MA 105 D3 Lecture 1

Ravi Raghunathan

Department of Mathematics

August 7, 2023

Course objectives

Welcome to IIT Bombay.

- To help the students achieve a better and more rigourous understanding of the calculus of one variable.
- To introduce the ideas and theorems in the calculus of several variables.
- To help students achieve a working knowledge of the tools and techniques of the calculus of several variables with a view to the applications they are likely to encounter in the future.

For details about the syllabus, tutorials, assignments, quizes, exams and procedures for evaluation please refer to the course booklet. The course booklet can also be found on moodle:
`http://moodle.iitb.ac.in/login/index.php`

The emphasis of this course will be on the underlying ideas and methods rather than very intricate problem solving involving formal manipulations (of course, there will be plenty of problems - just not many with lots of algebra tricks). The aim is to get you to think about calculus, in particular, and mathematics in general.

Ask questions! There is a good chance that if you don't understand something, many other people also do not understand it.

So, any questions before we start?

Sequences

Definition: A sequence in a set X is a function $a : \mathbb{N} \to X$, that is, a function from the natural numbers to X.

In this course X will usually be a subset of (or equal to) $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$, though we will also have occassion to consider sequences of functions sometimes. In later mathematics courses X may be the complex numbers $\mathbb{C}$ (MA 412), vector spaces (whatever those maybe) the set of continuous functions on an interval $\mathcal{C}([a, b])$ or other sets of functions (MA 110).

Rather than write the value of the function at n as $a(n)$, we often write a_n for the members of the sequence. A sequence is often specified by listing the first few terms

$$a_1, a_2, a_3, \ldots$$

or, more generally by describing the n^{th} term a_n. When we want to talk about the sequence as a whole we sometimes write $\{a_n\}_{n=1}^{\infty}$, but more often we once again just write a_n.

Examples of sequences

1. $a_n = n$ (here we can take $X = \mathbb{N} \subset \mathbb{R}$ if we want, and the sequence is just the identity function. Of course, we can also take $X = \mathbb{R}$).
2. $a_n = 1/n$ (here we can take $X = \mathbb{Q} \subset \mathbb{R}$ if we want, where $\mathbb{Q}$ denotes the rational numbers, or we can take $X = \mathbb{R}$ itself).
3. $a_n = \frac{n!}{n^n}$ ($X = \mathbb{Q}$ or $X = \mathbb{R}$).
4. $a_n = n^{1/n}$ (here the values taken by a_n are irrational numbers, so it best to take $X = \mathbb{R}$).
5. $a_n = \sin\left(\frac{1}{n}\right)$ (again the values taken by a_n are irrational numbers, so it best to take $X = \mathbb{R}$).

These are all examples of sequence of real numbers.

More examples

6. $a_n = \left(n^2, \frac{1}{n}\right)$ (here $X = \mathbb{R}^2$ or $X = \mathbb{Q}^2$).

   This is a sequence in $\mathbb{R}^2$.

7. $f_n(x) = \cos(nx)$ (here X is the set of continuous functions on any interval $[a, b]$ or even on $\mathbb{R}$).

   This is a sequence of functions. More precisely, it is a sequence of continuous functions.

Series

Given a sequence a_n of real numbers, we can manufacture a new sequence, namely its sequence of partial sums:

$$s_1 = a_1,\ s_2 = a_1 + a_2,\ s_3 = a_1 + a_2 + a_3, \ldots$$

More precisely, we have the sequence

$$s_n = \sum_{k=1}^{n} a_k.$$

8. We can take $a_n = r^n$, for some r, i.e., a geometric progression. Then $s_n = \sum_{k=0}^{n} r^k$.

9. $s_n(x) = \sum_{i=0}^{n} \frac{x^i}{i!}$, or writing it out $s_n(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$.

   We get a sequence of polynomial functions.

Monotonic sequences

For the moment we will concentrate on sequences in $\mathbb{R}$.

Definition: A sequence is said to be a monotonically increasing sequence if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$.

Definition: A sequence is said to be a monotonically decreasing sequence if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}$.

A monotonic sequence is one that is either monotonically increasing or monotonically decreasing.

From the examples in the previous slide, Example 1 is a monotonically increasing sequence, Example 2 is a monotonically decreasing sequence.
How about Example 3?

In Example 3 we notice that if $a_n = \frac{n!}{n^n}$,

$$a_{n+1} = \frac{(n+1)!}{(n+1)^{(n+1)}} = a_n \times \frac{(n+1)n^n}{(n+1)^{(n+1)}} \leq a_n,$$

so the sequence is monotonically decreasing.

Eventually monotonic sequences

In Example 4 ($a_n = n^{1/n}$), we note that

$$a_1 = 1 < 2^{1/2} = a_2 < 3^{1/3} = a_3,$$

(raise both a_2 and a_3 to the sixth power to see that $2^3 < 3^2$!). However, $3^1/3 > 4^{1/4} > 5^{1/5}$. So what do you think happens as n gets larger?

In fact, $a_{n+1} \leq a_n$, for all $n \geq 3$. Prove this fact as an exercise. Such a sequence is called an eventually monotonic sequence, that is, the sequence becomes monotonic(ally decreasing) after some stage. One can similarly define eventually monotonically increasing sequences.

Let us quickly run through the other examples. Example 5 - monotonically decreasing. Example 6 - is not a sequence of real numbers. Example 7 - is a sequence of real numbers if we fix a value of x . Can it be monotonic for some x? Example 8 is monotonic for any fixed value of r and so is Example 9 for any non-negative value of x.

Limits: Preliminaries

While all of you are familiar with limits, most of you have probably not worked with a rigourous definition. We will be more interested in limits of functions of a real variable (which is what arise in the differential calculus), but limits of sequences are closely related to the former, and occur in their own right in the theory of Riemann integration.

So what does it mean for a sequence to tend to a limit? Let us look at the sequence $a_n = 1/n^2$. We wish to study the behaviour of this sequence as n gets large. Clearly as n gets larger and larger, $1/n^2$ gets smaller and smaller and seems to approach the value 0, or more precisely

the distance between $1/n^2$ and 0 becomes smaller and smaller.

In fact (and this is the key point), by choosing n large enough, we can make the distance between $1/n^2$ and 0 smaller than any prescribed quantity.

Let us examine the above statement, and then try and quantify it.

More precisely:

The distance between $1/n^2$ and 0 is given by $|1/n^2 - 0| = 1/n^2$.

Suppose I require that $1/n^2$ be less that 0.1 (that is 0.1 is my prescribed quantity). Clearly, $1/n^2 < 1/10$ for all $n > 3$.

Similarly, if I require that $1/n^2$ be less than $0.0001 (= 10^{-4})$, this will be true for all $n > 100$.

We can do this for any number, no matter how small. If $\epsilon > 0$ is any number,

$$1/n^2 < \epsilon \iff 1/\epsilon < n^2 \iff n > 1/\sqrt{\epsilon}.$$

In other words, given any $\epsilon > 0$, we can always find a natural number N (in this case any $N > 1/\sqrt{\epsilon}$) such that for all $n > N$, $|1/n^2 - 0| < \epsilon$.

The rigourous definition of a limit

Motivated by the previous example, we define the limit as follows.

Definition: A sequence a_n tends to a limit l/converges to a limit l, if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - l| < \epsilon$$

whenever $n > N$.

This is what we mean when we write

$$\lim_{n \to \infty} a_n = l.$$

If we just want to say that the sequence has a limit without specifying what that limit is, we simply say $\{a_n\}_{n=1}^{\infty}$ converges, or that it is convergent.
A sequence that does not converge is said to diverge, or to be divergent.

Remarks on the definition

Remarks

1. Note that the N will (of course) depend on ϵ, as it did in our example, so it would have been more correct to write $N(\epsilon)$ in the definition of the limit. However, we usually omit this extra bit of notation.
2. We have already shown that $\lim_{n\to\infty} 1/n^2 = 0$. The same argument works for $\lim_{n\to\infty} 1/n^\alpha$, for any real $\alpha > 0$. We just take N to be any integer bigger than $1/\epsilon^{1/\alpha}$ for a given ϵ.
3. For a given ϵ, once one N works, any larger N will also work. In order to show that a sequence tends to a limit l we are not obliged to find the best possible N for a given ϵ, just some N that works. Thus, for the sequence $1/n^2$ and $\epsilon = 0.1$, we took $N = 3$, but we can also take $N = 10, 100, 1729$, or any other number bigger than 3.
4. Showing that a sequence converges to a limit l is not easy. One first has to guess the value l and then prove that l satisfies the definition. We will see how to get around this in various ways.

MA 105 D3 Lecture 2

Ravi Raghunathan

Department of Mathematics

August 8, 2023

Recap

Monotonic sequences

For the moment we will concentrate on sequences in $\mathbb{R}$.

Definition: A sequence is said to be a monotonically increasing sequence if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$.

Definition: A sequence is said to be a monotonically decreasing sequence if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}$.

A monotonic sequence is one that is either monotonically increasing or monotonically decreasing.
Definition: A sequence a_n is said to be eventually monotonically decreasing (resp. increasing) if there is an $N \in \mathbb{N}$ such that $a_n \geq a_{n+1}$ (resp. $a_n \leq a_{n+1}$ for all $n \geq \mathbb{N}$.

The rigourous definition of a limit

Definition: A sequence a_n tends to a limit l/converges to a limit l, if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - l| < \epsilon$$

whenever $n > N$.

This is what we mean when we write

$$\lim_{n\to\infty} a_n = l.$$

If we just want to say that the sequence has a limit without specifying what that limit is, we simply say $\{a_n\}_{n=1}^{\infty}$ converges, or that it is convergent.

A sequence that does not converge is said to diverge, or to be divergent.

Remarks on the definition

Remarks

1. Note that the N will (of course) depend on ϵ, as it did in our example, so it would have been more correct to write $N(\epsilon)$ in the definition of the limit. However, we usually omit this extra bit of notation.
2. We have already shown that $\lim_{n\to\infty} 1/n^2 = 0$. The same argument works for $\lim_{n\to\infty} 1/n^\alpha$, for any real $\alpha > 0$. We just take N to be any integer bigger than $1/\epsilon^{1/\alpha}$ for a given ϵ.
3. For a given ϵ, once one N works, any larger N will also work. In order to show that a sequence tends to a limit l we are not obliged to find the best possible N for a given ϵ, just some N that works. Thus, for the sequence $1/n^2$ and $\epsilon = 0.1$, we took $N = 3$, but we can also take $N = 10, 100, 1729$, or any other number bigger than 3.
4. Showing that a sequence converges to a limit l is not easy. One first has to guess the value l and then prove that l satisfies the definition. We will see how to get around this in various ways.

More examples of limits

Let us show that $\lim_{n\to\infty} \sin\left(\frac{1}{n}\right) = 0$.

For this we note that for $x \in [0, \pi/2]$, $0 \leq \sin x \leq x$ (try to remember why this is true).
Hence,

$$|\sin 1/n - 0| = |\sin 1/n| < 1/n.$$

Thus, given any $\epsilon > 0$, if we choose some $N > 1/\epsilon$, $n > N$ implies $1/n < 1/N < \epsilon$. It follows that $|\sin 1/n - 0| < \epsilon$.

Let us consider Exercise 1.1.(ii) of the tutorial sheet. Here we have to show that $\lim_{n\to\infty} 5/(3n+1) = 0$. Once again, we have only to note that

$$\frac{5}{3n+1} < \frac{5}{3n},$$

and if this is to be smaller than ϵ, we must have $n > N > 5/3\epsilon$.

Formulæ for limits

If a_n and b_n are two convergent sequences then

1. $\lim_{n\to\infty}(a_n \pm b_n) = \lim_{n\to\infty} a_n \pm \lim_{n\to\infty} b_n$
2. $\lim_{n\to\infty}(a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$.
3. $\lim_{n\to\infty}(a_n/b_n) = \lim_{n\to\infty} a_n / \lim_{n\to\infty} b_n$, provided $\lim_{n\to\infty} b_n \neq 0$

Implicit in the formulæ is the fact that the limits on left hand side exist if the limits on the right hand side exist.

Note that the constant sequence $a_n = c$ has limit c, so as a special case of (2) above we have

$$\lim_{n\to\infty}(c \cdot b_n) = c \cdot \lim_{n\to\infty} b_n.$$

Using the formulæ above we can break down the limits of more complicated sequences into simpler ones and evaluate them.

The Sandwich Theorem(s)

Theorem 1: If a_n, b_n and c_n are convergent sequences such that $a_n \leq b_n \leq c_n$ for all n, then

$$\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n \leq \lim_{n\to\infty} c_n.$$

A second version of the theorem is especially useful:
Theorem 2: Suppose $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n$. If b_n is a sequence satisfying $a_n \leq b_n \leq c_n$ for all n, then b_n converges and

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \lim_{n\to\infty} c_n.$$

Note that we do not assume that b_n converges in this version of the theorem - we get the convergence of b_n for free . Together with the rules for sums, differences, products and quotients, this theorem allows us to handle a large number of more complicated limits.

An example using the theorems above

Consider Exercise 1.2.(iii) on the tutorial sheet. We have to show that

$$\lim_{n\to\infty} \frac{n^3+3n^2+1}{n^4+8n^2+2}$$

exists and to evaluate it.

It is clear that

$$0 < \frac{n^3+3n^2+1}{n^4+8n^2+2} \le \frac{1}{n} + \frac{3}{n^2} + \frac{1}{n^4}.$$

How do we get this? Note that $n^3/(n^4+8n^2+2) < n^3/n^4 = 1/n$, and the other two terms can be handled similarly.

Hence, applying the Sandwich Theorem (Theorem 2) to the sequences

$$a_n = 0, \quad b_n = \frac{n^3+3n^2+1}{n^4+8n^2+2} \quad \text{and} \quad c_n = \frac{1}{n} + \frac{3}{n^2} + \frac{1}{n^4}$$

we see that the limit we want exists provided $\lim_{n\to\infty} c_n$ exists, so this is what we must concentrate on proving.

The limit $\lim_{n\to\infty} c_n$ exists provided each of the terms appearing in the sum has a limit and in that case it is equal to the sum of the limits (by the first formula). But each of these limits is quite easy to evaluate.

We already know that

$$\lim_{n\to\infty} 1/n = 0 = \lim_{n\to\infty} 1/n^4,$$

while

$$\lim_{n\to\infty} 3/n^2 = 3 \cdot \lim_{n\to\infty} 1/n^2 = 0$$

where we have used the special case of the second formula (limit of the product is the product of the limits) for the first equality in the equation above. Since all three limits converge to 0, it follows the given limit is $0 + 0 + 0 = 0$.

Bounded Sequences

The formulæ and theorems stated above can be easily proved starting from the definitions. We will prove the second formula and leave the other proofs as exercises.
Definition: A sequence a_n is said to be bounded if there is a real number $M > 0$ such that $|a_n| \leq M$ for every $n \in \mathbb{N}$. A sequence that is not bounded is called unbounded.

In our list of examples, Example 1 ($a_n = n$) is an example of an unbounded sequence, while Examples 2 - 5 ($a_n = 1/n, \sin(1/n), n!/n^n, n^{1/n}$) are examples of bounded sequences.

Bounded sequences don't necessarily converge - for instance $a_n = (-1)^n$. However,

Convergent sequences are bounded

Lemma: Every convergent sequence is bounded.

Proof: Suppose a_n converges to l. Choose $\epsilon = 1$. There exists $N \in \mathbb{N}$ such that $|a_n - l| < 1$ for all $n > N$. In other words, $l - 1 < a_n < l + 1$, for all $n > N$, which gives $|a_n| < |l| + 1$ for all $n > N$. Let

$$M_1 = \max\{|a_1|, |a_2|, \ldots, |a_N|\}$$

and let $M = \max\{M_1, |l| + 1\}$. Then $|a_n| \leq M$ for all $n \in \mathbb{N}$. $\square$

In the slides presented in class, I had forgotten to put absolute value signs in many places in the proof above and in the next slide. This has now been corrected.

We will use this Lemma to prove the product rule for limits.

The proof of the product rule

We wish to prove that $\lim_{n\to\infty} a_n b_n = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$.

Suppose $\lim_{n\to\infty} a_n = l_1$ and $\lim_{n\to\infty} b_n = l_2$. We need to show that $\lim_{n\to\infty} a_n b_n = l_1 l_2$.

Fix $\epsilon > 0$. We need to show that we can find $N \in \mathbb{N}$ such that $|a_n b_n - l_1 l_2| < \epsilon$, whenever $n > N$. Notice that

$$\begin{aligned} |a_n b_n - l_1 l_2| &= |a_n b_n - a_n l_2 + a_n l_2 - l_1 l_2| \\ &= |a_n(b_n - l_2) + (a_n - l_1) l_2| \\ &\leq |a_n||b_n - l_2| + |a_n - l_1||l_2|, \end{aligned}$$

where the last inequality follows from the triangle inequality. So in order to guarantee that the left hand side is small, we must ensure that the two terms on the right hand side together add up to less than ϵ. In fact, we make sure that each term is less than $\epsilon/2$.

The proof of the product rule, continued

Since a_n is convergent, it is bounded by the lemma we have just proved. Hence, there is an M such that $|a_n| < M$ for all $n \in \mathbb{N}$.

Given the quantities $\epsilon/2l_2$ and $\epsilon/2M$, there exist N_1 and N_2 such that

$$|a_n - l_1| < \epsilon/2l_2 \quad \text{and} \quad |b_n - l_2| < \epsilon/2M.$$

Let $N = \max\{N_1, N_2\}$. If $n > N$, then both the inequalities above hold. Hence, we have

$$|a_n||b_n - l_2| \leq M \cdot \frac{\epsilon}{2M} = \frac{\epsilon}{2} \quad \text{and} \quad |a_n - l_1||l_2| \leq l_2 \cdot \frac{\epsilon}{2l_2} = \frac{\epsilon}{2}.$$

Now it follows that

$$|a_n b_n - l_1 l_2| \leq |a_n||b_n - l_2| + |a_n - l_1||l_2| < \epsilon,$$

for all $n > N$, which is what we needed to prove. $\square$

The proofs of the other rules for limits are similar to the one we proved above. Try them as exercises.

A guarantee for convergence

As we mentioned earlier, proving that a limit exists is hard because we have to guess what its value might be and then prove that it satisfies the definition. The following theorem guarantees the convergence of a sequence without knowing the limit beforehand.
Definition: A sequence a_n is said to be bounded above (resp. bounded below) if $a_n < M$ (resp. $a_n > M$) for some $M \in \mathbb{R}$.
A sequence that is bounded both above and below is obviously bounded.

Theorem 3: A montonically increasing (resp. decreasing) sequence which is bounded above (resp. below) converges.

Remarks on Theorem 3

Theorem 3 clearly makes things very simple in many cases. For instance, if we have a monotonically decreasing sequence of positive numbers, it must have a limit, since 0 is always a lower bound!

Can we guess what the limit of a monotonically increasing sequence a_n bounded above might be?
It will be the supremum or least upper bound (lub) of the sequence.
This is the number, say M which has the following properties:

1. $a_n \leq M$ for all n and
2. If M_1 is such that $a_n \leq M_1$ for all n, then $M \leq M_1$.

The point is that a sequence bounded above may not have a maximum but will always have a supremum. As an example, take the sequence $1 - 1/n$. Clearly there is no maximal element in the sequence, but 1 is its supremum.

Another monotonic sequence

Let us look at Exercise 1.5.(i) which considers the sequence

$$a_1 = 3/2 \quad \text{and} \quad a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right).$$

$$\begin{aligned} a_{n+1} \quad < a_n \quad &\iff \quad \tfrac{1}{2}\left(a_n + \tfrac{2}{a_n}\right) < a_n \\ &\iff \quad \sqrt{2} < a_n. \end{aligned}$$

On the other hand,

$$\frac{1}{2}\left(a_n + \frac{2}{a_n}\right) \geq \sqrt{2}, \quad \text{(Why is this true?)}$$

so $a_{n+1} \geq \sqrt{2}$ for all $n \geq 1$ and $a_1 > \sqrt{2}$ is given.
Hence, $\{a_n\}_{n=1}^{\infty}$ is a monotonically decreasing sequence, bounded below by $\sqrt{2}$. By Theorem 3, it converges.
Exercise 1. What do you think is the limit of the above sequence (Refer to the supplement to Tutorial 1)?

More remarks on limits

Exercise 2. More generally, what is the limit of a monotonically decreasing sequence bounded below? How can you describe it? This number is called the infimum or greatest lower bound (glb) of the sequence.

The proof of Theorem 3 is not so easy and more or less involves understanding what a real number is. It is related to the notion of a Cauchy sequence about which I will try to say something a little later (again, refer to the supplement to Tutorial 1).

An important remark: If we change finitely many terms of a sequence it does not affect the convergence of a sequence or the fact that it is bounded or unbounded.
If it is convergent, the limit will not change. If it is bounded, it will remain bounded though the supremum may change. Thus, an eventually monotonically increasing sequence bounded above will converge (formulate the analogue for decreasing sequences).
Bottomline: From the point of view of the limit, only what happens for large n matters.

MA 105 D3 Lecture 3

Ravi Raghunathan

Department of Mathematics

August 10, 2023

Recap

Cauchy sequences: the definition

Limits of functions

The rigourous definition of a limit

Definition: A sequence a_n tends to a limit l/converges to a limit l, or that $\lim_{n\to\infty} a_n = l$, if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - l| < \epsilon$$

whenever $n > N$.

We also stated rules for limits and the Sandwich Theorems and proved the product rule. Once we have these rules, we can handle more complicated limits.

Lemma: Every convergent sequence is bounded.

Theorem 3: A montonically increasing (resp. decreasing) sequence which is bounded above (resp. below) converges to its supremum (resp. infimum).

Example

Let us look at Exercise 1.5.(i) which considers the sequence

$$a_1 = 3/2 \quad \text{and} \quad a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right).$$

$$\begin{aligned} a_{n+1} < a_n \quad &\iff \quad \tfrac{1}{2}\left(a_n + \tfrac{2}{a_n}\right) < a_n \\ &\iff \quad \sqrt{2} < a_n. \end{aligned}$$

On the other hand,

$$\frac{1}{2}\left(a_n + \frac{2}{a_n}\right) \geq \sqrt{2}, \quad \text{(Why is this true?)}$$

so $a_{n+1} \geq \sqrt{2}$ for all $n \geq 1$ and $a_1 > \sqrt{2}$ is given.
Hence, $\{a_n\}_{n=1}^{\infty}$ is a monotonically decreasing sequence, bounded below by $\sqrt{2}$. By Theorem 3, it converges.
Exercise 1. What do you think is the limit of the above sequence (Refer to the supplement to Tutorial 1)?

Cauchy sequences

As we have seen, it is not easy to tell whether a sequence converges or not because we have to first guess what the limit might be, and then try and prove that the sequence actually converges to this limit. For a monotonic sequence, we have a criterion, but what about more general sequences?

There is another very useful notion which allows us to decide whether the sequence converges by looking only at the elements of the sequence itself. We describe this below.
Definition: A sequence a_n in $\mathbb{R}$ is said to be a Cauchy sequence if for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - a_m| < \epsilon,$$

for all $m, n > N$.

Cauchy sequences: the theorem

Theorem 4: Every Cauchy sequence in $\mathbb{R}$ converges to a limit in $\mathbb{R}$.

Remark 1: One can now check the convergence of a sequence just by looking at the sequence itself!

