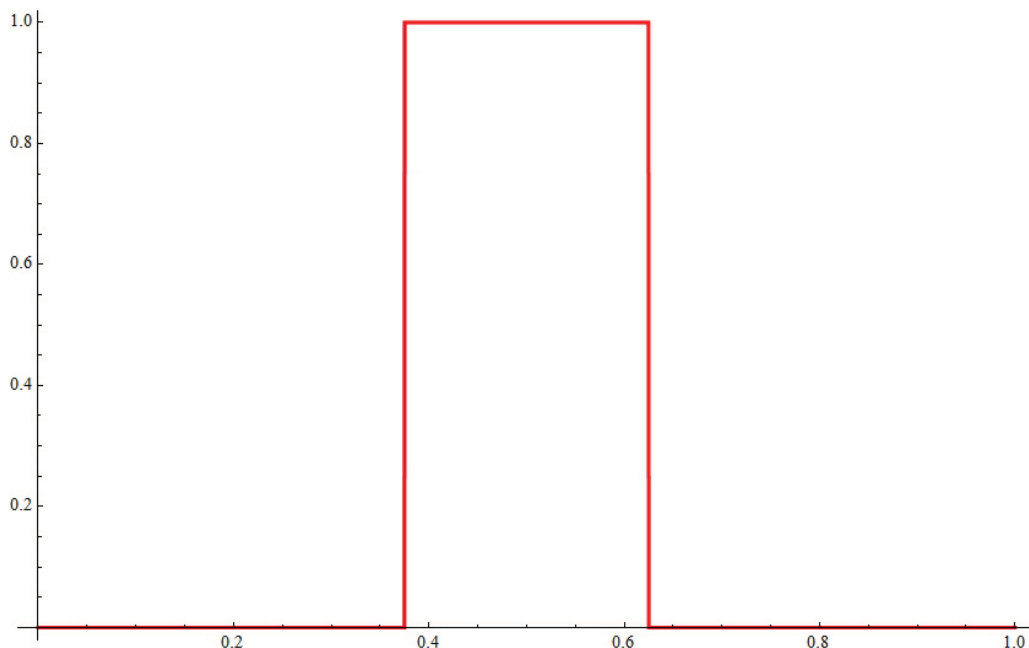


Figure 8: not a function

Example 2. Consider the logistic map $L(x) = 3x(1-x)$ for $x \in [0, 1]$. If we consider L as a map from $[0, 1]$ into $[0, 1]$, we see from the figure below that L is neither 1-1 nor onto. It is not 1-1 since there are 2 points $a, a' \in [0, 1]$ such that $f(a) = f(a') = 1/2$. It is not onto since there is no point $a \in [0, 1]$ such that $f(a) = 0.9$.

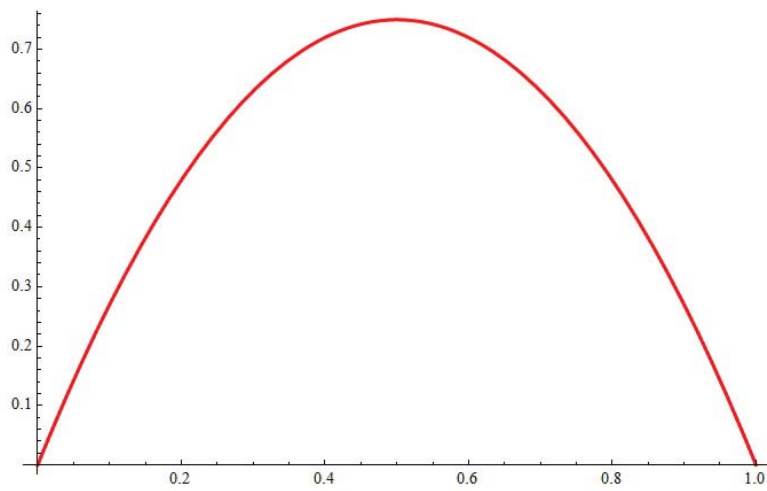


Figure 9: The logistic map $f(x) = 3x(x - 1)$.

Definition 8 Suppose $f : A \rightarrow B$ and $g : B \rightarrow C$. Then the **composition** $g \circ f : A \rightarrow C$ is defined by $(g \circ f)(x) = g(f(x))$, for all $x \in A$.

It is easily seen (and the reader should check) that this operation is **associative**; i.e., $f \circ (g \circ h) = (f \circ g) \circ h$. However this operation is not **commutative** in general; i.e., $f \circ g \neq g \circ f$ usually. For example, consider $f(x) = x^2$ and $g(x) = x + 1$, for $x \in \mathbb{R}$.

There is a **right identity** for the operation of composition of functions. If $f : A \rightarrow B$ and $I_A(x) = x, \forall x \in A$, then $f \circ I_A = f$. Similarly I_B is a **left identity** for f ; i.e., $I_B \circ f = f$.

Definition 9 If $f : A \rightarrow B$ is 1-1 and onto, it has an **inverse function** $f^{-1} : B \rightarrow A$ defined by requiring $f \circ f^{-1} = I_B$ and $f^{-1} \circ f = I_A$. If $f(a) = b$, then $f^{-1}(b) = a$.

Exercise 10 a) Prove that if $f : A \rightarrow B$ is 1-1 and onto, it has an **inverse function** f^{-1} .
 b) Conversely show that if f has an inverse function, then f must be 1-1 and onto.

Example 1. Let $f(x) = x^2$ map $[0, \infty)$ onto $[0, \infty)$. Then f is 1-1 and onto with the inverse function $f^{-1}(x) = \sqrt{x} = x^{1/2}$. Here of course we take the non-negative square root of $x \geq 0$. It is necessary to restrict f to non-negative real number in order for f to be 1-1.

Example 2. Define $f(x) = e^x$. Then f maps $(-\infty, \infty)$ 1-1 onto $(0, \infty)$. The inverse function is $f^{-1}(x) = \log x = \log_e x$. It is only defined for positive x . We discuss these functions in more detail later.

When you draw the graph of f^{-1} , you just need to reflect the graph of f across the line $y = x$.

4 Mathematical Induction

Notation:

| | | |
|----------------|-------------------------------------|-----------------------|
| \mathbb{Z}^+ | $\{1, 2, 3, 4, \dots\}$ | the positive integers |
| \mathbb{Z} | $\{0, \pm 1, \pm 2, \pm 3, \dots\}$ | the integers |
| \mathbb{R} | $(-\infty, +\infty)$ | real numbers |

We will assume that you are familiar with the integers as far as arithmetic goes. They satisfy most of the axioms that we will list later for the real numbers. In particular, \mathbb{Z} is closed under addition and multiplication (also subtraction but not division). This means $n, m \in \mathbb{Z}$ implies $n + m, n - m$ and $n * m$ are all unique elements of \mathbb{Z} . Moreover, one has an identity for $+$, namely 0, an identity for $*$, namely 1. Addition and multiplication are associative and commutative. There is an additive inverse in \mathbb{Z} for every $n \in \mathbb{Z}$ namely $-n$. But unless $n = \pm 1$, there is no multiplicative inverse for n in \mathbb{Z} .

One thing that differentiates \mathbb{Z} from the real numbers \mathbb{R} is the following axiom. Moreover there is an ordering of \mathbb{Z} which behaves well with respect to addition and multiplication. We will list the order axioms later, with one exception.

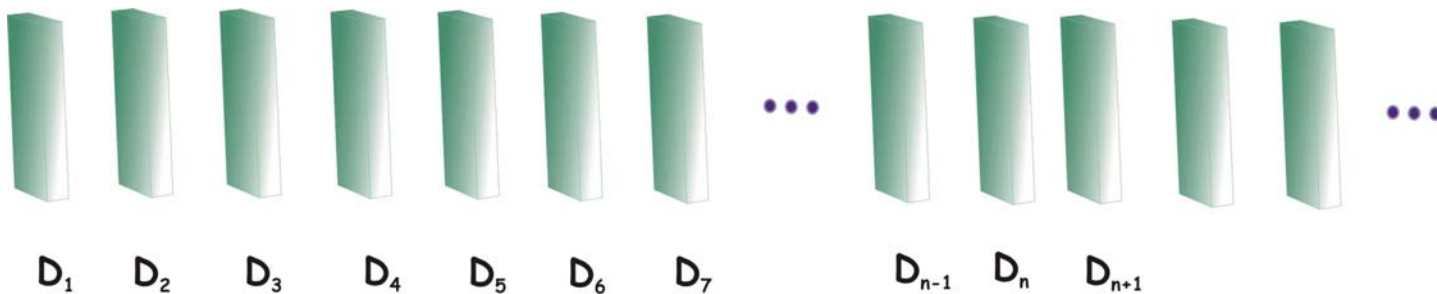
Axiom 11 The Well Ordering Axiom. If $S \subset \mathbb{Z}^+$, and $S \neq \emptyset$, then S has a least element $a \in S$ such that $a \leq x, \forall x \in S$.

This axiom says that any non-empty set of positive integers has a least element. We usually call such a least element a minimum. By an axiom, we mean that it is a basic unproved assumption.

G. Peano (1858-1932) wrote down the 5 Peano Postulates (or axioms) for the natural numbers $\mathbb{Z}^+ \cup \{0\}$. We won't list them here. See, for example, Birkhoff and MacLane, *Survey of Modern Algebra*. Once one has these axioms it would be nice to show that something exists satisfying the axioms. We will not do that here, feeling pretty confident that you believe \mathbb{Z} exists.

The most important fact about the well ordering axiom is that it is equivalent to mathematical induction.

Domino Version of Mathematical Induction. Given an infinite line of equally spaced dominos of equal dimensions and weight, in order to knock over all the dominos by just knocking over the first one in line, we should make sure that the n th domino is so close to the $(n+1)$ st domino that when the n th domino falls over, it knocks over the $(n+1)$ st domino. See Figure 10.



Make sure D_n
knocks over D_{n+1}
for all $n=1,2,3,\dots$

Figure 10: An infinite line of equally spaced dominos. If the n th domino is close enough to knock over the $n+1$ st domino, then once you knock over the 1st domino, they should all fall over.

Translating this to theorems, we get the following

Principle of Mathematical Induction I.

Suppose you want to prove an infinite list of theorems $T_n, n = 1, 2, \dots$. It suffices to do 2 things.

- 1) Prove T_1 .
- 2) Prove that T_n true implies T_{n+1} true for all $n \geq 1$.

Note that this works by the well ordering axiom. If $S = \{n \in \mathbb{Z}^+ | T_n \text{ is false}\}$, then either S is empty or S has a least element q . But we know $q > 1$ by the fact that we proved T_1 . And we know that T_{q-1} is true since q is the least element of S . But then by 2) we know T_{q-1} implies T_q , contradicting $q \in S$.

Example 1. T_n is the formula used by Gauss as a youth to confound his teacher:

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}, \quad n = 1, 2, 3, \dots$$

We follow our procedure.

First, prove T_1 . $1 = \frac{1(2)}{2}$. Yes, that is certainly true.

Second, assume T_n and use it to prove T_{n+1} , for $n = 1, 2, 3, \dots$

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}$$

Add the next term in the sum, namely, $n + 1$, to both sides of the equation:

$$n + 1 = n + 1$$

Obtain

$$1 + 2 + \dots + n + (n + 1) = \frac{n(n+1)}{2} + (n + 1)$$

and finish by noting that

$$\frac{n(n+1)}{2} + (n + 1) = (n + 1) \left(\frac{n}{2} + 1 \right) = (n + 1) \left(\frac{n+2}{2} \right),$$

which gives us formula T_{n+1} .

Note: I personally find this proof a bit disappointing. It does not seem to reveal the underlying reason for the truth of such a formula and it requires that you believe in mathematical induction. Of course there are many other proofs of this sort of thing. For example, look at

$$1 + 2 + \cdots + n$$

$$n + (n - 1) + \cdots + 1$$

When you add corresponding terms you always get $n + 1$. There are n such terms. Thus twice our sum is $n(n + 1)$.

Example 2. The formula relating $n!$ and Gamma.

Assume the definition of the gamma function given in the preceding section of these notes makes sense:

$$\Gamma(s) = \int_0^{\infty} e^{-t} t^{s-1} dt, \text{ for } s > 0.$$

We want to prove that for $n=0,1,2,\dots$, we have

Formula G_n : $\Gamma(n + 1) = n! = n(n - 1)(n - 2) \cdots 2 \cdot 1.$

Here $0!$ is defined to be 1.

Proof. by Induction.

Step 1. Check G_0 .

$$\Gamma(0 + 1) = \int_0^{\infty} e^{-t} t^{1-1} dt = \int_0^{\infty} e^{-t} dt = -e^{-t} \Big|_{t=0}^{\infty} = 1 - 0 = 0!.$$

Step 2. Check that G_n implies G_{n+1} .

Assume G_n which is

$$n! = \Gamma(n + 1) = \int_0^{\infty} e^{-t} t^n dt.$$

Recall integration by parts which says that

$$\int u dv = uv - \int v du.$$

Set $u = e^{-t}$ and $dv = t^n$. Then $du = -e^{-t} dt$ and $v = \frac{1}{n+1} t^{n+1}$. Plug this into the integration by parts formula and get

$$\begin{aligned} n! &= \Gamma(n + 1) = \int_0^{\infty} e^{-t} t^n dt = \frac{1}{n + 1} e^{-t} t^{n+1} \Big|_{t=0}^{\infty} + \frac{1}{n + 1} \int_0^{\infty} e^{-t} t^{n+1} dt \\ &= 0 + \frac{1}{n + 1} \Gamma(n + 2). \end{aligned}$$

This says $(n + 1)n! = \Gamma(n + 2)$, which is formula G_{n+1} . This completes our induction proof. ■

There is also a second induction principle. See Lang, p. 10. You should be able to translate it into something you believe about dominoes. Of course, you have never really seen or been able to draw an infinite collection of dominos. You might want to try to draw a domino picture for the second form of mathematical induction.

5 Finite and Denumerable or Countable Sets

Definition 12 A set S is *finite with n elements* iff there is a 1-1, onto map

$$f : S \rightarrow \{1, 2, 3, \dots, n\}.$$

Write $|S| = n$ if this is the case.

Examples.

- 1) The empty set \emptyset is a finite set with 0 elements.
- 2) If set A has n elements, set B has m elements, and $A \cap B = \emptyset$, then $A \cup B$ has $n + m$ elements.

Proposition 13 *Properties of $|S|$ for finite sets S, T .*

- 1) $T \subset S$ implies $|T| \leq |S|$.
- 2) $|S \cup T| \leq |S| + |T|$.
- 3) $|S \cap T| \leq \min\{|S|, |T|\}$.
- 4) $|S \times T| = |S| \cdot |T|$.
- 5) Define $T^S = \{f : S \rightarrow T\}$. Then $|T^S| = |T|^{|S|}$.

Discussion.

These facts are fairly simple to prove. It is elementary combinatorics. For example, we sketch a proof of property 5).

Let $S = \{s_1, \dots, s_n\}$ and $f \in T^S$ means we can write $f(s_i) = t_i$, with $t_i \in T$, for $i = 1, 2, \dots, n$. Then $f \in T^S$ corresponds to the element $(t_1, \dots, t_n) \in \underbrace{T \times \dots \times T}_n$. This correspondence is 1-1, onto.

Definition 14 A set S is *denumerable (or countable and infinite)* iff there is a 1-1, onto map $f : \mathbb{Z}^+ \rightarrow S$.

Examples.

- 1) \mathbb{Z} =the set of all integers is denumerable.
- 2) $2\mathbb{Z}$ =the set of even integers is denumerable.

Corresponding statements for infinite sets to those of the preceding proposition defy intuition. Cantor's set theory boggled many minds. Luckily we only need to note a few things from Cantor's theory. Later we will have a new way to think about the size of infinite sets. For example length of an interval or area of a region in the plane or volume of a region in 3-space.

Proposition 15 *Facts About Denumerable Sets.*

Fact 1) a) If S is a denumerable set, then there exists a proper subset $T \subset S$, proper meaning $T \neq S$, such that there is a bijection $f : S \rightarrow T$.

Fact 1b) Any infinite subset of a denumerable set is also denumerable.

Fact 2) If sets S and T are denumerable, then so is the Cartesian product $S \times T$.

Fact 3) If $\{S_n\}_{n \geq 1}$ is a sequence of denumerable sets, then the union $\bigcup_{n \geq 1} S_n$ is denumerable.

Fact 4) The following sets are all denumerable:

\mathbb{Z}^+ =the positive integers

\mathbb{Z} =the integers

$2\mathbb{Z}$ =the even integers

$\mathbb{Z} - 2\mathbb{Z}$ =the odd integers

$\mathbb{Q} = \left\{ \frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0 \right\}$ = the rational numbers

Fact 5) The set \mathbb{R} of all real numbers is not denumerable. Here we view \mathbb{R} as the set of all decimal expansions. We will have more to say about it in the next section.

Proof. (See the stories after this proof for a more amusing way to see these facts).

Fact 1a) Suppose that $h:\mathbb{Z}^+ \rightarrow S$ is 1-1, onto. Write $h(n) = s_n$, for $n = 1, 2, 3, \dots$. That is we can think of S as a sequence $\{s_n\}_{n \geq 1}$ with the property that $s_n = s_m$ implies $n = m$. So let $T = \{s_2, s_3, s_4, \dots\} = S - \{s_1\}$. That is, take one element s_1 out of S . Define the map $g:\mathbb{Z}^+ \rightarrow T$ by $g(n) = s_{n+1}$. It should be clear that g is 1-1, onto. Thus T is denumerable.

Fact 1b) Suppose that T is an infinite subset of the denumerable set S . Then writing S as a sequence as in 1a, we have $S = \{s_1, s_2, s_3, s_4, \dots\}$ and $T = \{s_{k_1}, s_{k_2}, s_{k_3}, \dots\}$, with $k_1 < k_2 < k_3 < \dots < k_n < k_{n+1} < \dots$. We can think of T as a subsequence of S and then mapping $h:\mathbb{Z}^+ \rightarrow S$ is defined by $h(n) = s_{k_n}$, for all $n \in \mathbb{Z}^+$. Again it should be clear that h is 1-1 and onto.

Fact 2) Let $S = \{s_n\}_{n \geq 1}$ and $T = \{t_n\}_{n \geq 1}$. Then $S \times T = \{(s_n, t_m)\}_{(n,m) \in \mathbb{Z}^+ \times \mathbb{Z}^+}$. So we have a 1-1, onto map $f:\mathbb{Z}^+ \times \mathbb{Z}^+ \rightarrow S \times T$ defined by $f(n, m) = (s_n, t_m)$. Thus to prove 2) we need only show that there is a 1-1, onto map $g:\mathbb{Z}^+ \rightarrow \mathbb{Z}^+ \times \mathbb{Z}^+$. For then $f \circ g$ is 1-1 from \mathbb{Z}^+ onto $S \times T$. We will do this as follows by lining up the points with positive integer coordinates in the plane using the arrows as in Figure 11.

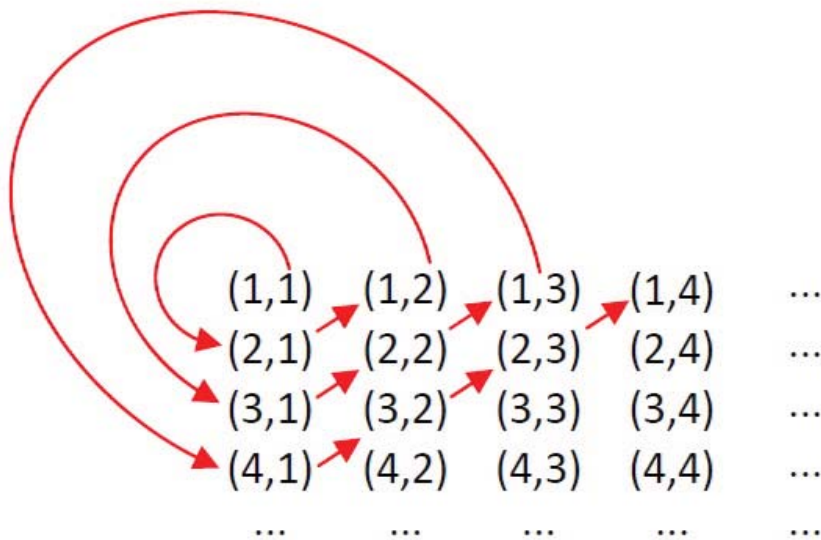


Figure 11: The arrows indicate the order to enumerate points with positive integer coordinates in the plane.

Start off at $(1, 1)$ and follow the arrows. Our list is

$$(1, 1), (2, 1), (1, 2), (3, 1), (2, 2), (1, 3), (4, 1), (3, 2), (2, 3), (1, 4), \dots$$

The general formula for the function $g^{-1} : \mathbb{Z}^+ \times \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ is

$$g(m, n) = \frac{(m + n - 2)(m + n - 1)}{2} + n.$$

Why? See Sagan, *Advanced Calculus*, p. 51.

The triangle with (m, n) in the line is depicted in Figure 12.

The triangle has $1 + 2 + 3 + \dots + [(m + n) - 2] = \frac{(m+n-2)(m+n-1)}{2}$ terms. This is the number of terms before $(m + n - 1, 1)$. Then to get to (m, n) you have to add n to this. Lang, *Undergraduate Analysis*, p. 13, uses a different function $g(m, n) = 2^m 3^n$. This does not map onto \mathbb{Z}^+ .

Fact 3) Let $S_m = \{s_{m1}, s_{m2}, s_{m3}, \dots\}$. Then $\bigcup_{m \geq 1} S_m = \{s_{m,n} | m, n \in \mathbb{Z}^+\}$. We can use arguments similar to those we just found to see that this set is denumerable. Of course we need to be careful since the map defined by $f(m, n) = s_{m,n}$ for $m, n \in \mathbb{Z}^+$ may not be 1-1.

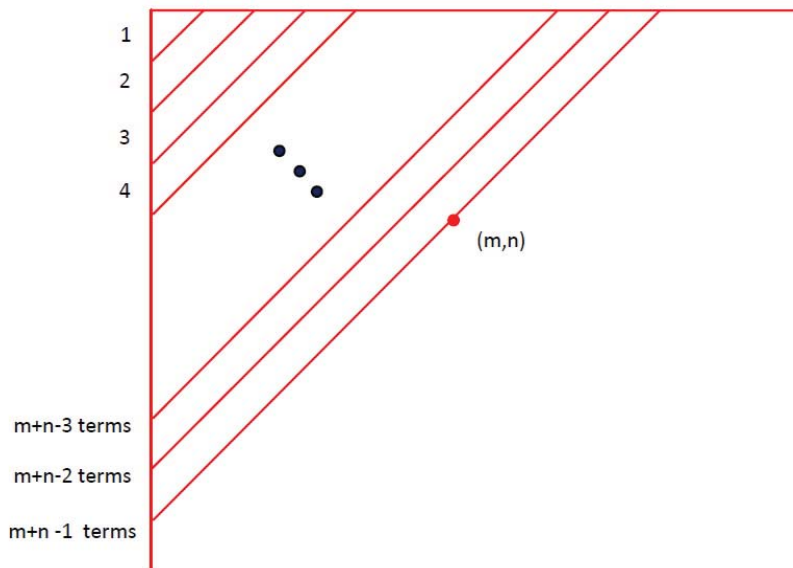


Figure 12: In the list to the right of the triangle we count the number of points with positive integer coordinates on each diagonal. The last diagonal goes through the point (m,n) .

Fact 4) These examples follow fairly easily from 1), 2) and 3). We will let the reader fill in the details.

Fact 5) Here we use **Cantor's diagonal argument** (Sagan, *Advanced Calculus*, p. 53) to see that the set of real numbers \mathbb{R} is not denumerable. This is a proof by contradiction. We will assume that \mathbb{R} is denumerable and deduce a contradiction. In fact, we look at the interval $[0, 1]$ and show that even this proper subset of \mathbb{R} is not denumerable. We can represent real numbers in $[0, 1]$ by decimal expansions like .12379285..... By this, we mean $\frac{1}{10} + \frac{2}{100} + \frac{3}{1000} + \frac{7}{10000} + \frac{9}{100000} + \dots$.

In general a real number in $[0, 1]$ has the decimal representation $\alpha = .a_1a_2a_3a_4\dots$, with $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Thus if $[0, 1]$ were denumerable, we'd have a list of decimals including every element of $[0, 1]$ exactly once as follows:

$$\begin{aligned} \alpha_1 &= .a_{11}a_{12}a_{13}a_{14}\dots \\ \alpha_2 &= .a_{21}a_{22}a_{23}a_{24}\dots \\ \alpha_3 &= .a_{31}a_{32}a_{33}a_{34}\dots \\ &\dots\dots\dots \\ &\dots\dots\dots \\ &\dots\dots\dots \end{aligned}$$

Here $a_j \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Look now at the stuff on the diagonal in this list. Then define a new decimal $\beta = .b_1b_2b_3b_4\dots$ by forcing it to differ from all of the α_j . This is done by defining

$$b_j = \begin{cases} 0, & \text{if } a_{jj} \neq 0 \\ 1, & \text{if } a_{jj} = 0. \end{cases}$$

We have thus found an element $\beta \in [0, 1]$ which is not on our great list. That is, we have found a contradiction to the existence of such a list. Why isn't β on the list? Well β cannot be equal to any α_j since the j th digit in β , namely b_j does not equal the j th digit in α_j , which is a_{jj} . ■

Moral: There are lots more real numbers than rational numbers.

6 Stories about the Infinite Motel - Interpretation of the Facts about Denumerable Sets

Reference: N. Ya Vilenkin, *Stories about Sets*.

Consider the mega motel of the galaxy with rooms labelled by the positive integers. See Figure 13. This motel extended across almost all of the galaxies.



Figure 13: The megamotel of the galaxy with denumerably many rooms.

A traveller arrived at the motel and saw that it was full. He began to be worried as the next galaxy was pretty far away. "No problem," said the manager and proceeded to move the occupant of room r_n to room r_{n+1} for all $n = 1, 2, 3, \dots$. Then room r_1 was vacant and the traveller was given that room. See Figure 14.

Moral: If you add or subtract an element from a denumerable set, you still have a denumerable set.

A few days later a denumerably infinite number of bears showed up at the motel which was still full. The angry bears began to growl. But the manager did not worry. He moved the guest in room r_n into room r_{2n} , for all $n = 1, 2, 3, \dots$. This freed up the odd numbered rooms for the bears. So bear b_n was placed in room r_{2n-1} , for all $n = 1, 2, 3, \dots$.

Moral. A union of 2 denumerable sets is denumerable.

The motel was part of a denumerable chain of denumerable motels. Later, when the galactic economy went into a depression, the chain closed all motels but 1. All the motels in the chain were full and the manager of the mega motel was told to find rooms for all the guests from the infinite chain of denumerable motels. This motel manager showed his cleverness again as Figure 15 indicates. He listed the rooms in all the hotels in a table so that hotel i has guests rooms $r_{i,1}, r_{i,2}, r_{i,3}, r_{i,4}, \dots$. Then, in a slightly different manner from the proof of Fact 2 in Proposition 15, he twined a red thread through all the rooms, lining up the guests so that he could put them into rooms in his motel.

Moral: The Cartesian product of 2 denumerable sets is denumerable.

But the next part of the story concerns a defeat of this clever motel manager. The powers that be in the commission of cosmic motels asked the manager to compile a list of all the ways in which the rooms of his motel could be occupied. This



Figure 14: A traveler t arrives at the full hotel. Guest g_n is moved to room r_{n+1} and the traveler t is put in room r_1 .



Figure 15: The manager of the mega motel has to put the guests from the entire chain of denumerable motels into his motel. He runs a red thread through the rooms to put the guests in order and thus into 1-1 correspondence with \mathbb{Z}^+ and the rooms in his motel.

list was supposed to be an infinite table. Each line of the table was to be an infinite sequence of 0's and 1's. At the n th position there would be a 1 if room r_n were occupied and a 0 otherwise. For example, the sequence 000000000000 would represent an empty motel. The sequence 1010101010101010..... would mean that the odd rooms were occupied and the even rooms empty.

The proof of Fact 5 in Proposition 15 (Cantor's diagonal argument) shows that this list is incomplete. For suppose the table is

$$\begin{aligned} \alpha_1 &= a_{11}a_{12}a_{13} \cdots a_{1n} \cdots \\ \alpha_2 &= a_{21}a_{22}a_{23} \cdots a_{2n} \cdots \\ \alpha_3 &= a_{31}a_{32}a_{33} \cdots a_{3n} \cdots \\ &\dots\dots\dots \\ \alpha_n &= a_{n1}a_{n2}a_{n3} \cdots a_{nn} \cdots \\ &\dots\dots\dots \end{aligned}$$

Now define $\beta = b_1b_2b_3 \cdots$ with $b_i \in \{0, 1\}$ by saying $b_n = 1$, if $a_{nn} = 0$, and $b_n = 0$, if $a_{nn} = 1$. Then β cannot be in our table at any row. For the n th entry in β cannot equal the n th entry in α_n , for any n .

So the set of all ways of occupying the motel is not denumerable. The motel manager failed this time.

Moral. The set of all sequences of 0s and 1s is not denumerable. Nor are the real numbers.

Part II

The Real Numbers

7 Pictures of our Cast of Characters

Figure 16 shows our favorite sets of numbers. Of course, they are all infinite sets and thus we cannot put all the elements in.

First we have $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, the integers, a discrete set of equally spaced points on a line, marching out to infinity.

Then we have the rational numbers $\mathbb{Q} = \{\frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0\}$. This set is everywhere dense in the real line; every open interval contains a rational. For example any open interval containing 0, must contain infinitely many of the numbers $\frac{1}{2^n}, n = 1, 2, 3, 4, \dots$. However, \mathbb{Q} is full of holes where $\sqrt{2}, e, \pi$ would be if they were rational but they are not.

The set of real numbers, \mathbb{R} , consists of all decimal expansions including that for

$$e = 2.71828 18284 59045 23536 02874 71352 66249 7757.....$$

It can be pictured as a continuous line, with no holes or gaps. You can think of the real numbers algebraically as decimals. By this we mean an infinite series:

$$\alpha = \sum_{j=-n}^{\infty} a_j 10^{-j}, \text{ with } a_j \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

In the usual decimal notation, we write $\alpha = a_1a_2a_3a_4a_5 \cdots$. This representation is not unique; for example, $0.999999 \cdots = 1$. To see this, use the **geometric series**

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \text{ for } |x| < 1.$$

Here I assume that you learned about infinite series in calculus. They are of course limits, which we have yet to define carefully. Anyway, back to our example, we have

$$0.999999 \cdots = \frac{9}{10} \sum_{n=0}^{\infty} \left(\frac{1}{10}\right)^n = \frac{9}{10} \frac{1}{1 - \frac{1}{10}} = 1.$$

\mathbb{Z} is the set of real numbers which have a decimal representation with only all 0's or all 9's after the decimal point. \mathbb{Q} is the set of real numbers with decimals that are repeating after a certain point. For example, $\frac{1}{3} = 0.33333333 \cdots$; $\frac{1}{7} = 0.142857142857142857 \cdots$. This too comes from the geometric series and the fact that $\frac{142857}{999999} = \frac{1}{7}$.

some integers



some rationals



some real numbers



Figure 16: The red dots indicate integers in the first line, rational numbers in the second line, and real numbers in the 3rd line. Of course we cannot actually draw all the rationals in an interval so we tried to indicate a cloud of points.

8 The Field Axioms for the Real Numbers

Given real numbers x, y, z we have unique real numbers $x + y, xy$ such that the following axioms hold $\forall x, y, z \in \mathbb{R}$.

A1. associative law for addition: $(x + y) + z = x + (y + z)$

A2. identity for addition: $\exists 0 \in \mathbb{R}$ s.t. $0 + x = x$

A3. inverses for addition: $\forall x \in \mathbb{R}, \exists -x \in \mathbb{R}$ s.t. $x + (-x) = (-x) + x = 0$.

A4. commutative law for addition: $x + y = y + x$

M1. associative law for multiplication: $x(yz) = (xy)z$

M2. identity for multiplication: $\exists 1 \in \mathbb{R}$ s.t. $1x = x1 = x$

M3. multiplicative inverses for non-zero elements: $\forall x \in \mathbb{R}$ s.t. $x \neq 0, \exists x^{-1} \in \mathbb{R}$ s.t. $xx^{-1} = 1 = x^{-1}x$

M4. commutative law for multiplication: $xy = yx$

D. distributive law: $x(y + z) = xy + xz$

Any set with 2 operations $+$ and \times that satisfy the preceding 9 axioms is called a **field**. The rational numbers \mathbb{Q} also satisfy these 9 axioms and are thus a field too. Mostly fields are topics studied in algebra, not analysis.

From these laws you can deduce the many facts that you know from school before college. For example we list a few facts.

Facts About \mathbb{R} that Follow from the Field Axioms.

Fact 1) $\forall x, y, z \in \mathbb{R}$, if $xy = xz$ and $x \neq 0$, then $y = z$.

Fact 2) $0 \cdot x = 0 \forall x \in \mathbb{R}$.

Fact 3) The elements 0 and 1 are unique.

Fact 4) $-(-x) = (-x)(-x) = x$.

Proof. Fact 1) Multiply the equation by x^{-1} which exists by M3. This gives

$$y = 1 \cdot y = (x^{-1} \cdot x)y = x^{-1}(xy) = x^{-1}(xz) = (x^{-1} \cdot x)z = 1 \cdot z = z.$$

Here we have used axioms M2, M3, M1.

Fact 2) Using our axioms M2, D, A2 we have

$$0 \cdot x + x = 0 \cdot x + 1 \cdot x = (0 + 1) \cdot x = 1 \cdot x = x.$$

It follows that $0 \cdot x + x = x$. Now subtract x from both sides (or equivalently add $-x$ to both sides) to get

$$(0 \cdot x + x) - x = x - x$$

which says by A1 and A3

$$0 \cdot x + (x - x) = 0.$$

Thus by A3 and A2, $0 \cdot x + 0 = 0$ and again by A2, we have $0 \cdot x = 0$.

Fact 3) and Fact 4) We leave these proofs to the reader. ■

9 Order Axioms for \mathbb{R}

The set \mathbb{R} has a subset P which we know as the set of positive real numbers. Then P satisfies the following **2 Order Axioms**:

Ord 1. $\mathbb{R} = P \cup \{0\} \cup (-P)$, where $-P = \{-x | x \in P\} =$ negative real numbers. Moreover, this union is disjoint; i.e., the intersection of any pair of the 3 sets is empty.

Ord 2. $x, y \in P$ implies $x + y$ and $xy \in P$.

Definition 16 For real numbers x, y we write $x < y$ iff $y - x \in P$. We write $x \leq y$ iff either $x < y$ or $x = y$.

All the usual properties of inequalities can be deduced from our 2 order axioms and this definition. We will do a few of these.

Facts about Order. $\forall x, y, z \in \mathbb{R}$

Fact 1) Transitivity. $x < y$ and $y < z$ implies $x < z$.

Fact 2) Trichotomy. For any $x, y, z \in \mathbb{R}$ exactly one of the following inequalities is true: $x < y$, $y < x$, or $x = y$.

Fact 3) $x < y$ implies $x + z < y + z$ for any $z \in \mathbb{R}$.

Fact 4) $0 < x$ iff $x \in P$.

Fact 5) If $0 < c$ and $x < y$, then $cx < cy$.

Fact 6) If $c < 0$ and $x < y$, then $cy < cx$.

Fact 7) $0 < 1$

Fact 8) If $0 < x < y$, then $0 < \frac{1}{y} < \frac{1}{x}$.

Proof. We will leave most of these proofs to the reader as Exercises. But let's do 1) and 7).

Fact 1) $x < y$ means $y - x \in P$. $y < z$ means $z - y \in P$. Then by Ord2 and the axioms for arithmetic in \mathbb{R} , we have $y - x + z - y = z - x \in P$. This says $x < z$.

Fact 7) note that by Ord 1, 1 must either be in P or $-P$, since $1 \neq 0$ (Why?). If $1 \in -P$, then $-1 \in P$ and Ord 2 says $(-1)(-1) = 1 \in P$, contradiction. ■

Definition 17 The *absolute value* $|x|$ of a real number x is defined by $|x| = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$

Equivalently, we can write

$$|x| = \sqrt{x^2}. \tag{5}$$

This is true because we always take the positive square root.

Properties of the Absolute Value.

For all $x, y, z \in \mathbb{R}$, we have the following properties.

Property 1) $|x| \geq 0 \forall x \in \mathbb{R}$ and $|x| = 0$ iff $x = 0$.

Property 2) $|xy| = |x| |y|$.

Property 3) the triangle inequality: $|x + y| \leq |x| + |y|$.

Proof. We leave all but 3) to the reader as **Exercises**.

Property 3) Using Formula (5), we have $|x + y|^2 = \left(\sqrt{(x + y)^2}\right)^2 = (x + y)^2 = x^2 + 2xy + y^2 = |x|^2 + 2xy + |y|^2 \leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2$.

Here we have used the fact that $x \leq |x|$ as well as property 2) of the absolute value. Now take square roots of both sides to finish the proof. To know that is legal you need to prove that $0 \leq x < y$ implies $0 \leq \sqrt{x} < \sqrt{y}$. We prove these 2 facts below. ■

Proofs of some Facts about Inequalities that were Needed in the Preceding Proof.

Fact 1) $x \leq |x|, \forall x \in \mathbb{R}$

Fact 2) $0 \leq x < y$ implies $0 \leq \sqrt{x} < \sqrt{y}$.

Proof. Fact 1) If $x \geq 0$, then $|x| = x$ and the result is clear. If $x < 0$, then $|x| = -x > 0 > x$, and the result holds by the transitivity property of $>$.

Fact 2) We prove it by contradiction. Assume $\sqrt{x} > \sqrt{y}$. Then $x = \sqrt{x}\sqrt{x} > \sqrt{x}\sqrt{y} > \sqrt{y}\sqrt{y} = y$. Again by transitivity, we have $x > y$, a contradiction to our hypothesis. ■

The absolute value is very useful. For example, it allows us to define the **distance** $d(x, y)$ between 2 real numbers x and y to be $d(x, y) = |x - y|$.

10 The Last Axiom for the Real Numbers

10.1 The Holes in \mathbb{Q}

We have stated 9 field axioms and 2 order axioms. Both of these axioms are also valid for the rational numbers; i.e., \mathbb{Q} is an ordered field just like \mathbb{R} . So what distinguishes \mathbb{Q} from \mathbb{R} ? We tried indicate this in Figure 16. \mathbb{Q} has holes like $\sqrt{2}$, π , e , while \mathbb{R} is a continuum. Of course the holes in \mathbb{Q} are as invisible as points. We will prove later that every interval on the real line contains a rational number.

There is a fairly simple axiom that allows us to say that \mathbb{R} has no holes. Before stating this axiom, let's explain why $\sqrt{2}$ is irrational. The Pythagoreans noticed this over 1000 years ago but kept it secret on pain of death. It seemed evil to them that the diagonal of a unit square or the hypotenuse of such a nice triangle as that in Figure 17 should be irrational.

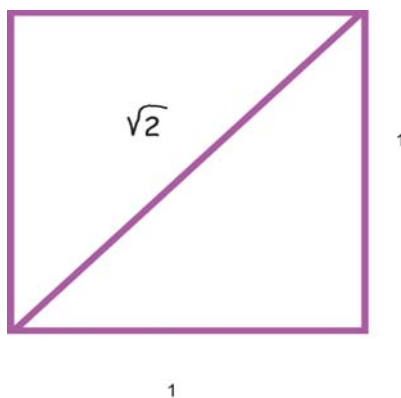


Figure 17: $\sqrt{2}$ is the length of the diagonal of a square each of whose sides has length 1.

Theorem 18 $\sqrt{2}$ is not rational.

Proof. (by contradiction)

Suppose $\sqrt{2}$ were rational and

$$\sqrt{2} = \frac{m}{n}, \text{ with } m, n \in \mathbb{Z}, n \neq 0. \quad (6)$$

We can assume that the fraction $\frac{m}{n}$ is in lowest terms; i.e., m and n have no common divisors. Square formula (6). This gives

$$2 = \frac{m^2}{n^2} \text{ and then } 2n^2 = m^2.$$

But then m must be even, since the square of an odd number is odd. So $m = 2r$, for some $r \in \mathbb{Z}$. Therefore

$$m^2 = 4r^2 = 2n^2.$$

Divide by 2 to see that n has to be even since n^2 is even. This is a contradiction since n and m now have a common divisor, namely, 2. ■

Similarly (or, better, using unique factorization of positive integers as a product of primes) one can show that \sqrt{m} is irrational for any positive integer m such that m is not the square of another integer. Thus $\sqrt{5}$, $\sqrt{6}$ are also irrational. You can do similar things for cube roots. It is harder to see that π and e are irrational. We will at least show e is irrational later. In fact, e and π are transcendental, meaning that they are not roots of a polynomial with rational coefficients. Of course, $\sqrt{2}$ is a root of $x^2 - 2$. See Hardy and Wright, *Theory of Numbers*, for more information.

A reference for weird facts about numbers (without proof) is David Wells, *The Penguin Dictionary of Curious and Interesting Numbers*. Here we learn that J. Lambert proved $\pi \notin \mathbb{Q}$ in 1766. And in 1882 Lindemann proved π to be transcendental. This implies that it is not possible to square the circle with ruler and compass - one of the 3 famous problems of antiquity. It asks for the ruler and compass construction of a square whose area equals that of a given circle.

The other 2 problems are angle trisection and cube duplication. To understand these problems you need to figure out the precise rules for ruler and compass constructions. Many undergraduate algebra books use Galois theory to show that all three problems are impossible.

Despite the provable impossibility of circle squaring, circle squarers abound. In 1897 the Indiana House of Representatives almost passed a law setting $\pi = \frac{16}{\sqrt{3}} \cong 9.2376$ - due to the efforts of a circle squarer.

Now many computers have been put to work finding more and more digits of π .

| year | digits | where |
|------|----------------|---------------------------|
| 1961 | 100,000 | U.S., Shanks and Wrench |
| 1967 | 500,000 | France |
| 1988 | 201 million | Japan, Y. Canada |
| 1989 | over 1 billion | U.S., Chudnovsky brothers |

What is the point of such calculations? Some believe that π is a normal number, which means that there is, in some sense, no pattern at all in the decimal expansion of π . But we digress into number theory. Anyway here are the first few digits:

$$\pi \cong 3.14159 26535 89793 23846 26433 83279 50288 41972 .$$

A reference is S. S. Hall, *Mapping the Next Millennium*, Chapter 13, which also contains a map obtained from trends in the first million digits of π as produced by the Chudnovsky brothers.

So, anyway we have lots of irrational numbers like $\sqrt{2}, \pi, e$. But we have another way to know that \mathbb{Q} is full of holes. Thinking of \mathbb{R} as the set of all infinite decimals, we know (by Cantor's diagonal argument) that \mathbb{R} is not denumerable, while \mathbb{Q} is denumerable. So \mathbb{R} is actually much much bigger than \mathbb{Q} .

But we only need one more axiom to fill in the holes in \mathbb{Q} .

10.2 Axiom C. The Completeness Axiom.

Before stating the completeness axiom, we need a definition.

Definition 19 We say that a real number a is the **least upper bound** (or **supremum**) of a set S of real numbers iff

$$x \leq a, \forall x \in S \text{ and if } x \leq b, \forall x \in S \text{ implies } a \leq b.$$

Notation: $a = l.u.b.S$ (or $a = \sup S$).

What is the *l.u.b.*? It is just what it claims to be; namely, the least of all the upper bounds for S (assuming S has upper bounds). There is also an analogous definition of **greatest lower bound** (*g.l.b.*) or **infimum**. We give the reader the job of writing down the definition. It is the greatest of all possible lower bounds.

If the set S is finite, there is no problem finding the *l.u.b.* or *g.l.b.* of S . Then you would say *l.u.b.S* is the maximum element of the finite set, for example. But when the set S is infinite, things become a lot less obvious. Why even should the *l.u.b.* of S exist? We will soon state an axiom that proclaims the existence of the *l.u.b.* or *g.l.b.* of a bounded set. Otherwise we would have no way to know. We will have proclaimed this existence by fiat. If we were kind, we would also produce a proof that the real numbers actually exist. I am sure you are not really worried about that. Or are you?

The *l.u.b.* of S is the left most real number to the right of all the elements of S . If you confuse right and left as much as I do, you may find this confusing.

Examples.

- 1) $l.u.b.\{\sqrt{2}, \pi, e\} = \pi$.
- 2) If $S = (0, 1) = \{x \in \mathbb{R} | 0 < x < 1\}$, then $l.u.b.S = 1$. This example shows that the least upper bound need not be an element of the set.
- 3) If $S = \{\frac{1}{n} | n \in \mathbb{Z}^+\}$, then $g.l.b.S = 0$. This example shows that the greatest lower bound need not be in the set.

The Completeness Axiom.

Suppose S is a non-empty set of real numbers and that S is bounded above; i.e., there is a real number B such that $x \leq B$ for all $x \in S$. Then there exists a real number $a = l.u.b.S$.

Assume $l.u.b. S = a \notin S$. This is the most interesting case. Figure 18 shows the picture of a least upper bound for a bounded set S . The red dots represent points of the infinite set S . Of course I cannot really put an infinite number of

points on a page in a lifetime. So you have to imagine them. Also points have no width. So what I am drawing is not really the points but a representation of what they would be if they had width. Again use your imagination. The point B (blue oval) at the right end is an upper bound for the set S . The point a (blue oval) is the least upper bound of S .

Assume $l.u.b. S = a \notin S$. One way to characterize $l.u.b. S$ is to note that, for a small positive ε , the interval $(a - \varepsilon, a)$ has infinitely many points from the set S . Why? Otherwise there would be a smaller $l.u.b. S$ than a . On the other hand, the interval $[a, a + \varepsilon)$ has no points from S , again assuming ε is small and positive. Why? Otherwise a would not even be an upper bound for S .

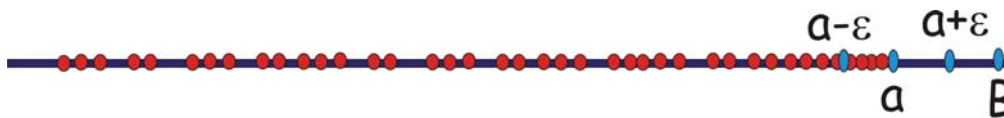


Figure 18: The infinite bounded set S is indicated with red dots. An upper bound B for S is indicated with a blue oval. The least upper bound a of S is indicated with a blue oval. The interval $(a - \varepsilon, a)$ contains infinitely many points of S . The interval $(a, a + \varepsilon)$ contains no points of S . Here $a - \varepsilon$ and $a + \varepsilon$ are also indicated with blue ovals. Since S is infinite we cannot really draw all its points. Moreover points are invisible really. So our figure is just a shadow of what is really happening. That you have to imagine.

The completeness axiom will imply all we need to know about limits and their existence. And that is what this course is about. Without it we would be in serious trouble.

The result of all this is that \mathbb{R} is characterized by 9+2+1 axioms; 9 field axioms, 2 order axioms, and 1 completeness axiom. To summarize that, we say that \mathbb{R} is a **complete ordered field**.

Next let us look at an example of how the completeness axiom can be used to fill in a hole in the rationals.

Example 1. A Set of Rational Numbers whose l.u.b. is $\sqrt{2}$ - found by Newton's Method.

See Figure 19.

Define $x_1 = 2$, $x_2 = \frac{1}{2}x_1 + \frac{1}{x_1}$, ..., $x_n = \frac{1}{2}x_{n-1} + \frac{1}{x_{n-1}}$, $n = 2, 3, 4, 5, \dots$. In this way, we have an inductive definition of an infinite set $S = \{x_n | n \in \mathbb{Z}^+\}$. The first few elements of S are $2, 1.5, 1.416666666\dots, 1.414215686, \dots$. We claim that the least upper bound of this set is $\sqrt{2}$. We will prove this claim later after noting that $\{x_n\}_{n \geq 1}$ is an decreasing sequence that is bounded below and thus must have a limit (which we are about to define), which is $\sqrt{2}$. This says that 2 is the $l.u.b.$ of S .

Note on Newton's Method.

This is a method which often approximates the root of a polynomial very well. In this case the polynomial is $x^2 - 2$. To find a root near $x_1 = 2$, you need to look at the tangent to the curve $y = f(x) = x^2 - 2$ at the point $(x_1, f(x_1))$. See Figure 19. The point where that tangent line intersects the x -axis is x_2 . To find it, look at the slope of the tangent at $(x_1, f(x_1))$ which is $f'(x_1) = 2x_1 = 4$. Then use the point-slope equation of a line to get:

$$\begin{aligned} 4 &= \frac{f(x_1) - 0}{x_1 - x_2} \\ &= \frac{2}{2 - x_2}. \end{aligned}$$

That says $x_2 = 1.5$.

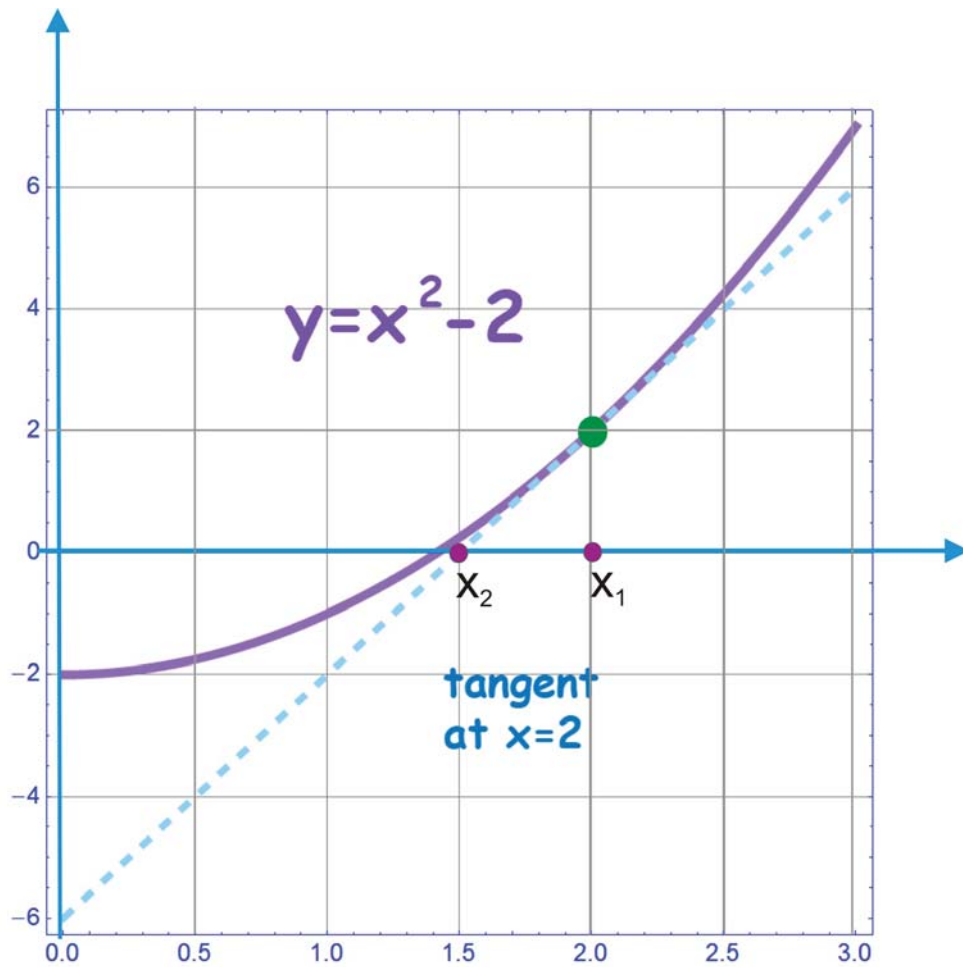


Figure 19: Graph showing the function $y = x^2 - 2$ as well as its tangent line at the point $x_1 = 2$. The 1st two Newton approximations to $\sqrt{2}$ are given. The first approximation is $x_1 = 2$. The second approximation is x_2 which is the intersection of the tangent line at x_1 and the x -axis.

Part III

Limits

11 Definition of Limits

Recall that a **sequence** $\{x_n\}_{n \geq 1} = \{x_1, x_2, x_3, x_4, \dots\}$ has indices n which are positive integers. We do not assume that the mapping from \mathbb{Z}^+ to $\{x_n\}_{n \geq 1}$ is 1-1. In fact a possible sequence has all x_n equal to the same number, say 2. No problem finding the limit of that sequence. You can think of a sequence as a vector with infinitely many components.

Definition 20 If $\{x_n\}_{n \geq 1}$ is a sequence of real numbers, we say that the real number L is the **limit** of x_n as n goes to ∞ and write $L = \lim_{n \rightarrow \infty} x_n$ iff, for every $\varepsilon > 0$ there is an $N \in \mathbb{Z}^+$ (with N depending on ε) such that $n \geq N$ implies $|x_n - L| < \varepsilon$.

In this definition you are supposed to think of ε as an arbitrarily small positive guy. Paul Erdős used to call children "epsilons." A computer might think $\varepsilon = 10^{-10}$. Such a small number of inches would be invisible. We have tried to draw a picture of a sequence approaching a limit. See Figure 20. Here we graph points (n, x_n) and $L = \lim_{n \rightarrow \infty} x_n$. Given a small positive number ε , we are supposed to be able to find a positive integer $N = N(\varepsilon)$ depending on ε so that every (n, x_n) , for $n \geq N(\varepsilon)$ is in the shaded box. Then you have to imagine taking an even smaller ε say $\frac{\varepsilon}{10^9}$ and there will be a new probably larger $N = N(\frac{\varepsilon}{10^9})$ so that all (n, x_n) with $n \geq N(\frac{\varepsilon}{10^9})$ are in the new smaller version of the shaded box. These x_n will be extremely close to a . And the point is that you can make all but a finite number of the x_n as close you want to a .

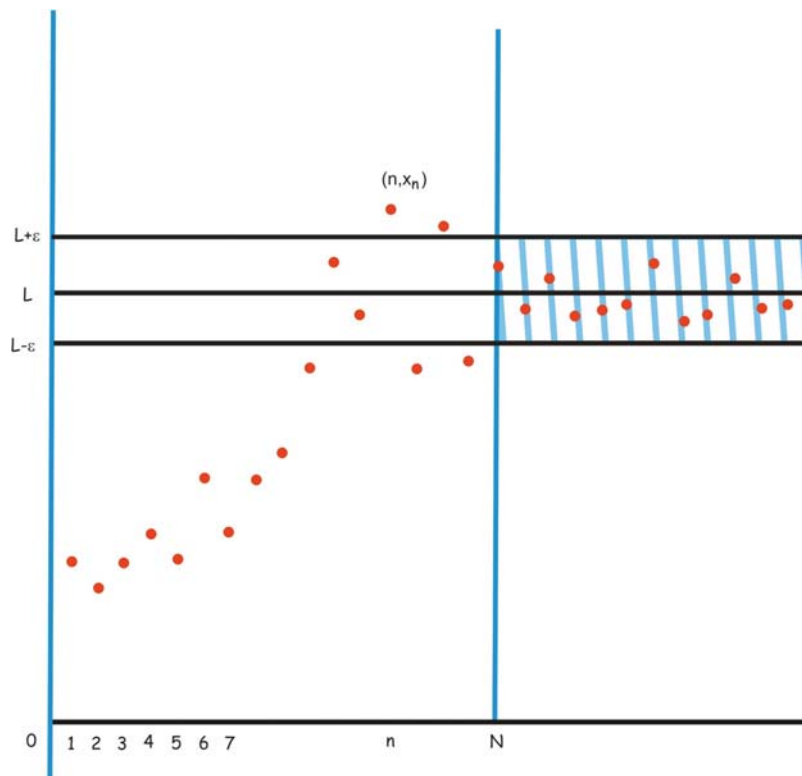


Figure 20: illustration of the definition of limit

This definition is one of the most important in the course. It was not given by I. Newton (1642-1727) or W. Leibniz (1646-1716) when they invented calculus in the 1600s. Our definition makes the idea of the sequence x_n approaching a real number L very precise by saying the distance between x_n and L is getting smaller than any small positive epsilon. Memorize it or stop reading now. We will have a similar definition of a limit of a function $f(x)$ as x approaches a finite real number a .

This definition of limit dates from the 1800s. It is due to B. Bolzano (1781-1848), A. Cauchy (1789-1857) and K. Weierstrass (1815-1897). Weierstrass was the advisor of the first woman math. prof. - Sonya Kovalevsky. The main fact about Cauchy that comes to my mind is that he lost that memoir of Galois. See Edna Kramer, *The Nature and Growth of Modern Mathematics*. An entertaining story book on the history of math. is E. T. Bell, *Men of Mathematics*.

Now we have a precise meaning to apply to our example above obtained using Newton's method to find a sequence of rationals approaching $\sqrt{2}$. We will do the proof later after we know more about limits. Let's consider another favorite example.

Example 2. A Sequence of Rational Numbers Approaching e .

Define e by the Taylor series for e^x . That is,

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

This means that e is the limit of the sequence of partial sums

$$s_n = \sum_{k=0}^n \frac{1}{k!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \cdots + \frac{1}{n!}.$$

So we have $s_1 = 1, s_2 = 2, s_3 = 1 + 1 + \frac{1}{2} = 2.5, s_4 = 1 + 1 + \frac{1}{2} + \frac{1}{6} \cong 2.6667, \dots$ This sequence does not converge as fast as that in example 1.

Example 3. The Simplest Limit (except for a constant sequence).

Define the sequence $x_n = \frac{1}{n}$, for all $n \in \mathbb{Z}^+$. I think it is obvious what the limit of this sequence is; namely,

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

To prove this from the definition of limit, we need to show that given $\varepsilon > 0$, we can find $N \in \mathbb{Z}^+$ so that $n \geq N$ implies $|\frac{1}{n} - 0| < \varepsilon$. That means, we need

$$\frac{1}{n} < \varepsilon, \text{ if } n \geq N.$$

The inequality is equivalent to saying $n > \frac{1}{\varepsilon}$. This means that we should take $N = \lfloor \frac{1}{\varepsilon} \rfloor + 1$. Here the **floor** of $x = \lfloor x \rfloor =$ the greatest integer $\leq x$. For then if $n \geq N \geq \frac{1}{\varepsilon} + 1 > \frac{1}{\varepsilon}$, we have $n > \frac{1}{\varepsilon}$.

To picture this you could graph the points $(n, \frac{1}{n})$ as in Figure 21. We seek the limit of the y-coordinates as the x-coordinates approach infinity.

Example 3 is this sort of limit that anyone could work out - even before having the wonderful Definition 20. More subtle limits however could confuse experts such as Cauchy himself. We will see more examples after considering the main facts about limits of sequences.

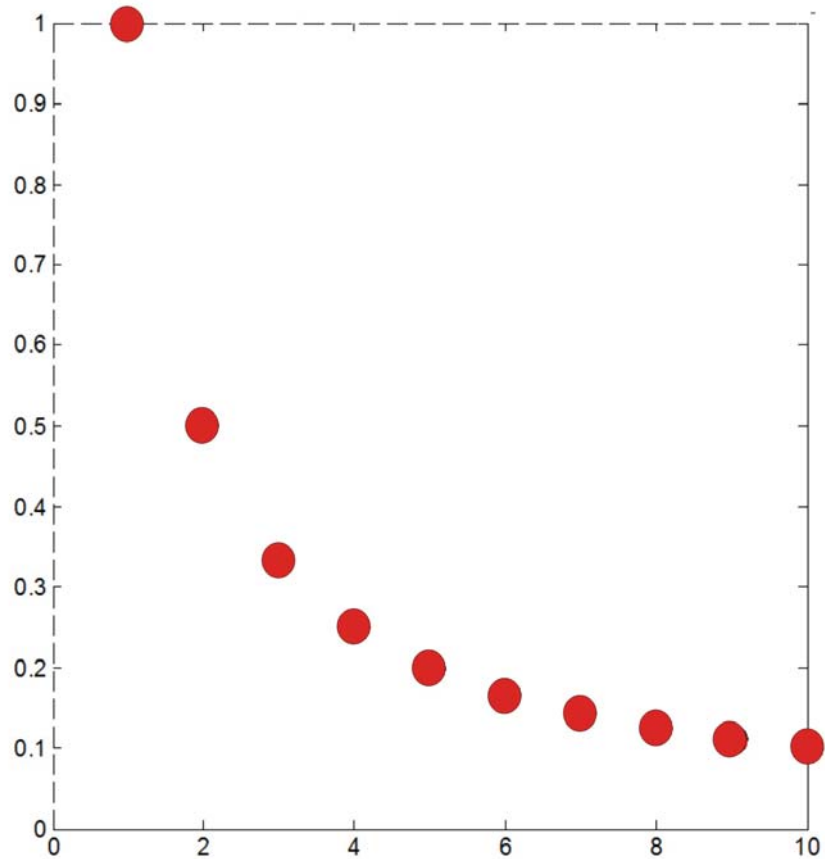


Figure 21: The red points are $(n, \frac{1}{n})$, for $n = 1, 2, \dots, 10$. It is supposed to be clear that the y-coordinates are approaching 0 as the x-coordinates march out to infinity. That is the points are getting closer and closer to the y-axis. Of course you cannot actually see what is happening at infinity.

12 Facts About Limits

Now that we have seen a few examples, perhaps we should prove a few things so that we can refer to them rather than reprove them every time.

Facts About Limits of Sequences of Real Numbers.

Fact 1) (Uniqueness) If the limit $a = \lim_{n \rightarrow \infty} x_n$ exists, then a is unique; i.e., if also $b = \lim_{n \rightarrow \infty} x_n$, then $a = b$.

Fact 2) (Limit Exists implies Sequence Bounded) If $a = \lim_{n \rightarrow \infty} x_n$ exists, then the set $\{x_n | n \in \mathbb{Z}^+\}$ is bounded above and below.

Fact 3) (Limit of a Sum is the Sum of the Limits) If we have limits $a = \lim_{n \rightarrow \infty} x_n$ and $b = \lim_{n \rightarrow \infty} y_n$, then $\lim_{n \rightarrow \infty} (x_n + y_n) = a + b$.

Fact 4) (Limit of a Product) If we have limits $a = \lim_{n \rightarrow \infty} x_n$ and $b = \lim_{n \rightarrow \infty} y_n$, then $\lim_{n \rightarrow \infty} (x_n y_n) = ab$.

Fact 5) (Limit of a Quotient) If we have limits $a = \lim_{n \rightarrow \infty} x_n$ and $b = \lim_{n \rightarrow \infty} y_n$, and in addition $b \neq 0$, then we have $y_n \neq 0$, for all n sufficiently large and $\frac{a}{b} = \lim_{n \rightarrow \infty} \frac{x_n}{y_n}$, assuming we throw out the finite number of n such that $y_n = 0$.

Fact 6) (Limit of an Increasing Bounded Sequence) Suppose that $x_n \leq x_{n+1} \leq B$, for all $n \in \mathbb{Z}^+$. Then the limit $a = \lim_{n \rightarrow \infty} x_n$ exists. Moreover, a is the least upper bound of the set of all elements in the sequence. That is,

$$a = \lim_{n \rightarrow \infty} x_n = l.u.b.\{x_n | n \in \mathbb{Z}^+\}.$$

There is an analogous result for decreasing sequences which are bounded below.

Fact 7) (Limits Preserve \leq) Suppose $a = \lim_{n \rightarrow \infty} x_n$ and $b = \lim_{n \rightarrow \infty} y_n$ and $x_n \leq y_n$, for all n . Then $a \leq b$.

Proof. Fact 1) Let us postpone this one to the end of this subsection. It may seem to be the most obvious but it requires a little thought.

Fact 2) Let $\varepsilon = 1$ be given. Then there exists N such that $n \geq N$ implies $|x_n - L| < 1$. By properties of inequalities, this implies $|x_n| = |x_n - L + L| \leq |x_n - L| + |L| < 1 + |L|$, if $n \geq N$. It follows that a bound on the sequence is $\max\{|x_1|, \dots, |x_{N-1}|, |L| + 1\}$.

Fact 3) Suppose we are given an $\varepsilon > 0$. Then we know we can find N and M depending on ε so that

$$n \geq N \text{ implies } |x_n - a| < \varepsilon \text{ and } n \geq M \text{ implies } |y_n - b| < \varepsilon. \quad (7)$$

Therefore if we take $K = \max\{N, M\}$ and $n \geq K$, we have

$$|x_n + y_n - (a + b)| \leq |x_n - a| + |y_n - b| < 2\varepsilon.$$

You may worry that we got 2ε rather than ε . The trick to get rid of 2, is to start out in formula (7) with $\frac{\varepsilon}{2}$ rather than ε . This is called an " $\frac{\varepsilon}{2}$ proof." Watch out for the $\frac{\varepsilon}{10}$ proofs!

Fact 4) The main idea is to start with what you need to prove: for n large enough

$$|x_n y_n - ab| < \varepsilon. \quad (8)$$

In order to get this, given our hypotheses, we use a trick that you may remember from calculus, if you ever proved the formula for the derivative of a product. We know $|x_n - a|$ is eventually small for large n , so we should be able to show $|x_n y_n - x_n b|$ is small. Thus we subtract $x_n b$ from $x_n y_n - ab$ and add it back in to obtain

$$|x_n y_n - ab| = |x_n y_n - x_n b + x_n b - ab|.$$

Now use the triangle inequality and the multiplicative property of the absolute value, finding that

$$|x_n y_n - ab| = |x_n y_n - x_n b + x_n b - ab| \leq |x_n y_n - x_n b| + |x_n b - ab| = |x_n| |y_n - b| + |b| |x_n - a|. \quad (9)$$

In order to make $|x_n| |y_n - b| + |b| |x_n - a|$ small, we actually need to know that $|x_n|$ is not blowing up as $n \rightarrow \infty$. Fact 2 tells us that there is a positive number K such that $|x_n| \leq K$, for all n . This, plus inequality (9), implies

$$|x_n y_n - ab| \leq K |y_n - b| + |b| |x_n - a|. \quad (10)$$

Given $\varepsilon > 0$, we can find N and M depending on ε so that we have an improved version of the inequalities (7)

$$n \geq N \text{ implies } |x_n - a| < \frac{\varepsilon}{1 + |b|} \text{ and } n \geq M \text{ implies } |y_n - b| < \frac{\varepsilon}{K}. \quad (11)$$

We divide by $1 + |b|$ rather than $|b|$ because b might be 0. We know $K > 0$ so we don't need to add 1 to K to avoid dividing by 0.

Now we combine inequalities (10) and (11) to get the desired inequality (8).

Fact 5) We will prove this sort of thing later. The reader should think about it though, using ideas similar to the proof of 4).

Fact 6) We know that the set $S = \{x_n | n \in \mathbb{Z}^+\}$ is non-empty and bounded. Therefore, by the completeness axiom, it has a least upper bound which we will call a . We want to show that $a = \lim_{n \rightarrow \infty} x_n$. Suppose that we have been given $\varepsilon > 0$. Look at $a - \varepsilon$. We know that $a - \varepsilon < a$ and thus $a - \varepsilon$ cannot be an upper bound for the set S of all $x_n, n \in \mathbb{Z}^+$. Look at Figure 22.

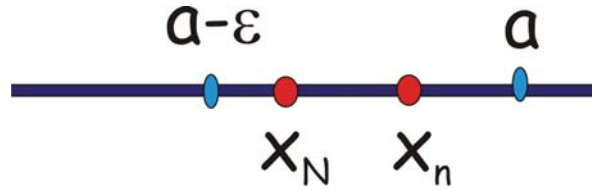


Figure 22: Picture of the proof of fact 6. $\{x_n\}$ is a bounded increasing sequence and $a = \text{lub}\{x_n | n \in \mathbb{Z}^+\}$. Here we assume $n \geq N$. We know N exists such that $x_N \in (a - \varepsilon, a]$ as $a - \varepsilon$ is not an upper bound for the set of all $x_k, k \in \mathbb{Z}^+$.

This means there is an $N \in \mathbb{Z}^+$ so that $a - \varepsilon < x_N \leq a$. Since $\{x_n\}$ is an increasing sequence, this means that for all $n \geq N$ we have

$$a - \varepsilon < x_N \leq x_n \leq a.$$

This implies that $|x_n - a| < \varepsilon$ if $n \geq N$. Therefore according to our definition of limit, $a = \lim_{n \rightarrow \infty} x_n$.

Fact 7) Using Facts 3 and 4, it suffices to look at $z_n = y_n - x_n \geq 0, \forall n$. We know that $\lim_{n \rightarrow \infty} z_n = b - a$ and we must show that $\lim_{n \rightarrow \infty} z_n \geq 0$. Do a proof by contradiction looking at Figure 23.

We leave the details to the reader as we will return to do a more general version of this argument later. Next we need 2 lemmas which will prove useful now and in the future. ■

Lemma 21 Suppose that a is a real number such that $|a| < \varepsilon$ for all $\varepsilon > 0$. Then $a = 0$.

Proof. We do a proof by contradiction again. If a is not 0, then $|a| > 0$ and then, we can take $\varepsilon = \frac{|a|}{2}$. This means $|a| < \frac{|a|}{2}$. But that is absurd as it implies $1 < \frac{1}{2}$ and thus $2 < 1$ and $1 < 0$. We have our contradiction. ■

Lemma 22 The set \mathbb{Z}^+ of positive integers is not bounded above.

Proof. Again we need a proof by contradiction. Suppose that a real number b is an upper bound for the set \mathbb{Z}^+ . By the completeness axiom, then \mathbb{Z}^+ has a least upper bound a . But this means that $a - 1$ is not an upper bound for \mathbb{Z}^+ and there is a positive integer n such that $a - 1 < n$. Therefore $a < n + 1$, by a property of inequalities. But then we have a contradiction as $n + 1 \in \mathbb{Z}^+$ and a was supposed to be an upper bound for \mathbb{Z}^+ . ■

Now we can prove our 1st fact about limits.

Proof. of Fact 1 About Limits.

If $a = \lim_{n \rightarrow \infty} x_n$ and $b = \lim_{n \rightarrow \infty} x_n$, we must show $a = b$. The number $a - b$ satisfies the hypothesis of Lemma 21. Why? We have

$$|a - b| = |a - x_n + x_n - b| \leq |a - x_n| + |x_n - b|.$$

Given, $\varepsilon > 0$, we know there must be positive integers N and M such that $n \geq N$ implies $|a - x_n| < \frac{\varepsilon}{2}$ and $n \geq M$ implies $|b - x_n| < \frac{\varepsilon}{2}$. It follows that $n \geq \max\{N, M\}$ implies

$$|a - b| = |a - x_n + x_n - b| \leq |a - x_n| + |x_n - b| < \varepsilon.$$

Thus, by Lemma 21, we see that $a = b$. Limits are unique. ■

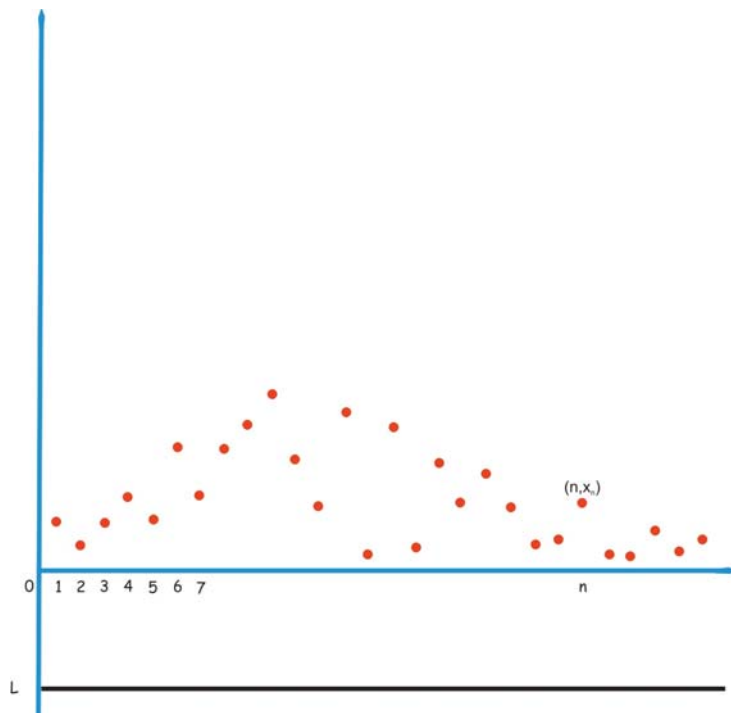


Figure 23: Picture of a proof by contradiction in which we assume that the sequence $\{x_n\}$ is non-negative while the limit L is negative. But then $|x_n - L| \geq |L| > 0$ for all n . This contradicts the definition of limit.

13 More Examples of Limits of Sequences

Example 1. $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$.

Here $x_n = \left(1 + \frac{1}{n}\right)^n$, for $n = 1, 2, 3, \dots$. So we have

$$x_1 = 2, x_2 = \left(1 + \frac{1}{2}\right)^2 = 2.25, x_3 = \left(1 + \frac{1}{3}\right)^3 \cong 2.3704, \dots, x_{30} = \left(1 + \frac{1}{30}\right)^{30} \cong 2.6743.$$

This is an increasing sequence bounded above by 4. It therefore has a limit. It can be shown (using l'Hopital's rule after taking the natural logarithm) that the limit is $e \cong 2.71828$. We will say more about this later. See Section 23.

Note that $y_n = \left(1 + \frac{1}{n}\right)^{n+1}$ is a decreasing sequence, also approaching e as n goes to infinity. We have, for example

$$y_1 = 4, y_2 = \left(1 + \frac{1}{2}\right)^3 \cong 3.375, y_3 = \left(1 + \frac{1}{3}\right)^4 \cong 3.1605, \dots, y_{30} = \left(1 + \frac{1}{30}\right)^{31} \cong 2.7635.$$

Example 2. $x_n = \cos(n\pi) = (-1)^n, n = 1, 2, 3, \dots$. This sequence has no limit as n goes to infinity. The sequence alternates between 1 and -1. Thus it cannot decide on a limit.

Proof. To prove there is no limit, proceed by contradiction. If $L = \lim_{n \rightarrow \infty} (-1)^n$, then according to the definition of limit we can take $\varepsilon = 1$ (or any positive number), and find N so that $n \geq N$ implies $|x_n - L| < 1$. This means that for n even and large we have

$$|1 - L| < 1 \text{ or equivalently } -1 < 1 - L < 1$$

and for n odd and large we have

$$|1 + L| = |-1 - L| < 1 \text{ or equivalently } -1 < 1 + L < 1.$$

But then we can add the inequalities and obtain $-2 < 2 < 2$. Contradiction. Thus the sequence has no limit. ■

Example 3. $\lim_{n \rightarrow \infty} \frac{n}{3n-2} = \frac{1}{3}$.

Proof. We need to show that the distance between $\frac{n}{3n-2}$ and $\frac{1}{3}$ is small for large n . To do this, look at the following

$$\left| \frac{n}{3n-2} - \frac{1}{3} \right| = \frac{1}{3} \left| \frac{3n - (3n-2)}{3n-2} \right| = \frac{1}{3} \left| \frac{2}{3n-2} \right| < \varepsilon.$$

This inequality is equivalent to

$$3n - 2 > \frac{2}{3\varepsilon},$$

which says

$$n > \frac{1}{3} \left(\frac{2}{3\varepsilon} + 2 \right).$$

Such an n exists by Lemma 22 above.

Another proof can be found using some high school algebra to see that

$$\frac{n}{3n-2} = \frac{n}{3n-2} \frac{\frac{1}{n}}{\frac{1}{n}} = \frac{1}{3 - \frac{2}{n}}.$$

Then use the fact that $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ which was proved above and Facts about limits stated above to prove the result. ■

Next recall the example of a sequence approaching $\sqrt{2}$ obtained using Newton's method.

Example 4. Define $x_1 = 2$, $x_2 = \frac{1}{2}x_1 + \frac{1}{x_1}$, ..., $x_{n+1} = \frac{1}{2}x_n + \frac{1}{x_n}$, $n = 1, 3, 4, 5, \dots$. Then $\lim_{n \rightarrow \infty} x_n = \sqrt{2}$.

To prove that this limit is indeed correct, using facts about limits proved above, it suffices for us to show the following.

Claim. $\{x_n\}$ is a decreasing sequence bounded below; i.e.,

$$x_n > x_{n+1} > 1, \text{ for all } n = 1, 2, 3, \dots$$

Once this claim is proved, it is not hard to see that the limit must exist (by the analog of Fact 6 for decreasing bounded sequences) and that it must be $\sqrt{2}$. If $L = \lim_{n \rightarrow \infty} x_n$, then look at the recursion

$$x_{n+1} = \frac{1}{2}x_n + \frac{1}{x_n}.$$

Take the limit as $n \rightarrow \infty$ to see that (using the appropriate facts about limits)

$$L = \frac{L}{2} + \frac{1}{L}.$$

Multiply by $2L$ to obtain $2L^2 = L^2 + 2$. Thus $L^2 = 2$. Therefore $L = \pm\sqrt{2}$. Why must L be positive? Use Fact 7.

Proof. of the Claim.

Here we use mathematical induction. We give the induction step, taking $a = x_n$ and $b = x_{n+1}$. We need to show that if $a^2 > 2$ and $a > 0$ then $b = \frac{a}{2} + \frac{1}{a}$ implies $0 < b < a$ and $b^2 > 2$ which implies $b > 1$.

To see this, note that

$$a - b = a - \left(\frac{a}{2} + \frac{1}{a} \right) = \frac{a}{2} - \frac{1}{a} = \frac{a^2 - 2}{2a}.$$

Then $a^2 > 2$ and $a > 0$ imply that $a - b > 0$.

Next look at

$$b^2 - 2 = \left(\frac{a}{2} + \frac{1}{a} \right)^2 - 2 = \frac{a^2}{4} + 1 + \frac{1}{a^2} - 2 = \left(\frac{a}{2} - \frac{1}{a} \right)^2 > 0.$$

So $b^2 > 2$. The proof of the claims is complete which finishes the proof that $\lim_{n \rightarrow \infty} x_n = \sqrt{2}$. ■

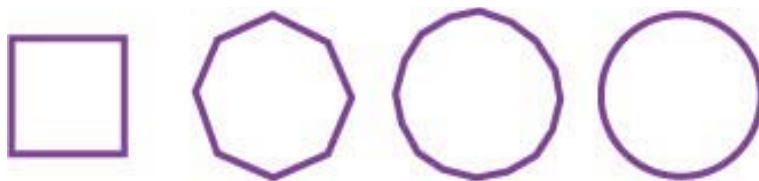


Figure 24: n-gons approaching a circle of radius 1 for $n = 4, 8, 16$.