Remark 2: One can easily check the converse:

Theorem 5: Every convergent sequence is Cauchy.
Remark 3: Remember that when we defined sequences we defined them to be functions from $\mathbb{N}$ to X, for any set X. So far we have only considered $X = \mathbb{R}$, but as we said earlier we can take other sets, for instance, susbets of $\mathbb{R}$. For instance, if we take $X = \mathbb{R} \setminus 0$, Theorem 4 is not valid. The sequence $1/n$ is a Cauchy sequence in this X but obviously does not converge in X. If we take $X = \mathbb{Q}$, the example given in 1.5.(i) ($a_{n+1} = (a_n + 2/a_n)/2$) is a Cauchy sequence in $\mathbb{Q}$ which does not converge in $\mathbb{Q}$. Thus Theorem 4 is really a theorem about real numbers.

The completeness of $\mathbb{R}$

A set in which every Cauchy sequence converges (to a limit which is also in the set) is called a a complete set. Thus Theorem 4 is sometimes rewritten as

Theorem 4': The real numbers are complete.

You will see other examples of complete sets in MA 110.

Sequences in $\mathbb{R}^2$ and $\mathbb{R}^3$

Most of our definitions for sequences in $\mathbb{R}$ are actually valid for sequences in $\mathbb{R}^2$ and $\mathbb{R}^3$. Indeed, the only thing we really need to define the limit is the notion of distance. Thus if we replace the modulus function $|\ |$ on $\mathbb{R}$ by the distance functions in $\mathbb{R}^2$ and $\mathbb{R}^3$ all the definitions of convergent sequences and Cauchy sequences remain the same.

For instance, a sequence $a(n) = (x_n, y_n)$ in $\mathbb{R}^2$ is said to converge to a point $l = (l_1, l_2)$ (in $\mathbb{R}^2$) if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\sqrt{(x_n - l_1)^2 + (y_n - l_2)^2} < \epsilon$$

whenver $n > N$. A similar defintion can be made in $\mathbb{R}^3$ using the distance function on $\mathbb{R}^3$. Some of our earlier rules will make sense and some won't.

Infinite series - a more rigourous treatment

Let us recall what we mean when we write

$$a + ar + ar^2 + \ldots = \frac{a}{1-r}.$$

Another way of writing the same expression is

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}.$$

The precise meaning is the following. Form the partial sums

$$s_n = \sum_{k=0}^{n} ar^k.$$

These partial sums $s_1, s_2, \ldots s_n, \ldots$ form a sequence and by $\sum_{k=0}^{\infty} ar^k = a/(1-r)$, we mean $\lim_{n\to\infty} s_n = a/1-r$.
So when we speak of the sum of an infinite series, what we really mean is the limit of its partial sums.

Convergence of the geometric series

So to justify our formula we should show that $\lim_{n\to\infty} s_n = a/1 - r$, that is, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\left| s_n - \frac{a}{1-r} \right| < \epsilon,$$

for all $n > N$.
In other words we need to show that

$$\left| \frac{a(1-r^{n+1})}{1-r} - \frac{a}{1-r} \right| = \left| \frac{ar^{n+1}}{1-r} \right| < \epsilon$$

if n is chosen large enough.
But $\lim_{n\to\infty} r^n = 0$, so there exists N such that $r^{n+1} < (1-r)\epsilon/a$ for all $n > N$, so for this N, if $n > N$,

$$\left| s_n - \frac{a}{1-r} \right| < \epsilon.$$

This shows that the geometric series converges to the given expression.

Limits of functions of a real variable

Since we have already defined the limit of a sequence rigourously, it will not be hard to define the limit of a real valued function $f : (a, b) \to \mathbb{R}$.

Definition: A function $f : (a, b) \to \mathbb{R}$ is said to tend to (or converge to) a limit l at a point $x_0 \in [a, b]$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - l| < \epsilon$$

for all $x \in (a, b)$ such that $0 < |x - x_0| < \delta$.In this case, we write

$$\lim_{x \to x_0} f(x) = l,$$

or $f(x) \to l$ as $x \to x_0$ which we read as "$f(x)$" tends to l as x tends to x_0". This is just the rigourous way of saying that the

distance between $f(x)$ and l can be made as small as one pleases by making the distance between x and x_0 sufficiently small.

A subtle point and the rules for limits

Notice that in the definition above, the point x_0 can be one of the end points a or b.

Thus the limit of a function may exist even if the function is not defined at that point.

The rules and formulæ for limits of functions are the same as those for sequence and can be proved in almost exactly the same way. If $\lim_{x\to x_0} f(x) = l_1$ and $\lim_{x\to x_0} g(x) = l_2$, then

1. $\lim_{x\to x_0} f(x) \pm g(x) = l_1 \pm l_2$.
2. $\lim_{x\to x_0} f(x)g(x) = l_1 l_2$.
3. $\lim_{x\to x_0} f(x)/g(x) = l_1/l_2$. provided $l_2 \neq 0$

As before, implicit in the formulæ is the fact that if the limits on the left hand side exist. We prove the first rule below.

The proof of the addition formula for limits

Proof: We first show that $\lim_{x\to x_0} f(x)+g(x) = l_1 + l_2$. Let $\epsilon > 0$ be arbitrary.

Since $\lim_{x\to x_0} f(x) = l_1$ and $\lim_{x\to x_0} g(x) = l_2$, there exist δ_1, δ_2 such that

$$|f(x) - l_1| < \frac{\epsilon}{2} \quad \text{and} \quad |g(x) - l_2| < \frac{\epsilon}{2}$$

whenever $0 < |x - x_0| < \delta_1$ and $0 < |x - x_0| < \delta_2$. If we choose $\delta = \min\{\delta_1, \delta_2\}$ and if $0 < |x - x_0| < \delta$ then both the above inequalities hold. Thus, if $|x - x_0| < \delta$, then

$$\begin{aligned} |f(x) + g(x) - (l_1 + l_2)| &= |f(x) - l_1 + g(x) - l_2| \\ &\leq |f(x) - l_1| + |g(x) - l_2| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

which is what we needed to prove. If we replace $g(x)$ by $-g(x)$ we get the second part of the first rule. $\square$

The Sandwich Theorem(s) for limits of functions

Theorem 5: As $x \to x_0$, if $f(x) \to l_1$, $g(x) \to l_2$ and $h(x) \to l_3$ for functions f, g, h on some interval (a, b) such that $f(x) \leq g(x) \leq h(x)$ for all $x \in (a, b)$, then

$$l_1 \leq l_2 \leq l_3.$$

As before, we have a second version.

Theorem 6: Suppose $\lim_{x \to x_0} f(x) = \lim_{x \to x_0} h(x) = l$ and If $g(x)$ is a function satisfying $f(x) \leq g(x) \leq h(x)$ for all $x \in (a, b)$, then $g(x)$ converges to a limit as $x \to x_0$ and

$$\lim_{x \to x_0} g(x) = l$$

Once again, note that we do not assume that $g(x)$ converges to a limit in this version of the theorem - we get the convergence of $g(x)$ for free .

Some examples

Let us look at Exercise 1.11. We will use this exercise to explore a few subtle points.
Let $c \in [a, b]$ and $f, g : (a, b) \to \mathbb{R}$ be such that $\lim_{x \to c} f(x) = 0$.
Prove or disprove the following statements.

(i) $\lim_{x \to c}[f(x)g(x)] = 0$.

(ii) $\lim_{x \to c}[f(x)g(x)] = 0$, if g is bounded.($g(x)$ is said to be bounded on (a, b) if there exists $M > 0$ such that $|g(x)| < M$ for all $x \in (a, b)$.

(iii) $\lim_{x \to c}[f(x)g(x)] = 0$, if $\lim_{x \to c} g(x)$ exists.

Before getting into proofs, let us guess whether the statements above are true or false.

(i) false

(ii) true

(iii) true.

(i) Notice that $g(x)$ is not given to be bounded - if this was not obvious before, you should suspect that such a condition is needed after looking at part (ii). So the most natural thing to do is to look for a counter-example to this statement by taking $g(x)$ to be an unbounded function. What is the simplest example of an unbounded function $g(x)$ on an open interval?

How about $g(x) = \frac{1}{x}$ on $(0, 1)$?
What would a candidate for $f(x)$ be - what is the simplest example of a function $f(x)$ which tends to 0 for some value of c in $[0, 1]$.

$f(x) = x$, and $c = 0$ is a pretty simple candidate.

Clearly $\lim_{x\to 0} f(x)g(x) = \lim_{x\to 0} 1 = 1 \neq 0$, which shows that (i) is not true in general.

Exercise 1: Can you find a counter-example to (i) with c in (a, b) (that is, c should not be one of the end points)? (Hint: Can you find an unbounded function on a closed interval $[a, b]$?)

Let us move to part (ii).
Suppose $g(x)$ is bounded on (a, b). This means that there is some real number $M > 0$ such that $|g(x)| < M$. Let $\epsilon > 0$. We would like to show that

$$|f(x)g(x) - 0| = |f(x)g(x)| < \epsilon,$$

if $0 < |x - c| < \delta$ for some $\delta > 0$.
Since $\lim_{x \to c} f(x) = 0$, there exists $\delta > 0$ such that $|f(x)| < \epsilon/M$ for all $|x - c| < \delta$. It follows that

$$|f(x)g(x)| = |f(x)||g(x)| < \frac{\epsilon}{M} \cdot M = \epsilon$$

for all $0 < |x - c| < \delta$, and this is what we had to show.

Part (iii) follows immediately from the product rule, but can one deduce part (iii) from (ii) instead?
Hint: Think back to the lemma on convergent sequences that we proved in Lecture 1: Every convergent sequence is bounded. What is the analogue for functions which converge to a limit at some point? Indeed, you can easily show the following

Lemma 7: Let $f : (a, b) \to \mathbb{R}$ be a function such that $\lim_{x \to c} f(x)$ exists for some $c \in [a, b]$. If $c \in (a, b)$, there exists an (open) interval $I = (c - \eta, c + \eta) \subset (a, b)$ such that $f(x)$ is bounded on I. If $c = a$, then there is an open interval $I_1 = (a, a + \eta)$ such that $f(x)$ is bounded on I_1. Similarly if $c = b$, there exists an open interval $I_2 = (b - \eta, b)$ such that $f(x)$ is bounded on I_2.

The proof of the lemma above is almost the same as the the lemma for convergent sequences. Basically, replace "N" by "δ" in the proof.

If one applies the Lemma above to $g(x)$, we see that $g(x)$ is bounded in some (possibly) smaller interval $(0, \eta)$. Now apply part (ii) to this interval to deduce that (iii) is true.

MA 105 D3 Lecture 4

Ravi Raghunathan

Department of Mathematics

August 14, 2023

Recap

Odds and ends about limits of functions of a real variable

Continuity

More about continuous functions

The rigourous definition of a limit of a function

Definition: A function $f : (a, b) \to \mathbb{R}$ is said to tend to (or converge to) a limit l at a point $x_0 \in [a, b]$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - l| < \epsilon$$

for all $x \in (a, b)$ such that $0 < |x - x_0| < \delta$. In this case, we write

$$\lim_{x \to x_0} f(x) = l,$$

or $f(x) \to l$ as $x \to x_0$ which we read as "$f(x)$" tends to l as x tends to x_0".

The the limit of a function may exist even if the function is not defined at that point because x_0 can be a or b.

Rules for limits

If $\lim_{x \to x_0} f(x) = l_1$ and $\lim_{x \to x_0} g(x) = l_2$, then

1. $\lim_{x \to x_0} f(x) \pm g(x) = l_1 \pm l_2$.
2. $\lim_{x \to x_0} f(x)g(x) = l_1 l_2$.
3. $\lim_{x \to x_0} f(x)/g(x) = l_1/l_2$. provided $l_2 \neq 0$

Theorem 5: As $x \to x_0$, if $f(x) \to l_1$, $g(x) \to l_2$ and $h(x) \to l_3$ for functions f, g, h on some interval (a, b) such that $f(x) \leq g(x) \leq h(x)$ for all $x \in (a, b)$, then

$$l_1 \leq l_2 \leq l_3.$$

As before, we have a second version.

Theorem 6: Suppose $\lim_{x \to x_0} f(x) = \lim_{x \to x_0} h(x) = l$ and If $g(x)$ is a function satisfying $f(x) \leq g(x) \leq h(x)$ for all $x \in (a, b)$, then $g(x)$ converges to a limit as $x \to x_0$ and

$$\lim_{x \to x_0} g(x) = l$$

Limits at infinity

There is one further case of limits that we need to consider. This occurs when we consider functions defined on open intervals of the form $(-\infty, b)$, (a, ∞) or $(-\infty, \infty) = \mathbb{R}$ and we wish to define limits as the variable goes to plus or minus infinity. The definition here is very similar to the definition we gave for sequences. Let us consider the last case.

Definition: We say that $f : \mathbb{R} \to \mathbb{R}$ tends to a limit l as $x \to \infty$ (resp. $x \to -\infty$) if for all $\epsilon > 0$ there exists $X \in \mathbb{R}$ such that

$$|f(x) - l| < \epsilon,$$

whenever $x > X$ (resp. $x < X$), and we write

$$\lim_{x\to\infty} f(x) = l \quad \text{or} \quad \lim_{x\to-\infty} f(x) = l.$$

or, alternatively, $f(x) \to l$ as $x \to \infty$ or as $x \to -\infty$, depending on which case we are considering.

Limits from the left and right

If $f : (a, b) \to \mathbb{R}$ is a function and $c \in (a, b)$, then it is possible to approach c from either the left or the right on the real line.

We can define the limit of the function $f(x)$ as x approaches c from the left (if it exists) as a number l^- such that for all $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - l^-| < \epsilon$ whenever $|x - c| < \delta$ and $x \in (a, c)$.

Our notation for this is $\lim_{x \to c-} f(x) = l^-$, and it is also called the left hand (side) limit.

Exercise 2: Write down a definition for the limit of a function from the right. We usually denote the right hand (side) limit by $lim_{x \to c+} f(x)$. Show, using the definitions, that $\lim_{x \to c} f(x)$ exists if and only if the left hand and right hand limits both exist and are equal.

We can also think of the left hand limit as follows. We restrict our attention to the interval (a, c), that is we think of f as a function only on this interval. Call this restricted function f_a. Then, another way of defining the left hand limit is

$$\lim_{x \to c-} f(x) = \lim_{x \to c} f_a(x).$$

It should be easy to see that it is the same as the definition before One can make a similar definition for the right hand limit.

The notions of left and right hand limits are useful because sometimes a function is defined in different ways to the left and right of a particular point. For instance, $|x|$ has different definitions to the left and right of 0.

Calculating limits explicitly

As with sequences, using the rules for limits of functions together with the Sandwich theorem allows one to treat the limits of a large number of expressions once one knows a few basic ones:

(i) $\lim_{x\to 0} x^\alpha = 0$ if $\alpha > 0$, (ii) $\lim_{x\to\infty} x^\alpha = 0$ if $\alpha < 0$,
(iii) $\lim_{x\to 0} \sin x = 0$, (iv) $\lim_{x\to 0} \sin x/x = 1$
(v) $\lim_{x\to 0}(e^x - 1)/x = 1$, (vi) $\lim_{x\to 0} \ln(1+x)/x = 1$
We have not concentrated on trying to find limits of complicated expressions of functions using clever algebraic manipulations or other techniques. However, I can't resist mentioning the following problem.
Exercise 3: Find

$$\lim_{x\to 0} \frac{\sin(\tan x) - \tan(\sin x)}{\arcsin(\arctan x) - \arctan(\arcsin x)}.$$

I will give the solution next time, together with the history of the problem. Six years ago, one needed to know a couple of keywords in order to get the solution through google. But as of three years ago, just typing in the formula above into google will lead you to the solution.

Continuity - the definition

Definition: If $f : [a, b] \to \mathbb{R}$ is a function and $c \in [a, b]$, then f is said to be continuous at the point c if and only if

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

Thus, if c is one of the end points we require only the left or right hand limit to exist.

A function f on (a, b) (resp. $[a, b]$) is said to be continuous if and only if it is continuous at every point c in (a, b) (resp. $[a, b]$).

If f is not continuous at a point c we say that it is dicontinuous at c, or that c is a point of discontinuity for f.

Intuitively, continuous functions are functions whose graphs can be drawn on a sheet of paper without lifting the pencil of the sheet of paper. That is, there should be no "jumps" in the graph of the function.

Continuity of familiar functions: polynomials

What are the functions we really know or understand? What does "knowing" or understanding a function $f(x)$ even mean? Presumably, if we understand a function f, we should be able to calculate the value of the function $f(x)$ at any given point x. But if you think about it, for what functions $f(x)$ can you really do this?

One class of functions is the polynomial functions. More generally we can understand rational functions, that is functions of the form $R(x) = P(x)/Q(x)$ where $P(x)$ and $Q(x)$ are polynomials, since we can certainly compute the values of $R(x)$ by plugging in the value of x. How do we show that polynomials or rational functions are continuous (on $\mathbb{R}$)?

It is trivial to show from the definition that the constant functions and the function $f(x) = x$ are continuous. Because of the rules for limits of functions, the sum, difference, product and quotient (with non-zero denominator) of continuous functions are continuous. Applying this fact we see easily that $R(x)$ is continuous whenever the denominator is non-zero.

Continuity of other familiar functions

What are the other (continuous) functions we know? How about the trignometric functions? Well, here it is less clear how to proceed. After all we can only calculate $\sin x$ for a few special values of x ($x = 0, \pi/6, \pi/4, \ldots$ etc.). How can we show continuity when we don't even know how to compute the function?

Of course, if we define $\sin x$ as the y-coordinate of a point on the unit circle it seems intuitively clear that the y-coordinate varies continuously as the point varies on the unit circle, but knowing the precise definition of continuity this argument should not satisfy you.

We will not prove the continuity of $\sin x$ in this course, though we will given an idea of how this is done next week. So let us assume from now on that $\sin x$ is continuous. How can we show that $\cos x$ is continuous?

The composition of continuous functions

Theorem 8: Let $f : (a, b) \to (c, d)$ and $g : (c, d) \to (e, f)$ be functions such that f is continuous at x_0 in (a, b) and g is continuous at $f(x_0) = y_0$ in (c, d). Then the function $g(f(x))$ (also written as $g \circ f(x)$ sometimes) is continuous at x_0. So the composition of continuous functions is continuous.

Exercise 4: Prove the theorem above starting from the definition of continuity.

Using the theorem above we can show that $\cos x$ is continuous if we show that $\sqrt{x}$ is continuous, since $\cos x = \sqrt{1 - \sin^2 x}$ and we know that $1 - \sin^2 x$ is continuous since it is the product of the sums of two continuous functions ($(1 + \sin x)$ and $(1 - \sin x)$!).

Once we have the continuity of $\cos x$ we get the continuity of all the rational trignometric functions, that is functions of the form $P(x)/Q(x)$, where P and Q are polynomials in $\sin x$ and $\cos x$, provided $Q(x)$ is not zero.

The continuity of the square root function

Thus in order to prove the continuity of $\cos x$ (assuming the continuity of $\sin x$) we need only prove the continuity of the square root function.

The main observation is that continuity is a local property, that is, only the behaviour of the function near the point being investigated is important.

Let $x_0 \in [0, \infty)$. To show that the square root function is continuous at x_0 we need to show that $\lim_{y \to x_0} \sqrt{y} = \sqrt{x_0}$, that is we need to show that $|\sqrt{y} - \sqrt{x_0}| < \epsilon$ whenever $0 < |y - x_0| < \delta$ for some δ. First assume that $x_0 \neq 0$. Then

$$|\sqrt{y} - \sqrt{x_0}| = \left| \frac{y - x_0}{\sqrt{y} + \sqrt{x_0}} \right| < \frac{|y - x_0|}{\sqrt{x_0}}.$$

If we choose $\delta = \epsilon \sqrt{x_0}$, we see that

$$|\sqrt{y} - \sqrt{x_0}| < \epsilon,$$

which is what we needed to prove ($x_0 = 0$, exercise!). $\square$

The intermediate value theorem

One of the most important properties of continuous functions is the Intermediate Value Property (IVP). We will use this property repeatedly to prove other results.

Theorem 9: Suppose $f : [a, b] \to \mathbb{R}$ is a continuous function. For every u between $f(a)$ and $f(b)$ there exists $c \in [a, b]$ there such that $f(c) = u$.

Functions which have this property are said to have the Intermediate Value Property. Theorem 9 can thus be restated as saying that continuous functions have the IVP.

We will not be proving this property - it is a consequence of the completeness of the real numbers. Intuitively, this is clear. Since one can draw the graph of the function without lifting one's pencil off the sheet of paper, the pencil must cut every line $y = e$ with e between $f(a)$ and $f(b)$.

The IVT in a picture

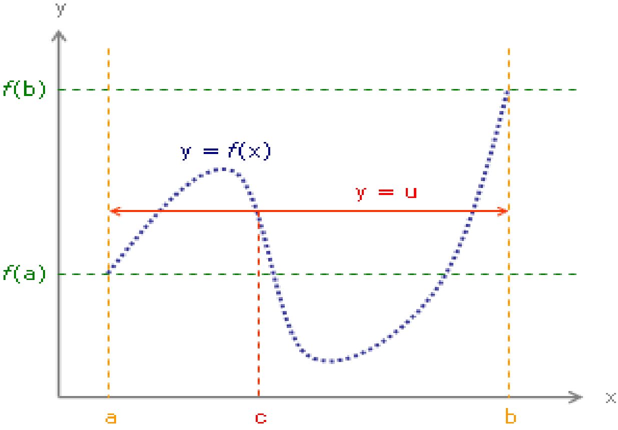


Image created by Enoch Lau see
http://en.wikipedia.org/wiki/File:Intermediatevaluetheorem.png

Zeros of functions

One of the most useful applications of the intermediate value property is to find roots of polynomials, or, more generally, to find zeros of continuous functions, that is to find points $x \in \mathbb{R}$ such that $f(x) = 0$.

Theorem 10: Every polynomial of odd degree has at least one real root.

Proof: Let $P(x) = a_n x^n + \ldots + a_0$ be a polynomial of odd degree. We can assume without loss of generality that $a_n > 0$. It is easy to see that if we take $x = b > 0$ large enough, $P(b)$ will be positive. On the other hand, by taking $x = a < 0$ small enough, we can ensure that $P(a) < 0$. Since $P(x)$ is continuous, it has the IVP, so there must be a point $x_0 \in (a, b)$ such that $f(x_0) = 0$. $\square$

The IVP can often be used to get more specific information. For instance, it is not hard to see that the polynomial $x^4 - 2x^3 + x^2 + x - 3$ has a root that lies between 1 and 2.

Continuous functions on closed, bounded intervals

The other major result on continuous functions that we need is the following. A closed bounded interval is one of the form $[a, b]$, where $-\infty < a$ and $b < \infty$.

Theorem 11: A continuous function on a closed bounded interval $[a, b]$ is bounded and attains its infimum and supremum, that is, there are points x_1 and x_2 in $[a, b]$ such that $f(x_1) = m$ and $f(x_2) = M$, where m and M denote the infimum and supremum respectively.

We defined infimum and supremum for sequences previously. The definition for functions of a real variable is the same: Let $X \subset \mathbb{R}$ and let $f : X \to \mathbb{R}$ be a function. A real number M is called the supremum of $f(x)$ (on X) if
1. If $f(x) \leq M$ for all $x \in X$.
2. If for some real number M_1 $f(x) \leq M_1$ for all $x \in X$, then $M \leq M_1$.

Relaxing the conditions

Again, we will not prove Theorem 11, but will use it quite often. Note the contrast with open intervals. The function $1/x$ on $(0, 1)$ does not attain a maximum - in fact it is unbounded. Similarly the function $1/x$ on $(1, \infty)$ does not attain a minimum, although, it is bounded below and the infimum is 0.

Exercise 5: In light of the above theorem, can you find a continuous function $g : (a, b) \to \mathbb{R}$ for part (i) of Exercise 1.11, with $c \in (a, b)$?

The function $\sin \frac{1}{x}$

Let us look at Exercise 1.13 part (i).
Consider the function defined as $f(x) = \sin \frac{1}{x}$ when $x \neq 0$, and $f(0) = 0$. Is this function is continuous at $x = 0$.
How about $x \neq 0$? Why is $f(x)$ continuous? Because it is a compostion continuous functions (sine and $1/x$).

Let us look at the sequence of points $x_n = 2/(2n+1)\pi$.
Clearly $x_n \to 0$ as $n \to \infty$.
For these points $f(x_n) = \pm 1$. This means that no matter how small I take my δ, there will be a point $x_n \in (0, \delta)$, such that $|f(x_n)| = 1$. But this means that $|f(x) - f(0)| = |f(x)|$ cannot be made smaller than 1 no matter how small δ may be. Hence, f is not continuous at 0. The same kind of argument will show that there is no value that we can assign $f(0)$ to make the function $f(x)$ continuous at 0.

You can easily check that $f(x)$ has the IVP. However, we have proved that it is not continuous. So IVP $\not\Rightarrow$ continuity.

MA 105 D3 Lecture 5

Ravi Raghunathan

Department of Mathematics

August 17, 2023

Recap: Continuity

The Prehistory of Limits

The derivative

Maxima and minima

Continuity - the definition

Definition: If $f : [a, b] \to \mathbb{R}$ is a function and $c \in [a, b]$, then f is said to be continuous at the point c if and only if

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

Thus, if c is one of the end points we require only the left or right hand limit to exist.

A function f on (a, b) (resp. $[a, b]$) is said to be continuous if and only if it is continuous at every point c in (a, b) (resp. $[a, b]$).

If f is not continuous at a point c we say that it is dicontinuous at c, or that c is a point of discontinuity for f.

Intuitively, continuous functions are functions whose graphs can be drawn on a sheet of paper without lifting the pencil of the sheet of paper. That is, there should be no "jumps" in the graph of the function.

The basic properties of continuous functions

The sums, differences, products and quotients of continuous functions are continuous (in the last case, the value of the continuous function in the denominator should be non-zero).

The composition of continuous functions is continuous.
Theorem 9: Suppose $f : [a, b] \to \mathbb{R}$ is a continuous function. For every u between $f(a)$ and $f(b)$ there exists $c \in [a, b]$ there such that $f(c) = u$.

Functions which have this property are said to have the Intermediate Value Property (IVP).

We will not be proving this property - it is a consequence of the completeness of the real numbers. Intuitively, this is clear. Since one can draw the graph of the function without lifting one's pencil off the sheet of paper, the pencil must cut every line $y = e$ with e between $f(a)$ and $f(b)$.

The IVT in a picture

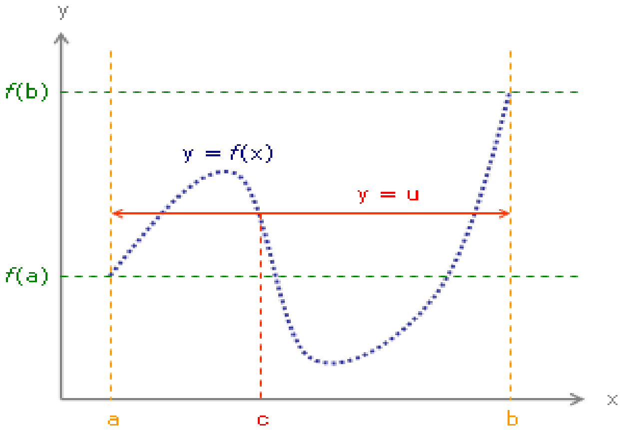


Image created by Enoch Lau see
http://en.wikipedia.org/wiki/File:Intermediatevaluetheorem.png
(Creative Commons Attribution-Share Alike 3.0 Unported license).

Continuous functions on closed, bounded intervals

The other major result on continuous functions that we need is the following. A closed bounded interval is one of the form $[a, b]$, where $-\infty < a$ and $b < \infty$.

Theorem 11: A continuous function on a closed bounded interval $[a, b]$ is bounded and attains its infimum and supremum, that is, there are points x_1 and x_2 in $[a, b]$ such that $f(x_1) = m$ and $f(x_2) = M$, where m and M denote the infimum and supremum respectively.

We defined infimum and supremum for sequences previously. The definition for functions of a real variable is the same: Let $X \subset \mathbb{R}$ and let $f : X \to \mathbb{R}$ be a function. A real number M is called the supremum of $f(x)$ (on X) if
1. If $f(x) \leq M$ for all $x \in X$.
2. If for some real number M_1 $f(x) \leq M_1$ for all $x \in X$, then $M \leq M_1$.

Relaxing the conditions

Again, we will not prove Theorem 11, but will use it quite often. Note the contrast with open intervals. The function $1/x$ on $(0, 1)$ does not attain a maximum - in fact it is unbounded. Similarly the function $1/x$ on $(1, \infty)$ does not attain a minimum, although, it is bounded below and the infimum is 0.

Exercise 5: In light of the above theorem, can you find a continuous function $g : (a, b) \to \mathbb{R}$ for part (i) of Exercise 1.11, with $c \in (a, b)$ such that $\lim_{x\to c} f(x)g(x) \neq 0$?
Let us look at Exercise 1.13 part (i).
Consider the function defined as $f(x) = \sin \frac{1}{x}$ when $x \neq 0$, and $f(0) = 0$. Is this function is continuous at $x = 0$.

Let us look at the sequence of points $x_n = 2/(2n + 1)\pi$. Clearly $x_n \to 0$ as $n \to \infty$ and $|f(x_n)| = 1$

You can easily check that $f(x)$ has the IVP.

Sequential continuity

The preceding example showed that in order to demonstrate that a function $f(x)$ is not continuous at a point x_0 it is enough to find a sequence x_n tending to x_0 such that the value of the function $|f(x_n) - f(x_0)|$ remains large. Suppose it is not possible to find such a sequence. Does that mean the function is continuous at x_0? Yes.

Theorem 12: A function $f(x)$ is continuous at a point a if and only if for every sequence $x_n \to a$, $\lim_{x_n \to a} f(x_n) = f(a)$.

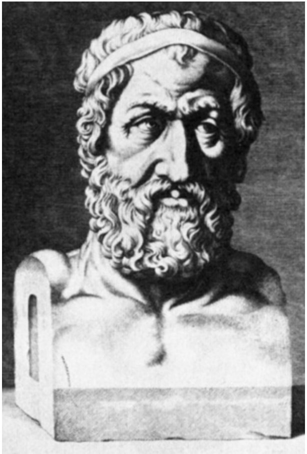
A function that satisfies the property that for every sequence $x_n \to a$, $\lim_{x_n \to a} f(x_n) = f(a)$ is said to be sequentially continuous. The theorem says that sequential continuity and continuity are the same thing. Indeed, it is clear that a continuous function is necessarily sequentially continuous. It is the reverse that is slightly harder to prove.

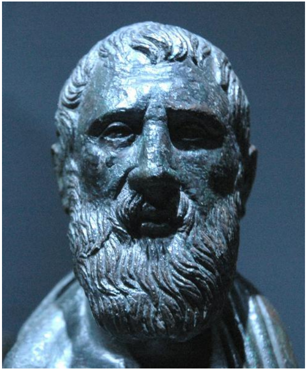
The first man to think about limits?

Zeno of Elea (490 - 460 BCE) was a famous Greek philosopher (source: Wikipedia)

Zeno of Elea

First let us record that we have no idea what Zeno looked like. The picture above was painted in the period 1588 - 1594 CE in Spain, about two thousand years after Zeno's time. Here are two more images of Zeno (also from Wikipedia)





Zeno's Paradoxes

I couldn't find out where the first statue came from and when it was made. The second seems to have come from Herculaneum in Italy (incidentally, Elea (modern Vilia) is a town in Italy). Now Herculaneum was destroyed by a volcanic erruption from the nearby volcano Vesuvius in 79 CE, so it looks like the bust was created within 500 years of Zeno's death. Maybe it was even made during his lifetime. Unfortunately, it is not clear whether this statue is one of Zeno of Elea or of another Zeno (of Citium) who lived about 150 years later.

The important about Zeno is that it would appear that he was the first human to think about limits and limiting processes, at least in recorded history. Most of what we know about him is through his paradoxes, nine of which survive in the works of another famous Greek philosopher Aristotle (384 - 322 BCE) , the official guru of Alexander the Great (aka Sikander in India).

Achilles and the tortoise

One of Zeno's motivations for stating his paradoxes seems to have been to defend his own guru Parmenides' philosophy (whatever that was). Anyway here is his most famous paradox as recorded by Aristotle.

Achilles and the tortoise:
In a race, the quickest runner can never overtake the slowest, since the pursuer must first reach the point whence the pursued started, so that the slower must always hold a lead.

Aristotle, Physics VI:9, 239b15

General knowledge question: Who was Achilles?

Zeno's paradox animated

Achilles and the Tortoise

A gateway to infinite series

Nowadays, this line of argument does not really bother us, since we understand that an infinite number of terms (in this case consisting of the time travelled in each segment or the distance travelled in each segment) can add up to something finite. Nevertheless there are other philosophical issues that continued to bother mathematicians and physicists for a long time. After all, this kind of discussion does lead us to question whether intervals of time and space can be infinitely subdivided, or if "instantaneous motion" makes sense.

Since we are learning mathematics, we won't speculate on physics or philosophy, but we note that Zeno's argument gives a good way to derive the sum of an infinite geometric series. The geometric series is one of the simplest examples of infinite series, so let us see how this is done.

Geometric series - the formula

Let us suppose that the speed of achilles is v and that the speed of the tortoise is rv for some $0 < r < 1$. We will assume that the tortoise was given a headstart of distance "a".

- The distance covered by Achilles in time t is vt.
- The distance covered by the tortoise in time t is rvt.
- Achilles catches up with the tortoise when $vt = a + rvt$, that is, at time $t = a/(v - rv)$ and when the total distance covered by Achilles is $vt = a/(1 - r)$.

On the other hand,

- Distance covered by the tortoise by the time Achilles has covered distance a is ar.
- Distance covered by the tortoise by the time Achilles has covered distance ar is ar^2 ....
- Total distance covered by Achilles when he has caught up with the tortoise is $a + ar + ar^2 + \dots$.
- Thus we get $a + ar + ar^2 + \cdots = a/(1 - r)$.

The defintion

For now, if you did not understand the rigourous definition of the limit, forget about it. You will be able to understand what follows as long as you remember your 11th standard treatment of limits. Recall that $f : (a, b) \to \mathbb{R}$ is said to be differentiable at a point $c \in (a, b)$ if

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

exists. In this case the value of the limit is denoted $f'(c)$ and is called the derivative of f at c. The derivative may also be denoted by $\frac{df}{dx}(c)$ or by $\frac{dy}{dx}|_c$, where $y = f(x)$.

In general, the derivative measures the rate of change of a function at a given point. Thus, if the function we are studying is the position of a particle on the x-coordinate, then $x'(t)$ is the velocity of the particle. If the function we are studying is the velocity $v(t)$ of the particle, then the derivative $v'(t)$ is the acceleration of the particle. If the function we are studying is the population of India, then the derivative measures the rate of change of the population.

The slope of the tangent

From the point of view of geometry, the derivative $f'(c)$ gives us the slope of the curve, that is, the slope of the tangent to the curve $y = f(x)$ at $(c, f(c))$. This becomes particularly clear if we rewrite the derivative as the following limit:

$$\lim_{y \to c} \frac{f(y) - f(c)}{y - c}.$$

The expression inside the limit obviously represents the slope of a line passing through $(c, f(c))$ and $(y, f(y))$, and as y approaches c this line obviously becomes tangent to $y = f(x)$ at the point $(c, f(c))$.

Another way of thinking of the derivative

Another way of thinking of the derivative of the function f at the point x_0 is as follows. If f is differentiable at x_0 we know that

$$\frac{f(x_0+h)-f(x_0)}{h} - f'(x_0) \to 0$$

as $h \to 0$. Since we are keeping x_0 fixed, we can treat the above quantity as a function of h. Thus we can write

$$\frac{f(x_0+h)-f(x)}{h} - f'(x_0) = o_1(h)$$

for some function $o_1(h)$ with the property that $o(h) \to 0$ as $h \to 0$. Taking a common denominator,

$$\frac{f(x_0+h)-f(x_0)-f'(x_0)h}{h} = o_1(h) \tag{1}$$

We can use the above equality to give an equivalent definition for the derivative. A function f is said to be differentiable at the point x_0 if there exists a real number (denoted $f'(x_0)$) such that (1) holds for some function $o_1(h)$ such that $o_1(h) \to 0$ as $h \to 0$.

The derivative as a linear map

We can rewrite equation (1) as

$$f(x_0 + h) = f(x_0) + f'(x_0)h + o_1(h)h$$

Thus, the derivative of $f(x)$ at a point x_0 can be viewed as that real number (if it exists) by which you have to multiply h so that the resulting remainder goes to 0 faster than h (that is, the remainder divided by h goes to 0 as h goes to 0).

A function $f : \mathbb{R} \to \mathbb{R}$ which has the property that $f(x + y) = f(x) + f(y)$ is called a linear function (or linear map). All such functions are given by multiplication by a real number, that is, every linear function has the form $f(x) = \lambda x$, for some real number λ. Thus the derivative may be regarded as a linear function (in the variable h). This point of view will be particularly useful in multivariable calculus.

Examples

Exercise 1.16: Let $f : (a, b) \to \mathbb{R}$ be a function such that

$$|f(x+h) - f(x)| \leq C|h|^{\alpha}$$

for all $x, x+h \in (a, b)$, where C is a constant and $\alpha > 1$. Show that f is differentiable on (a, b) and compute $f^{'}(x)$ for $x \in (a, b)$.

Solution:

$$\left| \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \right| \leq C \lim_{h \to 0} |h|^{\alpha - 1} = 0.$$

Note: Functions that satisfy the property above for some α (not necessarily greater than 1) are said to be Lipschitz continuous with exponent α.

Calculating derivatives

As with limits all of you are already familiar with the rule for calculating the sums, differences, products and quotients of derivatives. You should try and remember how to prove these.

You should also recall the chain rule for calculating the derivative of the composition of functions and try to prove it as an exercise using the $\epsilon - \delta$ definition of a limit.

Note that the proof of the chain rule given in some books involves writing

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

and then taking limits as $\Delta x \to 0$. This is not quite correct since Δu could be 0 even for infinitely many values of u.

Maxima and minima

Let $X \subset \mathbb{R}$ and let $f : X \to \mathbb{R}$ be a function (you can think of X as an open, closed or half-open interval, for instance).

Definition: The function f is said to attain a maximum (resp. minimum) at a point $x_0 \in X$ if $f(x) \leq f(x_0)$ (resp. $f(x) \geq f(x_0)$) for all $x \in X$.

Once again, I remind you that, in general, f may not attain a maximum or minimum at all on the set X. The standard example being $X = (0, 1)$ and $f(x) = 1/x$ (can you find an example on the closed interval $[0, 1]$?). However, if X is a closed bounded interval and f is a continuous function, Theorem 11 tells us that the maximum and minimum are actually attained. Theorem 11 is sometimes called the Extreme Value Theorem.

Maxima and minima and the derivative

If f has a maximum at the point x_0 and if it also differentiable at x_0, we can reason as follows. We know that $f(x_0 + h) - f(x_0) \leq 0$ for every $h > 0$ such that $x + h \in X$. Hence, we see that (one half of the Sandwich Theorem!)

$$\lim_{h \to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} \leq 0.$$

On the other hand, when $h < 0$, we get

$$\lim_{h \to 0^-} \frac{f(x_0 + h) - f(x_0)}{h} \geq 0.$$

Because f is assumed to be differentiable at x_0 we know that left and right hand limits must be equal. It follows that we must have $f'(x_0) = 0$. A similar argument shows that $f'(x_0) = 0$ if f has a minimum at the point x_0.

Local maxima and minima

The preceeding argument is purely local. Before explaining what this means, we give the following defintion.

Definition: Let $f : X \to \mathbb{R}$ be a function and x_0 be in X. Suppose there is an sub-interval $x_0 \in (c, d) \subset X$ such that $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $x \in (c, d)$, then f is said to have a local maximum (resp. local minimum) at x_0.

Sometimes we use the terms global maximum or global minimum instead of just maximum or minimum in order to emphasize the points are not just local maxima or minima. The argument of the previous slide actually proves the following

Theorem 13: If $f : X \to \mathbb{R}$ is differentiable and has a local minimum or maximum at a point $x_0 \in X$, $f'(x_0) = 0$.

MA 105 D3 Lecture 6

Ravi Raghunathan

Department of Mathematics

August 21, 2023

Recap

Maxima and minima

Properties of differentiable functions

The second derivative

Maxima and minima

Let $X \subset \mathbb{R}$ and let $f : X \to \mathbb{R}$ be a function (you can think of X as an open, closed or half-open interval, for instance).

Definition: The function f is said to attain a maximum (resp. minimum) at a point $x_0 \in X$ if $f(x) \leq f(x_0)$ (resp. $f(x) \geq f(x_0)$) for all $x \in X$.

If X is a closed bounded interval and f is a continuous function, the Extreme Value Theorem tells us that the maximum and minimum are actually attained.

Definition: Let $f : X \to \mathbb{R}$ be a function and x_0 be in X. Suppose there is an sub-interval $x_0 \in (c, d) \subset X$ such that $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $x \in (c, d)$, then f is said to have a local maximum (resp. local minimum) at x_0.

Theorem 13: (Fermat's Theorem) If $f : X \to \mathbb{R}$ is differentiable and has a local minimum or maximum at a point $x_0 \in X$, $f'(x_0) = 0$.

Rolle's Theorem

Theorem 13 is known as Fermat's theorem. It can be used to prove Rolle's Theorem.

Theorem 14: Suppose $f : [a, b] \to \mathbb{R}$ is a continuous function which is differentiable in (a, b) and $f(a) = f(b)$. Then there is a point x_0 in (a, b) such that $f'(x_0) = 0$.

Proof: Since f is a continuous function on a closed bounded interval Theorem 11 tells us that f must attain its minimum and maximum somewhere in $[a, b]$. If both the minimum and maximum are attained at the end points, f must be the constant function, in which case we know that $f'(x) = 0$ for all $x \in (a, b)$. Hence, we can assume that at least one of the minimum or maximum is attained at an interior point x_0 and Theorem 13 shows that $f'(x_0) = 0$ in this case. □

One easy consequence: If $P(x)$ is a polynomial of degree n with n real roots, then all the roots of $P'(x)$ are also real. (How do we know that polynomials are differentiable?)

Problems centered around Rolle's Theorem

Exercise 2.3: Let $f : [a, b] \to \mathbb{R}$ be continuous and suppose f is differentiable on (a, b). If $f(a)$ and $f(b)$ are of opposite signs and $f'(x) \neq 0$ for all $x \in (a, b)$, then there is a unique point x_0 in (a, b) such that $f(x_0) = 0$. Solution: Since the

Intermediate Value Theorem guarantees the existence of a point x_0 such that $f(x_0) = 0$, the real point of this exercise is the uniqueness. Suppose there were two points $x_1, x_2 \in (a, b)$

such that $f(x_1) = f(x_2) = 0$. Applying Rolle's Theorem, we see that there would exist $c \in (x_1, x_2)$ such that $f'(c) = 0$ contradicting our hypothesis, This proves the exercise. Let us

look at Exercise 2.8(i): Find a function f which satisfies all the given conditions, or else show that no such function exists: $f''(x) > 0$ for all $x \in \mathbb{R}$ and $f'(0) = 1, f'(1) = 1$. Solution:

Apply Rolle's Theorem to $f'(x)$ to conclude that such a function cannot exist.

The Mean Value Theorem

Rolle's theorem is a special case of the Mean Value Theorem (MVT).

Theorem 15: Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function and that f is differentiable in (a, b). Then there is a point x_0 in (a, b) such that

$$\frac{f(b) - f(a)}{b - a} = f'(x_0).$$

$\square$

Proof: Apply Rolle's Theorem to the function

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

(Why does one think of the function $g(x)$?)

Applications of the MVT

Here is an application of the MVT which you have probably always taken for granted:

Theorem 16: If f satisfies the hypotheses of the MVT, and further $f'(x) = 0$ for every $x \in (a, b)$, f is a constant function.

Indeed, if $f(c) \neq f(d)$ for some two points $c < d$ in $[a, b]$,

$$0 \neq \frac{f(d) - f(c)}{d - c} = f'(x_0),$$

for some $x_0 \in (c, d)$, by the MVT. This contradicts the hypothesis. $\square$

Applications of the MVT continued

Consider Exercise 2.6.:

Let f be continuous on $[a, b]$ and differentiable on (a,b). If $f(a) = a$ and $f(b) = b$, show that there exist distinct $c_1, c_2 \in (a, b)$ such that $f'(c_1) + f'(c_2) = 2$.

Solution: The idea is that the function clearly has an average rate of growth equal to 1 on the interval $[a, b]$. If the derivative at some point is less than 1, there must be another point where it is greater than 1 so that the sum adds up to 2. How to use this idea?

Split the interval into two pieces - $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$ - and apply the MVT to each interval.

Darboux's Theorem

Another interesting property of differentiable functions is that their derivatives have the IVP (intermediate value property). This fact is sometimes called Darboux's Theorem.

Theorem 17: Let $f : (a, b) \to \mathbb{R}$ be a differentiable function. If c, d, $c < d$ are points in (a, b), then for every u between $f'(c)$ and $f'(d)$, there exists an x in $[c, d]$ such that $f'(x) = u$.

Proof: We can assume, without loss of generality, that $f'(c) < u < f'(d)$, otherwise we can take $x = c$ or $x = d$. Define $g(t) = ut - f(t)$. This is a continuous function on $[c, d]$, and hence, by Theorem 11 must attain its extreme values. The maximum of g cannot occur at c or d since $g'(c) = u - f'(c) > 0$ and $g'(d) = u - f'(d) < 0$, which means that g is increasing at c, and decreasing at d. It follows that there exists $x \in (c, d)$ where g takes a maximum. By Fermat's Theorem $g'(x) = 0$ which yields $f'(x) = u$. $\square$

Continuity of the first derivative

We have just seen that the derivative satisfies the IVP. Can we find a function which is differentiable but for which the derivative is not continuous?

Here is the standard example:

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & \text{if } x \neq 0 \\ = 0 & \text{if } x = 0. \end{cases}$$

This function will be differentiable at 0 but its derivative will not be continuous at that point. In order to see this you will need to study the function in Exercise 1.13(ii). This will show that $f'(0) = 0$.

On the other hand, if we use the product rule when $x \neq 0$ we get

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

which does not go to 0 as $x \to 0$.

Back to maxima and minima

We will assume that $f : [a, b] \to \mathbb{R}$ is a continuous function and that f is differentiable on (a, b). A point x_0 in (a, b) such that $f'(x_0) = 0$ often called a stationary point. We will assume further that $f'(x)$ is differentiable at x_0, that is, that the second derivative $f''(x_0)$ exists. We formulate the Second Derivative Test below.