14 Some History. Dedekind Cuts.

Why is the completeness axiom necessary? Why do we need to understand limits? This idea is fundamental to most of applied mathematics. It is central to differential equations and therefore to the theory of earthquakes and cosmology. The idea of "gradually getting there," "tending towards," "approaching" is one of the most basic. We cannot actually draw a picture of what is happening to x_n when n has moved infinitely far out. The idea is subtle.

You can sometimes see it happen on a computer. Look for example at Figure 24.

Here an equilateral polygon approaches a circle. So its area must approach that of a circle. This gives a way to approximate π . Archimedes did this around 250 B.C. Assume the radius of the circle is 1. Then

4 sides give the area 2
 8 sides give the area 2.828
 16 sides give the area 3.061.

Around 1821 Cauchy had formulated a principle of convergence of sequences of real numbers. His idea was to use the idea of approximation. We will give the definition of Cauchy sequence soon. It gives a useful way to construct the real numbers as well as a criterion for convergence of a sequence. See V. Bryant, *Yet Another Introduction to Analysis* for more examples.

The need for clarification of the concept of a real number became apparent in 1826 when Abel corrected Cauchy's belief that a sequence of continuous functions must have a continuous limit function. This showed that intuition can be very misleading when investigating limits. See G. Temple, *100 Years of Math.*, for more discussion of the history of the concept of limit.

The completeness axiom for \mathbb{R} can be stated in a different way. If A and B are non-empty sets of real numbers such that $a \leq b$ for all $a \in A$ and $b \in B$, then there exists a real number ω so that $\omega \geq a$ for all $a \in A$ and $\omega \leq b$ for all $b \in B$. This axiom is another way of saying there are no holes in the real line. V. Bryant, *Yet Another Introduction to Analysis*, p. 11, calls ω "Piggy-in-the-middle." See Figure 25.

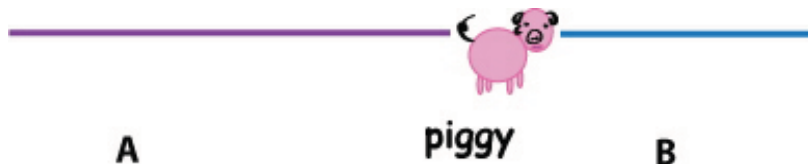


Figure 25: piggy in the middle - real number between 2 sets A and B such that $a \leq b$ for every $a \in A$ and $b \in B$.

In 1872 Dedekind used this sort of idea to construct the real numbers. A **Dedekind cut** consists of 2 sets A and B of rational numbers such that

- 1) $\mathbb{Q} = A \cup B$
- 2) $a \in A$ and $b \in B$ implies $a \leq b$.

15 Cauchy Sequences

Next we want to define Cauchy sequences. This gives a convergence criterion, a new version of the completeness axiom, and a way to construct the real numbers, the space of Lebesgue integrable functions, and Hensel's space of p -adic numbers for every prime p (as space which has many applications in number theory).

Definition 23 A sequence $\{x_n\}$ of real numbers is a **Cauchy sequence** iff for every $\varepsilon > 0$ there is an $N \in \mathbb{Z}^+$ such that $n, m \geq N$ implies $|x_n - x_m| < \varepsilon$.

In this definition, we just ask that the distance between x_n and x_m is less than ε for all but a finite number of n and m . To say this another way, we ask that the sequence elements x_n and x_m become arbitrarily close as $m, n \rightarrow \infty$. The useful thing about this convergence criterion is that it does not require you to know what the limit is.

Cauchy made this definition in 1821. He did not prove the following theorem.

Theorem 24 Every Cauchy sequence of real numbers has a limit.

This theorem is actually a consequence of our completeness axiom and we will soon give a proof. In fact, the theorem is logically equivalent to the completeness axiom. Thus one could construct \mathbb{R} (as Cantor did in 1883) as "limits of" Cauchy sequences of rational numbers. Here we identify 2 Cauchy sequences if they converge to the same limit. This gives a construction of \mathbb{R} which is analogous to that used to construct the spaces of Lebesgue integrable functions out of the space of continuous functions.

Before thinking about proving the preceding theorem, we need to think about subsequences.

Definition 25 Suppose that $\{x_n\}$ is a sequence. A **subsequence** $\{x_{n_k}\}$ is a sequence obtained by selecting out certain terms of the original sequence. Here

$$1 \leq n_1 < n_2 < n_3 < \dots < n_k < n_{k+1} < \dots$$

Example. Consider the sequence $x_n = (-1)^n$. One subsequence consists of the terms with even indices $x_{2n} = 1$. Another subsequence consists of the terms with odd indices $x_{2n+1} = -1$. Both of these subsequences converge (since they are constant) even though the original sequence does not converge. According the Fact 4 below this gives another proof that the original sequence does not converge.

Facts About Cauchy (and Other) Sequences of Real Numbers.

Fact 1. If a sequence of real numbers has a limit then it is a Cauchy sequence.

Fact 2. Cauchy sequences of real numbers are bounded.

Fact 3. Any bounded sequence of real numbers has a convergent subsequence.

Fact 4. For a Cauchy sequence of real numbers, if a subsequence converges to L , then the original sequence also converges to L .

Proof. Fact 1.

Suppose $\lim_{n \rightarrow \infty} x_n = L$. Given $\varepsilon > 0$, we know there is an $N \in \mathbb{Z}^+$ so that $n \geq N$ implies $|x_n - L| < \varepsilon$. Similarly $m \geq N$ implies $|x_m - L| < \varepsilon$. Therefore

$$|x_n - x_m| = |x_n - L + L - x_m| \leq |x_n - L| + |L - x_m| < 2\varepsilon. \quad (12)$$

Here we have used the triangle inequality. It follows that our sequence is Cauchy. Replace ε by $\frac{\varepsilon}{2}$ if you feel paranoid about the 2ε in formula (12).

Fact 2.

Suppose that $\{x_n\}$ is a Cauchy sequence of real numbers. Given $\varepsilon = 1$, we know by definition that there is an integer N_1 so that $n \geq N_1$ implies $|x_n - x_{N_1}| < 1$. Use the triangle inequality to see that $n \geq N_1$ implies

$$|x_n| = |x_n - x_{N_1} + x_{N_1}| \leq |x_n - x_{N_1}| + |x_{N_1}| < 1 + |x_{N_1}|.$$

This gives a bound on the sequence elements x_n such that $n \geq N_1$. That means we have a bound on all but a finite number of sequence elements. It is then not hard to get a bound on the entire sequence. Such a bound is

$$\max\{|x_1|, \dots, |x_{N_1-1}|, 1 + |x_{N_1}|\}.$$

Fact 3.

We need to show that any bounded sequence of real numbers has a convergent subsequence. We give the proof of Bryant in *Yet Another Introduction to Analysis*. There is also a proof in Lang, *Undergraduate Analysis* as a Corollary to the Bolzano-Weierstrass Theorem (p. 38).

Step 1. Any sequence has a subsequence which is either increasing or decreasing.

To prove this, we use the **Spanish or (La Jolla) Hotel Argument**. Consider a sequence of hotels placed on the real axis such that the n th hotel has height x_n , $n \in \mathbb{Z}^+$. See Figure 26.

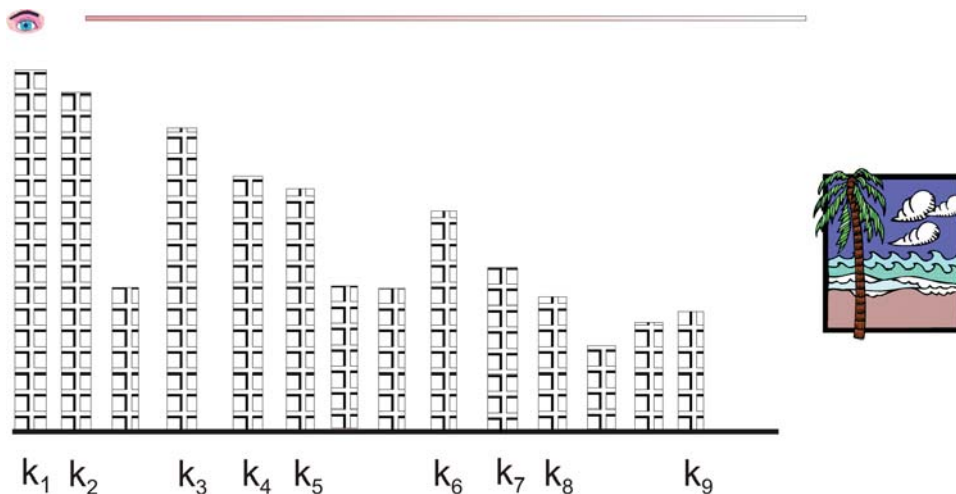


Figure 26: Picture of case A in the Spanish Hotel Argument. The indices $\{k_j\}$ correspond to hotels of strictly decreasing height so that an eyeball at the top of the hotels with label k_n can see the ocean and palm tree at infinity for every $n \in \mathbb{Z}^+$.

Note that if the sequence $\{x_n\}$ is not bounded above and below, you can easily find a subsequence that is either increasing or decreasing.

There are two possibilities.

Case A). There is an infinite sequence of hotels with unblocked views to the right in the direction of the sea at infinity. See Figure 26. This means that there is an infinite sequence of positive integers $k_1 < k_2 < k_3 < \dots < k_n < \dots$ such that from the top of the corresponding hotel, a person has an unblocked view to the right in the direction of the sea. If this is the case, then

$$x_{k_1} > x_{k_2} > x_{k_3} > \dots > x_{k_n} > x_{k_{n+1}} > \dots$$

Thus we have an infinite strictly decreasing sequence of hotel heights.

If Case A) is false, we must be in Case B).

Case B). In this case, after a finite number of hotels, every hotel has a blocked view. Let the finite number of hotels be indexed by k_j , $j = 1, \dots, N$. Then the next integer after that is $N + 1 = m_1$ with the property that there is a positive integer $m_2 > m_1$ such that $x_{m_2} \geq x_{m_1}$. This means hotel m_2 blocks the view of hotel m_1 . Continue to obtain a subsequence which is increasing:

$$x_{m_1} \leq x_{m_2} \leq x_{m_3} \leq \dots \leq x_{m_n} \leq x_{m_{n+1}} \leq \dots$$

See Figure 27.

Step 2. Convergence of Subsequence from Step 1.

To see the convergence, just recall that we are assuming in Fact 3 that our original sequence and thus any subsequence is bounded. So we just need to apply Fact 6 about limits. Any increasing or decreasing bounded sequence must converge.

Fact 4.

Let $\{x_n\}$ be a Cauchy sequence with a convergent subsequence $\{x_{n_k}\}$ such that $\lim_{k \rightarrow \infty} x_{n_k} = L$. So given $\varepsilon > 0$, there is a positive integer K such that $k \geq K$ implies $|x_{n_k} - L| < \varepsilon$.

We want to show that our original sequence x_n converges to L . For any $k \in \mathbb{Z}^+$ we have $n_k \geq k$. Since our sequence is Cauchy, we know there exists a positive integer N so that for $k \geq N$ we have $|x_k - x_{n_k}| < \varepsilon$. Therefore if $k \geq \max\{K, N\}$, we have

$$|x_k - L| \leq |x_k - x_{n_k}| + |x_{n_k} - L| < 2\varepsilon.$$

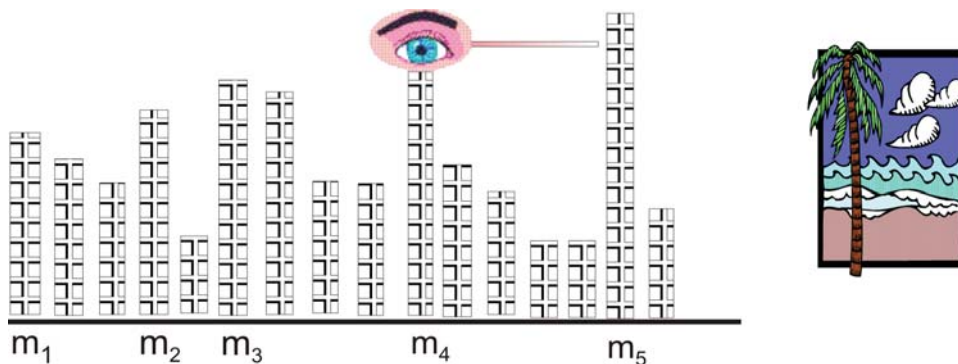


Figure 27: Case B of the Spanish Hotel Argument. The sequence of indices $\{m_k\}$ corresponds to hotels of increasing height.

Here we used the triangle inequality. It follows that the original sequence converges to L . Again you may want to replace ε by $\frac{\varepsilon}{2}$. ■

Proof of Theorem 24

Proof. We want to show that every Cauchy sequence of real numbers converges to a real number.

Fact 2 about Cauchy sequences says a Cauchy sequence is bounded.

Fact 3 about Cauchy sequences says a bounded sequence has a convergent subsequence.

Fact 4 about Cauchy sequences says that once you have a convergent subsequence the original sequence is forced to converge to the same limit as the subsequence.

So we are done. ■

One can base calculus on the concept of an infinitesimal rather than on the idea of a limit. This is called "non-standard analysis." It was created by A. Robinson. R. Rucker, *Infinity and the Mind*, p. 93, says that it is simpler to believe in the infinitely small "dx" rather than to let Δx approach 0. But, Rucker says: "So great is the average person's fear of infinity that to this day calculus all over the world is being taught as a study of **limit processes** instead of what it really is **infinitesimal analysis**." There are a few calculus texts based on non-standard analysis: H.J. Keisler, *Elementary Calculus* and Henle and Kleinberg, *Infinitesimal Calculus*. I will not pursue this subject at all in these notes. It seems harder to deal with than limits since so few people have actually tried to understand it.

Others argue that the universe is finite. See Greenspan, *Discrete Models*, where it is said that "It is unfortunate that so many scientists have been conditioned to believe that 10^{30} particles can **always** be well approximated by an **infinite** number of points." Classical applied math. views the vibrating string as a continuum like \mathbb{R} . Greenspan argues that we should perhaps replace the continuum with a large finite set of points. This replaces calculus with finite difference calculus or the finite element method. We will have nothing to say about that here, except to note that in the end usually one needs a computer to obtain an approximate solution to our applied math. problems and that leads us to replace derivatives with finite differences, for example.

Rucker, *Infinity and the Mind*, considers this question also. On p. 33, for example, he says: "The question of whether or not matter is infinitely divisible may never be decided. For whenever an allegedly minimal particle is exhibited, there will be those who claim that if a high enough energy were available, the particle could be decomposed; and whenever someone wishes to claim that matter is infinitely divisible, there will be some smallest known particle, which cannot be split."

There is also a very basic controversy in mathematics - that of constructivism. One aspect of the constructivist approach is to seek so-called "constructive proofs" which involve a new meaning for the mathematical word "or." Sets and numbers must be constructed. Constructive mathematicians seek a different approach to the completeness axiom and thus to the existence of limits. References are E. Bishop, *Foundations of Constructive Analysis* (where this course and graduate courses are done in a constructive manner) and Volume 39 of the Journal, *Contemporary Mathematics*, which was dedicated to Errett Bishop, including an article by Bishop titled "Schizophrenia in Contemporary Mathematics." I personally find that this constructive approach to the basics of the logic of our proofs does in fact lead my brain to so many twists and turns that schizophrenia might be a good description. I will say no more about it here.

Part IV

Limits of Functions

Now we want to consider the limit of a function $f(x)$ as x approaches a , denoted $\lim_{x \rightarrow a} f(x)$.

Definition 26 Suppose I is an open interval containing the point a and $f : I - \{a\} \rightarrow \mathbb{R}$. Then we say L is the **limit** of $f(x)$ (or $f(x)$ **converges** to L) as x approaches a and write $\lim_{x \rightarrow a} f(x) = L$ iff

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < |x - a| < \delta \implies |f(x) - L| < \varepsilon. \quad (13)$$

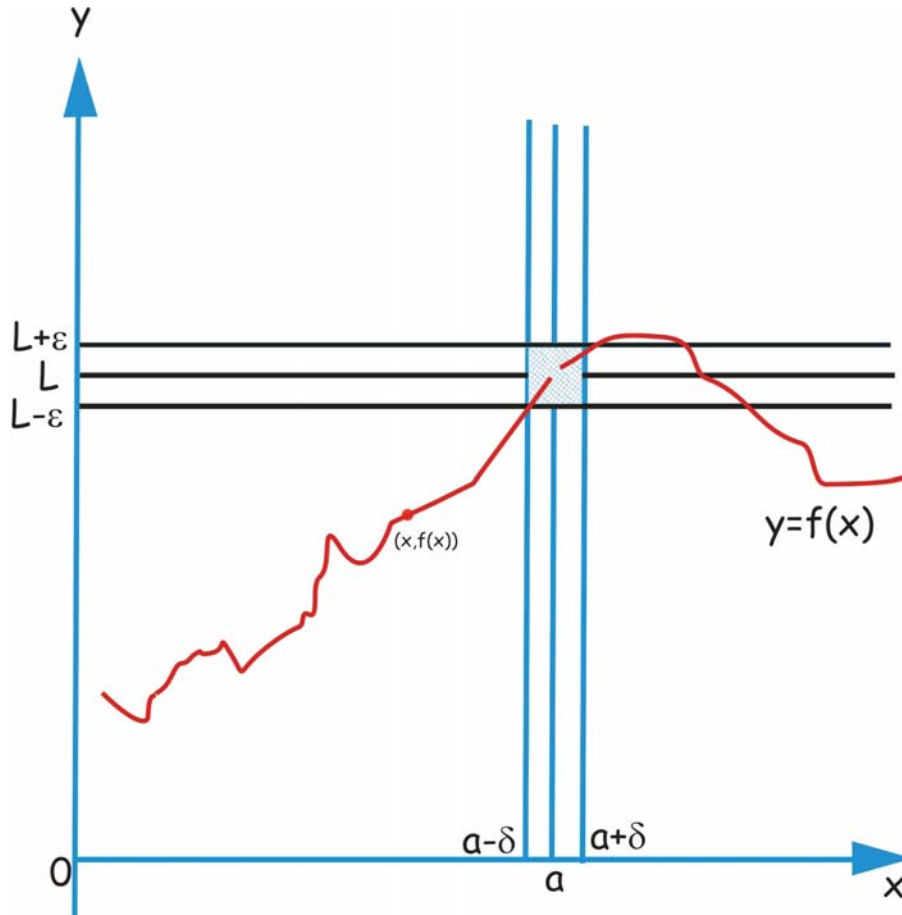


Figure 28: The graph of a function $y = f(x)$ is red. The definition of $\lim_{x \rightarrow a} f(x) = L$ says that given a positive ε we can find a positive δ (depending on ε) so that for $x \neq a$ in the interval $(a - \delta, a + \delta)$, the graph of the function must lie in the blue box of height 2ε and width 2δ , except perhaps for $(a, f(a))$. In the picture $f(a)$ is undefined and thus there is a hole in the graph at (a, L) .

Again you need to memorize this definition. And, yes, it is a pretty horrific sentence full of quantifiers $\exists \forall$. If you believe in the usual logic of mathematics, then you should be willing to write down the negation of this statement. If you are a constructivist, I do not know what you would do.

See Figure 28 for a picture of the definition of limit. Note that we do not assume that $f(a)$ is defined. Thus $|x - a| = 0$ is excluded from consideration in the statement in formula (13) of the definition of limit. It is assumed in the definition that δ is small enough that $0 < |x - a| < \delta$ implies $x \in I - \{a\}$ and thus $f(x)$ makes sense. If $f(a)$ is defined that is O.K. too. It is not required that $f(a) = L$ however. If $f(a) \neq L$, the point $(a, f(a))$ would be outside the little blue box for small enough ε in Figure 28.

Note. We could weaken the hypotheses in the definition of limit. Most authors assume that a is an **accumulation point** of the set S where f is defined. This means that for every $\delta > 0$, there is a point $x \neq a$ such that $x \in S \cap (a - \delta, a + \delta)$. This insures that there are points $x \neq a$ such that $0 < |x - a| < \delta$ and $f(x)$ makes sense. See Apostol, *Mathematical Analysis* or Sagan, *Advanced Calculus*. Lang, *Undergraduate Analysis*, does not do this, nor does he assume that he is only taking points $x \neq a$ in $(a - \delta, a + \delta)$. This allows $f(a)$ to have a bad definition (in which case you would not have a limit) or a to be an **isolated point** (i.e., any non-accumulation point) of the domain of f (in which case you would have a limit trivially). We give the more general definition of limit (with accumulation points) in Lectures, II. For now, our definition suffices.

Example 1. Suppose $f(x) = 3x - 1$, $x \in \mathbb{R}$. Then $\lim_{x \rightarrow 2} (3x - 1) = 5$. Of course, you do not need the definition to compute this limit since $f(x)$ is a continuous function and thus our limit is $f(2)$. But we have not proved anything about continuous functions yet, nor even defined them. So we prove that our limit is correct.

Proof. $|3x - 1 - 5| = |3x - 6| = 3|x - 2| < \varepsilon$ if $|x - 2| < \frac{\varepsilon}{3} = \delta$. ■

For an example of $\lim_{x \rightarrow a} f(x)$ where $f(a)$ is not defined, look at what we will later call a derivative.

Example 2. $\lim_{x \rightarrow a} \frac{x^2 - 1}{x - 1} = 2$.

To prove this, just note that as long as $x \neq 1$, we have the equality: $\frac{x^2 - 1}{x - 1} = x + 1$. Thus we can proceed as in Example 1 to obtain the proof. We leave this to the reader. Note that the graph of the function $\frac{x^2 - 1}{x - 1}$ is a straight line with a hole at the point $(1, 2)$.

If you want, you can also consider **right and left hand limits**. You just replace the open interval I in the definition of limit above with a half open interval $(a - \delta, a)$ or $(a, a + \delta)$, for small positive δ . For example consider the function known as **floor of x** $= [x] = L =$ the greatest integer $\leq x$. See Figure 29. Then we want to say

$$\lim_{\substack{x \rightarrow 1 \\ x < 1}} [x] = 0 \quad \text{and} \quad \lim_{\substack{x \rightarrow 1 \\ x > 1}} [x] = 1.$$

One also writes $\lim_{\substack{x \rightarrow 1 \\ x < 1}} [x] = \lim_{x \rightarrow 1^-} [x]$ and $\lim_{\substack{x \rightarrow 1 \\ x > 1}} [x] = \lim_{x \rightarrow 1^+} [x]$.

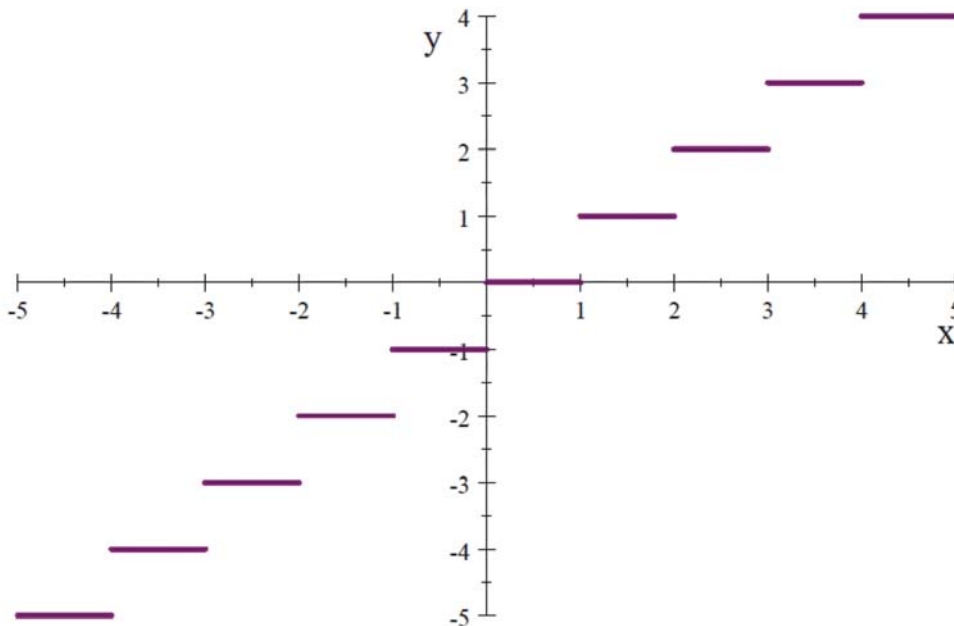


Figure 29: graph of the floor of $x = [x]$

Before looking at more examples, it will help to know the basics about limits.

Properties of Limits.

In the following, we always assume our functions are defined on an open interval I containing the point a , except perhaps at $x = a$.

Property 1) Sequential Definition of Limits. $\lim_{x \rightarrow a} f(x) = L$ iff for every sequence $\{x_n\}$ of points in $I - \{a\}$ such that $\lim_{n \rightarrow \infty} x_n = a$, we have $\lim_{n \rightarrow \infty} f(x_n) = L$.

Moral. If you hate δ for some reason, you can replace δ 's with sequences and thus N 's.

Property 2) Limits are unique. That is, $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} f(x) = K$ implies $K = L$.

Property 3) Limit of a sum is the sum of the limits. Suppose $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$. Then $\lim_{x \rightarrow a} (f(x) + g(x)) = L + M$.

Property 4) Limit of a product is the product of the limits. Suppose $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$. Then $\lim_{x \rightarrow a} (f(x)g(x)) = LM$.

Property 5) Limit of a quotient is the quotient of the limits. Suppose $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$ and, in addition, $M \neq 0$. Then $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{L}{M}$.

Property 6) Limits preserve inequalities. Suppose $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$ and, $f(x) \leq g(x)$, for all x in an open interval containing a , except perhaps when $x = a$. Then $L \leq M$.

Proof. Property 1). We leave this proof as an **exercise**.

Property 2). We leave this as an **exercise**. Imitate the analogous proof for limits of sequences.

Property 3). This proof is similar to that of the analogous fact for limits of sequences. We leave it as an **exercise**.

Property 4). We proceed as in the proof of the analogous fact for sequences. Note that

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

We need to bound $|f(x)|$ for x close to a in order to be able to make the first term in the last sum small. To do this use the definition of limit with $\varepsilon = 1$. This says there is a $\delta_1 > 0$ such that $0 < |x - a| < \delta_1$ implies $|f(x) - L| < 1$. Since $|f(x)| = |f(x) - L + L| \leq |f(x) - L| + |L|$, it follows that $0 < |x - a| < \delta_1$ implies $|f(x)| < 1 + |L|$. Thus $0 < |x - a| < \delta_1$ implies

$$|f(x)g(x) - LM| \leq (1 + |L|) |g(x) - M| + |f(x) - L| |M|. \quad (14)$$

We know that $\exists \delta_2 > 0$ s.t.

$$0 < |x - a| < \delta_2 \text{ implies } |g(x) - M| < \frac{\varepsilon}{2(1 + |L|)}. \quad (15)$$

And $\exists \delta_3 > 0$ s.t.

$$0 < |x - a| < \delta_3 \text{ implies } |f(x) - L| < \frac{\varepsilon}{2(1 + |M|)}. \quad (16)$$

Define $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Then $0 < |x - a| < \delta$ implies (combining inequalities (14), (15) and (16))

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

A shorter proof can be obtained using Property 1 and the corresponding fact for limits of sequences.

Property 5. By Property 4, it suffices to prove the special case that

$$\lim_{x \rightarrow a} g(x) = M, \text{ with } M \neq 0 \text{ implies } \lim_{x \rightarrow a} \frac{1}{g(x)} = \frac{1}{M}. \quad (17)$$

Since dividing by 0 is a "no no," we need to show that for x close enough to a , $g(x) \neq 0$. Since $M \neq 0$, we can take $\varepsilon = \frac{|M|}{2}$. So $\exists \delta_1 > 0$ s.t. $0 < |x - a| < \delta_1$ implies

$$||g(x)| - |M|| \leq |g(x) - M| < \frac{|M|}{2}.$$

(The first inequality here follows from the triangle inequality; i.e., $||A| - |B|| \leq |A - B|$. The proof of this is an **exercise**.)
 Therefore $-\frac{|M|}{2} < |g(x)| - |M| < \frac{|M|}{2}$ which implies, upon adding $|M|$, that

$$|g(x)| > \frac{|M|}{2} > 0, \text{ if } 0 < |x - a| < \delta_1. \tag{18}$$

Now we prove (17). To do this, we need to prove the following is $< \varepsilon$ when x is close enough to a :

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

From formula (18), we know $0 < |x - a| < \delta_1$ implies

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

This can be made $< \varepsilon$ since $\exists \delta < \delta_1$ s.t. $|M - g(x)| < \frac{\varepsilon}{2} |M|^2$ if $0 < |x - a| < \delta$. This completes the proof of (17).

Property 6. Look at $h(x) = g(x) - f(x)$. Then $h(x) \geq 0$, for all x in an open interval containing a , except perhaps when $x = a$. We know from earlier properties that $\lim_{x \rightarrow a} h(x) = M - L \doteq K$. Thus it suffices to show $K \geq 0$. Assume $K < 0$ and deduce a contradiction. Note first that for x in our open interval with $x \neq a$ we have $|h(x) - K| \geq |K| = -K > 0$. See Figure 26 below. But we know $\forall \varepsilon > 0 \exists \delta > 0$ s.t. $|h(x) - K| < \varepsilon$ for $0 < |x - a| < \delta$. This implies $|K| = -K \leq h(x) - K \leq |h(x) - K| < \varepsilon \forall \varepsilon > 0$. Lemma 21 says $|K| = 0$. This contradicts our hypothesis that $K < 0$ and we're done.

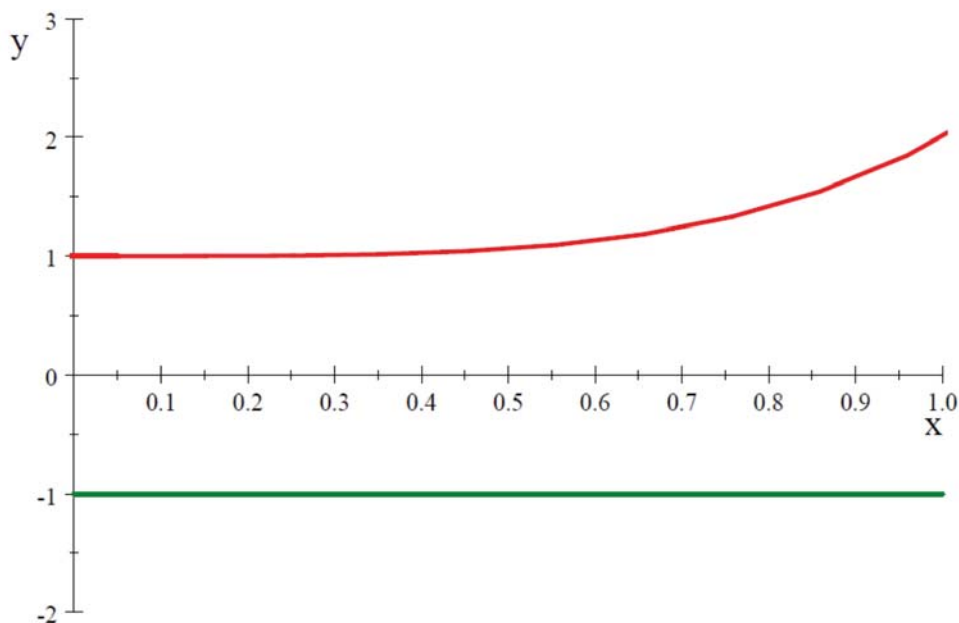


Figure 30: We plot $f(x) = x^4 + 1 \geq 0$ in red; $g(x) = -1$ in green. It should be clear that the distance between $f(x)$ and $g(x)$; i.e., $|f(x) - (-1)| \geq 1$ for all x and thus $f(x)$ cannot approach a negative number like -1 as $x \rightarrow a$.

■

Example 1. $\lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} = 2x.$

Here we are computing the derivative of the function $f(x) = x^2$. This is the sort of limit everyone (e.g., Newton and Leibniz) could do without knowing the precise definition of limit. The computation goes as follows:

$$\frac{(x+h)^2 - x^2}{h} = \frac{x^2 + 2xh + h^2 - x^2}{h} = 2x + h.$$

Taking limits as $h \rightarrow 0$ and using the properties of limits given above, we see that

$$\lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \rightarrow 0} (2x + h) = \lim_{h \rightarrow 0} 2x + \lim_{h \rightarrow 0} h = 2x.$$

Here we have used the fact that the limit of a sum is the sum of the limits, the limit of a constant function is the constant, $\lim_{h \rightarrow 0} h = 0$, and the limit of a product is the product of the limits. Note that x is a constant during our calculation.

Example 2.

Define $f(x) = 1$ if x is rational and $f(x) = 0$ if x is irrational. Then $\lim_{x \rightarrow a} f(x)$ does not exist for any real number a . To see this, you just have to note that any interval contains both rationals and irrationals. See the proof below. If $\lim_{x \rightarrow a} f(x) = L$, we know $\exists \delta > 0$ s.t. $0 < |x - a| < \delta$ implies $|f(x) - L| < \frac{1}{2}$. But there are points x such that $0 < |x - a| < \delta$ with $f(x) = 1$ and other points u such that $0 < |u - a| < \delta$ with $f(u) = 0$. Then $1 = |f(x) - f(u)| \leq |f(x) - L| + |L - f(u)| < 1$. But $1 < 1$ is impossible.

Theorem 27 Any open interval I contains both rational and irrational numbers.

Proof. Rationals in I.

It suffices to show that there is a rational number arbitrarily close to any real number. Given $\varepsilon > 0$ (our measure of closeness), we know by Lemma 22 there is a positive integer n such that $n > \frac{1}{\varepsilon}$. Therefore, it suffices to show that for any real number a there is a rational number q such that $|a - q| \leq \frac{1}{n}$.

Case 1. a is positive.

If $a > 0$, then, for n as in the preceding paragraph, look at the set $S = \{k \in \mathbb{Z}^+ | na < k\}$. This set has a least element m by the well ordering axiom for the positive integers. This means $(m - 1) \leq na$. Therefore

$$\frac{m}{n} - \frac{1}{n} \leq a < \frac{m}{n},$$

which says $|a - \frac{m}{n}| \leq \frac{1}{n} < \varepsilon$. Thus we have found a rational number within ε distance of a .

Case 2. a is negative.

In this case $-a$ is positive and we can use Case 1 to find a rational number q so that $|-a - q| < \varepsilon$. It follows that $|a - (-q)| < \varepsilon$. Of course q rational implies $-q$ is rational.

Case 3. $a=0$.

In this case, life is even easier as we have $|0 - \frac{1}{n}| = \frac{1}{n} < \varepsilon$.

Irrationals in I.

We need to show that there is an irrational number arbitrarily close to any given rational number q . We know from the preceding that we can choose a rational number r such that $|r - q\sqrt{2}| < \varepsilon$. This implies $|\frac{r}{\sqrt{2}} - q| < \frac{\varepsilon}{\sqrt{2}} < \varepsilon$. Since $\frac{r}{\sqrt{2}}$ is irrational (as otherwise $\frac{r}{\sqrt{2}} = u$ is rational and then so is $\sqrt{2} = \frac{r}{u}$ contradicting Theorem 18), we are done. ■

Before considering another example, we need a Lemma.

Lemma 28 Squeeze Lemma. Suppose $f(x) \leq g(x) \leq h(x)$ for all x in some open interval containing a . If $\lim_{x \rightarrow a} f(x) = L = \lim_{x \rightarrow a} h(x)$, then $\lim_{x \rightarrow a} g(x) = L$ also.

Proof. We know from Property 6 above that if $\lim_{x \rightarrow a} g(x)$ exists, it must be both $\leq L$ and $\geq L$ and, therefore $= L$. But why must the limit exist? Given $\varepsilon > 0$, $\exists \delta_1$ s.t. $0 < |x - a| < \delta_1$ implies $|f(x) - L| < \varepsilon$. Similarly $\exists \delta_2$ s.t. $0 < |x - a| < \delta_2$ implies $|h(x) - L| < \varepsilon$. It follows that if $\delta = \min\{\delta_1, \delta_2\}$, we have

$$0 < |x - a| < \delta \text{ implies } -\varepsilon < f(x) - L \leq g(x) - L \leq h(x) - L < \varepsilon.$$

Thus

$$0 < |x - a| < \delta \text{ implies } |g(x) - L| < \varepsilon.$$

■

Example 3. $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$

First look at Figure 31. That may convince you that the formula is correct.

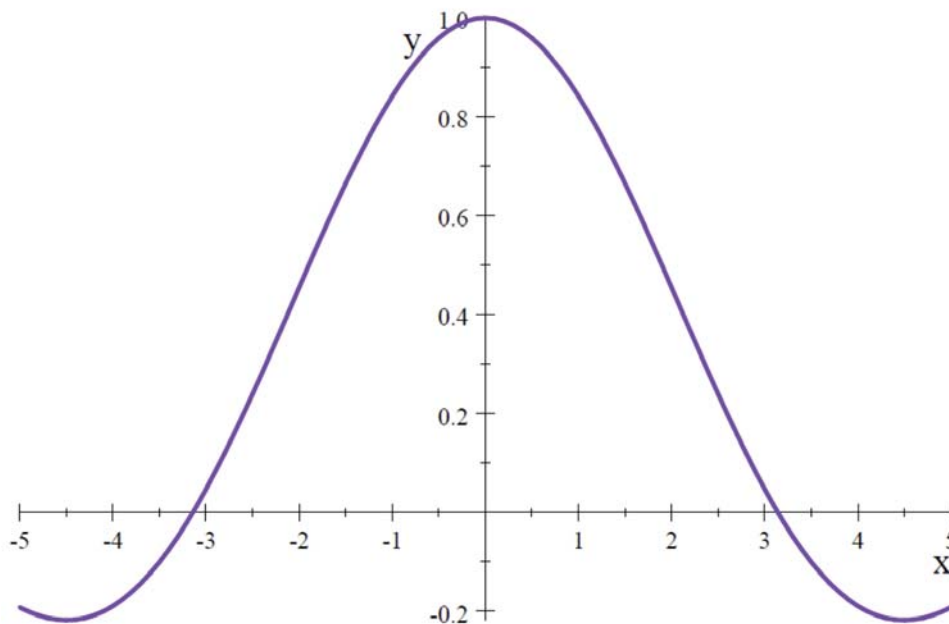


Figure 31: a graph of $\frac{\sin x}{x}$

How are we going to prove this formula? Later we will see that this limit is the derivative of $\sin x$ at $x = 0$ and thus it is $\cos 0 = 1$. Defining the sine and cosine as in your favorite calculus book, and measuring our angles in radians, we have Figure 32 which implies that

$$\frac{\sin x \cos x}{2} \leq \frac{x}{2} \leq \frac{\tan x}{2}.$$

Multiply by $\frac{2}{\sin x}$ to see that

$$\cos x \leq \frac{x}{\sin x} \leq \frac{1}{\cos x}.$$

Thus $\frac{\sin x}{x}$ is squeezed between 2 functions approaching 1 as $x \rightarrow 0$ and by the Squeeze Lemma (which was Lemma 28), it must approach 1 as well. Am I cheating by assuming that $\cos x$ is continuous at $x = 0$? We will say more about trigonometric functions later.

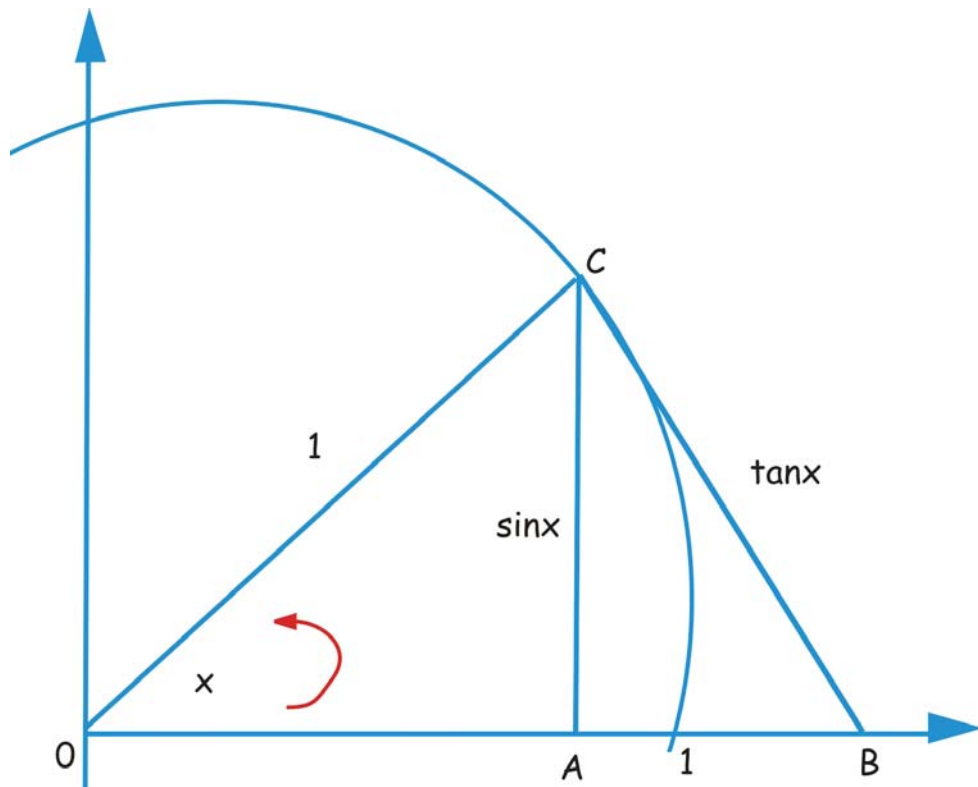


Figure 32: A circle of radius 1 is drawn with an acute angle x measured in radians; i.e., $0 \leq x < \frac{\pi}{2}$. By looking at the areas of the triangle OAC , the arc of the circle with angle x , the triangle OBC , we see that $\frac{\sin x \cos x}{2} \leq \frac{x}{2} \leq \frac{\tan x}{2}$.

16 Limits Involving Infinity

Until now we have been dealing with limits of functions $f(x)$ as x approaches a finite real number a . What happens if we let x approach ∞ ? This definition is similar to the definition of limit of a sequence. We can of course also let x approach $-\infty$.

Definition 29 Suppose $f : (a, \infty) \rightarrow \mathbb{R}$ for some a . Then $\lim_{x \rightarrow \infty} f(x) = L$ means $\forall \varepsilon > 0 \exists B > 0$ s.t. $x \geq B$ implies $|f(x) - L| < \varepsilon$.

Exercise. Show that $\lim_{x \rightarrow \infty} f(x) = L \iff \lim_{\substack{y \rightarrow 0 \\ y > 0}} f\left(\frac{1}{y}\right) = L$.

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

Proof. Given $\varepsilon > 0$, we need to find $B > 0$ s.t. $x \geq B$ implies $\left|\frac{1}{x} - 0\right| = \frac{1}{|x|} < \varepsilon$. Since $x \geq B > 0$, $|x| = x$ and thus we need $\frac{1}{x} < \varepsilon$. This is equivalent to $x > \frac{1}{\varepsilon}$. So we can take $B = 1 + \frac{1}{\varepsilon}$. Then $x \geq B$ implies $x > \frac{1}{\varepsilon}$ and we have the desired inequality. ■

It is possible to prove that these infinite limits have the usual properties of limits. We leave that to the reader to check. You can use the exercise after the definition to do this. That is, make the change of variables $y = 1/x$ and let y go to 0. That will also simplify the following examples.

Example 2. $\lim_{x \rightarrow \infty} \frac{1+x}{1-x} = 1$.

Proof. Note that $\frac{1+x}{1-x} = \frac{1+x}{1-x} \cdot \frac{\frac{1}{x}}{\frac{1}{x}} = \frac{1+\frac{1}{x}}{1-\frac{1}{x}}$. Using the various properties of limits this approaches $\frac{1+0}{1-0} = 1$, as $x \rightarrow \infty$. ■

Example 3. $\lim_{x \rightarrow -\infty} \frac{1}{x} = 0$.

Example 4. $\lim_{x \rightarrow -\infty} \frac{1+x}{1-x} = -1$.

Part V

Continuous Functions

Intuitively continuous functions defined on an open interval in \mathbb{R} are those whose graphs have no breaks or holes. They can have jagged peaks and valleys though. See Figure 33.

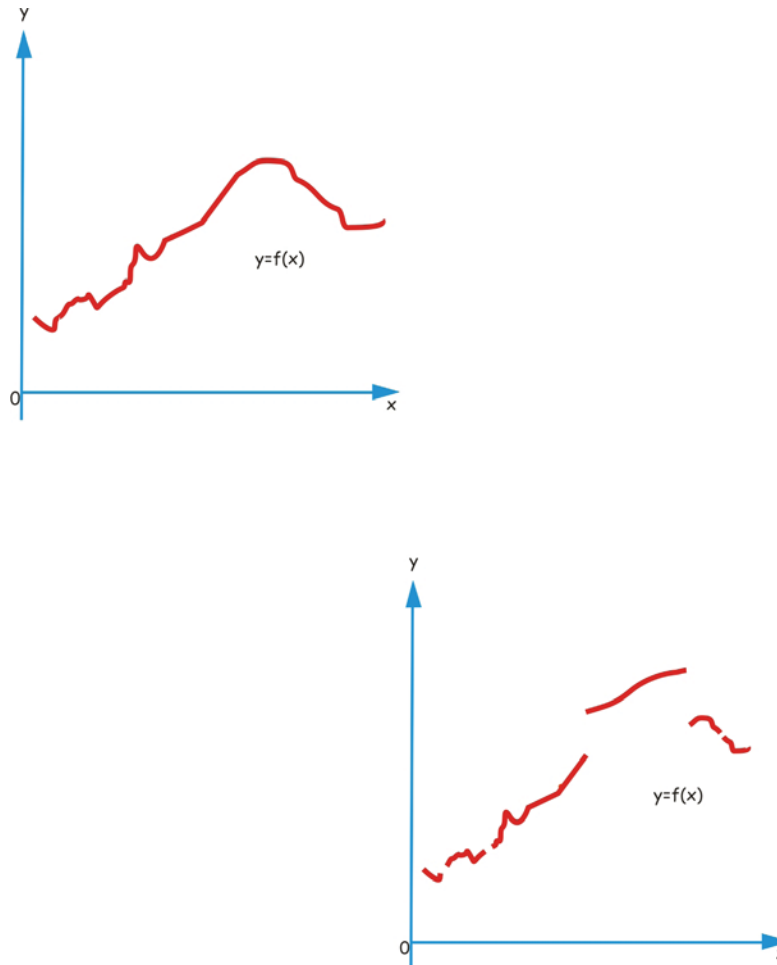


Figure 33: On the upper left is a continuous function on an interval, while on the lower right is a discontinuous function.

Definition 30 If $S \subset \mathbb{R}$, $f : S \rightarrow \mathbb{R}$ is *continuous* at $a \in S$

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \text{s.t.} \quad |x - a| < \delta \quad \text{implies} \quad |f(x) - f(a)| < \varepsilon. \quad (19)$$

We say that f is *continuous on the set* S if f is continuous at every point $a \in S$.

The reader should compare the $\varepsilon\delta$ definition in equation (19) with that in equation (13). The only difference is that here for a continuous function we let $x = a$ while before in the definition of limit as x approaches a we never allow x to equal a .

Note: Assuming that a is an accumulation point of S , f is continuous at a iff $\lim_{x \rightarrow a} f(x) = f(a)$ using the more general definition of limit mentioned above in the note after the definition of limit. If a is an isolated point of S , all functions defined on S are continuous at a . Thus all functions f whose domain of definition is \mathbb{Z} are continuous on \mathbb{Z} .

Example 1. Polynomials $p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, $(a_n, a_{n-1}, \dots, a_1, a_0 \in \mathbb{R})$ are continuous at all points in \mathbb{R} .

Proof. This follows from Fact 1 below as well as the fact that constant functions $f(x) = c$ are easily seen to be continuous everywhere as is the identity function $f(x) = x$ (**exercise**). ■

Example 2. Define $f(x) = \begin{cases} \frac{\sin x}{x}, & x \neq 0 \\ 1, & x = 0. \end{cases}$ This function is also continuous at all points in \mathbb{R} .

Proof. The continuity at non-zero points is easy from Fact 1 below. At $x = 0$, however, one must use the definition of continuity and Example 3 from the section on limits. See Figure 31. ■

Example 3. Define $f(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0. \end{cases}$ This function is also continuous at all points in \mathbb{R} . See Figure 34 below.

The unusual aspect of this function is that there are infinitely many peaks and valleys on the interval $[0, 1]$. Thus you can not really draw the graph in a finite amount of time.

Proof. Again the continuity at non-zero points is easy from Facts 1 and 2 below. But at $x = 0$ one must use the definition of continuity and the fact that $|\sin \theta| \leq 1$ for all angles θ . Thus $|x \sin \frac{1}{x}| \leq |x|$ and we can take $\varepsilon = \delta$ in the definition of continuity at 0. ■

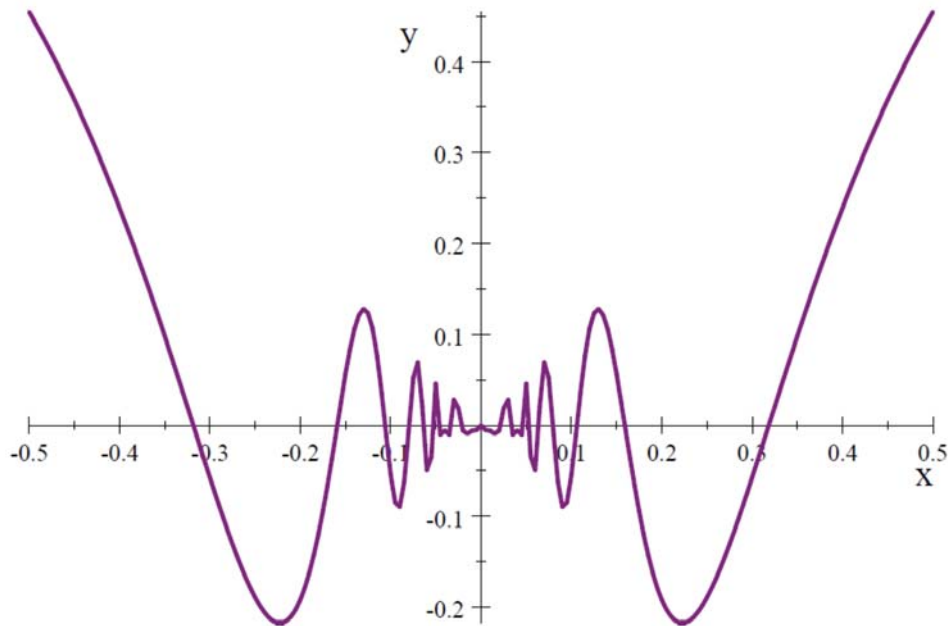


Figure 34: a graph of $x \sin \frac{1}{x}$. There are infinitely many wiggles in this curve near 0 although the function is everywhere continuous.

Facts About Continuous Functions.

Fact 1) Sum, Product, Quotient of Continuous Functions are Continuous. Suppose $f, g : S \rightarrow \mathbb{R}$ are both continuous at $a \in S$. Then the sum $f + g$ and product fg are also continuous at a . If $g(a) \neq 0$, then the quotient $\frac{f}{g}$ is also continuous at a .

Fact 2) Composite of Continuous Functions is Continuous. Suppose $f : S \rightarrow T$ and $g : T \rightarrow \mathbb{R}$ for subsets S, T of \mathbb{R} . If $a \in S$, then $b = f(a) \in T$. If f is continuous at a and g is continuous at $b = f(a)$, then the composite function $g \circ f : S \rightarrow \mathbb{R}$ is continuous at a . Recall that $(g \circ f)(x) = g(f(x))$.

Proof. Fact 1) These facts follow from the corresponding facts about limits.

Fact 2) Given $\varepsilon > 0$ we know there is $\delta > 0$ s.t. $|y - b| < \delta$ implies $|g(y) - g(b)| < \varepsilon$. For this very δ we know there is a $\delta_1 > 0$ s.t. $|x - a| < \delta_1$ implies $|f(x) - f(a)| < \delta$. Putting these 2 sentences together gives (since $b = f(a)$)

$$|x - a| < \delta_1 \text{ implies } |f(x) - f(a)| < \delta \text{ which implies } |g(f(x)) - g(b)| < \varepsilon.$$

■

The most important properties of a continuous function on a closed finite interval $[a, b]$ are the Intermediate Value Theorem and the Weierstrass Theorem on the existence of maxima and minima. We state these 2 theorems first and then prove them.

Theorem 31 Intermediate Value Theorem. Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous. Then for every γ between $f(a)$ and $f(b)$ there exists a point $c \in [a, b]$ such that $f(c) = \gamma$.

Roughly this says that the graph of f has no breaks or holes. It must cross every horizontal line $y = \gamma$ if γ is between $f(a)$ and $f(b)$. See Figure 35.

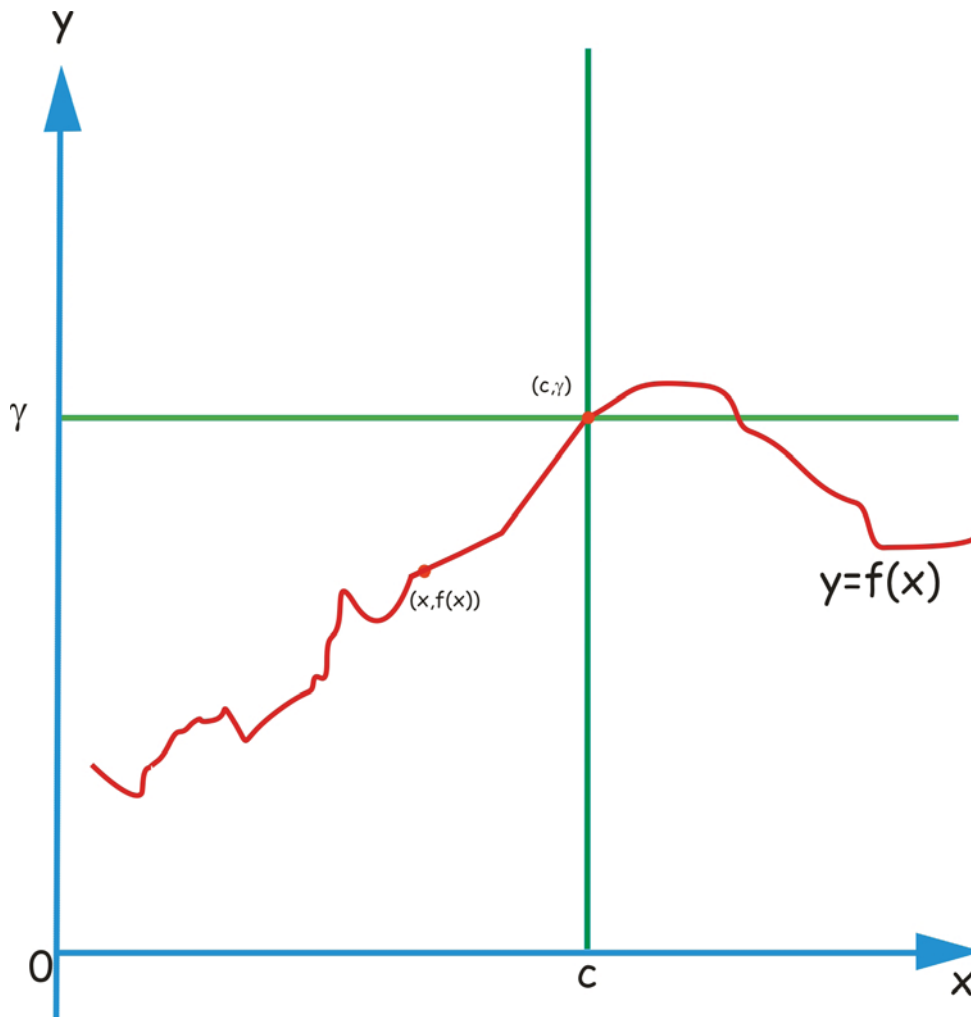


Figure 35: Assume f is continuous on the closed finite interval $[a, b]$. The intermediate value theorem says that for any γ in $f[a, b]$ there is a point where the horizontal line $y = \gamma$ intersects the graph of $y = f(x)$. We label the point (c, γ) on our graph.

Theorem 32 Weierstrass Theorem on the Existence of Maxima and Minima. Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous. Then there exists a point $c \in [a, b]$ such that $f(c)$ is the minimum value of f on $[a, b]$; i.e., $f(c) \leq f(x) \forall x \in [a, b]$. We write $f(c) = \min\{f(x) \mid x \in [a, b]\}$. Similarly there exists a point $d \in [a, b]$ such that $f(d)$ is the maximum value of f on $[a, b]$; i.e., $f(d) \geq f(x) \forall x \in [a, b]$. Write $f(d) = \max\{f(x) \mid x \in [a, b]\}$.

The Weierstrass theorem assures the existence of solutions to max-min problems for continuous functions on closed finite intervals. If you drop either the hypothesis that the interval is closed and finite or the hypothesis that the function is continuous on the interval, then the conclusion may be false. Ultimately it is possible to replace the closed interval with an

arbitrary compact set (see Lang, *Undergraduate Analysis*, for the definition) such as the Cantor dust (see Figure 36) obtained by repeatedly removing middle thirds of intervals in $[0, 1]$. We will see this example again later in the course. The actual Cantor set is invisible. We can only show pictures of approximations to the set. The Cantor set is a fractal. We will say more about fractals later. See Falconer, *Fractal Geometry* for more information on fractals.



Figure 36: 5 approximations to the Cantor dust, each approximation removing middle thirds of the intervals in the approximation above.

You may question how useful it is to know that a max or min exists without knowing how to find it. Sometimes existence does tell you everything. Of course once we have derivatives we will have more information on the location of a max or min.

The following example says BEWARE that you know your function and interval satisfy the Weierstrass hypotheses.
Example. Consider the function $f(x) = \frac{1}{x}$, for $x \in (0, 1]$. This function has no maximum value even though it is continuous on the interval $(0, 1]$. This is not a closed interval of course.

Given a function like $f(x) = x^5 - 3x + 129$, you probably know (at least if you have taken numerical analysis) there are many ways to solve $f(c) = 7$ approximately. Programs like Mathematica, Matlab, Scientific Workplace do these things with amazing speed. Scientific Workplace tells me the 5 solutions are approximately

$$\{2.1044 + 1.5040i, 2.1044 - 1.5040i, -0.78088 - 2.5059i, -0.78088 + 2.5059i, -2.6470\}.$$

Only the last root is real. That is the one we are talking about here. Yeah, this is real analysis not complex. So we ignore 4 out of 5 roots. In Figure 37 we plot the function $f(x) = x^5 - 3x + 129$ and the line $y = 7$ on the interval $[-3, 1]$. The intermediate value theorem tells us that the root c of $f(c) = 7$ exists since $f(-3) = -105$ and $f(1) = 127$.

In what follows we give a numerical analyst's proof of the Intermediate Value Theorem (from V. Bryant, *Yet Another Introduction to Analysis*). It tells you how to approximate a point $c \in [a, b]$ with $f(c) = \gamma$.

Proof of the Intermediate Value Theorem by the Bisection Method. This is a proof by induction.

Step 1. Set $r_1 = a$ and $s_1 = b$. Assume $f(r_1) < f(s_1)$. Then $f(r_1) < \gamma < f(s_1)$. Define $m_1 = \frac{r_1 + s_1}{2}$ = the **midpoint** of the interval $[r_1, s_1]$. To obtain the next subinterval $[r_2, s_2]$, there are 3 cases according to the location of $f(m_1)$. See the Figure 38.

Case a. If $f(m_1) < \gamma$, set $r_2 = m_1$ and $s_2 = s_1$.

Case b. If $f(m_1) > \gamma$, set $r_2 = r_1$ and $s_2 = m_1$.

Case c. $f(m_1) = \gamma$ and we are done as $m_1 = c$.

Induction Step. Assume that $[r_j, s_j]$ have been found for $j = 1, 2, \dots, n$. Define $[r_{n+1}, s_{n+1}]$ as follows. Look at the midpoint $m_n = \frac{r_n + s_n}{2}$. Again consider 3 cases according to the location of $f(m_n)$.

Case a. If $f(m_n) < \gamma$, set $r_{n+1} = m_n$ and $s_{n+1} = s_n$.

Case b. If $f(m_n) > \gamma$, set $r_{n+1} = r_n$ and $s_{n+1} = m_n$.

Case c. $f(m_n) = \gamma$ and we are done as $m_n = c$.

Assuming we are never in Case c (when we would be done after a finite number of steps), we create 2 infinite sequences $\{r_n\}$ and $\{s_n\}$ with the following properties:

$$a = r_1 \leq r_2 \leq \dots \leq r_n \leq r_{n+1} \leq s_{n+1} \leq s_n \leq \dots \leq s_2 \leq s_1 = b$$

and $s_n - r_n = \frac{b-a}{2^{n-1}}$.

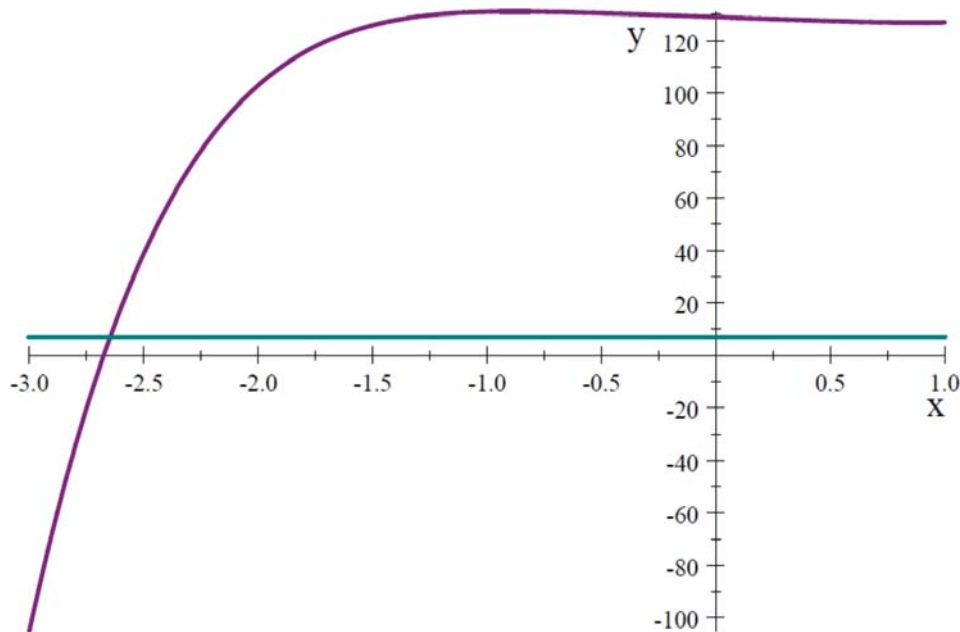


Figure 37: The graph of $f(x) = x^5 - 3x + 129$ and that of the line $y = 7$. The x-coordinate of the intersection of the 2 curves is approximately -2.647 .

Claim. $\lim_{n \rightarrow \infty} r_n = \lim_{n \rightarrow \infty} s_n = c$ and $f(c) = \gamma$.

Proof of Claim. The sequence $\{r_n\}$ is increasing and bounded. Thus it must have a limit which we will call c , using Fact 6 in the section on facts about limits of sequences. Similarly $\{s_n\}$ is decreasing and bounded and must have a limit which we will call c' . We also know that

$$s_n - r_n = \frac{b - a}{2^{n-1}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

It follows that $\lim_{n \rightarrow \infty} (s_n - r_n) = c' - c = 0$.

We know also that

$$f(r_n) \leq \gamma \leq f(s_n).$$

Since f is continuous, we have

$$f(c) = \lim_{n \rightarrow \infty} f(r_n) \leq \gamma \leq \lim_{n \rightarrow \infty} f(s_n) = f(c).$$

That completes the proof of the Intermediate Value Theorem. ■

Next we prove the earlier theorem about existence of maxima and minima. This will require us to remember the facts about limits of sequences proved using the Spanish Hotel argument.

Proof. of the Weierstrass Theorem that a Continuous Function Achieves its Maxima and Minima on a Finite Closed Interval $[a, b]$.

Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous. We want to find $c \in [a, b]$ so that $f(x) \leq f(c)$ for all $x \in [a, b]$. We need 3 steps.

Step 1. Any sequence in the closed interval $[a, b]$ has a subsequence converging to a limit in $[a, b]$.

This is Fact 3 in the list of facts about Cauchy (and other) sequences of real numbers. It was proved using the Spanish Hotel argument.

Step 2. The set $\{f(x) | x \in [a, b]\}$, is bounded above.

This is true since otherwise there is a sequence of points $x_n \in [a, b]$ such that $f(x_n) > n$, for all $n \in \mathbb{Z}^+$. By Step 1, we know that the sequence $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$ such that $\lim_{k \rightarrow \infty} x_{n_k} = c \in [a, b]$. Now we know that $f(x_{n_k}) > n_k \geq k$, for all $k \in \mathbb{Z}^+$. This implies that $\lim_{k \rightarrow \infty} f(x_{n_k}) = \infty$. But this contradicts the continuity of the function f which implies that $\lim_{k \rightarrow \infty} f(x_{n_k}) = f(c)$ which is a finite real number.

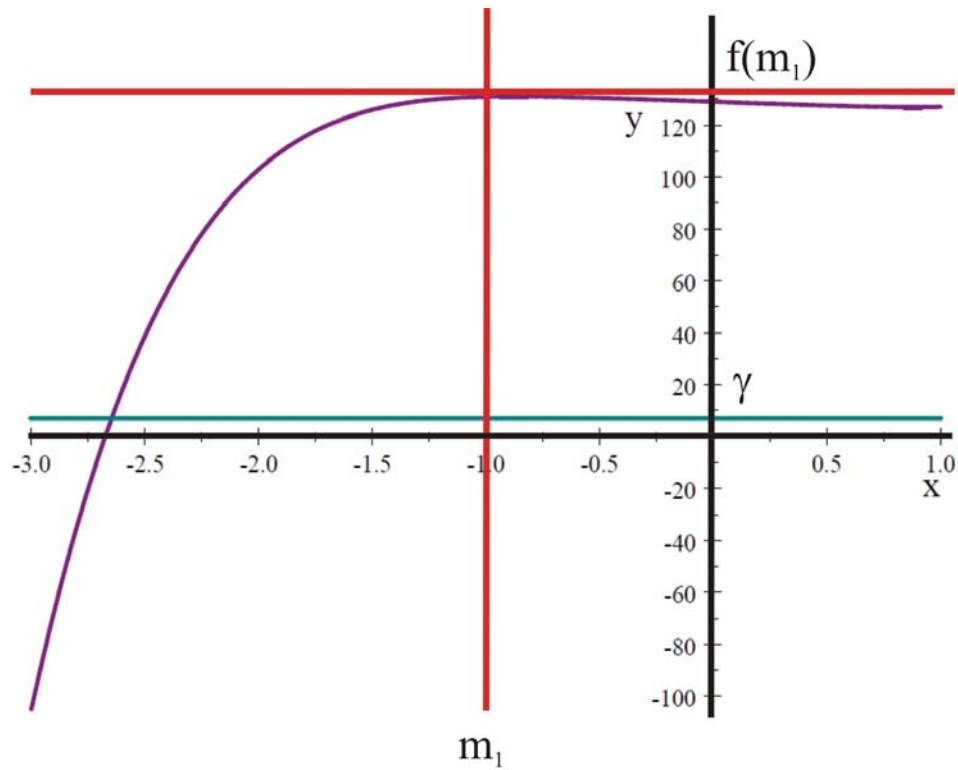


Figure 38: Here we visualize the 1st step in the bisection method with the 1st interval being $[a, b] = [r_1, s_1] = [-3, 1]$. We have $f(x) = x^5 - 3x + 129$, and seek to find c with $f(c) = 7$. This is the polynomial from Figure 37. The midpoint is $m_1 = -1$. Then clearly $f(m_1) > \gamma = 7$. So the next interval is $[r_2, s_2] = [-3, -1]$. Of course you can see from the graph approximately where the curve crosses the line $y = 7$. And we noted above that an approximation to c with $f(c) = 7$ is -2.6470 . The reader should find the next interval $[r_3, s_3]$, or even enough intervals to get our approximation to c .

Step 3. *l.u.b.* $\{f(x)|x \in [a, b]\} \doteq \beta = \max \{f(x)|x \in [a, b]\} = f(c)$ for some $c \in [a, b]$.

By Step 2 and the completeness axiom we know the least upper bound β exists. But why does $\beta = f(c)$ for some $c \in [a, b]$? By the definition of least upper bound, there is a point $u_n \in [a, b]$ such that

$$f(u_n) > \beta - \frac{1}{n}, \text{ for every } n \in \mathbb{Z}^+.$$

Once more by Fact 3 about Cauchy (and other) sequences, we know $\{u_n\}$ has a convergent subsequence $\{u_{n_k}\}$ with $\lim_{k \rightarrow \infty} u_{n_k} = c \in [a, b]$. We want to show that $f(c) = \beta$. To do this, note that

$$\beta \geq f(u_{n_k}) > \beta - \frac{1}{n_k} \geq \beta - \frac{1}{k} \text{ for every } k \in \mathbb{Z}^+.$$

Therefore by the Squeeze Lemma (which was Lemma 28) and the continuity of f ,

$$\beta = \lim_{k \rightarrow \infty} f(u_{n_k}) = f(\lim_{k \rightarrow \infty} u_{n_k}) = f(c).$$

This completes the proof of the Weierstrass Theorem on Existence of Maxima. We leave minima to the reader. It amounts to replacing f by $-f$. You should not have to do the entire proof over again. ■

The following Corollary follows from the 2 preceding theorems.

Corollary 33 *If a function $f : [a, b] \rightarrow \mathbb{R}$ is continuous, then the image $f[a, b] = \{f(x)|x \in [a, b]\}$ is also a closed interval.*

Part VI

Derivatives

17 Examples

Next we finally get to do some calculus (if by calculus, you mean derivatives and integrals only). We start with derivatives.

It would be hard to do applied mathematics without derivatives. Newton invented calculus (along with Leibniz) to formulate Newton's 3 laws of motion such as force = mass \times acceleration (1666). Of course acceleration is a second derivative. See E.T. Bell, *Men of Mathematics* for some of this story. The derivative can be used to represent all sorts of instantaneous rates of change - not just that of distance with respect to time. Thus one finds differential equations in physics, chemistry, economics, biology, ecology, weather modeling.

For example, the **predator-prey equations** describe the evolution of 2 interacting species such as cats and mice on a desert island. If $u(t)$ =number of mice at time t , $v(t)$ =number of cats at time t , the predator-prey equations say:

$$\begin{aligned}u' &= (a - bv - mu)u \\v' &= (cu - d - nv)v,\end{aligned}$$

where a, b, c, d, m, n are constants. References for these equations include H. Kocak, *Differential and Difference Equations through Computer Experiment*; or J. M. Maynard Smith, *Mathematical Ideas in Biology*. See K. Devlin, *Mathematics: The Science of Patterns*, Chapter 3, for more examples.

The difference between the graph of an everywhere differentiable function and a merely everywhere continuous function is quite visible. The more derivatives a function has the smoother the graph.

Example 1. An infinitely differentiable function The Gaussian or normal density.

Consider the function $G(x) = \frac{1}{\sqrt{\pi}}e^{-x^2}$ which is plotted in Figure 39. The n th derivative $G^{(n)}(x)$ exists for every $n = 1, 2, 3, \dots$ and every real point x . This graph is as smooth as they come.

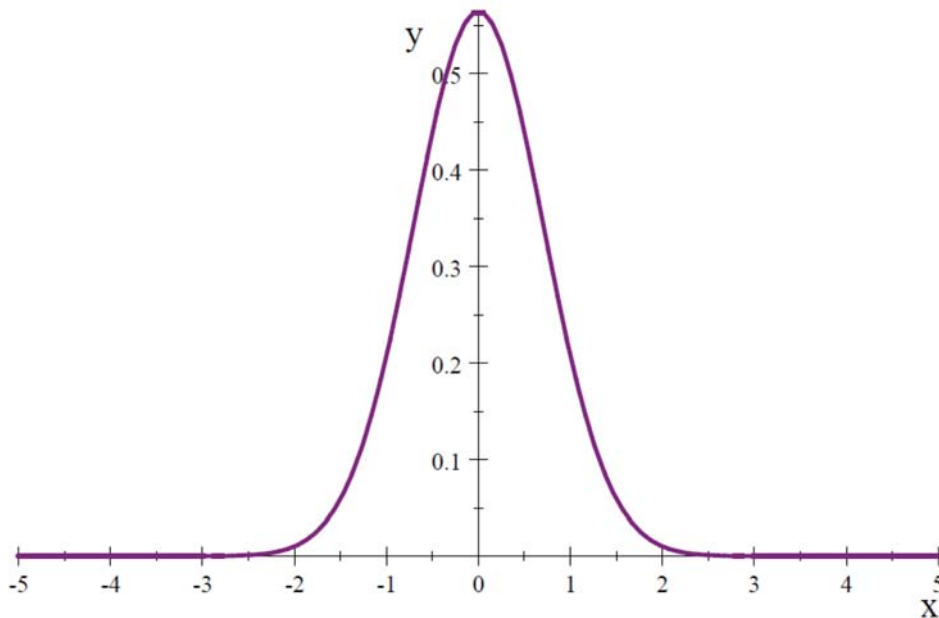


Figure 39: plot of $G(x) = \frac{1}{\sqrt{\pi}}e^{-x^2}$

Example 2. A function which is not differentiable at one point.

Consider the absolute value function $|x|$ whose graph is in Figure 40. This function has derivative 1 for $x > 0$, -1 for $x < 0$, but no derivative at $x = 0$. You can see the problem in the graph. There is a very sharp turn at $x = 0$.

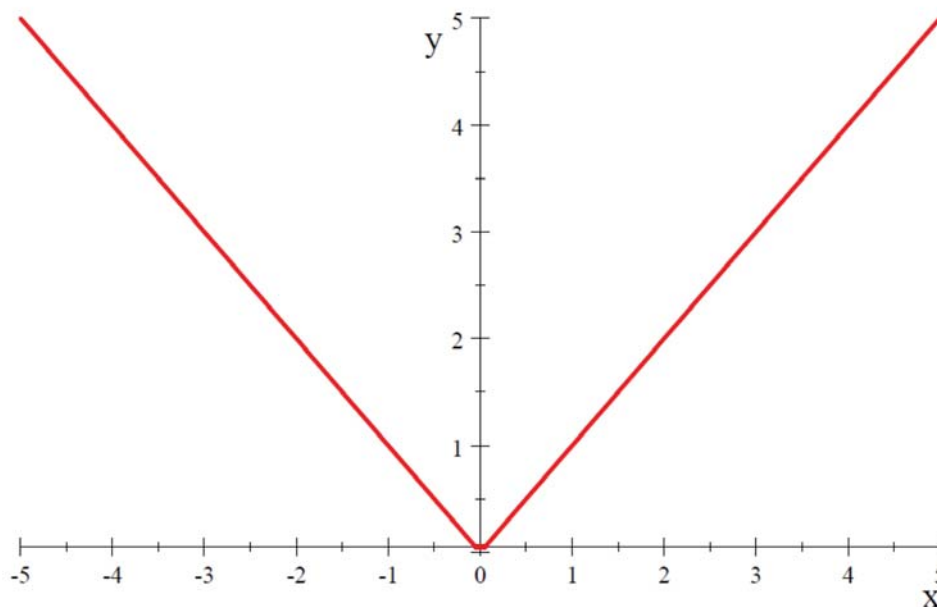


Figure 40: plot of $|x|$

Example 3. Weierstrass Continuous Nowhere Differentiable Function.

Here we examine a function that is as non-smooth as one can imagine. In fact, people did not believe that such a function could exist before Weierstrass. Hermite described these functions as a "dreadful plague." Poincaré wrote: "Yesterday, if a new function was invented it was to serve some practical end; today they are specially invented only to show up the arguments of our fathers, and they will never have any other use." Even as late as the 1960's, before "everyone" had a computer fast enough to graph these things, such examples were viewed as pathological monsters. Now there are thousands of websites with pictures of approximations of them.

Later we will examine the Weierstrass continuous nowhere differentiable function in more detail. If you are impatient, see Falconer, *Fractal Geometry*. The **Weierstrass function** is defined by an infinite series depending on 2 parameters λ and s :

$$f(x) = \sum_{k=0}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k x), \text{ for } \lambda > 1 \text{ and } 1 < s < 2.$$

We will prove later (in Lectures, II) that $f(x)$ is continuous. This is an easy consequence of the Weierstrass M-test for uniform convergence of series of functions. However it can be shown that the derivative $f'(x)$ does not exist at any point x . We will give exercises on this in a final exam.

Figures 41, 42, and 43 show graphs of approximations of such functions on the interval $[0, .5]$ (in which the infinite series is cut off at some finite number of terms). It turns out that the graph of the Weierstrass function is a fractal having box dimension s . We will define box dimension in a final exam. See Falconer, loc. cit., if you feel impatient. A smooth curve would have box dimension 1. The closer to 2 the dimension gets, the more the curve fills the 2D picture. In 41 we take $\lambda = 1.5$, $s = .7$, and cut off the sum at $k = 3000$. In Figure 42 we take $\lambda = 1.5$, $s = .9$, and cut off the sum at $k = 3000$. In Figure 43 we take $\lambda = 5$, $s = .9$, and cut off the sum at $k = 1000$.

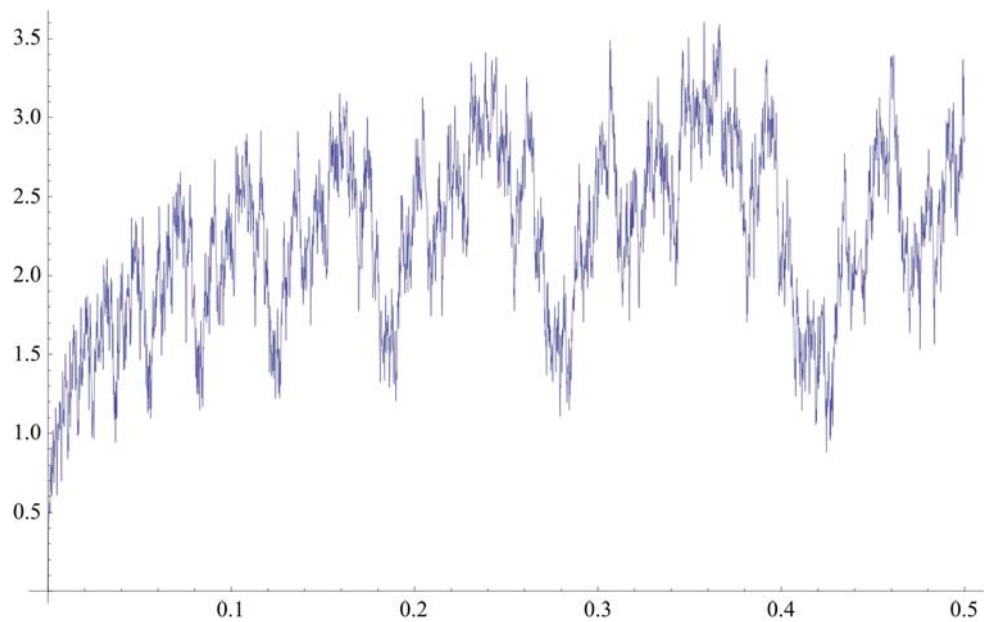


Figure 41: Mathematica plot of the sum $\sum_{k=0}^{3000} 1.5^{(1.7-2)k} \sin(1.5^k x)$.

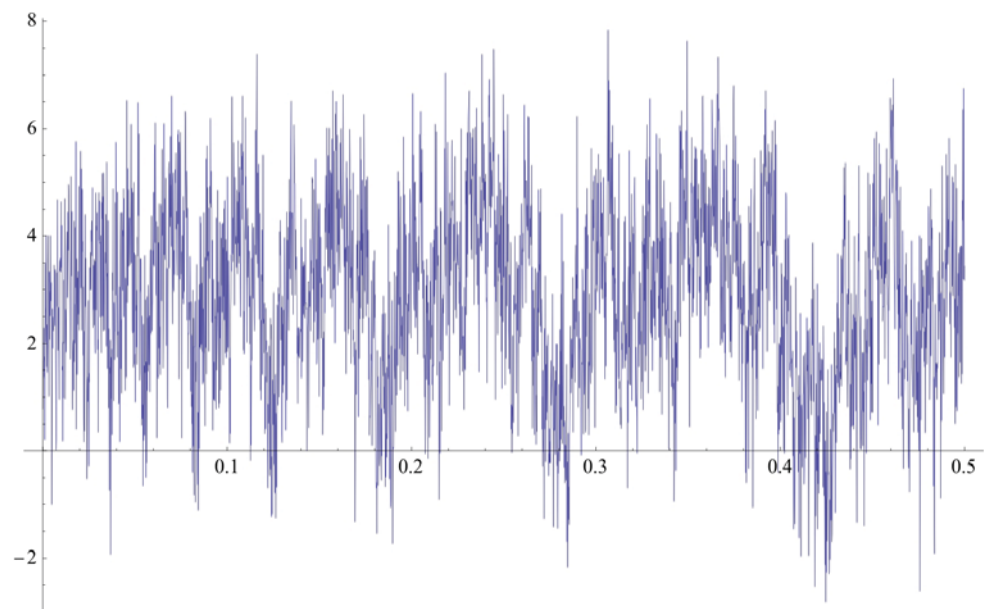


Figure 42: Mathematica plot of the sum $\sum_{k=0}^{3000} 1.5^{(1.9-2)k} \sin(1.5^k x)$.

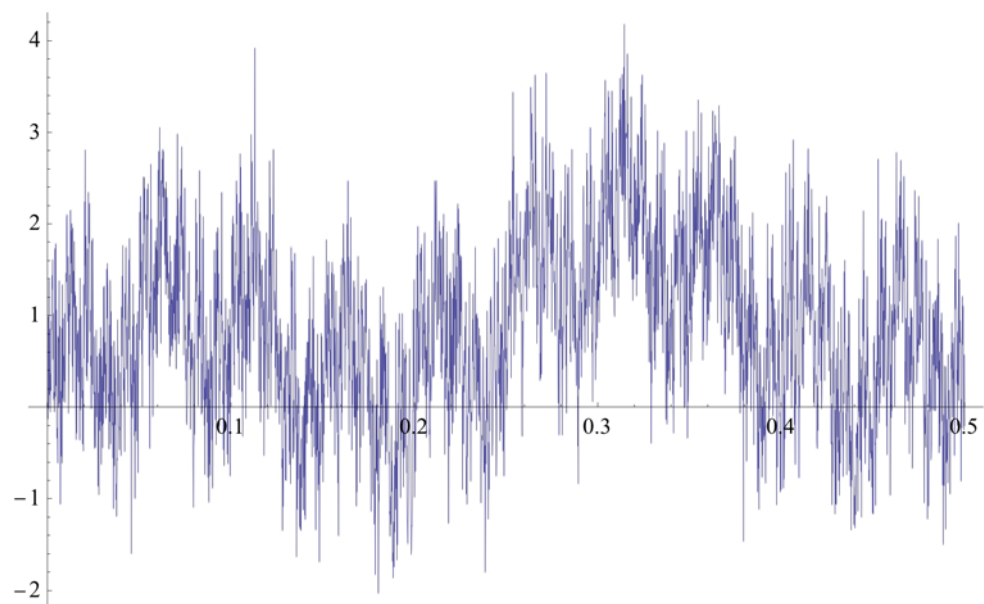


Figure 43: Mathematica plot of the sum $\sum_{k=0}^{1000} 5^{(1.9-2)k} \sin(5^k t)$

18 Definition of Derivative

The derivative $f'(x)$ can be thought of geometrically as the slope of the tangent line to the curve $y = f(x)$ at the point $(x, f(x))$. But what does that really mean? It means that you take limits of slopes of secant lines; i.e., lines through points $(x, f(x))$ and $(x + h, f(x + h))$ for small values of $|h|$. See Figure 44.

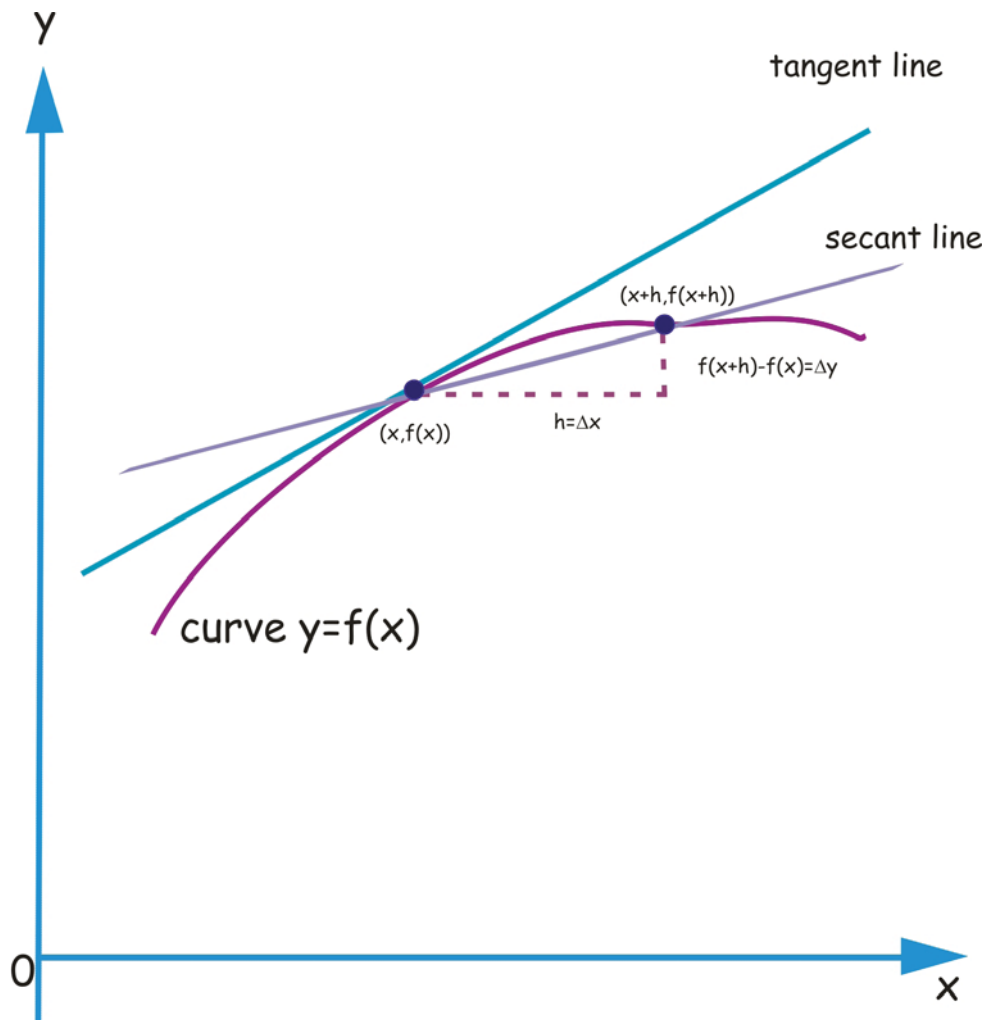


Figure 44: The derivative of $f(x)$ at the point x is the limit of the slopes $\frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}$ of the secant lines as $\Delta x \rightarrow 0$. Here by secant line, we mean the line connecting $(x, f(x))$ and $(x + \Delta x, f(x + \Delta x))$. The derivative $f'(x)$ is the slope of the tangent to the curve $y = f(x)$ at the point $(x, f(x))$.

If instead $t = \text{time}$ and $y = f(t) = \text{distance from origin at time } t$, then $\frac{\Delta y}{\Delta t} = \frac{f(t+\Delta t) - f(t)}{\Delta t} = \text{average velocity over the interval from } t \text{ to } t + \Delta t$. The instantaneous velocity at time t is the limit of $\frac{\Delta y}{\Delta t}$ as Δt approaches 0.

More precisely, suppose the function f is defined on an open interval I containing the point c

Definition 34 The *derivative* $f'(c)$ is defined by

$$f'(c) \doteq \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} \doteq \frac{df}{dx}(c).$$

If this limit exists we say that f is **differentiable** at c . When f is differentiable at every point in a set S , we say that f is **differentiable on S** .

Example 1. Consider $f(x) = x^3 - 2x + 1$. This is differentiable at any point x , as are all polynomials. Let's compute the derivative.

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} &= \lim_{h \rightarrow 0} \frac{((x+h)^3 - 2(x+h) + 1) - (x^3 - 2x + 1)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(x^3 + 3x^2h + 3xh^2 + h^3 - 2(x+h) + 1) - (x^3 - 2x + 1)}{h} \\ &= \lim_{h \rightarrow 0} \frac{3x^2h + 3xh^2 + h^3 - 2h}{h} = 3x^2 - 2. \end{aligned}$$

Example 2. Define the function $g(x) = \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q}. \end{cases}$ The graph looks like 2 horizontal lines even though one line has way fewer points, as the rationals are denumerable while the irrationals are not. We saw earlier that this is not a continuous function at any point x since any interval on the real line contains both rationals and irrationals. See the section on properties of limits. Later we will show that this implies the derivative cannot exist at any point. Thus this function is nowhere differentiable.

It is also possible to define 1-sided derivatives of a function f at a point c . For the right-hand derivative at c , you only need f to be defined on $[c, c + \delta)$ for some $\delta > 0$.

Definition 35 The *right-hand derivative* is defined by: $f'_+(c) = \lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{f(c+h) - f(c)}{h}$.

We leave it to the reader to define the left-hand derivative. The right- and left-hand derivatives must be equal at the point c in order for $f'(c)$, the 2-sided derivative to exist.

Example. **The absolute value.** The function $f(x) = |x|$ has right-hand derivative $f'_+(0) = 1$, while the left-hand derivative $f'_-(0) = -1$. Thus, the absolute value function does not have a 1-sided derivative. To see the right-hand formula, note that

$$f'_+(c) = \lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{|0+h| - |0|}{h} = \lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{|h|}{h} = \lim_{h > 0} \frac{h}{h} = 1.$$

We leave it to the reader to check the left-hand formula.

19 Alternative Definition of the Derivative: The Linear Approximation.

Definition 36 Suppose f is defined on the open interval I . If $c \in I$, then f is differentiable at c if and only if there is a real number L and a function ϕ defined on an open interval containing 0 such that

$$f(c+h) = f(c) + Lh + \phi(h), \quad \text{and} \quad \lim_{h \rightarrow 0} \frac{\phi(h)}{h} = 0. \quad (20)$$

In formula (20), $L = f'(c)$. Note that the first 2 terms on the right ($f(c) + Lh$) are a linear function of h (holding c fixed). We view $\phi(h)$ as a 2nd order term since $\frac{\phi(h)}{h}$ must approach 0 with h . This 2nd definition of derivative is the beginning of a Taylor expansion at c . See Figure 45 for an illustration of formula (20).

Set

$$\psi(h) = \begin{cases} \frac{\phi(h)}{h}, & h \neq 0 \\ 0, & h = 0. \end{cases} \quad (21)$$

Then ψ is continuous at 0 and $\phi(h) = h\psi(h)$ and we see that $\phi(h)$ approaches 0 faster than h as $h \rightarrow 0$. It is also clear that ϕ is continuous at 0.

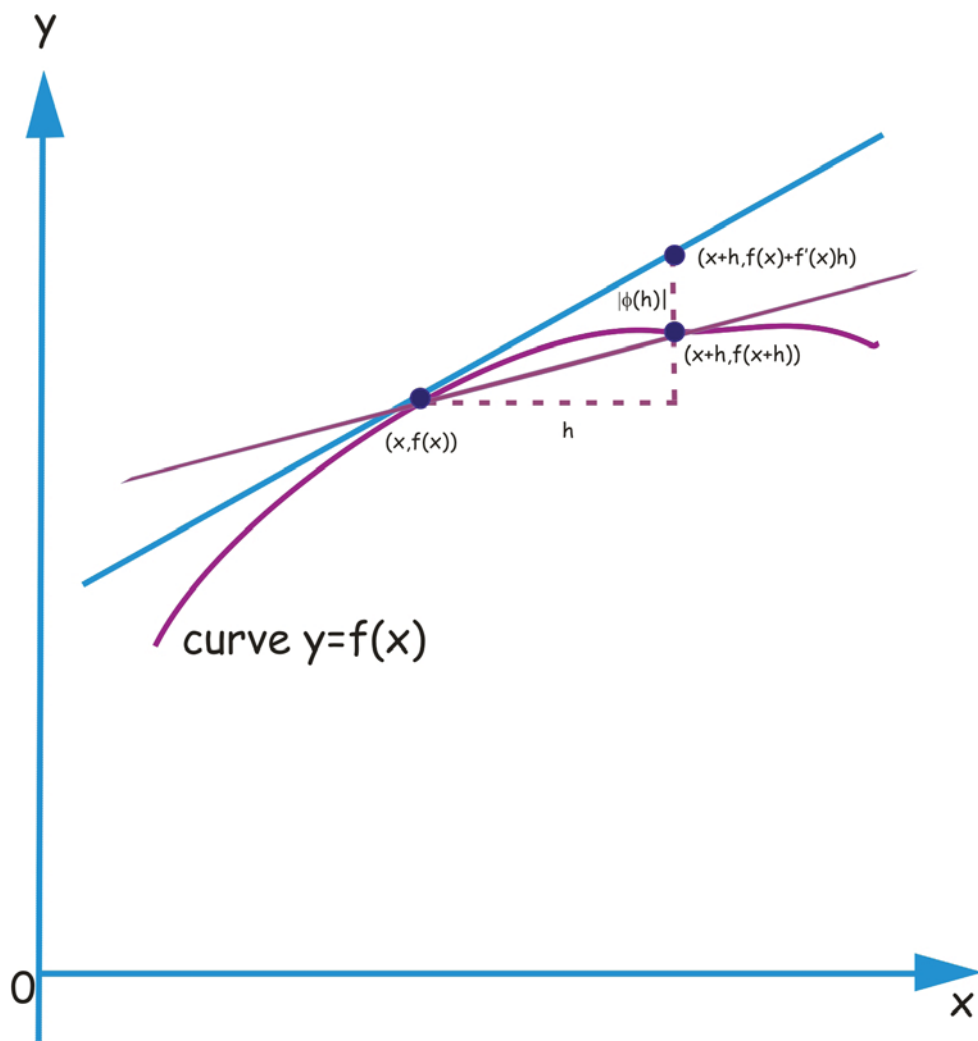


Figure 45: The linear approximation to the derivative (as a function of h holding x fixed) is shown here as $f(x) + f'(x)h$. The difference with $f(x + h)$ is the function $\phi(h)$. One has $f(x + h) - (f(x) + f'(x)h) = \phi(h)$.

We will need this 2nd definition of derivative in proving the chain rule. It is also a useful way to get a rough linear approximation to a non-linear function. Before showing the 2 definitions of derivative are the same, let's do an example.

Example. Let $f(x) = \sqrt{x}$. Then we can show that $f'(x) = \frac{1}{2\sqrt{x}}$.

$$f'(x) = \frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{\sqrt{x+h} - \sqrt{x}}{h} \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}} = \frac{x+h-x}{h(\sqrt{x+h} + \sqrt{x})} = \frac{1}{\sqrt{x+h} + \sqrt{x}} \rightarrow \frac{1}{2\sqrt{x}}, \text{ as } h \rightarrow 0.$$

Here we assume \sqrt{x} is continuous for $x \geq 0$. Exercise. Prove the preceding statement.

Consider using formula (20) to approximate $\sqrt{5} = \sqrt{4+1} \cong \sqrt{4} + L1$, where $L = f'(4) = \frac{1}{4}$. So we get $\sqrt{5} \cong 2.25$. Scientific Workplace says $\sqrt{5} = 2.2361$. Our result was not so bad.

Theorem 37 The usual definition 34 of derivative is equivalent to the linear approximation definition 36 with $f'(c) = L$.

Proof. Definition 34 implies 36. Suppose f is differentiable at the point c according to Definition 34. Set $L = f'(c)$ and set

$$\phi(h) = f(c+h) - f(c) - Lh.$$

Then

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\phi(h)}{h} &= \lim_{h \rightarrow 0} \frac{f(c+h) - f(c) - Lh}{h} = \lim_{h \rightarrow 0} \left(\frac{f(c+h) - f(c)}{h} - L \right) \\ &= \lim_{h \rightarrow 0} \left(\frac{f(c+h) - f(c)}{h} - L \right) = 0, \end{aligned}$$

since $L = f'(c)$. This proves 36.

Definition 36 implies 34. Suppose there is a real number L and there is a function ϕ as in Definition 36. So we have

$$f(c+h) = f(c) + Lh + \phi(h).$$

Subtract $f(c)$ and divide by h to obtain

$$\frac{f(c+h) - f(c)}{h} = \frac{Lh + \phi(h)}{h} = L + \frac{\phi(h)}{h}.$$

Since we know from Definition 36 that $\lim_{h \rightarrow 0} \frac{\phi(h)}{h} = 0$, It follows that $\lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = L$ and f is differentiable according to Definition 34 with $f'(c) = L$. ■

20 Properties of the Derivative.

To make our lives easier when trying to find the formula for the derivative of some function, it will help to know the properties of the derivative.

Just the Facts about Derivatives.

Fact 1) Differentiability implies Continuity. Suppose f is differentiable at c . Then f must be continuous at c .

Fact 2) Derivative is Linear. Suppose f and g are differentiable at c . And suppose k is a real number. Then $f + g$ and kf are differentiable at c and

$$(f + g)'(c) = f'(c) + g'(c); \quad (kf)'(c) = kf'(c).$$

Fact 3) Product Rule. Suppose f and g are differentiable at c . Then so is the product fg and

$$(fg)'(c) = f(c)g'(c) + f'(c)g(c).$$

Fact 4) Quotient Rule. Suppose f and g are differentiable at c and $g(c) \neq 0$. Then the quotient $\frac{f}{g}$ is differentiable at c and

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

Fact 5) Chain Rule. Suppose f is differentiable at c and g is differentiable at $f(c)$. Then the composite $(g \circ f)(x) = g(f(x))$ is differentiable at c and

$$(g \circ f)'(c) = g'(f(c))f'(c).$$

The Leibniz notation makes the chain rule easy to remember, setting $u = f(x)$ and $y = g(f(x)) = g(u)$, we have

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}.$$

It seems to say that you can just cancel the du 's.

Proof. Fact 1) We need to show that the existence of $f'(c)$ implies $f(c+h)$ approaches $f(c)$ as $h \rightarrow 0$. Use the linear approximation Definition 36 of the derivative. This says

$$f(c+h) = f(c) + f'(c)h + \phi(h), \quad \text{where } \frac{\phi(h)}{h} \rightarrow 0 \text{ as } h \rightarrow 0.$$

To finish the proof, you just need to convince yourself that $\phi(h) \rightarrow 0$ as $h \rightarrow 0$. We saw this after Formula (21) for example.

Fact 2) I leave these proofs to you as Exercises.

Fact 3) To show the product rule, we use the trick we needed in the proof of the limit of a product result, i.e., we add and subtract a term in between the 2 things in the numerator of the difference quotient for the derivative:

$$\begin{aligned} \frac{f(c+h)g(c+h) - f(c)g(c)}{h} &= \frac{f(c+h)g(c+h) - f(c+h)g(c) + f(c+h)g(c) - f(c)g(c)}{h} \\ &= \frac{f(c+h)g(c+h) - f(c+h)g(c)}{h} + \frac{f(c+h)g(c) - f(c)g(c)}{h} \\ &= f(c+h) \frac{g(c+h) - g(c)}{h} + \frac{f(c+h) - f(c)}{h} g(c). \end{aligned}$$

Using Fact 1 and properties of limits, we see that this last mess approaches $f(c)g'(c) + f'(c)g(c)$ as $h \rightarrow 0$.

Fact 4) To prove the quotient rule it suffices to do the case $f(x) = 1$, $\forall x$ in the domain of f and g , by the product rule. To do this, look at

$$\frac{\frac{1}{g(c+h)} - \frac{1}{g(c)}}{h} = \frac{1}{h} \frac{g(c) - g(c+h)}{g(c)g(c+h)} = \frac{1}{g(c)g(c+h)} \frac{g(c) - g(c+h)}{h}.$$

Since g is non-zero in an open interval containing c (because $g(c) \neq 0$ and g is continuous at c by Fact 1), the denominator is not vanishing for small enough values of h . Since g is continuous at c by Fact 1, we see that the last mess approaches $\frac{1}{g(c)^2} (-g'(c))$ as $h \rightarrow 0$.

Fact 5) It is easy to convince ourselves of the formula by writing

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \frac{\Delta u}{\Delta x}.$$

Here $\Delta y = g(x + \Delta x) - g(x)$ and $\Delta u = f(x + \Delta x) - f(x)$. The problem with this "proof" is that Δu might vanish at lots of points as $\Delta x \rightarrow 0$. Then you would be dividing by 0 (which is certainly a real no-no). So we need to devise a proof that avoids this pitfall. That is really why we like the 2nd definition of the derivative using the function ψ rather than ϕ ; see Formulas (36) and (21).

Let

$$k = k(h) = f(x+h) - f(x) = f'(x)h + h\psi_1(h) \tag{22}$$

and let $y = f(x)$. Then we have

$$g(f(x+h)) - g(f(x)) = g(y+k) - g(y) = g'(y)k + k\psi_2(k).$$

It follows that

$$\frac{g(f(x+h)) - g(f(x))}{h} = \frac{g'(y)k + k\psi_2(k)}{h} = g'(y) \frac{k}{h} + \frac{k}{h} \psi_2(k).$$

Now from formula (22) we see that $\frac{k}{h}$ approaches $f'(x)$ as $h \rightarrow 0$. We also use the fact that the functions $\psi_2(k)$ and k are continuous at $h = 0$ and both take the value 0 at $h = 0$. Therefore the limit of the last mess is $g'(f(x))f'(x)$ as $h \rightarrow 0$ and the chain rule is proved. ■

The Usual Examples.

Example 1. Let $f(x) = k = \text{constant}$, for all $x \in \mathbb{R}$. Then $f'(x) = 0$, for all $x \in \mathbb{R}$.

Proof.

$$\frac{f(x+h) - f(x)}{h} = \frac{k - k}{h} = 0.$$

Take the limit as $h \rightarrow 0$ and you get 0 (what else?) ■

Example 2. Let $f(x) = ax + b$, $\forall x \in \mathbb{R}$ where a, b are constants. Then $f'(x) = a \forall x \in \mathbb{R}$.

Proof.

$$\frac{f(x+h) - f(x)}{h} = \frac{(a(x+h) + b) - (ax + b)}{h} = \frac{ah}{h} = a.$$

Take the limit as $h \rightarrow 0$ and you get a , which is the derivative. You could also do this example by doing the case $a = 1, b = 0$ first and then using the linearity property of derivatives. ■

Example 3. Let $f(x) = x^n$, where $n = 1, 2, 3, 4, 5, \dots$. Then $f'(x) = nx^{n-1}$.

Proof. by Mathematical Induction.

The Case $n=1$. We have done the case $n = 1$ in Example 2. Here we get $f(x) = x$, $f'(x) = 1x^{1-1} = 1x^0 = 1$.

Induction Step. Use the fact that $x^n = x \cdot x^{n-1}$ and the product rule to see that

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

Here we used the $(n-1)$ st formula. This completes the proof of the Induction step. ■

We can use these examples and the rules for derivatives to deduce that all **polynomials**

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0, \text{ where } a_j \in \mathbb{R}$$

are everywhere differentiable. By the quotient rule we can differentiate **rational functions** $\frac{p(x)}{q(x)}$, for polynomials $p(x)$, $q(x)$, at any point $x \in \mathbb{R}$ such that $q(x) \neq 0$. Any computer with a program such as Scientific Workplace, Matlab or Mathematica can grind out these derivatives and lots more. We have to say more about e^x , $\log x$, $\sin x$, $\cos x$, x^a , for $x > 0$, $a \in \mathbb{R}$, before we can do more interesting derivatives. Then we can use the chain rule to get derivatives of functions like $\sqrt{1+\sqrt{x}}$ or x^x or $\sin(\frac{1}{x})$.

Our next goals are to figure out the mean value theorem and the formula for the derivative of the inverse of a function. Here we mean the inverse function for the operation of composition of functions, such as the derivative of \sqrt{x} knowing the derivative of x^2 .

21 The Mean Value Theorem and Applications

This theorem will be of great use in deducing properties of the function $f(x)$ from properties of its derivative $f'(x)$. It also helps us to prove that any function whose derivative is identically 0 on an interval must be a constant. This fact is important for integration. It explains why the antiderivative tables always have $+C$ at the end of the formulas.

Theorem 38 The Mean Value Theorem. Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) . Then there is a point $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The conclusion of the mean value theorem says that the tangent to the curve $y = f(x)$ at the point $(c, f(c))$ is parallel to the line through the points $(a, f(a))$ and $(b, f(b))$. See Figure 46.

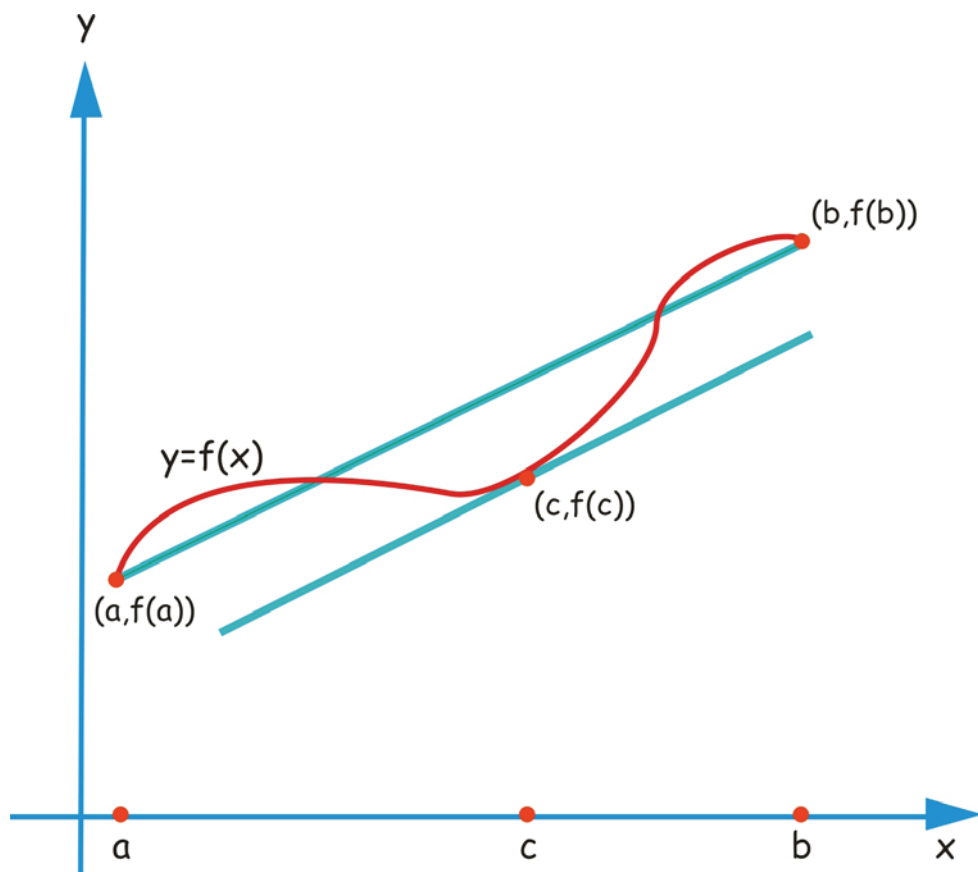


Figure 46: The mean value theorem says the slope of the line through $(a, f(a))$ and $(b, f(b))$ equals the slope of the tangent at some point c .

According to the C.H. Edwards, Jr., in *The Historical Development of the Calculus*, p. 314, our proof is due to O. Bonnet (1819-1890). The proof is first done in a special case and we need to know the first derivative test before proceeding.

Theorem 39 First Derivative Test. Suppose that $f : (a, b) \rightarrow \mathbb{R}$ is differentiable and

$$f(c) \geq f(x) \text{ for all } x \in (c - \delta, c + \delta) \subset (a, b) \text{ for some } \delta > 0.$$

Then we say that f has a **local maximum** at $x = c$. It follows that the derivative $f'(c) = 0$.

Proof. Consider the difference quotient

$$\frac{f(c+h) - f(c)}{h} = \begin{cases} \text{number} \leq 0, & \text{if } \delta > h > 0 \\ \text{number} \geq 0, & \text{if } -\delta < h < 0. \end{cases}$$

See Figure 47. For the derivative $f'(c)$ to exist, this must have a limit as $h \rightarrow 0$. This limit must be 0 as the right hand limit is ≤ 0 , while the left-hand limit is ≥ 0 (since limits preserve inequalities). ■

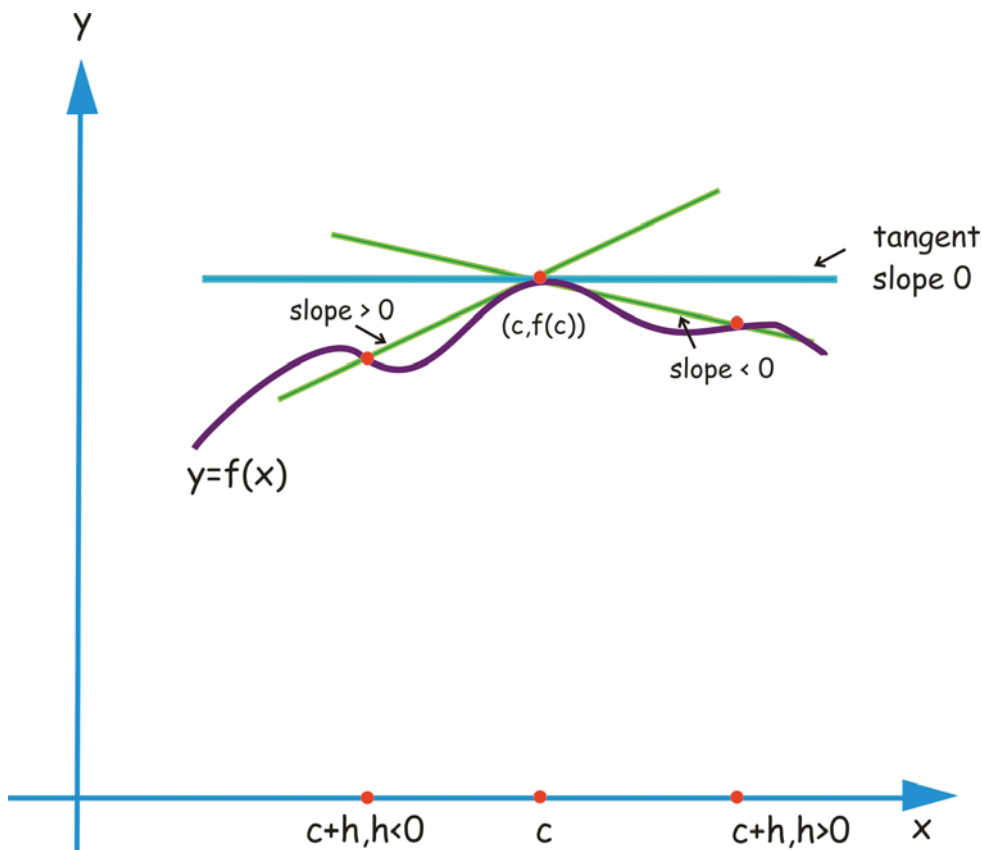


Figure 47: The first derivative test for a local maximum is pictured. The slopes of secant lines on the left are positive while those on the right are negative. So the slope of the tangent line must be 0.

Proof. of the Mean Value Theorem.

Special Case. $f(a) = f(b) = 0$. (also known as Rolle's Theorem).

By the hypothesis of the mean value theorem $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) . In this special case $\frac{f(b)-f(a)}{b-a} = 0$ and thus we need to find a point $c \in (a, b)$ such that $f'(c) = 0$. If $f(x) = \text{constant}$, then any point $c \in (a, b)$ will work. Otherwise we know $f(u) \neq f(a)$ for some $u \in (a, b)$. We may assume $f(u) > f(a)$. Otherwise we can make a similar argument which is left to the reader. Let c be the point in $[a, b]$ such that $f(c) = \max\{f(x) | x \in (a, b)\}$. We know that the maximum of a continuous function is attained on a closed finite interval by the Weierstrass theorem (see Theorem 32). It follows from the fact that $f(a) = f(b)$ that $c \neq a$ and $c \neq b$. Then by the first derivative test we know that $f'(c) = 0$. This completes the proof of the special case.

In general, $f(a) \neq f(b)$.

Then we pull a function out of a hat or bonnet:

$$g(x) \doteq f(x) - \frac{f(b) - f(a)}{b - a}(x - a).$$

The equation of the line joining the points $(a, f(a))$ and $(b, f(b))$ is $h(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$. So $g(x) - f(a)$ is the difference of $f(x)$ and $h(x)$. Then $g(x)$ satisfies the hypotheses of the special case (Rolle's theorem). For $g(a) = f(a) = g(b)$. Moreover $g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$. So the special case tells us that there is a point $c \in (a, b)$ such that $g'(c) = 0$. This means

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

which is exactly what the mean value theorem asserts in the general case and we are done. ■

We get the following corollary immediately since now knowing that the slopes of all tangent lines are positive implies the same for all secant lines.

Corollary 40 *Suppose that $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) . If $f'(x) > 0$ for all $x \in (a, b)$, then the function $f(x)$ is **monotone strictly increasing** on $[a, b]$; i.e., for every pair of points $u, v \in [a, b]$ such that $u < v$, we have $f(u) < f(v)$.*

Proof. By the mean value theorem

$$\frac{f(v) - f(u)}{v - u} = f'(c) \text{ for some } c \in (u, v). \quad (23)$$

Here we replace a, b with u, v , of course. But we know $f'(c) > 0$ and $v - u > 0$. It follows from equation (23) that $f(v) - f(u) > 0$. Thus $f(u) < f(v)$. ■

Similarly one can show that if $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) and $f'(x) < 0$ for all $x \in (a, b)$, then the function $f(x)$ is **monotone strictly decreasing** on $[a, b]$; i.e., for every pair of points $u, v \in [a, b]$ such that $u < v$, we have $f(u) > f(v)$. This is an **exercise**.

Yet another **exercise** is to show that if $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) and $f'(x) \geq 0$ for all $x \in (a, b)$, then the function $f(x)$ is **monotone increasing** on $[a, b]$; i.e., for every pair of points $u, v \in [a, b]$ such that $u \leq v$, we have $f(u) \leq f(v)$. Similarly if $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) and $f'(x) \leq 0$ for all $x \in (a, b)$, then the function $f(x)$ is **monotone decreasing** on $[a, b]$; i.e., for every pair of point $u, v \in [a, b]$ such that $u \leq v$, we have $f(u) \geq f(v)$.

A similar fact is important enough that we state it as a Corollary. We will need it in the theory of integration.

Corollary 41 *Suppose that $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) and $f'(x) = 0$ for all $x \in (a, b)$. Then $f(x)$ is constant.*

Proof. This is left to the reader. If $v \in (a, b)$, use the mean value theorem to see that $\frac{f(v) - f(a)}{v - a} = 0$. ■

In the next theorem we use a fact about the **second derivative** which is of course the derivative of the derivative; i.e., $f'' = (f')'$. We use the 2nd derivative to obtain a test for whether the graph of $f(x)$ is **strictly convex up**, meaning that the graph of $y = f(x)$, for $u \leq x \leq v$ lies below the **chord** (which is the finite part of the secant line) connecting $(u, f(u))$ and $(v, f(v))$ for every pair of points u, v in the domain of f . You can state this as an inequality by parameterizing the chord through $(u, f(u))$ and $(v, f(v))$ as the set of points $t(u, f(u)) + (1 - t)(v, f(v)) = (tu + (1 - t)v, tf(u) + (1 - t)f(v))$ for the parameter $t \in [0, 1]$. Then $f(x)$ is **strictly convex up** if

$$f(tu + (1 - t)v) < tf(u) + (1 - t)f(v), \text{ for all } t \in [0, 1].$$

See Figure 48 for an example.

Of course there will be a similar test telling whether the graph is **strictly convex down** (lying above all the secants). The terminology for this is often confusing. The word "convex" or its negative is often replaced by the word "concave," especially in older calculus books such as E. Purcell, *Calculus with Analytic Geometry*, where the definition is slightly different in another way. Purcell's definition involves the placement of the tangent lines rather than the secant lines.

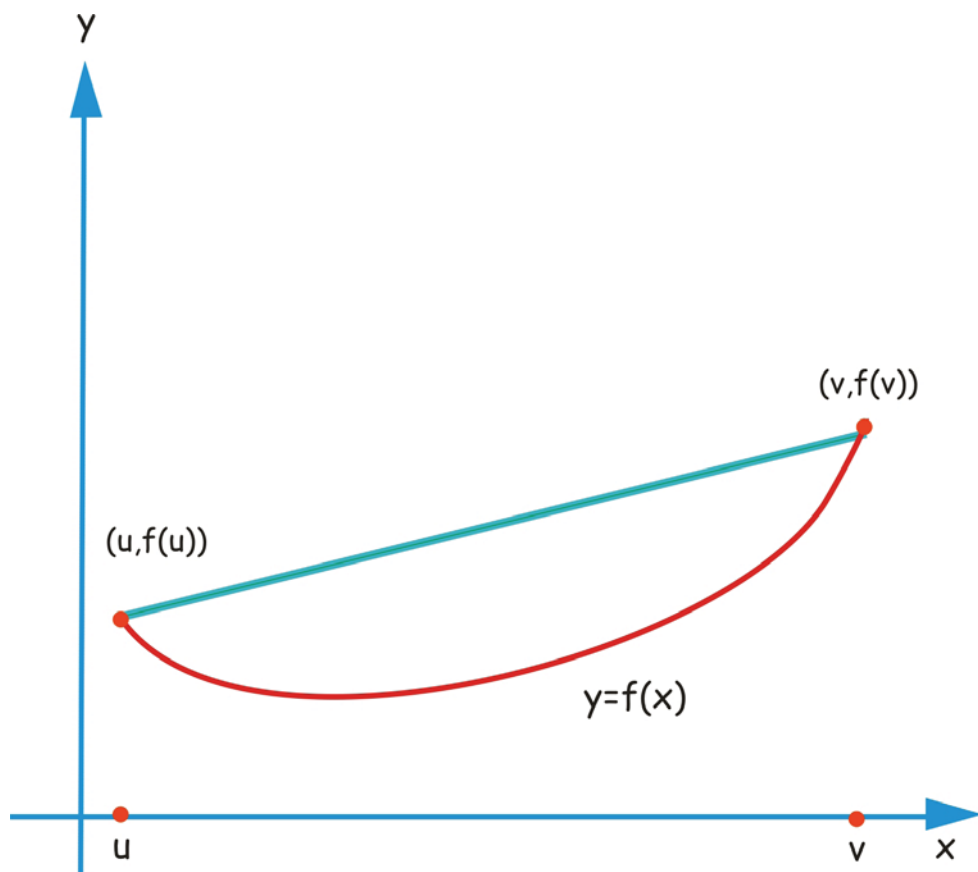


Figure 48: A curve is "convex up" if it always lies below the chord (or finite part) of the secant line for every pair of points u, v in the domain of the function. Here the chord is shown in turquoise while the curve is in red.

Theorem 42 Convexity Test. Suppose that $f : [a, b] \rightarrow \mathbb{R}$ is continuous and the second derivative $f''(x)$ exists and $f''(x) > 0$ for all $x \in (a, b)$. Then $f(x)$ is **strictly convex up**.

Proof. The equation of the secant line is given by

$$L(x) = f(u) + \frac{f(v) - f(u)}{v - u}(x - u).$$

Set $g(x) = L(x) - f(x)$. Our goal is to show that $g(x) > 0$ for all $x \in (u, v)$. Note that $g(u) = g(v) = 0$.

Use the mean value theorem for f on the interval (u, v) to see that $g'(x) = f'(d) - f'(x)$, for some $d \in (u, v)$. Here d depends only on u and v , and is independent of x .

Then $g''(x) = -f''(x)$. The hypothesis that $f''(x) > 0$ for all $x \in (a, b)$ implies that g' is strictly decreasing on $[u, v]$. We see that $g'(x) = \frac{f(v) - f(u)}{v - u} - f'(x)$. We know that $g'(d) = 0$ for some $d \in (u, v)$. So we must have $g'(x) > 0$ for $u < x < d$ and $g'(x) < 0$ for $d < x < v$. It follows from our discussion after the mean value theorem that $g(x)$ strictly increases from 0 to some positive value as x goes from u to d and then $g(x)$ decreases from this positive value back to 0 as x goes from d to v . This is just what we wanted as it shows $g(x) > 0$ for all $x \in (u, v)$. ■

Theorem 43 Second Derivative Test. If $f : [a, b] \rightarrow \mathbb{R}$, suppose $f'(x)$ exists at any x in an open interval containing the point c , and $f'(c) = 0$, while $f''(c) > 0$ for some $c \in (a, b)$. Then $f(c)$ is a **local or relative minimum** of $f(x)$; meaning that $\exists \delta > 0$ s.t. $f(c) \leq f(x)$ for all $x \in [c - \delta, c + \delta]$.

Proof. By the preceding theorem, if we know $f''(x)$ exists and is positive on an open interval containing c , then $f(x)$ is strictly convex up on $[c - \delta, c + \delta]$ for some $\delta > 0$. This makes $f(c)$ a local minimum since $f(x)$ must be greater than $f(c)$ for all $x \in [c - \delta, c + \delta]$.

But we don't need to know that $f''(x)$ exists except at the point $x = c$. For we have

$$0 < f''(c) = \lim_{h \rightarrow 0} \frac{f'(c+h) - f'(c)}{h} = \lim_{h \rightarrow 0} \frac{f'(c+h)}{h}.$$

Therefore $f'(x)$ must be positive on $(c, c + \delta_1)$ for some small $\delta_1 > 0$ and $f'(x)$ must be negative on $(c - \delta_2, c)$ for some small $\delta_2 > 0$. This means that taking $\delta = \frac{1}{2} \min\{\delta_1, \delta_2\}$, we have our result. For then f is decreasing on $[c - \delta, c]$ and increasing on $[c, c + \delta]$. That makes $f(c)$ a minimum on $[c - \delta, c + \delta]$. ■

22 Inverse Functions and Their Derivatives

We seek a rule to tell us how to differentiate $\log x$ assuming we know the derivative of e^x , (or the derivative of \sqrt{x} , assuming we know the derivative of x^2), or (the derivative of $\arcsin(x)$ assuming we know the derivative of $\sin(x)$). This is the inverse function theorem in 1 variable. We will prove a special case.

Theorem 44 Inverse Function Theorem. Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) with $f'(x) > 0$ for all $x \in (a, b)$. Then there is an inverse function $g : [f(a), f(b)] \rightarrow \mathbb{R}$ meaning that $g(f(x)) = x$ for all $x \in [a, b]$ and $f(g(y)) = y$ for all $y \in [f(a), f(b)]$. We often write $g = f^{-1}$ although this is not to be confused with $\frac{1}{f}$. The inverse function g is differentiable and

$$g'(y) = \frac{1}{f'(g(y))}. \quad (24)$$

Note that the Leibniz notation for derivatives makes this theorem look like high school algebra as it says $\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$. Moreover, if we know the inverse function to be differentiable, then the formula for its derivative follows from the chain rule as $g(f(x)) = x$ implies $g'(f(x))f'(x) = 1$ and thus as $x = g(y)$, we get formula (24). Unfortunately we need to show that the inverse function is differentiable. Before doing that, let's look at a few examples.

Example 1. Consider the function

$$y = f(x) = x^n = \underset{n \text{ times}}{x \cdots x}, \quad n = 1, 2, 3, 4, \dots$$

The inverse function is $g(y) = y^{\frac{1}{n}} = \sqrt[n]{y}$, defined for $y \geq 0$ if n is even and all y if n is odd. Then $g'(y) = \frac{1}{n}y^{\frac{1}{n}-1}$. No surprise that it follows the general formula for powers. But we need to prove this using the inverse function theorem. Assume $y > 0$ if n is even and in general that $y \neq 0$. Then by formula (24) we have, using the fact that $f'(x) = nx^{n-1}$,

$$g'(y) = \frac{1}{f'(g(y))} = \frac{1}{n(g(y))^{n-1}} = \frac{1}{n}y^{-\frac{n-1}{n}} = \frac{1}{n}y^{\frac{1}{n}-1}.$$

Here we've used the following definition in Example 2 of a rational power of y .

Example 2. Let $r \in \mathbb{Q}$ and $r = \frac{p}{q}$, with integers p and q , where $q > 0$. Then we can define

$$x^{\frac{p}{q}} = \left(x^{\frac{1}{q}}\right)^p, \tag{25}$$

assuming $x > 0$ if q is even. We leave it as an **exercise** for the reader to show that $f'(x) = rx^{r-1}$.

Proof. of the inverse function theorem.

Step 1. Define f^{-1} .

We are assuming $f : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable on (a, b) with $f'(x) > 0$ for all $x \in (a, b)$. Thus $f(x)$ is strictly increasing on $[a, b]$ by the mean-value theorem. This shows that f is 1-1 on $[a, b]$. By the intermediate value theorem we know that f maps $[a, b]$ 1-1, onto the interval $[f(a), f(b)]$. Thus the inverse function $g : [f(a), f(b)] \rightarrow [a, b]$ is well-defined, 1-1, and onto. Given $y \in [f(a), f(b)]$, we know by the intermediate value theorem that there is an $x \in [a, b]$ such that $f(x) = y$ and so we then define $g(y) = x$. Note that x is unique since f is 1-1.

Step 2. The inverse function g is also strictly increasing.

Suppose that $f(a) < r < s < f(b)$. We want to show that $g(r) < g(s)$. We do a proof by contradiction. If $g(s) \leq g(r)$, then since f is strictly increasing, we have $s = f(g(s)) \leq f(g(r)) = r$. But this says $s \leq r$ while $r < s$ giving us our contradiction. So g must be strictly increasing.

Step 3. The continuity of the inverse function.

We want to show that $g(y)$ is continuous at $r = f(c)$ for $c \in [a, b]$. Let us first assume that c is not an endpoint. Suppose we are given $\varepsilon > 0$ which is small enough that $c \pm \varepsilon \in (a, b)$. Let $g(r + \delta_2) = c + \varepsilon$ and $g(r - \delta_1) = c - \varepsilon$. Set $\delta = \min\{\delta_1, \delta_2\}$. Then we claim that

$$|y - f(c)| < \delta \text{ implies } |g(y) - c| < \varepsilon.$$

To see this, look at Figure 49 and note that

$$|y - f(c)| < \delta \iff -\delta < y - f(c) < \delta \iff f(c) - \delta < y < f(c) + \delta.$$

This implies $r - \delta_1 < y < r + \delta_2$. It follows upon applying g to this inequality, using the fact that g is strictly increasing, we have

$$c - \varepsilon = g(r - \delta_1) < g(y) < g(r + \delta_2) = c + \varepsilon.$$

Thus $|g(y) - c| < \varepsilon$ if $|y - r| < \delta$. This proves the continuity of g . We leave the case that $c = a$ or $c = b$ to the reader as an **exercise**.

Note that the graph of the inverse function g is obtained from that of the function f by reflecting the graph of f across the line $y = x$. Thus it is inconceivable that if the graph of f has no breaks, then the graph of g could have breaks.

Step 4. The formula for the derivative of the inverse function.

As in Step 3, let $f(c) = r$ and thus $g(r) = c$. Then

$$g'(r) = \lim_{y \rightarrow r} \frac{g(y) - g(r)}{y - r} = \lim_{y \rightarrow r} \frac{x - c}{f(x) - f(c)} = \lim_{y \rightarrow r} \frac{1}{\frac{f(x) - f(c)}{x - c}}.$$

This proof is over once we prove that if y approaches r , then $x = g(y)$ approaches $c = g(r)$ as the limit on the right as x approaches c is $\frac{1}{f'(x)} = \frac{1}{f'(g(y))}$. But the continuity of g which was proved in Step 3 says that if y approaches r , then $x = g(y)$ approaches $c = g(r)$. So we're done. ■

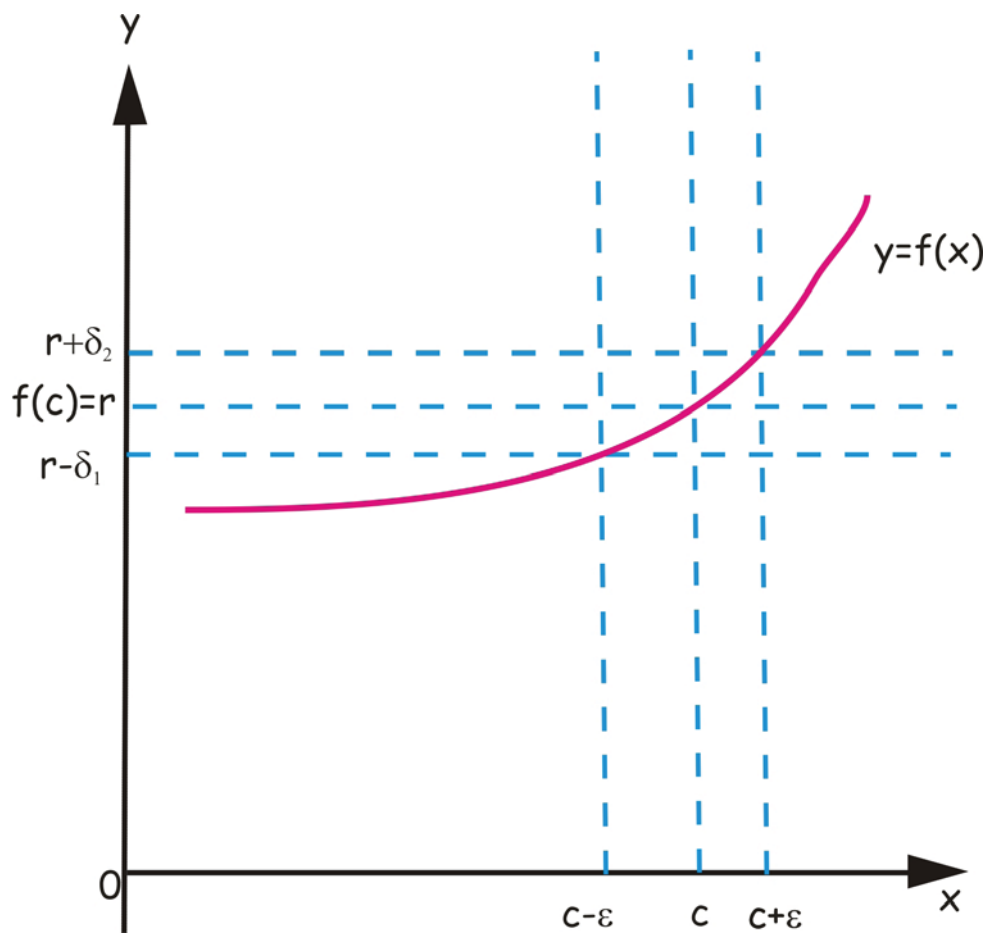


Figure 49: Figure for the proof of the continuity of the inverse of a function $y = f(x)$ with positive derivative on an interval. The inverse function is $x = g(y)$. For a point c let $r = f(c)$. Given small positive ε , define the positive numbers δ_1 and δ_2 by $r + \delta_2 = f(c + \varepsilon)$ and $r - \delta_1 = f(c - \varepsilon)$.

23 Favorite Functions and Their Derivatives

We assume you have seen the functions e^x , $\log x$, $\sin x$, $\cos x$, etc. before. But how do you define them precisely? Here I will take the approach that you believe in infinite series and I will define e^x by its Taylor series, even though we will not discuss series for a while. Then I will define $\log x$ to be the inverse function of e^x . Many calculus books take a different approach and define $\log x$ as an integral. That seems less natural to me and we haven't yet discussed integrals either. As to the trigonometric functions, I will say very little about them here.

23.1 Exponential

Definition 45 The *exponential function* is defined for any real number x by

$$e^x = \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = \lim_{M \rightarrow \infty} \sum_{n=0}^M \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots \quad (26)$$

The power series in formula (26) converges for all real x as we will see later using the ratio test. Assuming that it is legal to differentiate a power series term-by-term where it converges (as we will later prove), we see that

$$\frac{de^x}{dx} = 0 + 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots = e^x.$$

We also see that $e^0 = 1$. It follows that $y = f(x) = e^x$ satisfies the first order ordinary differential equation

$$\frac{dy}{dx} = y \quad \text{and the initial condition } y(0) = 1. \quad (27)$$

The ODE and Initial Condition in Formula (27) gives another way to define $y = e^x$ since the solution to this 1st order ODE with initial condition problem is unique. The good thing about our definition vs this one is that ours gives us a way to compute the function. Maybe that doesn't impress anyone with a calculator or a computer, but it used to make me very happy. The series converges pretty fast. Try computing e as $\sum_{n=0}^{10} \frac{1}{n!} = 2.7183$. This gives the first 5 digits by summing 11 terms of the series. Pretty cool.

Yes, I love this series. It even works for complex numbers $z = x + iy$, where $i = \sqrt{-1}$, for square matrices X , and for p -adic numbers so dearly loved by number theorists. Moreover there is a great generalization to Lie algebras, but we won't go there. The complex number version gives you a way to understand $\sin x$ and $\cos x$ without knowing trigonometry since $e^{i\theta} = \cos \theta + i \sin \theta$ (Euler's formula).

Next let's derive the basic facts about the exponential from the power series definition, making use of theorems about power series that will be proved later in the course. But you saw them in the power series part of the usual calculus class.

Facts About the Exponential Function.

Fact 1. *Exp* takes addition to multiplication.

$$e^{x+y} = e^x e^y.$$

Proof. Multiply the power series and use the binomial theorem. This gives

$$e^{x+y} = \sum_{n=0}^{\infty} \frac{x^n}{n!} \sum_{m=0}^{\infty} \frac{y^m}{m!} = \sum_{k=0}^{\infty} \sum_{\substack{m+n=k \\ m,n \geq 0}} \frac{x^n y^m}{n! m!}.$$

Here we treat the double series as we would a double integral. We change variables from m, n to $n, k = m + n$. It follows that

$$e^{x+y} = \sum_{k=0}^{\infty} \sum_{n=0}^k \frac{x^n y^{k-n}}{n!(k-n)!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^k \frac{k!}{n!(k-n)!} x^n y^{k-n} = \sum_{k=0}^{\infty} \frac{1}{k!} (x+y)^k,$$

by the binomial theorem. ■

Fact 2. *Exp* never vanishes. $e^x \neq 0$ and $e^{-x} = \frac{1}{e^x}$. In fact, $e^x > 0$ for all real numbers x .
Proof. By Fact 1, we know that $e^x e^{-x} = e^{x-x} = e^0 = 1$. This implies that $e^x \neq 0$ and $e^{-x} = \frac{1}{e^x}$. Since the power series defining e^x is positive when $x \geq 0$ as all the terms are ≥ 0 and the first term is 1, we see that $e^x > 0$ for all real numbers x . For $x < 0$, we know $e^x = \frac{1}{e^{-x}}$, and the reciprocal of a positive number is positive. ■

Fact 3. The exponential grows faster than any power of x as x goes to ∞ . More precisely, we have

$$\lim_{x \rightarrow \infty} \frac{e^x}{x^n} = \infty.$$

Proof. By our definition of e^x for $x > 0$, e^x is bounded below by any one term of the power series, e.g.,

$$e^x > \frac{x^{n+1}}{(n+1)!}.$$

Therefore, when $x > 0$, we have

$$\frac{e^x}{x^n} > \frac{x^{n+1}}{(n+1)!} \frac{1}{x^n} = \frac{x}{(n+1)!} \rightarrow \infty, \text{ as } x \rightarrow \infty, \text{ since } n \text{ is fixed.}$$

■

Now we can graph e^x . We know that it is strictly increasing for all real x , since it is its own derivative and it is always positive. The second derivative is the same and positive and thus the graph is convex up everywhere. The function e^x goes to ∞ as $x \rightarrow \infty$. Since $e^{-x} = \frac{1}{e^x}$, it follows that $e^x \rightarrow 0$ as $x \rightarrow -\infty$.

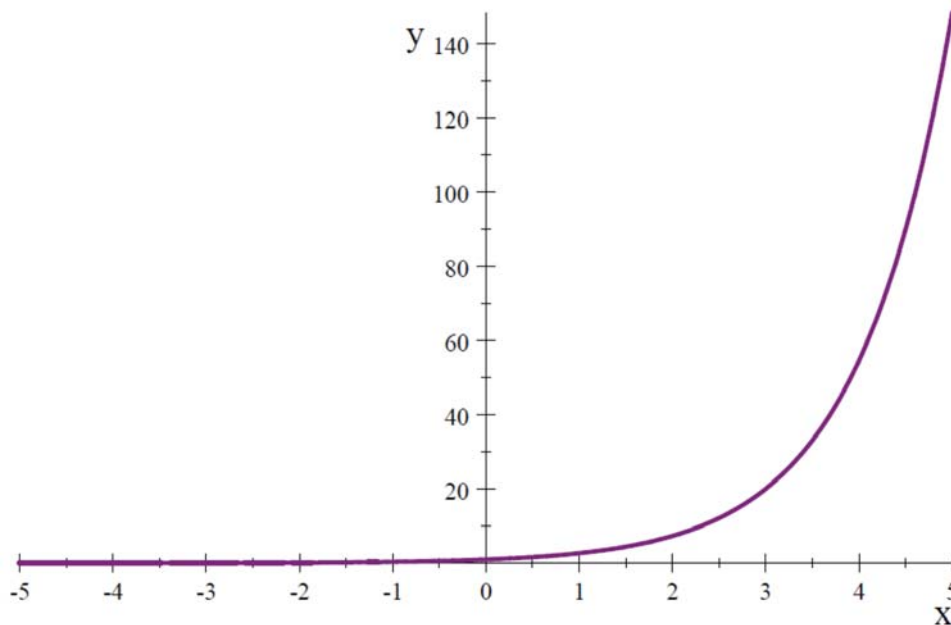


Figure 50: the graph of $y = e^x$

Fact 4. $(e^x)^y = e^{xy}$.

Proof. We postpone our proof until we have discussed $\log x$. ■

Fact 5. e is irrational.

Proof. This is a proof by contradiction. Suppose instead that $e = \frac{a}{b}$, where $a, b \in \mathbb{Z}^+$. Let $k \geq b$ and look at the k th partial sum of the power series for e . If we subtract the k th partial sum from e , we get

$$e - \sum_{n=0}^k \frac{x^n}{n!} = \frac{a}{b} - \sum_{n=0}^k \frac{x^n}{n!} = \sum_{n=k+1}^{\infty} \frac{x^n}{n!}.$$

Multiply this by $k!$ to find

$$k! \left(\frac{a}{b} - \sum_{n=0}^k \frac{x^n}{n!} \right) = k! \sum_{n=k+1}^{\infty} \frac{x^n}{n!}. \quad (28)$$

Now the left-hand side of this last equation is an integer while (by the argument below) the right-hand side is in the interval $(0, 1)$. This is a contradiction, proving that e is irrational.

To see that the right-hand side of equation (28) is in the interval $(0, 1)$, proceed as follows.

$$\begin{aligned} k! \sum_{n=k+1}^{\infty} \frac{x^n}{n!} &= \frac{k!}{(k+1)!} + \frac{k!}{(k+2)!} + \frac{k!}{(k+3)!} + \cdots \\ &= \frac{1}{k+1} + \frac{1}{(k+1)(k+2)} + \frac{1}{(k+1)(k+2)(k+3)} + \cdots \\ &< \frac{1}{k+1} + \frac{1}{(k+1)^2} + \frac{1}{(k+1)^3} + \cdots \end{aligned}$$

This last sum can be evaluated using the formula for the geometric series $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$, if $|x| < 1$. This gives

$$\sum_{n=1}^{\infty} \frac{1}{(k+1)^n} = \frac{1}{k+1} \frac{1}{1 - \frac{1}{k+1}} = \frac{1}{k}.$$

It follows that the right-hand side of equation (28) is in the interval $(0, 1)$ and we have our contradiction. ■

Lambert proved that e and π are irrational in 1761. Gelfond proved that e^π is irrational in 1929. See Hardy and Wright, *Introduction to the Theory of Numbers*, p. 46, for the proof of the irrationality of π . In fact, both e and π are transcendental; i.e., they are not roots of a polynomial with rational coefficients. Note that rational numbers $r = \frac{a}{b}$, with $a, b \in \mathbb{Z}$, $b \neq 0$, are roots of the first degree polynomial with rational coefficients $x - r$ or, if you want integer coefficients $bx - a$.

There is a whole branch of number theory devoted to such questions. In the 1970's a French mathematician named Apéry showed that the following value of the Riemann zeta function is irrational:

$$\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3}.$$

Apéry was an unknown older French mathematician when he found his proof and so no one believed him at first. See Van Der Poorten, "A Proof that Euler Missed ..." in *The Mathematical Intelligencer*, 1 (1979), 195-203. Euler showed that

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}, \quad \zeta(4) = \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90},$$

and similar formulas for $\zeta(2n)$, $n = 3, 4, 5, 6, \dots$ saying that $\zeta(2n) = r\pi^{2n}$, where r is rational. See the last section of Lectures, II for a proof of the 1st formula using Fourier series.

23.2 Logarithm and the Power Function

Here we reverse the treatment of \exp and \log found in most calculus books. Note that $\exp : \mathbb{R} \rightarrow \mathbb{R}^+ = (0, \infty)$ is 1-1 and onto. Why? It is 1-1 since it is its own derivative and it is positive, thus strictly increasing and its graph is convex up. It is onto since e^x certainly approaches ∞ as $x \rightarrow \infty$. Moreover as $x \rightarrow -\infty$, we see that $e^x = \frac{1}{e^{-x}} \rightarrow 0$. Since e^x is continuous, by the intermediate value theorem it must be onto $(0, \infty)$. So e^x satisfies the hypotheses of our theorem on inverse functions (Theorem 44).

The moral of the preceding paragraph is that we can define an inverse function to $y = e^x$.

Definition 46 We define the **logarithm** base e (natural logarithm) denoted $\log y$ by $g(y) = \log y = x$ iff $y = e^x$.

Note that most calculus books call this function $\ln y$ for natural logarithm to distinguish it from the logarithm base 10. We will have no need for base 10 logs and so we will probably hopefully never write $\ln y$.

Properties of the Logarithm

Property 0) The Differential Equation.

The function $g(y) = \log y$ maps $(0, \infty)$ 1-1 onto \mathbb{R} . It satisfies the differential equation $g'(y) = \frac{1}{y}$ with $g(1) = 0$.

Proof. We use Theorem 44 which says that if $y = f(x) = e^x$, then the derivative of the inverse function is obtained as follows:

$$g'(y) = \frac{1}{f'(g(y))} = \frac{1}{f(g(y))} = \frac{1}{y}.$$

Since $f(0) = 1$, we know that $g(1) = 0$. ■

The calculus book usually defines $\log y$ as:

$$\log y = \int_1^y \frac{1}{t} dt.$$

This will make more sense after we have proved the Fundamental Theorem of Calculus (which will be discussed soon). For then we will know that the derivative of the integral as a function of its upper endpoint is just the integrand. And, of course,

$$\int_0^0 \frac{1}{t} dt = 0.$$

Property 1) log takes multiplication to addition; i.e.,

$$\log u + \log v = \log(uv).$$

Proof. To see this, use Fact 1 for e^x . This says

$$e^{x+y} = e^x e^y. \tag{29}$$

Let $u = e^x$ and $v = e^y$. Then take logs of both sides of formula (29). That says

$$\log(e^{x+y}) = \log(e^x e^y).$$

Since log erases exp, we have

$$\log u + \log v = x + y = \log(e^x e^y) = \log(uv).$$

■

Now we are ready to graph $y = \log x$. Of course the graph is obtained by reflecting the graph of e^x across the line $y = x$. Since $\frac{dy}{dx} = \frac{1}{x} > 0$ if $x > 0$, we know that $\log x$ is increasing on $(0, \infty)$. And the second derivative $\frac{d^2y}{dx^2} = -\frac{1}{x^2} < 0$, for $x > 0$. This means that the graph is convex down. We also know that $\log 1 = 0$ and $\log e = 1$. Recall that $e \cong 2.7183$. So we get Figure 51. We also know $\lim_{y \rightarrow 0} \log y = -\infty$ and $\lim_{y \rightarrow \infty} \log y = \infty$ from the corresponding facts about e^x .

$$\begin{array}{l} y \rightarrow 0 \\ y > 0 \end{array}$$

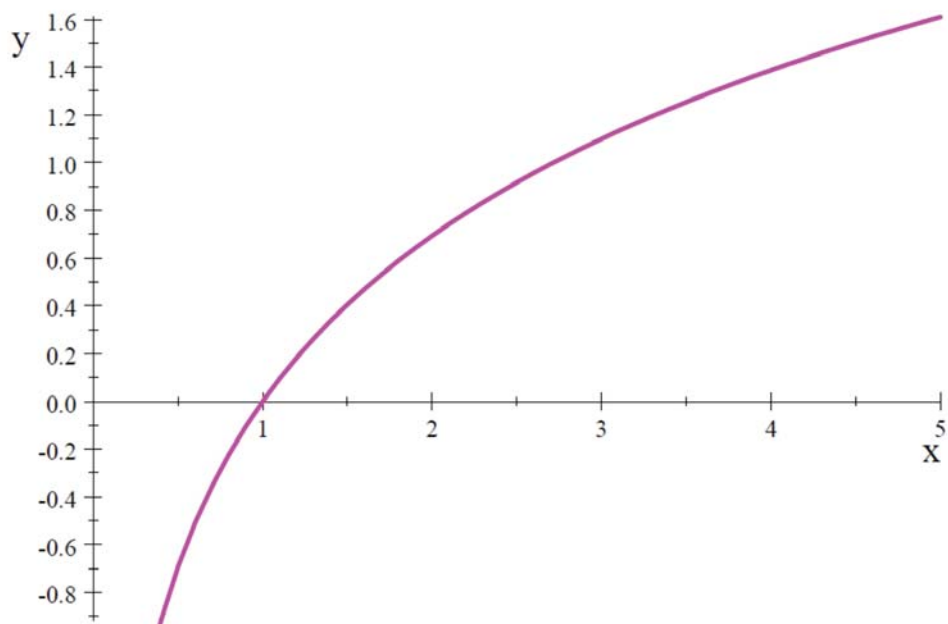


Figure 51: graph of $y = \log x$

Now we define a function that includes most of the powers that we might want to think about for real variables.

Definition 47 The **power function** is defined as follows. Suppose $x > 0$ and a is any real number. Then define $x^a = \exp(a \log x)$.

Proposition 48 (Definitions of Power Functions Agree for Rational Powers.) If $a = \frac{m}{n}$, for $m, n \in \mathbb{Z}, n > 0$, then x^a from Definition 47 gives the same answer as our earlier definition in formula (25). As in Definition 47, we are assuming that $x > 0$.

Proof. Now our old definition said $x^{\frac{m}{n}} = \left(x^{\frac{1}{n}}\right)^m$, for $x > 0$. First one should look at the case $n = 1$, when x^m is defined to be a product of x 's taken m times. Why is that the same as $\exp(m \log x)$? Note that $\log(x \cdots x) = \log x + \cdots + \log x = m \log x$. This proves the log of the desired formula. Since the log is 1-1, this proves the desired formula.

Now look at the special case $m = 1$. By $y = x^{\frac{1}{n}} = \sqrt[n]{x}$, we mean the inverse function to $x = y^n$. So we need to show that with our new definition, $\left(x^{\frac{1}{n}}\right)^n = x$. Well, we have from properties of $e^x = \exp x$, $\left(\exp\left(\frac{1}{n} \log x\right)\right)^n = \exp\left(n \frac{1}{n} \log x\right) = \exp \log x = x$. Note also that $\exp\left(\frac{1}{n} \log x\right)$ is positive.

Finally we leave it to the reader to do the general case as an **exercise**. ■

Property 2 of Log. $\log(x^a) = a \log x$.

Proof. By Definition 47 of x^a and the fact that e^x and \log are inverse functions, we have:

$$\log(x^a) = \log(\exp(a \log x)) = a \log x.$$

■

Facts about Powers.

Property 1 of Powers. $x^a x^b = x^{a+b}$.

Proof. Take the log of both sides. On the left, you get

$$\log(x^a x^b) = \log x^a + \log x^b = a \log x + b \log x = (a + b) \log x.$$

On the right, you get

$$\log x^{a+b} = (a + b) \log x.$$

Since the 2 logs are equal, so are the original entities, because \log is 1-1. ■

Property 2 of Powers. $(x^a)^b = x^{ab}$.

Proof. Again take the log of both sides. On the left, you get

$$\log\left((x^a)^b\right) = b \log(x^a) = ba \log x.$$

On the right, you get

$$\log(x^{ab}) = ab \log x.$$

Since the 2 logs are equal, so are the originals. ■

Example 1. $\lim_{h \rightarrow 0} (1 + h)^{\frac{1}{h}} = e$.

Proof. Take logs. Set $g(y) = \log y$. Recall that $g'(y) = \frac{1}{y}$ and therefore

$$\lim_{h \rightarrow 0} g\left[(1 + h)^{\frac{1}{h}}\right] = \lim_{h \rightarrow 0} \frac{\log(1 + h)}{h} = g'(1) = \frac{1}{1} = 1.$$

So erasing g by exponentiating will compute the desired limit as $f(x) = e^x$ is continuous. Thus

$$f\left(\lim_{h \rightarrow 0} g\left[(1 + h)^{\frac{1}{h}}\right]\right) = \lim_{h \rightarrow 0} \left(f\left(g\left[(1 + h)^{\frac{1}{h}}\right]\right)\right) = \lim_{h \rightarrow 0} \left[(1 + h)^{\frac{1}{h}}\right] = f(1) = e.$$

■

Example 2. $\lim_{y \rightarrow \infty} \frac{\log y}{y} = 0$.

Proof. Set $y = e^x$. Note that $y \rightarrow \infty$ iff $x \rightarrow \infty$. The stated formula is therefore a result of the fact that $\frac{\log y}{y} = \frac{x}{e^x} \rightarrow 0$, as $x \rightarrow \infty$. ■

23.3 Complex Numbers and Trigonometric Functions

Since -1 is not the square of a real number, to get a root of $x^2 + 1 = 0$, we need to enlarge our world beyond the real line. Let i denote a creature whose square is -1 . Thus $i = \sqrt{-1}$ is not real, yes, it's imaginary, but not really more imaginary than most of mathematics. Of course the equation $x^2 + 1 = 0$ also has the root $-i$. And it is really impossible to say which of the 2 roots we are calling i . But worry not. We just call one of them i .

Define the set \mathbb{C} of **complex numbers** to be:

$$\mathbb{C} = \{z = x + iy \mid x, y \in \mathbb{R}\}$$

If $z = x + iy$, with $x, y \in \mathbb{R}$, we say that $x = \operatorname{Re} z$ is the **real part** of z and $y = \operatorname{Im} z$ is the **imaginary part** of z .

We can visualize the complex numbers $z = x + iy$, with $x, y \in \mathbb{R}$ as points (x, y) in the plane. See Figure 52.

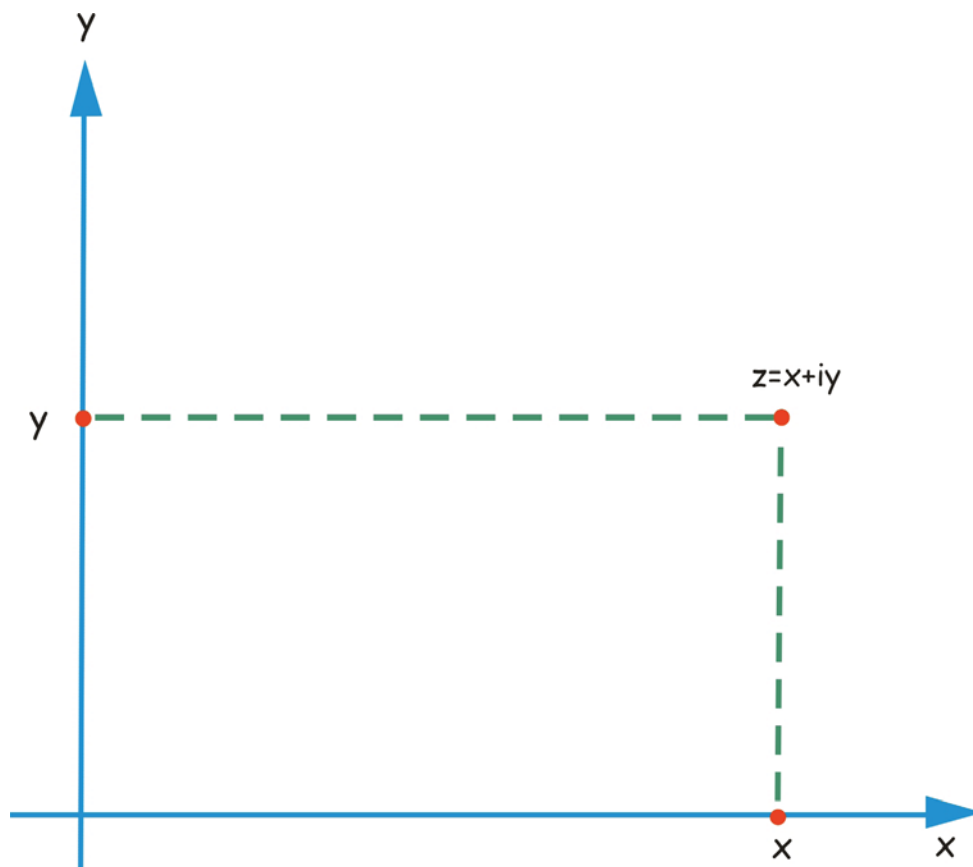


Figure 52: The complex number $z = x + iy$, with x, y real can be pictured as a point in the plane with coordinates (x, y) , or thought of as a vector from the origin to (x, y) .

The set \mathbb{C} forms a field; i.e., it satisfies the 9 field axioms stated in the earlier section on the field axioms for \mathbb{R} , where we have defined sum and product of complex numbers as follows if $z = x + iy$ and $w = u + iv$, with $x, y, u, v \in \mathbb{R}$

$$z + w = (x + u) + i(y + v) \quad \text{and} \quad zw = (x + iy)(u + iv) = xu - yv + i(xv + yu).$$

And if $w \neq 0$, you can divide by w to get

$$\frac{z}{w} = \frac{x + iy}{u + iv} = \frac{x + iy}{u + iv} \cdot \frac{u - iv}{u - iv} = \frac{xu + yv + i(yu - xv)}{u^2 + v^2} = \frac{xu + yv}{u^2 + v^2} + i \frac{yu - xv}{u^2 + v^2}.$$

Here we just recalled an old trick from high school algebra, multiplying by 1 in the form of the conjugate $u - iv$ of the denominator over $u - iv$.

To say that \mathbb{C} forms a field means that you can add, subtract, multiply and divide by non-0 numbers with the usual laws of associativity, distributivity, commutativity, etc. Note however that \mathbb{C} is not an ordered field, just a field. The identity for addition is the complex number $0 = 0 + i0$. A complex number is 0 iff both its real and imaginary parts are 0.

Next we define the complex absolute value.

Definition 49 *Absolute value* of a complex number z is $|z| = \sqrt{x^2 + y^2}$ if $z = x + iy$, with $x, y \in \mathbb{R}$.

Complex conjugate of z is $c(z) = \bar{z} = x - iy$.

Then $|z|^2 = z\bar{z}$. Also $c(z + w) = c(z) + c(w)$ and $c(zw) = c(z)c(w)$. The map $c(z)$ is a continuous function which maps \mathbb{C} 1-1 onto \mathbb{C} fixing elements of \mathbb{R} .

The properties of complex absolute value are the same as those for the real absolute value.

Properties of Complex Absolute Value.

Property 1) $|z| \geq 0$, for all complex numbers z and $|z| = 0$ iff $z = 0$.

Property 2) $|zw| = |z||w|$.

Property 3) triangle inequality $|z + w| \leq |z| + |w|$.

We leave the proofs of these properties to the reader. We will prove these things later in these notes, when we consider normed vector spaces. This time there really is a triangle in the triangle inequality. See Figure 53.