Theorem 18: With the assumptions above:

1. If $f''(x_0) > 0$, the function has a local minimum at x_0.
2. If $f''(x_0) < 0$, the function has a local maximum at x_0.
3. If $f''(x_0) = 0$, no conclusion can be drawn.

The proof of the Second Derivative Test

Proof: The proofs are straightforward. For instance, to prove the first part we observe that

$$0 < f''(x_0) = \lim_{h \to 0} \frac{f'(x_0 + h)}{h}.$$

It follows that for $|h|$ small enough, $f'(x_0 + h) < 0$, if $h < 0$ and $f'(x_0 + h) > 0$ if $h > 0$. It follows that $f(x_0)$ is decreasing to the left of x_0 and increasing to the right of x_0. Hence, x_0 must be a local minimum. A similar argument yields the second case. □

If the third case of the theorem above occurs, the function may be changing from concave to convex. In this case x_0 is called a point of inflection. An example of this phenomenon is given by $f(x) = x^3$ at $x = 0$.

Concavity and convexity

Let I denote an interval (open or closed or half-open).
Definition: A function $f : I \to \mathbb{R}$ is said to be concave (or sometimes concave downwards) if

$$f(tx_1 + (1-t)x_2) \geq tf(x_1) + (1-t)f(x_2)$$

for all x_1 and x_2 in I and $t \in [0, 1]$.
Similarly, a function is said to be convex (or concave upwards) if

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

By replacing the $\geq$ and $\leq$ signs above by strict inequalities we can define strictly concave and strictly convex functions.

Note that if $f(x)$ is a concave function, $-f(x)$ is a convex function, so it is really enough to study one class or the other. Convex functions occur in many areas of mathematics.

Examples of concave and convex functions

Here are some examples of convex functions.

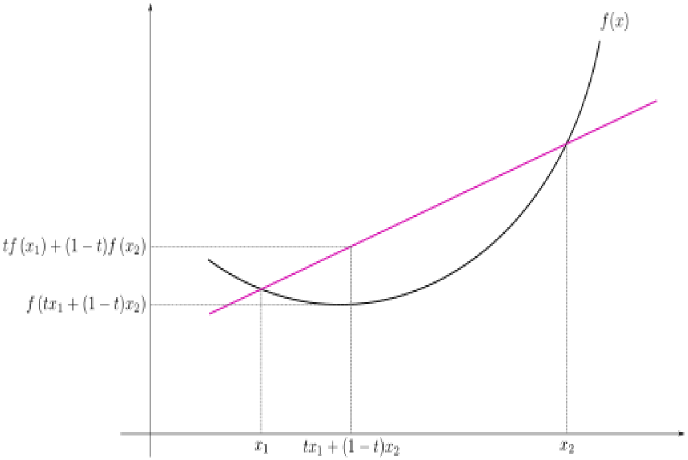
1. $f(x) = x^2$ on $\mathbb{R}$.
2. $f(x) = x^3$ on $[0, \infty)$.
3. $f(x) = e^x$ on $\mathbb{R}$.

Examples of concave functions include

1. $f(x) = -x^2$
2. $f(x) = x^3$ on $(-\infty, 0]$
3. $f(x) = \log x$ on $(0, \infty)$.

For a convex function f and point $c \in (x_1, x_2)$, the point $(c, f(c))$ always lies below the line joining $(x_1, f(x_1))$ and $(x_2, f(x_2))$.

Convexity illustrated graphically



Propeties of Convex functions

Convex functions have many nice properties. For instance, it is easy to show that convex functions are continuous (do this!). More is true.

Exercise 1. Every convex function is Lipschitz continuous (a function is Lipschitz continuous if it satisfies the inequality given in Exercise 1.16 but with $\alpha = 1$). In fact, much more is true. A convex function is actually differentiable at all but at most countably many points.

A differentiable function is convex if and only if its derivative is monotonically increasing. Moreover, if a function is both differentiable and convex, it is continuously differentiable, that is, its derivative is continuous (feel free to try proving these facts).

Convexity and the second derivative (not yet covered in class)

A twice differentiable function on an interval will be convex if its second derivative is everywhere non-negative. If the second derivative is positive, the function will be strictly convex.

However, the converse of the second statement above is not true. Can you give a counter-example to the converse of the second statement?

How about $f(x) = x^4$?

Definition: A point of inflection x_0 for a function f is a point where the function changes its behavior from concave to convex (or vice-versa). At such a point $f''(x_0) = 0$, but this is only a necessary, not a sufficient condition. (Why?)
If further, we also assume that the lowest order (≥ 2) non-zero derivative is odd, then we get a sufficient condition.

MA 105 D3 Lecture 7

Ravi Raghunathan

Department of Mathematics

August 22, 2023

Recap

Towards Taylor's Theorem - higher derivatives

Arnold's problem

Concavity and convexity

Let I denote an interval (open or closed or half-open).
Definition: A function $f : I \to \mathbb{R}$ is said to be concave (or sometimes concave downwards) if

$$f(tx_1 + (1-t)x_2) \geq tf(x_1) + (1-t)f(x_2)$$

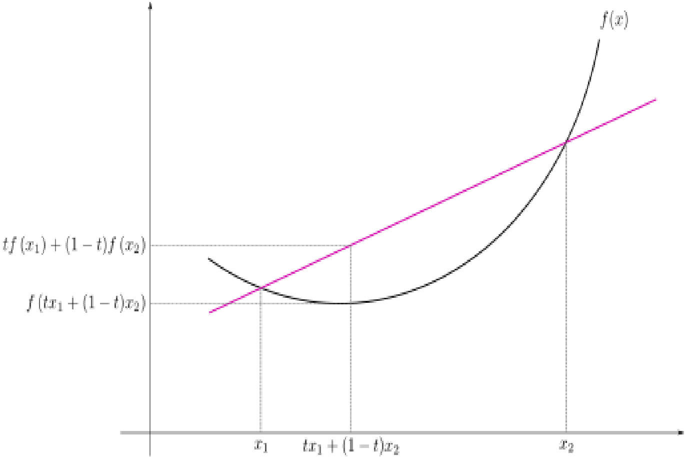
for all x_1 and x_2 in I and $t \in [0,1]$.
Similarly, a function is said to be convex (or concave upwards) if

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

By replacing the $\geq$ and $\leq$ signs above by strict inequalities we can define strictly concave and strictly convex functions.

Note that if $f(x)$ is a concave function, $-f(x)$ is a convex function, so it is really enough to study one class or the other. Convex functions occur in many areas of mathematics.

Convexity illustrated graphically



Convexity and the second derivative

A twice differentiable function on an interval will be convex if its second derivative is everywhere non-negative. If the second derivative is positive, the function will be strictly convex.

However, the converse of the second statement above is not true. Can you give a counter-example to the converse of the second statement?

How about $f(x) = x^4$?

Definition: A point of inflection x_0 for a function f is a point where the function changes its behavior from concave to convex (or vice-versa). At such a point $f''(x_0) = 0$, but this is only a necessary, not a sufficient condition. (Why?)
If further, we also assume that the lowest order (≥ 2) non-zero derivative is odd, then we get a sufficient condition.

Smooth functions

We will now introduce some notation. The space $\mathcal{C}^k(I)$, will denote the space of k times continuously differentiable functions on an (open) interval I, for some fixed $k \in \mathbb{N}$, that is, the space of functions for which k derivatives exist and such that the k-th derivative is a continuous functions.

The space $\mathcal{C}^\infty(I)$ will consist of functions that lie in $\mathcal{C}^k(I)$ for every $k \in \mathbb{N}$. Such functions are called smooth or infinitely differentiable functions.

From now on we will denote the k-th derivative of a function $f(x)$ by $f^{(k)}(x)$.

Our aim will be to enlarge the class of functions we understand using the polynomials as stepping stones.

The Taylor polynomials

Given a function $f(x)$ which is n times differentiable at some point x_0 in an interval I, we can associate to it a family of polynomials $P_0(x), P_1(x), \ldots P_n(x)$ called the Taylor polynomials of degrees $0, 1, \ldots n$ at x_0 as follows.

We let $P_0(x) = f(x_0)$,

$$P_1(x) = f(x_0) + f^{(1)}(x_0)(x - x_0),$$

$$P_2(x) = f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2!}(x - x_0)^2$$

We can continue in this way to define

$$P_n(x) = f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

Taylor's Theorem

The Taylor polynomials are rigged exactly so that the degree n Talyor polynomial has the same first n derivatives at the point x_0 as the function $f(x)$ has, that is, $P^{(k)}(x_0) = f^{(k)}(x_0)$ for all $0 \leq k \leq n$, where $f^{(0)} = f(x)$ by convention.

Taylor's Theorem says that we can recover a lot of information about the function from the Taylor polynomials.

Theorem 19: Let I be an open interval and suppose that $[a, b] \subset I$. Suppose that $f \in \mathcal{C}^n(I)$ $(n \geq 0)$ and suppose that $f^{(n)}$ is differentiable on I. Then there exists $c \in (a, b)$ such that

$$f(b) = P_n(b) + \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1},$$

where $P_n(x)$ denotes the Taylor polynomial of degree n at a.

Remarks on Taylor's Theorem and some examples

Remark 1: When $n = 0$ in Taylor's Theorem we get the MVT. When $n = 1$, Taylor's Theorem is called the Extended Mean Value Theorem.

Remark 2: The Taylor polynomials are nothing but the partial sums of the Taylor Series associated to a $\mathcal{C}^\infty$ function about (or at) the point a:

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(b-a)^k.$$

We can show that this series converges provided we know that the difference $f(x) - P_n(x) = R_n(x)$ can be made less than any $\epsilon > 0$ when n is sufficiently large. We will see how to do this for certain simple functions like e^x or $\sin x$.

The Taylor series for e^x

Let us show that the Taylor series for the function e^x about the point 0 is a convergent series for any value of $x = b \geq 0$ and that it converges to the value e^b (a similar proof works for $b < 0$).

In this case, at any point a, $f^{(n)}(a) = e^a$, so at $a = 0$ we obtain $f^{(n)}(0) = 1$. Hence the series about 0 is

$$\sum_{k=0}^{\infty} \frac{b^k}{k!}.$$

If we look at $R_n(b) = e^b - s_n(b)$ we obtain

$$|R_n(b)| = \frac{e^c b^{n+1}}{(n+1)!} \leq \frac{e^b b^{n+1}}{(n+1)!},$$

since $c \leq b$. As $n \to \infty$ this clearly goes to 0. This shows that the Taylor series of e^b converges to the value of the function at each real number b.

Defining functions using Taylor series

Instead of finding the Taylor series of a given function we can reverse the process and define functions using convergent series. Thus, one can define the function e^x as

$$e^x := \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

In this case, we have to first show that the series on the right hand side converges for a given value of x, in which case the definition above makes sense.

We show the convergence of such series by showing that they are Cauchy series. This means that we do not have to guess at a value of the limit.

Power series

As we have explained in the previous slide the "correct" (both from the point of view of proofs and of computation) way to define a function like e^x is via convergent series involving non-negative integer powers of x. Such series are called power series and such functions should be viewed as the natural generalizations of polynomials.

The nice thing about power series is that once we know that they converge in some interval $(a-r, a+r)$ around a, it is not hard to show that the functions that they define are continuous functions. In fact, it is not too hard to show that they are smooth functions (that is, that all their derivatives exist). Thus when functions are given by convergent power series, we can automatically conclude they are smooth. This is the advantage of defining functions in this way.

Calculating the values of functions

As we have also mentioned several times, calculators and computers calculate the values of various common functions like trigonometric polynomials and expressions in $\log x$ and e^x by using Taylor series.

The great advantage of Taylor series is that one can estimate the error since we have a simple formula for the error which can be easily estimated. For instance, for the function $\sin x$, the n-th derivative is either $\pm \sin x$ or $\pm \cos x$, so in either case $|f^{(n}(x)| \leq 1$. Hence,

$$|R_n(x)| \leq \frac{|x|^{n+1}}{(n+1)!}$$

If we take $x = 1$, and we want to compute $\sin 1$ to an error of less than 10^{-16}, we need only make sure that $(n+1)! > 10^{16}$, which is achieved when $n \geq 21$. (Can you find a value of n which works for any value of x?)

Arnold's problem

Recall that you were asked to find the following limit (Exercise 3 of the previous lecture).

$$\lim_{x\to 0} \frac{\sin(\tan x) - \tan(\sin x)}{\arcsin(\arctan x) - \arctan(\arcsin x)}.$$

The problem above was posed by the Russian mathematician Vladimir I. Arnold (see his book "Huygens and Barrow, Newton and Hooke") as an example of a problem that seventeenth century mathematicians could solve very easily, but that modern mathematicians, even with all their extra machinery and knowledge can't. In fact, it was his habit to put up this problem while lecturing to eminent mathematicians in leading universities and challenge them to solve it within ten minutes.

Notice that the limit that we have to calculate has the form

$$\frac{f(x) - g(x)}{g^{-1}(x) - f^{-1}(x)}.$$

The solution to Arnold's problem

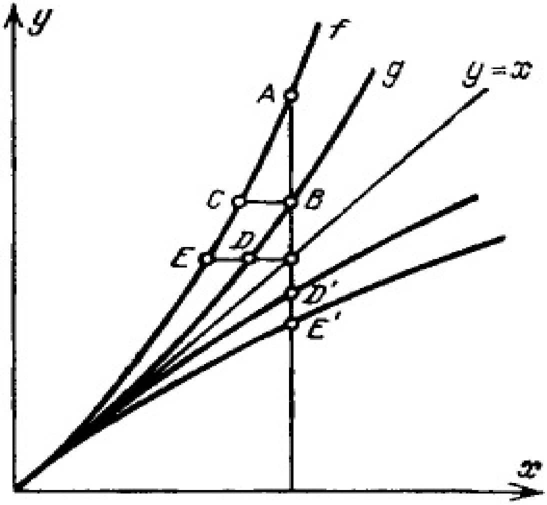


Fig. 37.
Calculation of the limit |AB|/|D′E′|

V. I. Arnold

The preceeding picture was taken from V. I. Arnold's book "Huygens and Barrow, Newton and Hooke (Birkhauser 1990)

V. I. Arnold (1936-2008) worked in geometry, differential equations and mathematical physics. He was one of the most important mathematicians of the twentieth century.

MA 105 D3 Lecture 8

Ravi Raghunathan

Department of Mathematics

August 24, 2023

The proof of Taylor's theorem

A very explicit calculation

Analytic functions

The Taylor polynomials

Given a function $f(x)$ which is n times differentiable at some point x_0 in an interval I, we can associate to it a family of polynomials $P_0(x), P_1(x), \ldots P_n(x)$ called the Taylor polynomials of degrees $0, 1, \ldots n$ at x_0 as follows.

We let $P_0(x) = f(x_0)$,

$$P_1(x) = f(x_0) + f^{(1)}(x_0)(x - x_0),$$

$$P_2(x) = f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2!}(x - x_0)^2$$

We can continue in this way to define

$$P_n(x) = f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

Taylor's Theorem

The Taylor polynomials are rigged exactly so that the degree n Talyor polynomial has the same first n derivatives at the point x_0 as the function $f(x)$ has, that is, $P^{(k)}(x_0) = f^{(k)}(x_0)$ for all $0 \leq k \leq n$, where $f^{(0)} = f(x)$ by convention.

Taylor's Theorem says that we can recover a lot of information about the function from the Taylor polynomials.

Theorem 19: Let I be an open interval and suppose that $[a, b] \subset I$. Suppose that $f \in \mathcal{C}^n(I)$ $(n \geq 0)$ and suppose that $f^{(n)}$ is differentiable on I. Then there exists $c \in (a, b)$ such that

$$f(b) = P_n(b) + \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1},$$

where $P_n(x)$ denotes the Taylor polynomial of degree n at a.

The proof of Taylor's theorem

Proof: From the definition, we see that

$$P_n(b) = f(a) + f^{(1)}(a)(b-a) + \cdots \frac{f^{(n}(a)}{n!}(b-a)^n$$

Consider the function

$$F(x) = f(b) - f(x) - f^{(1)}(x)(b-x) - \frac{f^{(2)}(x)}{2!}(b-x)^2 - \ldots - \frac{f^{(n)}(x)}{n!}(b-x)^n.$$

Clearly $F(b) = 0$, and

$$F^{(1)}(x) = -\frac{f^{(n+1)}(x)(b-x)^n}{n!}. \tag{1}$$

We would like to apply Rolle's Theorem here, but $F(a) \neq 0$. So consider

$$g(x) = F(x) - \left(\frac{b-x}{b-a}\right)^{n+1} F(a)$$

(this is similar to the method by which we reduced the MVT to Rolle's Theorem), and we see that $g(a) = 0$. Applying Rolle's Theorem we see that there is a $c \in (a, b)$ such that $g'(c) = 0$.

This yields

$$F^{(1)}(c) = -(n+1)\left[\frac{(b-c)^n}{(b-a)^{n+1}}\right]F(a). \tag{2}$$

We can eliminate $F^{(1)}(c)$ using (1). This gives

$$-(n+1)\left[\frac{(b-c)^n}{(b-a)^{n+1}}\right]F(a) = -\frac{f^{(n+1)}(c)(b-c)^n}{n!},$$

from which we obtain

$$F(a) = \frac{(b-a)^{n+1}}{(n+1)!}f^{(n+1)}(c).$$

This proves what we want. $\square$

Power series

As we have explained in the previous slide the "correct" (both from the point of view of proofs and of computation) way to define a function like e^x is via convergent series involving non-negative integer powers of x. Such series are called power series and such functions should be viewed as the natural generalizations of polynomials.

The nice thing about power series is that once we know that they converge in some interval $(a-r, a+r)$ around a, it is not hard to show that the functions that they define are continuous functions. In fact, it is not too hard to show that they are smooth functions (that is, that all their derivatives exist), and in fact, the derivative is nothing but the power series obtained by differentiating term by term which also converges in $(a-r, a+r)$!

Thus when functions are given by convergent power series, we can automatically conclude that they are smooth.

Calculating the values of functions

As we have also mentioned several times, calculators and computers calculate the values of various common functions like trigonometric polynomials and expressions in $\log x$ and e^x by using Taylor series.

The great advantage of Taylor series is that one can estimate the error since we have a simple formula for the error which can be easily estimated. For instance, for the function $\sin x$, the n-th derivative is either $\pm \sin x$ or $\pm \cos x$, so in either case $|f^{(n}(x)| \leq 1$. Hence,

$$|R_n(x)| \leq \frac{|x|^{n+1}}{(n+1)!}$$

If we take $x = 1$, and we want to compute $\sin 1$ to an error of less than 10^{-16}, we need only make sure that $(n+1)! > 10^{16}$, which is achieved when $n \geq 21$. (Can you find a value of n which works for any value of x?)

Computing the vaues of $\sin x$

First, remember that $\sin x$ is periodic, so we only have to look at the values of x between $-\pi$ and π.

But we can do better, because $\sin(-x) = -\sin x$. So we only have to bother about the interval $[0, \pi]$.

We can do still better! Once we know $\sin x$ in $[0, \pi/2]$, we can easily figure out what it is in $[\pi/2, \pi]$.

So finally, it is enough to find the desired value of n for $x \in [0, \pi/2]$.

Computing the values of $\sin x$

We know that the remainder term satisfies

$$|R_n(x)| \leq \frac{|x|^{n+1}}{(n+1)!}.$$

Hence, we need

$$\frac{|x|^{n+1}}{(n+1)!} < 10^{-16}.$$

We know that it is enough to look at $x \leq \pi/2$. Let us be a little careless and allow $x \leq 2$ (so we won't get the best possible n, maybe).

We already know that $1/(n+1)! < 10^{-16}$ if $n \geq 21$. Now $|x|^{22} \leq 2^{22}$. If we take $n = 31$, we see that $|x|^{32} \leq 2^{22} \cdot 2^{10}$,

$$1/(n+1)! = 1/32! < 10^{-16} \cdot 10^{-10} \cdot 2^{-10}.$$

Smooth functions and Taylor series

Given a smooth function $f(x)$ on $a \in I \subset \mathbb{R}$ we can write down its associated Taylor polynomials $P_n(x)$ around any point a in $\mathbb{R}$. Here are some natural questions that arise. Let us take $a = 0$ in what follows.

Question 1. When $x = 0$, obviously $P_n(0) = f(0)$ for all n. Do the Taylor polynomials $P_n(x)$ (around 0, say) always converge as $n \to \infty$ for $x \neq 0, x \in I$? at least for all x in some sub-interval $(c, d) \ni 0$?

Question 2. If $P_n(x)$ converges as $n \to \infty$ does it necessarily converge to $f(x)$?

We will answer the second question.

Smooth but not analytic

The standard example is the function

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ e^{-1/x} & \text{if } x > 0 \end{cases}$$

Notice that $f^{(k)}(0) = 0$ for all $k \geq 0$. Hence $P_n(x) = 0$ for all $n \geq 0$! Hence, $\lim_{n\to\infty} P_n(x) = 0$. Thus the Taylor polynomials $P_n(x)$ around 0 converge to 0 for any $x \in \mathbb{R}$.

But obviously, they do not converge to the value of the function, since $f(x) > 0$ if $x > 0$!

In this case the Taylor series does a very poor job of approximating the function. Indeed, the remainder term $R_n(x) = f(x)$ for all $x > 0$.

Thus, when we use Taylor series to approximate a function in an interval I, we must make sure that $R_n(x) \to 0$ as $n \to \infty$, for all $x \in I$.

Analytic functions

We say that a function $f(x)$ is analytic in an (open(interval I, if for each point $a \in I$, the Taylor polynomials of the function $f(x)$ around a, converge to $f(x)$ in some (possibly smaller) interval containing a. This means that $R_n(x) \to 0$ for all x in some interval $a \in (c, d) \subseteq I$

The functions $\sin x$, $\cos x$ and e^x are analytic on all of $\mathbb{R}$. The function $\tan x$ is analytic in $(-\pi/2, \pi/2)$. The function $\log x$ is analytic in $(0, \infty)$.

The purpose of this discussion is to alert you to the fact that Taylor series may not always do a good job of approximating a given function. You have to prove something before using Taylor series.

Question 1 even more subtle: See
`https://math.stackexchange.com/questions/620290/`
`is-it-possible-for-a-function-to-be-smooth-everywhere-anal`

L'Hôpital's rule

Suppose f and g are $\mathcal{C}^1$ functions in an interval I containing 0. By the MVT, for $x \in I$,

$$f(x) = f(0) + f^{(1)}(c_1)x \quad \text{and} \quad g(x) = g(0) + g^{(1)}(c_2)x$$

for $0 < c_1, c_2 < x$.
If $f(0) = g(0) = 0$,

$$\lim_{x \to 0} f(x)/g(x) = \lim_{x \to 0} f^{(1)}(c_1)x/g^{(1)}(c_2)x = \lim_{x \to 0} f^{(1)}(c_1)/g^{(1)}(c_2).$$

But $f^{(1)}$ and $g^{(1)}$ are continuous functions and as $x \to 0$, $c_1, c_2 \to 0$. Hence,

$$\lim_{x \to 0} f^{(1)}(c_1)/g^{(1)}(c_2) = f^{(1)}(0)/g^{(1)}(0)$$

If the functions are in $\mathcal{C}^n$, and $f^{(k)}(0) = g^{(k)}(0) = 0$ for all $k < n$, we can apply the MVT repeatedly (or we can apply Taylor's theorem directly) to get $f^{(n)}(0)/g^{(n)}(0)$ as the limit.

About the Quiz:

You will need to enter two numbers (A and B) at the top of your paper below your roll number.