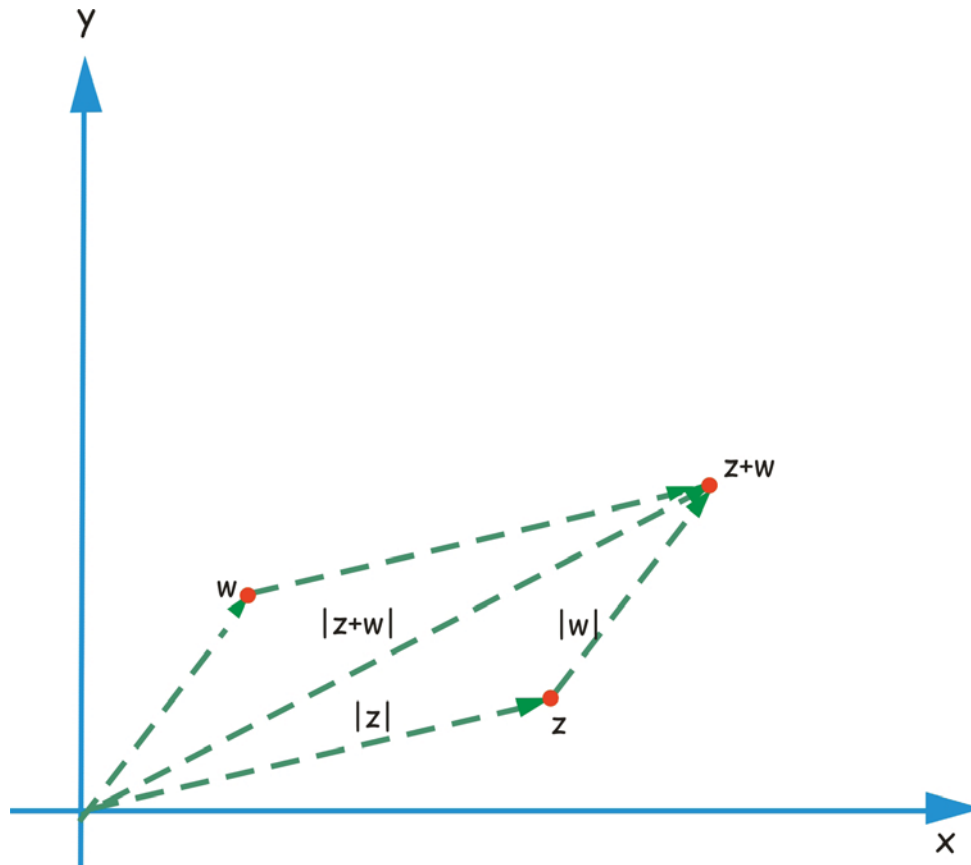


Figure 53: Since the complex absolute value $|z|$ is the length of the vector from the origin to the point representing z in the plane and since the sum of 2 complex numbers z and w corresponds to the sum of the 2 vectors corresponding to z and w , the triangle inequality says the sum of the lengths of 2 sides of a triangle is greater than or equal to the length of the third side.

We could redo all of the things about limits and derivatives for complex numbers, but we won't. That's another course. And we would fall asleep the 2nd time through. In particular series of complex numbers work just like series of real numbers.

If z is a complex number, we can define the **complex exponential** as a complex power series:

$$\exp(z) = e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

It turns out that this series converges for all complex numbers z . Then we have

$$e^{z+w} = e^z e^w \tag{30}$$

just as in the real case. You multiply complex power series just the same way you multiply real power series (carefully). And the binomial theorem is as true for complex numbers as for real numbers.

So we see that

$$e^{x+iy} = e^x e^{iy}.$$

What does this mean? Let's look at e^{iy} :

$$\begin{aligned} e^{iy} &= \sum_{n=0}^{\infty} \frac{(iy)^n}{n!} = 1 + iy + \frac{(iy)^2}{2} + \frac{(iy)^3}{6} + \frac{(iy)^4}{24} + \dots \\ &= 1 + iy - \frac{y^2}{2} - i\frac{y^3}{6} + \frac{y^4}{24} + \dots \\ &= \left\{ 1 - \frac{y^2}{2} + \frac{y^4}{24} + \dots \right\} + i \left\{ y - \frac{y^3}{6} + \frac{y^5}{120} + \dots \right\} \\ &= \cos y + i \sin y. \end{aligned}$$

Here we are defining $\sin x$ and $\cos x$ by their Taylor series. You can deduce the standard trigonometric identities from the properties of the complex exponential like formula (30).

One finds that $|e^{iy}| = 1$, for $y \in \mathbb{R}$. To see this note that $|e^{iy}|^2 = e^{iy} e^{-iy} = e^{iy-iy} = e^0 = 1$. Since $|e^{iy}|$ must be positive it follows that $|e^{iy}| = 1$ for all real y . This says $\sin^2 y + \cos^2 y = 1$, for $y \in \mathbb{R}$.

See Figure 54.

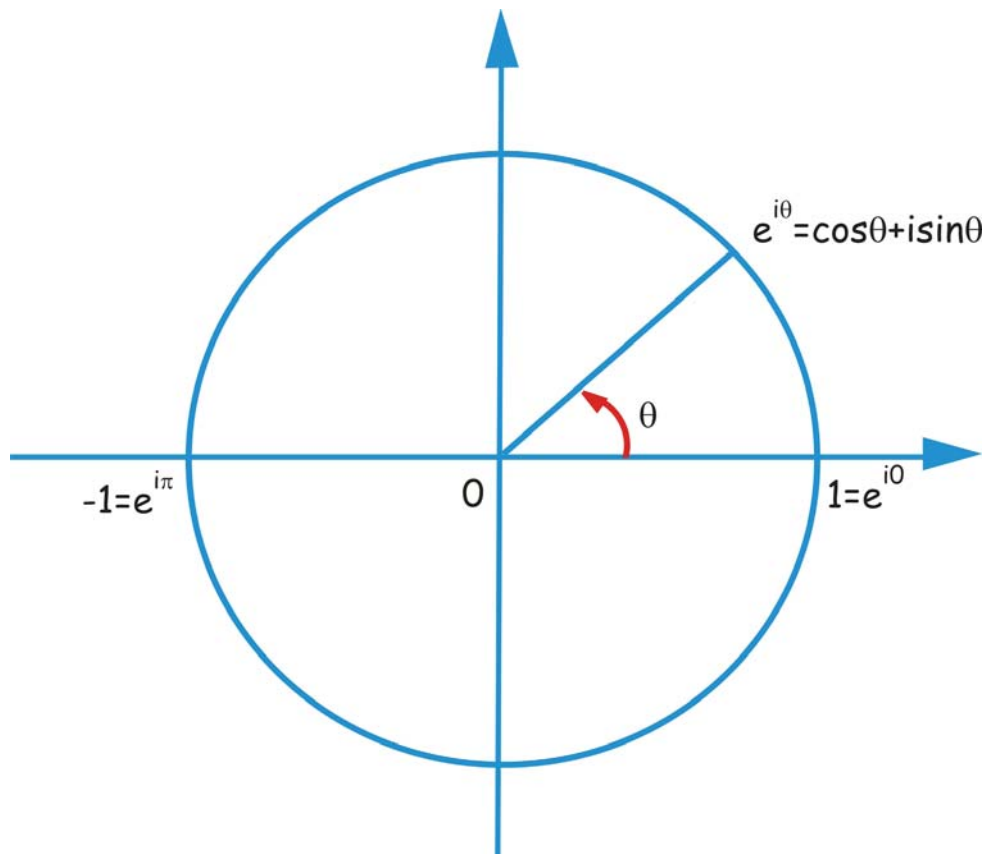


Figure 54: For an acute angle θ , $\cos \theta$ is the x -coordinate of $e^{i\theta}$ and $\sin \theta$ is the y -coordinate. In particular, the complex absolute value $|e^{i\theta}| = 1$.

If we define $\sin x$ and $\cos x$ by their Taylor series, we have

$$f(x) = \cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \quad \text{and} \quad g(x) = \sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}. \quad (31)$$

You can check (assuming it is legal to differentiate term-by-term) that $\sin x$ and $\cos x$ satisfy the following system of ODEs with initial conditions:

$$f'(x) = -g(x) \quad \text{and} \quad g'(x) = f(x), \quad f(0) = 1 \quad \text{and} \quad g(0) = 0. \quad (32)$$

It is possible to define \sin and \cos to be solutions of the initial value problems in equation (32). Lang, *Undergraduate Analysis*, defines the trig functions this way. I don't find this to be very satisfying since it does not tell you how to compute them. Of course you might prefer the geometric definition seen in Figure 54 for an acute angle x measured in radians. All definitions are the same once you show that whatever your definition of \sin and \cos it satisfies formula (32). For you learn in differential equations that solutions to such things are unique.

I prefer to deduce all the standard trig identities from our Taylor series for \sin and \cos , $e^{iy} = \cos y + i \sin y$, $e^{z+w} = e^z e^w$. It is a nice **exercise**. For example, you can easily see from the fact that all powers in the Taylor series for $\cos x$ are even, that $\cos(-x) = \cos x$.

To figure out the addition formula for $\cos x$, look at

$$e^{iu+iv} = e^{iu} e^{iv} = (\cos u + i \sin u) (\cos v + i \sin v) = \cos u \cos v - \sin u \sin v + i(\cos u \sin v + \sin u \cos v).$$

Therefore

$$\cos(u + v) = \cos u \cos v - \sin u \sin v.$$

I learned this treatment of elementary functions as by power series an undergrad reading P. Dienes, *The Taylor Series*, Chapter 4. I really like this approach since it allows you to compute the functions and minimizes the memorization of trig identities.

Of course we still have the problem of defining π . Lang defines π by saying that $\frac{\pi}{2}$ is the first positive zero of $\cos x$. It takes a bit of effort to see that such a zero exists if you are not allowed to look at Figure 54.

Then Lang goes on to show that \sin and \cos are periodic of period 2π . This is certainly not obvious from the Taylor series in formula (31). Any finite combination via sum, product, composite, difference, quotient, inverse of the functions seen up to now (polynomials, powers, exponentials, sines, cosine) are called **elementary functions**. Of course they might be very complicated such as

$$e^{\pi x} - 2001 \log \sqrt{x - 500x^{\frac{2}{3}} + \sqrt{x}}.$$

These are the functions considered in calculus.

There are other favorite functions with many applications; e.g., in statistics, physics,... A good reference for many such functions is N.N. Lebedev, *Special Functions*. Mathematica and Matlab know many of these functions: the error function, Bessel functions, Airy integrals,

We have already seen the gamma function in the introduction. The incomplete gamma function is another favorite, as are Legendre polynomials, also known as spherical harmonics because they arise in problems with spherical symmetry, such as the problem of understanding the vibrations of the earth after a large earthquake, the solution of the Schrödinger equation for the hydrogen atom, or the study of the sun's magnetic field.

Not every function that you can think of is expressible by its Taylor series. One example is the infinitely differentiable function defined by

$$f(x) = \begin{cases} \exp(\frac{-1}{x^2}), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

One can show that all the derivatives $f^{(n)}(0) = 0$, for $n = 0, 1, 2, 3, \dots$. This implies that the Taylor series of this function at the origin is identically 0. We will say more about that later when we discuss Taylor series. Similar functions can be used to glue 0 to 1 in an infinitely differentiable way (C^∞ - glue).

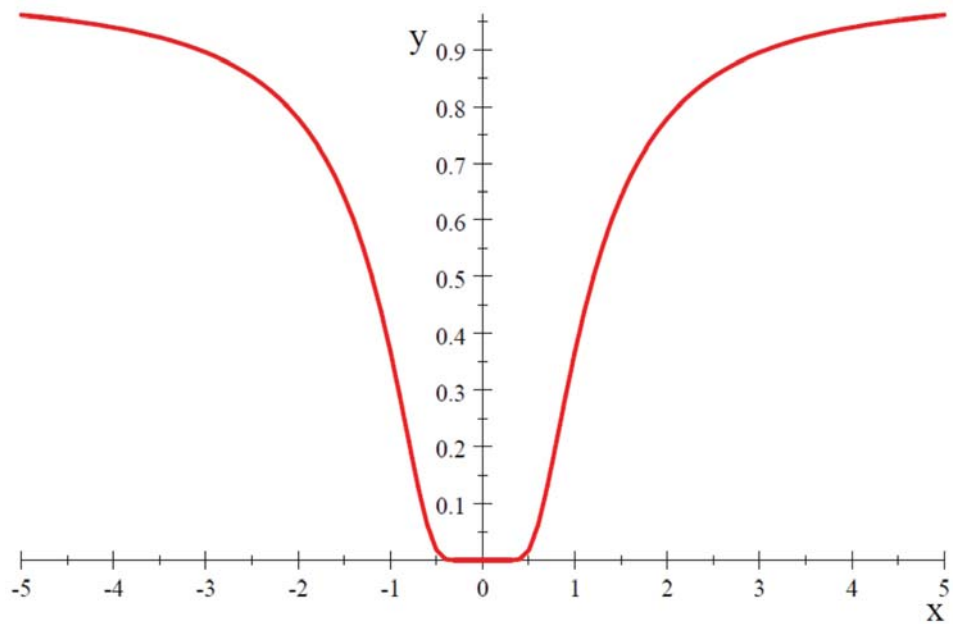


Figure 55: The function $f(x) = \begin{cases} \exp(-\frac{1}{x^2}), & x \neq 0 \\ 0, & x = 0 \end{cases}$ is an infinitely differentiable function not represented by its Taylor series at 0.

Basic Properties of Integrals

24 Axioms for Integrals and the Fundamental Theorem of Calculus

In this section we prove the basic properties of the Riemann integral of a continuous function on a finite interval from 2 basic axioms. We will not show that an integral satisfying these 2 axioms exists until much later, using a method which may seem different from that of limits of Riemann sums that you saw in calculus. It is an intermediate method between that usually used for Riemann integrals and that used for Lebesgue integrals. Both kinds of integrals give the same answer for continuous functions on finite intervals, or even piecewise continuous functions on finite intervals (i.e., functions with a finite number of jump discontinuities; that is, points where the right- and left- hand limits exist but are different). An example is seen in Figure 33.

Both Riemann and Lebesgue integrals were invented to discuss Fourier analysis. Yes, the Fourier coefficients in equation (2) are integrals. Both ideas of integrals start with area. The integral of a constant function $f(x) = c$, for all x in $[a, b]$ will be $c(b - a)$ which is the area under the curve. See Figure 56.

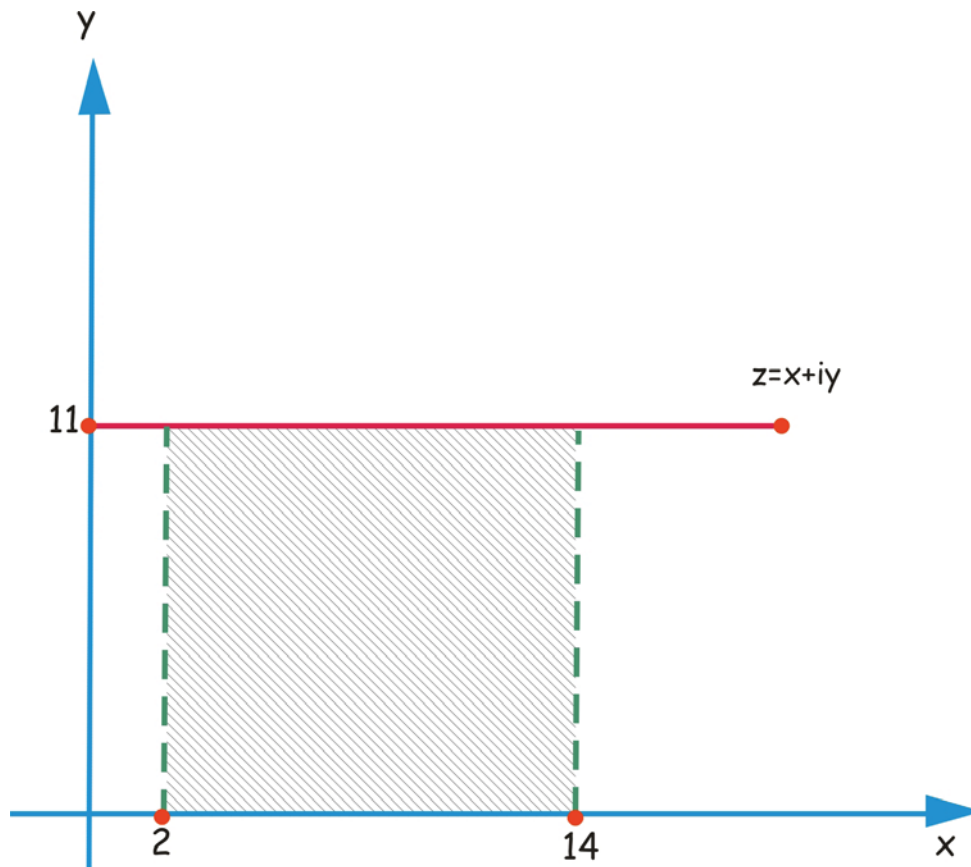


Figure 56: The integral of the constant function $f(x) = 11$ from 2 to 14 is $11 * 12 = 132$.

But, of course, there are many more reasons we need integrals; for example, to compute probabilities. Many tests from statistics come from computing $\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$.

Anyway, the amazing thing is that we can do most of the stuff you learned in calculus about integrals from 2 measly

axioms. We will write definite integrals as $I_c^d f$ or $\int_c^d f$ rather than $\int_c^d f(x)dx$ to emphasize that the definite integral is a real number depending on the function f , the interval $[c, d]$ and nothing else. The variable x is a "dummy variable." It is there to help us evaluate integrals; for example, when we want to make a substitution. But the definite integral is not a function of the variable x .

The 2 Simple Axioms for Integrals

Suppose that $f : [a, b] \rightarrow \mathbb{R}$ is continuous and $a < b$ are finite numbers. We write for $a < c < d < b$

$$I_c^d f = \int_c^d f = \int_c^d f(x)dx.$$

Axiom I. Suppose that $m \leq f(x) \leq M$, for all $x \in [a, b]$. Then

$$m(d - c) \leq I_c^d f \leq M(d - c).$$

See Figure 57

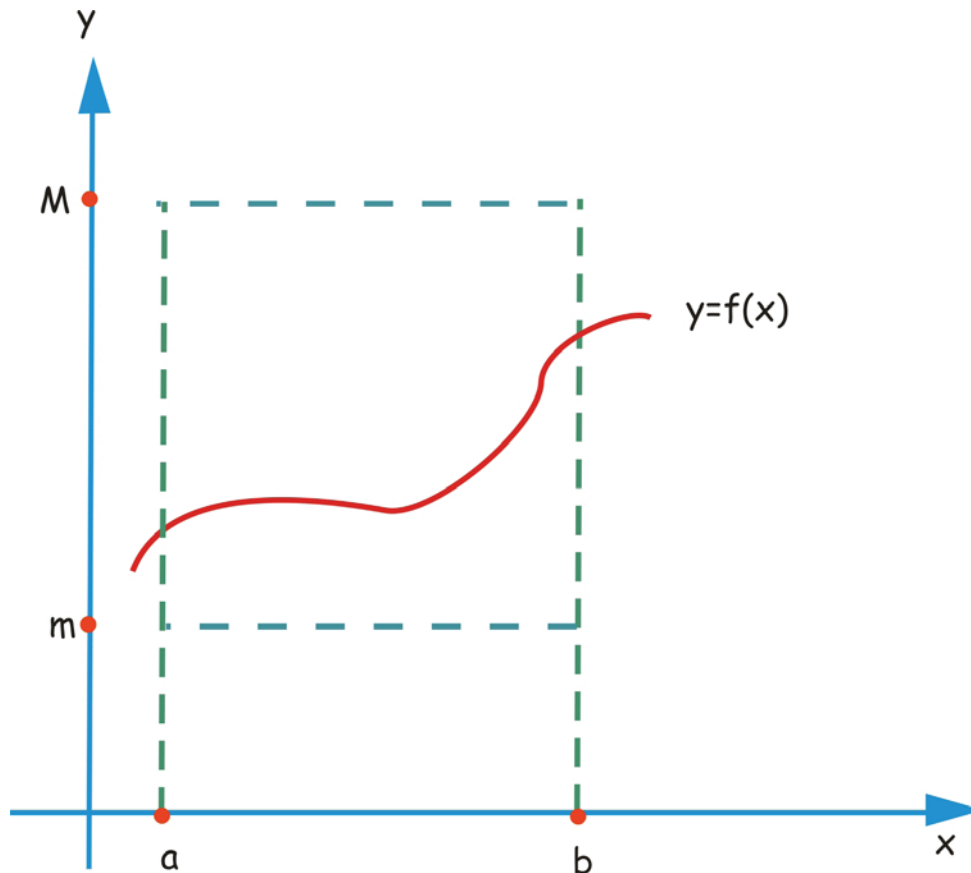


Figure 57: Illustration of Axiom 1 for integrals showing that when the function is between the positive numbers m and M , then you expect the integral over $[a, b]$ to be between the area of the large rectangle which is $M(b - a)$ and that of the small rectangle which is $m(b - a)$.

Axiom II. Suppose $c < r < d$. Then

$$I_c^d f = I_c^r f + I_r^d f.$$

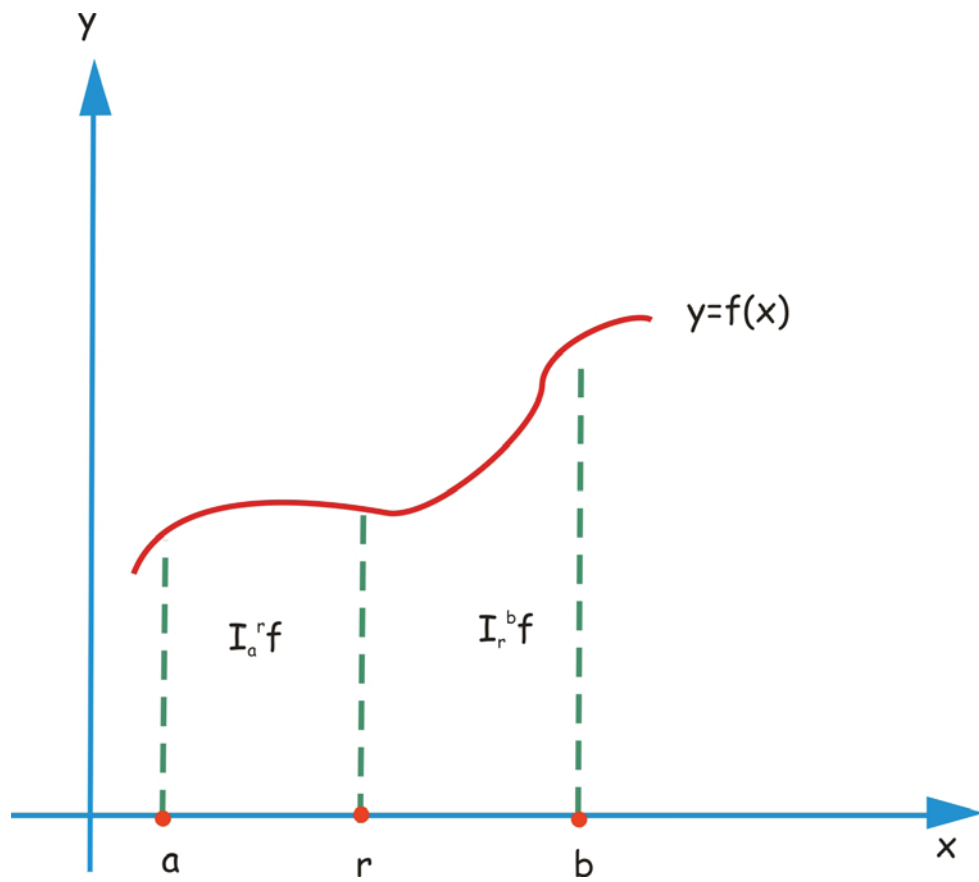


Figure 58: Illustration of Axiom 2 showing that the integral of f on $[a, b]$ should equal the sum of the integrals of f on $[a, r]$ and on $[r, b]$ if $a < r < b$.

It follows from Axiom I that $I_c^c f = 0$. If $c > d$, define $I_c^d f = -I_d^c f$.

Later we will construct the integral starting with integrals of step maps or piecewise constant functions on the finite interval $[a, b]$. A step map looks like $f(x) = w_i$, $a_{i-1} < x < a_i$, for $i = 1, \dots, n$. Here $a = a_0 < a_1 < a_2 < \dots < a_n = b$. Then we define

$$I_a^b f = \int_a^b f = \sum_{i=1}^n w_i (a_i - a_{i-1}).$$

See Figure 59.

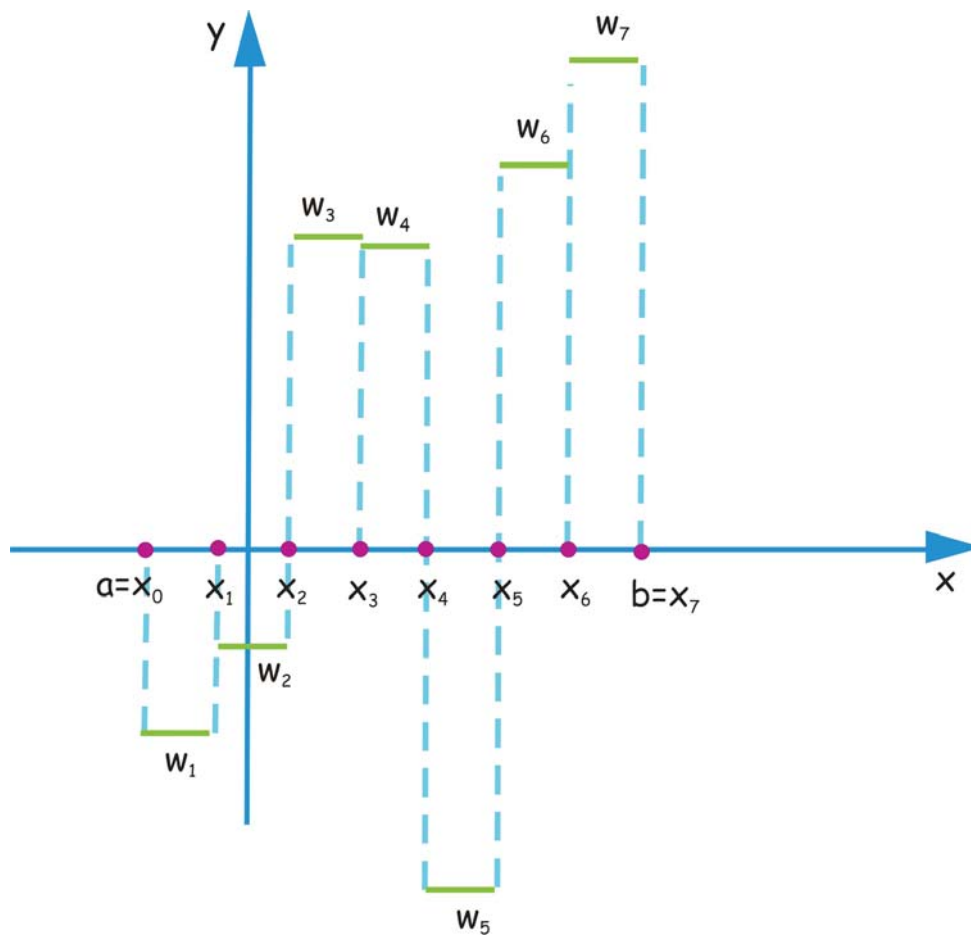


Figure 59: graph of a step function with values w_j on the j th subinterval of the partition of $[a, b]$

Integrals are extended to continuous functions and others by taking limits of Cauchy sequences of step functions using the absolute value norm

$$\|f\|_1 = \int_a^b |f(x)| dx.$$

But we have to talk about normed vector spaces first.

For the rest of this part of the lectures we assume that there exists an integral satisfying Axioms I and II. These 2 puny axioms are enough to be able to prove the Fundamental Theorem of Calculus, that famous theorem that Newton and Leibniz fought to own.

Theorem 50 Fundamental Theorem of Calculus. Suppose that we have an integral satisfying our 2 axioms able to integrate all continuous functions on finite intervals. Then if $f : [a, b] \rightarrow \mathbb{R}$ is continuous with $-\infty < a < b < \infty$, we have

$$\frac{d}{dx} \int_a^x f = \frac{d}{dx} \int_a^x f(t) dt = f(x) \quad \text{for all } x \in [a, b].$$

Proof. Case 1. The Right-Hand Derivative.

For the right-hand derivative, we need to investigate the difference quotient below with $h > 0$. We use Axiom II to see:

$$\frac{\int_a^{x+h} f - \int_a^x f}{h} = \frac{\int_x^{x+h} f}{h}.$$

Here we use the fact that $h > 0$ and thus $a < x < x + h < b$ when h is near enough to 0.

Let $f(u) = \min\{f(t) \mid x \leq t \leq x + h\}$ and $f(v) = \max\{f(t) \mid x \leq t \leq x + h\}$. We know that $u, v \in [x, x + h]$ exist by the Weierstrass theorem on the existence of maxima and minima for continuous functions on closed finite intervals. So now by Axiom I, we have

$$f(u) = \frac{f(u)h}{h} \leq \frac{\int_x^{x+h} f}{h} \leq \frac{f(v)h}{h} = f(v).$$

Now let h approach 0 from above. The quantity in the middle approaches the right-hand derivative of $\int_a^x f$ as a function of x , holding a fixed. The quantities on the outside of the inequalities approach $f(x)$ by the continuity of f . So by the Squeeze Lemma, the right-hand derivative is $f(x)$.

Case 2. The Left-Hand Derivative.

We leave it to the reader as an **exercise** to show that the left-hand derivative is also $f(x)$. This is a little tricky because the denominator in the difference quotient is now negative and $a < x + h < x < b$ for $h < 0$. One has, again using Axiom 2,

$$\frac{\int_a^{x+h} f - \int_a^x f}{h} = \frac{-\int_{x+h}^x f}{h} = \frac{\int_{x+h}^x f}{-h}.$$

Now use the same argument as in Case 1 to see that the left-hand derivative is $f(x)$, to complete the proof. ■

There was a huge quarrel over who invented calculus and "proved" the fundamental theorem - Newton or Leibniz. The effects may still persist. According to Wiener in 1949, "it became an act of faith and of patriotic loyalty for British mathematicians to use the less flexible Newtonian notation and to affect to look down on the new work done by the Leibnizian school on the Continent When the great continental school of the Bernoullis and Euler arose (not to mention Lagrange and Laplace who came later) there were no men (sic) of comparable calibre north of the Channel to compete with them"

See Edna E. Kramer, *The Nature and Growth of Mathematics*, p. 172.

Anyway it is relatively simple now to prove all the basic rules of calculus that you know and love, using our 2 axioms and the fundamental theorem. No more memorizing formulas! Next we show the basic fact which allows the evaluation of integrals by finding antiderivatives.

Corollary 51 *The integral is unique. Assume that $f : [a, b] \rightarrow \mathbb{R}$ is continuous with $-\infty < a < b < \infty$. If F is differentiable on (a, b) with $F' = f$, then for $x \in (a, b)$, we have*

$$\int_a^x f(t) dt = \int_a^x f = F(x) - F(a). \tag{33}$$

Proof. Well, by the fundamental theorem, we know that the derivative with respect to x of the left-hand side of formula (33) is $f(x)$. By Corollary 41 of the mean-value theorem, formula (33) is true up to a constant. That is, we have a constant C such that

$$\int_a^x f = F(x) + C.$$

To determine C , set $x = a$. We saw earlier that $\int_a^a f = 0$. Thus $0 = F(a) + C$, which implies $C = -F(a)$, proving our Corollary. ■

The Fundamental Theorem of Calculus says that the derivative erases the integral:

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

The next Corollary essentially says the integral erases the derivative. It is just a restatement of the preceding Corollary.

Corollary 52 *With the same hypotheses as the preceding Corollary, we have*

$$\int_a^x \frac{d}{dt} F(t) dt = F(x) - F(a).$$

As a result of the fundamental theorem and the preceding Corollary we see that the integral and the derivative are essentially inverse functions on functions. Thus the indefinite integral is called an antiderivative and written $\int f$. It is only determined up to a constant by Corollary 41 of the mean value theorem.

Examples.

Example 1. Constant Function.

Suppose $f(x) = c = \text{constant}$, for all $x \in [a, b]$. Then by Axiom I we have

$$c(b - a) \leq \int_a^b f \leq c(b - a).$$

It follows that

$$\int_a^b c = c(b - a). \tag{34}$$

If $c > 0$, this integral is the area of the rectangle bounded by $y = c$, the x -axis, $x = a$, and $x = b$.

If $c < 0$, the integral is the negative of the area of this rectangle. We could also prove formula (34) from the preceding Corollaries. For the derivative of $F(x) = cx$ is c and thus

$$\int_a^b c = F(b) - F(a) = cb - ca = c(b - a).$$

Example 2. Standard Examples from Calculus.

By the preceding Corollaries any derivative formula gives an integral formula when read backwards. Any table of derivatives is a table of integrals. So for example, we have the following formulas, assuming that $0 < a < b$ for the middle formula:

$$\int_a^b e^x dx = e^b - e^a, \quad \int_a^b \frac{1}{y} dy = \log b - \log a, \quad \int_a^b \cos x dx = \sin b - \sin a.$$

25 Rules for Integrals

Most of the following rules for integrals come from some property of derivatives and the fundamental theorem of calculus.

Rules for Integrals. Suppose that f and g are continuous on $[a, b]$.

Rule 1) Linearity. If $\alpha, \beta \in \mathbb{R}$, then

$$\int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g.$$

Rule 2) Integrals Preserve \leq . Suppose that $f(x) \leq g(x)$ for all $x \in [a, b]$. Then

$$\int_a^b f \leq \int_a^b g.$$

It follows that

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

Rule 3) Positivity. Suppose that $f(x) \geq 0$ for all $x \in [a, b]$. If there is a point $c \in [a, b]$ such that $f(c) > 0$, then

$$\int_a^b f > 0.$$

Rule 4) Substitution.

Suppose $g'(x)$ is continuous on $[a, b]$. Then, if f is continuous on $g[a, b]$,

$$\int_{g(a)}^{g(b)} f(u) du = \int_a^b f(g(x)) g'(x) dx.$$

Rule 5) Integration by Parts.

Suppose that f' and g' are continuous on $[a, b]$. Then

$$\int_a^b f g' = f(b)g(b) - f(a)g(a) - \int_a^b f' g.$$

Proof. Rules 1), 4), 5) come from corollary 51 and corresponding rules for derivatives. So let's do these first.

Rule 1). We want to show that for all $x \in [a, b]$:

$$\int_a^x (\alpha f + \beta g) = \alpha \int_a^x f + \beta \int_a^x g.$$

Take derivatives of the function of x on the left using the fundamental theorem. You get $(\alpha f + \beta g)(x)$. But the derivative of the function of x on the right is also $(\alpha f + \beta g)(x)$, using the linearity of derivatives.

So now we know that the left hand function of x has the same derivative as the right hand function of x . But then corollary 41 of the mean value theorem says that the two functions differ by a constant K ; i.e.,

$$\int_a^x (\alpha f + \beta g) = \alpha \int_a^x f + \beta \int_a^x g + K.$$

What is K ? Set $x = a$ and use the fact that $\int_a^a h = 0$. This says K must be 0 and Rule 1 is proved.

Rule 4). We want to prove that for all $x \in [a, b]$:

$$\int_{g(a)}^{g(x)} f(u) du = \int_a^x f(g(t))g'(t) dt.$$

Again we differentiate the left hand function of x , using the fundamental theorem of calculus and the chain rule, obtaining $f(g(x))g'(x)$. When we differentiate the right hand side as a function of x , we get the same answer using just the fundamental theorem. Again Corollary 41 of the mean value theorem tells us that the left and right hand side must differ by a constant. Plug in $x = a$ to see that the constant must be 0.

Rule 5). Prove this as an **exercise** imitating the proof of Rule 1 using the rule for the derivative of a product.

Rule 2). We want to show assuming $a < b$ and $f(x) \leq g(x)$ for all $x \in [a, b]$ that $\int_a^b f \leq \int_a^b g$. To do this, look at

$h = g - f$. Then h is ≥ 0 on $[a, b]$ and Axiom 1 for integrals says $\int_a^b h \geq 0(b-a) = 0$. By the linearity property just proved,

then $\int_a^b g - \int_a^b f = \int_a^b h \geq 0$. This implies the result.

Rule 3). We want to prove that $\int_a^b f > 0$. For this, you should draw a picture. See Figure 60. By the continuity of f at c , if we're given $\varepsilon = f(c)/2$, then there is a δ so that $|x - c| < \delta$ implies $|f(x) - f(c)| < f(c)/2$. This means that

$$\frac{-f(c)}{2} < f(x) - f(c) < \frac{f(c)}{2}.$$

Add $f(c)$ to this and get, for $x \in (c - \delta, c + \delta)$,

$$0 < \frac{f(c)}{2} < f(x) < \frac{3f(c)}{2}.$$

It follows using Axiom 2 for integrals and the fact that the integral preserves \leq

$$\int_a^b f = \int_a^{c-\delta} f + \int_{c-\delta}^{c+\delta} f + \int_{c+\delta}^b f \geq \int_{c-\delta}^{c+\delta} \frac{f(c)}{2} = \frac{f(c)}{2}(2\delta) > 0.$$

Here we have assumed $a < c < b$. If $a = c$ or $c = b$, the result still works. We leave this to you to prove as an **exercise**. ■

26 Questions and History

A Question to Ponder. Can you find a formula for any indefinite integral?

For example, consider $\int_0^x \frac{dt}{(1-t^2)^{\frac{1}{4}}}$ and $\int_0^x e^{-t^2} dt$.

Answer.

Joseph Liouville (1809-1882) proved that such integrals as

$$Erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt = \text{the error function}$$

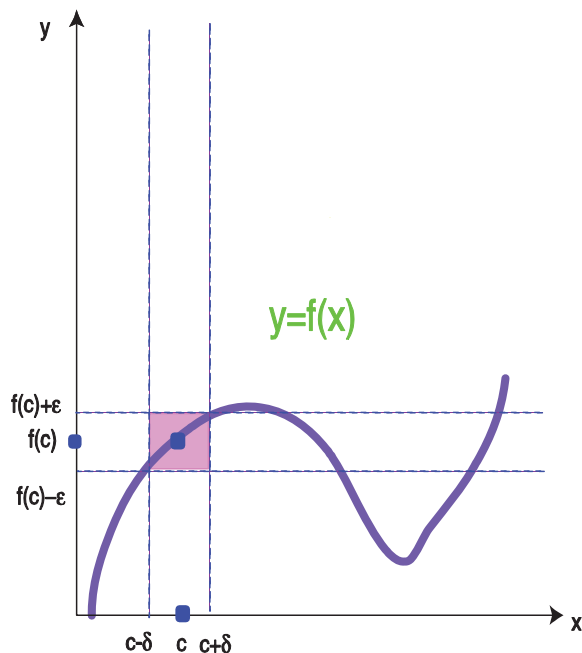


Figure 60: Picture proof that a non-negative continuous function $f(x)$ on an interval $[a, b]$ such that $f(c) > 0$ at some point c must be positive in a subinterval $(c - \delta, c + \delta)$. Here take $\varepsilon \leq f(c)/2$ and the corresponding δ with the graph of $y = f(x)$ above the interval $(c - \delta, c + \delta)$ in the pink box.

and

$$F(x|m) = \int_0^x (1 - m \sin^2 t)^{-\frac{1}{2}} dt = \text{the elliptic integral of the 1st kind}$$

cannot be expressed in terms of a finite number of elementary functions. A reference is: D.G. Mead, *American Math. Monthly*, 68 (1961), 152-156. The error function comes up in statistics and in the solution of various partial differential equations. The elliptic integrals come up in various sorts of applied math. problems as well as in number theory.

Mathematica, Matlab, Maple, etc. attempt to do all "doable" integrals. But of course these programs have to recognize that some integrals simply are not doable. See Wolfram's book to accompany Mathematica for some discussion of this problem.

Leibniz search for closed form expressions of integrals. J. Stillwell, *Math. and its History*, p. 110, says: "The search for closed forms was a wild goose chase but, like many efforts to solve intractable problems, it led to worthwhile results in other directions. Attempts to integrate rational functions raised the problem of factorization of polynomials and led ultimately to the fundamental theorem of algebra Attempts to integrate $(1 - x^4)^{-\frac{1}{2}}$ led to the theory of elliptic functions.... the problem of deciding which algebraic functions may be integrated in closed form has been solved only recently, though not in a form suitable for calculus textbooks, which continue to remain oblivious to most of the developments since Leibniz." See J. H. Davenport, *On the Integration of Algebraic Functions, Lecture Notes in Computer Science*, Springer-Verlag, 1981.

Unlike Leibniz, Newton evaluated integrals by expanding functions in power series and integrating the power series term-by-term. Isaac Newton offered his works on calculus to the British Royal Society and Cambridge University Press but the works were rejected. Thus Leibniz managed to publish the first paper on calculus. As Stillwell notes, loc. cit., p. 109: "This led to Leibniz's initially receiving credit for the calculus and later to a bitter dispute with Newton and his followers over the question of priority for the discovery." Stillwell finds (p. 110): "One thing has changed: it is now much easier to publish a calculus book than it was for Newton."

27 Taylor's Formula with Remainder

Next we consider the Taylor formula which is a finite version of Taylor series. In 1715 Brook Taylor (1685-1731) was the first person to publish the formula.

Theorem 53 Taylor's Formula with Remainder.

Suppose that J is an interval and $f : J \rightarrow \mathbb{R}$ has n continuous derivatives on the interval. If $a, x \in J$, we have

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (x-a)^k + R_n,$$

where the remainder is

$$R_n = \frac{1}{(n-1)!} \int_a^x f^{(n)}(t) (x-t)^{n-1} dt.$$

Proof. Induction on n .

Start with the fundamental theorem of calculus. This says

$$f(x) - f(a) = \int_a^x f'(t) dt,$$

which is the case $n = 1$ of Taylor's formula.

For the induction step, assume the n th formula and deduce the $(n+1)$ st. Start with R_n , the n th formula for the remainder, and use integration by parts to find

$$R_n = \frac{f^{(n)}(a)}{n!} (x-a)^n + R_{n+1}.$$

Let's do the details. Integration by parts says $\int u dv = uv - \int v du$. So let $u = f^{(n)}(t)$ and $dv = (x-t)^{n-1} dt$ to obtain $du = f^{(n+1)}(t) dt$ and $v = \frac{-1}{n} (x-t)^n$ and thus:

$$\begin{aligned} R_n &= \frac{1}{(n-1)!} \int_a^x f^{(n)}(t) (x-t)^{n-1} dt \\ &= \frac{1}{(n-1)!} \left\{ \frac{-1}{n} (x-t)^n f^{(n)}(t) \Big|_a^x + \frac{1}{n} \int_a^x (x-t)^n f^{(n+1)}(t) dt \right\} \\ &= \frac{1}{n!} f^{(n)}(a) (x-a)^n + R_{n+1}. \end{aligned}$$

This completes the proof of the induction step. ■

It is much easier to remember the alternate formula for the remainder. It looks much like the next term in the Taylor series.

Corollary 54 Taylor's Formula with Alternate Remainder. Under the same hypotheses as for Taylor's formula, we can find a point c between a and x such that the remainder in Taylor's formula has the form:

$$R_n = \frac{1}{n!} f^{(n)}(c) (x-a)^n.$$

Proof. Case 1. $a < x$.

We are assuming that $f^{(n)}$ is continuous on the interval $[a, x]$. So by the Weierstrass theorem on maxima and minima we have constants m and M such that

$$m \leq f^{(n)}(x) \leq M \quad \text{for all } x \in [a, b].$$

In this case, since integrals (in the right direction) preserve inequalities, we have

$$\frac{m}{n!}(x-a)^n \leq \frac{m}{(n-1)!} \int_a^x (x-t)^{n-1} dt \quad (35)$$

$$\leq R_n = \frac{1}{(n-1)!} \int_a^x f^{(n)}(t)(x-t)^{n-1} dt \quad (36)$$

$$\leq \frac{M}{(n-1)!} \int_a^x (x-t)^{n-1} dt = \frac{M}{n!}(x-a)^n. \quad (37)$$

Therefore

$$m \leq \frac{R_n n!}{(x-a)^n} \leq M. \quad (38)$$

By the intermediate value theorem for $f^{(n)}$, we know there is a point c in $[a, x]$ such that

$$\frac{R_n n!}{(x-a)^n} = f^{(n)}(c).$$

Solve for R_n to obtain the desired formula.

Case 2. $x < a$.

We leave the details of this case to the reader as an **exercise**. Note that $\int_a^x = -\int_x^a$ and that for n even, $-(x-t)^{n-1} \geq 0$

for $t \in [x, a]$. The inequality is reversed if n is odd. You have to take account of that to get formula (35). But in the latter case another sign switch will occur when you derive formula (38). The joy of inequalities. ■

28 Examples of Taylor's Formula

Example 1. The Taylor Formula Remainder need not always approach 0 as $n \rightarrow \infty$.

Define

$$f(x) = \begin{cases} e^{-1/x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

This function has the property that $f^{(n)}(x)$ exists for all values of x . Moreover it can be shown that $f^{(n)}(0) = 0$ for all $n = 0, 1, 2, 3, \dots$. See the exercises. So the Taylor formula says

$$\begin{aligned} f(x) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (x-a)^k + R_n \\ &= 0 + R_n. \end{aligned}$$

Since $f(x) = e^{-1/x} > 0$ if $x > 0$, we see that if we let $n \rightarrow \infty$, the Taylor series does not represent the function for $x > 0$. Thus, for positive x , the remainder R_n cannot approach 0 as $n \rightarrow \infty$.

We say that such a function is "not analytic". Weierstrass defined analytic functions around 1870 and found many applications. His student Sonya Kovalevsky (alias Sofia Kovalevskaya) wrote a thesis about power series (in 2 or more variables) solutions of partial differential equations.

Example 2. The Exponential.

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

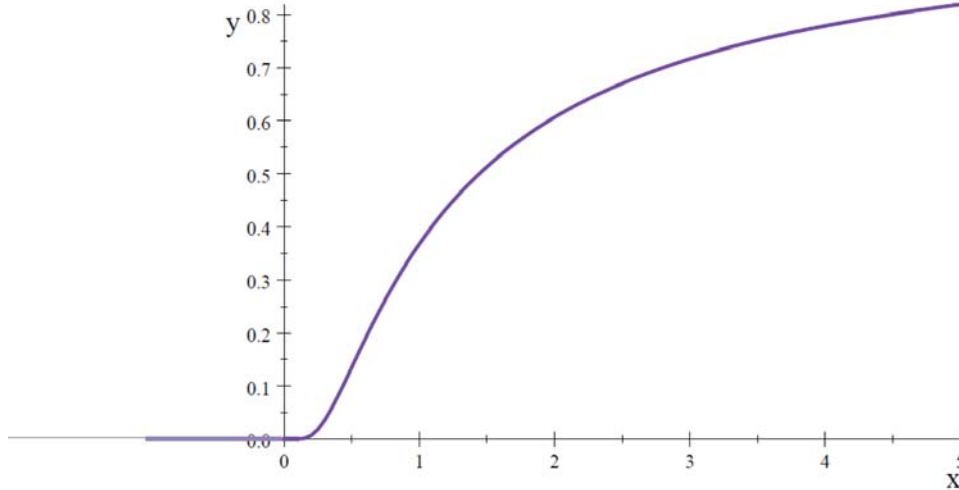


Figure 61: The function $f(x) = \begin{cases} e^{-1/x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$

The ratio test (which we consider later) shows that the remainder (2nd form) goes to 0 as $n \rightarrow \infty$. **Stirling's formula** for $n!$ would make this even clearer as it says

$$n! \sim \sqrt{2\pi n} n^n e^{-n}, \text{ as } n \rightarrow \infty, \text{ meaning that } \lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1.$$

Read " \sim " as "is asymptotic to." A proof of Stirling's formula can be found in Lang, *Undergraduate Analysis*.

As we noted earlier $e^{ix} = \cos x + i \sin x$, where $i = \sqrt{-1}$. So you can get the Taylor series for $\cos x$ and $\sin x$ out of that for e^{ix} .

Example 3. Geometric Series.

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, \text{ for } |x| < 1.$$

Example 4. The Logarithm.

$$-\log(1-x) = \sum_{n=1}^{\infty} \frac{x^n}{n}, \text{ for } |x| < 1.$$

Example 5. The Binomial Series.

$$(1+x)^s = \sum_{n=0}^{\infty} \binom{s}{n} x^n, \text{ for } |x| < 1.$$

Here the **generalized binomial coefficient** is

$$\binom{s}{n} = \frac{s(s-1)\cdots(s-n+1)}{n(n-1)\cdots 2 \cdot 1} = \frac{\Gamma(s+1)}{\Gamma(s)\Gamma(n+1)},$$

where $\Gamma(s)$ is the gamma function from the introduction.

You can plug an integer value of s into the binomial series and obtain the binomial theorem. In that case when n becomes bigger than s the coefficient is 0. When n is negative $\Gamma(n+1)$ is infinite and you have to use the first formula for the binomial coefficient. For example, when $s = -1$, you get the geometric series with x replaced by $-x$. When $s = \frac{1}{2}$, you get $\sqrt{1+x} = \exp\left\{\frac{1}{2} \log(1+x)\right\}$.

Example 6. Legendre Polynomials (Spherical Harmonics).

The Legendre polynomials can be defined by

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

This is called the Rodrigues formula for the Legendre polynomials. The first few are:

$$P_0(x) = 1, P_1(x) = x, P_2(x) = \frac{1}{2} (3x^2 - 1), P_3(x) = \frac{1}{2} (5x^3 - 3x) .$$

Mathematica knows them and will graph them quickly. The Mathematica command `Plot[Table[LegendreP[i, x], {i, 30}], {x, -1, 1}]` produced the following figure.

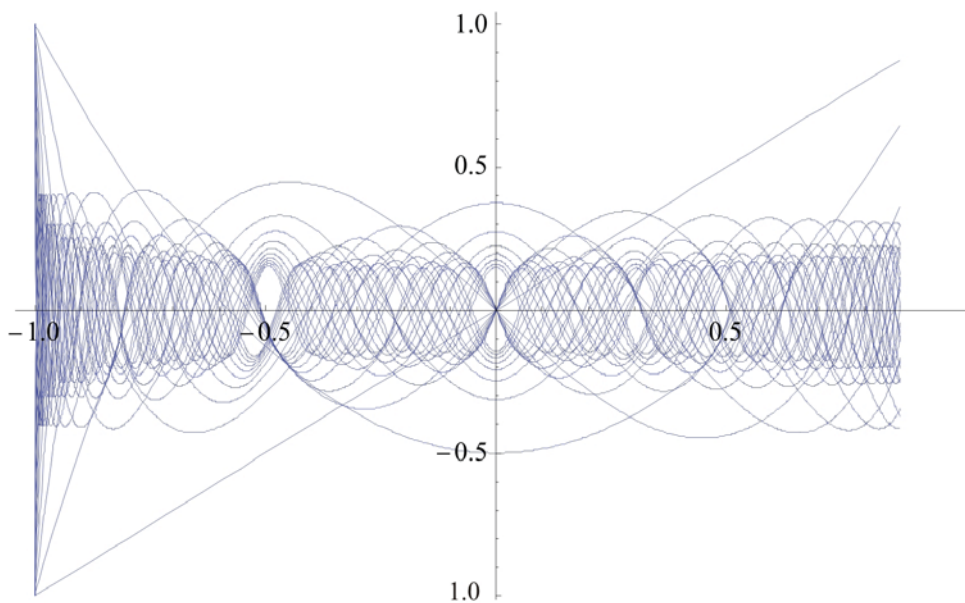


Figure 62: Graph of Legendre polynomials P_n for $n = 0, \dots, 30$.

There is a way of putting all the Legendre polynomials together in what is called a **generating function** (beloved of combinatorists):

$$(1 - 2xt + t^2)^{-\frac{1}{2}} = \sum_{n=0}^{\infty} P_n(x)t^n, \text{ for } |t| < 1 \text{ and } |x| \leq 1.$$

It is possible to use the generating function to derive properties of the Legendre polynomials quickly. For example, you can see that $P_n(1) = 1$.

References for Legendre polynomials include: N.N. Lebedev, *Special Functions*, S. Wolfram, *Mathematica*, R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. I*.

The Legendre polynomials arise in problems of applied math. which have spherical symmetry. They form a system of orthogonal polynomials on the interval $[-1, 1]$ with the **inner product**

$$\langle f, g \rangle = \int_{-1}^1 f(t)g(t)dt.$$

The Legendre polynomials are pairwise orthogonal for this inner product; i.e., $\langle P_n, P_m \rangle = 0$ if $n \neq m$. It is possible to expand arbitrary continuous functions on $[-1, 1]$ as series of Legendre polynomials (a generalized Fourier series) convergent in the sense of the norm coming from this inner product $\|f\| = \sqrt{\langle f, f \rangle}$. We will say much more about such things in the

last section of Lectures, II. The Legendre polynomials are important for obtaining explicit solutions of partial differential equations with spherical symmetry; e.g., Schrödinger's equation for the hydrogen atom.

Stillwell, *Mathematics and its History*, p. 107 says: "It is misleading ... to describe Newton as a founder of calculus unless one understands calculus, as he did, as an algebra of infinite series. In this calculus, differentiation and integration are carried out term by term on powers of x and hence are comparatively trivial."

Newton studied the binomial series and used it to obtain the power series for arcsine in 1669. Mercator found the power series for $\log(1-x)$ in 1668.

There is also

$$\arctan x = \sum_{n=0}^{\infty} (-1)^{n+1} \frac{x^{2n+1}}{2n+1}, \text{ if } |x| < 1.$$

By a theorem of Abel, this converges for $x = 1$ and gives the Leibniz formula:

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots$$

Of course there are many applications of power series. For example, in the theory of perturbations, small deviations from the state of a physical system, one plugs a power series into the differential equation describing the system and computes as many coefficients of the solution as possible, in order to deduce information about the system.

29 A Look at the History of the Integral

References: G. Temple, *100 Years of Mathematics*; J. Korevaar, *Mathematical Methods*.

In 1854 Riemann began studying integrals in order to understand Fourier series of periodic functions. He based his theory on Riemann sums $S(f, P)$ which approach the **Riemann integral** $\int_a^b f(x)dx$ as you refine the partition $P = \{a = x_0 < x_1 < \dots < x_n = b\}$ of the interval $[a, b]$ on the x -axis. Here, for $s_k \in [x_k, x_{k+1})$, the **Riemann sum** is

$$S(f, P) = \sum_{k=1}^n f(s_k)(x_k - x_{k-1}).$$

"Refining the partition" means that the lengths of all subintervals $[x_k, x_{k+1})$ get smaller. Assume that f is continuous or piecewise continuous on $[a, b]$. You may view this as approximating the function $f(x)$ by the step function with value $f(s_k)$ on the k th subinterval $[x_k, x_{k+1})$. See Figure 63.

In 1875 Darboux modified the Riemann sums by taking $f(s_k)$ to be either the max or min of $f(x)$ on the k th subinterval. He thus got upper and lower sums. For a function $f(x)$ to be Riemann integrable one wants the upper and lower sums to approach the same limit as the partition P is refined. The common limit of the upper and lower Darboux sums is called the

Riemann integral $\int_a^b f(x)dx$. This is fine for piecewise continuous functions.

Our Axioms 1 and 2 for the integral actually come out of upper and lower Darboux sums. We will not go this route however. Instead we will prove the existence of the integral for continuous functions (or piecewise continuous functions) by a method closer to that of Lebesgue (1902). Lebesgue's methods led to a better theory of Fourier series and to the modern theory of probability measures.

If you are only interested in the calculation of the sort of integrals that arise in calculus, then it does not matter what definition of integral you use. For a continuous function on a closed finite interval or a piecewise continuous function (i.e., a function that is mostly continuous but has a finite number of jump discontinuities) the Lebesgue integral equals the Riemann integral. However, the Lebesgue integral will integrate more functions than the Riemann integral. For a positive function on

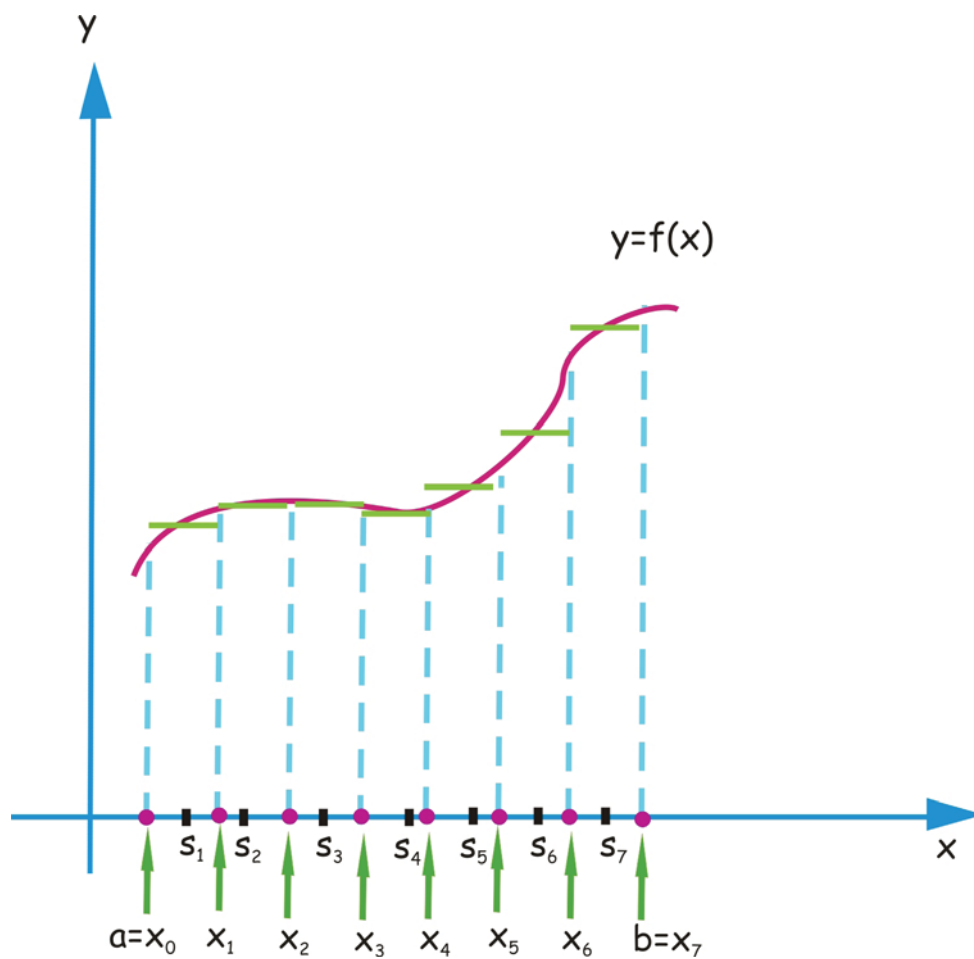


Figure 63: The Riemann sum is obtained by summing terms $f(s_k)(x_k - x_{k-1})$ over $k = 1, \dots, 7$. Since our function is non-negative, this means we add up the areas of the rectangles with dotted sides, green tops and bases on the x -axis.

an infinite interval (or a positive unbounded function) the Lebesgue integral is the same as the improper Riemann integral, when the latter exists.

For theoretical work, the Lebesgue integral is much nicer than the Riemann integral. For example, the interchange of integral and limit is legal in many more situations for the Lebesgue integral. The hypotheses of the Lebesgue dominated convergence theorem are quite weak in comparison to those that we will discuss later for the Riemann integral. Similarly the story of repeated integrals is much simpler for Lebesgue integrals (see the statement of Fubini's theorem).

There are at least 2 ways to approach the Lebesgue integral.

Lebesgue's Way. Define the Lebesgue measure $\mu(S)$ of a (measurable) subset $S \subset \mathbb{R}$ where $\mu[a, b] = b - a$. Suppose the bounded function f has range contained in the interval $[c, d]$ on the y -axis. Then partition the interval on the y -axis rather than the x -axis by partition $Q = \{c = y_0 < y_1 < \dots < y_n = d\}$. Thus obtain step functions approximating f involving sets $S_k = f^{-1}[y_{k-1}, y_k]$. Assume the function is such that all these sets are Lebesgue measurable. Let $\gamma_k \in [y_{k-1}, y_k]$. Then the step function $\phi(x)$ is defined to have the constant value γ_k on S_k , for $k = 1, \dots, n$. The **Lebesgue integral of step function** $\psi(x)$ is defined to be

$$\sum_{k=1}^n \gamma_k \mu(S_k).$$

See Figure 64.

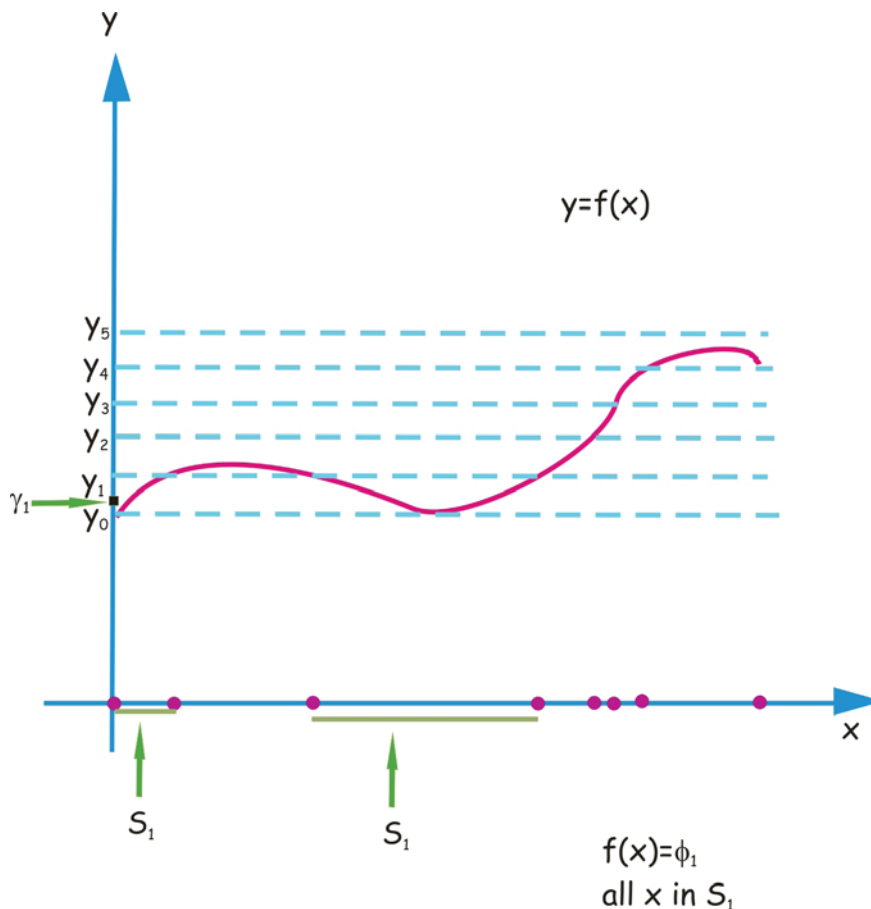


Figure 64: To get a Lebesgue integral create a step function by partitioning the y -axis. Then define the step function ϕ to be constant on the inverse image of the intervals on the y -axis. For example, $\phi(x) = \gamma_1$ for all $x \in f^{-1}[y_0, y_1] = S_1$.

The New-Fangled Way. Use Cauchy sequences of step functions with respect to the L^1 -norm on the space of nice

functions on $[a, b]$

$$\|f\|_1 = \int_a^b |f(x)| dx.$$

In order to do this, we must think about normed vector spaces (infinite dimensional) of functions. The method is the same as that used to create the real numbers as limits of Cauchy sequences of rational numbers.

The Lebesgue measure μ of a set S of real numbers starts out with the idea that an interval $I = [a, b]$ or (a, b) or $[a, b)$ or $(a, b]$ has measure $\mu(I) = b - a$. Measurable sets are created by taking countable unions, complements, countable intersections. Then the measure is assumed to have certain properties on measurable sets. In particular, the Lebesgue should be non-negative and countably additive meaning that if one has a infinite sequence S_n of pairwise disjoint ($n \neq m \implies S_n \cap S_m = \emptyset$) measurable sets then

$$\mu\left(\bigcup_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} \mu(S_n).$$

We will not say more about the Lebesgue integral or measure here.

Exercises I

Exercises 1

Draw a picture to illustrate each problem, if possible.

- Let A, B, C be sets.
 - Show that $(A \cap B) \cap C = A \cap (B \cap C)$.
 - Show that if \emptyset denotes the empty set, then $A \cup \emptyset = A$.
 - Let $A \subset B$ and $B \subset C$. Show that if we denote the complement of B in C by $B^c = \{x \in C \mid x \notin B\}$ and the complement of A in C by A^c , then $B^c \subset A^c$.
 - Show that $(A \cap B)^c = A^c \cup B^c$.
- Define the function $L: [0, 1] \rightarrow [0, 1]$ by $L(x) = \frac{1}{2}x(1-x)$.
 - Show that L is not 1-1.
 - Show that L is not onto.
 - Compute $(L \circ L)(x) = L(L(x))$.
- Suppose that a function $f: A \rightarrow B$ is 1-1 and onto.
 - Define the inverse function $f^{-1}: B \rightarrow A$.
 - Show that $f = (f^{-1})^{-1}$.
 - Look at the example with $A = B = [0, 1]$ and $f(x) = x^2$. Show that this function is 1-1 and onto. What is the inverse function f^{-1} ?
- Suppose A is a finite set and $|A|$ = the number of elements in A .
 - Show that if $f: A \rightarrow B$ is 1-1, then $|A| \leq |B|$.
 - Show that if $f: A \rightarrow B$ is onto, then $|B| \leq |A|$.
- Use mathematical induction to prove that if $n! = n \cdot (n-1) \cdots 2 \cdot 1$, then $2^n < n!$, for all $n = 4, 5, 6, 7, \dots$

Exercises 2

- Explain why the set \mathbb{Q} of rational numbers is denumerable.
 - Suppose that $g: S \rightarrow T$ is 1-1 and onto. Show that S is denumerable if and only if T is denumerable.
- Which of the following sets is denumerable:
 $\{2^n \mid n = 1, 2, 3, \dots\}$ or the interval $[-2, 1]$?
Explain your answer.
- Suppose that a is a real number and n, m are positive integers. Prove that $a^{n+m} = a^n \cdot a^m$.
 - Prove the formula in part a) if both n and m are negative integers.
 - Again suppose that a is real and n, m are positive integers. Show that $(a^n)^m = a^{n \cdot m}$.
- Suppose that a is a positive real number. Show that there is a positive integer n such that $\frac{1}{n} < a$.
 - Show that $\sqrt{7}$ is irrational.
- Use the 2 order axioms ORD1 and ORD2 listed on p. 22 of the lectures plus the definition of $a < b$ to deduce that $x < y$ implies $x+z < y+z$ (which is Fact 3 in our list of facts about order).
 - Using the properties of inequalities (see our list of facts about order) and the definition of absolute value, find the set of real numbers x such that $|x^2 - 1| < \frac{1}{4}$. Draw a picture of the set.
- Prove that $||x| - |y|| \leq |x - y|$ for all real numbers x and y .

Exercises 3

- 1) Tell whether the sequences below have a limit. Find the limit, if possible.
Prove that your answer is correct.
- a) $1+(-1)^n$ b) $n!/n^n$ c) $(1-n)/(1+n)$
- 2) Assume that $\{x_n\}$ is a sequence of real numbers. Show that if $a = \lim_{n \rightarrow \infty} x_n$ exists, then the set $\{x_1, x_2, x_3, \dots\}$ is bounded (both above and below).
- 3) Assume that $\{x_n\}$ and $\{y_n\}$ are sequences of real numbers. Show that if $a = \lim_{n \rightarrow \infty} x_n$ exists and $b = \lim_{n \rightarrow \infty} y_n$ exists, then $ab = \lim_{n \rightarrow \infty} x_n y_n$ exists.
- 4) **True-False.** State whether the following are true or false. Give a brief reason for your answer (such as a reference to some fact in these notes or a textbook).
- a) Every bounded sequence of real numbers is convergent to a real number.
b) Every bounded sequence of real numbers is Cauchy.
c) Suppose that $a_n \leq b_n \leq c_n$ for $n \geq n_0$. If the limit $L = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n$ exists and is the same for both $\{a_n\}$ and $\{c_n\}$, then $\{b_n\}$ has a limit which is also the same; i.e. $L = \lim_{n \rightarrow \infty} b_n$.
- 5) a) Define $\infty = \lim_{n \rightarrow \infty} a_n$.
b) Suppose that $\{a_n\}$ is an unbounded sequence of non-negative real numbers. Show that $\{a_n\}$ has a subsequence converging to ∞ according to your definition in part a).

Exercises 4

- 1) Find the following limits. Then explain your answer using the $\epsilon\delta$ definition of limit. This means given ϵ , find δ as a function of ϵ .
- a) $\lim_{x \rightarrow 0} (5x-1)$ b) $\lim_{x \rightarrow -1} x^2$ c) $\lim_{x \rightarrow 0} \sqrt{x}$.
 $x > 0$
- 2) **True-False.** Tell whether the following statements are true or false. Give a brief reason for your answer.
- a) $\lim_{x \rightarrow a} f(x) = L$ implies $\lim_{h \rightarrow 0} f(a+h) = L$.
b) Suppose $f: [a, \infty) \rightarrow \mathbb{R}$. Then $\lim_{x \rightarrow a} f(x) = L$ implies $L = f(a)$.
 $x > a$
- Hint: Draw a picture of the graph of $y=f(x)$ near $x=a$.
- 3) Prove or disprove:
- a) $\lim_{x \rightarrow 3a} f(x) = 3 \lim_{x \rightarrow a} f(x)$ b) $\lim_{x \rightarrow 3a} f(x) = \lim_{x \rightarrow a} f(3x)$.
- 4) Prove parts 1,2,3 in the Proposition about limits.

5) Prove that if $f:(0,\infty)\rightarrow\mathbb{R}$, then $\lim_{x\rightarrow\infty} f(x) = L \Leftrightarrow \lim_{x\rightarrow 0} f\left(\frac{1}{x}\right) = L$.

$x > 0$

Exercises 5

1) Determine where the following functions are continuous. Explain your answers briefly. Draw the graphs.

a) $f(x) = \lfloor x \rfloor =$ floor of $x =$ greatest integer $\leq x$.

b) $f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{otherwise.} \end{cases}$

c) $f: [-1,1] \rightarrow [-1,1]$ defined by setting $f(0)=0$ and then connecting the point $\left(\frac{1}{2k}, 0\right)$ to two points

$\left(\frac{1}{2k+1}, \frac{1}{2k+1}\right)$ and $\left(\frac{1}{2k-1}, \frac{1}{2k-1}\right)$ for all integers $k \in \mathbb{Z}$ with $k \neq 0$. So the function consists

of infinitely many line segments. This example implies that it is wrong to say that continuity means being able to draw the graph without lifting the pen from the paper. You cannot really draw this graph.

2) a) Use the Intermediate Value Theorem to show that any polynomial of odd degree has a real root. Why doesn't this work for polynomials of even degree?

b) Suppose that $f:[a,b] \rightarrow [a,b]$ is continuous. Show that there is an $x \in [a,b]$ such that $f(x)=x$. The point x is called a fixed point of f . Hint: Look at $g(x)=f(x)-x$.

3) **True-False.** Tell whether the following statements are true or false. Give a brief reason for your answer.

a) Every function $f:[0,1] \rightarrow \mathbb{R}$ has a maximum value.

b) Every continuous function $f:[0,1] \rightarrow \mathbb{R}$ has a minimum value.

c) Every continuous function $f:(0,1) \rightarrow \mathbb{R}$ has a maximum value.

4) Suppose that $f:[a,b] \rightarrow \mathbb{R}$ is continuous and strictly increasing; i.e., $x < y$, for x,y in $[a,b]$ implies $f(x) < f(y)$. Show that if $c=f(a)$ and $d=f(b)$, the inverse function $f^{-1}: [c,d] \rightarrow \mathbb{R}$ exists and is continuous and strictly increasing as well. Draw a picture.

$\infty \rightarrow \infty$

Exercises 6

$\infty \rightarrow \infty$

1) a) Assuming the derivative of e^x is e^x , find the derivative of $\log(x) = \log_e(x) = \ln(x)$ using the theorem about derivatives of inverse functions. (In this course, e is the only base we use.)

b) Compute the derivative of the following function using properties of derivatives such as the chain rule: $f(x) = x^{1/x}$. Hint: Recall that the definition of x^a for $x > 0$ is $x^a = e^{a \log x}$.

2) Assume that f and g are functions on the open interval (a,b) . Assume that both f and g are differentiable at $x \in (a,b)$. Suppose c is a real constant. Prove using the definition of derivative that:

a) $(f+g)'(x) = f'(x) + g'(x)$. b) $(cf)' = cf'$ if c is a constant.

c) Using mathematical induction, prove the formula for the derivative of x^n , for $n=1,2,3,\dots$.

3) a) Define $f(x) = x \sin(1/x)$ if $x \neq 0$ and $f(0)=0$. Show that $f(x)$ is not differentiable at $x=0$ but

$f(x)$ is continuous at $x=0$.

- b) Discuss the differentiability of the floor function $[x]=\lfloor x \rfloor = \text{the largest integer } \leq x$, for all real numbers x . Does the function have right- and left- hand derivatives at some points? Are they equal?
- 4) a) Suppose that a continuous function f on $[a,b]$ is differentiable on (a,b) and that the derivative $f'(x)$ is bounded (above and below) on (a,b) . Prove that then f is uniformly continuous on $[a,b]$.
Hint. Use the mean value theorem.
- b) **First Derivative Test.** Consider the function $f(x)=x^{1/x}$ from problem 1b). Find all relative and absolute extrema for this function on $(0,\infty)$. Also find the glb and lub of $\{f(x)|x\}$, if possible. Sketch the function.

Exercises 7

- 1) Prove the **Generalized Mean Value Theorem also known as Cauchy's Mean Value Theorem**. This says the following. Assume that f and g are differentiable on (a,b) and continuous on $[a,b]$. Show there exists a point $c \in (a,b)$ such that
- $$[f(b)-f(a)]g'(c)=[g(b)-g(a)]f'(c).$$
- Hint. As in the proof of the mean value theorem, consider a function to which you can apply Rolle's theorem. Try the function: $h(x)=[f(b)-f(a)][g(x)-g(a)]-[g(b)-g(a)][f(x)-f(a)]$.

- 2) Prove **l'Hôpital's rule**. This says the following. Assume that f and g are differentiable on an open interval (a,c) . Suppose also that $g(x)$ and $g'(x)$ do not vanish in (a,c) . Finally assume that both $f(x)$ and $g(x)$ approach 0 as x goes to c , with $x < c$. Then l'Hôpital's rule says:

$$\text{If } \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)} = k, \text{ then } \lim_{x \rightarrow c} \frac{f(x)}{g(x)} = k.$$

$x < c$ $x < c$

- 3) a) Define $g(x) = \begin{cases} e^{-1/(x-1)}, & \text{for } x > 1 \\ 0, & \text{for } x \leq 1 \end{cases}$. Prove that $g(x)$ is continuous at $x=1$.

- b) Show that the function $g(x)$ is differentiable everywhere. You have to consider 3 cases: $x < 1$, $x = 1$, and $x > 1$.

- 4) a) Show that the function $g(x)$ from problem 3 has a second derivative $g''(x)$ for all x . Again there are 3 cases.
- b) Using mathematical induction, consider the n th derivative $g^{(n)}(x)$ for the function $g(x)$ of problem 1. Show $g^{(n)}(x)$ exists for all $n \in \mathbb{Z}^+$ and all real numbers x . Again there are 3 cases.

Hint: the induction hypothesis should be something like the following:

$$g^{(n)}(x)=0, \text{ for } x \leq 1 \text{ and } g^{(n)}(x) = P_n \left(\frac{1}{x-1} \right) e^{-1/(x-1)}, \text{ for } x > 1,$$

where $P_n(u)$ is a polynomial.

- 5) Sketch a graph of the function $g(x)$ from problem 3. What is the Taylor series for $g(x)$ centered at $x=1$? Does it represent $g(x)$?

Exercises 8

1) Prove that $\frac{\pi}{6} \leq \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} \sin x \, dx \leq \frac{\pi}{3}$ without evaluating the integral.

2) Assume that $a < b < c$. Note that if a function $f(x)$ is continuous on $[a, b]$ and also continuous on $[b, c]$, but has a jump discontinuity at $x = b$, then we can extend the integral to say f is integrable on $[a, c]$ and

$$\int_a^c f(t) dt = \int_a^b f(t) dt + \int_b^c f(t) dt.$$

Show that this integral still satisfies the our axioms I and II.

3) Define $\text{sign}(x) = 1$ if $x > 0$, $\text{sign}(x) = -1$ if $x < 0$, and $\text{sign}(0) = 0$. Show that for any $a < b$, we have (using the preceding problem):

$$\int_a^b \text{sign}(x) dx = |b| - |a|. \quad \text{Why is } |x| \text{ not an antiderivative of } \text{sign}(x)?$$

4) **The Mean Value Theorem for Integrals.** Suppose $f: [a, b] \rightarrow \mathbb{R}$ is continuous. Show that then there is a

point c in $[a, b]$ such that $\int_a^b f(t) dt = f(c)(b - a)$.

Hint. First note that f has a maximum M and minimum m on $[a, b]$. Why? Then use the axioms for

integrals to see that $m(b - a) \leq \int_a^b f(t) dt \leq M(b - a)$. Then $\frac{1}{b - a} \int_a^b f(t) dt$ is in the interval $[m, M]$.

Use the intermediate value theorem to finish the problem.

5) Compute $\frac{d}{dx} \left(\int_{-x}^x e^{t^2} dt \right)$.

6) What is wrong with the formula $\int_{-1}^1 \frac{1}{t^2} dt = \left. \frac{-1}{t} \right|_{-1}^1 = -2$?

7) Write out Taylor's formula with remainder as in Lang p. 109 with $a = 0$ and $n = 6$ for the function $f(x) = \log(1 - x)$.

8) **Cauchy-Schwarz Inequality.**

Suppose that f and g are continuous on $[a, b]$. Show that $\left| \int_a^b f(t)g(t) dt \right|^2 \leq \int_a^b f(t)^2 dt \int_a^b g(t)^2 dt$.

Hint: Let x be a fixed real number and look at $\left| \int_a^b (f(t) - xg(t))^2 dt \right|^2 = Ax^2 + Bx + C$. This is always non-negative. What does that say about $B^2 - 4AC$?

9) Show using the fundamental theorem of calculus that, assuming the derivatives f', g' are continuous on an open interval containing $[a, b]$, then for all x in $[a, b]$, we have:

$$\int_a^x f'(t)g(t) dt = f(x)g(x) - f(a)g(a) - \int_a^x f(t)g'(t) dt.$$

I, Practice Exam 1

1) Define and **give an example**:

a) function $f:S \rightarrow T$; 1-1 function; onto function; b) denumerable set; c) lub; d) $\lim_{n \rightarrow \infty} x_n$;

e) If $f:(c,d) \rightarrow \mathbb{R}$ and $a \in (c,d)$, define the 2-sided limit $\lim_{x \rightarrow a} f(x)$;

$$x \neq a, x \in (c,d)$$

f) Cauchy sequence; g) subsequence.

2) a) State the well ordering axiom for \mathbb{Z}^+ .

b) Define the absolute value $|x|$ for a real number x and state its 3 basic properties.

c) State the completeness axiom for the real numbers.

d) State the negation of definition of the 2-sided limit $\lim_{x \rightarrow a} f(x)$ for the function $f:(c,d) \rightarrow \mathbb{R}$

$$x \in (c,d), x \neq a$$

with $a \in (c,d)$. Hint: You need recall how to negate a statement with the quantifiers \forall, \exists .

3) a) Prove that an increasing sequence of real numbers which is bounded above has a limit.

b) Show that a Cauchy sequence is bounded.

c) Show that if a sequence $\{x_n\}$ is Cauchy and it has a convergent subsequence, then $\{x_n\}$ converges to the same limit as the subsequence.

d) Prove that the set \mathbb{Z}^+ does not have an upper bound in \mathbb{R} .

e) Suppose that $\lim_{n \rightarrow \infty} x_n = L$ and $\lim_{n \rightarrow \infty} y_n = M$. Prove that $\lim_{n \rightarrow \infty} x_n y_n = LM$.

4) a) Prove that $\sqrt{5}$ is irrational.

b) Show that \mathbb{Q} is denumerable.

c) Show that \mathbb{R} is not denumerable.

d) State and prove the formula for the number of functions $f:S \rightarrow T$, if S and T are finite sets.

e) Prove by induction that for each positive integer n and for any real x with $x \geq -1$, we have

$$(1+x)^n \geq 1+nx.$$

This is (Jacob) Bernoulli's inequality. Where did you use $x \geq -1$? Is it really necessary?

5) Define a sequence inductively as follows: $a_1 = 1, a_{n+1} = \sqrt{1+a_n}$. Show that

$L = \lim_{n \rightarrow \infty} a_n$ exists and find a formula for the limit. Answer: The limit is the golden ratio:

$$L = \frac{1+\sqrt{5}}{2}.$$

Hint: First show that $0 < a_n < 2$ for all n , by induction. Then show that the sequence is strictly increasing.

Then note that L must satisfy $L^2 = 1+L$. Why?

6) State whether the following sequences have limits. If they do, find the limit and prove that the sequence approaches that limit using the definition of limit. If they don't, explain why they don't.

a) $\lim_{n \rightarrow \infty} \sin(n\pi)$; b) $\lim_{n \rightarrow \infty} \frac{n}{n^2 + 1}$; c) $\lim_{n \rightarrow \infty} \frac{n^2 - n}{n^2 + 1}$; d) $\lim_{n \rightarrow \infty} \frac{1}{2^n}$.

7) **True-False.**

Tell whether the following statements are true or false. Give a brief reason for your answer.

- a) The only real number a which satisfies the inequality $|a| < \epsilon$, for every $\epsilon > 0$, is the number $a=0$.
- b) The sum of two irrational numbers is irrational.
- c) For all real numbers x,y we have $|x-y| \leq |x| + |y|$.
- d) If $\lfloor x \rfloor$ denotes the largest integer less than or equal to x , then $\lim_{x \rightarrow 1} \lfloor x \rfloor = 1$.
- e) Any decreasing sequence of real numbers has a limit.
- f) If $f:S \rightarrow T$ and A is a subset of set T , define the inverse image $f^{-1}(A) = \{x | x \in S \text{ and } f(x) \in A\}$. Then $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$ for any subsets A and B of the set T . We do not assume here that f is 1-1 and onto.
- g) If $f:S \rightarrow T$ is 1-1 and onto then there is an inverse function $g:T \rightarrow S$ such that the composition $f \circ g = \text{identity on } T$ and $g \circ f = \text{identity on } S$.
- h) If $x > 0$ and y is any real number, there is a positive integer n such that $nx > y$.
- i) A sequence of real numbers can have 2 different limits.

8. Find the lub and glb for the following sets, if possible.

- a) $\left\{ \frac{m}{2m+1} \mid m \in \mathbb{Z}^+ \right\}$;
- b) $\{x \in \mathbb{Q} \mid x > 0 \text{ and } x^2 < 5\}$;
- c) the positive integers
- d) the open interval $(0,3)$.

9) Find the following limits, if possible. Then explain your answer using the definition of limit.

- a) $\lim_{x \rightarrow 0} \sqrt{x}$;
 $x > 0$
- b) Define $f(x)=0$ if x is irrational and $f(x)=1$ if x is rational. Find $\lim_{x \rightarrow a} f(x)$ for any rational number a .
 $x \neq a$



I, Practice Exam 2

- 1)
 - a) Define and give an example of a function $f(x)$ which is continuous at $x=c$.
 - b) Give 2 definitions for the derivative $f'(c)$. Draw sketches to illustrate the meaning of both defns.
 - c) Define a^x for $a>0$ and explain why this generalizes a^n for $n \in \mathbb{Z}^+$.
 - d) State the 2 axioms for the integral of a continuous function on $[a,b]$.

- 2) State and prove:
 - a) mean value theorem
 - b) fundamental theorem of calculus
 - c) Weierstrass theorem on the existence of maxima and minima
 - d) intermediate value theorem

- 3) Prove the following properties of the derivative
 - a) chain rule
 - b) linearity
 - c) product rule

- 4) Prove the following properties of the integral of a continuous function:
 - a) $\int_a^b f'(x)dx = f(b) - f(a)$, assuming f' continuous on $[a,b]$.
 - b) linearity
 - c) integration by parts
 - d) substitution formula
 - e) integrals preserve \leq

- 5) **True-False.**
Tell whether the following statements are true or false. If true, give a brief reason for your answer. If false, give a counterexample.
 - a) Suppose $f:(a,b) \rightarrow \mathbb{R}$ is continuous. Then there is a point c in (a,b) such that $f(x) \leq f(c)$ for all $x \in (a,b)$.
 - b) $\int_a^b fg = f \int_a^b g$ for functions $f,g:[a,b] \rightarrow \mathbb{R}$ continuous.
 - c) Suppose $f:[a,b] \rightarrow \mathbb{R}$ is continuous and f is differentiable on (a,b) with $f'(x) = 0$ for all $x \in (a,b)$. Then $f(x) = \text{constant}$, for all $x \in [a,b]$.
 - d) Recall that a function $f(x)$ defined for $x \in \mathbb{R}$ is even iff $f(-x) = f(x)$ for all $x \in \mathbb{R}$. We say $f(x)$ is odd iff $f(-x) = -f(x)$, for all $x \in \mathbb{R}$. Then the derivative $f'(x)$ of an even function $f(x)$ is an odd function.
 - e) Suppose that $x=g(y)$ is the inverse function to $y=f(x)$. Then $g'(y) = 1/f'(x)$.
 - f) f,g continuous on $[a,b]$ implies $\max\{f,g\}$ continuous on $[a,b]$.
 - g) f continuous at $x=c$ implies f differentiable at $x=c$.

- 6) a) Compute $g'(21)$ when $x=g(y)$ is the inverse function for $y=f(x)=2x^3+1$.
 b) Find the set of points where the following function $f(x)$ is continuous:

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

c) Compute $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$.

- d) Suppose that f is a differentiable function on the whole real line such that $f'(x)=2xf(x)$ for all $x \in \mathbb{R}$.

Show that there is a constant C so that $f(x) = Ce^{x^2}$ for all x .

e) Compute $\frac{d}{dt} \int_0^t g(x-t) dx$.

- 7) a) Suppose that $f'(x) > 0$ for all x in (a,b) . Why does $y=f(x)$ have an inverse function $x=g(y)$? What is the formula for the derivative of the inverse function? Prove it.
 b) State Taylor's formula with remainder (2nd version). Apply it to get the formula for $\log(1-x)$ using the first 3 terms of the Taylor series plus remainder.
 c) Define e^x by its Taylor series and prove that $e^u e^v = e^{u+v}$.
 d) Define $\log(y)$ as the inverse function to e^x and show that $\log(uv) = \log u + \log v$.

8) Define a function $f(x) = \begin{cases} e^{-1/x}, & \text{for } x > 0 \\ 0, & \text{for } x \leq 0 \end{cases}$. Show that $f^{(n)}(0)$ exists for all $n \in \mathbb{Z}^+$. Sketch this

function. Is $f(x)$ represented by its Taylor series around the point 0? Why?

- 9) Show that the following function gives an example of a continuous function whose graph cannot be drawn **without lifting pen from paper**. Define $f(x)$ inductively as a map from $[-1,1]$ to $[-1,1]$ consisting of straight line segments connecting the point $(1/(2k),0)$ to the 2 points $(1/(2k+1),1/(2k+1))$ and $(1/(2k-1),1/(2k-1))$ for $k=\pm 1, \pm 2, \pm 3, \dots$ Sketch the function.



I, Discussion of Final, Fractals:

from Wikipedia: list of fractals by Hausdoff dimension

Sierpinski Triangle



3D Cantor Dust



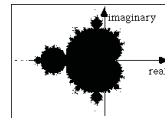
Lorenz attractor



Coastline of Great Britain



Mandelbrot Set



What makes a fractal?

I'm using 2 references:

Fractal Geometry by Kenneth Falconer
Encounters with Chaos by Denny Gulick



- 1) A fractal is a subset of \mathbb{R}^n with non integer dimension.
Of course this make no sense without a definition of dimension.
- 2) Fractal contains copies of itself at many scales.
- 3) It is too irregular to be described by traditional language.

Mandelbrot coined the term in 1977 though many examples were known.
 "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line."(Mandelbrot, 1983).

Other references: the web, books by Mandelbrot
 M. Barnsley, *Fractals Everywhere*

1) Cantor Dust.

From the interval $[0,1]$ remove the middle third. Then remove the middle third from the remaining 2 intervals $[0,1/3]$ and $[2/3,1]$.

Keep going with this removal of middle thirds forever

You end up with the Cantor dust. Impossible to draw it. We give the first 5 steps. Later we will see the Cantor Dust has box dimension $\ln 2 / \ln 3 \cong .63$.

Higher dimension than a point but smaller than an interval. There are many interesting facts about the Cantor dust. For example, the set is uncountable, but if you integrate the function that is 1 on the Cantor set and 0 off the set (using the Lebesgue integral), you get 0. The Riemann integral cannot deal with this.



Defn. Start with $C_0=[0,1]$. Let $C_1=[0,1/3] \cup [2/3,1]$. Continue in this way, defining C_k as a union of 2^k subintervals, each of length 3^{-k} obtained by removing middle thirds of the intervals in C_{k-1} . The **Cantor set C** is the intersection of all the C_k , for k running over all integers ≥ 0 .

Problem 1. Show that the Cantor set can be identified with all the real numbers in $[0,1]$ that can be represented in the form

$$x = \sum_{n=1}^{\infty} \frac{d_n}{3^n}, \text{ with } d_n \in \{0,2\}.$$

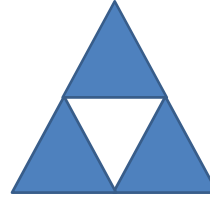
This includes, for example, $20/27 = 2/3 + 2/27 = 0.202000\dots$. Here, if a number has 2 expansions, we are saying 1 and only 1 of these expansions has no d_n taking the value 1. Thus $1/3 \in C$ since even though $1/3 = .10000\dots$, since we also have $1/3 = .0222\dots$

This implies there are as many points in the Cantor set as there are real numbers. The Cantor set is uncountable.

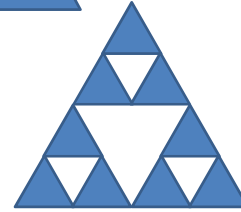
Hint. The numbers requiring a 1 in the 1st place of their ternary expansion lie in the interval $(1/3, 2/3)$. The numbers requiring a 1 in the 2nd place of their ternary expansion lie in the union of the intervals $(1/9, 2/9)$ and $(7/9, 8/9)$.

2) Sierpinski Triangle.

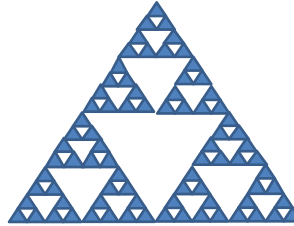
Start with an equilateral triangle and remove the center triangle.



Remove the center triangles from each of the 3 remaining triangles.



Keep going forever.



Fractal dimensions give a way of comparing fractals. Fractal dimensions can be defined in connection with real world data, such as the coastline of Great Britain. It turns out to have fractal dimension approximately 1.2. Here we will only look at the box dimension. It is only one of a wide variety of notions of fractal dimension. (The oldest and perhaps most important is the Hausdorff dimension. It is harder to calculate and you will need to look at the references such as Falconer to find out what that is.)

Definition of Box Dimension.

Defn. Suppose S is a subset of \mathbb{R}^n , $n=1,2,3$. By an n -box, we mean a closed interval if $n=1$; a square if $n=2$; a cube if $n=3$.

Defn. $\forall \epsilon > 0$ let $N(\epsilon)$ be the smallest number of n -boxes of side length ϵ needed to cover S .

Example.

It takes 8 boxes of length $1/3$ to cover the unit square with center square of side $1/3$ removed.



Defn. Box dimension of $S \subset \mathbb{R}^n$ is defined to be the following limit if it exists:

$$\dim_B S = \lim_{\epsilon \rightarrow 0} \frac{\ln N(\epsilon)}{\ln 1/\epsilon}$$

here we use $\ln(x)$ rather than $\log(x)$

Note. The box dimension is often called capacity. It sometimes differs from the Hausdorff dimension which we do not define here. But both definitions of dimension agree on most of the simple examples. For an m -dimensional surface in \mathbb{R}^n , $n=1,2,3$, the box dimension can be shown to be m . Thus, for example, the box dimension of the surface of the sphere is 2. See Falconer, p. 44.

Problem 2. Show that the box dimension of the unit square is 2.

Theorems to aid us in computing the box dimension.

Theorem 1. Let S be a subset of \mathbb{R}^n , $n=1,2,3$. If $0 < r < 1$, then the limit defining $\dim_B S$ exists iff the following limit exists and then they are equal

$$\dim_B S = \lim_{k \rightarrow \infty} \frac{\ln N(r^k)}{\ln \frac{1}{r^k}}.$$

Proof Sketch. Given $\varepsilon > 0$ and r with $0 < r < 1$, you just have to find a positive integer k such that

$$r^{k+1} < \varepsilon \leq r^k. \quad \text{Then } N(r^k) \leq N(\varepsilon) < N(r^{k+1}).$$

$$\text{As } \ln(x) \text{ monotone } \uparrow, \quad \ln\left(\frac{1}{r^k}\right) \leq \ln\left(\frac{1}{\varepsilon}\right) < \ln\left(\frac{1}{r^{k+1}}\right),$$

$$\text{and} \quad \ln N(r^k) \leq \ln N(\varepsilon) \leq \ln N(r^{k+1}).$$

$$\text{So} \quad \frac{\ln N(r^k)}{\ln(1/r^{k+1})} \leq \frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)} \leq \frac{\ln N(r^{k+1})}{\ln(1/r^k)}.$$



Example.

The Box dim of the Cantor set C .



Recall that C is obtained from the interval $[0,1]$ by continually removing middle thirds. At each stage there are twice as many intervals as the preceding stage. And each interval has length $1/3$ that of the preceding stage. Since $N(1/3) = 2$, we find that, by induction, $N(1/3^k) = 2^k$, for each $k \geq 1$. Let $r = 1/3$ in the preceding theorem and find

$$\begin{aligned} \dim_B C &= \lim_{k \rightarrow \infty} \frac{\ln N\left(\frac{1}{3^k}\right)}{\ln\left(\frac{1}{1/3^k}\right)} = \lim_{k \rightarrow \infty} \frac{\ln 2^k}{\ln 3^k} \\ &= \lim_{k \rightarrow \infty} \frac{k \ln 2}{k \ln 3} = \frac{\ln 2}{\ln 3} \cong 0.63. \end{aligned}$$

Problem 3. Show that the box dimension of the set of rational numbers in $[0,1]$ is 1.



Defn. Suppose our set S is a subset of \mathbb{R}^n , $n=1,2,3$. The distance between 2 points x,y is denoted $\|x-y\|$. Suppose that a function $f:S \rightarrow S$ has the property that for some constant r with $0 < r < 1$, $\|f(x)-f(y)\| = r \|x-y\|$, for all x,y in S . Then we call f a **similarity** of S . The constant r is called the **similarity constant**.

Example. Let $S=[0,1]$ & $f(x)=2/3 + x/3$. Then f is a similarity with constant $1/3$. It maps $[0,1]$ onto $[2/3,1]$.

Defn. If there are m similarities f_1, \dots, f_m of S such that $S=f_1(S) \cup \dots \cup f_m(S)$, and the images are non-overlapping except possibly for boundaries, we say that S is a **self-similar set**. It is composed of m (shrunk) copies of itself.

Example. The Cantor set C . Define $g(x)=x/3$ and $f(x)= 2/3 + x/3$. Then $C=f(C) \cup g(C)$. Both f and g are similarity functions with similarity constant $1/3$.

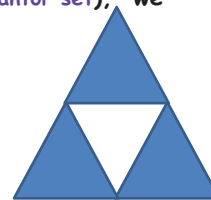
Problem 4. Show that the Sierpinski triangle is a self-similar set. Use this to see that the box dimension of the Sierpinski triangle is $\ln 3 / \ln 2 \cong 1.58$ using the following theorem. You need to define functions of vectors in the plane. First put the origin at the left hand base point of the big triangle. Then figure out what function shrinks the big triangle to the small one at the origin. Next what vector must you add to that function to shift the left small triangle over to the right one at the base?

Another Theorem making it easy to compute the box dimension.

Theorem 2. Suppose that S is a self similar set in \mathbb{R}^n , $n=1,2,3$; i.e., $S=f_1(S) \cup \dots \cup f_m(S)$, non-overlapping, and such that each similarity function f_i has the same similarity constant r . Then $\dim_B S = \ln m / \ln(1/r)$.

Proof. We use Theorem 1. Since $N(r^k) = cm^k$, for some positive constant c (Why? is [Problem 5](#). Hint. Think about the Cantor set), we have

$$\lim_{k \rightarrow \infty} \frac{\ln N(r^k)}{\ln(1/r)^k} = \lim_{k \rightarrow \infty} \frac{\ln m^k}{\ln(1/r)^k} = \frac{\ln m}{\ln(1/r)}.$$



Defn. A set M of real numbers is said to have **Lebesgue measure 0** iff for every $\varepsilon > 0$ there is a sequence of intervals E_n such that

$$M \subset \bigcup_{n \geq 1} E_n \text{ and } \sum_{n \geq 1} \text{length}(E_n) < \varepsilon.$$

Sets of Lebesgue measure 0 are considered negligible in the theory of the Lebesgue integral. One can ignore what a function does on such sets when computing Lebesgue integrals. And one identifies functions that are equal except on a set of Lebesgue measure 0.

Problem 6. Show that any countable set M of real numbers has Lebesgue measure 0.

Problem 7. Show that the Cantor set has Lebesgue measure 0.

Hint. To do this recall that setting $C_0 = [0, 1]$. Let $C_2 = [0, 1/3] \cup [2/3, 1]$. Continue in this way, defining C_k as a union of 2^k subintervals, each of length 3^{-k} obtained by removing middle thirds of the intervals in C_{k-1} . The Cantor set C is the intersection of all the C_k , for k running over all integers ≥ 0 . Show that C_k has length $(2/3)^k$.



Defn. The Devil's Staircase or the Almost Perfect Sneak.

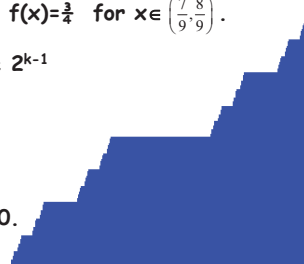
Define a function on the interval $[0, 1]$ as follows. Write the complement of the Cantor set in $[0, 1]$ as a union of intervals. First define f on the complement of the Cantor set.

Define $f(x) = \frac{1}{2}$ for $x \in (\frac{1}{3}, \frac{2}{3})$; $f(x) = \frac{1}{4}$ for $x \in (\frac{1}{9}, \frac{2}{9})$, $f(x) = \frac{3}{4}$ for $x \in (\frac{7}{9}, \frac{8}{9})$.

Then in the k th step, going from left to right on the 2^{k-1} subintervals left out of the Cantor set define

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

Define $f(x)$ for $x \in C =$ the Cantor set by $f(x) = \text{l.u.b.}\{f(t) \mid t < x, t \in [0, 1] - C\}$ and set $f(0) = 0$.



Problem 8. Prove that $f(x)$ is increasing, continuous and has derivative $f'(x) = 0$ except on the Cantor set C , which has Lebesgue measure 0. But $f(0) = 0$ and $f(1) = 1$. So the **fundamental theorem of calculus fails** for this function:

$$1 = f(1) - f(0) = \int_0^1 f'(t) dt = 0.$$

Hint. If f were not continuous, f would have a jump. Then there would have to be an open subinterval of $[0, 1]$ containing no value of f . But the range of f contains all numbers of the form $(2n+1)/2^k$ in $[0, 1]$.



A person moving toward you according to the Devil's staircase law $y=f(x)$ would cover a unit distance in a unit of time, but you might never see him or her move even if you were watching all the time. Thus Korevaar, *Mathematical Methods*, p. 404, calls this function "the almost perfect sneak."

In 1872 Weierstrass found functions that were continuous everywhere but nowhere differentiable. This shocked many famous mathematicians who had thought such a function impossible.

Hermite described these functions as a "dreadful plague."



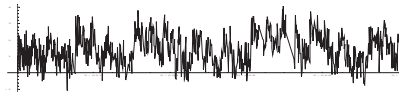
Poincaré wrote: "Yesterday, if a new function was invented it was to serve some practical end; today they are specially invented only to show up the arguments of our fathers, and they will never have any other use." Even as late as the 1960's, before "everyone" had a computer fast enough to graph these things, such examples were viewed as pathological monsters. Now there are thousands of websites with pictures of approximations of them.

The Weierstrass Nowhere Differentiable Function.

Weierstrass published this construction in 1872. It too is a fractal.

Defn. The Weierstrass function f . The definition involves parameters $\lambda > 1$ and $1 < s < 2$. Then $f: [0, 1] \rightarrow \mathbb{R}$ is defined by

$$f(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t).$$



Theorem 3. If λ is large enough, the box dimension of the graph of $f(t)$ in \mathbb{R}^2 is s . By the graph of f , we mean the set of points $(t, f(t))$, for all t in $[0, 1]$.

There are lots of pictures of this graph on the web. For example

http://en.wikipedia.org/wiki/Weierstrass_function.

<http://planetmath.org/encyclopedia/WeierstrassFunction.html>

Or see <http://www.math.washington.edu/%7Econroy/>

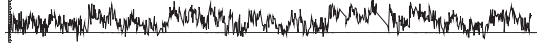
for an animation zooming in on the Weierstrass function

$$f(t) = \sum_{k=1}^{\infty} \frac{1}{2^k} \sin(2^k t).$$

Problem 9. Show that

$$f(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t) \text{ is a}$$

continuous function on $[0,1]$.



The following pictures of the Weierstrass function with $\lambda=2$ and $s=1.9$ were produced using the Mathematica commands below. I summed the first 150 terms of the Fourier series for the Weierstrass function and then plotted the result on the interval $[0,1]$. Why do 150 terms suffice? Look at the geometric series with $x=2^{-1}$. To get $x^k < 10^{-4}$, you need $\log(x^k) < \log(10^{-4})$ to get 4 significant digits. Then we want the next term which is the estimate for the remainder to be

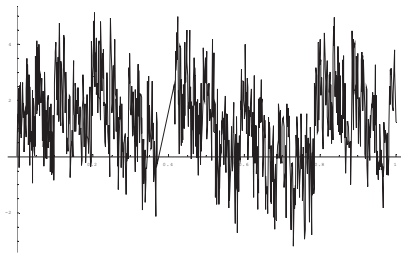
$$\begin{aligned} k \log(2^{-1}) &< -4 \log 10 &\Leftrightarrow & -0.1 k \log 2 < -4 \log 10 \\ \Leftrightarrow k &> 40 \log 10 / \log 2 \end{aligned}$$

$$40 * \text{Log}[10] / \text{Log}[2] // N \cong 132.877$$

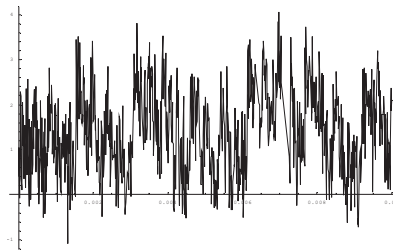
The commands to sum the first 150 terms and make a function of it. Then plot on the interval $[0,1]$.

```
fun[t_]:=fun[t]=Sum[2^(-k*(2-1.9))*Sin[t*2^k},{k,1,150}]
Plot[fun[t],{t,0,1}];
```

Weierstrass function $\lambda=2$ and $s=1.9$ plotted on the interval $[0,1]$.



Same function plotted on the interval $[0,0.01]$.



The essence of a fractal - the same behavior at all scales. This function also looks nowhere differentiable.

Preparations for computing the box dimension of a graph of a function.

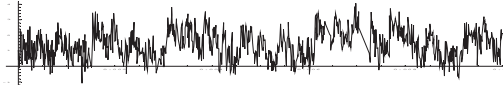
Defn. $R_f[a,b]=\sup\{|f(t)-f(u)|, t,u \text{ in } [a,b]\}$.

Proposition 1. Let $f:[0,1]\rightarrow\mathbb{R}$, be continuous. Suppose that $0<\delta<1$, and m is the least integer greater than or equal to $1/\delta$. If $N(\delta)$ is the number of squares of side δ that intersect the graph of f , we have

$$\delta^{-1} \sum_{k=1}^{m-1} R_f[k\delta, (k+1)\delta] \leq N(\delta) \leq 2m + \delta^{-1} \sum_{k=1}^{m-1} R_f[k\delta, (k+1)\delta].$$

Proof.

The number of squares of side δ in the column above the interval $[k\delta, (k+1)\delta]$ that intersect the graph of f is at least $R_f[k\delta, (k+1)\delta]/\delta$. It is at most $2 + R_f[k\delta, (k+1)\delta]/\delta$, using the fact that f is continuous. Sum over all the intervals to finish the proof.



Problem 10. Fill in the details in this proof. Draw a picture.

Lemma 1. Let $f:[0,1]\rightarrow\mathbb{R}$ be continuous. Assume the box dimension of the graph of f exists.

1) Suppose that $\exists c>0$ and $s, 1\leq s\leq 2$, such that $|f(t)-f(u)| \leq c|t-u|^{2-s}, \forall t,u \text{ in } [0,1]$. Then $\dim_B \text{graph} f \leq s$. This remains true if the inequality only holds for $|t-u|<\delta$, for some $\delta>0$.

2) Suppose that $\exists c>0, \delta_0>0$ and $s, 1\leq s\leq 2$ such that $\forall t \text{ in } [0,1]$, and $\forall \delta$ with $0<\delta\leq\delta_0 \exists u$ such that $|t-u|<\delta$ and $|f(t)-f(u)|\geq c\delta^{2-s}$. Then $\dim_B \text{graph} f \geq s$.



Proof.

1) From the hypothesis of 1) we see that $R_f[a,b]\leq c|a-b|^{2-s}$, for a,b in $[0,1]$. Using the notation of the proposition, we see that

$$m < (1 + \delta^{-1}) \text{ and thus } N(\delta) \leq (1 + \delta^{-1})(2 + c\delta^{-1}\delta^{2-s}) \leq c_1 \delta^{-s},$$

where c_1 is independent of δ . The result follows from the definition of box dimension.

2) Similarly the hypothesis of 2) implies that $R_f[a,b]\geq c|a-b|^{2-s}$. Since $\delta^{-1}\leq m$, we have from Proposition 1 that $N(\delta) \geq \delta^{-1} \delta^{-1} c \delta^{2-s} = c\delta^{-s}$. Again, the result follows from the definition of box dimension.

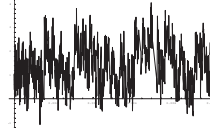


Problem 11. Suppose $f:[a,b] \rightarrow \mathbb{R}$ has a continuous derivative. Show $\dim_B \text{graph} f = 1$. Hint. Use the mean value theorem and Lemma 1.

Now we prove Theorem 3 which says that λ large enough implies that the box dimension of the graph of the Weierstrass function is s .

Given h in $(0,1)$, let N be the integer such that $\lambda^{-(N+1)} \leq h < \lambda^{-N}$.

Use our definition $f(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t)$.



Splitting sums up into 2 parts we get:

$$\begin{aligned} |f(t+h) - f(t)| &= \left| \sum_{k=1}^{\infty} \lambda^{(s-2)k} [\sin(\lambda^k(t+h)) - \sin(\lambda^k t)] \right| \\ &\leq \sum_{k=1}^N \lambda^{(s-2)k} |\sin(\lambda^k(t+h)) - \sin(\lambda^k t)| + \sum_{k=N+1}^{\infty} \lambda^{(s-2)k} |\sin(\lambda^k(t+h)) - \sin(\lambda^k t)| \\ &\leq \sum_{k=1}^N \lambda^{(s-2)k} \lambda^k h + \sum_{k=N+1}^{\infty} 2\lambda^{(s-2)k}. \end{aligned}$$

Here we used the mean value theorem on the first N terms and that $|\sin x| \leq 1$ for the rest of the terms. Then we sum the geometric series to see that

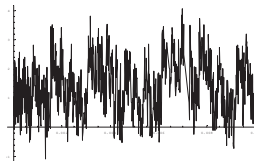
$$|f(t+h) - f(t)| \leq \frac{h\lambda^{(s-1)N}}{1 - \lambda^{1-s}} + 2 \frac{\lambda^{(s-2)(N+1)}}{1 - \lambda^{s-2}} \leq ch^{2-s},$$

where c is independent of h . This implies that the $\dim_B(\text{graph } f) \leq s$ by the Lemma above.

To go the other way, take our sum defining $f(t+h)-f(t)$ and split it into 3 parts, the first $N-1$ terms, the N th term, and the rest. This implies that

if $\lambda^{-(N+1)} \leq h < \lambda^{-N}$, then

$$\begin{aligned} & \left| f(t+h) - f(t) - \lambda^{(s-2)N} \sin(\lambda^N(t+h)) - \sin(\lambda^N t) \right| \\ (*) & \leq \frac{\lambda^{(s-2)N-s+1}}{1 - \lambda^{1-s}} + \frac{2\lambda^{(s-2)(N+1)}}{1 - \lambda^{s-2}}. \end{aligned}$$



Suppose that $\lambda > 2$ is large enough that

$$\text{right hand side of (*) is } < \frac{1}{20} \lambda^{(s-2)N}, \quad \forall N.$$

For $\delta < 1/\lambda$, take N such that $\lambda^{-N} \leq \delta < \lambda^{-(N-1)}$.

$\forall t, \exists h$, such that $\lambda^{-(N+1)} \leq h < \lambda^N$

and such that $\left| \sin[\lambda^N(t+h)] - \sin[\lambda^N t] \right| > \frac{1}{10}$.

All this implies that

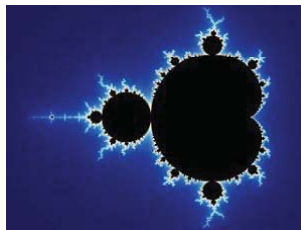
$$|f(t+h) - f(t)| \geq \frac{1}{20} \lambda^{(s-2)N} \geq \frac{1}{20} \lambda^{s-2} \delta^{2-s}.$$

It follows from the preceding Lemma that $\dim_b(\text{graph}f) \geq s$

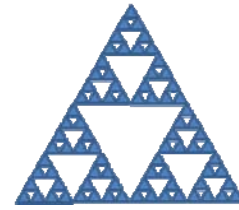


Problem 12. Show that any function satisfying condition 2 of Lemma 1 with $s > 1$ must be nowhere differentiable. It follows that the Weierstrass function is continuous but nowhere differentiable.

There is lots more to say about fractals, but we will stop here. I leave you with a picture of the Mandelbrot set from Wikipedia.



I, FINAL



1) Page 2 of fractals lecture. Show that the Cantor set can be expressed as certain triadic expansions and thus is uncountable.



2) Page 3 of fractals lecture. Show that the box dimension of the unit square is 2.

3) Page 5 of fractals lecture. Show that the box dimension of the set of rational numbers is 1. Compare with the box dimension of the Cantor set. How can it be that the Cantor set has a smaller box dimension than the rationals even though the Cantor set is uncountable?

4) Page 5 of fractals lecture. Show that the Sierpinski triangle is a self-similar set. Show using Thm. 2 that the box dimension of the Sierpinski triangle is $\ln 3 / \ln 2$.

5) Answer the Why? in the Proof of Theorem 2 on p.5 of the fractals lecture.

6) Page 6 of fractals lecture. Show that any countable set M of real numbers has Lebesgue measure 0.

7) Page 6 of fractals lecture. Show that the Cantor C set (fractals lecture p. 2) has Lebesgue measure 0. Hint. To do this show that C_k in the definition of C has length $(2/3)^k$.

8) Page 6 of fractals lecture. Prove that the Devil's Staircase function $f(x)$ is increasing, continuous and has derivative $f'(x)=0$ except on the Cantor set C , which has Lebesgue measure 0.

9) Page 8 of fractals lecture. Show that the Weierstrass function is a continuous function on $[0,1]$.

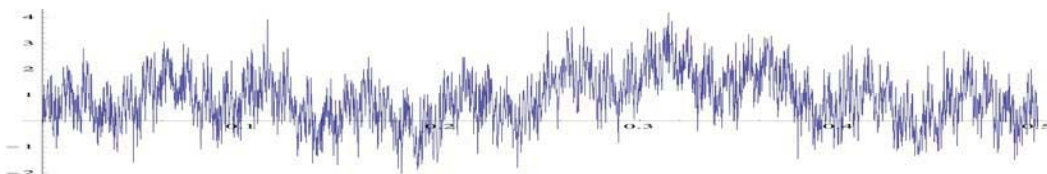
$$f(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t).$$

Hint. Use a convergence test (often called the Weierstrass M-Test) from calculus which we will do in the next part of the notes.

10) Page 9 of fractals lecture. Fill in the details of the proof of Proposition 1, page 9 of the fractals lecture.

11) Page 10 of fractals lecture. Suppose $f:[a,b] \rightarrow \mathbb{R}$ has a continuous derivative. Show $\dim B_{\text{graph}} f = 1$. Hint. Use the mean value theorem and Lemma 1.

12) Page 11 of fractals lecture. Show that any function satisfying condition 2 of Lemma 1 must be nowhere differentiable. It follows that the Weierstrass function is continuous but nowhere differentiable.



Lectures on Advanced Calculus with Applications, II

Audrey Terras

December, 2010

Part I

Normed Vector Spaces

1 Motivation

Recall Section 1.1 of Lectures I where we noted that Fourier needed to express a function $f(x)$ as a Fourier series:

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}.$$

Here $e^{ix} = \cos x + i \sin x$, where $i = \sqrt{-1}$ (which is not a real number). This means you can rewrite the series of complex exponentials as 2 series - one involving cosines and the other involving sines. The Fourier coefficients are

$$a_n = \int_0^1 f(y) e^{-2\pi i n y} dy.$$

In order to investigate Fourier series as well as integrals, we will need to look at normed vector spaces.

2 Why worry about infinite dimensional normed vector spaces?

We want to understand the integral from a more modern perspective rather than that of your calculus book. Secondly we want to understand convergence of series of functions - something that proved problematic for Cauchy in the 1800s. These things are important for many applications in physics, engineering, statistics. We will be able to study vibrating things such as violin strings, drums, buildings, bridges, spheres, planets, stock values. Quantum physics, for example, involves Hilbert space, which is a type of normed vector space with a scalar product where all Cauchy sequences of vectors converge.

The theory of such normed vector spaces was created at the same time as quantum mechanics - the 1920s and 1930s. So with this approach we are moving ahead hundreds of years from Newton and Leibnitz, perhaps 70 years from Riemann.

Fourier series involve orthogonal sets of vectors in an infinite dimensional normed vector space:

$$C[a, b] = \{f : [a, b] \rightarrow \mathbb{R} \mid f \text{ continuous}\}.$$

The L^2 -norm of a continuous function f in $C[a, b]$ is

$$\|f\|_2 = \left(\int_a^b |f(x)|^2 dx \right)^{1/2}.$$

This is an analog of the usual idea of length of a vector $f = (f(1), \dots, f(n)) \in \mathbb{R}^n$:

$$\|f\|_2 = \left(\sum_{j=1}^n |f(j)|^2 \right)^{1/2}.$$

There are other natural norms for $f \in C[a, b]$ such as:

$$\|f\|_1 = \int_a^b |f(x)| dx.$$

$$\|f\|_\infty = \max_{a \leq x \leq b} |f(x)|.$$

On **finite** dimensional vector spaces such as \mathbb{R}^n it does not matter what norm you use when you are trying to figure out whether a sequence of vectors has a limit. However, in infinite dimensional normed vector spaces, convergence can disappear if a different norm is used. Not all norms are equivalent in infinite dimensions. We will discuss this in detail later.

Note that $C[a, b]$ is infinite dimensional since the set $\{1, x, x^2, x^3, \dots, x^n, \dots\}$ is an infinite set of linearly independent vectors. Prove this as follows. Suppose that we have a linear dependence relation $\sum_{j=0}^n c_j x^j = 0$, for all x in $[a, b]$. This implies all the constants $c_j = 0$. **Why? Exercise.**

Infinite dimensional vector spaces are thus more interesting than finite dimensional ones. Each (inequivalent) norm leads to a different notion of convergence of sequences of vectors.

3 What is a Normed Vector Space?

In what follows we define normed vector space by 5 axioms. We will not put arrows on our vectors. We will try to keep vectors and scalars apart by mostly using Greek letters for scalars. Our scalars will be real in this section. However, at the end of these lectures, we will allow complex scalars. It simplifies Fourier series.

Definition 1 A **vector space** V is a set of vectors $v \in V$ which is closed under addition and closed under multiplication by scalars $\alpha \in \mathbb{R}$. This means \forall vectors $u, v \in V$, there is a unique sum $u + v \in V$ and \forall scalar $\alpha \in \mathbb{R}$, there is a unique product $\alpha v \in V$. Moreover the following 5 axioms must hold for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{R}$:

VS1. $u + (v + w) = (u + v) + w$

VS2. $\exists 0 \in V$ s.t. $0 + v = v$

VS3. $\forall v \in V \exists v' \in V$ s.t. $v + v' = 0$.

VS4. $v + u = u + v$

VS5. $1v = v$, $\alpha(\beta v) = (\alpha\beta)v$, $(\alpha + \beta)v = \alpha v + \beta v$, $\alpha(u + v) = \alpha u + \alpha v$.

You may say we cheated by putting 4 axioms into VS5.

Definition 2 A vector space V is a **normed vector space** if there is a norm function mapping V to the non-negative real numbers, written $\|v\|$, for $v \in V$, and satisfying the following 3 axioms:

N1. $\|v\| \geq 0 \forall v \in V$ and $\|v\| = 0$ if and only if $v = 0$.

N2. $\|\alpha v\| = |\alpha| \|v\|$, $\forall v \in V$ and $\forall \alpha \in \mathbb{R}$. Here $|\alpha|$ = absolute value of α .

N3. $\|u + v\| \leq \|u\| + \|v\|$, $\forall u, v \in V$. Triangle Inequality.

Definition 3 The **distance between 2 vectors** u, v in a normed vector space V is defined by

$$d(u, v) = \|u - v\|.$$

Example 1. 3-Space.

$$\mathbb{R}^3 = \left\{ \left(\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right) \mid x_1, x_2, x_3 \in \mathbb{R} \right\}.$$

Define addition and multiplication by scalars as usual:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix}.$$

$$\alpha \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{pmatrix}, \quad \forall \alpha \in \mathbb{R}.$$

The usual norm is

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2} \quad \text{if } x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Other norms are possible:

$$\|x\|_1 = |x_1| + |x_2| + |x_3| \quad \text{or} \quad \|x\|_\infty = \max\{|x_1|, |x_2|, |x_3|\}.$$

The proof that these definitions make \mathbb{R}^3 a normed vector space is tedious. So we make it an **exercise**.

Example 2. The space of continuous functions on an interval.

$$C[a, b] = \{f : [a, b] \rightarrow \mathbb{R} \mid f \text{ continuous}\}.$$

For $f, g \in C[a, b]$, define $(f + g)(x) = f(x) + g(x)$ for all $x \in [a, b]$ and define for $\alpha \in \mathbb{R}$ $(\alpha f)(x) = \alpha f(x)$ for all $x \in [a, b]$. We leave it as an **exercise** to check the axioms for a vector space. The most interesting part of the exercise is to show that $f + g$ and αf are both continuous functions on $[a, b]$.

Again there are many possible norms. We will look at 3:

$$\|f\|_2 = \left(\int_a^b |f(x)|^2 dx \right)^{1/2}.$$

$$\|f\|_1 = \int_a^b |f(x)| dx.$$

$$\|f\|_\infty = \max_{a \leq x \leq b} |f(x)|.$$

Most of the axioms for norms are easy to check. Let's do it for the $\|f\|_1$ norm.

Proof of N1. $\|v\| \geq 0 \quad \forall v \in V$ and $\|v\| = 0$ if and only if $v = 0$.

Since $|f(x)| \geq 0$ for all x we know that the integral is ≥ 0 , because the integral preserves inequalities (see Lectures I).

Suppose that $\int_a^b |f(x)| dx = 0$. Since f is continuous (as is $|f|$), this implies $f(x) = 0$ for all $x \in [a, b]$ by the positivity of the integral in Lectures I.

Proof of N2. $\|\alpha v\| = |\alpha| \|v\|, \quad \forall v \in V \quad \text{and} \quad \forall \alpha \in \mathbb{R}.$

Also for any $\alpha \in \mathbb{R}$ and $f \in C[a, b]$, we have $\|\alpha f\|_1 = \int_a^b |\alpha f(x)| dx = \int_a^b |\alpha| |f(x)| dx = |\alpha| \int_a^b |f(x)| dx = |\alpha| \|f\|_1$. This

proves N2 for norms. Here we used the multiplicative property of absolute value as well as the linearity of the integral (i.e., scalars come out of the integral from Lectures, I).

Proof of N3. $\|u + v\| \leq \|u\| + \|v\|$, $\forall u, v \in V$.

Using the definition of the 1-norm, and the triangle inequality for real numbers as well as the fact that the integral preserves \leq , we see that

$$\|f + g\|_1 = \int_a^b |f(x) + g(x)| dx \leq \int_a^b (|f(x)| + |g(x)|) dx.$$

To finish the proof, use the linearity of the integral to see that

$$\int_a^b (|f(x)| + |g(x)|) dx = \int_a^b |f(x)| dx + \int_a^b |g(x)| dx = \|f\|_1 + \|g\|_1.$$

Putting it all together gives $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$.

4 Scalar Products.

You have seen the dot (or scalar or inner) product in \mathbb{R}^3 . It is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \bullet \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = x_1y_1 + x_2y_2 + x_3y_3.$$

It turns out there is a similar thing for $C[a, b]$. First let's define the scalar product on a vector space and see how to get a norm if, in addition, the scalar product is positive definite.

Definition 4 A (positive definite) **scalar product** is a function mapping $(v, w) \in V \times V$ to $\langle v, w \rangle \in \mathbb{R}$ such that:

- SP1. $\langle v, w \rangle = \langle w, v \rangle$, $\forall v, w \in V$ (symmetry)
- SP2. $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$, $\forall u, v, w \in V$
- SP3. $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$, $\forall v, w \in V$ and $\forall \alpha \in \mathbb{R}$
- SP4. $\langle v, v \rangle \geq 0$, $\forall v \in V$ and $\langle v, v \rangle = 0 \iff v = 0$ (positive definite)

Axioms SP1,2,3 imply that $\langle v, w \rangle$ is **linear** in each variable holding the other variable fixed. Axiom SP4 says the scalar product is **positive definite**. We will always want to assume SP4 because we want to be able to get a norm out of the scalar product via the following definition.

Definition 5 If V is a vector space with a (positive definite) scalar product $\langle v, w \rangle$ for $v, w \in V$, define the associated **norm** by $\|v\| = \sqrt{\langle v, v \rangle}$, for all $v \in V$

Before proving that this really gives a norm, let's look at some examples.

Example 1. In \mathbb{R}^3 the scalar product is

$$\langle x, y \rangle = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \bullet \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = x_1y_1 + x_2y_2 + x_3y_3.$$

It is easy to check the axioms. For example, the positive definiteness follows from the fact that squares of real numbers are ≥ 0 and sums of non-negative numbers are non-negative:

$$\langle x, x \rangle = x_1^2 + x_2^2 + x_3^2 \geq 0.$$

And $0 = \langle x, x \rangle \geq x_i^2$ implies $x_i = 0$ for all i and thus $x = 0$.

Example 2. $C[a, b] = \{f : [a, b] \rightarrow \mathbb{R} \mid f \text{ is continuous}\}$

For f, g in $C[a, b]$, define the **scalar product** by

$$\langle f, g \rangle = \int_a^b fg.$$

Once more, it is not hard to use the properties of the integral to check axioms SP1, SP2, SP3 (**exercise**). To see SP4, note that $f(x)^2 \geq 0$ for all $x \in [a, b]$ implies by the fact that integrals preserve \geq that

$$\langle f, f \rangle = \int_a^b f(x)^2 dx \geq 0.$$

Now suppose that $\langle f, f \rangle = 0$. By the positivity property of the integral, we know that $f(x)^2 = 0$ for all $x \in [a, b]$ which says that f is the 0 function (the identity for addition in our vector space $C[a, b]$).

Then the norm associated to this scalar product is $\|f\|_2 = \left(\int_a^b |f(x)|^2 dx \right)^{1/2}$.

The following theorem is so useful people from lots of countries got their names attached.

Theorem 6 Cauchy-Schwarz (Bunyakovsky) Inequality

Suppose that V is a vector space with scalar product $\langle v, w \rangle$. Then, defining the norm $\|v\| = \sqrt{\langle v, v \rangle}$, we have for all $v, w \in V$:

$$|\langle v, w \rangle| \leq \|v\| \|w\|.$$

Proof. Let $t \in \mathbb{R}$ and look at

$$f(t) = \langle v + tw, v + tw \rangle.$$

By properties of the scalar product, we have $0 \leq f(t) = \langle v, v \rangle + 2t \langle v, w \rangle + t^2 \langle w, w \rangle$.

As a function of f , we see that $f(t) = At^2 + Bt + C$, where $A = \langle w, w \rangle$, $B = 2 \langle v, w \rangle$ and $C = \langle v, v \rangle$. So the graph of $f(t)$ is that of a parabola above or touching the t -axis. For example, in Figure 1, we have drawn a parabola touching the t -axis at one point.

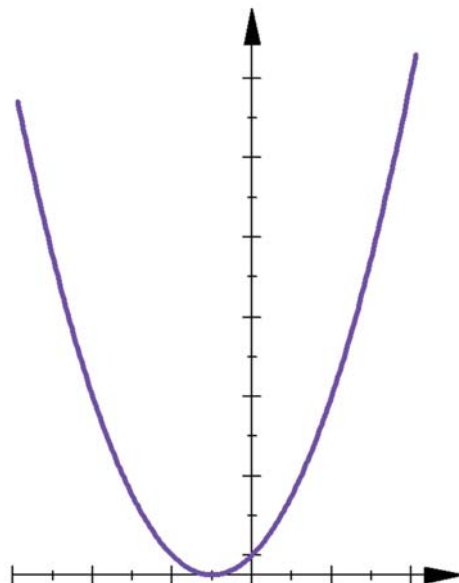


Figure 1: graph of a parabola with positive leading coefficient and non-positive discriminant

Recall the quadratic formula for the roots r_{\pm} of $f(t) = At^2 + Bt + C = 0$,

$$r_{\pm} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

Since we have at most one real root it follows that

$$B^2 - 4AC \leq 0.$$

Now plug in $A = \langle w, w \rangle$, $B = 2 \langle v, w \rangle$ and $C = \langle v, v \rangle$. This gives the Cauchy-Schwarz inequality. ■

Corollary 7 Under the hypotheses of the preceding theorem, using Definition 5, $\|v\| = \sqrt{\langle v, v \rangle}$ defines a norm on V .

Proof. We must prove:

N1. $\|v\| \geq 0 \quad \forall v \in V$ and $\|v\| = 0$ if and only if $v = 0$.

N2. $\|\alpha v\| = |\alpha| \|v\|$, $\forall v \in V$ and $\forall \alpha \in \mathbb{R}$.

N3. Triangle Inequality. $\|u + v\| \leq \|u\| + \|v\|$, $\forall u, v \in V$.

We get N1 from SP4.

We get N2 from SP3. For then $\|\alpha v\|^2 = \langle \alpha v, \alpha v \rangle = \alpha^2 \langle v, v \rangle = |\alpha|^2 \|v\|^2$, $\forall v \in V$ and $\forall \alpha \in \mathbb{R}$.

To prove the triangle inequality N3, we need to use the Cauchy-Schwarz inequality. This proof goes as follows. By the linearity and symmetry of the scalar product we see that

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle = \langle u, u \rangle + 2 \langle u, v \rangle + \langle v, v \rangle \\ &= \|u\|^2 + 2 \langle u, v \rangle + \|v\|^2 \\ &\leq \|u\|^2 + 2 |\langle u, v \rangle| + \|v\|^2 \quad \text{as } x \leq |x| \\ &\leq \|u\|^2 + 2 \|u\| \|v\| + \|v\|^2 \quad \text{by Cauchy - Schwarz} \\ &= (\|u\| + \|v\|)^2. \end{aligned}$$

Now use the fact that the square root $\sqrt{\quad}$ preserves inequalities to finish the proof of the triangle inequality. ■

What's the good of all this? Now we can happily define limits of sequences of vectors $\{v_n\}$ in our normed vector space V . **Can you guess the definition of** $\lim_{n \rightarrow \infty} v_n = L \in V$?

Answer: $\forall \varepsilon > 0, \exists N_{\varepsilon} \in \mathbb{Z}^+$ s.t. $n \geq N_{\varepsilon}$ implies $\|v_n - L\| < \varepsilon$. That is, just replace absolute value in the old definition of limit with the norm.

Similarly we can define a Cauchy sequence $\{v_n\}$ in the normed vector space V . We will do all this in detail later, but you should be able to guess what we will say.

Another use of the scalar product is to define orthogonal vectors in a vector space V with a scalar product.

Definition 8 Two vectors $v, w \in V$, a vector space V with scalar product \langle, \rangle , are defined to be **orthogonal** if the scalar product $\langle v, w \rangle = 0$.

In a vector space with scalar product, you can also **define the angle θ between 2 vectors** $v, w \in V$, by

$$\langle v, w \rangle = \|v\| \|w\| \cos \theta.$$

From this, the preceding definition makes sense as 2 vectors are orthogonal iff the cosine of the angle between them is 0. This definition of cosine agrees with the usual one if $V = \mathbb{R}^2$.

What is the cosine law? Using the triangles in Figure 2, it says

$$\|v - w\|^2 = \|v\|^2 - 2 \|v\| \|w\| \cos \theta + \|w\|^2.$$

You also need to see, using the axioms for scalar product, that

$$\|v - w\|^2 = \langle v - w, v - w \rangle = \|v\|^2 - 2 \langle v, w \rangle + \|w\|^2.$$

Putting these last 2 formulas together yields $\langle v, w \rangle = \|v\| \|w\| \cos \theta$.

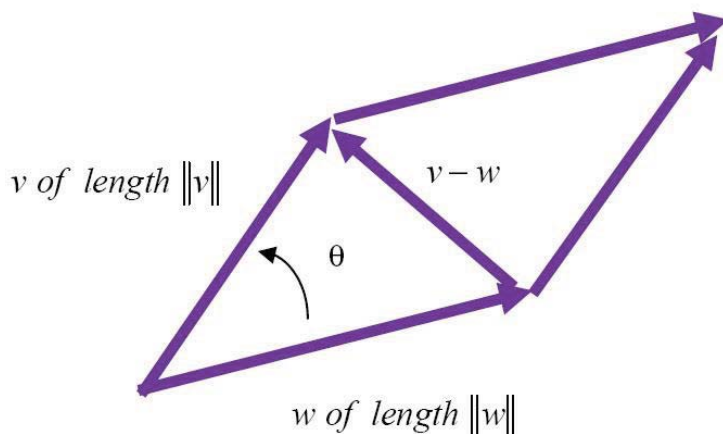


Figure 2: Visualizing vectors in a normed vector space and the angle between them. The cosine law says $\|v - w\|^2 = \|v\|^2 - 2\|v\|\|w\|\cos\theta + \|w\|^2$.

5 Comparison of Norms

Suppose $\{v_n\}$ is a sequence in a normed vector space V with norm $\|\cdot\|$. We will say $\lim_{n \rightarrow \infty} v_n = L$ and " v_n converges to $L \in V$ " $\iff \lim_{n \rightarrow \infty} \|v_n - L\| = 0$. Note that the last sequence $\|v_n - L\|$ is a sequence of real numbers.

Question. We know that there are lots of norms on V . How can we guarantee that 2 different norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ produce the same convergent sequences in V ?

The answer is that equivalent norms produce the same convergent sequences where we define equivalent as follows.

Definition 9 2 norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ on the vector space V are **equivalent** iff there are constants $A, B > 0$ such that for all $v \in V$ we have

$$A\|v\|_\alpha \leq \|v\|_\beta \leq B\|v\|_\alpha.$$

In the preceding definition we are assuming that the constants A and B are independent of $v \in V$.

Why do equivalent norms lead to the same convergent sequences?

Answer. Suppose $\{v_n\}$ is a convergent sequence for the $\|\cdot\|_\alpha$ -norm; i.e., for some $L \in V$ we have $\lim_{n \rightarrow \infty} \|v_n - L\|_\alpha = 0$.

And suppose $\|\cdot\|_\beta$ is an equivalent norm. Then

$$A\|v_n - L\|_\alpha \leq \|v_n - L\|_\beta \leq B\|v_n - L\|_\alpha.$$

Since the outside sequences go to 0 as $n \rightarrow \infty$, it follows by the squeeze lemma that the guy in the middle has to go to 0 as well.

Similarly $\lim_{n \rightarrow \infty} \|v_n - L\|_\beta = 0$ implies $\lim_{n \rightarrow \infty} \|v_n - L\|_\alpha = 0$.

Moral. It does not matter which of 2 equivalent norms you use to test a sequence for convergence.

Theorem 10 All norms on \mathbb{R}^n are equivalent.

Proof. See Lang, *Undergraduate Analysis*, p. 145. ■

Thus, for our purposes, it does not matter which norm you use on finite dimensional vector spaces. You get the same definition of convergence of sequences. However, things are very different for infinite dimensional vector spaces. See Figure 3 below for a sequence of functions f_n in $C[0, 1]$ such that

$$\lim_{n \rightarrow \infty} \|f_n - 0\|_1 = 0 \quad \text{using the norm} \quad \|f\|_1 = \int_0^1 |f(x)| dx.$$

However,

$$\|f_n - 0\|_\infty = 1 \text{ for all } n, \text{ using the norm} \quad \|f\|_\infty = \max_{a \leq x \leq b} |f(x)|.$$

It follows that the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ on $C[0, 1]$ are not equivalent.

Exercise. Using the same example, show that $\lim_{n \rightarrow \infty} \|f_n - 0\|_2 = 0$ using the norm $\|f\|_2 = \left(\int_0^1 |f(x)|^2 dx \right)^{1/2}$. It follows that the norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$ on $C[0, 1]$ are not equivalent.

Proposition 11 The norms $\|f\|_1 = \int_a^b |f(x)| dx$ and $\|f\|_2 = \left(\int_a^b f(x)^2 dx \right)^{1/2}$ are not equivalent. However, we do have the inequality

$$\|f\|_1 \leq \sqrt{b-a} \|f\|_2.$$

Proof. To prove the inequality, use the Cauchy-Schwarz inequality on the functions $|f|$ and $g(x) = 1$ for all $x \in [a, b]$. This gives

$$|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2.$$

So we have $|\langle |f|, 1 \rangle| \leq \| |f| \|_2 \|1\|_2$. Now this really means

$$\begin{aligned} \|f\|_1 &= \int_a^b |f(x)| dx \leq \left(\int_a^b f(x)^2 dx \right)^{1/2} \left(\int_a^b 1 dx \right)^{1/2} \\ &= \sqrt{b-a} \left(\int_a^b f(x)^2 dx \right)^{1/2} = \sqrt{b-a} \|f\|_2. \end{aligned}$$

To see that $\|\cdot\|_1$ and $\|\cdot\|_2$ are not equivalent norms, we take $a = 0$ and $b = 1$. Then we look at the following example. Define as in Lang, *Undergraduate Analysis*, p. 147:

$$g_n(x) = \begin{cases} \sqrt{n}, & \text{for } 0 \leq x \leq \frac{1}{n}; \\ \frac{1}{\sqrt{x}}, & \text{for } \frac{1}{n} \leq x \leq 1. \end{cases}$$

Note that g_n is continuous. Why? **Exercise.** Show also that

$$\|g_n\|_1 = 2 - \frac{1}{\sqrt{n}}$$

and

$$\|g_n\|_2 = \sqrt{1 + \log n}.$$

It follows that there cannot be a constant $C > 0$ such that $\|v\|_2 \leq C \|v\|_1$ at least on the interval $[0, 1]$. Can you extend this idea to arbitrary intervals $[a, b]$? ■

6 Limits in a Normed Vector Space

Let V be a normed vector space with norm $\|\cdot\|$. Recall our definition.

Definition 12 We say that a sequence of vectors $v_n \in V$ **converges** to $L \in V$ and write $\lim_{n \rightarrow \infty} v_n = L \iff$

$$\lim_{n \rightarrow \infty} \|v_n - L\| = 0.$$

Examples.

Example 1) Let $V = \mathbb{R}^2$ and consider $v_n = (\frac{1}{n}, \frac{1}{n^2})$. Using whatever norm is your favorite, $\lim_{n \rightarrow \infty} v_n = 0 = (0, 0)$. Prove this as an **exercise**.

Example 2) Define a sequence of functions on $[0, 1]$ by the picture in Figure 3. One can show (**exercise**) that

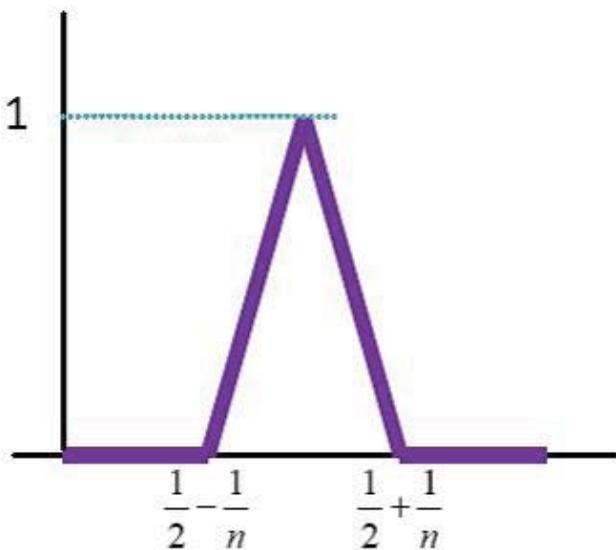


Figure 3: Define a function $f_n(x), 0 \leq x \leq 1$ by this graph.

$\|f_n - 0\|_1 \rightarrow 0$ as $n \rightarrow \infty$ while $\|f_n - 0\|_\infty = 1$ for all n . So the sequence has a limit function 0 using the 1-norm on $C[0, 1]$ but not using the ∞ -norm.

Definition 13 A **Cauchy sequence** $\{v_n\}$ of vectors in a normed vector space V means that

$$\forall \varepsilon > 0, \exists N \text{ s.t. } n, m \geq N \implies \|v_n - v_m\| < \varepsilon.$$

Let's list some facts about limits of sequences of vectors in V . Compare with the analogs for sequences of real numbers and you will see that mostly we replace the absolute value in the real version with the norm in our new version. We did these things in Lectures I. Sometimes we write exactly the same formulas as before, though of course we mean something different as now we are talking about sequences of vectors in V which may in fact be sequences of functions. The proofs of the following facts will be mostly the same as before, after replacing absolute value with norm.

Facts

1) **Uniqueness of Limits.**

$$\lim_{n \rightarrow \infty} v_n = L \text{ and } \lim_{n \rightarrow \infty} v_n = M \implies L = M.$$

2) **Cauchy implies Bounded**

If the sequence $\{v_n\}$ is Cauchy then $\{v_n\}$ is bounded; i.e., contained in a ball of radius r and center 0; meaning that $\|v_n\| \leq r$ for all $n \in \mathbb{Z}^+$.

3) Linearity of Limits.

$$\lim_{n \rightarrow \infty} v_n = L \quad \text{and} \quad \lim_{n \rightarrow \infty} w_n = M$$

implies

$$\lim_{n \rightarrow \infty} (v_n + w_n) = L + M$$

and for all $\alpha \in \mathbb{R}$.

$$\lim_{n \rightarrow \infty} (\alpha v_n) = \alpha L$$

4) **Convergent implies Cauchy.** $\{v_n\}$ converges $\implies \{v_n\}$ Cauchy.

5) **Subsequences of Convergent Sequences Converge to Same Limit as the Original Sequence.**

$\lim_{n \rightarrow \infty} v_n = L$ implies for any subsequence $\{v_{n_k}\}_{k \geq 1}$ we have $\lim_{k \rightarrow \infty} v_{n_k} = L$.

6) **Limits are the same for Equivalent Norms.**

If $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are 2 equivalent norms on V , then

$$\lim_{n \rightarrow \infty} \|v_n - L\|_\alpha = 0 \iff \lim_{n \rightarrow \infty} \|v_n - L\|_\beta = 0.$$

Proofs of the Facts. Compare with analogous proofs in \mathbb{R} .

1) $\|L - M\| = \|L - v_n + v_n - M\| \leq \|L - v_n\| + \|v_n - M\|$ by the triangle inequality.

We know that $\|L - v_n\| \rightarrow 0$ and $\|v_n - M\| \rightarrow 0$ as $n \rightarrow \infty$. This means $0 \leq \|L - M\| \leq 0$. Thus $\|L - M\| = 0$. By the first property N1 of norms, this implies $L - M = 0$.

2) Look at the proof of the analogous fact in Lectures I. Replace the absolute value in the old proof with the norm and you have proved Fact 2. Let's do it for practice.

Take $\varepsilon = 1$ and find N_0 such that $n \geq N_0$ implies $\|v_n - v_{N_0}\| < 1$. Then

$$\|v_n\| \leq \|v_n - v_{N_0} + v_{N_0}\| \leq \|v_n - v_{N_0}\| + \|v_{N_0}\| < 1 + \|v_{N_0}\|.$$

Take $B = \max\{\|x_1\|, \dots, \|x_{N_0-1}\|, \|v_{N_0}\| + 1\}$. This gives the bound B on all the v_n .

3) Use the triangle inequality to see that

$$\|v_n + w_n - (L + M)\| \leq \|v_n - L\| + \|w_n - M\|.$$

So given $\varepsilon > 0$, we can choose N_1 and N_2 so that

$$n \geq N_1 \quad \text{implies} \quad \|v_n - L\| < \varepsilon/2$$

and

$$n \geq N_2 \quad \text{implies} \quad \|w_n - M\| < \varepsilon/2.$$

Thus $n \geq \max\{N_1, N_2\}$ implies

$$\|v_n + w_n - (L + M)\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Exercise. Prove the second part of Fact 3) and then generalize to show that if $\{\alpha_n\}$ denotes a sequence of real numbers such that $\lim_{n \rightarrow \infty} \alpha_n = \alpha$, then $\lim_{n \rightarrow \infty} \alpha_n v_n = \alpha L$, assuming $\lim_{n \rightarrow \infty} v_n = L$ for the sequence $\{v_n\}$ of vectors in V .

4) The analogous proof was in Lectures 1. Replace all the absolute values in that proof with our norm and you have the proof of Fact 4.

If $\lim_{n \rightarrow \infty} v_n = L$ and $\varepsilon > 0$ is given, there exists $N_\varepsilon \in \mathbb{Z}^+$ such that $n \geq N_\varepsilon$ implies $\|v_n - L\| < \varepsilon/2$. So if $m \geq N_\varepsilon$, we have $\|v_m - L\| < \varepsilon/2$. It follows from the triangle inequality that for $n, m \geq N_\varepsilon$ we have

$$\begin{aligned} \|v_n - v_m\| &\leq \|v_n - L + L - v_m\| \\ &\leq \|v_n - L\| + \|L - v_m\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

This means that our sequence is Cauchy.

5) See Lectures I. Suppose that $\{v_{n_k}\}$ is a subsequence of $\{v_n\}$ and $\lim_{n \rightarrow \infty} v_n = L$. We need to prove that $\lim_{k \rightarrow \infty} v_{n_k} = L$. For this we do the usual trick with the triangle inequality:

$$\|v_{n_k} - L\| = \|v_{n_k} - v_n + v_n - L\| \leq \|v_{n_k} - v_n\| + \|v_n - L\|.$$

Suppose we are given $\varepsilon > 0$. Since our sequence $\{v_n\}$ is convergent, it is Cauchy (by Fact 4) and there is an N_1 so that

$$n, m \geq N_1 \text{ implies } \|v_m - v_n\| < \varepsilon.$$

If $k \geq N_1$, then $n_k \geq k$ implies $n_k \geq N_1$ and thus $\|v_{n_k} - v_k\| < \varepsilon$.

Since $\lim_{n \rightarrow \infty} v_n = L$, we know there is an N_2 such that

$$k \geq N_2 \text{ implies } \|v_k - L\| < \varepsilon.$$

It follows that, setting $N = \max\{N_1, N_2\}$, we have the result we want, namely

$$k \geq N \implies \|v_{n_k} - L\| \leq \|v_{n_k} - v_k\| + \|v_k - L\| < 2\varepsilon.$$

6) We proved this in a previous section.



7 Completeness of Normed Vector Spaces

Definition 14 Suppose that V is a normed vector space. We say that V is **complete** iff every Cauchy sequence $\{v_n\}$ of vectors $v_n \in V$ converges to a limit $L \in V$.

Theorem 15 Examples of Complete and Incomplete Normed Vector Spaces.

1) \mathbb{R}^k is a complete normed vector space.

2) The space of continuous functions on a finite interval $[a, b]$ which we denote $C[a, b]$ is complete using the norm $\|\cdot\|_\infty$ but not complete using the norms $\|\cdot\|_1$ or $\|\cdot\|_2$.

Proof.

1) We did the case $k = 1$ in Lectures I. Let's take $k = 2$ for simplicity here. Suppose we use the norm $\|\cdot\|_\infty$ for our proof. That is, if $x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix}$, set $\|x\|_\infty = \max\{|x^{(1)}|, |x^{(2)}|\}$. This is legal as all norms on \mathbb{R}^2 are equivalent (see Lang,

Undergraduate Analysis, p. 145). Suppose $\{v_n\}$ is a Cauchy sequence of vectors in \mathbb{R}^2 . Write $v_n = \begin{pmatrix} v_n^{(1)} \\ v_n^{(2)} \end{pmatrix}$.

Claim. $\{v_n^{(j)}\}$ is a Cauchy sequence in \mathbb{R} for $j = 1$ and $j = 2$.

Proof of Claim. Since $\{v_n\}$ is a Cauchy sequence and $\|v_n - v_m\|_\infty = \max\{|v_n^{(1)} - v_m^{(1)}|, |v_n^{(2)} - v_m^{(2)}|\}$, for $j = 1$ or $j = 2$, we have

$$\forall \varepsilon > 0, \exists N \text{ s.t. } n, m \geq N \implies |v_n^{(j)} - v_m^{(j)}| \leq \|v_n - v_m\|_\infty < \varepsilon.$$

Q.E.D. Claim.

Since \mathbb{R} is complete and Cauchy implies convergent in \mathbb{R} , we know there exist $L^{(1)}, L^{(2)} \in \mathbb{R}$ such that for $j = 1$ and 2 ,

$$L^{(j)} = \lim_{n \rightarrow \infty} v_n^{(j)}.$$

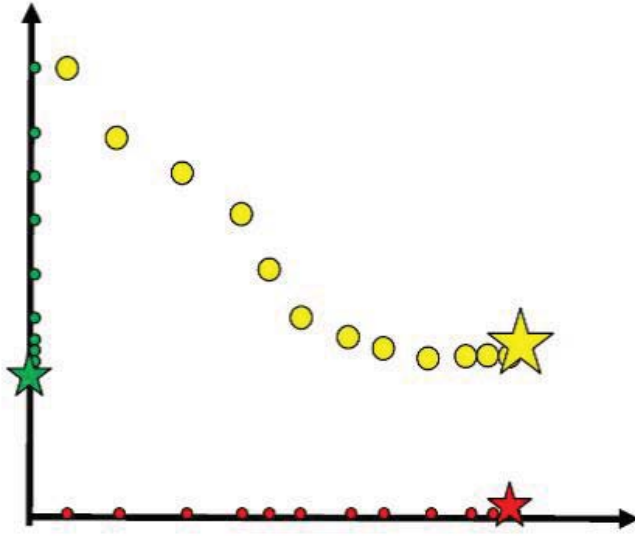


Figure 4: Here is a picture of the proof that a Cauchy sequence of vectors in the plane must converge because the projections of the sequence to points on the 2 axes converge to limits in \mathbb{R} , the red star and the green star.

Let $L = \begin{pmatrix} L^{(1)} \\ L^{(2)} \end{pmatrix}$.

Claim. $\lim_{n \rightarrow \infty} v_n = L$.

Proof. See Figure 4.

$\forall \varepsilon > 0, \exists N_j$ s.t. $n \geq N_j \implies |v_n^{(j)} - L^{(j)}| < \varepsilon$. Let $N = \max\{N_1, N_2\}$. Then $n \geq N$ implies

$$\|v_n - L\|_\infty = \max\left\{|v_n^{(1)} - L^{(1)}|, |v_n^{(2)} - L^{(2)}|\right\} < \varepsilon.$$

Q.E.D. Claim.

2) We postpone the proof that $C[a, b]$ is complete in the $\|\cdot\|_\infty$ norm.

To see that $C[0, 1]$ is not complete with respect to the $\|\cdot\|_1$ norm, we must find a Cauchy sequence of continuous functions that does not converge to an element of $C[0, 1]$ in this norm.

Example.

Consider the function $f_n(x)$ defined by

$$f_n(x) = \begin{cases} 0, & 0 \leq x \leq \frac{1}{2} - \frac{1}{n} \\ 1 + n(x - \frac{1}{2}), & \frac{1}{2} - \frac{1}{n} \leq x \leq \frac{1}{2} \\ 1, & \frac{1}{2} \leq x. \end{cases}$$

See Figure 5.

Define the function $L(x) = \begin{cases} 0, & 0 \leq x < \frac{1}{2} \\ 1, & \frac{1}{2} \leq x \leq 1. \end{cases}$

Since $L(x)$ is not continuous, it is not in the space $C[0, 1]$. But

$$\|f_n - L\|_1 = \text{area of triangle in Figure 6} = \frac{1}{2n} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Thus the sequence $\{f_n\}$ approaches a limit in the $\|\cdot\|_1$ -norm, but not a limit that is continuous. So we have a Cauchy sequence approaching a limit $L \notin C[0, 1]$. This says $C[0, 1]$ is not complete. It is like the rationals, **full of holes**. To make

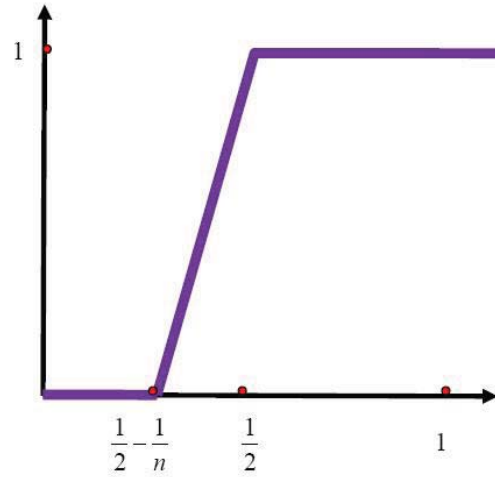


Figure 5: The function $f_n(x)$ designed not to converge to a continuous function with respect to the $\|\cdot\|_1$ norm as $n \rightarrow \infty$.

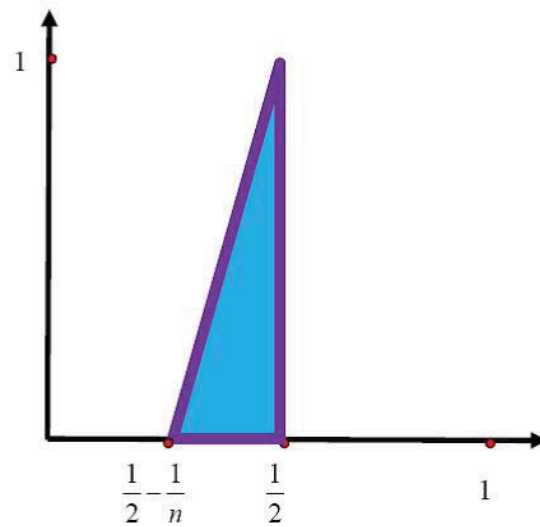


Figure 6: Triangle whose area is $\|f_n - L\|_1$

a space containing limits of all Cauchy sequences in $C[0, 1]$, you need to add all Lebesgue integrable functions on $[0,1]$. We will not say much about the Lebesgue integral in this course. If you are interested, you could look at Lang, *Undergraduate Analysis*, p. 262, Apostol, *Mathematical Analysis*, or Korevaar, *Mathematical Methods*.

Figure 7 shows the difference between $\|f - g\|_1$ and $\|f - g\|_\infty$. Let f be the purple function and g be the blue one. Then $\|f - g\|_\infty$ is the maximum length of the pink dotted lines while $\|f - g\|_1$ is the area between the 2 curves.

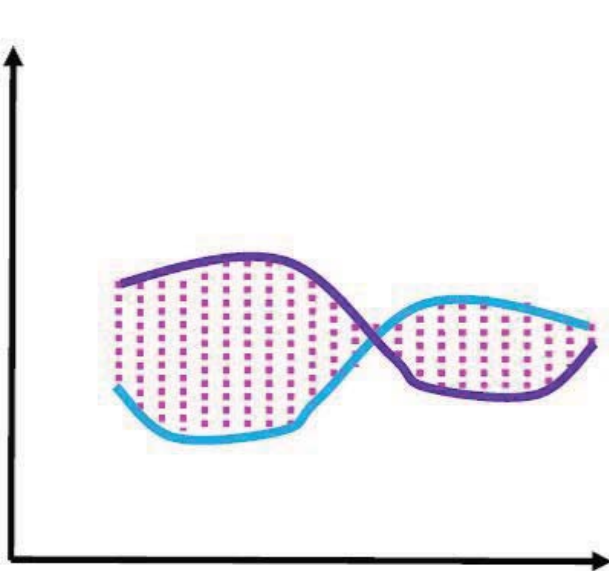


Figure 7: This figure can be used to see the difference between $\|f - g\|_1$ and $\|f - g\|_\infty$.

8 Open and Closed Sets in a Normed Vector Space and Other Definitions.

Here we give a very brief discussion of some concepts from point set topology, in particular open and closed sets. You could probably live without more definitions (unless you plan to go to grad school in math.). But I find that in discussing continuity, these ideas clarify things. If you always hated $\varepsilon\delta$, you will be happy to learn that these ideas allow you to forget about $\varepsilon\delta$.

Definition 16 An *open set* U in a normed vector space V is a subset of V such that $\forall a \in U, \exists r > 0$ such that the *open ball* of radius r and center a , $B(a, r) = \{x \in V \mid \|x - a\| < r\} \subset U$.

This means that U has no hard edges - no boundary points. As an example, if $V = \mathbb{R}$ with the usual absolute value as norm, the open interval (a, b) is an open set. In any normed vector space V , the open ball $B(a, r)$ is an open set.

A *closed set* F in a normed vector space V is a subset of V such that the complement F^c is an open set. Closed sets have hard edges. For example, if $V = \mathbb{R}$ with the usual absolute value as norm, the closed interval $[a, b]$ is a closed set. In any normed vector space V , the closed ball $\overline{B(a, r)} = \{x \in V \mid \|x - a\| \leq r\}$ is a closed set.

See Figure ?? for pictures of open and closed sets.

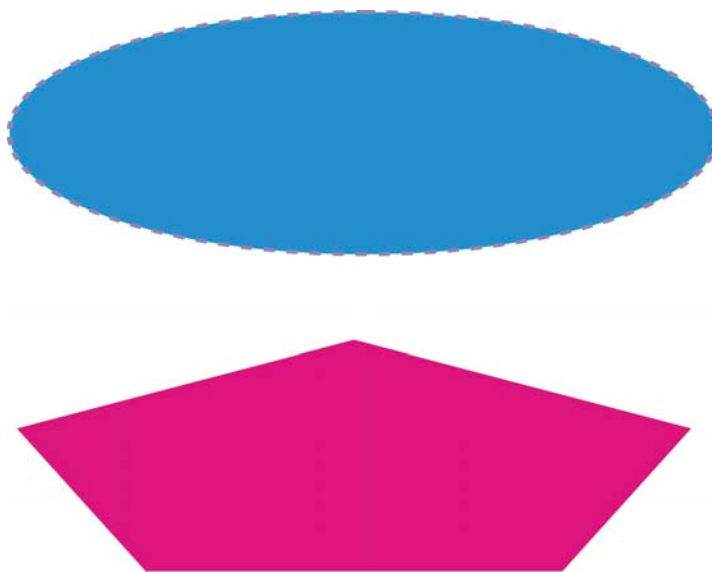


Figure 8: pictures of an open set (top) and a closed set (bottom). The boundary of the open set is dashed to indicate that it is not part of the set.