1. Fill in the numbers "A" and "B" above as follows:
   If the last digit a of your roll number satisfies $0 \leq a \leq 4$, let $A = a + 5$. If $5 \leq a \leq 9$, let $A = a$.
   If the second-last digit b of your roll number satisfies $0 \leq b \leq 4$, let $B = b + 5$. If $5 \leq b \leq 9$, let $B = b$. Thus $5 \leq A, B \leq 9$.
   Example: Your Roll number is 23B0092. Then $A = 7$ and $B = 9$.

2. **You must use these values of A and B below. Using the wrong value of A or B even in one question may lead to the loss of all marks in this quiz.**

MA 105 Lecture 9

Ravi Raghunathan

Department of Mathematics

August 28

The Darboux integral

A Feynman story

Riemann integration

Partitions

Definition: Given a closed interval $[a, b]$, a partition P of $[a, b]$ is simply a collections of points

$$P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}.$$

We can think of the points of the partition as dividing the original interval $[a, b]$ into sub-intervals $I_j = [x_{j-1}, x_j]$, $1 \leq j \leq n$. Indeed $I = \cup_j I_j$ and if two sub-interval intersect, they have at most one point in common. Hence, the notation "partition".

Definition: A partition $P' = \{a = x'_0 < x'_1 < \ldots < x'_m = b\}$ is said to be a refinement of the partition P if for each $x_i \in P$, there exists an $x'_j \in P'$ such that $x_i = x'_j$.

Intuitively, a refinement P' of a partition P will break some of the sub-intervals in P into smaller sub-intervals. Any two partitions have a common refinement - why?.

Lower and Upper sums

Given a partition $P = \{a = x_0 < x_1 < \ldots < x_{b-1} < x_n = b\}$ and a function $f : [a, b] \to \mathbb{R}$, we define two associated quantities. First we set:

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x) \quad \text{and} \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \quad 1 \leq i \leq n$$

Defintion: We define the Lower sum as

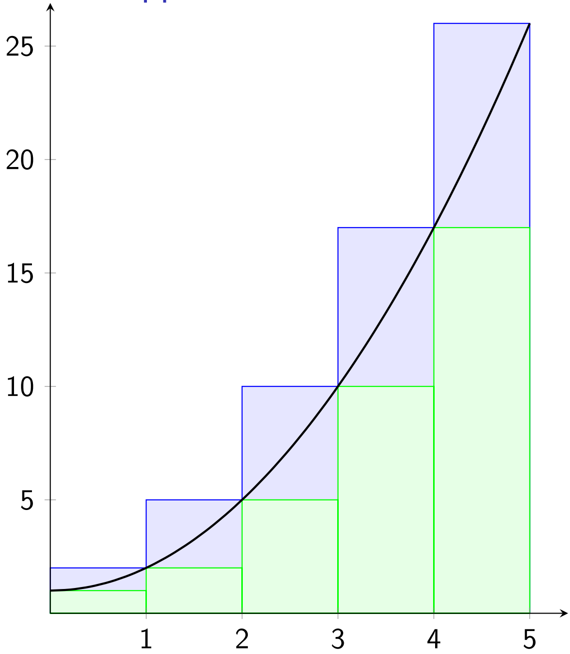
$$L(f, P) = \sum_{j=1}^{n} m_j (x_j - x_{j-1}).$$

Similarly, we can define the Upper sum as

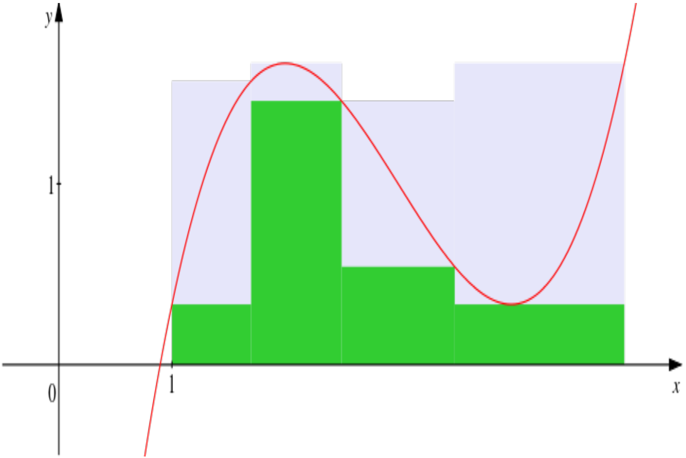
$$U(f, P) = \sum_{j=1}^{n} M_j (x_j - x_{j-1}).$$

In case the words "infimum" and "supremum" bother you, you can think "minimum" and "maximum most of time since we will usually be dealing with continuous functions on $[a, b]$.

Lower and Upper Darboux sums



A picture for a non-monotonic function



https://upload.wikimedia.org/wikipedia/commons/thumb/5/59/Darboux.svg/700px-Darboux.png

One basic example

In order to illustrate what we are saying we will take the following basic example. Let $[a, b] = [0, 1]$ and let $f(x) = x$.

One of the most natural partitions on an interval is a partition that divides the interval into sub-intervals of equal length. For $[0, 1]$, this is

$$P_n = \{0 < 1/n < 2/n < \ldots < (n-1)/n < 1\}.$$

On the interval $I_j = [\frac{j-1}{n}, \frac{j}{n}]$, where does the function $f(x) = x$ take its minimum? its maximum?

Clearly, the minimum $m_j = \frac{j-1}{n}$ is attained at $\frac{j-1}{n}$ and the maximum $M_j = \frac{j}{n}$ at $\frac{j}{n}$. And finally, $\frac{j}{n} - \frac{j-1}{n} = \frac{1}{n}$, for all $1 \leq j \leq n$.

An example of a refinement of P_n is P_{2n}, or, more generally, P_{kn} for any natural number k.

The Darboux integrals

We now define the lower Darboux integral of f by

$$L(f) = \sup\{L(f, P) \colon P \text{ is a partition of } [a, b]\},$$

where the supremum is taken over all partitions of $[a, b]$.

and similarly the upper Darboux integral of f by

$$U(f) = \inf\{U(f, P) \colon P \text{ is a partition of } [a, b]\},$$

and again the infimum is over all partitions of $[a, b]$. (This time there is no escaping inf and sup!)

If $L(f) = U(f)$, then we say that f is Darboux-integrable and define

$$\int_a^b f(t)\, dt := U(f) = L(f).$$

This common value of the two integrals is called the Darboux integral.

Back to the example

Let us calculate $L(f, P_n)$ and $U(f, P_n)$ in the example we gave.

$$L(f, P_n) = \sum_{j=1}^{n} \frac{(j-1)}{n} \cdot \frac{1}{n} = \sum_{j=0}^{n-1} \frac{j}{n^2}.$$

This can be evaluated explicitly:

$$L(f, P_n) = \frac{n(n-1)}{2} \cdot \frac{1}{n^2} = \frac{1}{2} - \frac{1}{2n}.$$

Similary, we can check that

$$U(f, P_n) = \frac{n(n+1)}{2} \cdot \frac{1}{n^2} = \frac{1}{2} + \frac{1}{2n}.$$

Can we conclude that the Darboux integral is $1/2$ by letting $n \to \infty$? Unfortunately, no.

A diversion: How to calculate powers of e in your head?

From Richard Feynman's "Surely you're joking Mr. Feynman!" (pages 173-174):

One day at Princeton I was sitting in the lounge and overheard some mathematicians talking about the series for e to the x power which is $1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$. Each term you get by multiplying the preceding term by x and dividing by the next number. For example, to get the next term after $x^3/3!$ you multiply that term by x and divide by 4. It's very simple.

When I was a kid I was excited by series, and had played with this thing. I had computed e using that series, and had seen how quickly the new terms became very small.

I mumbled something about how it was easy to calculate e to any power using that series (you just substitute the power for x).
"Oh yeah?" they said. "Well, the what's e to the 3.3?" said some joker - I think it was Tukey.
I say, "That's easy. It's 27.11."

Feynman's anecdote continued

Tukey knows it isn't so easy to compute all that in your head.
"Hey! How'd you do that?"
Another guy says, "You know Feynman, he's just faking it. It's not really right."
They go to get a table, and while they're doing that, I put on a few more figures: "27.1126," I say.
They find it in the table. "It's right! But how'd you do it!"
"I just summed the series."
"Nobody can sum the series that fast. You must just happen to know that one. How about e^3?"
"Look," I say. "It's hard work! Only one a day!"
"Hah! It's a fake!" they say, happily.
"All right," I say, "It's 20.085."
They look in the book as I put a few more figures on. They're all excited now, because I got another one right.

Here are these great mathematicians of the day, puzzled at how I can compute e to any power! One of them says, “He just can’t be substituting and summing - it’s too hard. There’s some trick. You couldn’t do just any old number like e to the 1.4.”
I say, “It’s hard work, but for you, OK. It’s 4.05.”
As they’re looking it up, I put on a few more digits and say, “And that’s the last one for the day!” and walk out.

What happened was this: I happened to know three numbers - the logarithm of 10 to the base *e* (needed to convert numbers from base 10 to base *e*), which is 2.3026 (so I knew that *e* to the 2.3 is very close to 10), and because of radioactivity (mean-life and half-life), I knew the log of 2 to the base *e*, which is .69315 (so I also knew that *e* to the .7 is nearly equal to 2). I also knew *e* (to the 1), which is 2.71828.

The first number they gave me was *e* to the 3.3, which is *e* to the 2.3 (10) times *e*, or 27.18. While they were sweating about how I was doing it, I was correcting for the extra .0026 - 2.3026 is a little high.

I knew I couldn't do another one; that was sheer luck. But then the guy said *e* to the 3: that's *e* to the 2.3 times *e* to the .7, or ten times two. So I knew it was 20.something, and while they were worrying how I did it, I adjusted for the .693.

Now I was sure I couldn't do another one, because the last one was again by sheer luck. But the guy said *e* to the 1.4 which is *e* to the .7 times itself. So all I had to do is fix up 4 a little bit!

They never did figure out how I did it.

https://en.wikipedia.org/wiki/Richard_Feynman Richard Feynman (1918-1988)

Useful properties of the Darboux sums

Since, for any partition P, $L(f,P) \leq U(f,P)$, we have

$$L(f) \leq U(f).$$

In fact, for any two partitions P_1 and P_2, we have

$$L(f,P_1) \leq U(f,P_2).$$

This is easy to see - the lower sum computes the sum of the areas of rectangles that lie entirely below the curve while the upper sum computes the sum of the areas of rectangles whose “tops” lie above the curve.

One of the most useful properties of the Darboux sums is the following. If P' is a refinement of P then obviously

$$L(f,P) \leq L(f,P') \leq U(f,P') \leq U(f,P).$$

Riemann Sums

There is another way of getting at the integral due to Riemann which may be a little more intuitive and is better for calculation. This is done via Riemann sums.

To define the notion of a Riemann sum we need one more piece of data. Suppose that for each of the intervals I_j we are given a point $t_j \in I_j$. We will denote the collection of points t_j by t. The pair (P, t) is sometimes called a tagged partition.

Definition: We define the Riemann sum associated to the function f, and the tagged partition (P, t) by

$$R(f, P, t) = \sum_{j=1}^{n} f(t_j)(x_j - x_{j-1}).$$

The norm of a partition

As must be clear, the Lower sum, Upper sum and Riemann sum all give approximations to the area between the lines $x = a$ and $x = b$ and between the curve $y = f(x)$ and the x-axis and

$$L(f, P) \leq R(f, P, t) \leq U(f, P).$$

The point is to make this statement quantitatively precise.

We define the norm of a partition P (denoted $\|P\|$) by

$$\|P\| = \max_j \{|x_j - x_{j-1}|\}, \quad 1 \leq j \leq n.$$

The norm gives some measure of the "size" of a partition, in particular, it allows us to say whether a partition is big or small.

When the size of the partition is small, it means that every interval in the partition is small.

MA 105 D3 Lecture 10

Ravi Raghunathan

Department of Mathematics

August 29

Recap: The Darboux integral

Riemann integration

Partitions

Definition: Given a closed interval $[a, b]$, a partition P of $[a, b]$ is simply a collections of points

$$P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}.$$

We can think of the points of the partition as dividing the original interval $[a, b]$ into sub-intervals $I_j = [x_{j-1}, x_j]$, $1 \leq j \leq n$. Indeed $I = \cup_j I_j$ and if two sub-interval intersect, they have at most one point in common. Hence, the notation "partition".

Definition: A partition $P' = \{a = x'_0 < x'_1 < \ldots < x'_m = b\}$ is said to be a refinement of the partition P if for each $x_i \in P$, there exists an $x'_j \in P'$ such that $x_i = x'_j$ (more compactly, $P \subseteq P'$).

Intuitively, a refinement P' of a partition P will break some of the sub-intervals in P into smaller sub-intervals. Any two partitions P_1, P_2 have a common refinement - $P_1 \cup P_2$.

Lower and Upper sums

Given a partition $P = \{a = x_0 < x_1 < \ldots < x_{b-1} < x_n = b\}$ and a function $f : [a, b] \to \mathbb{R}$, we define two associated quantities. First we set:

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x) \quad \text{and} \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \quad 1 \leq i \leq n$$

Defintion: We define the Lower sum as

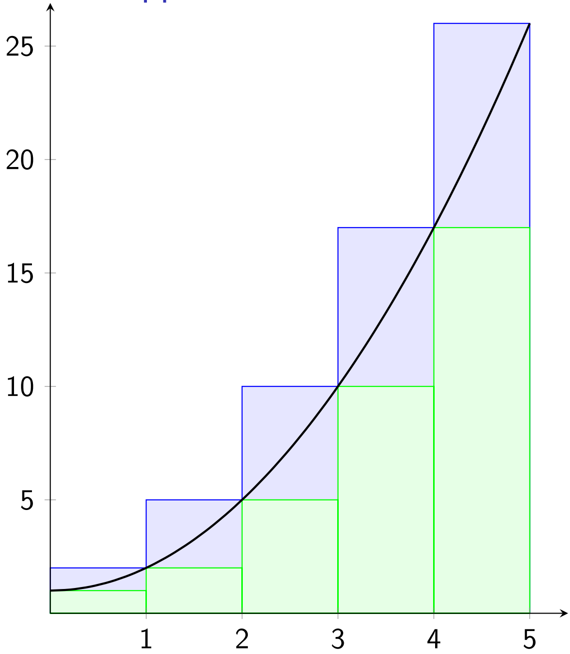
$$L(f, P) = \sum_{j=1}^{n} m_j(x_j - x_{j-1}).$$

Similarly, we can define the Upper sum as

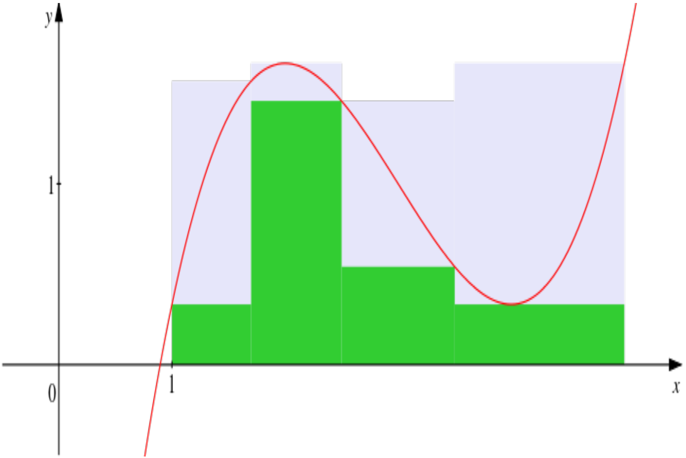
$$U(f, P) = \sum_{j=1}^{n} M_j(x_j - x_{j-1}).$$

In case the words "infimum" and "supremum" bother you, you can think "minimum" and "maximum most of time since we will usually be dealing with continuous functions on $[a, b]$.

Lower and Upper Darboux sums



A picture for a non-monotonic function



https://upload.wikimedia.org/wikipedia/commons/thumb/5/59/Darboux.svg/700px-Darboux.svg.png

The Darboux integrals

We now define the lower Darboux integral of f by

$$L(f) = \sup\{L(f, P)\colon P \text{ is a partition of } [a, b]\},$$

where the supremum is taken over all partitions of $[a, b]$.

and similarly the upper Darboux integral of f by

$$U(f) = \inf\{U(f, P)\colon P \text{ is a partition of } [a, b]\},$$

and again the infimum is over all partitions of $[a, b]$. (This time there is no escaping inf and sup!)

If $L(f) = U(f)$, then we say that f is Darboux-integrable and define

$$\int_a^b f(t)\,dt := U(f) = L(f).$$

This common value of the two integrals is called the Darboux integral.

Useful properties of the Darboux sums

Since, for any partition P, $L(f, P) \leq U(f, P)$, we have

$$L(f) \leq U(f).$$

In fact, for any two partitions P_1 and P_2, we have

$$L(f, P_1) \leq U(f, P_2).$$

One of the most useful properties of the Darboux sums is the following. If P' is a refinement of P then obviously

$$L(f, P) \leq L(f, P') \leq U(f, P') \leq U(f, P).$$

Riemann Sums

There is another way of getting at the integral due to Riemann which may be a little more intuitive and is better for calculation. This is done via Riemann sums.

To define the notion of a Riemann sum we need one more piece of data. Suppose that for each of the intervals I_j we are given a point $t_j \in I_j$. We will denote the collection of points t_j by t. The pair (P, t) is sometimes called a tagged partition.

Definition: We define the Riemann sum associated to the function f, and the tagged partition (P, t) by

$$R(f, P, t) = \sum_{j=1}^{n} f(t_j)(x_j - x_{j-1}).$$

The norm of a partition

As must be clear, the Lower sum, Upper sum and Riemann sum all give approximations to the area between the lines $x = a$ and $x = b$ and between the curve $y = f(x)$ and the x-axis and

$$L(f, P) \leq R(f, P, t) \leq U(f, P).$$

The point is to make this statement quantitatively precise.

We define the norm of a partition P (denoted $\|P\|$) by

$$\|P\| = \max_j \{|x_j - x_{j-1}|\}, \quad 1 \leq j \leq n.$$

The norm gives some measure of the "size" of a partition, in particular, it allows us to say whether a partition is big or small.

When the size of the partition is small, it means that every interval in the partition is small.

The Riemann integral

Definition 1: A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\delta > 0$ such that

$$|R(f, P, t) - R| < \epsilon,$$

whenever $\|P\| < \delta$. In this case R is called the Riemann integral of the function f on the interval $[a, b]$.

In other words, for all sufficiently "small" or "fine" partitions, the Riemann sums must be within ϵ of R.

Notice, that as long as $\|P\|$ is small, it doesn't matter exactly where the x_j's or the t_j's are in the interval $[a, b]$.

Also notice that if P' is a refinement of P, then $\|P'\| \leq \|P\|$.

The Riemann integral continued

Intuitively, we can see that the smaller or finer the partition, the better the area under the curve is represented by the Riemann sum.

The reason that the Riemann integral is useful is because the definition we have given is actually equivalent to the following apparently weaker definition.

Definition 2: A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists a partition P such that for every tagged refinement of (P', t') of P with $\|P'\| \leq \delta$,

$$|R(f, P', t') - R| < \epsilon.$$

The nice thing about the above definition is that one only has to check that $|R(f, P', t') - R|$ is small for for refinements of a fixed partition, and not for all partitions.

Back to our example

Using Definition 2 of the Riemann integral it is easy to see that the function $f(x) = x$ is Riemann integrable.

Let $\epsilon > 0$ be arbitrary. For our fixed partition we take $P = P_n$ where $n > \frac{1}{\epsilon}$ is some fixed number. Moreover, if (P', t') is any refinement of P_n we have

$$L(f, P_n) \leq L(f, P') \leq R(f, P', t') \leq U(f, P') \leq U(f, P_n),$$

whence it follows that (remember $U(f, P_n) - 1/2 + 1/2n$ and $L(f, P_n) = 1/2 - 1/2n$)

$$|R(f, P', t') - \frac{1}{2}| < \epsilon.$$

The example continued

As the preceding example shows, Definition 2 of the Riemann integral is really easy to work with. Why do we then care about Definition 1 or the Darboux integral?

The reason is that while Definition 2 is good for showing that a given function is Riemann integrable, the other definitions are often better for proving the *abstract properties* of integrals.

In fact, this will be clear in the tutorial exercises. You will see that sometimes the Darboux integral is better than the Riemann integral.

Before going any further we will formally state what we have already been referring to for several slides.

Comparison with the Darboux integral

Theorem 20: The Riemann integral (using either definition) exists if and only if the Darboux integral exists and in this case the two integrals are equal.

With this theorem in hand, we see that the function $f(x) = x$ is also Darboux integrable.

How does one prove Theorem 20? It is not too hard but it takes some work and is roundabout.

The easiest way is to proceed as follows. It is clear that if f is Riemann integrable in the sense of Definition 1, it is Riemann integrable in the sense of Definition 2. Next, one shows that if f is Riemann integrable in the sense of Definition 2, then it is Darboux integrable. And finally, one can show that if the Darboux integral exists, then the Riemann integral exists in the sense of Definition 1. An interested student can try this as an exercise.

The main theorem for Riemann integration

From now on we will use any of the three definitions - the Darboux definition, Definition 1 and Definition 2 for the integral interchangeably and we will use only the words Riemann integral.

The main theorem of Riemann integration is the following:

Theorem 21: Let $f : [a, b] \to \mathbb{R}$ be a function that is bounded, and continuous at all but finitely many points of $[a, b]$. Then f is Riemann integrable on $[a, b]$.

In fact, one can allow even countably many discontinuities and the Theorem will remain true.
Exercise 1: Those of you who have an extra interest in the course should think about trying to prove both Theorem 21 and the extension to countably many discontinuities (Warning: there is one crucial fact about continuous functions that we have not covered that you will have to discover for yourself).

An example of a function that is not Darboux integrable

Here is a function that is not Darboux integrable of $[0, 1]$. Let

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational.} \end{cases}$$

It should be clear that no matter what partition one takes the infimum on any sub-interval in the partition will be 0 and the supremum will be 1.

From this one can see immediately that

$$L(f, P) = 0 \neq 1 = U(f, P),$$

for every P, and hence that $L(f) = 0 \neq 1 = U(f)$.

Another property of the Riemann Integral

Theorem 23: Suppose f is Riemann integrable on $[a, b]$ and $c \in [a, b]$. Then

$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt.$$

Proof: First we note that if $c = a$ or $c = b$, there is nothing to prove.

Next, if $c \in (a, b)$ we proceed as follows. If P_1 is a partition of $[a, c]$ and P_2 is a partition of $[c, b]$, then $P_1 \cup P_2 = P'$ is obviously a partition of $[a, b]$. Thus, partitions of the form $P_1 \cup P_2$ constitute a subset of the set of all partitions of $[a, b]$. For such partitions P' we have

$$L(f, P') = L(f, P_1) + L(f, P_2).$$

Let us denote by $L(f)_{[a,c]}$ (resp. $L(f)_{[c,b]}$) the Darboux lower integral of f on the interval $[a, c]$ (resp. $[c, b]$).

MA 109 D3 Lecture 11

Ravi Raghunathan

Department of Mathematics

August 31, 2023

Recap: Darboux and Riemann integration

The fundamental theorem of calculus

The Riemann integral

Definition 1: A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\delta > 0$ such that

$$|R(f, P, t) - R| < \epsilon,$$

whenever $\|P\| < \delta$. In this case R is called the Riemann integral of the function f on the interval $[a, b]$.

Definition 2: A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\delta > 0$ and a partition P, such that for every tagged refinement (P', t') of P with $\|P'\| \leq \delta$,

$$|R(f, P', t') - R| < \epsilon.$$

The nice thing about the above definition is that one only has to check that $|R(f, P', t') - R|$ is small for for refinements of a fixed partition, and not for all partitions.

Comparison with the Darboux integral

Theorem 20: The Riemann integral (using either definition) exists if and only if the Darboux integral exists and in this case the two integrals are equal.

Theorem 21: Let $f : [a, b] \to \mathbb{R}$ be a function that is bounded, and continuous at all but finitely many points of $[a, b]$. Then f is Riemann integrable on $[a, b]$.

Because of Theorem 20, we are now free to use Darboux integrability and Riemann integrability interchangably.

An example of a function that is not Darboux/Riemann integrable

Here is a function that is not Darboux integrable of $[0, 1]$. Let

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational.} \end{cases}$$

It should be clear that no matter what partition one takes the

infimum on any sub-interval in the partition will be 0 and the supremum will be 1.

From this one can see immediately that

$$L(f, P) = 0 \neq 1 = U(f, P),$$

for every P, and hence that $L(f) = 0 \neq 1 = U(f)$.

Another property of the Riemann Integral

Theorem 23: Suppose f is Riemann integrable on $[a, b]$ and $c \in [a, b]$. Then

$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt.$$

Proof: First we note that if $c = a$ or $c = b$, there is nothing to prove.

Next, if $c \in (a, b)$ we proceed as follows. If P_1 is a partition of $[a, c]$ and P_2 is a partition of $[c, b]$, then $P_1 \cup P_2 = P'$ is obviously a partition of $[a, b]$. Thus, partitions of the form $P_1 \cup P_2$ constitute a subset of the set of all partitions of $[a, b]$. For such partitions P' we have

$$L(f, P') = L(f, P_1) + L(f, P_2).$$

Let us denote by $L(f)_{[a,c]}$ (resp. $L(f)_{[c,b]}$) the Darboux lower integral of f on the interval $[a, c]$ (resp. $[c, b]$).

If we take the supremum over all partitions P_1 of $[a, c]$ and P_2 of $[c, b]$ we get

$$\sup_{P'} L(f, P') = L(f)_{[a,c]} + L(f)_{[c,b]}.$$

Now the supremum on the left hand side is taken only over partitions P' having the special form $P_1 \cup P_2$. Hence it is less than or equal to $\sup_P L(f, P)$ where this supremum is taken over all partitions P. We thus obtain

$$L(f)_{[a,c]} + L(f)_{[c,b]} \leq L(f).$$

On the other hand, for any partition $P = \{a < x_1 < \ldots x_{n-1} < b\}$ we can consider the partition $P' = P \cup \{c\}$. This will be a refinement of the partition P and can be written as a union of two partitions P_1 of $[a, c]$ and P_2 of $[c, b]$.

Remarks not made in class

Note that if f is Riemann integrable on $[a, b]$ it is automatically Riemann integrable on $[a, c]$ and $[c, b]$. With notation as above, note that

$$0 \leq U(f, P_1) - L(f, P_1) \leq U(f, P') - L(f, P'),$$

so if the second expression is small, the first one will be too. Thus f will be integrable on $[a, c]$, and similarly for $[c, b]$.
Two students (including Arrol - unfortunately I did not get the other student's name) have pointed this out and have also pointed out that one can then be more efficient with the proof. My thanks to them. However, I prefer to not to revisit the proof on these slides.

By the property for refinements for Darboux sums we know that $L(f, P) \leq L(f, P')$.

Thus, given any partition P of $[a, b]$, there is a refinement P' which can be written as the union of two partitions P_1 and P_2 of $[a, c]$ and $[c, b]$ respectively, and by the above inequality,

$$\sup_P L(f, P) \leq \sup_{P'} L(f, P'),$$

where the first supremum is taken over all partitions of $[a, b]$ and the second only over those partitions P' which can be written as a union of two partitions as above. This shows that

$$L(f) \leq L(f)_{[a,c]} + L(f)_{[c,b]},$$

so, together with the previous inequality, we get

$$L(f) = L(f)_{[a,c]} + L(f)_{[c,b]}.$$

The same kind of reasoning applies to the upper sums which allows us to prove the required property.

Motivation

The Fundamental Theorem of calculus allows us to relate the process of Riemann integration to the process of differentiation. Essentially, it tells us that integrating and differentiating are inverse processes. This is a tremendously useful theorem for several reasons.

It turns out that (Riemann) integrating even simple functions is much harder than differentiating them (if you don't believe me, try integrating $(\tan x)^3$ via Riemann sums!). In practice, however, integration is what we need to do to solve physical problems. Usually, when we are studying the motion of a particle or a planet what we find is that the position of a particle, which is a function of time, satisfies some differential equation. Solving the differential equation involves performing the inverse operation of taking some combination of derivatives. The simplest such inverse operation is taking the inverse of the first derivative, which the Fundamental Theorem says, is the same as integrating.

Calculating Integrals

Thus, calculating integrals is one of the basic things one needs to do for solving even the simplest physics and engineering problems. The problem is that this is quite difficult to do.

Once we know the derivatives of some basic functions (polynomials, trigonometric functions, exponentials, logarithms) we can differentiate a wide class of functions using the rules for differentiation , especially the product and chain rules. By contrast, the only rule for Riemann integration that can be proved from the basic definitions is the sum rule.

The Fundamental Theorem solves this problem (partially) because it allows us to deduce formalæ for the integrals of the products and the composition of functions from the corresponding rules for derivatives.

The Fundamental Theorem - Part I

Theorem 24 (Part I): Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and let

$$F(x) = \int_a^x f(t)dt$$

for any $x \in [a, b]$. Then $F(x)$ is continuous on $[a, b]$, differentiable on (a, b) and

$$F'(x) = f(x),$$

for all $x \in (a, b)$.

Proof: We know that $f(t)$ is Riemann integrable for any $x \in [a, b]$ because of Theorem 21 (every continuous function is Riemann integrable).

The proof of Part I continued

By Theorem 23 we know that

$$\int_a^{x+h} f(t)dt = \int_a^x f(t)dt + \int_x^{x+h} f(t)dt,$$

for $x + h \in [a, b]$. Hence

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h} \cdot \int_x^{x+h} f(t)dt.$$

We know that if $f(t) \leq g(t)$ on $[a, b]$, then $\int f(t)dt \leq \int g(t)dt$. We apply this to the three functions $m(h)$, f and $M(h)$, where $m(h)$ and $M(h)$ are the constant functions given by the minimum and maximum of the function f on $[x, x + h]$ to get:

$$m(h) \cdot h \leq \int_x^{x+h} f(t)dt \leq M(h) \cdot h.$$

Dividing by h and taking the limit gives

$$\lim_{h\to 0} m(h) \leq \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \leq \lim_{h\to 0} M(h).$$

But f is a continuous function, so $\lim_{h\to 0} m(h) = \lim_{h\to 0} M(h) = f(x)$. By the Sandwich theorem for limits (use version 2), we see that limit in the middle exists and is equal to $f(x)$, that is $F'(x) = f(x)$. This proves the first part of the Fundamental Theorem of Calculus. $\square$

This first form of the Fundamental Theorem allows us to compute definite integrals. Keeping the notation as in the Theorem we obtain

Corollary:

$$\int_c^d f(t)dt = F(d) - F(c),$$

for any two points $c, d \in [a, b]$.

The Fundamental Theorem of Calculus Part 2

Theorem 24 (Part II): Let $f : [a, b] \to \mathbb{R}$ be given and suppose there exists a continuous function $g : [a, b] \to \mathbb{R}$ which is differentiable on (a, b) and which satisfies $g'(x) = f(x)$. Then, if f is Riemann integrable on $[a, b]$,

$$\int_a^b f(t)dt = g(b) - g(a).$$

Note that this statement does not assume that the function $f(t)$ is continuous, and is thus stronger than the corollary just stated.

Proof: We can write:

$$g(b) - g(a) = \sum_{i=1}^{n} [g(x_i) - g(x_{i-1})],$$

where $\{a = x_0, x_1, \ldots, x_n = b\}$ is an arbitrary partition of $[a, b]$. Using the mean value theorem for each of the intervals $I_j = [x_j, x_{j-1}]$, we can write

The proof of the Fundamental Theorem part II continued

$$g(x_i) - g(x_{i-1}) = g'(c_i)(x_i - x_{i-1}).$$

where $c_i \in (x_{i-1}, x_i)$.

Substituting this in the previous expression and using the fact that $g'(c_i) = f(c_i)$, we get

$$g(b) - g(a) = \sum_{i=1}^{n} [f(c_i)(x_i - x_{i-1})].$$

The calculation above is valid for any partition The right hand side obviously represents a Riemann sum. By hypothesis f is Riemann integrable. It follows (using Definition 1, for example) that as $\|P\| \to 0$, the right hand side goes to the Riemann integral. □

MA 109 D3 Lecture 12

Ravi Raghunathan

Department of Mathematics

September 4, 2023

A remark on Definition 2 of Riemann integrability

Recall that the second definition of Riemann integrability was the following:

Definition 2: A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\delta > 0$ and a partition P, such that for every tagged refinement (P', t') of P with $\|P''\| < \delta$,

$$|R(f, P', t') - R| < \epsilon. \qquad (1)$$

Suppose f is Riemann integrable in the sense above and let P be the chosen partition. Let P' be a refinement of P such that $\|P'\| < \delta$. If P'' is any refinement of P', it is a refinement of P, $\|P''\| < \delta$ and (1) holds for P''. Thus, by replacing P by P' we can reformulate Definition 2 as follows:

A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$, a partition P, such that for every tagged refinement (P', t') of P

$$|R(f, P', t') - R| < \epsilon.$$

The Fundamental Theorem - Part I

Theorem 24 (Part I): Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and let

$$F(x) = \int_a^x f(t)dt$$

for any $x \in [a, b]$. Then $F(x)$ is continuous on $[a, b]$, differentiable on (a, b) and

$$F'(x) = f(x),$$

for all $x \in (a, b)$.

Theorem 24 (Part II): Let $f : [a, b] \to \mathbb{R}$ be given and suppose there exists a continuous function $g : [a, b] \to \mathbb{R}$ which is differentiable on (a, b) and which satisfies $g'(x) = f(x)$. Then, if f is Riemann integrable on $[a, b]$,

$$\int_a^b f(t)dt = g(b) - g(a).$$

Tutorial Problem 4.4

Exercise 4.4 Compute

(a) $\frac{d^2y}{dx^2}$, if

$$x = \int_0^y \frac{dt}{\sqrt{1+t^2}}$$

(b) $\frac{dF}{dx}$, if for $x \in \mathbb{R}$

$$(i) F(x) = \int_1^{2x} \cos(t^2) dt$$

and

$$(ii) F(x) = \int_0^{x^2} \cos(t) dt.$$

Problem 4.5

Let p be a real number and let f be a continuous function on $\mathbb{R}$ that satisfies the equation $f(x+p) = f(x)$ for all $x \in \mathbb{R}$. Show that the integral

$$\int_a^{a+p} f(t)dt$$

has the same value for every real number a.

(Hint: Consider $F(a) = \int_a^{a+p} f(t)dt$.)

The logarithm

For $x \in (0, \infty)$ we define

$$f(x) = \int_1^x \frac{1}{t} dt.$$

Then, for any y, define $g(x) = f(xy)$

Differentiating with respect to x we see that $g'(x) = f'(x)$ Hence,

$$f(x) = g(x) + C,$$

for some constant C. Set $x = 1$ to obtain $C = -f(y)$. Thus,

$$f(xy) = f(x) + f(y).$$

The logarithm and exponential functions

The function $f(x)$ is usually denoted $\ln x$. Since $f'(x) = \frac{1}{x} > 0$, whenever $x > 0$, we see that f is (strictly) monotonic increasing and concave.

By computing the Darboux lower sums associated to $\ln x$, we can easily check that $\ln x > 1$ if $x \geq 3$. By the intermediate value theorem, it follows that there exists a real number e, such that $\ln e = 1$.

It is not hard to see that f must have an inverse function. This is the exponential function sometimes denoted $\exp(x)$. Clearly $\exp(x + y) = \exp(x) \cdot \exp(y)$. Again, it requires some work to see that

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

When $x = 1$ we will obtain a formula for e!

The Mean Value Theorem for Integration

Let $f : [a, b] \to \mathbb{R}$ be a continuous function. The slide that was projected in the last class unnecessarily had the condition that f was differentiable. We apply the Mean Value Theorem to the function

$$F(x) = \int_a^x f(t)dt.$$

This says that there exists $c \in (a, b)$ such that

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

But this is the same as saying

$$\int_a^b f(t)dt = f(c)(b - a).$$

This is the Mean Value Theorem for integration.

Functions with range contained in $\mathbb{R}$

We will be interested in studying functions $f : \mathbb{R}^m \to \mathbb{R}$, when $m = 2, 3$. We have already mentioned how limits of such functions can be studied in the first few lectures. Before doing this in detail, however, we will study certain other features of functions in two and three variables.

The most basic thing one needs to understand about a function is the domain on which it is defined. Very often a function is given by a formula which makes sense only on some subset of $\mathbb{R}^m$ and not on the whole of $\mathbb{R}^m$. When studying functions of two or more variables given by formulæ it makes sense to first identify this subset, which is sometimes call the natural domain of the function, and to describe it geometrically if possible.

Exercise 5.1: Find the natural domains of the following functions:
(i) $\frac{xy}{x^2-y^2}$

Clearly this function is defined whenever the denominator is not zero, in other words when $x^2 - y^2 \neq 0$.
The natural domain is thus

$$\mathbb{R}^2 \setminus \{(x,y) \,|\, x^2 - y^2 = 0\},$$

that is, $\mathbb{R}^2$ minus the pair of straight lines with slopes ± 1.
(ii) $f(x,y) = \log(x^2 + y^2)$
This function is defined whenever $x^2 + y^2 \neq 0$, in other words, in $\mathbb{R}^2 \setminus \{(0,0)\}$.

Level curves and contour lines

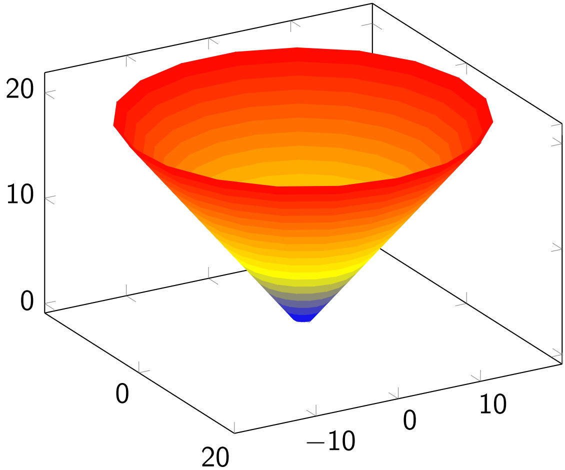
The second thing one should do with a function from $\mathbb{R}^2 \to \mathbb{R}$ is to study its range. This is done in different ways.

One way is to study the level sets of the functions. These are the sets of the form $\{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$, where c is a constant. The level set "lives" in the xy-plane.

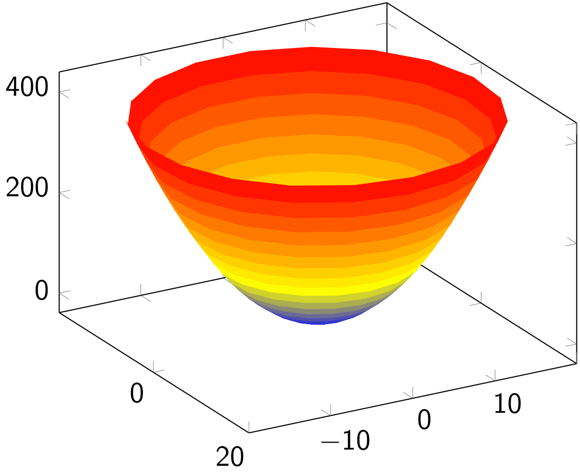
One can also plot (in three dimensions) the surface $z = f(x, y)$. By varying the value of c in the level curves one can get a good idea of what the surface looks like.

When one plots the $f(x, y) = c$ for some constant c one gets a curve. Such a curve is usually called a contour line (the contour "lives" in the $z = c$ plane).

I have a couple of pictures in the next two slides to illustrate the point.



This is the graph of the function $z = \sqrt{x^2 + y^2}$ lying above the xy-plane. It is a right circular cone.
The contour lines $z = c$ give circles lying on planes parallel to the xy-plane.The curves given by $z = f(x, 0)$ and $z = f(0, y)$ give pairs of straight lines in the planes $y = 0$ and $x = 0$.



This is the graph of the function $z = x^2 + y^2$ lying above the xy-plane. It is a paraboloid of revolution.
The contour lines $z = c$ give circles lying on planes parallel to the xy-plane.The curves $z = f(x, 0)$ or $z = f(y, 0)$ give parabolæ lying in the planes $y = 0$ and $x = 0$. Exercise 5.2.(ii).

Limits

We have already said what it means for a function of two or more variables to approach a limit. We simply have to replace the absolute value function on $\mathbb{R}$ by the distance function on $\mathbb{R}^m$. We will do this in two variables. The three variable definition is entirely analogous. We will denote by U a set in $\mathbb{R}^2$.

Definition: A function $f : U \to \mathbb{R}$ is said to tend to a limit l as $x = (x_1, x_2)$ approaches $c = (c_1, c_2)$ if for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x) - l| < \epsilon,$$

whenever $0 < \|x - c\| < \delta$ with $x \in U$.

We recall that

$$\|x\| = \sqrt{x_1^2 + x_2^2}.$$

Continuity

Before talking about continuity we remark the following. In the plane $\mathbb{R}^2$ it is possible to approach the point c from infinitely many different directions - not just from the right and from the left. In fact, one may not even be approaching the point c along a straight line! Hence, to say that a function from $\mathbb{R}^2$ to $\mathbb{R}$ possesses a limit is actually imposing a strong condition - for instance, the limits along all possible curves leading to the point must exist and all these (infinitely many) limits must be equal.

Once we have the notion of a limit, the definition of continuity is just the same as for functions of one variable.

Definition: The function $f : U \to \mathbb{R}$ is said to be continuous at a point c if

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

The rules for limits and continuity

The rules for addition, subtraction, multiplication and division of limits remain valid for functions of two variables (or three variables for that matter). Nothing really changes in the statements or the proofs.

Using these rules, we can conclude, as before, that the sum, difference, product and quotient of continuous functions are continuous (as usual we must assume that the denominator of the quotient is non zero).

MA 109 D3 Lecture 13

Ravi Raghunathan

Department of Mathematics

September 5, 2023

Functions of severable variables

Limits and continuity

Differentiation

Functions with range contained in $\mathbb{R}$

We want to study functions from $\mathbb{R}^n \to \mathbb{R}^m$ (in our course $1 \leq m, n \leq 3$.

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a function, we see that $f(x) \in \mathbb{R}^m$ for $x \in \mathbb{R}^n$, so $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$, where f_i, $1 \leq i \leq m$ is a function from $\mathbb{R}^n \to \mathbb{R}$. Thus, studying functions from $\mathbb{R}^n \to \mathbb{R}^m$ is the same as studying an m-tuple of functions of functions from $\mathbb{R}^n$ to $\mathbb{R}$.

For now, we study functions $f : \mathbb{R}^n \to \mathbb{R}$, or more commonly, functions $f : U \to \mathbb{R}$, where $U \subseteq \mathbb{R}^n$. The function f will often be given by a formula and there will be some natural domain on which the formula makes sense.

Level curves and contour lines

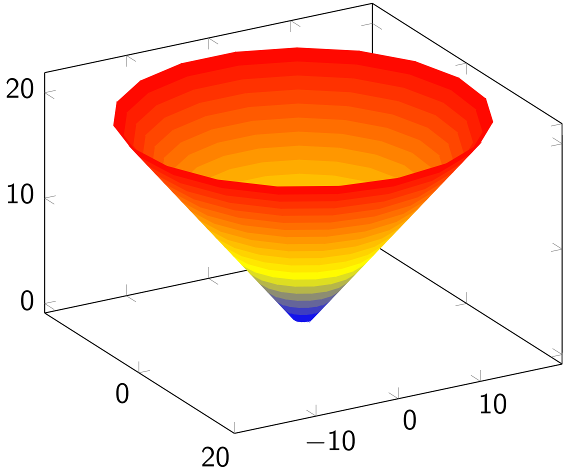
The second thing one should do with a function from $\mathbb{R}^2 \to \mathbb{R}$ is to study its range. This is done in different ways.

One way is to study the level sets of the functions. These are the sets of the form $\{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$, where c is a constant. The level set "lives" in the xy-plane.

One can also plot (in three dimensions) the surface $z = f(x, y)$. By varying the value of c in the level curves one can get a good idea of what the surface looks like.

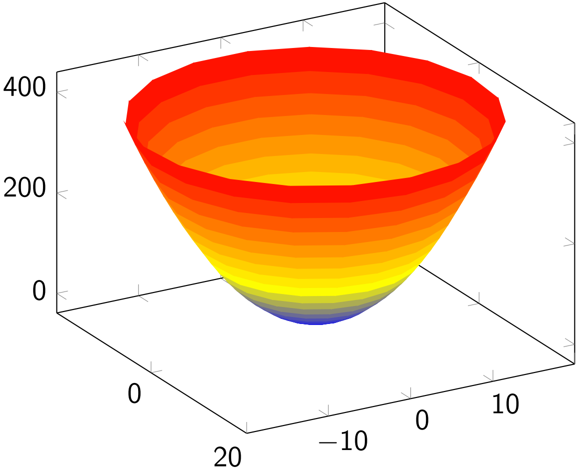
When one plots the $f(x, y) = c$ for some constant c one gets a curve. Such a curve is usually called a contour line (the contour "lives" in the $z = c$ plane).

I have a couple of pictures in the next two slides to illustrate the point.



This is the graph of the function $z = \sqrt{x^2 + y^2}$ lying above the xy-plane. It is a right circular cone.

The contour lines $z = c$ give circles lying on planes parallel to the xy-plane.The curves given by $z = f(x, 0)$ and $z = f(0, y)$ give pairs of straight lines in the planes $y = 0$ and $x = 0$.



This is the graph of the function $z = x^2 + y^2$ lying above the xy-plane. It is a paraboloid of revolution.
The contour lines $z = c$ give circles lying on planes parallel to the xy-plane.The curves $z = f(x, 0)$ or $z = f(y, 0)$ give parabolæ lying in the planes $y = 0$ and $x = 0$. Exercise 5.2.(ii).

Limits

We have already said what it means for a function of two or more variables to approach a limit. We simply have to replace the absolute value function on $\mathbb{R}$ by the distance function on $\mathbb{R}^m$. We will do this in two variables. The three variable definition is entirely analogous. We will denote by U a set in $\mathbb{R}^2$.

Definition: A function $f : U \to \mathbb{R}$ is said to tend to a limit l as $x = (x_1, x_2)$ approaches $c = (c_1, c_2)$ if for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x) - l| < \epsilon,$$

whenever $0 < \|x - c\| < \delta$ with $x \in U$.

We recall that

$$\|x\| = \sqrt{x_1^2 + x_2^2}.$$

Notice that the set

$$B_\delta(c) = \{x \in \mathbb{R}^2 \mid \|x - c\| < \delta\}$$

is a circular disc/ball of radius δ. Thus, $f(x)$ is close to l (that is, within a distance ε) whenever x lies in a sufficiently small disc (of radius δ).