Theorem 17 *Properties of Open & Closed Sets in a normed vector space V*

- 1) The empty set \emptyset and V are both open and closed.
- 2) Finite intersections of open sets are open.
Arbitrary intersections of closed sets are closed.
- 3) Arbitrary unions of open sets are open.
Finite unions of closed sets are closed.

Proof. I will leave most of these proofs as **exercises**. Let me do parts of 2) and 3).

2) Suppose that U_1, \dots, U_n are open sets in V . If $p \in \bigcap_{i=1}^n U_i$, then $\forall i, \exists \delta_i > 0$ such that $B(p, \delta_i) \subset U_i$. Let $\delta = \min\{\delta_1, \dots, \delta_n\}$. Then $B(p, \delta) \subset \bigcap_{i=1}^n U_i$. Thus $\bigcap_{i=1}^n U_i$ is open. This proof is pictured for $n = 2$ in Figure 9.

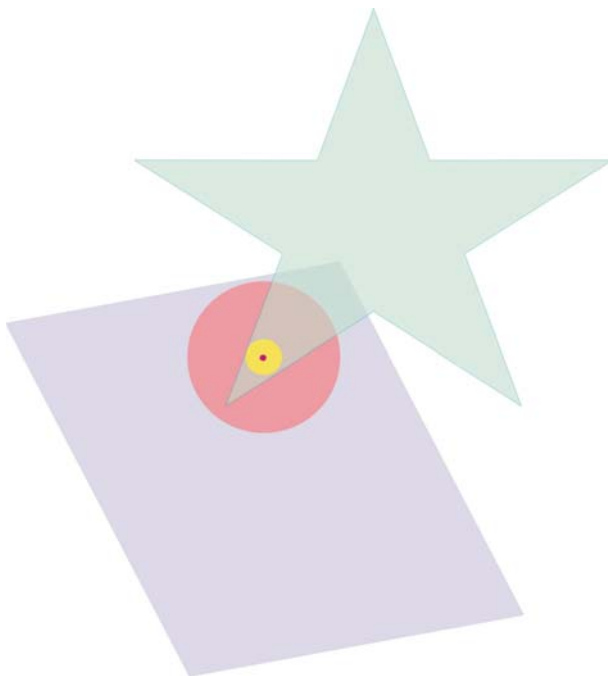


Figure 9: An open turquoise star is intersected with an open purple rhombus. The red point is in the intersection of the 2 sets. Then 2 discs with centers at the red point are found, the yellow disc staying inside the star and the pink disc staying inside the rhombus.

3) Suppose $\{U_i\}_{i \in I}$ denotes an arbitrary collection of open sets in V . If $p \in \bigcup_{i \in I} U_i$, then $p \in U_i$, for some i and thus $\exists \delta_i > 0$ such that $B(p, \delta) \subset U_i \subset \bigcup_{i \in I} U_i$. It follows that $\bigcup_{i \in I} U_i$ is open.

The facts about closed sets can be derived from the facts about open sets using $(A \cup B)^c = A^c \cap B^c$. ■

A set may be neither open nor closed. For example, consider the half-open interval $(a, b]$.

Definition 18 The **closure** \bar{A} of a subset A of a complete normed vector space V is the set of points $p \in V$ such that $\forall \delta > 0, \exists x \in A$ such that $\|x - p\| < \delta$. This says that $\forall \delta > 0, B(p, \delta) \cap A \neq \emptyset$.

You should picture a point in the closure of a set as a point sticking to the set. Some call points in the closure of A adherent to A for that reason. See Figure 10.

Proposition 19 A point p is in \bar{A} iff \exists a sequence $\{x_n\}$ of points $x_n \in A$ such that $p = \lim_{n \rightarrow \infty} x_n$.

Proof. \implies If $p \in \bar{A}$, for every $n \in \mathbb{Z}^+$, $\exists x_n \in B(p, \frac{1}{n}) \cap A$. It follows that $p = \lim_{n \rightarrow \infty} x_n$.

\impliedby If $p = \lim_{n \rightarrow \infty} x_n$, for a sequence $\{x_n\}$ of points $x_n \in A$, it follows that $\forall \delta > 0, \exists N$ s.t. $n \geq N$ implies $x_n \in B(p, \delta) \cap A$.

This implies $p \in \bar{A}$. ■

Examples.

1) Any point in S is in the closure of S ; i.e., $S \subset \bar{S}$. Just take $x = p$ in the definition of closure.

2) If $V = \mathbb{R}^2$, using any of our favorite norms, any point in the closed ball $\{x \in \mathbb{R}^2 \mid \|x\| \leq r\}$ is in the closure of the open ball $B(0, r) = \{x \in \mathbb{R}^2 \mid \|x\| < r\}$.

If a point x is in the complement of the closed ball, then it is not in the closure of $B(0, r)$.

Thus $\bar{B}(0, r) = \{x \in \mathbb{R}^2 \mid \|x\| \leq r\}$.

3) What is the closure of \mathbb{Q} =the rationals? Answer. \mathbb{R} .

4) What is the closure of $\{\frac{1}{n} \mid n \in \mathbb{Z}^+\}$? Answer. $\{0\} \cup \{\frac{1}{n} \mid n \in \mathbb{Z}^+\}$.

5) What is the closure of the interval (a, b) in \mathbb{R} ? Answer. $[a, b]$.

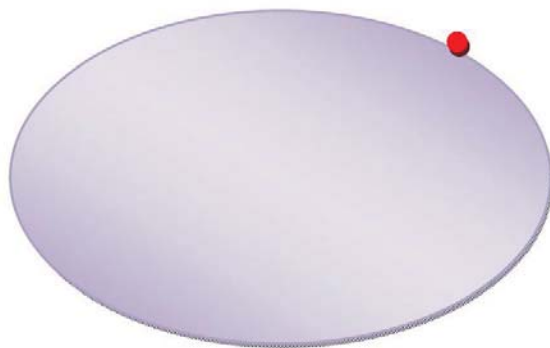


Figure 10: The red point is an element of the closure of the purple set.

6) What is the closure of the set $\{ \text{polynomials } p(x) = a_n x^n + \dots + a_1 x + a_0 \mid a_j \in \mathbb{R} \}$ in the space $C[a, b]$ of continuous functions on a finite interval $[a, b]$ using the $\| \cdot \|_\infty$ norm? Answer. $C[a, b]$ by the Weierstrass Theorem proved in a later section saying that any continuous function on a finite interval can be uniformly approximated by polynomials.

Proposition 20 A set S in a normed vector space is closed iff $S = \overline{S}$.

Proof. Suppose S is closed. Then S^c is open. First we know that $S \subset \overline{S}$. Now to go the other way, suppose $p \in S^c$. Then $\exists \delta > 0$ such that $B(p, \delta) \subset S^c$. This means that p is not in \overline{S} . That shows S closed implies $S = \overline{S}$.

To go the other way, suppose $S = \overline{S}$. We want to show that S^c is open. Suppose $p \in S^c$. Then we know that p is not in \overline{S} . By definition, this means $\exists \delta > 0$ such that $B(p, \delta) \subset S^c$. That is what we needed to show S^c open. ■

Corollary 21 S is closed in a normed vector space $V \iff \forall$ sequence $\{x_n\}$ of points $x_n \in S$ having a limit $p = \lim_{n \rightarrow \infty} x_n$ in V , it follows that $p \in S$.

Proposition 22 A closed subset A of a complete normed vector space V is complete.

Proof. Suppose that $\{x_n\}$ is a Cauchy sequence in A . Since V is complete, we know $\exists p \in V$ such that $\lim_{n \rightarrow \infty} x_n = p$. Since A is closed, it follows from the preceding Corollary that $p \in A$. ■

It can be useful to learn the following definition.

Definition 23 A point a in the set $\overline{S} - \{a\}$ is an **accumulation point** of the set S . This means that for every $\delta > 0$, the deleted ball $B(a, \delta) - \{a\} = \{x \mid 0 < \|x - a\| < \delta\}$ contains a point of S .

A set S is closed iff it contains all its accumulation points.

Examples.

Example 1) The set \mathbb{Z} of integers has no accumulation points. For $n \in \mathbb{Z}$, one sees that $(B(n, \frac{1}{2}) - \{n\}) \cap \mathbb{Z} = \emptyset$.

Example 2) What is the set of accumulation points of \mathbb{Q} =the rationals? Answer. \mathbb{R} .

Example 3) What is the set of accumulation points of $\{\frac{1}{n} \mid n \in \mathbb{Z}^+\}$? Answer. $\{0\}$.

Example 4) What is the set of accumulation points of the open ball $B(a, r)$ in \mathbb{R}^2 using the usual norm? Answer. The closed ball $\overline{B(a, r)} = \{x \in \mathbb{R}^2 \mid \|x - a\|_2 \leq r\}$.

Another important word in the mathematician's vocabulary is "compact." I will give a definition that works for normed vector spaces but not more general spaces that mathematicians like to consider. We will not say much more about compact sets except to note that in finite dimensional normed vector spaces a set is compact iff it is closed and bounded. Here **bounded** means contained in a ball of radius r and center 0. This is false for infinite dimensional normed vector spaces, however. A ball of radius r in infinite dimensions is not compact. Very inconvenient.

Definition 24 A subset S of a normed vector space V is (sequentially) **compact** iff every sequence $\{x_n\}$ in S has a subsequence $\{x_{n_k}\}$ converging to an element of S .

See Lang, *Undergraduate Analysis* or Apostol, *Mathematical Analysis*, for more information on compact sets.

9 Limits of Functions in Normed Vector Spaces

This brings us to the story of limits of functions in and on normed vector spaces. Why are we interested in this question? We want to know when we can interchange limit and integral, limit and derivative. We want to know whether a series of functions (such as a Taylor series or a Fourier series) converges. In fact, we need to know precisely what we mean by convergence of a series of functions. We want to think of the definite integral as a function on the space $C[a, b]$ of continuous functions on a finite closed interval $[a, b]$. What sort of function is it? Linear? Continuous?

In fact, proving things about limits of functions on normed vector spaces is no harder than it was for functions on the real line. We will basically copy the proofs from those of the analogous results for functions on the real line. Suppose that V and W are normed vector spaces. I will use the same symbol for the norm on V and the norm on W . Suppose $S \subset V$ and $f : S \rightarrow W$. I will make a more general hypothesis on the nature of a and S in $\lim_{x \rightarrow a} f(x) = L$, however.

Definition 25 Suppose $S \subset V$ and $f : S \rightarrow W$. Here V and W are normed vector spaces. We will denote both norms by $\|\cdot\|$ though they may be different. Assume a is an accumulation point of S . Define the **limit of f as x approaches a** :

$$\lim_{\substack{x \rightarrow a \\ x \in S}} f(x) = \lim_{x \rightarrow a} f(x) = L \in W$$

to mean that for every $\varepsilon > 0$, there exists $\delta > 0$ (depending on ε) such that $x \in S$ and $0 < \|x - a\| < \delta$ implies $\|f(x) - L\| < \varepsilon$.

Examples.

Example 1) Let $V = C[a, b]$, the space of continuous real-valued functions on the finite interval $[a, b]$. Define $I : V \rightarrow \mathbb{R}$ by $I(f) = \int_a^b f(x) dx$. Suppose our norms are $\|f\|_1 = \int_a^b |f(x)| dx$ on V and the usual absolute value on \mathbb{R} . Does $\lim_{f \rightarrow 0} I(f) = 0$?

Answer: Yes. In fact, here $\delta = \varepsilon$, since (using the fact that integrals preserve inequalities), we see that $\|f - 0\|_1 < \varepsilon$ implies

$$|I(f) - I(0)| = \left| \int_a^b f(x) dx - 0 \right| \leq \int_a^b |f(x)| dx = \|f\|_1 < \varepsilon.$$

Example 2) Set $f(x, y) = \frac{y^2 - x^2}{y^2 + x^2}$, for $(x, y) \neq (0, 0)$ in \mathbb{R}^2 . Does $\lim_{(x, y) \rightarrow (0, 0)} f(x, y) = 0$? Here you can use any of our favorite norms on \mathbb{R}^2 and ordinary absolute value on \mathbb{R} .

Answer. No. The function $f(x, y)$ is -1 on the x -axis and $+1$ on the y -axis. Set $y = kx$. Then $f(x, kx) = \frac{k^2 - 1}{k^2 + 1}$. Thus $f(x, y)$ has different values on various lines through the origin. It follows that there is no limit as (x, y) approaches the origin. Figure 11 shows a 3D graph of the surface $z = f(x, y)$ with contours down on the x, y -plane. Along the lines (x, kx) the values are $\frac{k^2 - 1}{k^2 + 1}$. So the contours are straight lines.

It is hard to see how badly the function fails to have a limit as (x, y) approaches $(0, 0)$. Perhaps it is better just to think about what has to happen to the surface $z = f(x, y)$ if z has to be -1 on the x -axis and $+1$ on the y -axis.

There are more examples in the exercises. The properties of limits are similar to those for functions on the reals.

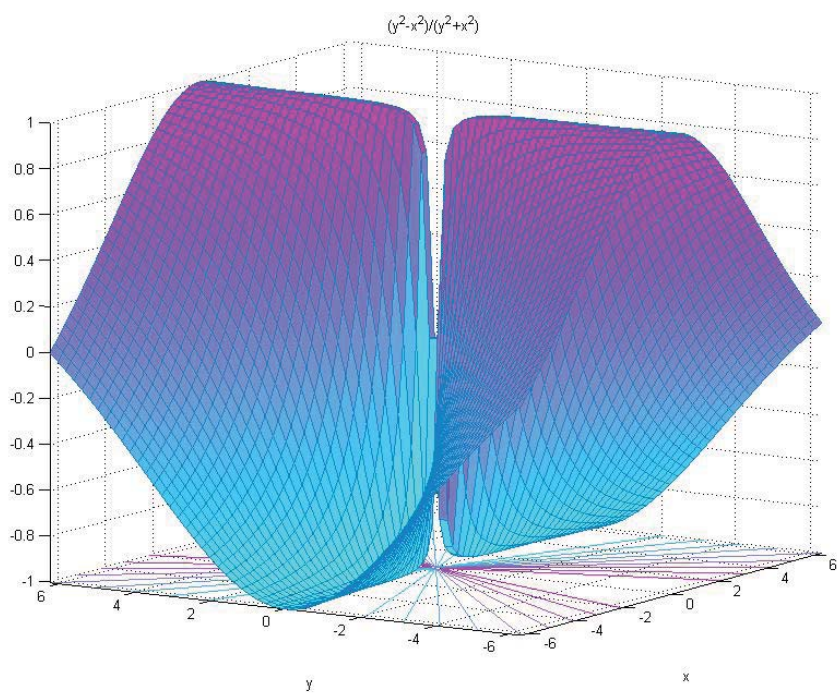


Figure 11: 3D plot of $z = f(x, y) = \frac{y^2 - x^2}{y^2 + x^2}$.

Notes on the fine points of the definition of limit.

Suppose $f : S \rightarrow W$. Before defining the limit $\lim_{x \rightarrow a} f(x)$, we need to make sure that there are points $x \in S$ which are arbitrarily close to a . If we don't assure that, then the limit could be anything. Thus we should assume minimally that $a \in \bar{S}$. Then $S \cap \{x \mid x \in V, \|x - a\| < \delta\} \neq \emptyset$ for all $\delta > 0$. This is what is done by Lang, *Undergraduate Analysis*, p. 160. However, this does not assure the non-emptiness of the intersection with the punctured ball: $S \cap \{x \mid x \in V, 0 < \|x - a\| < \delta\} \neq \emptyset$ for all $\delta > 0$. For example Lang seems to allow $S = \{a\}$ in the definition of limit. Moreover, Lang leaves out the assumption $0 < \|x - a\|$ in the definition of limit. I cannot make myself do that. See also Dieudonné, *Foundations of Modern Analysis, Volume I*. Sagan, *Advanced Calculus* and Apostol, *Mathematical Analysis*, p. 77, assume that a is an accumulation point of S . So I will do that. It won't really matter for the examples we want to consider in infinite dimensional normed vector spaces where our sets will be like $C[a, b]$.

Properties of Limits in Normed Vector Spaces.

We assume that V, W are normed vector spaces, $S \subset V$, $f, g : S \rightarrow W$ and we assume that a is an accumulation point of S .

Property 1) Uniqueness.

Suppose $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} f(x) = M$. Then $L = M$.

Property 2) Linearity.

Suppose $\alpha, \beta \in \mathbb{R}$ and $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$. Then $\lim_{x \rightarrow a} (\alpha f(x) + \beta g(x)) = \alpha L + \beta M$.

Property 3) Composite.

Suppose we have 3 normed vector spaces V, W, Z and functions $f : S \rightarrow T$, $g : T \rightarrow Z$, with the usual assumptions on the sets $S \subset V$ and $T \subset W$ as well as the points a and L . If $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{y \rightarrow L} g(y) = M$, then $\lim_{x \rightarrow a} g(f(x)) = M$.

Property 4) Inequalities.

Suppose $f, g : S \rightarrow \mathbb{R}$ (with the absolute value making \mathbb{R} a normed vector space) and $f(x) \leq g(x)$ for all $x \in S$. If $\lim_{x \rightarrow a} f(x) = L$ and $\lim_{x \rightarrow a} g(x) = M$, then $L \leq M$.

Property 5) Independence of Which Equivalent Norm is Used.

Using equivalent norms on either V or W leads to the same definition of $\lim_{x \rightarrow a} f(x) = L$.

Some Proofs.

Proof. Property 1) First we see that

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

Then, for any $\varepsilon > 0$, $\exists \delta_1$ such that $0 < \|x - a\| < \delta_1$ implies $\|L - f(x)\| < \frac{\varepsilon}{2}$. And $\exists \delta_2$ such that $0 < \|x - a\| < \delta_2$ implies $\|f(x) - M\| < \frac{\varepsilon}{2}$. Thus taking $\delta = \min\{\delta_1, \delta_2\}$ we see that $0 < \|x - a\| < \delta$ implies

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

Since ε was arbitrary this means $\|L - M\| = 0$ by a result from Lectures I, Section 11. By the first axiom of norms, this implies $L - M = 0$. ■

Proof. Property 2) First we note that

$$\begin{aligned} \|(\alpha f(x) + \beta g(x)) - (\alpha L + \beta M)\| &= \|\alpha f(x) - \alpha L + \beta g(x) - \beta M\| \\ &\leq \|\alpha f(x) - \alpha L\| + \|\beta g(x) - \beta M\| \\ &= |\alpha| \|f(x) - L\| + |\beta| \|g(x) - M\|. \end{aligned}$$

Now $\exists \delta_1$ so that $0 < \|x - a\| < \delta_1$ implies $\|f(x) - L\| < \frac{\varepsilon}{2(1+|\alpha|)}$. And $\exists \delta_2$ so that $0 < \|x - a\| < \delta_2$ implies $\|g(x) - M\| < \frac{\varepsilon}{2(1+|\beta|)}$. Then let $\delta = \min\{\delta_1, \delta_2\}$ so that $0 < \|x - a\| < \delta$ implies

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

■

Proof. Property 3) $\forall \varepsilon \exists \delta_1$ such that $0 < \|y - L\| < \delta_1$ implies $\|g(y) - M\| < \varepsilon$.

And $\exists \delta_2$ such that $0 < \|x - a\| < \delta_2$ implies $\|f(x) - L\| < \delta_1$.
 Then $0 < \|x - a\| < \delta_2$ implies, setting $y = f(x)$, $\|g(y) - M\| < \varepsilon$ since $0 < \|y - L\| < \delta_1$. ■

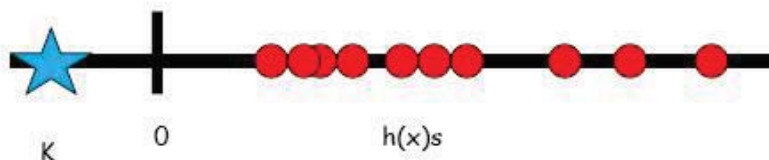


Figure 12: Picture of the proof of property 4 of limits. Here we assume $h(x) \geq 0$ and $\lim_{x \rightarrow a} h(x) = K < 0$. The $h(x)$ values are red circles to the right of 0. The limiting value K is a blue star to the left of 0. No way can this happen since the red circles can never get close to the blue star.

Proof. Property 4) Define $h(x) = g(x) - f(x)$. Then $h(x) \geq 0$ for all $x \in S$. Let $K = M - L$. If $K < 0$, we can get a contradiction. We know from Property 1 of Limits that $K = M - L = \lim_{x \rightarrow a} h(x)$. See Figure 12. The red circles are the values of $h(x)$ to the right of 0 and the blue star is K , to the left of 0. Take $\varepsilon = \frac{|K|}{2}$. Then $\exists \delta$ such that $0 < \|x - a\| < \delta$ implies $|h(x) - K| < \frac{|K|}{2}$.

Proof. Then, since K is negative, $h(x) - K < \frac{-K}{2}$ and, adding K to both sides, $h(x) < \frac{K}{2} < 0$, a contradiction to $h(x) \geq 0$. ■ ■

Exercise. Prove Property 5) of Limits.

We won't discuss limits of products in general. You can refer to Lang, *Undergraduate Analysis*, p.162-3 for the general story of limits of products. You get the joy of considering a special case in an exercise. There are lots of other cases one could look at; e.g., scalar valued function times vector valued function, matrix valued function times matrix valued function,.....

If you hate $\varepsilon\delta$ stuff, you will love the following theorem, which allows you to think about limits of sequences instead.

Theorem 26 Sequential Definition of Limits. Suppose $S \subset V$ and $f : S \rightarrow W$, where V and W are normed vector spaces. Assume that a is an accumulation point of the set S . Then the existence of $\lim_{x \rightarrow a} f(x) = L$ is equivalent to saying that for every sequence of vectors $\{x_n\}$ in S such that $\lim_{n \rightarrow \infty} x_n = a$, we have $\lim_{n \rightarrow \infty} f(x_n) = L$ exists.

Proof. \implies We leave this part as an **exercise**.

\impliedby Proof by Contradiction. Suppose $\forall \{x_n\}$ in S s.t. $\lim_{n \rightarrow \infty} x_n = a$, we have $\lim_{n \rightarrow \infty} f(x_n) = L$. If, by contradiction, $\lim_{x \rightarrow a} f(x)$ does not equal L , then (using the rules for negating a statement involving lots of $\forall \exists$), we see that $\exists \varepsilon > 0$ s.t. $\forall n \in \mathbb{Z}^+, \exists x_n \in S$ with $0 < \|x_n - a\| < \frac{1}{n}$ and $\|f(x_n) - L\| \geq \varepsilon$. Since then $\lim_{n \rightarrow \infty} x_n = a$, this is a contradiction to $\lim_{n \rightarrow \infty} f(x_n) = L$. ■

Maybe we should try to draw a picture of the definition of limit in higher dimensions. The problem is that it is hard to draw the graph of a function unless it maps a subset of the plane into the reals. Here the graph of $z = f(x, y)$ is 3-dimensional. Just plot the points $(x, y, f(x, y))$ in 3-space. Of course in the infinite dimensional case good luck drawing pictures. Even drawing the graph of a function from the plane to the plane requires 4 dimensional pictures. You can still project them down to 2 dimensions as you would in the case of a real valued function of 2 variables. Or you can make a movie of the graph being rotated.

10 Continuous Functions in Normed Vector Spaces

Suppose that V, W are normed vector spaces and S is a subset of V .

Definition 27 $f : S \rightarrow W$ is **continuous** at $c \in S$ iff for every $\varepsilon > 0$, there exists $\delta > 0$ (depending on ε) such that $x \in S$ and $\|x - c\| < \delta$ implies $\|f(x) - f(c)\| < \varepsilon$.

We see (assuming that c is an accumulation point of S) that $f : U \rightarrow W$ is **continuous** at $c \in U$ iff $\lim_{x \rightarrow c} f(x) = f(c)$.

The following theorem allows you to erase $\varepsilon\delta$ from your vocabulary. For a proof, see Apostol, *Mathematical Analysis*, p. 82.

Theorem 28 Suppose that V, W are normed vector spaces and S is a subset of V . We say that $f : S \rightarrow W$ is **continuous** at $c \in S$ iff for every open set $Y \subset W$, the inverse image $f^{-1}(Y) = \{x \in V \mid f(x) \in Y\}$ is open in V .

The definition includes some slightly crazy continuous functions. If $V = W = \mathbb{R}$ and $S = \mathbb{Z}$, every point of \mathbb{Z} is **isolated** meaning that there is a $\delta > 0$ such that $B(p, \delta) \cap \mathbb{Z} = \{p\}$. Take $\delta = \frac{1}{2}$. Thus every function $f : \mathbb{Z} \rightarrow \mathbb{R}$ is continuous.

When $V = W = \mathbb{R}$, we view continuity to mean that the graph of $y = f(x)$ does not break up at $x = c$. When $V = \mathbb{R}^2$, you can think a similar thing about the surface $z = f(x, y)$ in 3-space. But recalling Figure 11 of the function $f(x, y) = \frac{y^2 - x^2}{y^2 + x^2}$, it is even hard to see the break up for a function of 2 variables. It is easier to recall that we saw that $f(x, y)$ has a different value on various lines through the origin. It is 1 on the y -axis and -1 on the x -axis, for example.

We can use the properties of limits to deduce the following properties of continuous functions.

Properties of Continuous Functions.

Property 1) Linearity. Suppose that $f, g : U \rightarrow W$, where $U \subset V$ and V, W are normed vector spaces. Let α, β be (real) scalars. Then f and g continuous at $c \in U$ implies that $(\alpha f + \beta g)$ is continuous at c .

Property 2) Composition. Suppose that V, W, Z are normed vector spaces with $U \subset V$ and $T \subset W$. Let $c \in U$. Suppose that $f : U \rightarrow T$ and $g : T \rightarrow Z$. Suppose f is continuous at c and g is continuous at $f(c)$. Then $g \circ f$ is continuous at c .

Property 3) Sequential Definition of Continuity.

Assume V, W are normed vector spaces with $U \subset V$. The function $f : U \rightarrow W$ is continuous at $c \in U$ iff \forall sequence $\{v_n\}$ of vectors in V such that $\lim_{n \rightarrow \infty} v_n = c$, we have $\lim_{n \rightarrow \infty} f(v_n) = f(c)$.

For the proofs, you just have to look at the proofs of the corresponding properties of limits. We leave it to you as an exercise.

Examples (the same as those in the section on limits).

Example 1) Let $V = C[a, b]$, the space of continuous real-valued functions on the finite interval $[a, b]$. Define $I : V \rightarrow \mathbb{R}$ by $I(f) = \int_a^b f(x) dx$. Suppose we use $\|f\|_1 = \int_a^b |f(x)| dx$ on V and the usual absolute value on \mathbb{R} . We showed earlier that the linear function $I(f)$ is actually continuous at $f = 0$. Now I claim $I(f)$ is continuous on V ; i.e., continuous everywhere. Why? Using properties of the integral on continuous functions that we proved in Lectures I,

$$|I(f) - I(g)| = |I(f - g)| \leq I(|f - g|) = \|f - g\|_1.$$

This means that given $\varepsilon > 0$, we can take $\delta = \varepsilon$ and then $\|f - g\|_1 < \varepsilon$ implies $|I(f) - I(g)| < \varepsilon$ (which is the $\varepsilon\delta$ definition of continuity at g (or f). In fact, since δ depends only on ε and not on f or g , we have proved that the function $I(f)$ is uniformly continuous - a concept we are about to define.

Exercise. Is $I(f)$ still continuous when we replace the norm $\|f\|_1$ on V with $\|f\|_2$? Explain your answer.

Example 2) Look again at the function $f(x, y) = \frac{y^2 - x^2}{y^2 + x^2}$, for $(x, y) \neq (0, 0)$ in \mathbb{R}^2 . We know that this function cannot be continuous at $(0, 0)$ since it has no limit as $(x, y) \rightarrow (0, 0)$. This function is continuous at every other point though. There are more such examples in the exercises.

Definition 29 Suppose that V, W are normed vector spaces and $U \subset V$. We say that $f : U \rightarrow W$ is **uniformly continuous** on U iff $\forall \varepsilon > 0 \exists \delta > 0$ (with δ depending only on ε) such that $\forall u, v \in U, \|u - v\| < \delta$ implies $\|f(u) - f(v)\| < \varepsilon$.

The point in this definition is that δ does **not** depend on $u, v \in U$.

Example 1 just considered is an example of a uniformly continuous function where in fact $\delta = \varepsilon$. We get lots more examples using the following theorem.

Theorem 30 Suppose K is a compact subset of a normed vector space V and W is any normed vector space. A continuous function $f : K \rightarrow W$ must be uniformly continuous on K .

For a proof of the preceding theorem, see p. 198 of Lang, *Undergraduate Analysis*.

More Examples.

Example 3) V -normed vector space. Let $f(x) = \|x\|$. Then f is uniformly continuous on V . For all $x, y \in V$, the triangle inequality implies (as it did for ordinary absolute value in an early homework problem from part 1),

$$\left| \|x\| - \|y\| \right| \leq \|x - y\|.$$

This says we can take $\varepsilon = \delta$ again.

Example 4) Suppose that $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear. Then take e_j to be the standard basis vector $e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, with a 1

in the j th row and the rest of the entries being 0. Every vector $v \in \mathbb{R}^n$, can be written uniquely in the form:

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_{j-1} \\ v_j \\ \vdots \\ v_n \end{pmatrix} = \sum_{j=1}^n v_j e_j.$$

It follows from linearity of L , that

$$Lv = \sum_{j=1}^n v_j L e_j.$$

Write $L e_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{j-1,j} \\ a_{jj} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{R}^m$. So we see that

$$Lv = \sum_{j=1}^n v_j \begin{pmatrix} a_{1j} \\ \vdots \\ a_{j-1,j} \\ a_{jj} \\ \vdots \\ a_{mj} \end{pmatrix} = Av,$$

where we multiply the matrix A whose entries are a_{ij} with the vector v . As an **exercise**, show that the linear function L is uniformly continuous. You can see this by using the infinity norm on \mathbb{R}^n and \mathbb{R}^m and showing that there is a constant $C > 0$ so that $\|Lx\|_\infty \leq C \|x\|_\infty$. The constant C depends on the entries a_{ij} of the matrix A . If you take the $K = \max |a_{ij}|$, then $C = nK$ should work.

When a normed vector space V is not complete, one can always obtain a larger complete space W containing V . This is done by taking all the Cauchy sequences $\{x_n\}$ of elements $x_n \in V$ modulo the equivalence relation $\{x_n\} \simeq \{y_n\}$ iff $\|x_n - y_n\| \rightarrow 0$ as $n \rightarrow \infty$. This larger space W is then called the **completion** of V . See Lang, *Undergraduate Analysis* for more information on completions.

11 Completeness of $C[a, b]$ with respect to the ∞ Norm.

We promised earlier to prove the following theorem.

Theorem 31 *The normed vector space $C[a, b]$ of continuous real valued functions on the finite interval $[a, b]$ is complete with respect to the norm $\|f\|_\infty = \max_{x \in [a, b]} |f(x)|$.*

Proof. Recall that V "complete" means every Cauchy sequence in V converges to an element of V . So let $\{f_n\}$ be a Cauchy sequence in $C[a, b]$ using the norm $\|f\|_\infty$. This means for every $x \in [a, b]$, the sequence $\{f_n(x)\}$ of real numbers is Cauchy; as $\forall \varepsilon \exists N_\varepsilon$ s.t.

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty < \varepsilon \quad \text{when } n, m \geq N_\varepsilon. \quad (1)$$

We showed in Lectures I that Cauchy sequences of real numbers converge to a limit in \mathbb{R} . Thus $\forall x \in [a, b]$ there is a function $f(x) = \lim_{n \rightarrow \infty} f_n(x)$. Now we need to show that f_n converges uniformly to f on $[a, b]$.

Let $\varepsilon > 0$ be given. There is $M = M(x, \varepsilon) \geq N_\varepsilon$ so that $m \geq M$ implies $|f_m(x) - f(x)| < \varepsilon$. Then for $n \geq N_\varepsilon$ we have the following sneaky formula by adding and subtracting $f_m(x)$ and using the triangle inequality:

$$|f(x) - f_n(x)| \leq |f(x) - f_m(x)| + |f_m(x) - f_n(x)| < \varepsilon + \|f_n - f_m\|_\infty < 2\varepsilon. \quad (2)$$

We chose N_ε so that formula (1) holds. This implies $\|f - f_n\|_\infty < 2\varepsilon$ for $n \geq N_\varepsilon$ which is uniform convergence of f_n to f on $[a, b]$ since N_ε does not depend on x .

Next we need to show that f is continuous on $[a, b]$. To see this, note that for $x, y \in [a, b]$, using the triangle inequality in a sneaky way again (this time adding and subtracting $f_n(x) - f_n(y)$):

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x) + f_n(x) - f_n(y) + f_n(y) - f(y)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)|. \end{aligned}$$

We know that for $n \geq N_\varepsilon$ the 1st and 3rd terms are $< 2\varepsilon$ by formula (2). Since f_n is continuous, there is a positive δ , depending on n, ε and y such that $|x - y| < \delta$ implies the middle term is also $< \varepsilon$. So the final result is that $|f(x) - f(y)| < 5\varepsilon$, if $|x - y| < \delta$. Replace ε by $\varepsilon/5$, if you like. ■

Part II

Series in a Normed Vector Space

12 The Basics

In this section I will be more sketchy than usual. Hope that is OK and that you remember some of the series part of calculus. We are mostly interested in series of functions like power series and Fourier series. That is, we are interested in series in a normed vector space like $C[a, b]$ that is infinite dimensional. And we want to know: can we interchange limit and \sum , derivative and \sum , integral and \sum ? The answer will depend on the norm used. Before doing the infinite dimensional theory of series we'd better review the theory of series of real numbers.

First define what we mean by convergence of a series in a normed vector space; i.e., convergence of the sequence of partial sums.

Definition 32 Suppose V is a normed vector space. Let $\{v_n\}$ be a sequence of vectors in V . Then we say the series $\sum_{n=1}^{\infty} v_n$ **converges** to s and write $\sum_{n=1}^{\infty} v_n = s$ iff $s = \lim_{n \rightarrow \infty} \sum_{k=1}^n v_k$.

So if $\sum_{n=1}^{\infty} v_n = s$, the sequence of **partial sums** $s_n = \sum_{k=1}^n v_k$ is a Cauchy sequence, which means $\forall \varepsilon \exists N$ s.t. $\|s_n - s_m\| < \varepsilon$ if $n, m \geq N$. But if we assume $n > m \geq N$, then $s_n - s_m = \sum_{k=1}^n v_k - \sum_{k=1}^m v_k = \sum_{k=m+1}^n v_k$. Thus we have $\left\| \sum_{k=m+1}^n v_k \right\| < \varepsilon$. Now take $n = m+1 \geq N$. Then we see that $\|v_n\| < \varepsilon$. We have proved the following necessary condition for convergence of a series.

Necessary but NOT Sufficient Condition for Convergence. The series $\sum_{n=1}^{\infty} v_n$ converges implies that the terms approach 0; i.e. $\lim_{n \rightarrow \infty} v_n = 0$.

This condition is not sufficient for convergence. The integral test will give us many examples (e.g., the harmonic series).

13 Tests for Convergence of Series of Non-Negative Real Numbers.

Test 1. Suppose $a_n \geq 0$ for all $n \in \mathbb{Z}^+$. Then the series $\sum_{n=1}^{\infty} a_n$ converges iff the partial sums $s_n = \sum_{k=1}^n a_k$ are bounded.

Proof. The partial sums form an increasing sequence. We know from Lectures I that an increasing sequence s_n converges iff it is bounded and then it converges to the least upper bound of the set $\{s_1, s_2, s_3, \dots\}$. ■

Test 2. Comparison Test. Suppose $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are series of real numbers and there exists a positive constant C (independent of n) such that $0 \leq a_n \leq C b_n$ for all $n \geq N$.

- a) If $\sum_{n=1}^{\infty} b_n$ converges, then so does $\sum_{n=1}^{\infty} a_n$.
- b) If $\sum_{n=1}^{\infty} a_n$ diverges, then so does $\sum_{n=1}^{\infty} b_n$.

Proof. You can ignore all the a_n for $n < N$. They cannot affect the convergence of $\sum_{n=1}^{\infty} a_n$. Explain why as an **exercise**.

a) So we will just assume the first N terms a_n are 0. Then we can show that the partial sums $\sum_{k=1}^n a_k$ are bounded and use Test 1 to complete the proof. For we have

$$\sum_{k=1}^n a_k \leq C \sum_{k=1}^n b_k \leq C \sum_{k=1}^{\infty} b_k.$$

b) **Exercise.** ■

Test 3. Ratio Test. Suppose that $0 \leq a_n$.

- a) If \exists a positive constant C such that $0 < C < 1$ and $a_{n+1} \leq C a_n$, for all $n \geq N$, then $\sum_{n=1}^{\infty} a_n$ converges.
- b) If \exists a positive constant C such that $C > 1$ and $a_{n+1} \geq C a_n$, for all $n \geq N$, then $\sum_{n=1}^{\infty} a_n$ diverges.

Proof. a) Compare with the geometric series $\sum_{n=0}^{\infty} C^n = \frac{1}{1-C}$. We can, as before, ignore the first N terms of the series.

Then, by induction, we can show that $a_{N+k} \leq C^k a_N$, for $k = 1, 2, \dots$. Again you can do this as an **exercise**. The comparison test will give the conclusion of part a).

b) Again compare with the geometric series. Once more, ignore the first N terms and obtain by induction $a_{N+k} \geq C^k a_N$, for $k = 1, 2, \dots$. The comparison test will then imply the conclusion of part b) since for $C > 1$, $\sum_{n=0}^{\infty} C^n$ diverges. In fact, the terms C^n do not approach 0. ■

Question: Where is the ratio in the ratio test? It is concealed in the hypothesis: $a_{n+1} \leq C a_n$. Divide by a_n if $a_n \neq 0$ and you have a ratio.

Exercise. State and prove the n th root test.

Test 4. The Integral Test.

Suppose $0 \leq f(x)$ for all $x \geq 1$ and that $f(x)$ is monotone decreasing and continuous. Then

$$\sum_{n=1}^{\infty} f(n) \text{ converges iff } \int_1^{\infty} f(x) dx \text{ converges.}$$

Proof. Draw a picture.

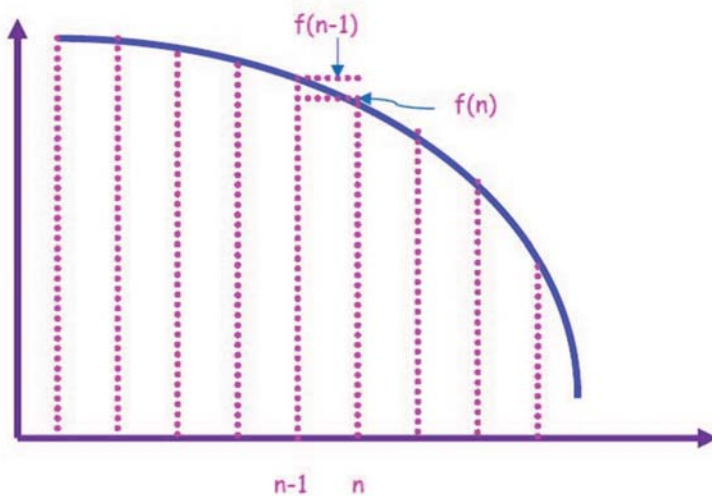


Figure 13: a picture proof of the integral test

From Figure 13 you see that

$$f(n) \leq \int_{n-1}^n f(x) dx \leq f(n-1).$$

Add these inequalities up from $n = 2$ to N and obtain

$$\sum_{n=2}^N f(n) \leq \int_1^N f(x) dx \leq \sum_{n=1}^{N-1} f(n).$$

We let you finish the argument as an **exercise**. ■

There is yet one more such test - the n th root test. We leave this one to you to remember or Google.

Example.

My favorite function is the **Riemann zeta function**.

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

The integral test says that this converges if $s > 1$ and diverges if $s \leq 1$.

The case $s = 1$ is the divergent **harmonic series**. The cases $s = 2n$ an even integer > 1 were evaluated by Euler as rational numbers times π^{2n} . Some have wasted much time thinking about the case of $s = 3$ and the like producing some horrendous formulas but not anything like Euler's formula.

Riemann wrote a paper about 1850 showing how to make sense of $\zeta(s)$ for all complex numbers s except for $s = 1$. Why was Riemann interested in this? He was interested in the distribution of prime numbers 2, 3, 5, 7, 11, Euler had showed the product formula that for $\text{Re } s > 1$,

$$\zeta(s) = \prod_{p=\text{prime}} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Here we define the infinite product as the limit of the partial products. We won't prove this formula. It relies on the geometric series and the fundamental theorem of arithmetic (every $n \geq 1$ is a product of powers of primes and the product is unique up to order).

Thanks to complex analysis, it turns out that the location of the complex numbers s such that $\zeta(s) = 0$ gives information about the distribution of primes. About 50 years after Riemann's work Hadamard and de la Vallée Poussin used what was known about the location of zeta zeros to show the **prime number theorem** which says that the number of primes $\leq x$ is asymptotic to $\frac{x}{\log x}$ as $x \rightarrow \infty$.

Riemann stated the Riemann hypothesis saying that the non-real zeros of $\zeta(s)$ must have $\text{Re } s = \frac{1}{2}$. You win 1 million dollars from the Clay Math. Institute if you can prove it. However, I think there was a deadline for the proof and that deadline has passed. A reference for zeta is Edwards, *Riemann's Zeta Function*.

14 Non-Absolute Convergence

Theorem 33 The Alternating Series Test. Suppose $0 \leq a_n \leq a_{n-1}$ for all n and $\lim_{n \rightarrow \infty} a_n = 0$. Then the series $\sum_{n=1}^{\infty} (-1)^{n-1} a_n$ converges. This sort of convergence is called **conditional convergence** if $\sum_{n=1}^{\infty} a_n$ diverges.

Proof. Consider the even and odd partial sums separately.

Case 1. The Odd partial sums

$$s_{2n-1} = a_1 - (a_2 - a_3) - \cdots - (a_{2n-2} - a_{2n-1})$$

$$\begin{aligned} s_{2n+1} &= a_1 - (a_2 - a_3) - \cdots - (a_{2n-2} - a_{2n-1}) - (a_{2n} - a_{2n+1}) \\ &= s_{2n-1} - (a_{2n} - a_{2n+1}) \\ &\leq s_{2n-1} \end{aligned}$$

This implies (since the terms a_n are decreasing) that the **odd partial sums are decreasing**.

Case 2. The even partial sums

$$s_{2n} = (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2n-1} - a_{2n}).$$

$$\begin{aligned} s_{2n+2} &= (a_1 - a_2) + (a_3 - a_4) + \cdots + (a_{2n-1} - a_{2n}) + (a_{2n+1} - a_{2n+2}) \\ &= s_{2n} + (a_{2n+1} - a_{2n+2}) \\ &\geq s_{2n}. \end{aligned}$$

This says the **even partial sums are increasing**.

We also know the **relation between even and odd partial sums**, given by

$$s_{2n} = s_{2n-1} - a_{2n} \leq s_{2n-1}$$

So the picture of the partial sums is

$$s_2 \leq s_4 \leq \dots \leq s_{2n} \leq s_{2n-1} \leq \dots \leq s_3 \leq s_1.$$

The even partial sums are an increasing sequence of real numbers which is bounded above and thus they have a limit

$$L = \lim_{n \rightarrow \infty} s_{2n}.$$

using stuff from Lectures I (the existence of the *l.u.b.*). The odd partial sums are a decreasing sequence bounded below and thus have a limit

$$M = \lim_{n \rightarrow \infty} s_{2n+1}.$$

To see that

$$L = M = \lim_{n \rightarrow \infty} s_n,$$

use the relation between even and odd partial sums $s_{2n} = s_{2n-1} - a_{2n}$ and the hypothesis that $\lim_{n \rightarrow \infty} a_n = 0$ to obtain (using properties of limits):

$$L = \lim_{n \rightarrow \infty} s_{2n} = \lim_{n \rightarrow \infty} (s_{2n-1} - a_{2n}) = \lim_{n \rightarrow \infty} s_{2n-1} - \lim_{n \rightarrow \infty} a_{2n} = \lim_{n \rightarrow \infty} s_{2n-1} - 0 = \lim_{n \rightarrow \infty} s_{2n-1} = M.$$

■
Example. $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$ converges by the alternating series test. With some effort, using the Taylor series for $\log(1-x) = \sum_{n=1}^{\infty} \frac{x^n}{n}$, which has radius of convergence 1, it is possible to show that

$$\log 2 = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + - \dots$$

If one rearranges this series to sum one odd then two evens:

$$s = 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + - \dots,$$

with a little algebra, one can show that $s = \frac{\log 2}{2}$. That is, we get half the sum of the original series with alternating signs.

Theorem 34 Riemann's Rearrangement Theorem. You can rearrange a series that converges conditionally to make it converge to anything or diverge to ∞ or $-\infty$. Rearrangement of the series $\sum_{n=1}^{\infty} b_n$ means take a 1-1, onto map σ of the positive integers onto the positive integers and look at the series $\sum_{n=1}^{\infty} b_{\sigma(n)}$.

This moral of this theorem is that conditionally or non-absolutely convergent series are quite nasty. Absolute convergence of $\sum_{n=1}^{\infty} v_n$ means $\sum_{n=1}^{\infty} \|v_n\|$ converges in \mathbb{R} and will be discussed in the next section. For a proof of Riemann's Rearrangement Theorem, see Sagan, *Advanced Calculus*.

15 Absolute Convergence in Normed Vector Spaces

Suppose that V is a complete normed vector space (recall complete means contains limits of all Cauchy sequences which must converge to limits in V).

Theorem 35 *Absolute Convergence \implies Convergence in a Complete Normed Vector Space.*

Suppose V is a complete normed vector space and $v_n \in V$ for all n . Then

$$\sum_{n=1}^{\infty} \|v_n\| \text{ converges in } \mathbb{R} \text{ implies } \sum_{n=1}^{\infty} v_n \text{ converges to a limit } s \text{ in } V.$$

This is called **absolute convergence** of $\sum_{n=1}^{\infty} v_n$.

Proof. We will show that the sequence $s_n = \sum_{k=1}^n v_k$ of partial sums is Cauchy and then completeness of V says that it must converge to a limit $s \in V$. To see this, assume $n > m$ and use the triangle inequality to obtain

$$\|s_n - s_m\| = \left\| \sum_{k=m+1}^n v_k \right\| \leq \sum_{k=m+1}^n \|v_k\|.$$

We can choose N to make the last sum $< \varepsilon$ for $n > m \geq N$ by the hypothesis that $\sum_{n=1}^{\infty} \|v_n\|$ converges in \mathbb{R} . ■

Example. Suppose X is an $m \times m$ real matrix. Define $e^X = \exp(X) = \sum_{n=0}^{\infty} \frac{1}{n!} X^n$. Use the operator norm on X defined by

$$\|X\| = \max \{ \|Xv\| \mid v \in \mathbb{R}^m, \|v\| = 1 \}.$$

Here $\|v\|$ denotes any of our favorite norms on \mathbb{R}^m . We know we can say *max* rather than *l.u.b.* or *sup* by Theorem 2.2, p. 197, of Lang, *Undergraduate Analysis*, because the unit sphere in \mathbb{R}^m is compact and the function on \mathbb{R}^m produced by taking vector v to Xv is continuous linear. More about matrix norms can be found in Strang, *Linear Algebra and its Applications*, Chapter 7 or Horn and Johnson, *Matrix Analysis*, Chapter 5.

Then one can show that

$$\|XY\| \leq \|X\| \|Y\|.$$

Since $\sum_{n=0}^{\infty} \frac{1}{n!} \|X\|^n$ always converges, we know that the matrix exponential series always converges.

Theorem 36 *If $\sum_{n=1}^{\infty} v_n$ converges absolutely then any rearrangement converges to the same limit.*

Proof. See Lang, *Undergraduate Analysis*, p. 227. ■

For the following convergence test recall that $C[a, b]$ is the space of continuous functions on a finite closed interval $[a, b]$ and that this space is complete using the norm $\|f\|_{\infty} = \max_{X \in [A, B]} |f(x)|$. Recall that we needed this next test in the final exam from Lectures I when we looked at the fractal nature of the Weierstrass nowhere differentiable continuous function. This test gave the continuity.

Theorem 37 *Weierstrass M-Test.*

Theorem 38 *Let $f_n \in C[a, b]$. Suppose $\|f_n\|_{\infty} \leq M_n$ for all $n = 1, 2, 3, \dots$ and suppose that $\sum_{n=1}^{\infty} M_n$ converges in \mathbb{R} .*

Then $\sum_{n=1}^{\infty} f_n$ converges in the $\|\cdot\|_{\infty}$ norm (i.e., uniformly) to a continuous function $f \in C[a, b]$.

Proof. We make our usual argument to see that the sequence of partial sums of $\sum_{n=1}^{\infty} f_n$ is Cauchy with respect to the $\|\cdot\|_{\infty}$ norm. Then completeness of $C[a, b]$ completes the proof that $\sum_{n=1}^{\infty} f_n$ converges to a continuous function f .

That is, the n th partial sum $s_n = \sum_{k=1}^n f_k$ implies $s_n - s_m = \sum_{k=1}^n f_k - \sum_{k=1}^m f_k = \sum_{k=m+1}^n f_k$. By the triangle inequality and the hypothesis of the Weierstrass M-Test, $\forall \varepsilon > 0, \exists N$ s.t. $n > m \geq N$ implies

$$\left\| \sum_{k=m+1}^n f_k \right\| \leq \sum_{k=m+1}^n \|f_k\| \leq \sum_{k=m+1}^n M_k < \varepsilon.$$

It follows that the sequence s_n is Cauchy and the series $\sum_{n=1}^{\infty} f_n$ converges uniformly as we have proved that $C[a, b]$ is complete in the $\|\cdot\|_{\infty}$ -norm. ■

Example 1. $\sum_{n=1}^{\infty} \frac{\sin(nx)}{n^2}$ converges uniformly and absolutely on any interval $[a, b]$.

Here use the M-Test with $M_n = \frac{1}{n^2}$. The series $\sum_{n=1}^{\infty} M_n = \sum_{n=1}^{\infty} \frac{1}{n^2}$ converges by the integral test.

If we replace $\frac{1}{n^2}$ with $\frac{1}{n}$, the question of convergence is more delicate. We will think about it when we look at Fourier series later.

Example 2. Let $0 < c < 1$. Then $\sum_{n=0}^{\infty} x^n$ converges uniformly and absolutely to $\frac{1}{1-x}$ for all $x \in [-c, c]$.

Here use the M-Test with $M_n = c^n$. The series $\sum_{n=0}^{\infty} M_n = \sum_{n=0}^{\infty} c^n = \frac{1}{1-c}$ (using the geometric series).

Example 3. Integral Operators on $C[0, 1] = V$.

Let $\|f\|$ denote any of our 3 favorite norms on $C[0, 1]$. Suppose that $K(x, y)$ is a continuous function of 2 variables for $x, y \in [0, 1]$. As an example, consider $\sin(xy)$. Define the **integral operator** L with kernel K to be

$$Lf(x) = \int_0^1 K(x, y)f(y)dy.$$

Then $L : V \rightarrow V$ is linear and can be shown to be continuous. We can define a norm on the space $\mathcal{L}(V, V)$ of such operators L to be $\|L\| = \text{lub} \{\|Lf\| \mid f \in V, \|f\| = 1\}$. This is called the **operator norm**.

One has an **operator-valued geometric series** for $\lambda \in \mathbb{R}$: $\sum_{n=0}^{\infty} \lambda^n L^n = (I - \lambda L)^{-1}$. One can show, using properties of the operator norm that this series converges when $\|\lambda L\| = |\lambda| \|L\| < 1$. This fact is quite useful when considering the spectral theory of integral and differential operators. See Courant and Hilbert, *Methods of Mathematical Physics, Vol. I* and Stakgold, *Green's Functions and Boundary Value Problems* for more information

Part III

Power Series

16 Radius of Convergence

Here we consider convergence and differentiation and integration of power series $\sum_{n=0}^{\infty} a_n(x-p)^n$. I will mostly assume $p = 0$ for simplicity. We consider $x \in \mathbb{R}$ mostly. The same arguments should work for complex numbers and matrices, with a

little thought.

Theorem 39 a) Suppose $\{a_n\}$ is a sequence of real numbers and $\sum_{n=0}^{\infty} a_n x^n$ converges for some $x \neq 0$. Then $\sum_{n=0}^{\infty} a_n u^n$ converges absolutely for all u such that $|u| < |x|$.

b) If $\sum_{n=0}^{\infty} a_n x^n$ diverges, then $\sum_{n=0}^{\infty} a_n u^n$ diverges if $|u| > |x|$.

Proof. a) $\sum_{n=0}^{\infty} a_n x^n$ converges implies $\lim_{n \rightarrow \infty} a_n x^n = 0$. Since convergent implies bounded, we have $|a_n x^n| \leq M$ for all n . This implies if $|u| < |x|$, we can apply the comparison test by writing

$$|a_n u^n| = \left| a_n x^n \frac{u^n}{x^n} \right| \leq M \left| \frac{u^n}{x^n} \right|.$$

So we can compare the series $\sum_{n=0}^{\infty} |a_n u^n|$ with the convergent geometric series $\sum_{n=0}^{\infty} \left| \frac{u^n}{x^n} \right|$.

We leave the proof of b) as an **exercise**. Hint: Use part a). ■

We are calling the following positive number the radius of convergence because we (the definers) are really thinking of power series in the complex plane where we actually get an open circle of radius R where the power series converges. On the real line where this course mostly lives, we only get an interval.

Definition 40 Now define the **radius of convergence** R of $\sum_{n=0}^{\infty} a_n x^n$

$$R = \text{l.u.b.} \left\{ r \mid r \geq 0, \sum_{n=0}^{\infty} |a_n| r^n \text{ converges} \right\}.$$

I am using the convention here that \mathbb{R} (blackboard bold face font) is the set of real numbers and should not be confused with ordinary capital R . This convention comes to us from Nicolas Bourbaki (the unreal mathematician).

Corollary 41 Let S denote the set of real numbers x such that $\sum_{n=0}^{\infty} a_n x^n$ converges. Then S must be of the following forms: $\{0\}$, $(-R, R)$, $(-R, R]$, $[-R, R)$, $[-R, R]$, or $(-\infty, \infty) = \mathbb{R}$.

Theorem 42 Formulas for the Radius of Convergence R of a Power Series $\sum_{n=0}^{\infty} a_n x^n$.

We assume $a_n \neq 0, \forall n$.

Formula 1) Assume the limit $\lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right|$ exists. Then it is the radius of convergence R for $\sum_{n=0}^{\infty} a_n x^n$. That is

$$R = \lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right|.$$

Formula 2) Again, assuming the limit involved exists, we have the formula:

$$\frac{1}{R} = \lim_{n \rightarrow \infty} |a_n|^{\frac{1}{n}}.$$

Proof. Formula 1) Use the ratio test for absolute convergence of $\sum_{n=0}^{\infty} a_n x^n$. Setting $\lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right| = c$, we see that the ratio of the terms in the power series approaches a limit:

$$\left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| = \left| \frac{a_{n+1}}{a_n} x \right| \rightarrow \frac{|x|}{c}, \text{ as } n \rightarrow \infty$$

The ratio test proved above says the series converges if $|x| < c$ and diverges if $|x| > c$. **Why? (Exercise)**. So $c = R$, the radius of convergence. This comes from the definition of radius of convergence as a least upper bound. If $|x| > R$, the series diverges and if $|x| < R$, the series converges. That is what we needed to show.

Formula 2) We leave this to the reader as an **Exercise**. ■

Example 1). Our friend the **geometric series** $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$. The ratio of convergence is 1 by any of the formulas since $a_n = 1$, for all n .

Example 2). Our best friend the **exponential series** $\sum_{n=0}^{\infty} \frac{1}{n!} x^n = e^x$. Use the first formula for the radius of convergence. Obtain

$$R = \lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right| = R = \lim_{n \rightarrow \infty} \left| \frac{\frac{1}{n!}}{\frac{1}{(n+1)!}} \right| = \lim_{n \rightarrow \infty} \left| \frac{(n+1)!}{n!} \right| \lim_{n \rightarrow \infty} \left| \frac{(n+1)}{1} \right| = \infty.$$

This means the series converges everywhere.

The nice thing about this example is that the exponential series also converges for all complex numbers and for all $n \times n$ matrices. This gives a nice way to get the power series for sine and cosine out of Euler's formula

$$e^{i\theta} = \cos \theta + i \sin \theta.$$

Example 3). Here's a bad one: $\sum_{n=0}^{\infty} n! x^n$.

The radius of convergence is the reciprocal of that for Example 2. Thus you get $R = 0$. This series only converges at $x = 0$.

Exercise. Replace \mathbb{R} in Examples 1 and 2 with the space $\mathbb{R}^{n \times n}$ of $n \times n$ real matrices. Obtain the matrix geometric series and the matrix exponential. Where do they converge? Use the definition $\|L\| = \text{lub} \{ \|Lv\| \mid v \in \mathbb{R}^n, \|v\| = 1 \}$ to get a norm on the space of matrices. See Apostol, *Calculus, Vol. II* for many applications of the matrix exponential to differential equations.

17 Differentiation and Integration of Power Series

We want to legally interchange \int_a^b and limit (or $\sum_{n=0}^{\infty}$) so that we can integrate a power series term-by-term at least if we

stay inside the radius of convergence. Similarly we want to legally interchange $\frac{d}{dx}$ and limit (or $\sum_{n=0}^{\infty}$),

Example.

Integrate the geometric series to get the power series for $\log(1-x)$. This is legal if $|x| < 1$ using the corollary of the following theorem.

$$-\log(1-x) = \int_0^x \frac{1}{1-t} dt = \int_0^x \sum_{n=0}^{\infty} t^n dt = \sum_{n=0}^{\infty} \int_0^x t^n dt = \sum_{n=0}^{\infty} \frac{t^{n+1}}{n+1} \Big|_0^x = \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1}.$$

Note that the formula makes sense if $x = -1$, although the following theorem and corollary do not justify the equality at -1 .

Theorem 43 (Interchange of limit and integral). Suppose that $f_n : [a, b] \rightarrow \mathbb{R}$ is continuous $\forall n$ and converges uniformly on $[a, b]$ to f . Then

$$\lim_{n \rightarrow \infty} \int_a^b f_n = \int_a^b \lim_{n \rightarrow \infty} f_n = \int_a^b f.$$

Proof. We already know that f must be continuous on $[a, b]$. So we can integrate it. See Lectures, I. Using properties of the integral from Lectures I, we find that

$$\left| \int_a^b f - \int_a^b f_n \right| = \left| \int_a^b (f - f_n) \right| \leq \int_a^b |f - f_n| \leq \int_a^b \|f - f_n\|_\infty = \|f - f_n\|_\infty (b - a).$$

It follows that given $\varepsilon > 0$, if we take N so that $n \geq N$, implies $\|f - f_n\|_\infty < \frac{\varepsilon}{b-a}$ (as we can by uniform convergence), then we have

$$\left| \int_a^b f - \int_a^b f_n \right| < \varepsilon.$$

■

As a corollary, we see that it is legal to integrate a series of functions term-by-term on a finite interval if the series converges uniformly on that interval.

Corollary 44 Suppose $g_n : [a, b] \rightarrow \mathbb{R}$ is continuous $\forall n$ and $\sum_{n=0}^{\infty} g_n(x)$ converges uniformly on $[a, b]$ to $s(x)$. Then $s(x)$ is continuous on $[a, b]$ and

$$\sum_{n=0}^{\infty} \int_a^b g_n = \int_a^b \sum_{n=0}^{\infty} g_n = \int_a^b s.$$

Proof. Exercise. ■

The proof of the last theorem was essentially one line. But differentiation requires more effort, as well as more hypotheses. Why should that be?

Example. Consider $f_n(x) = \frac{\sin(nx)}{\sqrt{n}}$. This sequence converges uniformly to 0 on $[-\pi, \pi]$. Why? However,

$$f'_n(x) = \sqrt{n} \cos(nx).$$

This derivative has problems converging at all, much less to 0. For example, it diverges at $x = 0$. In fact, it can be shown to diverge everywhere. Suppose $x \neq 0$. You just need to see that if $x \neq 0$, $\exists N$ such that $|\cos(Nx)| < \frac{1}{2}$ and thus

$$|\cos(2Nx)| = |2\cos^2(Nx) - 1| = 1 - 2\cos^2(Nx) > \frac{1}{2}.$$

It follows that there is a subsequence satisfying $|\sqrt{n_k} \cos(n_k x)| > \frac{1}{2} \sqrt{n_k}$ and thus diverging.

Exercise. Find an example to show that we need uniform convergence of f_n to f in the hypothesis of the preceding theorem. Pointwise convergence at each x in $[a, b]$ does not suffice for the interchange of integral and limit.

Hint. Look at $f_n(x) = n^2 x(1-x)^n$, on $[0, 1]$.

Theorem 45 (Interchange of Derivative and Limit). Suppose $\{f_n\}$ is a sequence of continuously differentiable functions on $[a, b]$. Suppose, also that the sequence of derivatives $\{f'_n\}$ converges uniformly to g on $[a, b]$. And, finally assume there exists one point $x_0 \in [a, b]$ such that $f_n(x_0)$ converges.

Then

$$g(x) = \lim_{n \rightarrow \infty} \frac{d}{dx} f_n(x) = \frac{d}{dx} \lim_{n \rightarrow \infty} f_n(x).$$

and $f_n(x)$ converges uniformly to $f(x)$, with $f'(x) = g(x)$, for all $x \in [a, b]$.

Proof. By the fundamental theorem of calculus from Lectures I, $\forall n, \exists c_n$ s.t.

$$f_n(x) = \int_a^x f'_n + c_n. \quad (3)$$

Here $c_n = f_n(a)$.

Let $x = x_0$ and take the limit of formula (3) as $n \rightarrow \infty$. This says that $\exists c = \lim_{n \rightarrow \infty} c_n$.

Next take the limit of formula (3) for general $x \in [a, b]$ as $n \rightarrow \infty$ using the preceding Theorem (which we can since the sequence of derivatives converges uniformly). This says, the sequence $\{f_n\}$ converges pointwise (i.e., for each fixed x in $[a, b]$) to

$$f(x) = \int_a^x g + c.$$

To finish the proof of this theorem, we just need to see $\{f_n\}$ converges uniformly to f . Well, try this, using our favorite triangle inequality and properties of integrals from Lectures I,

$$\begin{aligned} |f_n(x) - f(x)| &= \left| \int_a^x f'_n + c_n - \int_a^x g - c \right| \\ &\leq \left| \int_a^x f'_n - g \right| + |c_n - c| \leq \int_a^x |f'_n - g| + |c_n - c| \\ &\leq (b-a) \|f'_n - g\|_\infty + |c_n - c|. \end{aligned}$$

So if you insist on giving me an $\varepsilon > 0$, I can certainly find an N (depending only on ε and not on x) such that $n > N$ implies $|f_n(x) - f| < \varepsilon$. That is the meaning of uniform convergence. ■

Corollary. Suppose $g_n(x) [a, b] \rightarrow \mathbb{R}$ is continuously differentiable, $\forall n$ and $\sum_{n=0}^{\infty} g'_n(x)$ converges uniformly on $[a, b]$ to $s(x)$. And assume there is one point $x_0 \in [a, b]$ such that $\sum_{n=0}^{\infty} g_n(x_0)$ converges. Then we have $\sum_{n=0}^{\infty} g_n(x)$ converges to $r(x)$ uniformly on $[a, b]$ and $s(x) = r'(x)$; i.e.,

$$\sum_{n=0}^{\infty} \frac{d}{dx} g_n(x) = \frac{d}{dx} \sum_{n=0}^{\infty} g_n(x).$$

Proof. Exercise. Deduce this corollary from the preceding theorem. ■

Examples. For our power series, these 2 theorems, allow us to integrate and differentiate term by term on closed subintervals of $(-R, R)$, where R is the radius of convergence. I find this somewhat surprising, since the power series you get by differentiation term-by-term blows up the term by a factor of n :

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

It is also a bit surprising that integration does not produce a series that converges in a larger region since

$$\int_0^x a_n t^n dt = a_n \frac{x^{n+1}}{n+1}, \quad \text{for } n \geq 0.$$

Theorem 46 Suppose $f(x) = \sum_{n=0}^{\infty} a_n x^n$ with radius of convergence R . Then for x in $(-R, R)$, we have the following facts.

1)

$$\int_0^x f(t) dt = \sum_{n=0}^{\infty} a_n \frac{x^{n+1}}{n+1}.$$

2)

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1}.$$

The integrated and differentiated series have the same radius of convergence R .

Proof. 2) Note that if R is its radius of convergence, a power series converges uniformly on closed subintervals of $(-R, R)$. So we just have to convince ourselves that the extra factor of n in $n a_n$ does not affect the radius of convergence.

To see this, note that if $0 < |x| < c < R$, we know by the definition of radius of convergence R on the first page of this part of the lectures, that $\sum_{n=0}^{\infty} |a_n| c^n < \infty$. Thus $\lim_{n \rightarrow \infty} a_n c^n = 0$. It follows that there is a bound M such that $|a_n c^n| \leq M$, for all n .

Therefore we have:

$$n |a_n| |x|^{n-1} = \frac{n}{c} \left(\frac{|x|}{c}\right)^{n-1} |a_n| c^n \leq M \frac{n}{c} \left(\frac{|x|}{c}\right)^{n-1}.$$

So $\sum_{n=0}^{\infty} n |a_n| |x|^{n-1}$ can be compared with $\sum_{n=0}^{\infty} n \left(\frac{|x|}{c}\right)^{n-1}$. This series converges by the ratio test as the ratios are

$$\frac{(n+1) \left(\frac{|x|}{c}\right)^n}{n \left(\frac{|x|}{c}\right)^{n-1}} = \frac{n+1}{n} \frac{|x|}{c} \rightarrow \frac{|x|}{c} < 1, \text{ as } n \rightarrow \infty.$$

1) is left to the reader for **Exercise**. ■

Examples. The power series for e^x , $\sin(x)$, $\cos(x)$ converge absolutely and uniformly on closed and bounded sets in the real line. So they are differentiable and integrable term by term. The same holds if we replace \mathbb{R} with \mathbb{C} or $n \times n$ matrices.

18 Taylor Series

In Lectures, I, we proved Taylor's formula with remainder:

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (x-a)^k + R_n.$$

We had 2 formulas for the remainder. The easiest to remember is $R_n = \frac{f^{(k)}(c)}{k!} (x-a)^k$, for some c between a and x . If one can show that $\lim_{n \rightarrow \infty} R_n = 0$, then the function $f(x)$ is represented by its Taylor series within the radius of convergence. That is, for $|x-a| < R$,

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

See Wikipedia (Taylor series article) for animated pictures showing the convergence of the Taylor series to e^x . We proved last quarter that our favorite functions: e^x , $\sin(x)$, $\cos(x)$, $\log(1-x)$ are represented by their Taylor series within the radius of convergence interval. Of course we cheated and defined e^x by its Taylor series. So we have:

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n, \text{ for all } x \in \mathbb{R};$$

$$\log(1 - x) = - \sum_{n=0}^{\infty} \frac{1}{n} x^n, \text{ for } |x| < 1.$$

Figure 14 shows the sum of the first 7 terms of the Taylor series for e^x , namely $f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{5 \cdot 24} + \frac{x^6}{30 \cdot 24}$ in red and e^x in blue on the interval $[-5, 5]$.

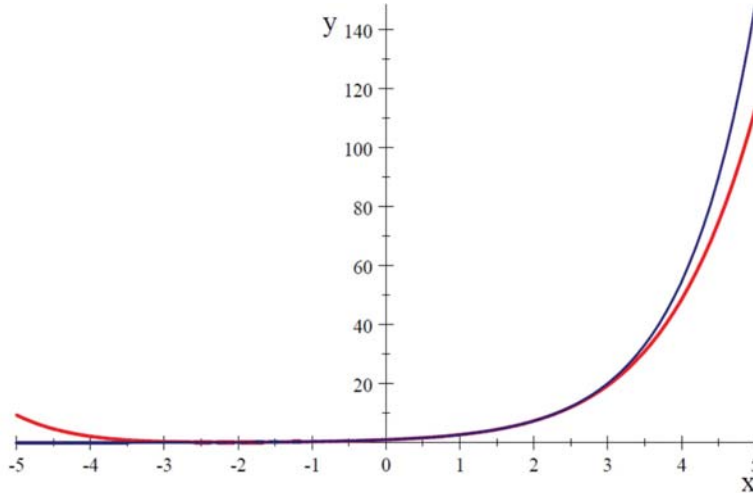


Figure 14: Plot of e^x (blue) and 1st 7 terms of its Taylor series about 0 (red).

Another example is the **Binomial Series**:

$$(1 + x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n, \tag{4}$$

where we define the generalized **binomial coefficient**

$$\binom{\alpha}{n} = \frac{\alpha(\alpha - 1) \cdots (\alpha - n + 1)}{n!}.$$

As usual, define $0! = 1$. When α is a positive integer, this is the ordinary binomial theorem. Otherwise you need to restrict x to have $|x| < 1$. Why?

Exercise. Prove formula (4) using Taylor's formula from Lectures, I. Find the radius of convergence. Why is this the binomial theorem when α is a positive integer?

We also saw earlier that there are functions $f(x)$ not represented by their Taylor series; e.g., $f(0) = 0$, and for $x \neq 0$, define $f(x) = e^{-\frac{1}{x^2}}$. For this function, one can show that $f^{(n)}(0) = 0$, for all n . This means the Taylor series for $f(x)$ around $x = 0$ is 0 even though $f(x)$ itself is positive except at $x = 0$. For $x \neq 0$,

$$e^{-\frac{1}{x^2}} = f(x) \neq \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = 0.$$

See Figure 15.

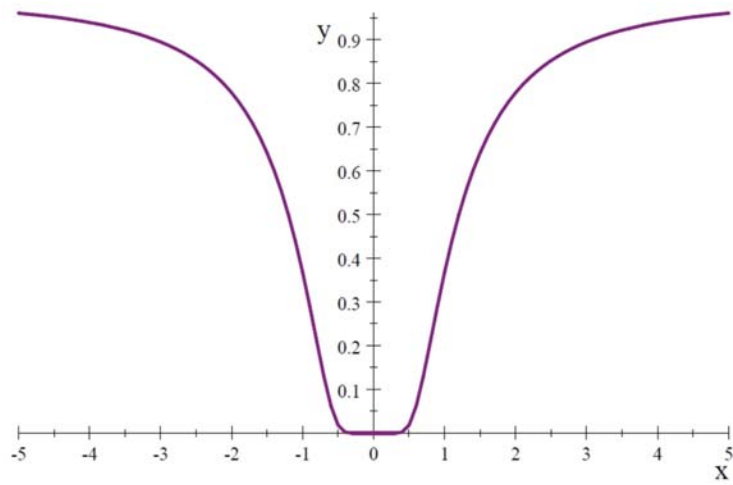


Figure 15: The plot of $f(x) = e^{-\frac{1}{x^2}}$. This function is not represented by its Taylor series about 0 which is 0 for all x .

Part IV

The Integral Exists!

19 Introduction

Recall our Lectures, I, Part 7. We assumed that for any finite interval $[a, b]$ there exists an integral $\int_a^b f$ of a continuous function $f : [a, b] \rightarrow \mathbb{R}$. And we assumed the integral satisfies the 2 axioms:

$$\text{INT 1. } m \leq f(x) \leq M, \forall x \in [a, b] \implies m(b-a) \leq \int_a^b f \leq M(b-a).$$

$$\text{INT 2. } a < c < b \implies \int_a^b f = \int_a^c f + \int_c^b f.$$

We deduced all the other basic facts about integrals from these 2 axioms; e.g., the fundamental theorem of calculus, Taylor's formula.

However, **we never showed that such an integral $\int_a^b f$ exists!** I doubt that anyone was too worried. But now we will finally prove the existence of the integral. We will in fact be able to integrate more general functions than just those that are continuous on $[a, b]$.

Note: We should also have proved \mathbb{R} exists. That is done by forming the completion of \mathbb{Q} (the space of Cauchy sequences $\{x_n\}$ of rationals mod the equivalence relation $\{x_n\} \simeq \{y_n\}$ iff $\lim_{n \rightarrow \infty} |x_n - y_n| = 0$).

How will we create $\int_a^b f$? It will be somewhat reminiscent of calculus. Recall the Riemann sums, where you partition the interval $[a, b]$ with points $a = a_0 < a_1 < \dots < a_n = b$. Then you form rectangles the sum of whose areas approximates the integral. In Figure 16 we approximate $\int_0^1 x^2 dx$ as in calculus by dividing up the x -axis interval $[0, 1]$ into 5 equal parts. The blue lines show the tops of the rectangles and also provide the graph of a step function approximating $f(x) = x^2$. Our approximation for the integral is

$$.2 * ((.2)^2 + (.4)^2 + (.6)^2 + (.8)^2 + 1^2) \cong 0.44$$

The actual integral is $x^3/3|_0^1 = 1/3 \cong .333333$.

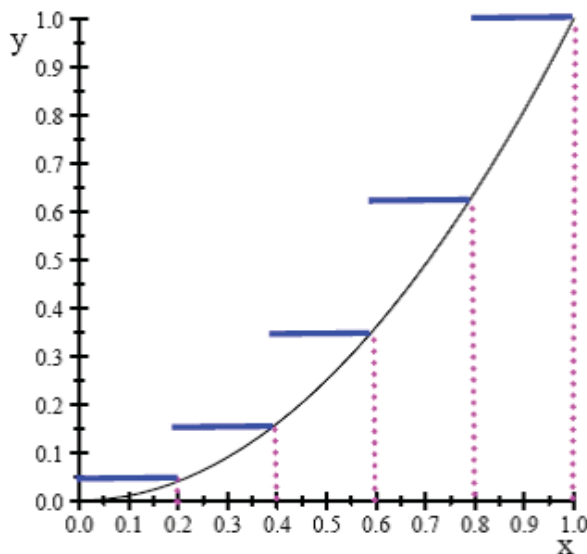


Figure 16: The calculus way of approximating the integral $\int_0^1 x^2 dx$. Divide the x -axis into 5 equal parts and take the value of the function at the right- (or left-) hand end point of the subinterval for the height of the rectangle.

Our newfangled way of finding an approximation for the same integral $\int_0^1 x^2 dx$ is illustrated in Figure 17. We need to find a step function $s(x)$ such that $\|s - f\|_\infty = \text{l.u.b.}_{0 \leq x \leq 1} |s(x) - f(x)|$ is small. To do this, one should 1st divide up the y -axis rather than the x -axis. Although in the end we divide up both axes.

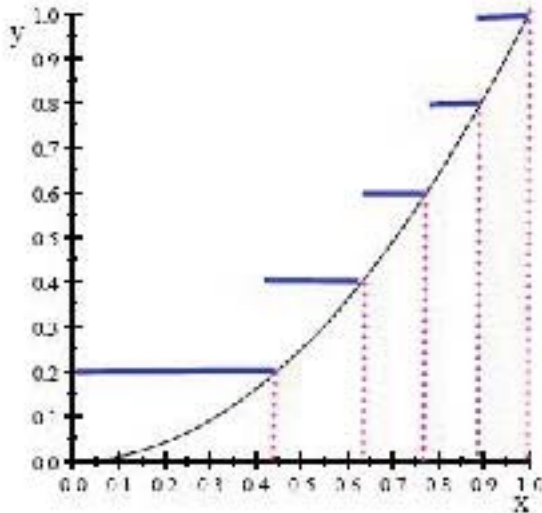


Figure 17: The newfangled way of approximating the integral $\int_0^1 x^2 dx$. We find a step function in blue $s(x)$ such that $\|x^2 - s(x)\| \leq \frac{1}{5}$.

For $\int_0^1 x^2 dx$, we divide the y -axis interval (also $[0, 1]$) into 5 equal subdivisions and go down to the x -axis via the inverse function \sqrt{x} to our original function x^2 . Now our approximation for the integral is:

$$.2 * (\sqrt{.2} - 0) + .4 * (\sqrt{.4} - \sqrt{.2}) + .6 * (\sqrt{.6} - \sqrt{.4}) + .8 * (\sqrt{.8} - \sqrt{.6}) + 1 * (1 - \sqrt{.8}) \cong 0.45026.$$

Numerically this is not impressive. But it will give us a way of constructing integrals that throws a new light on the theory. We will be able to get the Lebesgue integral simply by changing our norm from $\| \cdot \|_\infty$ norm to the $\| \cdot \|_1$ -norm.

Our method of integrating x^2 requires us to say the integral $\int_0^1 x^2 dx$ is the limit as $n \rightarrow \infty$ of $\int_0^1 s_n(x) dx$ for a sequence of step function $s_n(x)$ such that $\lim_{n \rightarrow \infty} \|s_n(x) - x^2\|_\infty = 0$. More generally we write $St[a, b]$ = the space of step functions on the interval $[a, b]$. We will be able to integrate functions f on $[a, b]$ that are uniform limits of sequences of step functions $s_n(x)$. These functions are in $\overline{St[a, b]}$ = the closure of the space of step functions with respect to the $\| \cdot \|_\infty$ norm. The official name for $\overline{St[a, b]}$ is the space of **regulated functions**. Yes, you can Google it.

In the exercises, we give an example of a non-regulated function. This function can't decide whether it is 1, -1 or 0 on any small interval containing 0. It does not have a right-hand limit at 0. The function $\frac{x \sin(1/x)}{|x \sin(1/x)|}$ is pictured in figure 18. This is not quite the function in the exercises as it does not have a value when $x \sin(1/x) = 0$.

20 Continuous Linear Extension Theorem.

We will extend the integral from step functions to piecewise continuous functions and further. To do this, we need the continuous linear extension theorem. Suppose that F and E are normed vector spaces. Assume that E is **complete**; i.e., all Cauchy sequences in E converge to a limit in E . We will write $\| \cdot \|$ for the norms on both F and E . In our case $E =$ the real numbers $= \mathbb{R}$ with norm the usual absolute value.

Suppose that F_0 is a **subspace** of F ; i.e., a subset which is a vector space using the same operations of $+$ and multiplication by scalars as in F . In our case, F_0 will be $St[a, b]$, the step functions on the finite interval $[a, b]$.

Let $L : F_0 \rightarrow E$ be **linear**; i.e., $L(\alpha x + \beta y) = \alpha L(x) + \beta L(y)$, for all x, y in F_0 and $\alpha, \beta \in \mathbb{R}$. In our case L will be the integral over a finite interval $[a, b]$.

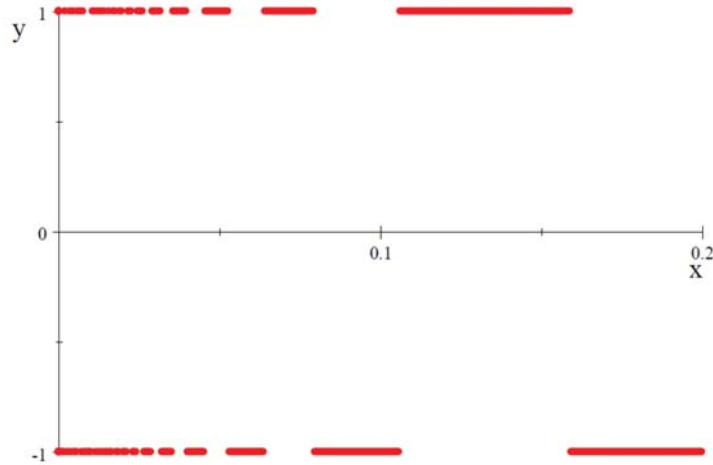


Figure 18: Plot of $\frac{x \sin(1/x)}{|x \sin(1/x)|}$ on $(0, .2)$

We also assume that L is continuous for the norms on E and F_0 . We will use the ∞ -norm on $St[a, b]$; i.e., $\|f\|_\infty = \text{l.u.b. } |f(x)|$.

Lemma 47 Suppose $L : F \rightarrow E$ is a linear function, where E, F are normed vector spaces. Then L is continuous $\Leftrightarrow \exists$ constant $C > 0$ such that $\|L(x)\| \leq C \|x\|$, $\forall x \in F$.

Proof. \Rightarrow L continuous implies it is continuous at $x = 0$. Thus, taking $\varepsilon = 1$, we see that $\exists \delta > 0$ such that $\|x\| < \delta$ implies $\|L(x)\| < 1$.

Now we use a trick coming from properties of the norm. If $x \neq 0$, (which we may assume without any problem since that case is pretty clear as $L(0) = 0$), we have $\left\| \frac{\delta x}{2\|x\|} \right\| = \frac{\delta}{2} < \delta$. Since L is linear, we see that

$$1 > \left\| L \left(\frac{\delta x}{2\|x\|} \right) \right\| = \frac{\delta}{2\|x\|} \|L(x)\|.$$

It follows that $\|L(x)\| < \frac{2}{\delta} \|x\|$. So our constant $C = \frac{2}{\delta}$.

\Leftarrow Now suppose $\|L(x)\| \leq C \|x\|$, $\forall x \in F$. Using the linearity of L , we see that $\|L(x) - L(y)\| = \|L(x - y)\| \leq C \|x - y\| < \varepsilon$, if $\|x - y\| < \frac{\varepsilon}{C} = \delta$. It follows that in fact L is uniformly continuous on F . ■

Lemma 48 If F_0 is a subspace of a normed vector space F , then the closure $\overline{F_0}$ is also a subspace of F .

Proof. Recall the fact from the earlier section which says that the closure $\overline{F_0}$ consist of points of F which are limits of sequences of points from F_0 . So we get this lemma from properties of limits. Let $x, y \in \overline{F_0}$, and $\{x_n\} \{y_n\}$ be sequences of points from F_0 such that $\lim_{n \rightarrow \infty} x_n = x$ and $\lim_{n \rightarrow \infty} y_n = y$. Suppose $\alpha, \beta \in \mathbb{R}$. We need to show that

$$\lim_{n \rightarrow \infty} (\alpha x_n + \beta y_n) = \alpha x + \beta y.$$

This is an **exercise** in the properties of limits. It follows that $\alpha x + \beta y \in \overline{F_0}$. That means $\overline{F_0}$ is a vector space. ■

Theorem 49 Continuous Linear Extension Theorem

Assume F =normed vector space, F_0 = subspace of F , E = complete normed vector space, $L:F_0 \rightarrow E$, is continuous linear. Then L can be extended to a unique continuous linear function $\overline{L} : \overline{F_0} \rightarrow E$. Here $\overline{F_0}$ is the closure of F_0 , as in Lemma 48. If C is the constant in Lemma 47 for L , then it also works for \overline{L} .

Proof. Suppose $x \in \overline{F_0}$. Then there is a sequence of points $\{x_n\}$ from F_0 such that $\lim_{n \rightarrow \infty} x_n = x$.

Claim 1. $\{L(x_n)\}$ is a Cauchy sequence in the complete normed vector space E .

Proof of Claim 1.

Using the linearity and continuity of L and Lemma 47, we know there is a positive constant C such that

$$\|L(x_n) - L(x_m)\| = \|L(x_n - x_m)\| \leq C \|x_n - x_m\| < \varepsilon$$

if n, m are sufficiently large, since $\{x_n\}$ is Cauchy.

Q.E.D. claim 1.

Since $L(x_n) \in E$, a complete normed vector space, we know

$$\exists \lim_{n \rightarrow \infty} L(x_n) = w \in E.$$

We want to **define**

$$w = \overline{L}(x).$$

Claim 2. w is unique and thus $w = \overline{L}(x)$ is a legal definition of a function.

Proof of Claim 2. Suppose we take another sequence $\{u_n\}$ from F_0 such that $\lim_{n \rightarrow \infty} u_n = x$. Then

$$\lim_{n \rightarrow \infty} L(u_n) = v \in E.$$

We want to show that $v = w$.

Using the triangle inequality and the linearity and continuity of L with the existence of its positive constant C , we see that

$$\begin{aligned} \|v - w\| &\leq \|v - L(u_n)\| + \|L(u_n) - L(x_n)\| + \|L(x_n) - w\| \\ &\leq \|v - L(u_n)\| + C \|x_n - u_n\| + \|L(x_n) - w\|. \end{aligned}$$

But each of the 3 terms on the right approaches 0 as $n \rightarrow \infty$. By the Squeeze Lemma then $\|v - w\| = 0$.

Q.E.D. claim 2.

Claim 3. The function $\overline{L}(x)$ is linear.

Exercise. Prove this using properties of limits and the linearity of L .

Claim 4. The function $\overline{L}(x)$ is continuous with the same constant C as for L .

Proof of Claim 4.

By the definition of $\overline{L}(x)$ and the continuity of the norm, if $\lim_{n \rightarrow \infty} x_n = x$, for a sequence $\{x_n\}$ from F_0 , we have

$$\|\overline{L}(x)\| = \left\| \lim_{n \rightarrow \infty} L(x_n) \right\| = \lim_{n \rightarrow \infty} \|L(x_n)\|.$$

We know that $\|L(x_n)\| \leq C \|x_n\| \forall n$. Since the limit preserves \leq , we see (using the continuity of $\|\cdot\|$) that

$$\|\overline{L}(x)\| \leq \lim_{n \rightarrow \infty} C \|x_n\| = C \|x\|.$$

Q.E.D. claim 4. ■

This completes our proof of the continuous linear extension theorem - the hardest part of our construction of the integral. Next we proceed to use the theorem.

Part V

Construction of the Integral Continued

21 The Integral of a Step Map on $[a, b]$

We assume $-\infty < a < b < \infty$. A **step map or function** is just what it says; i.e., a function whose graph looks like a bunch of steps. More precisely, we define $f : [a, b] \rightarrow \mathbb{R}$ to be a step function if we can partition $[a, b]$ by $\mathcal{P} = \{a_0 < a_1 < \dots < a_n\}$

so that f is constant on each subinterval (a_{i-1}, a_i) , for each $i = 1, \dots, n$:

$$f(x) = w_i, \quad \forall x \in (a_{i-1}, a_i).$$

We don't really care what values are assigned at the endpoints of the subintervals. These endpoints correspond to subintervals of length 0. Even if you assign a value to the function on such subintervals, it will contribute nothing to the Riemann integral defined below. See Figure 19.

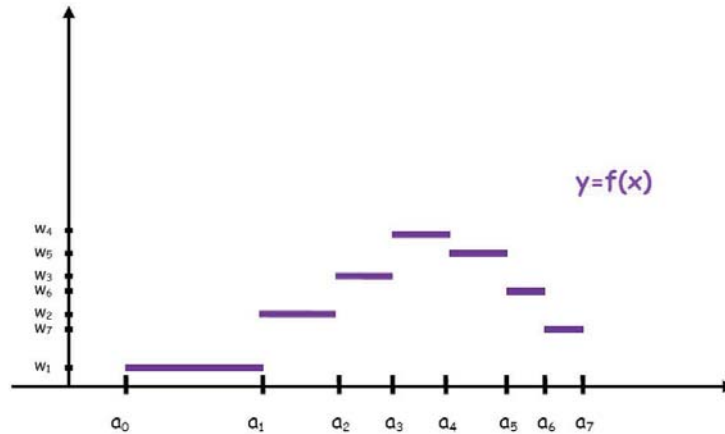


Figure 19: A step function $y = f(x)$.

Definition 50 The *Riemann integral of the step function* f is defined to be

$$\int_a^b f = I(f, \mathcal{P}) = \sum_{i=1}^n w_i(a_i - a_{i-1}).$$

The sum $\sum_{i=1}^n w_i(a_i - a_{i-1})$ should look like a Riemann sum. Here $w_i = f(c_i)$ for any choice of points $c_i \in (a_{i-1}, a_i)$.

We define a partition $\mathcal{Q} = \{b_0 < b_1 < \dots < b_m\}$ to be a **refinement** of the partition $\mathcal{P} = \{a_0 < a_1 < \dots < a_n\}$ if the set of points $\{a_0, a_1, \dots, a_n\}$ is contained in the set of points $\{b_0, b_1, \dots, b_m\}$.

So, for example, the points $\mathcal{P} = \{\frac{i}{n} \mid i = 0, \dots, n\}$ form a partition (regular) of $[0, 1]$. The set $\mathcal{Q} = \{\frac{i}{2n} \mid i = 0, \dots, 2n\}$ is a refinement of \mathcal{P} .

Now we need to worry about the possibility that the integral of a step function might depend on what partition we choose.

Lemma 51 Suppose f is a step function with respect to 2 partitions

$$\mathcal{P} = \{a_0 < a_1 < \dots < a_n\} \quad \text{and} \quad \mathcal{Q} = \{b_0 < b_1 < \dots < b_m\}$$

of $[a, b]$. Then $I(f, \mathcal{P}) = I(f, \mathcal{Q})$.

Proof. Step 1. The partitions \mathcal{P} and \mathcal{Q} have a common refinement \mathcal{R} whose subinterval endpoints are the union of the endpoints from the partitions \mathcal{P} and \mathcal{Q} ; i.e., the points $\{a_0, a_1, \dots, a_n\} \cup \{b_0, b_1, \dots, b_m\}$. Call the common refinement $\mathcal{R} = \{a = r_0 < r_1 < \dots < r_k = b\}$.

Step 2. We claim that $I(f, \mathcal{P}) = I(f, \mathcal{R}) = I(f, \mathcal{Q})$.

To see this, look at Figure 20.

This shows that if we look at any subinterval of partition \mathcal{P} , like (a_{i-1}, a_i) and further subdivide it by inserting a point c ; $a_{i-1} < c < a_i$, then since f is the constant w_i on the subinterval (a_{i-1}, a_i) , the contribution to $I(f, \mathcal{P})$ from (a_{i-1}, a_i) is

$$w_i(a_i - a_{i-1}) = w_i(c - a_{i-1}) + w_i(a_i - c).$$

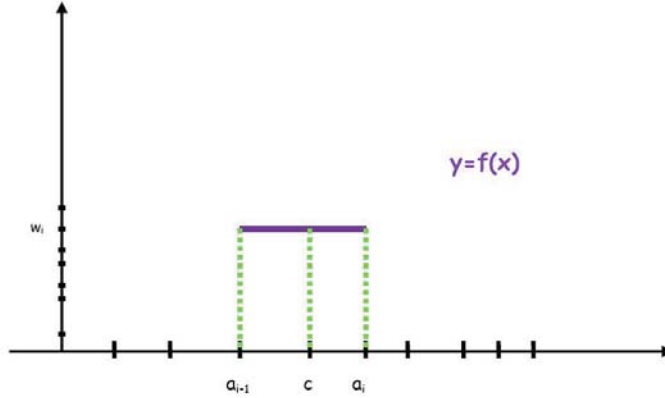


Figure 20: Inserting a partition point in the i th subinterval of \mathcal{P} replaces a term in $I(f, \mathcal{P}) = \int_a^b f$ by the sum of 2 terms representing the sums of the areas of the 2 rectangles pictured (if the function is ≥ 0 on the i th subinterval of \mathcal{P} anyway) but this does not change the final result.

The right hand side is the sum associated to the 2 subintervals (a_{i-1}, c) and (c, a_i) in our new partition refined by adding one point to the i th subinterval. Keep doing this for as many points as you want to add to \mathcal{P} to get the refinement \mathcal{R} . Hopefully this convinces you that $I(f, \mathcal{P}) = I(f, \mathcal{R})$. Similarly, since \mathcal{R} is also a refinement of \mathcal{Q} , we see that $I(f, \mathcal{Q}) = I(f, \mathcal{R})$. It follows that $I(f, \mathcal{P}) = I(f, \mathcal{Q})$. ■

Lemma 52 Suppose that $St[a, b]$ denotes the set of all step functions defined on the interval $[a, b]$. Define the integral $\int_a^b f$ for $f \in St[a, b]$ as in Definition 50. By Lemma 51, the integral is independent of the partition \mathcal{P} used to define the step function f . The integral then has the following properties on step functions.

Lemma 53 a) Integrals Preserve \leq .

$$f, g \in St[a, b] \text{ and } f(x) \leq g(x), \text{ for all } x \in [a, b], \text{ then } \int_a^b f \leq \int_a^b g.$$

b) Integral is Continuous Linear. The map $I(f) = \int_a^b f$ is a continuous linear map from $St[a, b]$ into \mathbb{R} using the $\|\cdot\|_\infty$ norm on $St[a, b]$ and the usual absolute value on \mathbb{R} . In particular,

$$\left| \int_a^b f \right| \leq (b - a) \|f\|_\infty.$$

Proof. a) Integrals Preserve \leq .

Given step function f with corresponding partition \mathcal{P} and step function g with partition \mathcal{Q} , we take a common refinement \mathcal{R} of \mathcal{P} and \mathcal{Q} . Both f and g are step functions with respect to \mathcal{R} . Let $\mathcal{R} = \{a_0 < a_1 < \dots < a_n\}$. Suppose that for each $i = 1, \dots, n$:

$$\begin{aligned} f(x) &= w_i, \quad \forall x \in (a_{i-1}, a_i) \\ g(x) &= v_i, \quad \forall x \in (a_{i-1}, a_i). \end{aligned}$$

If $f(x) \leq g(x)$, for all $x \in [a, b]$, then $w_i \leq v_i$, $\forall i$. It follows that

$$I(f, \mathcal{P}) = \int_a^b f = \sum_{i=1}^n w_i (a_i - a_{i-1}) \leq I(g, \mathcal{P}) = \int_a^b g = \sum_{i=1}^n v_i (a_i - a_{i-1}).$$

b) Integral is Continuous Linear. Note that if f is a step function with respect to the partition $\mathcal{P} = \{a_0 < a_1 < \dots < a_n\}$ of $[a, b]$ and for each $i = 1, \dots, n$:

$$f(x) = w_i, \quad \forall x \in (a_{i-1}, a_i),$$

then

$$I(f, \mathcal{P}) = \int_a^b f = \sum_{i=1}^n w_i (a_i - a_{i-1}).$$

We have $\|f\|_\infty = \max_{i=1, \dots, n} |w_i|$. It follows, using the triangle inequality and the fact that the lengths of the subintervals of the partition add up to $b - a$, that

$$\begin{aligned} \left| \int_a^b f \right| &\leq \sum_{i=1}^n |w_i| (a_i - a_{i-1}) \leq \sum_{i=1}^n \|f\|_\infty (a_i - a_{i-1}) \\ &= \|f\|_\infty \sum_{i=1}^n (a_i - a_{i-1}) = (b - a) \|f\|_\infty. \end{aligned}$$

Continuity at $f = 0$ follows from this inequality. Once we have proved linearity of the integral, the (uniform) continuity will also follow since then we can say

$$\left| \int_a^b f - \int_a^b g \right| = \left| \int_a^b (f - g) \right| \leq \left| \int_a^b |f - g| \right| \leq (b - a) \|f - g\|_\infty.$$

For the last 2 inequalities, we are using the fact that the integral on step functions preserves inequalities. One quickly finds a δ as a function of ε and independent of f and g .

Linearity. Let $\alpha, \beta \in \mathbb{R}$. Given step function f with corresponding partition \mathcal{P} and step function g with partition \mathcal{Q} , we take a common refinement \mathcal{R} of \mathcal{P} and \mathcal{Q} . Both f and g are step functions with respect to \mathcal{R} . Let $\mathcal{R} = \{a_0 < a_1 < \dots < a_n\}$. Suppose that for each $i = 1, \dots, n$:

$$\begin{aligned} f(x) &= w_i, \quad \forall x \in (a_{i-1}, a_i) \\ g(x) &= v_i, \quad \forall x \in (a_{i-1}, a_i). \end{aligned}$$

Then $\alpha f + \beta g$ is a step function for partition \mathcal{R} with values for each $i = 1, \dots, n$:

$$(f + g)(x) = \alpha w_i + \beta v_i, \quad \forall x \in (a_{i-1}, a_i).$$

This shows that $St[a, b]$ is indeed a normed vector space with norm $\|\cdot\|_\infty$.

It follows from our definition of the integral on step functions that

$$\begin{aligned} I(\alpha f + \beta g, \mathcal{R}) &= \int_a^b (\alpha f + \beta g) = \sum_{i=1}^n (\alpha w_i + \beta v_i) (a_i - a_{i-1}) \\ &= \alpha \sum_{i=1}^n w_i (a_i - a_{i-1}) + \beta \sum_{i=1}^n v_i (a_i - a_{i-1}) \\ &= \alpha \int_a^b f + \beta \int_a^b g. \end{aligned}$$

So the integral is linear. ■

Exercise. Show that it does not matter what (finite) values the function f takes on the partition points.

22 The Integral on Step Functions satisfies the 2 Axioms for Integrals from Lectures I.

Recall from Lectures I that we needed 2 axioms for integrals in order to prove the fundamental theorem of calculus. Now we prove those axioms for step functions $f \in St[a, b]$.

Axiom 1. If $m \leq f(x) \leq M$ for all $x \in [a, b]$, then $m(b-a) \leq \int_a^b f \leq M(b-a)$.

Axiom 2. If $a < c < b$, then $\int_a^b f = \int_a^c f + \int_c^b f$.

Proof. Axiom 1 for Step Functions. Since $m \leq f(x) \leq M$ for all $x \in [a, b]$ and we have proved that the integral preserves inequalities, we have

$$m(b-a) = \int_a^b m \leq \int_a^b f \leq \int_a^b M = M(b-a).$$

■

Proof. Axiom 2 for Step Functions. For f a step function on $[a, b]$, we can always add a point like c to our partition $\mathcal{P} = \{a_0 < a_1 < \dots < a_n\}$ of $[a, b]$ defining f and assume $c = a_j$ for some j . Suppose that for each $i = 1, \dots, n$:

$$f(x) = w_i, \quad \forall x \in (a_{i-1}, a_i).$$

■

Proof. Then according to our definition of the integral of a step function

$$\begin{aligned} I(f, \mathcal{P}) &= \int_a^b f = \sum_{i=1}^n w_i(a_i - a_{i-1}) \\ &= \sum_{i=1}^j w_i(a_i - a_{i-1}) + \sum_{i=j+1}^n w_i(a_i - a_{i-1}) \\ &= \int_a^c f + \int_c^b f. \end{aligned}$$

■

This completes our proofs of the properties of the integral on step functions. In the next section we prove a result which will help us with our scheme to integrate continuous functions.

23 Uniform Approximation of Continuous Functions by Step Functions

Theorem 54 $f : [a, b] \rightarrow \mathbb{R}$ continuous implies f is uniformly continuous on $[a, b]$.

Proof. (by Contradiction)

Otherwise $\exists \varepsilon > 0$ s.t. $\forall n \in \mathbb{Z}^+ \exists x_n, y_n \in [a, b]$ with $|x_n - y_n| < \frac{1}{n}$ and $|f(x_n) - f(y_n)| \geq \varepsilon$.

But we showed in Lectures I (using the Spanish Hotel argument) that any bounded sequence of real numbers has a convergent subsequence. Thus there is an infinite subset J of \mathbb{Z}^+ such that

$$\begin{aligned} \exists \lim_{n \rightarrow \infty} x_n &= x \in [a, b] \\ n &\in J \end{aligned}$$

and

$$\begin{aligned} \exists \lim_{n \rightarrow \infty} y_n &= y \in [a, b]. \\ n &\in J \end{aligned}$$

But then $|x_n - y_n| < \frac{1}{n}$ implies using the continuity of absolute value:

$$\begin{aligned} |x - y| &= \left| \lim_{\substack{n \rightarrow \infty \\ n \in J}} x_n - \lim_{\substack{n \rightarrow \infty \\ n \in J}} y_n \right| \\ &= \lim_{\substack{n \rightarrow \infty \\ n \in J}} |x_n - y_n| \leq \lim_{\substack{n \rightarrow \infty \\ n \in J}} \frac{1}{n} = 0, \end{aligned}$$

which means $x = y$.

On the other hand, since $|f(x_n) - f(y_n)| \geq \varepsilon$ and $x = y$, we have

$$\begin{aligned} 0 &= |f(x) - f(y)| = \left| \lim_{\substack{n \rightarrow \infty \\ n \in J}} f(x_n) - \lim_{\substack{n \rightarrow \infty \\ n \in J}} f(y_n) \right| \\ &= \lim_{\substack{n \rightarrow \infty \\ n \in J}} |f(x_n) - f(y_n)| \geq \varepsilon. \end{aligned}$$

This contradiction proves the theorem. ■

Now we are ready to uniformly approximate continuous functions by step functions.

Theorem 55 *For every continuous function $f : [a, b] \rightarrow \mathbb{R}$, there exists a sequence of step function $s_n \in St[a, b]$ such that $\|f - s_n\|_\infty \rightarrow 0$ as $n \rightarrow \infty$.*

Proof. Given $\varepsilon > 0$, the preceding theorem says

$$\exists \delta > 0 \quad \text{s.t.} \quad |x - y| < \delta \quad \text{implies} \quad |f(x) - f(y)| < \varepsilon. \quad (5)$$

Now let n be so large that $\frac{b-a}{n} < \delta$ and let $\mathcal{P} = \{a_0 < a_1 < \dots < a_n\}$ be a partition of $[a, b]$ such that $a_i - a_{i-1} = \frac{b-a}{n} < \delta$. Define the step function s_n on the subinterval (a_{i-1}, a_i) to take any value $f(c)$ for $c \in (a_{i-1}, a_i)$. Then by formula (5)

$$|f(x) - s_n(x)| < \varepsilon \quad \text{for all } x \in (a_{i-1}, a_i).$$

This means $\|f - s_n\|_\infty < \varepsilon$ for n sufficiently large. ■

The preceding theorem says that the closure of the space of step functions with respect to the $\|\cdot\|_\infty$ norm, written $\overline{St[a, b]}$, contains all continuous functions on $[a, b]$. It clearly also contains all **piecewise continuous functions** (which are allowed a finite number of jump discontinuities).

The space $\overline{St[a, b]}$ is called the space of **regulated functions**. We saw earlier (and in the exercises) that not all bounded functions on $[a, b]$ are regulated. A function $f(x)$ must have right and left hand limits at all $x \in [a, b]$ to be regulated.

24 Extension of the Integral to the Regulated Functions on $[a, b]$.

We can use the continuous linear extension theorem to extend the integral from $St[a, b]$ to $\overline{St[a, b]}$. Every $f \in \overline{St[a, b]}$ is a limit with respect to the $\|\cdot\|_\infty$ norm of a sequence $s_n \in St[a, b]$. Then the continuous linear extension theorem tells us to define

$$\int_a^b f = \lim_{n \rightarrow \infty} \int_a^b s_n.$$

We need to show that this extended integral satisfies the same 2 axioms that the step functions were shown to satisfy.

Theorem 56 *Properties of the Integral on the Space $\overline{St[a, b]}$ of Regulated Functions (which includes all continuous, even piecewise continuous functions on $[a, b]$).*

1) $\int_a^b f$ is a linear map from $f \in \overline{St[a, b]}$ into $\int_a^b f \in \mathbb{R}$ and

$$\left| \int_a^b f \right| \leq (b-a) \|f\|_\infty.$$

2) $\int_a^b f$ preserves inequalities; i.e., $f, g \in \overline{St[a, b]}$ with $f(x) \leq g(x) \quad \forall x \in [a, b]$ implies

$$\int_a^b f \leq \int_a^b g.$$

3) $a < c < b$ implies $\int_a^b f = \int_a^c f + \int_c^b f$.

Proof. 1) This is just the continuous linear extension theorem from the last lecture.

2) Let $h(x) = g(x) - f(x) \quad \forall x \in [a, b]$. Then $h(x) \geq 0 \quad \forall x \in [a, b]$. Suppose $s_n \in St[a, b]$ s.t. $\lim_{n \rightarrow \infty} \|s_n - h\|_\infty = 0$. Suppose s_n is a step function for the partition $\mathcal{P} = \{a_0 < a_1 < \dots < a_m\}$ of $[a, b]$ and for $i = 1, \dots, m$,

$$s_n(x) = w_i, \quad \forall x \in (a_{i-1}, a_i).$$

If $w_i < 0$ for some i , we can replace w_i by 0 and make a new step function s_n^* which is even closer to h than s_n was (because $h(x) \geq 0$ for all x). Now we see that

$$\int_a^b g - \int_a^b f = \int_a^b h = \lim_{n \rightarrow \infty} \int_a^b s_n^* \geq 0.$$

3) Let s_n be a sequence of step maps converging to f in the ∞ -norm on $[a, b]$. Then s_n must also converge to f in the ∞ -norm on $[a, c]$ and on $[c, b]$. We showed that for step functions we have:

$$\int_a^b s_n = \int_a^c s_n + \int_c^b s_n.$$

Take the limit as $n \rightarrow \infty$ to get property 3 using the basic properties of limits from last quarter. ■

This completes our discussion of the existence of an integral with the properties Int1 and Int 2 from Lectures I. From this we deduced in Lectures I all the basic properties of integrals such as the fundamental theorem of calculus, integration by parts, the formula for substituting in an integral, the formulas like

$$\int_a^b x^n dx = \frac{x^{n+1}}{n+1} \Big|_{x=a}^b = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1},$$

when $n \neq -1$ and assuming that if $n < -1$, 0 is not in the interval $[a, b]$.

To extend the fundamental theorem of calculus from continuous functions to piecewise continuous functions or regulated functions, requires a little more effort. Lang does this in *Undergraduate Analysis*, Chapter 10 and produces theorems legalizing differentiation under the integral sign. In the last part of the book he extends the theory to integrals in several variables. I leave it to you to read these things.

Part VI

Dirac and Weierstrass

One goal of this part of the lectures is to solve the following problem. Given a continuous function $f : [a, b] \rightarrow \mathbb{R}$, find a polynomial p such that $\|f - p\|_\infty$ (where $\|g\|_\infty = \max_{a \leq x \leq b} |g(x)|$) is arbitrarily small. You might think we had already solved this problem with Taylor's formula and series. But Taylor polynomials do not give arbitrarily good approximations unless the function f has lots of derivatives. Moreover, we know that there are infinitely differentiable functions that are not represented by their Taylor series. See Figure 15. Thus we must find a new method to get polynomial approximations. At the same time we will create the mechanism to find other sorts of approximations that we will need when we discuss Fourier series in the last section.

25 A New Kind of Product of Functions

You are familiar with the pointwise product of functions defined by $(f \cdot g)(x) = f(x) \cdot g(x)$. You just take the product of the real numbers $f(x)$ and $g(x)$. Thus if we define $f(x) = 1$ for all x , we get $(f \cdot g)(x) = g(x)$. So $f(x) = 1$ is the identity for pointwise product.

Now we want to define a new kind of product, well, not that new if you got to convolution in the Laplace transform section of ODEs courses. And not so new if you are taking probability or statistics courses where given independent random variables with densities f and g , the density of the sum of the random variables is the convolution product $f * g$. For this product, $f * 1 \neq f$.

Our aim is to use convolution in order to uniformly approximate continuous functions f on $[a, b]$ by polynomials $\sum_{n=0}^N a_n x^n$ (or continuous functions of period 1 by trigonometric polynomials $\sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$ for Fourier series).

We will assume that our functions are piecewise continuous and that at least one of the functions in $f * g$ vanishes off a bounded interval so that we know the integrals exist without thinking too hard about the meaning of

$$\int_{-\infty}^{\infty} = \lim_{A \rightarrow -\infty} \left(\lim_{B \rightarrow \infty} \int_A^B \right).$$

Definition 57 The *convolution* of f and g is defined for all x by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt.$$

One can read $f * g$ as **f splat g** since it does splat the properties of the 2 functions together, preserving the best properties. If one function is discontinuous, but the other is a polynomial, the convolution is a polynomial.

Example. Let $g(x) = 1 - x^2$, for all x . Define

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

See Figures 21 and 22.

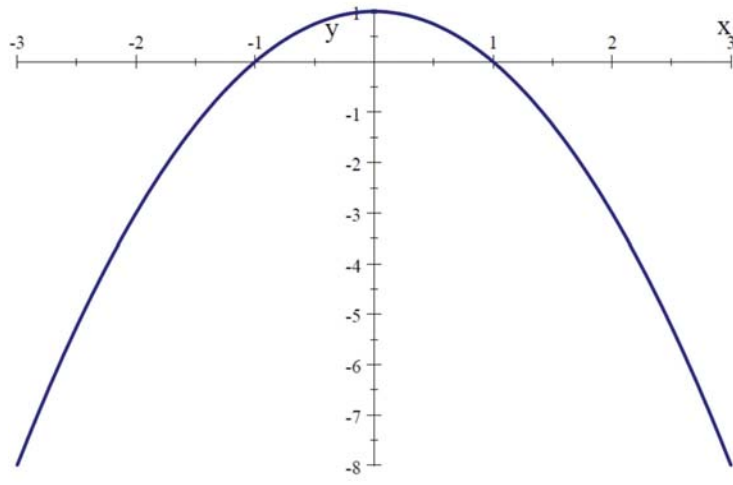


Figure 21: $g(x) = 1 - x^2$

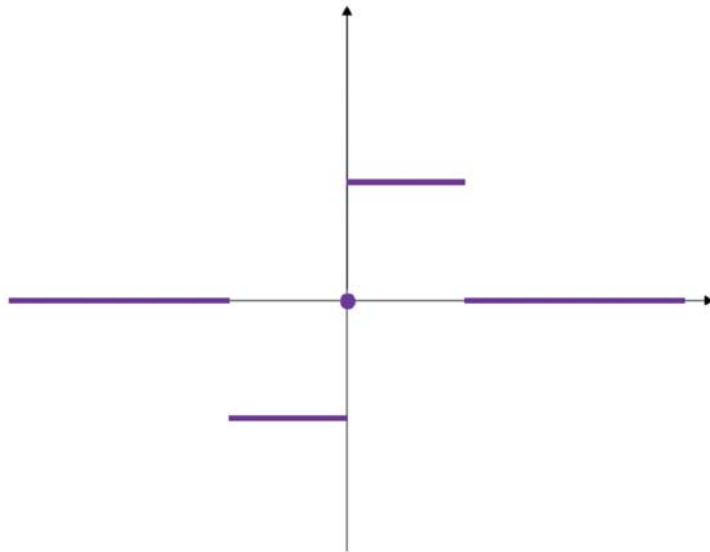


Figure 22: The $f(x) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ -1, & \text{if } -1 < x < 0 \\ 0, & \text{if } x = 0 \\ 0, & \text{if } x \geq 1 \text{ or } x \leq -1. \end{cases}$

Then, since f has differing formulas on different intervals (and mercifully g does not), we get

$$\begin{aligned}
 (f * g)(x) &= -1 \int_{-1}^0 g(x-t) dt + \int_0^1 g(x-t) dt \\
 &= -\int_{-1}^0 (1 - (x-t)^2) dt + \int_0^1 (1 - (x-t)^2) dt \\
 &= \frac{-2x^3}{3} + \frac{(x+1)^3}{3} + \frac{(x-1)^3}{3} = 2x.
 \end{aligned}$$

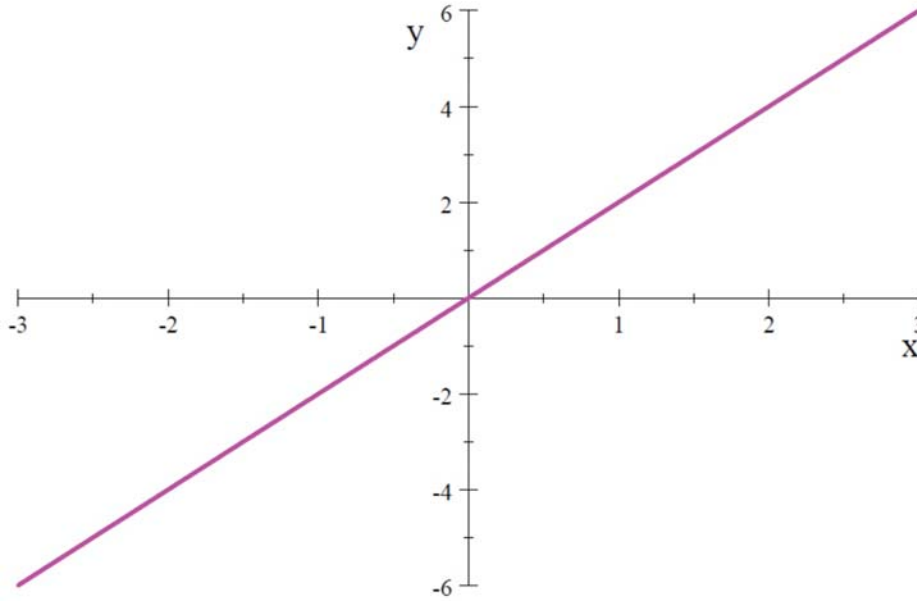


Figure 23: $f * g$, where f is from Figure 21 and g is from Figure 22.

So we see that even though f is discontinuous, when convolved with a polynomial, we get a polynomial.

Theorem 58 Properties of Convolution.

Assume f, g, h are piecewise continuous and at least one vanishes off a bounded interval when necessary to make an integral exist. Let $\alpha \in \mathbb{R}$. Then we have the following properties.

- 1) $f * g = g * f$
- 2) $f * (g + h) = f * g + f * h$
- 3) $(\alpha f) * g = \alpha (f * g)$
- 4) $f * (g * h) = (f * g) * h$
- 5) Suppose that c and d are both positive. If $f(x) = 0$ when $|x| > c$. and $g(x) = 0$ when $|x| > d$. Then $(f * g)(x) = 0$ when $|x| > c + d$.
- 6) If $g(x)$ is a polynomial then $(f * g)(x)$ is a polynomial.
- 7) $(f * g)' = f * (g')$ assuming g continuously differentiable.

Proof. 1) Make the change of variables $u = x - t$. Then $t = x - u$ and $du = -dt$ and the order of integration changes so we get:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt = -\int_{\infty}^{-\infty} f(x-u)g(u)du = \int_{-\infty}^{\infty} g(u)f(x-u)du = (g * f)(x).$$

2)-4) are **exercises**.

$$5) \int_{-\infty}^{\infty} f(t)g(x-t)dt = \int_{-c}^c f(t)g(x-t)dt.$$

Now if $x > c + d$, and $-c \leq t \leq c$, we see that $x - t > c + d - c = d$ and $g(x - t) = 0$. So $x > c + d$ implies $(f * g) = 0$. If $x < -c - d$ and $-c \leq t \leq c$, we see that $x - t < -c - d + c = -d$ and $g(x - t) = 0$. Thus $x < -c - d$ implies $(f * g) = 0$.

6) It suffices using 2) and 3) to consider the case that $g(x) = x^n$. Then supposing $f(x) = 0$ if $|x| > c$, we have the following, using the binomial theorem

$$\begin{aligned} (f * g)(x) &= \int_{-\infty}^{\infty} f(t)g(x-t)dt = \int_{-c}^c f(t)(x-t)^n dt \\ &= \int_{-c}^c f(t) \sum_{k=0}^n \binom{n}{k} x^{n-k} (-t)^k dt = \sum_{k=0}^n \binom{n}{k} x^{n-k} \int_{-c}^c f(t)(-t)^k dt. \end{aligned}$$

Let

$$\int_{-c}^c f(t)(-t)^k dt = c_k.$$

Then

$$(f * g)(x) = \sum_{k=0}^n \binom{n}{k} c_k x^{n-k},$$

which is a polynomial.

7) This is an **exercise** using a result legalizing interchange of derivative and integral which can be found in Lang, *Undergraduate Analysis*, p. 276, Theorem 7. The hypothesis requires the uniform continuity of $\frac{\partial}{\partial x}(f(t)g(x-t))$. ■

This concludes our discussion of convolution. It is important in the theory of probability and Fourier analysis, Laplace transforms. In fact the Fourier and Laplace transforms change convolution into ordinary pointwise product of the transformed functions. Convolution is also used to smooth data thanks to property 7). We want to use it to approximate continuous functions by polynomials.

26 A Function Which is Not an Ordinary Function

We want to think about something called the Dirac delta "function" denoted $\delta(x)$. This is not to be confused with our earlier use of the Greek letter δ to mean a small positive number. The Dirac delta function is used in physics to represent an impulse. Examples are a point mass and a point charge. It is often said to be a function that is 0 for $x \neq 0$ and ∞ at $x = 0$. At least that is what I remember as an undergrad taking physics classes for my minor. It used to drive me crazy, since as a mathematics major, I learned that such a function could not exist. Laurent Schwartz whose theory of distributions (1944) legalized delta describes how the formulas involving delta drove him crazy in 1935 in his autobiography. See L. Schwartz, *A Mathematician Grappling with his Century*, p. 218. He also says: "... it's a good thing that theoretical physicists do not wait for mathematical justification before going ahead with their theories!"

The graph usually associated with δ shows a unit arrow or spike at the origin - not a point at infinite height. See Figure 24.

Another "definition" of δ is that for any continuous function $f(x)$ it is supposed to give the formula

$$\int_{-\infty}^{\infty} f(t)\delta(t)dt = f(0).$$

We show in the exercises that this formula forces $\delta(x) = 0$ for $x \neq 0$. But if delta were really a function this would force

$$\int_{-\infty}^{\infty} f(t)\delta(t)dt = 0 \text{ and not } f(0).$$

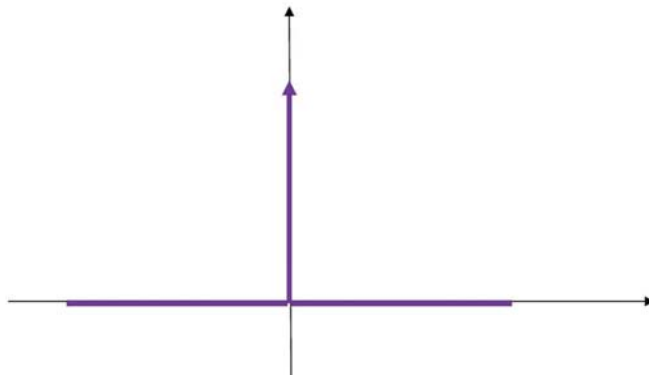


Figure 24: Dirac delta

Yet another definition of delta is that it is the identity for convolution. That is just as bad.

This disturbed mathematicians mightily in the 1930's and 1940's. Engineers and physicists just happily used delta and even its derivative or infinite sums of deltas. Finally Laurent Schwartz and others legalized delta and its relatives by calling it a distribution or generalized function and developing the calculus of distributions. See my book *Harmonic Analysis on Symmetric Spaces and Applications, I*, for an introduction. Other references are Korevaar, *Mathematical Methods*, Stakgold, *Green's Functions and Boundary Value Problems*, and Schwartz, *Math. for the Physical Sciences*. The subject of analysis on distributions does not seem to appear in undergrad math. courses.

So what to do? In this section we consider sequences of functions that have the behavior of delta in the limit as $n \rightarrow \infty$. These are called Dirac sequences or approximate identities (for convolution). In fact, Dirac did this himself.

Definition 59 A *Dirac sequence (of positive type)* $K_n(x)$ is a sequence of functions $K_n : \mathbb{R} \rightarrow \mathbb{R}$ such that

Dir 1. $K_n(x) \geq 0$, for all x and n .

Dir 2. $\int_{-\infty}^{\infty} K_n(x) dx = 1$, for all n .

Dir 3. $\forall \varepsilon > 0$ and $\forall \delta > 0$, $\exists N \in \mathbb{Z}^+$ s.t. $n \geq N$ implies

$$\int_{-\infty}^{-\delta} K_n(x) dx + \int_{\delta}^{\infty} K_n(x) dx < \varepsilon.$$

The 3 properties say that the area under the curve (which is 1) $y = K_n(x)$ becomes more and more concentrated at the origin as $n \rightarrow \infty$. You might think it hard to find such a sequence but we have the following examples.

Example 1. Let $K(x)$ be such that $K(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} K(x) dx = 1$. Then $K_n(x) = nK(nx)$ is a Dirac sequence.

Example 2. The Landau Kernel.

This kernel is named for the number theorist Edmund Landau (1877-1938) who was alive at the same time that Dirac wrote his important article introducing his function (1926-7).

Define the Landau kernel to be

$$L_n(x) = \begin{cases} \frac{(1-t^2)^n}{c_n}, & -1 \leq t \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad \text{where} \quad c_n = \int_{-1}^1 (1-t^2)^n dt.$$

It is possible to use integration by parts to find that

$$c_n = \frac{(n!)^2 2^{2n+1}}{(2n)!(2n+1)}.$$

See Courant and Hilbert, *Methods of Mathematical Physics*, Vol. I, p. 84. Unless you know Stirling's formula, this is not too helpful in proving the Landau kernel gives a Dirac sequence. Lang gives a simple inequality which is all we need. In Figure 25, we plot $L_n(t) = (1 - t^2)^n \frac{(2n)!(2n+1)}{(n!)^2 2^{2n+1}}$, for $t \in [-1, 1]$, when $n = 10$ in blue; 30 in magenta; 60 in green; 90 in turquoise, 120 in purple. Figure 25 shows the area under the curve (which is 1) begins to concentrate at the origin as n increases.

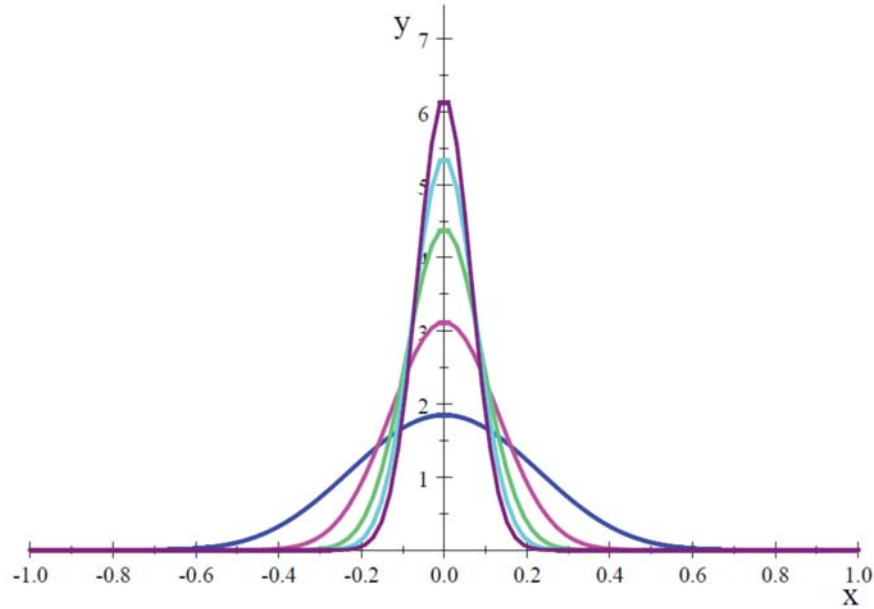


Figure 25: $L_n(t) = (1 - t^2)^n \frac{(2n)!(2n+1)}{(n!)^2 2^{2n+1}}$, for $t \in [-1, 1]$, when $n = 10$ in blue; 30 in magenta; 60 in green; 90 in turquoise, 120 in purple

Lemma. $c_n \geq \frac{2}{n+1}$.

Proof.

$$\frac{c_n}{2} = \int_0^1 (1 - t^2)^n dt = \int_0^1 (1 - t)^n (1 + t)^n dt \geq \int_0^1 (1 - t)^n dt = \frac{1}{n+1}.$$

■

It is clear that the Landau kernel has the first 2 properties of a Dirac sequence. To prove the 3rd property, we argue as in Lang, *Undergraduate Analysis*. Given $\varepsilon > 0$ and $0 < \delta < 1$, we need to find N so that $n \geq N$ makes the following integral $< \varepsilon$:

$$\frac{1}{c_n} \int_{\delta}^1 (1 - t^2)^n dt \leq \frac{n+1}{2} \int_{\delta}^1 (1 - t^2)^n dt \leq \frac{n+1}{2} \int_{\delta}^1 (1 - \delta^2)^n dt = \frac{n+1}{2} (1 - \delta^2)^n (1 - \delta).$$

We can make the stuff on the right $< \varepsilon$ for large n , since we can show that because $0 < 1 - \delta^2 < 1$, $\lim_{n \rightarrow \infty} \frac{n+1}{2} (1 - \delta^2)^n = 0$. To see this, you could use l'Hôpital's rule or just remember the appropriate fact about exponentials (if $c < 0$, then $xe^{cx} \rightarrow 0$, as $x \rightarrow \infty$).

27 A Dirac Sequence Approaches The Dirac Delta

We want to prove that any Dirac sequence K_n behaves like an identity for convolution in the limit as $n \rightarrow \infty$. Some people call Dirac sequences "approximate identities" for this reason.

Theorem 60 Suppose f is a bounded piecewise continuous function on \mathbb{R} . Let I be any finite interval on which f is continuous. Define $\|g\|_\infty = \max_{x \in I} |g(x)|$. Then $\lim_{n \rightarrow \infty} \|f - K_n * f\|_\infty = 0$. This says that the sequence $K_n * f$ converges uniformly to f on the interval I . Here the norm is taken on the interval I and we are assuming that the kernel vanishes off I .

Proof. Using the 1st and 2nd properties of Dirac sequences plus the fact that integrals preserve \leq , we have:

$$\begin{aligned} |(K_n * f)(x) - f(x)| &= \left| \int_{-\infty}^{\infty} K_n(t) f(x-t) dt - f(x) \int_{-\infty}^{\infty} K_n(t) dt \right| \\ &= \left| \int_{-\infty}^{\infty} K_n(t) (f(x-t) - f(x)) dt \right| \\ &\leq \int_{-\infty}^{\infty} K_n(t) |f(x-t) - f(x)| dt. \end{aligned}$$

Since f is uniformly continuous on I by Theorem 54, given ε there is a δ such that $|f(x-t) - f(x)| < \varepsilon$ when $|t| < \delta$. Since f is bounded, there is a bound M so that $|f(x)| \leq M$ for all x .

Now we can break up the last integral

$$\int_{-\infty}^{\infty} K_n(t) |f(x-t) - f(x)| dt = \int_{-\infty}^{-\delta} K_n(t) |f(x-t) - f(x)| dt + \int_{-\delta}^{\delta} K_n(t) |f(x-t) - f(x)| dt + \int_{\delta}^{\infty} K_n(t) |f(x-t) - f(x)| dt.$$

For the first 2 integrals, use the bound on f and the 3rd property of Dirac sequences to see that, for large enough n , they are less than $2M\varepsilon$ or even ε , if you prefer.

For the last integral, use the uniform continuity of f to see that the integral is

$$\leq \varepsilon \int_{-\delta}^{\delta} K_n(t) dt \leq \varepsilon \int_{-\infty}^{\infty} K_n(t) dt = \varepsilon.$$

This completes the proof that $|(K_n * f)(x) - f(x)| \leq (2M + 1)\varepsilon$. You can replace ε by $\frac{\varepsilon}{2M+1}$ if you are paranoid. ■

Corollary 61 Weierstrass Theorem. Any continuous function f on a finite closed interval $[a, b]$ can be uniformly approximated by polynomials.

Proof. We can use the preceding theorem for the interval $[0, 1]$ along with the Landau kernel. See Lang, *Undergraduate Analysis*, for a general interval. Suppose f is continuous on $[0, 1]$ and vanishes outside $[0, 1]$. We still need to see that $L_n * f$ is a polynomial. Suppose $x \in [0, 1]$. Then

$$\int_{-\infty}^{\infty} L_n(x-t) f(t) dt = \int_0^1 L_n(x-t) f(t) dt.$$

Also we see that for $x, t \in [0, 1]$, we have $-1 \leq x-t \leq 1$ and so $L_n(x-t) = (1 - (x-t)^2)^n$ and the integral is

$$\frac{1}{c_n} \int_0^1 (1 - (x-t)^2)^n f(t) dt.$$

This is a polynomial using the same reasoning as in our proof in the 1st section that convolution of f with any polynomial is a polynomial. ■

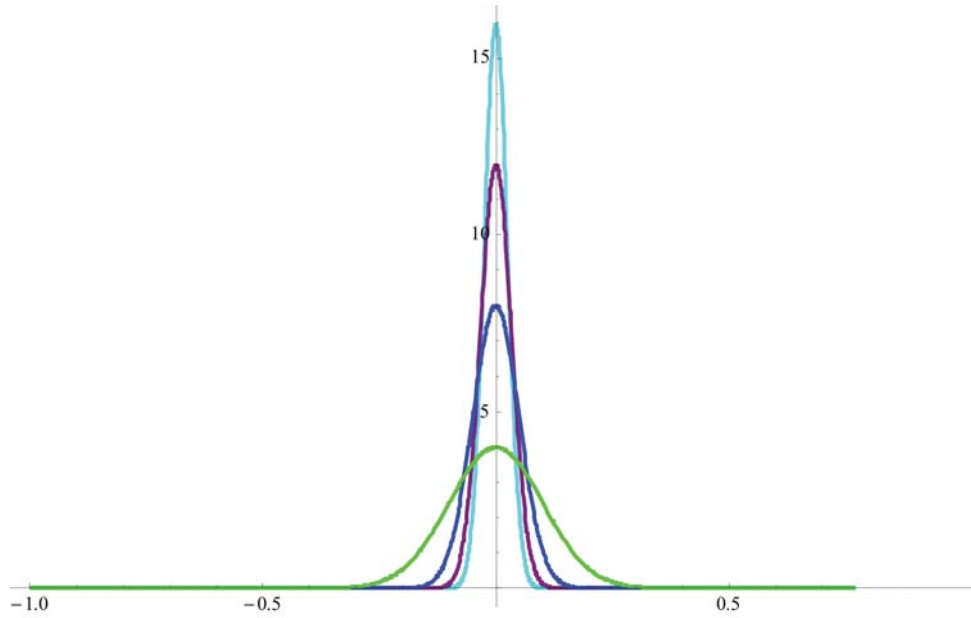


Figure 26: $G_\sigma(x) = \frac{1}{2\pi\sigma} e^{-\frac{x^2}{2\sigma^2}}$, where $\sigma = 1/10, 1/20, 1/30, 1/40$.

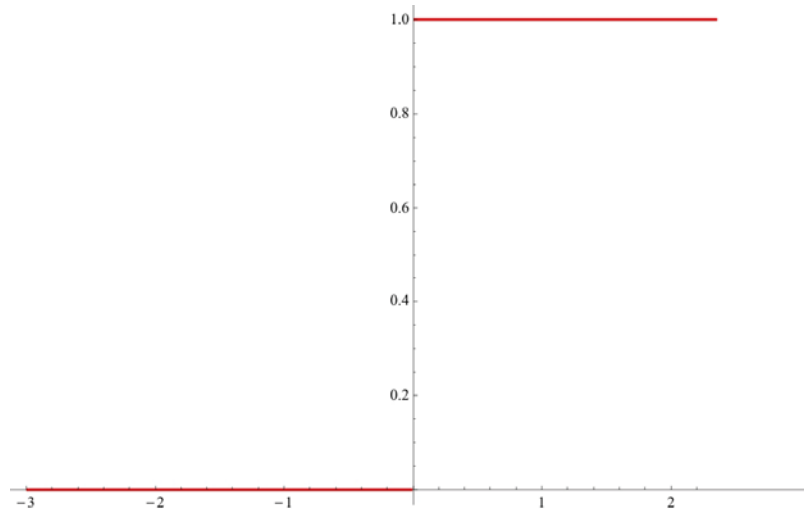


Figure 27: The unit step function $U(x) = 1$ if $x > 0$, $U(x) = 0$, otherwise.

Example. The Gauss Kernel.

One can show (exercise) that the **Gauss kernel (or normal density)** $G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ approaches the Dirac delta as $\sigma \rightarrow 0$. Figure 26 shows the kernel for $\sigma = 1/10, 1/20, 1/30, 1/40$.

As an example, let's approach the Unit Step function in Figure 27.

Convolutions of the Gaussians from Figure 26 and the Unit Step function are shown in Figure 28.

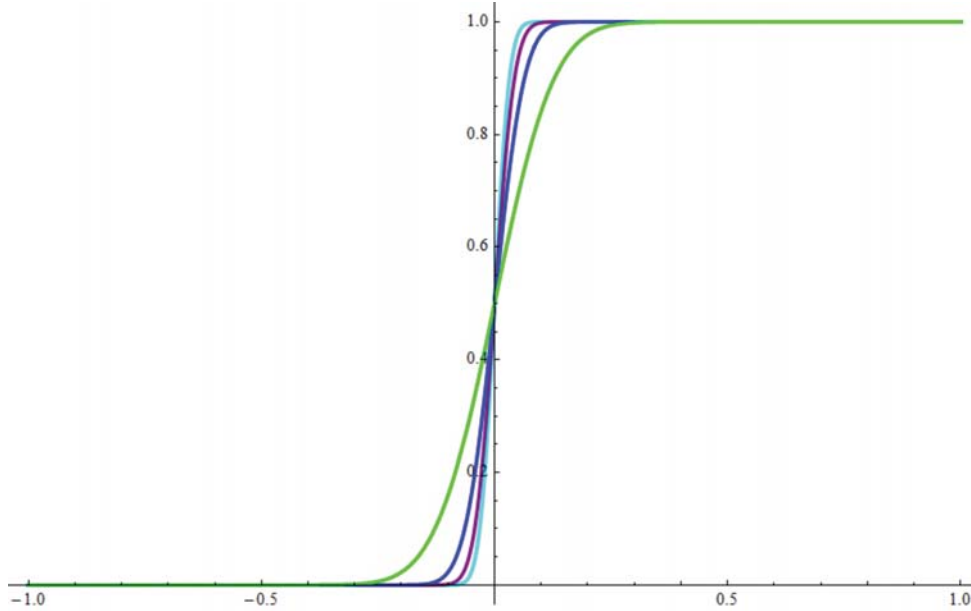


Figure 28: $G_\sigma * U$, for $\sigma = \frac{1}{10}, \frac{1}{20}, \frac{1}{30}, \frac{1}{40}$.

Part VII

Fourier Series

We said a bit about Fourier series in the introduction. Fourier's original paper was published in 1822. The publication was delayed by mathematicians such as Lagrange. See the book on Fourier by Grattan-Guinness and Ravetz. It may seem odd that people worried so much about Fourier expansions when they did not worry about Taylor expansions.

It will be slightly easier if we allow complex-valued functions like $f(x) = e^{2\pi i n x}$, where $i = \sqrt{-1}$. This means we need to allow vector spaces with complex scalars.

28 Inner Product Spaces with Complex Scalars

We want to look at vector spaces V with complex (or real) scalars. This means that V has all the axioms for addition and multiplication by scalars given earlier. Such a vector space is said to have a **Hermitian scalar product** $\langle v, w \rangle$ for $v, w \in V$ if it satisfies the following axioms.

Axioms for a Hermitian Scalar Product

For all $z, w, u \in V$ and $\alpha, \beta \in \mathbb{C}$ we have:

Axiom P1. $\langle z, w \rangle \in \mathbb{C}$ and $\langle z, w \rangle = \overline{\langle w, z \rangle}$. Here $\alpha = u + iv \in \mathbb{C}$ has complex conjugate $\bar{\alpha} = u - iv$.

Axiom P2. $\langle \alpha z + \beta w, u \rangle = \alpha \langle z, u \rangle + \beta \langle w, u \rangle$.

Axiom P3. $\langle z, z \rangle \geq 0$ and $\langle z, z \rangle = 0 \iff z = 0$.

Definition 62 As in the case of real scalars, we will say that 2 vectors $z, w \in V$ are **orthogonal** if $\langle z, w \rangle = 0$.

Theorem 63 If V is a vector space with complex scalars and $\langle z, w \rangle$ denotes an Hermitian scalar product on V , then $\|z\| \doteq \sqrt{\langle z, z \rangle}$ gives a norm on V according to our usual axioms for norms on vector spaces.

Proof. We need to check the 3 norm axioms.

- N1. $\|v\| \geq 0 \forall v \in V$ and $\|v\| = 0$ if and only if $v = 0$.
 N2. $\|\alpha v\| = |\alpha| \|v\|$, $\forall v \in V$ and $\forall \alpha \in \mathbb{R}$. Here $|\alpha| =$ complex absolute value of α .
 N3. $\|u + v\| \leq \|u\| + \|v\|$, $\forall u, v \in V$. Triangle Inequality.

The first follows from Property 3 of the scalar product. To see the second, note that if $\alpha \in \mathbb{C}$

$$\|\alpha z\| = \sqrt{\langle \alpha z, \alpha z \rangle} = \sqrt{\alpha \bar{\alpha}} \|z\|.$$

We are now done since the complex absolute value of α is $\sqrt{\alpha \bar{\alpha}}$.

In order to prove N3, we need the following theorem. The proof will then be an **exercise**. ■

Theorem 64 Cauchy-Schwarz Inequality for Complex scalar Product Spaces. Suppose $\langle z, w \rangle$ is an Hermitian scalar product on the complex vector space V . Then for all $z, w \in V$, we have:

$$|\langle z, w \rangle| \leq \|z\| \|w\|.$$

Proof. First note that if v, w are non-0 vectors in V , then there is a unique scalar $\alpha \in \mathbb{C}$ such that $v - \alpha w$ is orthogonal to w . The scalar is $\alpha = \frac{\langle v, w \rangle}{\langle w, w \rangle}$. It is an **exercise** to check this statement.

Next note that we have the **Pythagorean Theorem** for this general situation. It says that if 2 vectors u, v are orthogonal, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$. This is another **exercise**, requiring you to expand $\|u + v\|^2 = \langle u + v, u + v \rangle$ and make use of orthogonality of u and v .

To prove Cauchy-Schwarz, write $v = v - \alpha w + \alpha w$, where $\alpha = \frac{\langle v, w \rangle}{\langle w, w \rangle}$. Then

$$\|v\|^2 = \|v - \alpha w\|^2 + \|\alpha w\|^2 \geq \|\alpha w\|^2.$$

This means

$$\|v\|^2 \geq |\alpha|^2 \|w\|^2 = \frac{\langle v, w \rangle^2}{\|w\|^4} \|w\|^2 = \frac{\langle v, w \rangle^2}{\|w\|^2}.$$

The Theorem follows immediately upon multiplying by $\|w\|^2$. ■

Example 1. Let $\mathbb{C}^2 = \left\{ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mid z_j \in \mathbb{C} \right\}$. Define the sum of 2 vectors $z, w \in \mathbb{C}^2$ by

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} z_1 + w_1 \\ z_2 + w_2 \end{pmatrix}$$

and the product with $\alpha \in \mathbb{C}$ by

$$\alpha \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \alpha z_1 \\ \alpha z_2 \end{pmatrix}.$$

The Hermitian scalar product on \mathbb{C}^2 is defined by

$$\left\langle \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \right\rangle = z_1 \bar{w}_1 + z_2 \bar{w}_2.$$

We leave it as an exercise for you to check that this satisfies the 3 axioms for an Hermitian scalar product.

Example 2. Let $V = C[0, 1] = \{f : [0, 1] \rightarrow \mathbb{C}\}$. If $f, g \in V$ and $\alpha \in \mathbb{C}$, define $(f + g)(x) = f(x) + g(x)$ and $(\alpha f)(x) = \alpha f(x)$, for all $x \in [0, 1]$. Then define the Hermitian scalar product by

$$\langle f, g \rangle = \int_0^1 f(x) \overline{g(x)} dx.$$

It is easy to integrate a complex valued function $h(x) = u(x) + iv(x)$, with $u(x)$ and $v(x)$ real valued. The definition is

$$\int_a^b h = \int_a^b (u + iv) = \int_a^b u + i \int_a^b v.$$

It is an **exercise** to check the axioms for a Hermitian scalar product.

29 Complex Exponentials

Our Fourier series will involve complex exponentials because it would be twice as much work to write series of sines and cosines. We can define for $z \in \mathbb{C}$, the **complex exponential** by the same old Taylor series, except that now all the terms are complex:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

This series converges absolutely and uniformly in the complex plane by the complex version of the Weierstrass M-Test. The same proof that worked for real exponentials shows that $e^{z+w} = e^z e^w$. One also has $(e^z)^w = e^{zw}$. And, as before, if y is real, $e^{iy} = \cos y + i \sin y$.

Complex conjugation $c(z) = \bar{z}$ is a continuous function satisfying $\overline{z\bar{w}} = \bar{z}w$. This means that if $p(z)$ is a polynomial with real coefficients, then $\overline{p(z)} = p(\bar{z})$. Furthermore $\overline{e^z} = e^{\bar{z}}$. It is an **exercise** to prove the preceding statements.

The complex exponentials $e_n(x) = e^{2\pi i n x}$, $n = 0, \pm 1, \pm 2, \pm 3, \dots$ form a set of mutually orthogonal functions in $C[0, 1] = \{f : [0, 1] \rightarrow \mathbb{C}\}$ using the scalar product defined in Example 2 of the last section. To see this, suppose $m \neq n$ and do the integral:

$$\langle e_n, e_m \rangle = \int_0^1 e^{2\pi i n x} \overline{e^{2\pi i m x}} dx = \int_0^1 e^{2\pi i (n-m)x} dx = \frac{1}{2\pi i (n-m)} e^{2\pi i (n-m)x} \Big|_{x=0}^1 = 0.$$

Here we are assuming a complex version of the fundamental theorem of calculus, but you can check that it works given our definitions of these integrals. It is certainly simpler to do things this way than to show the analogous properties of $\cos(2\pi n x)$ and $\sin(2\pi m x)$, using trigonometric identities. We also need to know that $1 = e^{2\pi i n}$, for any integer n .

If $n = m$, we find that

$$\|e_n\|^2 = \langle e_n, e_n \rangle = \int_0^1 e^{2\pi i n x} \overline{e^{2\pi i n x}} dx = \int_0^1 1 dx = 1.$$

Our goal is to expand a function $f \in V = C[0, 1]$ in a **Fourier series**

$$f(x) = \sum_{n=-\infty}^{\infty} \gamma_n e^{2\pi i n x}, \quad \text{with} \quad \gamma_n = \langle f, e_n \rangle = \int_0^1 f(t) e^{-2\pi i n t} dt. \quad (6)$$

Fourier claimed to be able to expand "arbitrary" functions in his series. That will not be possible if we want the series in Formula (6) to converge pointwise or better yet uniformly. For applications we might want even better convergence in that we may want to be able to differentiate term-by-term. Generally we will find that the more derivatives f has, the faster the Fourier series will converge.

If you hate complex numbers and you want to keep everything real, use $e^{ix} = \cos x + i \sin x$ and write the Fourier series as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(2\pi n x) + b_n \sin(2\pi n x)),$$

where for $n = 1, 2, 3, \dots$

$$a_n = \int_0^1 f(t) \cos(2\pi n t) dt \quad \text{and} \quad b_n = \int_0^1 f(t) \sin(2\pi n t) dt. \quad (7)$$

Exercise. Using Taylor series prove that $e^{iy} = \cos y + i \sin y$.

Exercise. Show that if a_n, b_n , and c_n are defined as in equations (6) and (7), then they are related as follows for $n = 1, 2, 3, 4, \dots$

$$c_n = \frac{1}{2} (a_n - i b_n), \quad c_{-n} = \frac{1}{2} (a_n + i b_n).$$

30 Facts about Generalized Fourier Series in Hermitian scalar Product Spaces

We do some infinite dimensional linear algebra here. You should be familiar with the finite dimensional version from calculus. The set-up is that V is a complex vector space with Hermitian scalar product $\langle v, w \rangle$. Given an infinite set $\{v_1, v_2, v_3, \dots\}$ of pairwise orthogonal vectors in V , we want to express an arbitrary vector as an infinite sum $\sum_{n=1}^{\infty} \gamma_n v_n$, with scalars $\gamma_n \in \mathbb{C}$.

Theorem 65 Facts about Generalized Fourier Series in Infinite Dimensional scalar Product Spaces

Suppose $\{v_1, v_2, v_3, \dots\}$ are non-zero elements of V and pairwise orthogonal; i.e. $\langle v_n, v_m \rangle = 0$ if $n \neq m$.

Fact 1. Non-0 Orthogonal elements of V are linearly independent.

For any K , if there are scalars $\gamma_n \in \mathbb{C}$ such that $\sum_{n=1}^K \gamma_n v_n = 0$, then $\gamma_n = 0$ for all $n = 1, 2, \dots, K$. This says that

$\{v_1, v_2, v_3, \dots\}$ is a linearly independent set.

Fact 2. Formula for the Fourier Coefficients.

And suppose that a vector $z \in V$ has an expression

$$z = \sum_{n=1}^{\infty} \gamma_n v_n, \text{ with } \gamma_n \in \mathbb{C},$$

(converging with respect to the norm $\|w\| = \sqrt{\langle w, w \rangle}$). Then the n th generalized Fourier coefficient is

$$\gamma_j = \frac{\langle z, v_j \rangle}{\langle v_j, v_j \rangle}.$$

Fact 3. Bessel's Inequality. Assume that the set is **orthonormal**, meaning that $\|v_n\| = 1$, for all n . Then we have **Bessel's inequality**

$$\sum_{n=0}^{\infty} |\gamma_n|^2 \leq \|z\|^2.$$

Fact 4. Parseval's Equality (or the Plancherel Theorem).

Assume that the set is orthonormal. Also assume that $\{v_1, v_2, v_3, \dots\}$ is **complete**; that is, any vector $z \in V$ has an expression $z = \sum_{n=1}^{\infty} \gamma_n v_n$, with $\gamma_n \in \mathbb{C}$. Then we have **Parseval's equality**:

$$\sum_{n=0}^{\infty} |\gamma_n|^2 = \|z\|^2.$$

Proof. Fact 1. If $\sum_{n=1}^K \gamma_n v_n = 0$, take the scalar product with v_j to get

$$0 = \left\langle \sum_{n=1}^K \gamma_n v_n, v_j \right\rangle = \sum_{n=1}^K \gamma_n \langle v_n, v_j \rangle = \gamma_j \langle v_j, v_j \rangle.$$

The 1st equality holds since $\langle 0, w \rangle = 0$ for any vector w . The 2nd holds by the linearity of the scalar product in the 1st variable. The last equality holds by the pairwise orthogonality of the vectors v_n . Since $v_j \neq 0$, it follows that $\gamma_j = 0$.

Fact 2. Look at the scalar product $\langle z, v_j \rangle$:

$$\langle z, v_j \rangle = \left\langle \sum_{n=1}^{\infty} \gamma_n v_n, v_j \right\rangle = \sum_{n=1}^{\infty} \gamma_n \langle v_n, v_j \rangle = \gamma_j \langle v_j, v_j \rangle.$$

We can move limits such as infinite sums in and out of the scalar product $\langle v, w \rangle$ because the scalar product is continuous in v holding w fixed (**exercise** using the Cauchy-Schwarz inequality). Again, we have used the linearity of the scalar product in the 1st variable and the pairwise orthogonality of the vectors v_n . Solve for γ_j to finish the proof.

Fact 4. Look at the scalar product $\langle z, z \rangle$:

$$\langle z, z \rangle = \left\langle \sum_{n=1}^{\infty} \gamma_n v_n, \sum_{m=1}^{\infty} \gamma_m v_m \right\rangle = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \gamma_n \overline{\gamma_m} \langle v_n, v_m \rangle = \sum_{n=0}^{\infty} |\gamma_n|^2.$$

Here we used the continuity and linearity of the scalar product in each variable holding the other variable fixed as well as the pairwise orthogonality of the v_n plus the fact that $\|v_n\| = 1, \forall n$.

We leave the proof of Fact 3 as an **exercise**. ■

Example. Let $V = C[0, 1] = \{f : [0, 1] \rightarrow \mathbb{C} \mid f \text{ is continuous}\}$. Note that we can identify this space with $C[-\frac{1}{2}, \frac{1}{2}]$ or $C(I)$ for any other closed interval I of length 1.

Define the scalar product as $\langle f, g \rangle = \int_0^1 f(t) \overline{g(t)} dt$. We have an orthonormal set $S = \{e_n(t) = e^{2\pi i n t} \mid n \in \mathbb{Z}\}$. The

norm induced on this space by the scalar product is the **L^2 -norm** given by

$$\|f\|_2 = \sqrt{\int_0^1 |f(t)|^2 dt}.$$

Thanks to the preceding theorem, once we have proved that S is a complete orthonormal set, then we will know that any function $f \in C[0, 1]$ has a Fourier series converging in the L^2 -norm. That is

$$f(x) = \sum_{n=-\infty}^{\infty} \gamma_n e^{2\pi i n x}, \text{ with } \gamma_n = \langle f, e_n \rangle = \int_0^1 f(t) e^{-2\pi i t} dt.$$

For the applications to partial differential equations that we want to discuss, we will need the Fourier series to converge pointwise at each x , along with all the necessary derivatives. We will need extra hypotheses on f to insure such nice convergence.

You can also view the k th partial sum of the generalized Fourier series for z in Theorem 65 as giving best approximations to z in the subspace of $\sum_{n=1}^K \alpha_n v_n, \alpha_n \in \mathbb{C}$, spanned by the v_n 's. Here "best" means that the mean square error $\left\| z - \sum_{n=1}^K \alpha_n v_n \right\|$ is smallest when $\alpha_j = \gamma_j = \langle z, v_j \rangle$. Another way to say this is to say that the **Fourier coefficients give the best least squares fit to z** . We leave the proof of this as an **exercise** for the reader.

31 When do we have a Complete Orthonormal Set of Vectors in our Scalar Product Space?

First recall our definition from the preceding section.

Definition 66 A **complete orthonormal set** $\{v_n\}_{n \geq 1}$ in an scalar product space V means that $\{v_n\}_{n \geq 1}$ is an orthonormal set of vectors such that every vector $z \in V$ can be expanded in a generalized Fourier series

$$z = \sum_{n=1}^{\infty} \gamma_n v_n, \text{ where } \gamma_n = \langle z, v_n \rangle.$$

Here convergence of the series means convergence with respect to the norm $\|w\| = \sqrt{\langle w, w \rangle}$.

Theorem 67 **3 Equivalent Conditions for an Orthogonal Set to be Complete.**

Suppose $\{v_n\}_{n \geq 1}$ is an orthonormal set in an scalar product space V . The following are equivalent:

1) $\{v_n\}_{n \geq 1}$ is complete.

2) Parseval's Equality holds: $\sum_{n=0}^{\infty} |\gamma_n|^2 = \|z\|^2$ if $\gamma_n = \langle z, v_n \rangle$.

3) No non-0 vector $w \in V$ is orthogonal to all the v_n , for $n = 1, 2, 3, \dots$

Exercise. Prove the preceding theorem. **Hint:** You need only show $1) \implies 2) \implies 3) \implies 1)$.

In order to prove that our favorite sequence $\{e^{2\pi i n x}, n \in \mathbb{Z}\}$ is complete for the scalar product space $C[0, 1]$, we will need to use the theory of convolution and Dirac sequences studied earlier. But now we need different kernels from the Landau kernel or the Gauss kernel.

Definition 68 The *n th partial sum* of the Fourier series of $f \in C[0, 1]$ is defined to be

$$s_n(x) = \sum_{k=-n}^n \gamma_k e^{2\pi i k x}, \text{ with } \gamma_k = \langle f, e_k \rangle = \int_0^1 f(t) e^{-2\pi i k t} dt.$$

Note that the n th partial sum is always taken to be the symmetric sum of terms from $-n$ to $+n$. There are $2n + 1$ terms in the n th partial sum.

Definition 69 The *Dirichlet kernel* is defined to be

$$D_n(x) = \sum_{k=-n}^n e^{2\pi i k x}.$$

Definition 70 The *Fejér (or Cesàro) kernel* is

$$K_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} D_k(x).$$

Theorem 71 Properties of the Dirichlet and Fejér Kernels.

Property 1. Suppose $f \in C[-\frac{1}{2}, \frac{1}{2}]$. Then if we assume that convolution is over the interval $[-\frac{1}{2}, \frac{1}{2}]$ and that both K_n and f vanish off this interval, we find that the convolution of f and the Dirichlet kernel is the n th partial sum of the Fourier series of f ; i.e.,

$$(f * D_n)(x) = \sum_{k=-n}^n e^{2\pi i k x} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t) e^{-2\pi i k t} dt = s_n.$$

The convolution of f and the Fejér kernel is the average of the first n partial sums of the Fourier series of f ; i.e.

$$f * K_n = \frac{1}{n} \{s_0 + s_1 + \cdots + s_{n-1}\}.$$

Property 2.

$$D_n(x) = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}; \quad K_n(x) = \frac{1}{n} \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)^2.$$

Proof. Property 1.

$$(f * D_n)(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t) \sum_{k=-n}^n e^{2\pi i k(x-t)} dt = \sum_{k=-n}^n e^{2\pi i k x} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t) e^{-2\pi i k t} dt.$$

We leave the rest as an **exercise**.

Property 2. We can use the formula for the sum of a geometric progression to see that

$$\begin{aligned} D_n(x) &= \sum_{k=-n}^n e^{2\pi i k x} = e^{-2\pi i n x} \sum_{k=-n}^n e^{2\pi i(k+n)x} = e^{-2\pi i n x} \sum_{j=0}^{2n} e^{2\pi i j x} \\ &= e^{-2\pi i n x} \frac{e^{(2n+1)2\pi i x} - 1}{e^{2\pi i x} - 1} = e^{-2\pi i n x} e^{-\pi i x} \frac{e^{(2n+1)2\pi i x} - 1}{e^{\pi i x} - e^{-\pi i x}} \\ &= \frac{e^{(2n+1)\pi i x} - e^{-(2n+1)\pi i x}}{e^{\pi i x} - e^{-\pi i x}} = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}. \end{aligned}$$

We leave it as an **exercise** to prove the formula for K_n . ■

A plot of $D_{10}(x) = \frac{\sin((2*10+1)\pi x)}{\sin(\pi x)}$ is in Figure 29. Plots of $D_n(x) = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}$, $n = 1, 2, \dots, 10$ are in Figure 30. A plot of $K_{10}(x) = \frac{1}{10} \left(\frac{\sin(10\pi x)}{\sin(\pi x)} \right)^2$ is in Figure 31. Plots of $K_n(x) = \frac{1}{n} \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)^2$, $n = 1, 2, \dots, 10$, are in Figure 32.

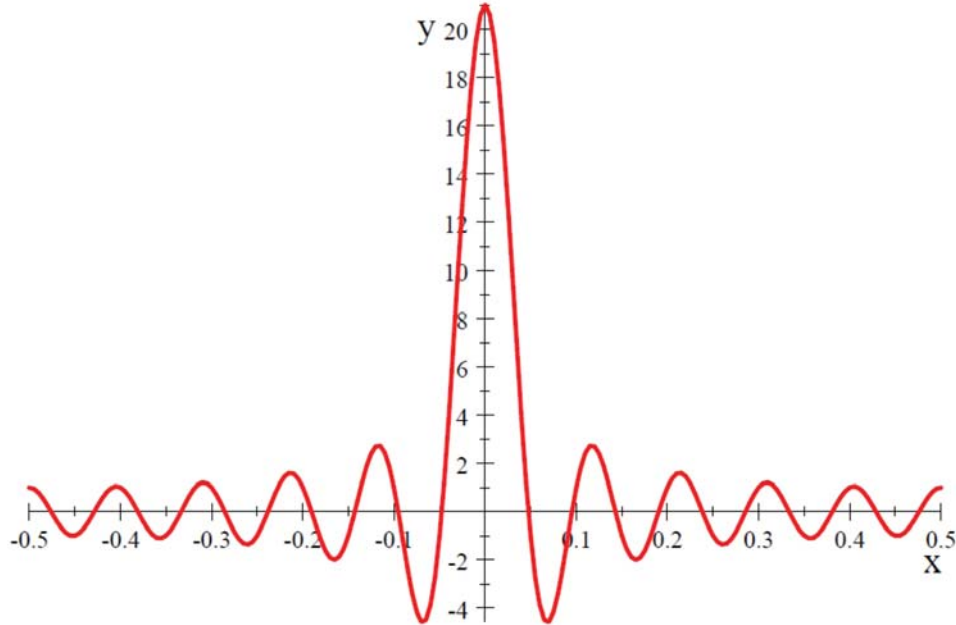


Figure 29: A plot of $D_{10}(x) = \frac{\sin((2*10+1)\pi x)}{\sin(\pi x)}$.

Figure 32 may make you believe the following Theorem.

Theorem 72 *The Fejér kernels form a Dirac sequence of positive type.*

Proof. Using

$$K_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=-k}^k e^{2\pi i j x} = \frac{1}{n} \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)^2,$$

we need to check 3 things from Definition 59. These are:

Dir 1. $K_n(x) \geq 0$, for all x and n .

Dir 2. $\int_{-\infty}^{\infty} K_n(x) dx = 1$, for all n .