In the plane $\mathbb{R}^2$ it is possible to approach the point c from infinitely many different directions - not just from the right and from the left. In fact, one may not even be approaching the point c along a straight line! Hence, to say that a function from $\mathbb{R}^2$ to $\mathbb{R}$ possesses a limit is actually imposing a strong condition - for instance, the limits along all possible curves leading to the point must exist and all these (infinitely many) limits must be equal.

Continuity

Once we have the notion of a limit, the definition of continuity is just the same as for functions of one variable.

Definition: The function $f : U \to \mathbb{R}$ is said to be continuous at a point c if

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

Exercise: Formulate the definition of a limit and of continuity for functions from $\mathbb{R}^3$ to $\mathbb{R}$.

The rules for limits and continuity

The rules for addition, subtraction, multiplication and division of limits remain valid for functions of two variables (or three variables for that matter). Nothing really changes in the statements or the proofs.

Using these rules, we can conclude, as before, that the sum, difference, product and quotient of continuous functions are continuous (as usual we must assume that the denominator of the quotient is non zero).

Continuity through examples

Once again, we emphasise that continuity at a point c is a very powerful condition (since the existence of a limit is implicit).

Exercise 5.3.(i) asks whether the function

$$f(x,y) = \begin{cases} \frac{x^3 y}{x^6+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

is continuous at $(0,0)$.

Solution: Let us look at the sequence of points $z_n = (\frac{1}{n}, \frac{1}{n^3})$, which goes to 0 as $n \to \infty$. Clearly $f(z_n) = \frac{1}{2}$ for all n, so

$$\lim_{n\to\infty} f(z_n) = \frac{1}{2} \neq 0.$$

This shows that f is not continuous at 0.
But does the limit exist?

Iterated limits

When evaluating a limit of the form $\lim_{(x_1,x_2)\to(c_1,c_2)} f(x_1,x_2)$ one may naturally be tempted to let x_1 go to c_1 first, and then let x_2 go to c_2. Does this give the limit in the previous sense?

Exercise 5.5: Let

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

we have

$$\lim_{x\to 0}\lim_{y\to 0} \frac{x^2y^2}{x^2y^2 + (x-y)^2} = \lim_{x\to 0} 0 = 0$$

Similarly, one has $\lim_{y\to 0}\lim_{x\to 0} f(x,y) = 0$.

However, choosing $z_n = (\frac{1}{n}, \frac{1}{n})$, shows that $f(z_n) = 1$ for all $n \in \mathbb{N}$. Now choose $z_n = (\frac{1}{n}, \frac{1}{2n})$ to see that the limit cannot exist.

Partial Derivatives

As before, U will denote a subset of $\mathbb{R}^2$. Given a function $f : U \to \mathbb{R}$, we can fix one of the variables and view the function f as a function of the other variable alone. We can then take the derivative of this one variable function.

To make things precise, fix x_2.
Definition: The partial derivative of $f : U \to \mathbb{R}$ with respect to x_1 at the point (a, b) is defined by

$$\frac{\partial f}{\partial x_1}(a, b) := \lim_{x_1 \to a} \frac{f((x_1, b)) - f((a, b))}{x_1 - a}.$$

Similarly, one can define the partial derivative with respect to x_2. In this case the variable x_1 is fixed and f is regarded only as a function of x_2:

$$\frac{\partial f}{\partial x_2}(a, b) := \lim_{x_2 \to b} \frac{f((a, x_2)) - f((a, b))}{x_2 - b}.$$

Directional Derivatives

The partial derivatives are special cases of the directional derivative. Let $v = (v_1, v_2)$ be a unit vector. Then v specifies a direction in $\mathbb{R}^2$.

Definition: The directional derivative of f in the direction v at a point $x = (x_1, x_2)$ is denoted by $\nabla_v f(x)$ and is defined as

$$\lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} = \lim_{t \to 0} \frac{f((x_1 + tv_1, x_2 + tv_2)) - f((x_1, x_2))}{t}.$$

$\nabla_v f(x)$ measures the rate of change of the function f at x along the path $x + tv$.

If we take $v = (1, 0)$ in the above definition, we obtain $\frac{\partial f}{\partial x_1}(x)$, while $v = (0, 1)$ yields $\frac{\partial f}{\partial x_2}(x)$.

Consider the function

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 = 0 \text{ or if } x_2 = 0 \\ 0 & \text{otherwise.} \end{cases}$$

It should be clear to you that since this function is constant along the two axes,

$$\frac{\partial f}{\partial x_1}(0,0) = 0 \quad \text{and} \quad \frac{\partial f}{\partial x_2}(0,0) = 0$$

On the other hand, $f(x_1, x_2)$ is not continuous at the origin! Thus, a function may have both partial derivatives (and, in fact, any directional derivative - see the next slide) but still not be continuous. This suggests that for a function of two variables, just requiring that both partial derivatives exist is not a good or useful definition of "differentiability".

Recall again, the following function from Exercise 5.5:

$$\frac{x^2y^2}{x^2y^2+(x-y)^2} \quad \text{for} \quad (x,y) \neq (0,0).$$

Let us further set $f(0,0)=0$. You can check that every directional derivative exists and is equal to 0, except along $y=x$ when the directional derivative is not defined. However, we have already seen that the function is not continuous at the origin since we have shown that $\lim_{(x,y)\to 0} f(x,y)$ does not exist. For an example with directional derivatives in all directions see Exercise 5.3(i).

Conclusion: All directional derivatives may exist at a point even if the function is discontinuous.

Let us go back and examine the notion of differentiability for a function of $f(x)$ of one variable. Suppose f is differentiable at the point x_0, What is the equation of the tangent line through $(x_0, f(x_0))$?

$$y = f(x_0) + f'(x_0)(x - x_0)$$

If we consider the difference $f(x) - f(x_0) - f'(x_0)(x - x_0)$ we get the distance of a point on the tangent line from the curve $y = f(x)$. Writing $h = (x - x_0)$, we see that the difference can be rewritten

$$f(x_0 + h) - f(x_0) - f'(x_0)h$$

The tangent line is close to the function f - how close?- so close that even after dividing by h the distance goes to 0. A few lectures ago we wrote this as

$$|f(x_0 + h) - f(x_0) - f'(x_0)h| = \varepsilon_1(h)|h|$$

where $\varepsilon_1(h)$ is a function that goes to 0 as h goes to 0. So $o(h) = \varepsilon_1(h)|h|$ is function that "goes to zero faster than h").

The preceding idea generalises to two (or more) dimensions. Let $f(x, y)$ be a function which has both partial derivatives. In the two variable case we need to look at the distance between the surface $z = f(x, y)$ and its tangent plane.

Let us first recall how to find the equation of a plane passing through the point $P = (x_0, y_0, z_0)$. It is the graph of the function

$$z = g(x, y) = z_0 + a(x - x_0) + b(y - y_0).$$

Let us determine the tangent plane to $z = f(x, y)$ passing through a point $P = (x_0, y_0, z_0)$ *on the surface*. In other words, we have to determine the constants a and b.

If we fix the y variable and treat $f(x, y)$ only as a function of x, we get a curve. Similarly, if we treat $g(x, y)$ as function only of x, we obtain a line. The tangent to the curve must be the same as the line passing through (x_0, y_0, z_0), and, in any event, their slopes must be the same. Thus, we must have

$$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial g}{\partial x}(x_0, y_0) = a.$$

Arguing in exactly the same way, but fixing the x variable and varying the y variable we obtain

$$\frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial g}{\partial y}(x_0, y_0) = b.$$

Hence, the equation of the tangent plane to $z = f(x, y)$ at the point (x_0, y_0) is

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

Differentiability for functions of two variables

We now define differentiability for functions of two variables by imitating the one variable definition, but using the "$o(h)$" version.
We let $(x, y) = (x_0, y_0) + (h, k) = (x_0 + h, y_0 + k)$
Definition A function $f : U \to \mathbb{R}$ is said to be differentiable at a point (x_0, y_0) if $\frac{\partial f}{\partial x}(x_0, y_0)$, and $\frac{\partial f}{\partial y}(x_0, y_0)$ exist and

$$\lim_{(h,k)\to 0} \frac{\left| f(x_0 + h, y_0 + k) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \right|}{\|(h, k)\|} = 0,$$

This is saying that the distance between the tangent plane and the surface is going to zero even after dividing by $\|(h, k)\|$. We could rewrite this as

$$\left| f((x_0, y_0) + (h, k)) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \right| = \varepsilon(h, k)\|(h, k)\|$$

where $\varepsilon(h, k)$ is a function that goes to 0 as $\|(h, k)\| \to 0$.This form of differentiability now looks exactly like the one variable version case (put $o(h, k) = \varepsilon(h, k)\|(h, k)\|$).

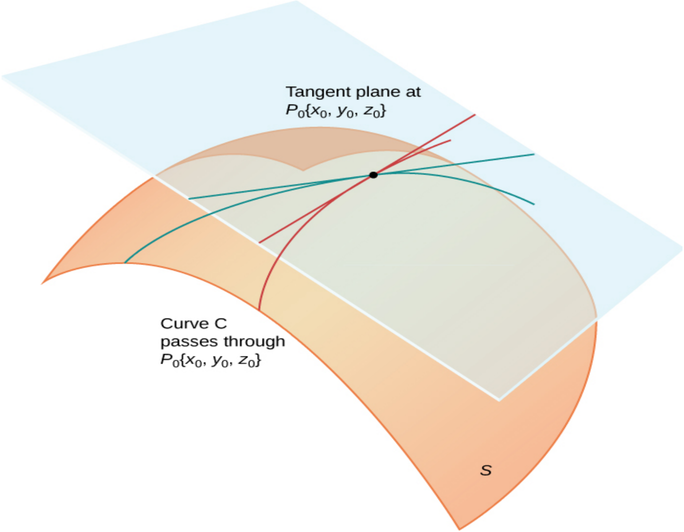
MA 105 D3 Lecture 14

Ravi Raghunathan

Department of Mathematics

September 7, 2023

The tangent plane in a picture



https://openstax.org/books/calculus-volume-3/pages/4-4-tangent-planes-and-linear-approximations

The tangent plane

Let $f(x, y)$ be a function which has both partial derivatives. In the two variable case we need to look at the distance between the surface $z = f(x, y)$ and its tangent plane.

Let us first recall how to find the equation of a plane passing through the point $P = (x_0, y_0, z_0)$. It is the graph of the function

$$z = g(x, y) = z_0 + a(x - x_0) + b(y - y_0).$$

Let us determine the tangent plane to $z = f(x, y)$ passing through a point $P = (x_0, y_0, z_0)$ *on the surface*. In other words, we have to determine the constants a and b.

If we fix the y variable and treat $f(x, y)$ only as a function of x, we get a curve. Similarly, if we treat $g(x, y)$ as function only of x, we obtain a line. The tangent to the curve must be the same as the line passing through (x_0, y_0, z_0), and, in any event, their slopes must be the same. Thus, we must have

$$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial g}{\partial x}(x_0, y_0) = a.$$

Arguing in exactly the same way, but fixing the x variable and varying the y variable we obtain

$$\frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial g}{\partial y}(x_0, y_0) = b.$$

Hence, the equation of the tangent plane to $z = f(x, y)$ at the point (x_0, y_0) is

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

The tangent plane to the sphere

Exercise: Find the equation of the tangent plane to the hemisphere $z = f(x, y) = \sqrt{1 - x^2 - y^2}$ at a point (x_0, y_0).

Solution: The partial derivatives are

Differentiability for functions of two variables

We now define differentiability for functions of two variables by imitating the one variable definition, but using the "$o(h)$" version. We let $(x, y) = (x_0, y_0) + (h, k) = (x_0 + h, y_0 + k)$

Definition A function $f : U \to \mathbb{R}$ is said to be differentiable at a point (x_0, y_0) if $\frac{\partial f}{\partial x}(x_0, y_0)$, and $\frac{\partial f}{\partial y}(x_0, y_0)$ exist and

$$\lim_{(h,k)\to 0} \frac{\left| f(x_0 + h, y_0 + k) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \right|}{\|(h, k)\|} = 0,$$

This is saying that the distance between the tangent plane and the surface is going to zero even after dividing by $\|(h, k)\|$. We could rewrite this as

$$\left| f((x_0, y_0) + (h, k)) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \right| = \varepsilon(h, k)\|(h, k)\|$$

where $\varepsilon(h, k)$ is a function that goes to 0 as $\|(h, k)\| \to 0$.This form of differentiability now looks exactly like the one variable version case (put $o(h, k) = \varepsilon(h, k)\|(h, k)\|$).

The derivative as a linear map

We can rewrite the differentiability criterion once more as follows. We define the 1×2 matrix

$$Df(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}.$$

A 1×2 matrix can be multiplied by a column vector (which is 2×1 matrix) to give a real number. In particular:

$$\begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

that is,

$$Df(x_0, y_0) \begin{pmatrix} h \\ k \end{pmatrix} = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k$$

The definition of differentiability can thus be reformulated using matrix notation.

Definition: The function $f(x, y)$ is said be differentiable at a point (x_0, y_0) if there exists a matrix denoted $Df((x_0, y_0))$ with the property that

$$f((x_0, y_0) + (h, k)) - f(x_0, y_0) - Df(x_0, y_0)\begin{pmatrix} h \\ k \end{pmatrix} = p(h, k)\|(h, k)\|,$$

for some function $p(h, k)$ which goes to zero as (h, k) goes to zero. Viewing the derivative as a matrix allows us to view it as a linear map from $\mathbb{R}^2 \to \mathbb{R}$. Given a 1×2 matrix A and two column vectors v and w, we see that

$$A \cdot (v + w) = A \cdot v + A \cdot w \quad \text{and} \quad A \cdot (\lambda v) = \lambda(A \cdot v),$$

for any real number λ. As we have seen before, functions satisfying the above two properties are called linear functions or linear maps. Thus, the map $v \to A \cdot v$ gives a linear map from $\mathbb{R}^2$ to $\mathbb{R}$.

The matrix $Df(x_0, y_0)$ is called the Derivative matrix of the function $f(x, y)$ at the point (x_0, y_0).

The Gradient

When viewed as a row vector rather than as a matrix, the Derivative matrix is called the gradient and is denoted $\nabla f(x_0, y_0)$. Thus

$$\nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right).$$

In terms of the coordinate vectors **i** and **j** the gradient can be written as

$$\nabla f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)\mathbf{i} + \frac{\partial f}{\partial y}(x_0, y_0)\mathbf{j}.$$

A criterion for differentiability

Before we state the criterion, we note that with our definition of differentiability, every differentiable function is continuous.

Theorem 26: Let $f : U \to \mathbb{R}$. If the partial derivatives $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ exist and are continuous in a neighbourhood of a point (x_0, y_0) (that is in a region of the plane of the form $\{(x, y) \mid \|(x, y) - (x_0, y_0)\| < r\}$ for some $r > 0$. Then f is differentiable at (x_0, y_0).

We omit the proof of this theorem. However, we note that a function whose partial derivatives exist and are continuous is said to be continuously differentiable or of class $\mathcal{C}^1$. The theorem says that every function that is $\mathcal{C}^1$ in a small disc around a point is differentiable at that point.

Three variables

For the next few slides, we will assume that $f : U \to \mathbb{R}$ is a function of three variables, that is, U is a subset of $\mathbb{R}^3$. In this case, if we denote the variables by x, y and z, we get three partial derivatives as follows: we hold two of the variables constant and vary the third. For instance if y and z are kept fixed while x is varied, we get the partial derivative with respect to x at the point (a, b, c):

$$\frac{\partial f}{\partial x}(a, b, c) = \lim_{x \to a} \frac{f(x, b, c) - f(a, b, c)}{x - a}.$$

In a similar way we can define the partial derivatives

$$\frac{\partial f}{\partial y}(a, b, c) \quad \text{and} \quad \frac{\partial f}{\partial z}(a, b, c).$$

Once we have the three partial derivatives we can once again define the gradient of f:

$$\nabla f(a, b, c) = \left(\frac{\partial f}{\partial x}(a, b, c), \frac{\partial f}{\partial y}(a, b, c), \frac{\partial f}{\partial z}(a, b, c) \right).$$

Differentiability in three variables

Exercise 1: Formulate a definition of differentiability for a function of three variables.

Exercise 2: Formulate the analogue of Theorem 26 for a function of three variables.

We can also define differentiability for functions from $\mathbb{R}^m$ to $\mathbb{R}^n$ where m and n are any positive integers. We will do this in detail in this course when m and n have the values 1 and 2 and 3.

Finally, the rules for the partial derivatives of sums, differences, products and quotients of functions $f, g : U \to \mathbb{R}$, ($U \subset \mathbb{R}^m$, $m = 2, 3$) are exactly analogous to those for the derivative of functions of one variable.

The Chain Rule

We now study the situation where we have composition of functions. We assume that $x, y : I \to \mathbb{R}$ are differentiable functions from some interval (open or closed) to $\mathbb{R}$. Thus the pair $(x(t), y(t))$ defines a function from I to $\mathbb{R}^2$. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is differentiable. We would like to study the derivative of the composite function $z(t) = f(x(t), y(t))$ from I to $\mathbb{R}$.

Theorem 27: With notation as above

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

For a function $w = f(x, y, z)$ in three variables the chain rule takes the form

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt}$$

Clarifications on the notation

The form in which I have written the chain rule is the standard one used in many books (both in engineering and mathematics). However, it is not very good notation. For instance, in the formula

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

the letter z is being used for two different functions: both as a function $z(t)$ from $\mathbb{R}$ to $\mathbb{R}$ on the left hand side, and as a function $z(x, y)$ from $\mathbb{R}^2$ to $\mathbb{R}$. If one wants to be precise one should write the chain rule as

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Similarly, for the function w we should write

$$\frac{dw}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

Verifying the chain rule in a simple case

Example: Let us verify this rule in a simple case. Let $z = xy$, $x = t^3$ and $y = t^2$.

Then $z = t^5$ so $z'(t) = 5t^4$. On the other hand, using the chain rule we get

$$z'(t) = y \cdot 3t^2 + x \cdot 2t = 3t^4 + 2t^4 = 5t^4.$$

Example: A continuous mapping $c : I \to \mathbb{R}^n$ of an interval I to $\mathbb{R}$ is called a curve in $\mathbb{R}^n$, $(n = 2, 3)$.
In what follows, we will assume that all the curves we have are actually differentiable, not just continuous. We will say what this means below.

An application to tangents of curves

Let us consider a curve $c(t)$ in $\mathbb{R}^3$. Each point on the curve will be given by a triple of coordinates which will depend on t. That is, the curve can be described by a triple of functions $(g(t), h(t), k(t))$. Saying that $c(t)$ is a differentiable function of t, means that each of $g(t), h(t), k(t)$ are differentiable functions from $\mathbb{R} \to \mathbb{R}$. If we write

$$c(t) = g(t)\mathbf{i} + h(t)\mathbf{j} + k(t)\mathbf{k}, \quad \text{then} \quad c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k},$$

represents its tangent or velocity vector at the point $c(t_0)$.

Tangents to curves on surfaces

So far our example has nothing to do with the chain rule. Suppose $z = f(x, y)$ is a surface, and $c(t) = (g(t), h(t), f(g(t), h(t)))$ lies on the $z = f(x, y)$. (Here we are assuming that $f : \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function!) Let us compute the tangent vector to the curve at $c(t_0)$. It is given by

$$c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k},$$

where $k(t) = f(g(t), h(t))$. Using the chain rule we see that

$$k'(t_0) = \frac{\partial f}{\partial x}(g(t_0), h(t_0))g'(t_0) + \frac{\partial f}{\partial y}(g(t_0), h(t_0))h'(t_0).$$

We can further show that this tangent vector lies on the tangent plane to the surface $z = f(x, y)$. Indeed we have already seen that the tangent plane has the equation

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0).$$

A normal vector to this plane is given by

$$\left(-\frac{\partial f}{\partial x}(x_0, y_0), -\frac{\partial f}{\partial y}(x_0, y_0), 1\right).$$

Thus, to verify that the tangent vector lies on the plane, we need only check that its dot product with normal vector is 0. But this is now clear.

Just to give a concrete example of what we are talking about, take a curve $(g(t), h(t))$ in the unit disc $x^2 + y^2 \leq 1$ in the xy plane. Then $\left(g(t), h(t), \sqrt{1 - g(t)^2 - h(t)^2}\right)$ lies on the upper hemisphere $z = \sqrt{1 - x^2 - y^2}$. For concreteness, we can take $I = \left[0, \frac{1}{\sqrt{2}}\right]$, $g(t) = t$ and $h(t) = t^2$.

MA 105: D3 Lecture 15

Ravi Raghunathan

Department of Mathematics

September 11, 2023

The Chain Rule

The Chain Rule and gradients

Problems involving the gradient

The total derivative for $f : U \to \mathbb{R}^n$

The Chain Rule

We now study the situation where we have composition of functions. We assume that $x, y : I \to \mathbb{R}$ are differentiable functions from some interval (open or closed) to $\mathbb{R}$. Thus the pair $(x(t), y(t))$ defines a function from I to $\mathbb{R}^2$. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is differentiable. We would like to study the derivative of the composite function $z(t) = f(x(t), y(t))$ from I to $\mathbb{R}$.

Theorem 27: With notation as above

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

For a function $w = f(x, y, z)$ in three variables the chain rule takes the form

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt}$$

Another application: Directional derivatives

Let $U \subset \mathbb{R}^3$ and let $f : U \to \mathbb{R}$ be differentiable. We want to relate the directional derivative to the gradient,

We consider the (differentiable) curve $c(t) = (x_0, y_0, z_0) + tv$, where $v = (v_1, v_2, v_3)$ is a unit vector. We can rewrite $c(t)$ as $c(t) = (x_0 + tv_1, y_0 + tv_2, z_0 + tv_3)$. We apply the chain rule to compute the derivative of the function $f(c(t))$:

$$\frac{d(f \circ c)}{dt} = \frac{\partial f}{\partial x} v_1 + \frac{\partial f}{\partial y} v_2 + \frac{\partial f}{\partial z} v_3.$$

But the left hand side is nothing but the directional derivative in the direction v. Hence,

$$\nabla_v f = \frac{d(f \circ c)}{dt} = \nabla f \cdot v.$$

Of course, the same argument works when $U \subset \mathbb{R}^2$ and f is a function of two variables.

The Chain Rule and Gradients

The preceding argument is a special case of a more general fact. Let $c(t)$ be any curve in $\mathbb{R}^3$. Then, clearly by the chain rule we have

$$\frac{d(f \circ c)}{dt} = \nabla f(c(t)) \cdot c'(t).$$

I leave this to you as a simple exercise.

Going back to the directional derivative, we can ask ourselves the following question. In what direction is f changing fastest at a given point (x_0, y_0, z_0)? In other words, in which direction does the directional derivative attain its largest value?