Dir 3. $\forall \varepsilon > 0$ and $\forall \delta > 0$, $\exists N \in \mathbb{Z}^+$ s.t. $n \geq N$ implies

$$\int_{-\frac{1}{2}}^{-\delta} K_n(x) dx + \int_{\delta}^{\frac{1}{2}} K_n(x) dx < \varepsilon.$$

Dir1 is clear. We leave Dir 2 to the reader as an **exercise**. That leaves Dir 3. Since $K_n(-x) = K_n(x)$ and we are assuming everything vanishes outside $[-\frac{1}{2}, \frac{1}{2}]$, we just need to note that

$$\frac{1}{n} \int_{\delta}^{\frac{1}{2}} \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)^2 dx \leq \frac{1}{n} \int_{\delta}^{\frac{1}{2}} \frac{1}{\sin^2(\pi x)} dx \rightarrow 0, \text{ as } n \rightarrow \infty.$$

■

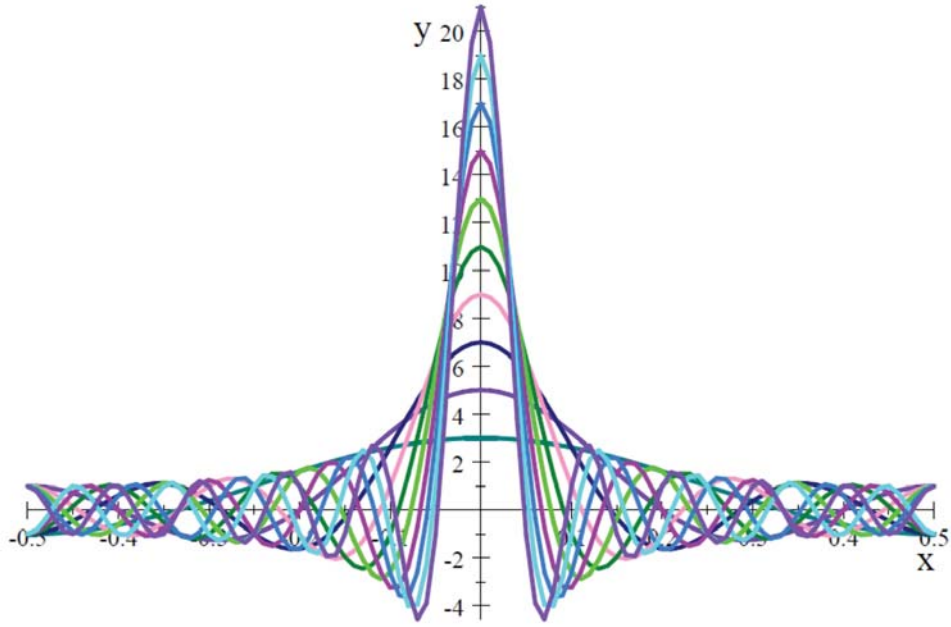


Figure 30: Plots of $D_n(x) = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}$, $n = 1, 2, \dots, 10$.

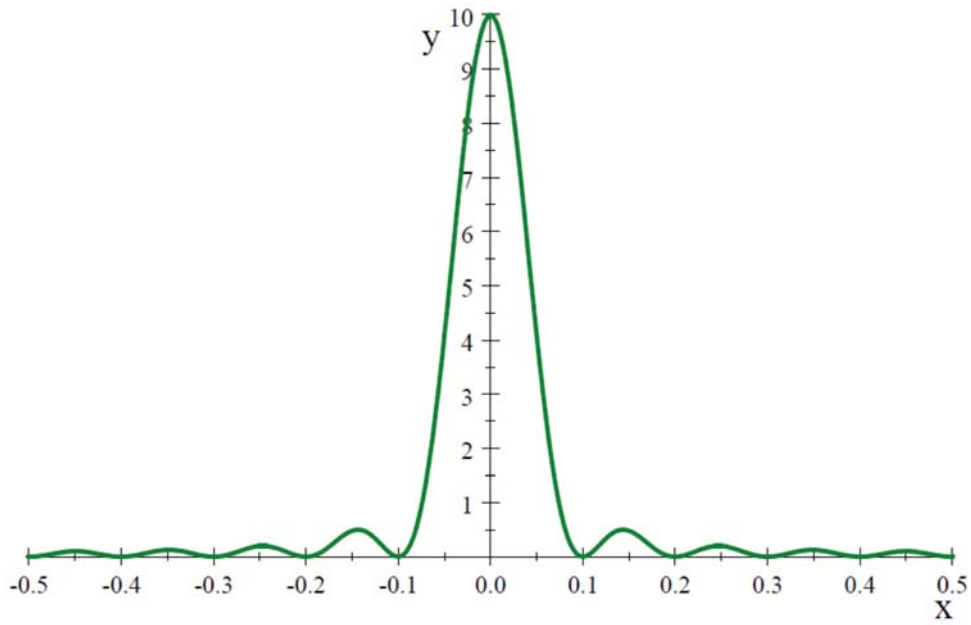


Figure 31: A plot of $K_{10}(x) = \frac{1}{10} \left(\frac{\sin(10\pi x)}{\sin(\pi x)} \right)^2$.

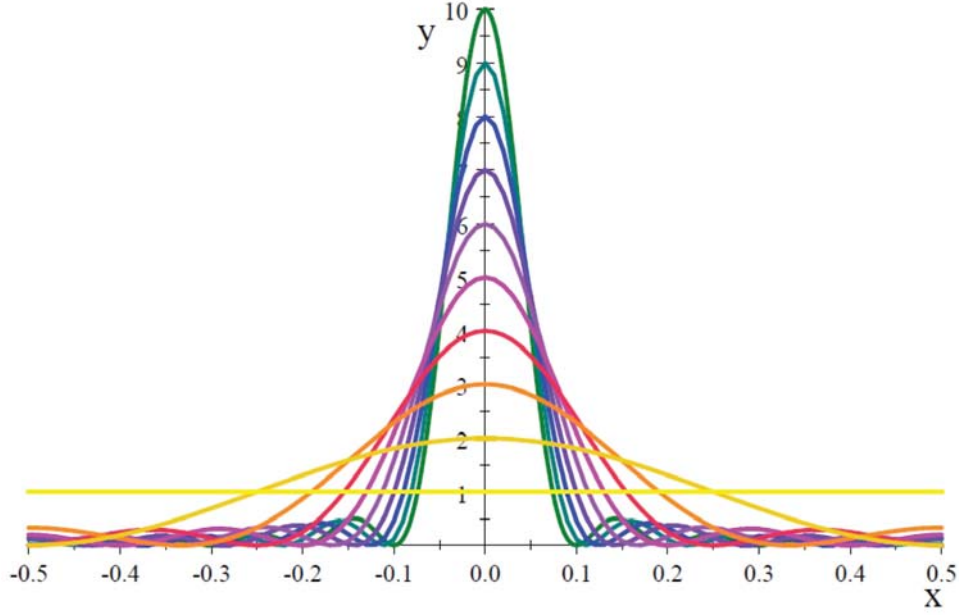


Figure 32: Plots of $K_n(x) = \frac{1}{n} \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)^2$, $n = 1, 2, \dots, 10$.

Theorem 73 (Fejér's Theorem, 1904) Suppose that $f \in C[0, 1]$ and $f(0) = f(1)$. Then if K_n denotes the Fejér kernel,

$$f * K_n \text{ approaches } f \text{ uniformly on } [0, 1] \text{ as } n \rightarrow \infty.$$

This means $\lim_{n \rightarrow \infty} \|f * K_n - f\|_\infty = 0$. Here, for bounded functions g on $[0, 1]$, we define $\|g\|_\infty = \text{lub}\{|g(x)| \mid x \in [0, 1]\}$. One says that the Fourier series of f is Cesàro summable to f where f is continuous.

Proof. This follows from the fact that K_n is a Dirac sequence of positive type by Theorem 60. ■

It follows from Fejér's Theorem 73 that if $f \in C[0, 1]$ and $f(0) = f(1)$ then f can be uniformly approximated by trigonometric polynomials of the form

$$\sum_{j=-k}^k \gamma_j e^{2\pi i j x}, \text{ for } \gamma_j \in \mathbb{C}.$$

For

$$(f * K_n)(x) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=-k}^k \gamma_j e^{2\pi i j x}, \text{ where } \gamma_j = \int_0^1 f(x) e^{-2\pi i j x} dx.$$

Corollary 74 The set $\{e^{2\pi i n x} \mid n \in \mathbb{Z}\}$ is a complete orthonormal set for $C[0, 1]$.

Proof. If $w \in C[0, 1]$ is orthogonal to $e^{2\pi i n x}$ for all n , it follows from Fejér's theorem that w must be 0. ■

Uniform convergence implies L^2 -convergence. Thus, if $f \in C[0, 1]$ satisfies $f(0) = f(1)$, then it can be uniformly approximated by trigonometric polynomials which implies that it can be approximated in L^2 -norm by trigonometric polynomials. But this means

$$f(x) = \sum_{n=-\infty}^{\infty} \gamma_n e^{2\pi i n x}, \text{ with } \gamma_n = \langle f, e_n \rangle = \int_0^1 f(t) e^{-2\pi i n t} dt. \quad (8)$$

Here convergence of the series is with respect to the L^2 -norm.

Warning: When we say that the convergence of the Fourier series in formula (8) is in the L^2 -norm, this does not suffice usually for what we are thinking is happening. We always hope to have uniform convergence, but we know that convergence with respect to $\|\cdot\|_\infty$ is harder to achieve.

Maybe we should write some other symbol than $=$ as we have not claimed that the Fourier series converges for every point $x \in [0, 1]$. That is false.

Note that we need not assume that f has period 1 at least if all we want is $\|\cdot\|_2$ -norm convergence, since it is possible to replace any $f \in C[0, 1]$ with a $g \in C[0, 1]$ such that $g(0) = g(1)$ and such that $\|g - f\|_2$ is arbitrarily small. This is an **exercise**.

The following Corollaries follow from Theorem 65 and Corollary 74.

Corollary 75 (Parseval's Equality for Fourier Series) Suppose that $f(x)$ is piecewise continuous on $[0, 1]$. Then

$$\int_0^1 |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |\gamma_n|^2,$$

$$\text{where } \gamma_n = \langle f, e_n \rangle = \int_0^1 f(t) e^{-2\pi i n t} dt.$$

Corollary 76 (Parseval's Equality for scalar Product)

Suppose that f and g are piecewise continuous on $[0, 1]$. Then

$$\int_0^1 f(t) \overline{g(t)} dt = \sum_{n=-\infty}^{\infty} \alpha_n \overline{\beta_n},$$

$$\text{if } \alpha_n = \langle f, e_n \rangle = \int_0^1 f(t) e^{-2\pi i n t} dt \quad \text{and} \quad \beta_n = \langle g, e_n \rangle = \int_0^1 g(t) e^{-2\pi i n t} dt.$$

Exercise. Prove the preceding Corollary.

32 Pointwise Convergence of Fourier Series

For the applications we need pointwise convergence of the Fourier series and even the ability to differentiate the Fourier series term-by-term. When is this legal? To figure such things out, consider the Dirichlet kernel defined in the last section:

$$D_n(x) = \sum_{k=-n}^n e^{2\pi i k x} = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}.$$

Recall that we showed in the last section that the convolution of f with the Dirichlet kernel is the n th partial sum of the Fourier series of f when the period interval has length 1:

$$f * D_n = s_n = \sum_{k=-n}^n e^{2\pi i k x} \gamma_k, \quad \text{where } \gamma_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t) e^{-2\pi i k t} dt.$$

However, the Dirichlet kernel is not as nice as the Fejér kernel because it is not positive.

Before proving some theorems, let's look at a few examples.

Example 1. The **step function** $f(x) = \begin{cases} 1, & 0 < x \leq \pi, \\ -1, & -\pi \leq x < 0. \end{cases}$ Make this periodic of period 2π on the real line.

You need to make the change of variables $x = 2\pi y$ here. This changes the Fourier series and coefficients to

$$f(x) = \sum_{n=-\infty}^{\infty} \gamma_n e^{inx}, \quad \text{with } \gamma_n = \langle f, e_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int} dt.$$

Here $\gamma_0 = 0$ and the Fourier coefficient for $n \neq 0$ is

$$\begin{aligned}\gamma_n &= \frac{-1}{2\pi} \int_{-\pi}^0 e^{-int} dt + \frac{1}{2\pi} \int_0^{\pi} e^{-int} dt = \frac{1}{2\pi} \left(\left. \frac{-e^{-int}}{-in} \right|_{-\pi}^0 + \left. \frac{e^{-int}}{-in} \right|_0^{\pi} \right) \\ &= \frac{i}{2\pi n} (-1 + e^{in\pi} + e^{-in\pi} - 1).\end{aligned}$$

Then for $n = 1, 2, 3, \dots$

$$\gamma_n e^{inx} + \gamma_{-n} e^{-inx} = \frac{4(-1)^n \sin(nx)}{n\pi}.$$

Figure 33 shows a plot of the function $f(x)$ at the top left, then it shows the n th partial sums of the Fourier series for $n = 10, 50, 100$.

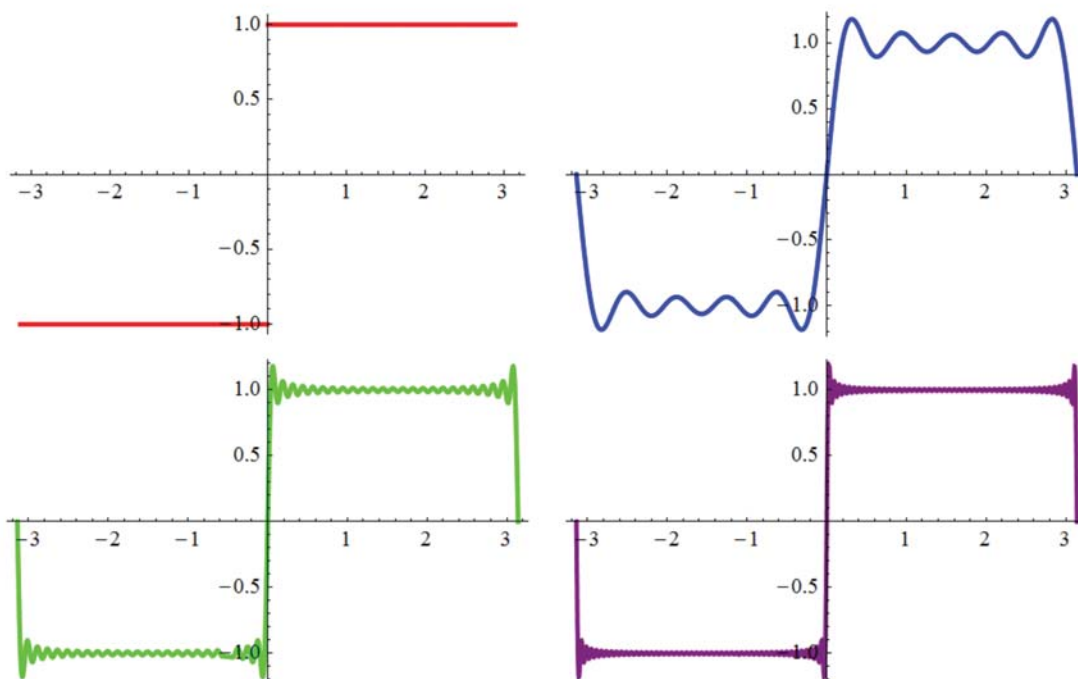


Figure 33: The function $f(x) = \begin{cases} 1, & 0 < x \leq \pi, \\ -1, & -\pi \leq x < 0. \end{cases}$ is shown at top left. The n th partial sums of the Fourier series of f for $n = 10, 50, 100$ are shown in the next 3 plots. The Gibbs phenomenon is revealed.

This reveals the Gibbs phenomenon which arises because the function has a jump discontinuity at 0. This sort of thing always happens at a jump discontinuity. It does not help to take more terms of the Fourier series. There will always be an overshoot of about 9%. See Dym and McKean, *Fourier Series and Integrals* for more details.

The phenomenon had been noted before Gibbs by various people. Gibbs explained the phenomenon, replying to Michaelson (in a letter to *Nature* in 1898). Michaelson had become angry that his machine for computing the 1st 80 terms in a Fourier series gave a result that was not close enough to the function near the jumps.

The next example shows that convergence is very nice when there are no jumps.

Example 2. Consider the function $f(x) = |x|$ for $|x| \leq \pi$. Make this periodic of period 2π on the real line.

The graphs in Figure 34 show the function plus the 10th, 50th and 100th partial sums of its Fourier series. The convergence is quite rapid.

Example 3. Let $f(x) = x$, if $x \in [0, 1)$ and make it periodic of period 1 elsewhere. This produces a sawtooth graph. See Figure 35.

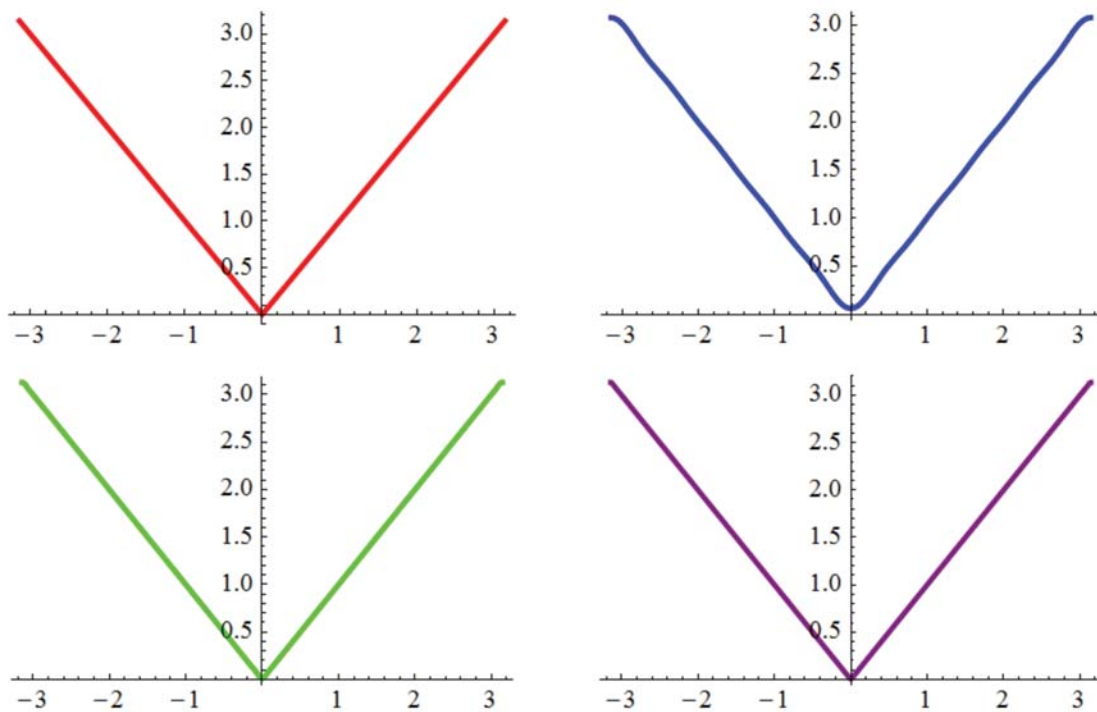


Figure 34: Define $f(x) = |x|$ for $|x| \leq \pi$. Make this periodic of period 2π on the real line. The graphs show at top left $f(x)$ on the period interval, then the 10th, 50th and 100th partial sums of its Fourier series.

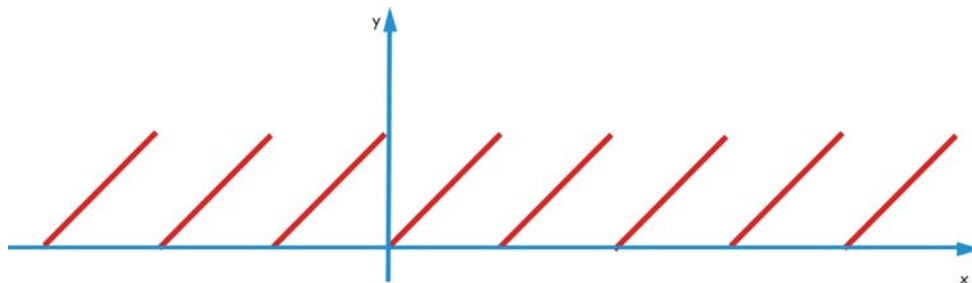


Figure 35: The sawtooth function - a function of period 1 defined by $f(x) = x$, for $x \in [0, 1)$.

One finds that the Fourier coefficients of the sawtooth function are

$$\gamma_n = \int_0^1 x e^{-2\pi i n x} dx = \begin{cases} \frac{1}{2}, & n = 0, \\ \frac{-1}{2\pi i n}, & n \neq 0. \end{cases}$$

Then the Parseval identity implies

$$\frac{1}{3} = \int_0^1 x^2 dx = \sum_{n=-\infty}^{\infty} |\gamma_n|^2 = \frac{1}{4} + \frac{2}{4\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Therefore we have Euler's formula

$$\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2} = \zeta(2).$$

The 1st step in showing the convergence of Fourier series of differentiable functions is the following Lemma which is also useful in quantum mechanics. We could weaken the hypothesis to assume only that the function f is Lebesgue integrable on $[0, 1]$.

Lemma 77 Riemann-Lebesgue Lemma. *Suppose that $f(x)$ is a piecewise continuous function on $[0, 1]$. Then*

$$\lim_{n \rightarrow \infty} \int_0^1 f(x) e^{2\pi i n x} dx = 0.$$

That is, the n th Fourier coefficient of f approaches 0 as n approaches ∞ .

Proof. Parseval's equality for f says that $\int_0^1 |f(x)|^2 dx = \sum_{n=-\infty}^{\infty} |\gamma_n|^2$. Since the integral is finite, the series must converge. It follows that the terms of the sum approach 0 as $n \rightarrow \infty$. ■

Note that the hypothesis of the Riemann-Lebesgue Lemma could be weakened considerably were we to discuss Lebesgue integrable functions.

Now we investigate the speed of convergence of the Fourier series to the function. We show that the more continuous derivatives f has, the faster the Fourier series converges to f . This sort of theorem goes back to Dirichlet (1829).

Theorem 78 Smoothness of f increases speed of convergence of Fourier series of f .

Suppose $f : [0, 1] \rightarrow \mathbb{C}$, with $f(0) = f(1)$. If the p th derivative of f is continuous for some $p \geq 1$, then the n th partial sum of the Fourier series of f converges uniformly to f ; i.e.,

$$s_n = \sum_{k=-n}^n e^{2\pi i k x} \gamma_k, \text{ where } \gamma_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t) e^{-2\pi i k t} dt$$

$$\|f - s_n\|_{\infty} \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Moreover there is a positive constant b such that

$$\|f - s_n\|_{\infty} \leq \frac{b}{n^{p-\frac{1}{2}}}.$$

Proof. We will take our interval to be $[-\frac{1}{2}, \frac{1}{2}]$. Since our functions have period 1, this is no problem. It is an **exercise** to show that if $D_n(x) = \sum_{k=-n}^n e^{2\pi i k x} = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}$, is the Dirichlet kernel, then

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} D_n(t) dt = 1.$$

By the same trick that was used in proving Fejér's theorem, we have

$$\begin{aligned} (s_n - f)(x) &= (D_n * f - f)(x) = \int_{-\frac{1}{2}}^{-\frac{1}{2}} (f(x-y) - f(x)) D_n(y) dy \\ &= \int_{-\frac{1}{2}}^{-\frac{1}{2}} \left\{ \frac{f(x-y) - f(x)}{\sin(\pi y)} \right\} \sin(\pi(2n+1)y) dy. \end{aligned}$$

Next note that the quantity in brackets is continuous at $y = 0$ (the only place in the interval $[-\frac{1}{2}, \frac{1}{2}]$ where something could go wrong), since $f'(x)$ exists and is continuous. The reason is that

$$\frac{f(x-y) - f(x)}{\sin(\pi y)} = \frac{f(x-y) - f(x)}{y} \frac{y}{\sin(\pi y)},$$

which approaches $-f'(x) \cdot \frac{1}{\pi}$ as $y \rightarrow 0$. Then we can apply the Riemann-Lebesgue Lemma to see that $(s_n - f)(x) \rightarrow 0$, as $n \rightarrow \infty$.

How do we estimate $\|f - s_n\|_\infty$? To do this use the fact that we now know that we have pointwise convergence of the Fourier series; i.e.,

$$f(x) = \sum_{n=-\infty}^{\infty} \gamma_n e^{2\pi i n x}, \text{ with } \gamma_n = \langle f, e_n \rangle = \int_0^1 f(t) e^{-2\pi i n t} dt.$$

If we write $\gamma_n = \widehat{f}(n)$, then it is an **exercise** to show that

$$\widehat{f^{(p)}}(n) = (2\pi i n)^p \widehat{f}(n). \tag{9}$$

Now we obtain our estimate as follows making use of the Cauchy-Schwarz inequality for the scalar product space of sequences $\{\alpha_n\}_{n \in \mathbb{Z}}$, where $\langle \{\alpha_n\}, \{\beta_n\} \rangle = \sum_{n=-\infty}^{\infty} \alpha_n \overline{\beta_n}$ and $\|\{\alpha_n\}\| = \sum_{n=-\infty}^{\infty} |\alpha_n|^2$. Cauchy-Schwarz says

$$|\langle \{\alpha_n\}, \{\beta_n\} \rangle| \leq \|\{\alpha_n\}\| \|\{\beta_n\}\|.$$

Now let's finish the proof. Note that using the triangle inequality and formula (9)

$$|(s - f_n)(x)| = \left| \sum_{|k|>n} \widehat{f}(k) e^{2\pi i k x} \right| \leq \sum_{|k|>n} |\widehat{f}(k)| = \sum_{|k|>n} \frac{|\widehat{f'}(k)|}{2\pi |k|}.$$

Now the Cauchy-Schwarz inequality and Bessel's inequality say that

$$|(s - f_n)(x)|^2 \leq \frac{1}{4\pi^2} \sum_{|k|>n} |\widehat{f'}(k)|^2 \sum_{|k|>n} \frac{1}{|k|^2} \leq \frac{1}{4\pi^2} \|f'\|_2^2 \sum_{|k|>n} \frac{1}{|k|^2}.$$

We know that $\sum_{|k|>n} \frac{1}{|k|^2} \leq \frac{1}{n}$ by the proof of the integral test. This completes the proof when $p = 1$. We leave it as an **exercise** to prove the result for general $p \geq 1$. ■

We can, in fact, allow the function f to have a jump discontinuity at a point c . Then, assuming that f has right- and left-hand derivatives at c , the Fourier series converges to $\frac{1}{2}(f(c^+) + f(c^-))$, where $f(c^+)$ denotes the right-hand limit and $f(c^-)$, the left-hand limit at c . We saw this in Figure 33. There are actually weaker hypotheses that suffice. See the many references on Fourier series for these things.

Exercises II

Exercises 1

- 1) a) Suppose $\|x\|_\alpha$ is a norm on a vector space E and $\|y\|_\beta$ is a norm on a vector space F . Show that you can make the Cartesian product $E \times F = \{(x,y) \mid x \in E \text{ and } y \in F\}$ into a vector space.
 b) Show that the Cartesian product $E \times F$ has a norm given by $\|(x,y)\| = \|x\|_\alpha + \|y\|_\beta$.
 c) Consider the special case that $E=F=\mathbb{R}$ with $\|x\|_\alpha = \|x\|_\beta = |x|$ = the ordinary absolute value of x .
 Using the definition of norm in b), draw a picture of the region in the plane given by the set of points $\{(x,y) \in \mathbb{R}^2 \mid \|(x,y) - (1,2)\| < 3\}$.

- 2) a) Suppose the E is an inner product space with inner product denoted by $\langle x,y \rangle$ for x,y in E .
 Then use the norm $\|v\| = (\langle v,v \rangle)^{1/2}$. Prove that $\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2$.

b) Let $E = \mathbb{R}^2$ and draw a picture to explain why the formula in part a) is called the parallelogram law.

- 3) Suppose that $E = C[0,1]$, the space of continuous real-valued functions on the interval $[0,1]$.

a) Prove that the L^1 norm $\|f\|_1 = \int_a^b |f(x)| dx$ is a norm.

b) Explain why $C[0,1]$ is an infinite dimensional vector space.

- 4) Take the vector space E of problem 3. Show that if $\|f\|_2 = \sqrt{\int_0^1 |f(x)|^2 dx}$, then we have

$\|f\|_1 \leq \|f\|_2$. Hint. Use the Cauchy-Schwarz inequality. What happens to the norm $\|f\|_2$ if we replace the interval $[0,1]$ with the interval $[a,b]$?

- 5) Let $E = \mathbb{R}^2$ using the usual inner product (i.e., the dot product) and the usual Euclidean norm. Under what conditions on 2 vectors x,y in \mathbb{R}^2 do we have equality in the Cauchy-Schwarz inequality $|\langle x,y \rangle| \leq \|x\| \|y\|$? Prove your answer.

- 6) Which of the following are norms on the 1-dimensional vector space $E = \mathbb{R}$? Give reasons for your answers.

- | | | |
|---|---|------------------|
| a) $ x $ = ordinary absolute value of x in \mathbb{R} . | b) x^2 . | c) $ x /(1+ x)$ |
| d) $2 x $ | e) $\ x\ =1$ if $x \neq 0$ and $\ x\ =0$ if $x=0$ | f) x^3 . |

- 7) Which of the following give equivalent norms on $C[a,b]$? Why?

$$\|f\|_1 = \int_a^b |f(x)| dx, \quad \|f\|_2 = \sqrt{\int_a^b |f(x)|^2 dx}, \quad \|f\|_\infty = \max_{x \in [a,b]} |f(x)|.$$

- 8) Define for $x=(x_1,x_2)$ in \mathbb{R}^2 , the p -norm $\|x\|_p = (|x_1|^p + |x_2|^p)^{1/p}$ for any $p=1,2,3,4,\dots$. Prove that

$$\lim_{p \rightarrow \infty} \|x\|_p = \|x\|_\infty = \max\{|x_1|, |x_2|\}.$$

Exercises 2

1) Suppose that $v_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$ is a sequence of vectors in \mathbb{R}^2 . Show that if we define limits for sequences of vectors in \mathbb{R}^2 using the norm $\|v\|_\infty$, then

$$\lim_{n \rightarrow \infty} v_n = L = \begin{pmatrix} a \\ b \end{pmatrix} \text{ if and only if } \lim_{n \rightarrow \infty} x_n = a \text{ and } \lim_{n \rightarrow \infty} y_n = b.$$

2) Consider the sequence of functions on $[0,1]$ given by $f_n(x) = x^n$, for $0 \leq x \leq 1$.

a) Show that for each $x \in [0,1]$ we have $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ exists (defining $f(x)$ to be this limit). This is called a **pointwise limit**. Note that $f(x)$ is only piecewise continuous on $[0,1]$.

b) Show that the limit in part a) is not uniform: i.e., $\|f - f_n\|_\infty$ does not $\rightarrow 0$ as $n \rightarrow \infty$.

c) What happens to $\|f - f_n\|_1$ and $\|f - f_n\|_2$ as $n \rightarrow \infty$? Here both integrals are over $[0,1]$.

3) a) Show that the sequence of partial sums of the series $\sum_{n=0}^{\infty} x^n(1-x)$ converges pointwise (i.e., for each x in $[0,1]$) but not uniformly on $[0,1]$.

b) Show that the sequence of partial sums of the series $\sum_{n=0}^{\infty} (-1)^n x^n(1-x)$ converges uniformly on $[0,1]$.

4) Consider the sequence of functions defined by $f_n(x) = 1/(x+n)$, for $n=1,2,3,4, \dots$ with $x \in (0, \infty)$. We have 4 notions of convergence: pointwise for each fixed $x \in (0, \infty)$ and convergence with respect to the norms $\|\cdot\|_\infty, \|\cdot\|_1$, and $\|\cdot\|_2$. Here the integrals are computed on the interval $(0, \infty)$. Under which definition of convergence does f_n converge to the 0-function as $n \rightarrow \infty$?

5) Define $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ to be **linear** iff $L(\alpha x + \beta y) = \alpha Lx + \beta Ly$ for all x, y in \mathbb{R}^n and α, β in \mathbb{R} . Show that L is uniformly continuous on \mathbb{R}^n .

Exercises 3

1) Suppose that $\langle \cdot, \cdot \rangle$ is a scalar product on a vector space V and suppose that we have a subset S of V with a vector a adherent to S . Let f and g be functions mapping S into V such that the two following limits exist

$$\lim_{\substack{x \rightarrow a \\ x \in S}} f(x) = L \text{ and } \lim_{\substack{x \rightarrow a \\ x \in S}} g(x) = M. \text{ Prove that then } \lim_{\substack{x \rightarrow a \\ x \in S}} \langle f(x), g(x) \rangle = \langle L, M \rangle.$$

2) Define $C[a,b]$ to be the space of continuous real valued functions on the finite interval $[a,b]$. Define the

mapping $I:C[a,b] \rightarrow \mathbb{R}$ by $I(f) = \int_a^b f(x)dx$ for $f \in C[a,b]$.

a) Is I continuous when we use the $\| \cdot \|_\infty$ norm on $C[a,b]$ and ordinary absolute value as our norm on \mathbb{R} ?

Why?

b) What if we use the $\| \cdot \|_1$ norm on $C[a,b]$? Why?

3) Consider the following functions on \mathbb{R}^2

$$f(x,y) = \begin{cases} \frac{2x^2y}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases} \quad g(x,y) = \begin{cases} \frac{(y^2-x^2)^2}{x^4+y^4}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

a) Does $\lim_{x \rightarrow 0} \lim_{y \rightarrow 0} h(x,y) = \lim_{y \rightarrow 0} \lim_{x \rightarrow 0} h(x,y) = L$ when $h=f$ or $h=g$?

b) Do the repeated limits in part a) equal the limit as $(x,y) \rightarrow (0,0)$ using any of the favorite norms on \mathbb{R}^2 ?

That is, does $\lim_{(x,y) \rightarrow (0,0)} h(x,y) = L$?

c) Are either of the functions f,g continuous at the origin?

4) Let ℓ^2 denote the set of all sequences $a = \{a_n\}_{n \geq 1}$ of real numbers such that $\sum_{n=1}^{\infty} a_n^2$ converges.

a) Show that you can define addition and scalar multiplication to make ℓ^2 a vector space.

b) Define for $a = \{a_n\}_{n \geq 1}$ and $b = \{b_n\}_{n \geq 1}$ in ℓ^2 , $\langle a,b \rangle = \sum_{n=1}^{\infty} a_n b_n$. Show that this series converges and has the properties of a scalar product. You will need to use the Cauchy-Schwarz inequality on the partial sums.

Exercises 4

1) State whether the following series converge and give a reason for your answer.

a) $\sum_{n=0}^{\infty} \frac{1}{2^n}$ b) $\sum_{n=1}^{\infty} \frac{1}{n}$ c) $\sum_{n=1}^{\infty} \frac{1}{n^2}$ d) $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$.

2) Suppose that the series $\sum_{n=1}^{\infty} a_n$ converges absolutely. Show that $\sum_{n=1}^{\infty} a_n^2$ also converges absolutely.

3) Test for convergence and explain your answer.

a) $\sum_{n=0}^{\infty} n^3 e^{-n}$ b) $\sum_{n=2}^{\infty} \left(\frac{1}{\log n} \right)^{\log n}$.

4) a) Does the series $\sum_{n=1}^{\infty} \frac{n}{n+1} x^n$ converge pointwise for each x in $(0,1)$? Does this series converge uniformly on $(0,1)$?

b) Show that the series $\sum_{n=1}^{\infty} \frac{x^n}{3^{nx}}$ converges uniformly for $x \in [0,C]$.

Exercises 5

1) Find the radius of convergence of the following power series.

a) $\sum_{n=1}^{\infty} nx^n$

b) $\sum_{n=1}^{\infty} \frac{1}{n} x^n$

c) $\sum_{n=1}^{\infty} n^n x^n$

d) $\sum_{n=1}^{\infty} \frac{1}{2^n} x^n$.

2) What functions $f(x)$ are represented by the power series in parts a), b) and d) of problem 1. Compute the power series for the derivatives $f'(x)$ of these 3 functions. For what values of x is it legal to differentiate term by term?

3) Find the Taylor series at 0 for the function $\sqrt{1-x}$.

4) For an integer n , the Bessel function $J_n(x)$ is defined by the power series

$$J_n(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(n+k)!} \left(\frac{x}{2}\right)^{2k+n}.$$

Find the radius of convergence of this power series and then show that the Bessel function satisfies Bessel's equation

$$x^2 J_n''(x) + x J_n'(x) + (x^2 - n^2) J_n(x) = 0.$$

Exercises 6

For most of these problems, it should help to draw a picture.

1) Suppose V is a normed vector space with norm denoted by $\| \cdot \|$. For any subset A of V define the **closure** of A to be the set

$$\bar{A} = \left\{ x \in V \mid \forall \delta > 0, \exists a \in A \text{ s.t. } \|x - a\| < \delta \right\}.$$

Show that if $A \subset B$, then $\bar{A} \subset \bar{B}$.

2) Using the definition of closure of a set in problem 1, show that

$$\overline{\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid x < y \right\}} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid x \leq y \right\}.$$

- 3) Let $f(x)=x^2$, for all $x \in [0,1]$. Find a step function $s(x)$ on the interval $[0,1]$ such that $\|f-s\|_\infty < 1/10$.
- 4) Define the function $f(x) = x \sin(1/x)$, when $x \neq 0$, and $f(0) = 0$. Define $g(x) = 1$ if $x > 0$, $g(x) = -1$ if $x < 0$, and $g(0) = 0$. Show that $h(x) = g(f(x))$ is not the (uniform) limit of a sequence of step functions on the interval $[0,1]$, using the norm $\| \cdot \|_\infty$.
- 5) Show that for 2 equivalent norms $\| \cdot \|_\alpha$ and $\| \cdot \|_\beta$ on a vector space V that the open sets, closed sets and continuous functions are the same no matter which of the 2 norms you use.

Exercises 7

- 1) a) **A Mean Value Theorem.** Suppose that $f, g: [a,b] \rightarrow \mathbb{R}$ are continuous on $[a,b]$ and $g(x) \geq 0$ for all $x \in [a,b]$. Show that there exists $c \in [a,b]$ such that

$$\int_a^b fg = f(c) \int_a^b g.$$

- b) Show that the result of problem 1 need not be true if $g(x)$ can assume both positive and negative values on $[a,b]$.

- 2) Define $\int_b^a f = -\int_a^b f$. Show that $\int_a^b f = \int_a^c f + \int_c^b f$ no matter what the ordering of the points a, b, c is.

There are 6 possible orderings starting with $a < c < b$, $a < b < c$, etc.

- 3) Let $g: [a,b] \rightarrow \mathbb{R}$ be continuous and $g(x) \geq 0$ for all $x \in [a,b]$. Show that if $\int_a^b g = 0$ then $g(x) = 0 \quad \forall x \in [a,b]$.

- 4) Let $f: [0,1] \rightarrow \mathbb{R}$ be defined by $f(x) = 1$ for all x in $[0,1]$. Define F by

$$F(x) = \begin{cases} 1, & x = 0 \\ x, & 0 < x < 1 \\ 0, & x = 1. \end{cases}$$

Clearly $F'(x) = f(x)$ for all $x \in (0,1)$. However $F(1) - F(0) \neq \int_0^1 f$.

What went wrong with the fundamental theorem of calculus here?

Exercises 8

- 1) a) Suppose that $f(x)=1$, for $0 \leq x \leq 1$ and $f(x)=0$, otherwise. Compute $(f * f)(x)$. Draw the graphs of f and $f * f$.
- b) Suppose that $f(x)$ is as in part a). Define $g(x) = 1-x^2$, for all x . Compute $f * g$. Draw the graphs of g and $f * g$.
- 2) Suppose that α is a real number and f, g, h are piecewise continuous functions such that h vanishes outside the interval $[-c, c]$. Show that
- $(f+g) * h = f * h + g * h$
 - $(\alpha g) * h = \alpha(g * h)$.
 - Suppose that f, g, h are piecewise continuous functions that vanish outside an interval $[-c, c]$. Show that $(f * g) * h = f * (g * h)$.

- 3) (Delta is not a function). Suppose that there is a piecewise continuous function δ such that

$$\int_{-\infty}^{\infty} f(t)\delta(t)dt = f(0) \text{ for all continuous functions } f \text{ which vanish outside an interval } [-c, c] \text{ for some } c.$$

Show that this implies $\delta(x)=0$, for all $x \neq 0$ and thus that $\int_{-\infty}^{\infty} f(t)\delta(t)dt = 0$. A similar result could be

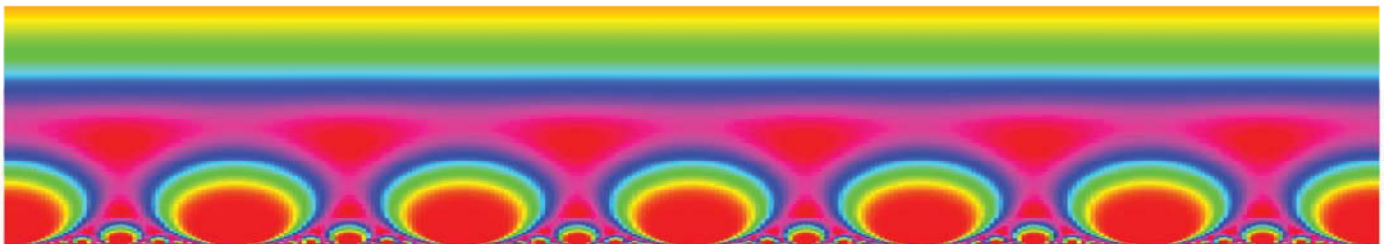
proved assuming only that δ is a Lebesgue integrable function on finite intervals. (In fact, δ is not an ordinary function. It is a generalized function or distribution.)

- 4) Define the Gaussian by $G_t(x) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}$. Show that $G_{1/n}$, for $n=1,2,3,\dots$ is a Dirac sequence.

Exercises 9

- 1) Compute the Fourier series of the function $f(x)=x^2$ for x in $[0,1]$ and extended to all real numbers to be periodic of period 1. What does the Parseval identity say for this function?

- 2) Suppose that f is a continuous function on $[0,1]$. Show that if $\int_0^1 f(x)x^n dx = 0$ for all $n=0,1,2,3, \dots$, then f must be identically 0. You will need to remember the Weierstrass theorem on uniform approximation here.



II, PRACTICE EXAM 1

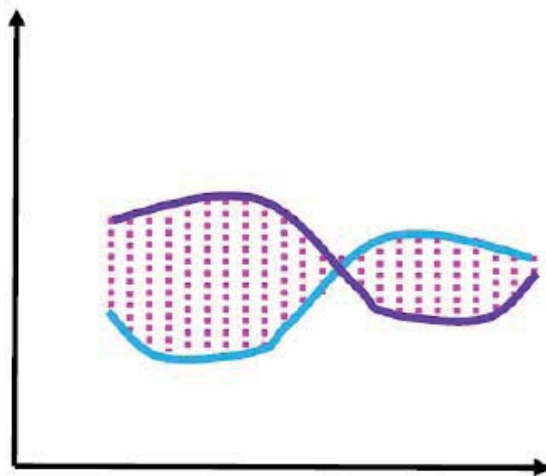
- 1) Define and give an example:
- a) norm; b) scalar product; c) equivalent norms;
 - d) limit of a sequence $\{v_n\}$ in a normed vector space; e) Cauchy sequence in a normed vector space;
 - f) complete normed vector space; g) closure of a set in a normed vector space;
 - h) $\lim_{x \rightarrow c} f(x) = L$ for function $f: D \rightarrow W$, with $V \supset \{x \mid 0 < \|x-c\| < \delta\}$ for some $\delta > 0$.
Here V, W are normed vector spaces.
 - i) continuous function f mapping a set in normed vector space V to another normed vector space;
 - j) uniform convergence of a sequence $\{f_n\}$ of functions $f_n \in C[a,b]$, the space of continuous real-valued functions on an interval $[a,b]$
 - k) $\|f\|_\infty, \|f\|_1, \|f\|_2$ for $f \in C[a,b] = \{\text{continuous real valued functions on the interval } [a,b]\}$
 - l) the cosine of the angle between 2 non-0 vectors v, w in a vector space V with a scalar product
- 2) **True - False.** Tell whether the following statements are true or false. Give a brief reason for your answer.
- a) $C[a,b]$ the space of continuous real-valued functions on an interval $[a,b]$, is a finite dimensional vector space.
 - b) For $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2$, $\|v\| = v_1^2 + v_2^2$ defines a norm on vectors in the plane.
 - c) The function $f(x,y) = \frac{x^2}{x^2 + y^2}$, for $(x,y) \neq (0,0)$ and $f(0,0) = 0$ is continuous on the plane \mathbb{R}^2 .
 - d) Suppose that f_n is a sequence of continuous functions on the interval $[a,b]$. Suppose for every x in $[a,b]$ we have $\lim_{n \rightarrow \infty} f_n(x) = f(x)$. Then $f(x)$ is continuous on $[a,b]$.
 - e) \mathbb{R}^2 is a complete normed vector space.
- 3) State and prove the Cauchy-Schwarz inequality. What norm is being used in this inequality? Would it still be true if we replaced that norm with some other one?
- 4) Suppose that $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are equivalent norms on a vector space V . Show that if $\{v_n\}$ is a sequence of vectors in V and $L \in V$, we have
- $$\lim_{n \rightarrow \infty} \|v_n - L\|_\alpha \Leftrightarrow \lim_{n \rightarrow \infty} \|v_n - L\|_\beta.$$
- 5) Suppose that $\langle v, w \rangle$ denotes a scalar product of 2 vectors v, w in the vector space V . Show that $\langle v, w \rangle$ is a continuous function of v , holding w fixed. Is it uniformly continuous?
- 6) Suppose that $L: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a linear map. Show that it is continuous.

7) Show that $f:U \rightarrow W$, where U is a subset of a normed vector space V and W is a normed vector space, is continuous at a point a in U if and only if for every sequence v_n in U such that $\lim_{n \rightarrow \infty} v_n = a$ we have $\lim_{n \rightarrow \infty} f(v_n) = f(a)$.

8) The function I mapping $C[a,b]$ =the space of continuous real valued functions on the interval $[a,b]$ into \mathbb{R} defined by $I(f) = \int_a^b f$ is continuous with respect to the $\| \cdot \|_\infty$ norm on $C[a,b]$ and the usual absolute value on \mathbb{R} .

9) Show that the norms $\| \cdot \|_\infty$ and $\| \cdot \|_1$ on the space $C[a,b]$ =the space of continuous real valued functions on the interval $[a,b]$ into \mathbb{R} are not equivalent.

10) Explain how the picture below can be used to see the difference between $\|f-g\|_\infty$ and $\|f-g\|_1$ assuming that f is the purple function which starts at the top and g is the blue function which starts at the bottom.



II, PRACTICE EXAM 2

1) Define and give an example:

- a) radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$ b) adherent point to a set A in a normed vector space V
 c) continuous linear map $L:V \rightarrow W$, where V, W are vector spaces d) step function on the interval $[a, b]$
 e) Riemann integral of a step map f) partition of an interval $[a, b]$
 g) convergence of a series $\sum v_n$ in a normed vector space

2) a) State and prove the integral test.
 b) State and prove the comparison test.

3) Which of the series below converge and why?

- a) $\sum_{n=2}^{\infty} \frac{1}{n(n-1)}$ b) $\sum_{n=2}^{\infty} \frac{(-1)^n}{\log(n)}$ c) $\sum_{n=2}^{\infty} \frac{1}{\sqrt{n(n-1)}}$
 d) $\sum_{n=1}^{\infty} \frac{\sin(n)}{2n^2 - n}$ e) $\sum_{n=1}^{\infty} \frac{n}{2^n}$

4) **True - False.** Tell whether the following statements are true or false. Give a brief reason for your answer.

- a) $\sum_{n=0}^{\infty} a_n x^n$ converges implies $\sum_{n=0}^{\infty} a_n u^n$ converges for all u such that $|u| < |x|$.
 b) $\sum_{n=1}^{\infty} \frac{1}{n} (x-2)^n$ converges if $|x| < 1$.
 c) f_n continuous on $[a, b]$ and $\lim_{n \rightarrow \infty} f_n(x) = f(x) \forall x \in [a, b] \Rightarrow \lim_{n \rightarrow \infty} \int_a^b f_n = \int_a^b f$.

d) Define $f(x)=0$ for x irrational and $f(x) = 1$ for x rational. Then $f \in \overline{St[a, b]}$ = the space of regulated functions on $[a, b]$.

e) Any bounded function on $[a, b]$ can be uniformly approximated by step functions and is thus integrable.

f) Any piecewise continuous function on $[a, b]$ can be uniformly approximated by step functions and is thus integrable.

g) If $\lim_{n \rightarrow \infty} a_n = 0$, then $\sum_{n=1}^{\infty} a_n$ is convergent.

5) a) State and prove a formula for the radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$.

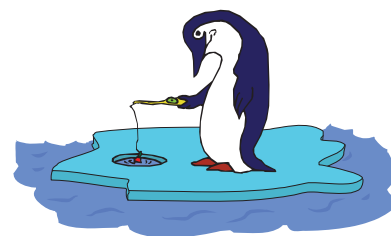
b) Apply your formula to find the radius of convergence of $\sum_{n=1}^{\infty} \frac{1}{n} x^n$.

c) What function does this power series represent within the interval where it converges?

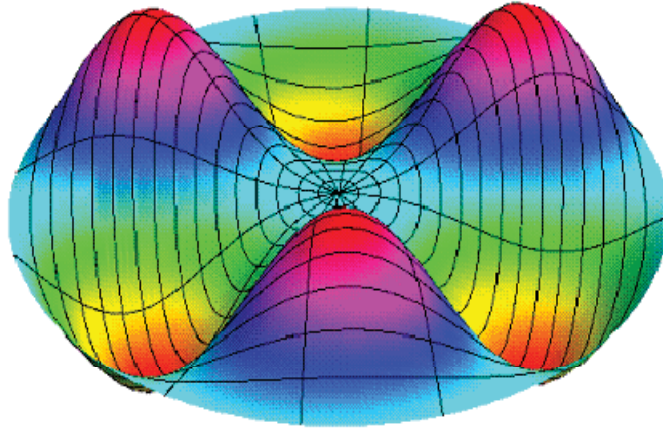
- 6) $a \in \text{the closure of } A = \overline{A} \Leftrightarrow \exists \text{ a sequence } a_n \in A \text{ s.t. } a = \lim_{n \rightarrow \infty} a_n.$
- 7) a) Define $\|g\|_\infty = \text{l.u.b. } \{ |g(x)| \mid x \in [0,1] \}$, where l.u.b.=least upper bound. Consider the function $f(x) = x^3$. Find a step function s on $[0,1]$ such that $\|f-s\|_\infty \leq .25$.
- b) Compute $\int_0^1 s$.
- 8) a) State and prove the continuous linear extension theorem.
b) How did we use the continuous linear extension theorem to extend the integral from step functions to continuous functions?
- 9) Suppose that f is a step function on $[a,b]$ with respect to partition \mathcal{P} and g is a step function on $[a,b]$ with respect to partition \mathcal{Q} . Show that if \mathcal{R} is a refinement of \mathcal{P} and \mathcal{Q} , then both f and g are step functions with respect to \mathcal{R} and $I(f+g, \mathcal{R}) = I(f, \mathcal{R}) + I(g, \mathcal{R})$.
- 10) Show that the map which sends step function f to $\int_a^b f$ is continuous with respect to the $\|\cdot\|_\infty$ norm on the vector space $\text{St}[a,b]$ of step functions on $[a,b]$.
- 11) Show that any continuous function on a finite **closed** interval is uniformly continuous on that interval.
- 12) a) Define the integral $\int_a^b f$ for any function f which is a uniform limit of a sequence s_n of step functions on $[a,b]$.
b) Explain why you know that this integral is a continuous linear function of f , where continuity is with respect to the $\|\cdot\|_\infty$ norm on the space of bounded functions on $[a,b]$.
- 13) Explain why the integral has the following 2 properties on the space $C[a,b]$ of continuous functions on the finite interval $[a,b]$.
- a) the integral preserves \leq
b) for any c with $a < c < b$ we have $\int_a^b f = \int_a^c f + \int_c^b f$.
- 14) Fundamental Theorem of Calculus. Suppose $f: [a,b] \rightarrow \mathbb{R}$ is continuous. Show that then

$$F(x) - F(a) = \int_a^x f \text{ for every } x \in [a,b] \Leftrightarrow F'(x) = f(x), \forall x \in [a,b].$$

- 15) Let $a < b < c$. Suppose that $f(x)$ is continuous on $[a,b]$ and we define $g(x)$ to be equal to $f(x)$ for all x in $[a,c)$ and for all x in $(c,b]$ but we define $g(c) = f(c) + 100$. Show that $\int_a^b f = \int_a^b g$.



II, Discussion of Final



Vibrating Things

Resonance is what happens when something is forced to vibrate at certain frequencies - - unbounded oscillations.

In "real" life this would cause the spring to self destruct. Such things can happen in bridges or other structures.

Most texts say the Tacoma Narrows bridge disaster was an example of such resonance. However, the latest research questions this conclusion.

The original Tacoma Narrows Bridge opened to traffic on July 1, 1940. It collapsed just four months later during a 42-mile-per-hour wind storm on Nov. 7.

M. Braun, Differential Equations and their Applications:
 "When the bridge began heaving violently, the authorities notified Professor F. B. Farquharson of the University of Washington. Professor Farquharson had conducted numerous tests on a simulated model of the bridge and had assured everyone of its stability. The professor was the last man on the bridge. Even when the span was tilting more than twenty-eight feet up and down, he was making scientific observations with little or no anticipation of the imminent collapse of the bridge. "

"A large sign near the bridge approach advertised a local bank with the slogan 'as safe as the Tacoma Bridge.'"

"After the collapse of the Tacoma Bridge, the governor of the state of Washington made an emotional speech in which he declared 'We are going to build the exact same bridge, exactly as before.' Upon hearing this, the noted engineer Von Karman sent a telegram to the governor stating 'If you build the exact same bridge exactly as before, it will fall into the exact same river exactly as before.'"

Non Forced Vibration of a String.

Let's look at the wave equation which describes the motion of a **vibrating string**. Assume the string has constant density ρ and constant tension τ . Then one can derive the following PDE known as the **wave equation**, using Newton's law of motion of the principle of least action:

$$(1) \quad \frac{\partial^2 u}{\partial t^2} = \frac{\tau}{\rho} \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < \pi, \quad 0 < t.$$

One assumes, for example, that the string is tied down at the boundary points giving the **boundary conditions**:

$$(2) \quad u(0, t) = u(\pi, t) = 0, \quad \text{for all } t > 0.$$

And one may assume **initial conditions**

$$(3) \quad u(x, 0) = f(x), \quad \frac{\partial u}{\partial t}(x, 0) = 0 \quad \text{for } 0 < x < \pi.$$

The **method of separation of variables of Daniel Bernoulli** says: look for a solution of the PDE in formula (1) of the form $u(x,t)=X(x)T(t)$.

If you want this to satisfy (2), assume $X(0)=X(\pi)$. If you want it to satisfy (3), you are in trouble for the 1st part, but the 2nd part becomes $T'(0)=0$.

Now plug $u(x,t)=X(x)T(t)$ into formula (1) $\frac{\partial^2 u}{\partial t^2} = \frac{\tau}{\rho} \frac{\partial^2 u}{\partial x^2}$
You get (setting $c=\tau/\rho$)

$$X(x)T''(t) = cT(t)X''(x).$$

Divide both sides by $X(x)T(t)$ (hoping you are not dividing by 0). This gives:

$$(4) \quad \frac{T''(t)}{T(t)} = c \frac{X''(x)}{X(x)}$$

This implies each side is constant. Call the constant λ . It is often called the **separation constant**. It is an eigenvalue in the 1st ODE below.

Exercise 1. Prove that each side in equation (4) must be constant.

Thus we now have 2 ODEs to solve. **Here assume $c=1$ in (4).**

$$\text{ODE 1. } X''(x) = \lambda X(x), \quad 0 = X(0) = X(\pi).$$

$$\text{ODE 2. } T''(t) = \lambda T(t), \quad T'(0) = 0.$$

Look at ODE1. The general solution from Math. 20D is

$$X(x) = a_1 \exp(x\sqrt{\lambda}) + a_2 \exp(-x\sqrt{\lambda})$$

with constants a_i . To satisfy the boundary conditions, we need $\lambda < 0$. This means, since $e^{ix} = \cos x + i \sin x$ when $i = (-1)^{\frac{1}{2}}$, that we should write, for $\lambda = -\mu^2$,

$$X(x) = c_1 \cos(\mu x) + c_2 \sin(\mu x)$$

with constants c_j .

In order to satisfy the 1st boundary condition, we need $X(0) = c_1 = 0$. This makes $X(x) = c_2 \sin(\mu x)$. The 2nd boundary condition is $X(\pi) = c_2 \sin(\mu\pi) = 0$. This says

$$\mu\pi = n\pi, \quad \text{for some integer } n=1,2,3,\dots$$

So we find the separation constant in equation (4) is

$$\lambda = -\mu^2 = -n^2, \quad \text{for some integer } n=1,2,3,\dots$$

Then

$$(5) \quad X(x) = c_2 \sin(nx), \quad \text{for } n=1,2,3,\dots$$

Look at ODE2.

Math 20D says that the general solution may be taken to be

$$T(t) = b_1 \cos(nt) + b_2 \sin(nt)$$

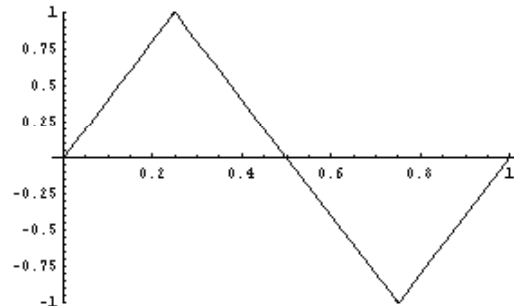
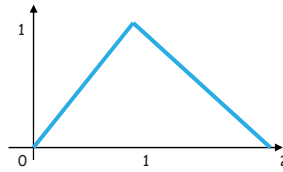
Since $T'(0) = nb_2 = 0$, we see that

$$(6) \quad T(t) = b_1 \cos(nt), \quad \text{for } n=1,2,3,\dots$$

Now turn to the problem of the 1st part of the initial condition (3). For this you need to be able to write the function

$$f(x) = X(x)T(0) = c_2 \sin(nx).$$

But, what should we do if the initial shape of the string is not a sine function; e.g., the plucked string pictured below.



To solve the plucked string problem you need to represent the function $f(x)=u(x,0)$ as a **Fourier sine series**:

$$(7) \quad f(x) = \sum_{n \geq 1} c_n \sin(nx)$$

Then the constants c_n are given by the formula

$$(8) \quad c_n = \int_0^{\pi} f(y) \sin(ny) dy.$$

This is proved in Lang, p. 318, for sufficiently smooth functions f . Of course, our plucked string function does not look smooth, just continuous. It has that sharp point, remember.

Anyway our **final solution to the vibrating string problem** is:

$$(9) \quad u(x,t) = \sum_{n \geq 1} c_n \sin(nx) \cos(nt).$$

Here the constants c_n are from formula (8) and we assume $\tau=\rho$ in the PDE (1). **Exercise 2. Check (9) solves the PDE.**

Forced Motion of a Vibrating String

Assume the vibrating string is as before but now apply an external force of the form $f(x)\cos(\omega t)$. This leads to the PDE:

$$u_{tt} - \frac{\tau}{\rho} u_{xx} = f(x) \cos(\omega t)$$

$$u(0,t) = u(\pi,t) = 0; \quad u(x,0) = u_t(x,0) = 0.$$

Assume $1=\tau/\rho$, for simplicity. So our problem is now

$$(10) \quad \begin{aligned} u_{tt} - u_{xx} &= f(x) \cos(\omega t) \\ u(0,t) = u(\pi,t) &= 0; \quad u(x,0) = u_t(x,0) = 0. \end{aligned}$$

To solve (10), plug in

$$u(x,t) = \sum_{n \geq 1} c_n(t) \sin(nx).$$

Define the inner product for piecewise continuous functions f , g on $[0, \pi]$ by

$$(12) \quad \langle f, g \rangle = \int_0^{\pi} f(x)g(x)dx.$$

Then

$$c_n(t) = \left\langle u(x,t), \frac{\sin(nx)}{\sqrt{\pi/2}} \right\rangle = \int_0^{\pi} u(x,t) \frac{\sin(nx)}{\sqrt{\pi/2}} dx$$

Here we use the fact that

$$\int_0^{\pi} \sin^2(nx) dx = \frac{\pi}{2}.$$

Exercise 3. Prove the last formula and then use it to show that

$$v_n(x) = \frac{\sin(nx)}{\sqrt{\pi/2}}, \quad n=1,2,3,\dots \text{ forms an orthonormal family for}$$

the inner product space of piecewise continuous functions on $[0, \pi]$ using the inner product defined by formula (12).

If we plug formula (11) into (10) without worrying about our issues of interchange of derivative and summation, we see that we need

$$\begin{aligned} c_n''(t) &= \int_0^{\pi} u_n(c,t)v_n(x)dx = \int_0^{\pi} (u_{xx}(x,t) + f(x)\cos(\omega t))v_n(x)dx \\ &= \int_0^{\pi} u(x,t)v_n''(x)dx + \cos(\omega t) \int_0^{\pi} f(x)v_n(x)dx. \end{aligned}$$

Exercise 4. Show this last formula holds using integration by parts and the boundary conditions.

So from our last formula, we have

$$\begin{aligned} c_n''(t) &= -n^2 \int_0^{\pi} u(x,t)v_n(x)dx + \cos(\omega t) \int_0^{\pi} f(x)v_n(x)dx \\ &= -n^2 c_n(t) + \langle f, v_n \rangle \cos(\omega t). \end{aligned}$$

So now we need to solve an ODE :

$$c_n''(t) = -n^2 c_n(t) + \langle f, v_n \rangle \cos(\omega t), \quad c_n(0) = c_n'(0) = 0$$

You can solve this using methods from Math. 20D - for example, variation of parameters. The result is:

$$c_n(t) = \frac{\langle f, v_n \rangle}{n^2 - \omega^2} (\cos(\omega t) - \cos(nt)).$$

Exercise 5. Check this answer.

Thus a solution for the problem posed by formula (10) is

$$(13) \quad u(x,t) = \sum_{n=1}^{\infty} a_n \frac{\cos(\omega t) - \cos(nt)}{n^2 - \omega^2} v_n(x), \quad a_n = \langle f, v_n \rangle, \quad v_n(x) = \frac{\sin(nx)}{\sqrt{\pi/2}}.$$

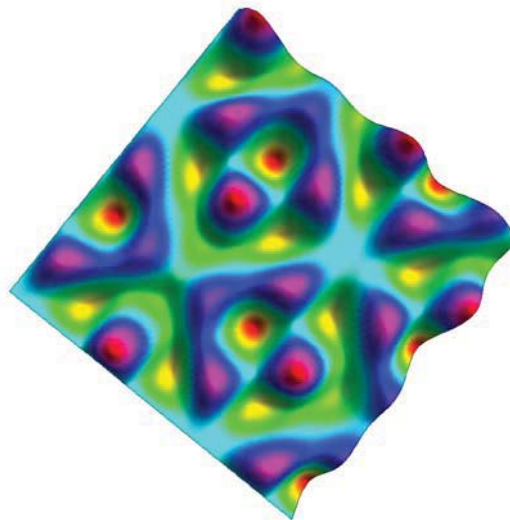
Exercise 6. What happens to formula (13) as $\omega \rightarrow \pm n$???
You need to compute

$$\lim_{\omega \rightarrow n} \frac{\cos(\omega t) - \cos(nt)}{\omega^2 - n^2}.$$

There is a similar result as $\omega \rightarrow -n$. You should be able to deduce from this that unless $a_n = 0$, the function $u(x,t)$ blows up as t goes to infinity. This is resonance.

Mathematica plots a vibrating square drum:

```
Animate[Plot3D[
(Sin[5*Pi*x]*Sin[3*Pi*y]-Sqrt[2/3]*Sin[3*Pi*x]*Sin[5*Pi*y])*Cos[Sqrt[2]*Pi*t/10000],
{x,0,3/2},{y,0,3/2},PlotPoints->80,ColorFunction->Hue,Mesh->False,FaceGrids->None,Axes-
->False,Boxed->False],{t,0,10000}]
```





II, Final

The test concerns Fourier series and an application to resonance.

Part I. Vibrating Bridges & Resonance.

Read the part of the lectures telling the story of Fourier series. It might also help to remember some things from the ODEs part of calculus. For example, in problem 18 on page 215 of Boyce and DePrima the ODE represents the forced motion of a vibrating spring: $u''+u = 3\cos(ft.)$. $u(0)=0=u'(0)$. You first find the solution assuming f is not 1 or -1. Then use l'Hopital's rule to see what happens if you take the limit as f approaches 1. **You should see the spring blow up!**

This phenomenon is called "**resonance.**" That is what happens when something is forced to vibrate at certain frequencies. One sees unbounded oscillations. In "real" life this would cause the spring to self destruct. Such things can happen in bridges or other structures. Most texts say the Tacoma Narrows bridge disaster was an example of such resonance. However, the latest research questions this conclusion. The Millenium pedestrian bridge in London showed such resonance and was closed for a time and fixed up to keep it from resonating with the many pedestrians walking across. Soldiers are required not to walk in step across bridges for similar reasons.



The original Tacoma Narrows Bridge opened to traffic on July 1, 1940. It collapsed just four months later during a 42-mile-per-hour wind storm on Nov. 7. There are many web sites devoted to the subject.

See a movie of the collapse

<http://www.archive.org/details/SF121>

or

<http://www.ketchum.org/bridgecollapse.html>

Or see Judith Packer's website at the U. of Colorado

<http://spot.colorado.edu/~packer/Fourier09.html>

The **London Millennium Footbridge** is a pedestrian-only steel suspension bridge crossing the River Thames in London, England. The bridge opened on June 10, 2000 but unexpected resonant lateral vibration caused the bridge to be closed on June 12 for modifications. The movements were produced by the sheer numbers of pedestrians (90,000 users in the first day, with up to 2,000 on the bridge at any one time).



M. Braun, *Differential Equations and their Applications*,

notes: "There were many humorous and ironic incidents associated with the collapse of the Tacoma Bridge. When the bridge began heaving violently, the authorities notified Professor F. B. Farquharson of the University of Washington. Professor Farquharson had conducted numerous tests on a simulated model of the bridge and had assured everyone of its stability. The professor was the last man on the bridge. Even when the span was tilting more than twenty-eight feet up and down, he was making scientific observations with little or no anticipation of the imminent collapse of the bridge. "

"A large sign near the bridge approach advertised a local bank with the slogan 'as safe as the Tacoma Bridge.'"

"After the collapse of the Tacoma Bridge, the governor of the state of Washington made an emotional speech in which he declared 'We are going to build the exact same bridge, exactly as before.'"

Upon hearing this, the noted engineer Von Karman sent a telegram to the governor stating 'If you build the exact same bridge exactly as before, it will fall into the exact same river exactly as before.'"

Another famous bridge failure due to resonance: Angers Bridge, in Angers, France, 16 April 1850. The collapse was due to resonance from marching soldiers.

Part II. Short course in series expansions and PDEs like the wave equation.

Applied Math. involves solving certain partial differential equations, for example, the wave equation, heat equation, Schrödinger's equation. One way to find solutions for such PDEs is the method of separation of variables. Usually this leads to Fourier series or generalized Fourier series. Let's look at the wave equation which describes the **motion of a vibrating string**. Assume the string has constant density ρ and constant tension τ . Then one can derive the following PDE known as the **wave equation**, using Newton's law of motion or the principle of least action:

$$(1) \quad \frac{\partial^2 u}{\partial t^2} = \frac{\tau}{\rho} \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < \pi, \quad 0 < t.$$

One assumes, for example, that the string is tied down at the boundary points giving the boundary conditions:

$$(2) \quad u(0,t) = u(\pi,t) = 0, \quad \text{for all } t > 0.$$

And one may assume initial conditions

$$(3) \quad u(x,0) = f(x), \quad \text{for} \quad \frac{\partial u}{\partial t}(x,0) = 0 \quad 0 < x < \pi.$$

The method of **separation of variables** of Daniel Bernoulli says: look for a solution of the PDE in formula (1) of the form $u(x,t) = X(x)T(t)$. If you want this to satisfy (2), assume $X(0) = X(\pi) = 0$. If you want it to satisfy (3), you are in trouble for the 1st part, but the 2nd part becomes $T'(0) = 0$.

Now plug $u(x,t) = X(x)T(t)$ into formula (1). You get (setting $c = \tau/\rho$)

$$X(x)T''(t) = cT(t)X''(x),$$

Divide both sides by $X(x)T(t)$ (hoping you are not dividing by 0). This gives:

$$(4) \quad \frac{T''(t)}{T(t)} = c \frac{X''(x)}{X(x)}.$$

This implies each side is constant. Call the constant λ . It is often called the separation constant. It is an eigenvalue in the 1st ODE below.

Exercise 1. Prove that each side in equation (4) must be constant.

Thus we now have 2 ODEs to solve, **assuming $c=1$ (or $\tau=\rho$ in formula (1))**.

ODE 1. $X''(x) = \lambda X(x), \quad 0 = X(0) = X(\pi).$

ODE 2. $T''(t) = \lambda T(t), \quad T'(0) = 0.$

Look at ODE1. The general solution from Math. 20D is

$$X(x) = a_1 \exp(x\sqrt{\lambda}) + a_2 \exp(-x\sqrt{\lambda}),$$

with constants a_i . To satisfy the boundary conditions, we need $\lambda < 0$. This means, since $e^{ix} = \cos x + i \sin x$ when $i = (-1)^{\frac{1}{2}}$, that we should write, for $\lambda = -\mu^2$,

$$X(x) = c_1 \cos(\mu x) + c_2 \sin(\mu x),$$

with constants c_j .

In order to satisfy the 1st boundary condition, we need $X(0) = c_1 = 0$. This makes $X(x) = c_2 \sin(\mu x)$. The 2nd boundary condition is $X(\pi) = c_2 \sin(\mu\pi) = 0$. This says

$$\mu\pi = n\pi, \quad \text{for some integer } n = 1, 2, 3, \dots$$

So we find the separation constant in equation (4) is

$$\lambda = -\mu^2 = -n^2, \text{ for some integer } n=1,2,3,\dots$$

Then

$$(5) \quad X(x) = c_2 \sin(nx), \text{ for } n=1,2,3,\dots$$

Look at ODE2.

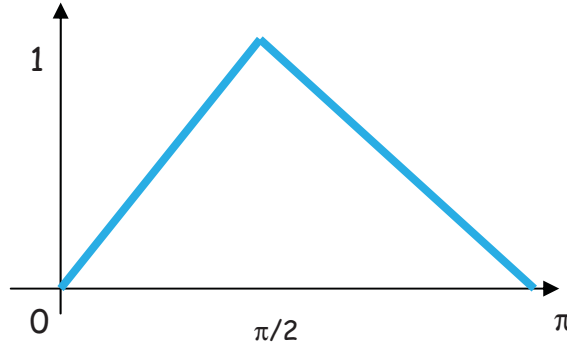
Math 20D says that the general solution may be taken to be

$$T(t) = b_1 \cos(nt) + b_2 \sin(nt).$$

Since $T'(0) = nb_2 = 0$, we see that $b_2 = 0$ and

$$(6) \quad T(t) = b_1 \cos(nt), \text{ for } n=1,2,3,\dots$$

Now we turn to the problem of the 1st part of the initial condition (3). For this you need to be able to write the function $f(x) = X(x)T(0) = c_2 \sin(nx)$. But, what should we do if the initial shape of the string is not a sine function; e.g., the plucked string pictured below.



To solve the plucked string problem you need to represent the function $f(x) = u(x,0)$ as a Fourier sine series (**thinking that f is an odd function of period 2π**):

$$(7) \quad f(x) = \sum_{n \geq 1} c_n \sin(nx).$$

Then the constants c_n are given by the formula

$$(8) \quad c_n = \frac{2}{\pi} \int_0^{\pi} f(y) \sin(ny) dy.$$

This was proved in Lectures II, for sufficiently smooth functions f . Of course, our plucked string function does not look smooth, just continuous. It has that sharp point, remember.

Anyway our final solution to the vibrating string problem is:

$$(9) \quad u(x,t) = \sum_{n \geq 1} c_n \sin(nx) \cos(nt).$$

Here the constants c_n are from formula (8).

Exercise 2. Check that formula (9) solves the PDE (1) assuming $c=1$ (or $\tau=\rho$ in formula (1)). What assumptions do you need to make on the function f to know that the Fourier coefficients decrease rapidly enough to make the differentiated series converge uniformly so that it is legal to differentiate term-by-term?

Thus we see that Fourier series seem necessary if we want to understand vibrating things like strings. They are also useful in the analysis of almost any phenomenon; e.g., the stock market, heat diffusion, yearly San Diego rainfall measurements. Of course, many questions are raised in the mathematical brain here. We put some of them in the preceding exercise. You might also ask:

- 1) Are the solutions $u(x,t)$ to (1),(2), and (3) unique? Then whatever crazy method we use will lead to the same answer.
- 2) When is an infinite series of solutions to a PDE again a solution?

We won't answer these questions here.

The quest for solutions to more general boundary value problems such as that of a vibrating drum leads to Fourier series in 2 variables or series involving Bessel functions, depending on the shape of the drum (square, circular, ...). See Courant and Hilbert, *Methods of Mathematical Physics*, I, p. 302.

Sometimes the problem may lead to Fourier integrals; e.g., heat equation on an infinite metal wire placed along the positive real axis. We have no time to think about this. The references below will.

Other References on Fourier Analysis.

Dym and McKean, *Fourier Series and Integrals*;

Terras, *Harmonic Analysis on Symmetric Spaces and Applications*.

Fourier's paper "Théorie analytique de la chaleur" [Analytic theory of heat] appeared in 1822. He says in it that "there is no function $f(x)$ or part of a function which cannot be expressed by a trigonometric series." In this generality the statement is false. Dirichlet, Riemann, Lebesgue and many others have slaved to make Fourier theory rigorous. Now it helps to use the theory of distributions or generalized functions to keep us from worrying too much about interchanges of derivative and sum.

In 1876 DuBois Reymond gave an example of a continuous function whose Fourier series diverges on an everywhere dense set of points. Kolmogorov (1926) gave an example of a Lebesgue integrable function on $[0,1]$ whose Fourier series diverges everywhere. In 1966 Carleson show that a square integrable function on $[0,1]$ has a Fourier series that converges almost everywhere (i.e., except on a set of Lebesgue measure 0). Charles Fefferman (who at age 12 was in the class analogous to Math 142 that I took as an undergrad) has shown that Carleson's result does not generalize to more than 1 variable. Fefferman was awarded the Fields medal in 1978 for this.

Part III. Forced Motion of a Vibrating String.

Assume the vibrating string is as before but now apply an external force of the form $f(x)\cos(\omega t)$. This leads to the PDE:

$$u_{tt} - \frac{\tau}{\rho} u_{xx} = f(x) \cos(\omega t)$$

$$u(0,t) = u(\pi,t) = 0$$

$$u(x,0) = u_t(x,0) = 0.$$

Assume $1 = \tau/\rho$, for simplicity. So our problem is now

$$(10) \quad u_{tt} - u_{xx} = f(x) \cos(\omega t)$$

$$u(0,t) = u(\pi,t) = 0$$

$$u(x,0) = u_t(x,0) = 0.$$

To solve (10), plug in

$$(11) \quad u(x,t) = \sum_{n \geq 1} c_n(t) \sin(nx).$$

Define the inner product for piecewise continuous functions f, g on $[0,\pi]$ by

$$(12) \quad \langle f, g \rangle = \int_0^\pi f(x)g(x)dx.$$

Then $c_n(t) = \left\langle u(*, t), \frac{\sin(n*)}{\sqrt{\pi/2}} \right\rangle = \int_0^\pi u(x, t) \frac{\sin(nx)}{\sqrt{\pi/2}} dx.$

Here we use the fact that $\int_0^\pi \sin^2(nx) dx = \frac{\pi}{2}.$

Exercise 3. Prove the last formula and then use it to show that $v_n(x) = \frac{\sin(nx)}{\sqrt{\pi/2}}, n=1,2,3,\dots$ forms

an orthonormal family for the inner product space of piecewise continuous functions on $[0, \pi]$ using the inner product defined by formula (12).

If we plug formula (11) into (10) without worrying about our issues of interchange of derivative and summation, we see that we need

$$\begin{aligned} c_n''(t) &= \int_0^\pi u_{tt}(x, t) v_n(x) dx = \int_0^\pi (u_{xx}(x, t) + f(x) \cos(\omega t)) v_n(x) dx \\ &= \int_0^\pi u(x, t) v_n''(x) dx + \cos(\omega t) \int_0^\pi f(x) v_n(x) dx. \end{aligned}$$

Exercise 4. Show this last formula holds using integration by parts and the boundary conditions.

So from our last formula, we have

$$\begin{aligned} c_n''(t) &= -n^2 \int_0^\pi u(x, t) v_n(x) dx + \cos(\omega t) \int_0^\pi f(x) v_n(x) dx \\ &= -n^2 c_n(t) + \langle f, v_n \rangle \cos(\omega t). \end{aligned}$$

So now we have an ODE to solve

$$c_n''(t) = -n^2 c_n(t) + \langle f, v_n \rangle \cos(\omega t)$$

$$c_n(0) = c_n'(0) = 0.$$

You can solve this using methods from Math. 20D - for example, variation of parameters. The result is:

$$c_n(t) = \frac{\langle f, v_n \rangle}{n^2 - \omega^2} (\cos(\omega t) - \cos(nt)).$$

Exercise 5. Check this answer.

Thus a solution for the problem posed by formula (10) is

$$(13) \quad u(x, t) = \sum_{n=1}^{\infty} a_n \frac{\cos(\omega t) - \cos(nt)}{n^2 - \omega^2} v_n(x), \quad a_n = \langle f, v_n \rangle, \quad v_n(x) = \frac{\sin(nx)}{\sqrt{\pi/2}}.$$

Exercise 6. What happens to formula (13) as $\omega \rightarrow \pm n$??? You need to compute

$$\lim_{\omega \rightarrow n} \frac{\cos(\omega t) - \cos(nt)}{\omega^2 - n^2}.$$

You should be able to deduce from this that unless $a_n=0$, the function $u(x,t)$ blows up as t goes to infinity. This is resonance.

The phenomenon of resonance is quite general. It works for all sorts of vibrating objects, even with non-constant density and tension and even in higher dimensions. Of course it can be used for good as well as evil. Consider musical instruments for example. Can they be played for evil? I suppose you might be able to destroy a glass by making it resonate.