Using what we have just learnt, we are looking for a unit vector $v = (v_1, v_2, v_3)$ such that

$$\nabla f(x_0, y_0, z_0) \cdot v$$

is as large as possible

We rewrite the preceding dot product as

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\| \|v\| \cos\theta.$$

where θ is the angle between v and $\nabla f(x_0, y_0, z_0)$.
Since v is a unit vector this gives

$$\nabla f(x_0, y_0, z_0) \cdot v = \|\nabla f(x_0, y_0, z_0)\| \cos\theta.$$

The maximum value on the right hand side is obviously attained when $\theta = 0$, that is, when v points in the direction of ∇f. In other words the function is increasing fastest in the direction v given by ∇f. Thus the unit vector that we seek is

$$v = \frac{\nabla f(x_0, y_0, z_0)}{\|\nabla f(x_0, y_0, z_0)\|}.$$

Surfaces defined implicitly

So far we have only been considering surfaces of the form $z = f(x, y)$, where f was a function on a subset of $\mathbb{R}^2$. We now consider a more general type of surface S defined implicitly:

$$S = \{(x, y, z) \,|\, f(x, y, z) = b\},$$

where b is a constant. Most surfaces we have come across are usually described in this form, for instance, the sphere which is given by $x^2 + y^2 + z^2 = r^2$ or the right circular cone $x^2 + y^2 - z^2 = 0$. Let us try to understand what a tangent plane is more precisely.

If S is a surface, a tangent plane to S at a point $s \in S$ (if it exists) is a plane that contains the tangent lines at s to all curves passing through s and lying on S.

For instance, with the definition above, it is clear that a tangent plane to the right circular cone does not exist at the origin, since such a plane would have to contain the lines $x = 0, y = z$, $x = 0, y = -z$ and $y = 0, x = z$. Clearly no such plane exists.

If $c(t)$ is an curve on the surface S given by $f(x, y, z) = b$, we see that

$$\frac{d}{dt}(f \circ c)(t)) = 0.$$

On the other hand, by the chain rule,

$$0 = \frac{d}{dt}(f \circ c)(t)) = \nabla f(c(t)) \cdot c'(t).$$

Thus, if $s = c(t_0)$ is a point on the surface, we see that

$$\nabla f(c(t_0)) \cdot c'(t_0) = 0,$$

for every curve $c(t)$ on the surface S passing through t_0. Hence, if $\nabla f(c(t_0)) \neq 0$, then $\nabla f(c(t_0))$ is perpendicular to the tangent plane of S at s_0.

Let $\mathbf{r}$ denote the position vector

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

of a point $P = (x, y, z)$ in $\mathbb{R}^3$. Instead of writing $\|\mathbf{r}\|$, it is customary to write r. This notation is very useful. For instance, Newton's Law of Gravitation can be expressed as

$$\mathbf{F} = -\frac{GMm}{r^3} \cdot \mathbf{r},$$

where the mass M is assumed to be at the origin, $\mathbf{r}$ denotes the position vector of the mass m, G is a constant and $\mathbf{F}$ denotes the gravitational force between the two (point) masses.

A simple computation shows that

$$\nabla\left(\frac{1}{r}\right) = -\frac{\mathbf{r}}{r^3}.$$

Thus the gravitational force at any point can be expressed as the gradient of a function. Moreover, it is clear that

$$\left\|\nabla\left(\frac{1}{r}\right)\right\| = \left\|-\frac{\mathbf{r}}{r^3}\right\| = \frac{1}{r^2}.$$

Keeping our previous discussion in mind, we know that if $V = GMm/r$, $\mathbf{F} = \nabla V$.

What are the level surfaces of V? Clearly, r must be a constant on these level sets, so the level surfaces are spheres. Since $\mathbf{F}$ is a multiple of $-\mathbf{r}$, we see that F points towards the origin and is thus orthogonal to the sphere.

In order to make our notation less cumbersome, we introduce the notation f_x for the partial derivative $\frac{\partial f}{\partial x}$. The notations f_y and f_z will have the obvious meanings.

Since we know that the gradient of f is normal to the level surface S given by $f(x, y, z) = c$ (provided the gradient is non zero), it allows us to write down the equation of the tangent plane of S at the point $s = (x_0, y_0, z_0)$. The equation of this plane is

$$f_x(x_0, y_0, z_0)(x-x_0)+f_y(x_0, y_0, z_0)(y-y_0)+f_z(x_0, y_0, z_0)(z-z_0) = 0.$$

For the curve $f(x, y) = c$ we can similarly write down the equation of the tangent passing through (x_0, y_0):

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = 0.$$

Note that the fact that the gradient of f is normal to the level surface $f(x, y, z) = c$ is true only for implicitly defined surfaces. If the surface is given as $z = f(x, y)$, then we cannot simply take the gradient of f and make the same statement. We must first convert our explicit surface to the implicit surface S given by $g(x, y, z) = z - f(x, y) = 0$. Then ∇g will be normal to S.

Problems involving the gradient, continued

Exercise 3: Find $\nabla_u F(2,2,1)$ where $\nabla_u F$ denotes the directional derivative of the function $F(x,y,z) = 3x - 5y + 2z$ and u is the unit vector in the outward normal to the sphere $x^2 + y^2 + z^2 = 9$ at the point $(2,2,1)$.

Solution: The unit outward normal to the sphere $g(x,y,z) = 9$ at $(2,2,1)$ is given by

$$\frac{\nabla g(2,2,1)}{\|\nabla g(2,2,1)\|}.$$

We see that $\nabla g(2,1,1) = (4,4,2)$ so the corresponding unit vector is $(2,2,1)/3$.

To get the directional derivative we simply take the dot product of ∇F with u:

$$(3,-5,2) \cdot (2,2,1)/3 = -2/3$$

Comments: Also, there is no need to compute the gradient to find the normal vector to the sphere - it is obviously the radial vector at the point $(2,2,1)$!.

Problems involving the gradient, continued

Exercise 4: Find the equations of the tangent plane and the normal line to the surface

$$F(x, y, z) := x^2 + 2xy - y^2 + z^2 = 7$$

at $(1, -1, 3)$.

Solution: We first compute the gradient of F to get $\nabla F(x, y, z) = (2x + 2y, 2x - 2y, 2z)$. At $(1, -1, 3)$, this yields the vector $\lambda(0, 4, 6)$ which is normal to the given surface at $(1, -1, 3)$. By taking $\lambda = 1$, we see that the point $(1, 3, 9)$ also lies on the normal line so its equations are

$$x = 1, \frac{y + 1}{4} = \frac{z - 3}{6}.$$

The equation of the tangent plane is given by

$$4(y + 1) + 6(z - 3) = 0,$$

since it consists of all lines orthogonal to the normal and passing through the point $(1, -1, 3)$.

The proof of the chain rule

How does one actually prove the chain rule for a function $f(x, y)$ of two variables? We can write

$$f(x(t+h), y(t+h)) = f(x(t)+h[x'(t)+p_1(h)], y(t)+h[y'(t)+p_2(h)])$$

for functions p_1 and p_2 that go to zero as h goes to zero. Here we are simply using the differentiability of x and y as functions of t. Now we can write the right hand side as

$$f(x(t), y(t))+Df(x, y)(h[x'(t)+p_1(h)], h[y'(t)+p_2(h)])^T+p_3(h)h,$$

(where T denotes transpose, so we get a column vector) by using the differentiability of f, for some other function $p_3(h)$ which goes to zero as h goes to zero (you may need to think about this step a little).

Remember that ∇f is the same as Df, just written as a row vector rather than as a matrix. Multiplying a 1×2 matrix by a 2×1 column vector is the same as taking the dot product of the two, thinking of both of them as row vectors.

It is not too hard to figure out what $p_3(h)$ above is. This gives

$$f(x(t+h), y(t+h)) - f(x(t), y(t)) - f_x x'(t)h - f_y y'(t)h = p(h)h,$$

for some function $p(h)$ with $\lim_{h \to 0} p(h)$.

Functions from $\mathbb{R}^m \to \mathbb{R}^n$

So far we have only studied functions whose range was a subset of $\mathbb{R}$. Let us now allow the range to be $\mathbb{R}^n$, $n = 1, 2, 3, \ldots$. Can we understand what continuity, differentiability etc. mean?

Let U be a subset of $\mathbb{R}^m$ $(m = 1, 2, 3, \ldots)$ and let $f : U \to \mathbb{R}^n$ be a function. If $x = (x_1, x_2, \ldots, x_m) \in U$, $f(x)$ will be an n-tuple where each coordinate is a function of x. Thus, we can write $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$, where each $f_i(x)$ is a function from U to $\mathbb{R}$.

Functions which take values in $\mathbb{R}$ are called scalar valued functions, which functions which take values in $\mathbb{R}^n$, $n > 1$ are usually called vector valued functions.

For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we define

$$\|x\|_n = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Sometimes we will omit the subscript and write $\|x\|$ for $\|x\|_n$.

Continuity of vector valued functions

The definition of continuity is exactly the same as before.
Definition: The function f is said to be continuous at a point $c \in U$ if

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

How does one define the limit on the left hand side? The function

f takes values in $\mathbb{R}^n$, so its limit must be a point in $\mathbb{R}^n$, say $l = (l_1, l_2, \ldots, l_n)$.

Definition: We say that $f(x)$ tends to the limit l if given any $\epsilon > 0$, there exists $\delta > 0$ such that if $0 < \|x - c\|_m < \delta$, then

$$\|f(x) - l\|_n < \epsilon.$$

You can easily prove the following theorem yourself:

Theorem: The function $f : U \to \mathbb{R}^n$ is continuous if and only if each of the functions $f_i : U \to \mathbb{R}$, $1 \leq i \leq n$, is continuous.

The derivative for $f : U \to \mathbb{R}^n$

We now define the derivative for a function $f : U \to \mathbb{R}^n$, where U is a subset of $\mathbb{R}^m$.

The function f is said to be differentiable at a point x if there exists an $n \times m$ matrix $Df(x)$ such that

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

Here $x = (x_1, x_2, \ldots, x_m)$ and $h = (h_1, h_2, \ldots h_m)$ are vectors in $\mathbb{R}^m$.

The matrix $Df(x)$ is usually called the total derivative of f. It is also referred to as the Jacobian matrix. What are its entries?

From our experience in the 2×1 case we might guess (correctly!) that the entries will be the partial derivatives.

Here is the total derivative or the derivative matrix written out fully.

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_m}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_m}(x) \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \frac{\partial f_n}{\partial x_1}(x) & \frac{\partial f_n}{\partial x_2}(x) & \cdots & \frac{\partial f_n}{\partial x_m}(x) \end{pmatrix}$$

In the 2×2 case we get

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) \end{pmatrix}.$$

As before, the derivative may be viewed as a linear map, this time from $\mathbb{R}^m$ to $\mathbb{R}^n$ (or, in the case just above, from $\mathbb{R}^2$ to $\mathbb{R}^2$).

Rules for the total derivative

Just like in the one variable case, it is easy to prove that

$$D(f+g)(x) = Df(x) + Dg(x).$$

Somewhat harder, but only because the notation gets more cumbersome, is the Chain rule:

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x),$$

where $\circ$ on the right hand side denotes matrix multiplication. Theorem 26 holds in this greater generality - a function from $\mathbb{R}^m$ to $\mathbb{R}^n$ is differentiable at a point x_0 if all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ $1 \leq i \leq n$, $1 \leq j \leq m$, are continuous in a neighborhood of x_0 (define a neighborhood of x_0 in $\mathbb{R}^m$!).

MA 105: D3 Lecture 16

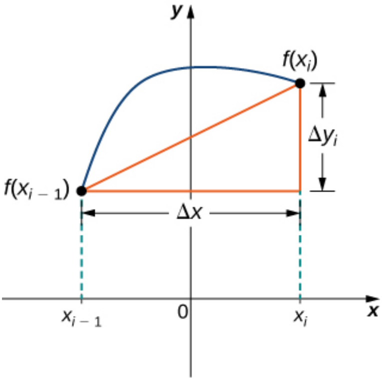
Ravi Raghunathan

Department of Mathematics

September 11, 2023

Multivariable calculus: Supplementary exercises (mostly from Stewart)

Arc Length



The formula for arc length

Let us denote the arc length of the curve $y = f(x)$ by S. The length of any given hypotenuse in the previous slide is given by the Pythagorean Theorem: $\sqrt{\Delta x^2 + \Delta y^2}$.
Intuitively, the sum of the lengths of the n hypotenuses appears to approximate S:

$$S \sim \sum_{i=1}^{n} \sqrt{\Delta x_i^2 + \Delta y_i^2} = \sum_{i=1}^{n} \sqrt{1 + \left(\frac{\Delta y_i}{\Delta x_i}\right)^2} \Delta x_i,$$

where "$\sim$" means approximately equal. We can use this idea to define the arc length as

$$S := \lim_{\Delta x_i \to 0} \sum_{i=1}^{\infty} \sqrt{1 + \left(\frac{\Delta y_i}{\Delta x_i}\right)^2} \Delta x_i = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx,$$

provided this limit exists (in particular, we demand that the limit is a finite number).

Exercise 4.10.(ii) Find the length of the curve

$$y(x) = \int_0^x \sqrt{\cos 2t}\, dt,\ 0 \leq x \leq \pi/4.$$

Solution: The formula for the arc length of a curve $y = f(x)$ between the points $x = a$ and $x = b$ is given by

$$\int_a^b \sqrt{1 + [f'(x)]^2} dx.$$

For the problem at hand this gives

$$\int_0^{\frac{\pi}{4}} \sqrt{1 + \cos 2x} dx = \sqrt{2} \int_0^{\pi/4} \cos(x) dx = 1.$$

Rectifiable curves

Not all curves have finite arc length! Here is an example of a curve with infinite arc length.

Example: Let $\gamma : [0,1] \to \mathbb{R}^2$ be the curve given by $\gamma(t) = (t, f(t))$, where

$$f(t) = \begin{cases} t\cos\left(\frac{\pi}{2t}\right), & \text{if } t \neq 0 \\ 0 \text{ if } t = 0 \end{cases}$$

If
http://math.stackexchange.com/questions/296397/nonrectifiable-curve
is correct, you should be able to check that this curve has infinite arc length. Try it as an exercise.

Notice that the curve above is given by a continuous function. Curves for which the arc length S is finite are called rectifiable curves. You can easily check that the graphs of piecewise $\mathcal{C}^1$ functions are rectifiable.

Things can get even stranger

In fact, there exist space filling curves, that is curves $\gamma : [0,1] \to [0,1] \times [0,1]$ which are continuous and surjective (but it is not injective!). Obviously the graph of this curve “fills up” the entire square. Such curves are not rectifiable (can you prove this - see
`https://en.wikipedia.org/wiki/Peano_curve`
for an example.

The existence of such curves should make you question whether your intuitive notion of dimension actually has any mathematical basis. If a line segment can be mapped continuously onto a square, is it reasonable to say that they have different dimensions? After all, this means we can describe any point on the square using just one number.

We will answer this question (without a proof) later in this course. We will also come back to arc length of a curve when studying multivariable calculus.

Comments

General comments based on today's interaction in class. To show that a limit exists you must either use the $\varepsilon - \delta$ definition or use the rules for limits when applicable. Choosing particular curves and approaching the limit point along these curves is a good strategy for showing a limit does not exist (or that a function is not continuous). It cannot be used to show limits exist.

Using the rules for limits will not work if the function you are given has a denominator which goes to 0 as (x, y) approaches the limit point. In this case, the only way to show that a limit exists to use some kind of inequality which shows that the numerator goes to zero at least as fast as the denominator.

The natural domain

Exercise 1. What is the natural domain of the following functions (try to describe the domain geometrically):
(a) $f(x,y) = \frac{\sqrt{x+y+1}}{x-1}$, (b) $f(x,y) = x\ln(y^2 - x)$.

Solution. (a)

$$D = \{(x,y) \,|\, x+y+1 \geq 0, x \neq 1\}.$$

This is the set of points that lie above the line $x+y+1=0$ but not on the line $x=1$.

(b)

$$D = \{(x,y) \,|\, x < y^2\}.$$

This this the set of points to the left of the parabola $x = y^2$.

Limits and Continuity

Exercise 2. Determine if the following limits exist. If they exist find them.

(a) $\lim_{(x,y)\to(1,0)} \ln\left(\frac{1+y^2}{x^2+xy}\right)$,

(b) $\lim_{(x,y)\to(0,0)} \frac{x^2ye^x}{x^2+4y}$,

(c) $\lim_{(x,y)\to(0,0)} \frac{x^2\sin y}{x^2+2y^2}$,

(d) $\lim_{(x,y)\to(1,-1)} e^{xy}\cos(x+y)$

(e) $\lim_{(x,y,z))\to(0,0,0)} \frac{xy+yz}{x^2+y^2+z^2}$.

Exercise 2 (a)

Solution. Determine if the following limits exist. If they exist find them.

(a) $\lim_{(x,y)\to(1,0)} \ln\left(\frac{1+y^2}{x^2+xy}\right)$:

The limit of a quotient is the quotient of the limits and $x \to \ln x$ is continuous. It follows that the limit exists.

Exercise 2 (b)

$\lim_{(x,y)\to(0,0)} \frac{x^2 \sin y}{x^2+2y^2}$:

If (x, y) such that $x^2 + y^2 < 1$,

$$|x^2 \sin y| < x^2|y| < (x^2 + 2y^2)|y|.$$

Hence, the quotient

$$\left|\frac{x^2 \sin y}{x^2 + 2y^2}\right| < |y|.$$

Thus, if $x^2 + y^2 < \delta = \varepsilon^2$, $|y| < |\varepsilon$, so $\left|\frac{x^2 \sin y}{x^2+2y^2} - 0\right| < \varepsilon$.

This shows that the limit is 0.

Exercise 2 (c)

(b) $\lim_{(x,y)\to(0,0)} \frac{x^2ye^x}{x^2+4y}$:

Let $f(x,y) = \frac{x^2ye^x}{x^2+4y}$. One solution offered in class was to say that the function is not defined for points such that $4y - x^2 = 0$. This is correct (although we did not pursue this fully in class). After all, if the function is not defined at points arbitrarily close to $(0,0)$, one can argue that the inequality $|f(x,y) - \ell| < \varepsilon$ cannot be satisfied for $\|x\| < \delta$ for any $\delta > 0$.

If this line of argument makes you uncomfortable, we can use the strategy from the class. The limit along the line $y = x$ is 0. Now take points (x,y) that lie on a curve close to the curve $x^2 + 4y = 0$. For instance, we can look at the curve $x^2 + 4y = x^2y$ or $y = \frac{x^2}{x^2-4}$. Along this curve $f(x,y) = e^x$, so the limit is simply 1 as $x \to 0$ (one of you gave me a similar curve in class, but I think that my example is a little simpler).

Thus, this limit does not exist.

Exercise 2 (d)

$\lim_{(x,y)\to(1,-1)} e^{xy}\cos(x+y)$

Again, use the rules for limits!

Exercise 2 (e)

(e) $\lim_{(x,y,z))\to(0,0,0)} \frac{xy+yz}{x^2+y^2+z^2}$.

Let $x = y$ and $z = y$. Then the quotient is $2/3$.
Let $x = y$ and $z = 0$. Then, the quotient is $1/2$.

Hence, the limit does not exist.

MA 105: Review of Taylor series and integration

Ravi Raghunathan

Department of Mathematics

September 9, 2023

General Advice

1. Concentrate on understanding the statements of the theorems. You will not be asked to reproduce long proofs.

2. When trying to understand a definition, make sure you know plenty of examples.

3. When trying to understand a theorem, make sure you know counter-examples to the conclusion of the theorem when you drop some of the hypotheses.

4. In general, the statement of the theorem is more important than its proof. And examples are more important than theorems!

Taylor's theorem

Taylor's theorem: Know how to compute the Taylor polynomials. Know the form of the remainder term. Recall that there are smooth functions for which the Taylor series about a point converges but does not converge to the function ($e^{-1/x}$).

Theorem 19: Let I be an open interval and suppose that $[a, b] \subset I$. Suppose that $f \in \mathcal{C}^n(I)$ ($n \geq 0$) and suppose that $f^{(n)}$ is differentiable on I. Then there exists $c \in (a, b)$ such that

$$f(b) = P_n(b) + \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1},$$

where

$$P_n(b) = f(a) + f^{(1)}(a)(b-a) + \frac{f^{(2)}(a)}{2!}(b-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(b-a)^n.$$

denotes the Taylor polynomial of degree n at a.

The term $R_n(b) = \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1}$ is sometimes called the remainder term.

Taylor series

If we have a smooth or $\mathcal{C}^\infty$-function f (that is, a function for which all derivatives exist) we can form its Taylor series about the point a.

$$T_f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

For sufficiently nice functions f,

1. $T_f(x)$ will be a convergent series for every vaule of x in some interval $(a-r, a+r)$, $r > 0$.
2. $T_f(x) = f(x)$ for all $x \in (a-r, a+r)$.

In general, neither 1. nor 2. need hold.

Exercise 1.

Find the first three terms of the Taylor series of the function $1/x^2$ at 1.

Solution: If the Taylor series of the function f at $x = a$ is
$$\sum_{n=0}^{\infty} b_n(x-a)^n, \text{ then } b_n = \frac{f^{(n)}(a)}{n!}.$$

Using these notations, for $f(x) = 1/x^2$ and $a = 1$, we get $b_0 = 1$, $b_1 = -2$ and $b_2 = 3$.

Exercise 2.

True or False (justify): Let $P_2(x)$ denote the Taylor polynomial of degree 2 about the point $a = 0$ for the function $f(x) = \log(1 + x)$. The inequality $|f(x) - P_2(x)| < 0.05$ holds for all x in $[0, 1/2]$.

True.

Applying Taylor's theorem for $n = 2$, we know that $R_2(x) = \frac{f^{(3)}(c)}{3!}(x - a)^3$ for some point $c \in [0, 1/2]$.

In our case $f(x) = \log(1 + x)$, so $f^{(3)}(c) = 2/(1 + c)^3$. If $c \in (0, 1/2)$, clearly $2/(1 + c)^3 < 2$.

If $x \in [0, 1/2]$ the maximum value that $x^3/3!$ can take is when $x = 1/2$.

Thus $0 \leq R_2(x) < 1/24 < 0.05$.

Integration

Remember what partitions and tagged partitions are.

Recall the definitions of the (Darboux) lower sums, upper sums, lower integrals, upper integrals and Riemann sums.

Learn all three definitions of the Riemann integral.

Basic fact: Bounded functions on closed intervals with at most a finite number of discontinuities are Riemann/Darboux integrable.

The Fundamental Theorem of calculus.

The basic definitions

Definition: Given a closed interval $[a,b]$, a partition P of $[a,b]$ is simply a collections of points

$$P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}.$$

If we are also giventags $t = \{t_0, \ldots, t_{n-1}\}$ with $t_j \in I_j = [x_{j-1}, x_j]$, the pair (P,t) is called a tagged partition.

Definition: A partition P' is called a refinement of a partition P if $P \subseteq P'$ (every point of P is also a point of P').

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x) \quad \text{and} \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \quad 1 \le i \le n$$

Defintion: We define the (Darboux) Lower sum and Upper sum by

$$L(f,P) = \sum_{j=1}^{n} m_j (x_j - x_{j-1}) \text{ and } U(f,P) = \sum_{j=1}^{n} M_j (x_j - x_{j-1}).$$

We always have $L(f,P_1) \le U(f,P_2)$ for any two partitions P_1 and P_2 of $[a,b]$ and if $P \subseteq P'$,

$$L(f,P) \le L(f,P') \le U(f,P') \le U(f,P).$$

The Darboux integrals

We now define the lower Darboux integral of f by

$$L(f) = \sup\{L(f,P)\colon P \text{ is a partition of } [a,b]\},$$

where the supremum is taken over all partitions of $[a,b]$.

and similarly the upper Darboux integral of f by

$$U(f) = \inf\{U(f,P)\colon P \text{ is a partition of } [a,b]\},$$

and again the infimum is over all partitions of $[a,b]$.

If $L(f) = U(f)$, then we say that f is Darboux-integrable and define

$$\int_a^b f(t)\,dt := U(f) = L(f).$$

This common value of the two integrals is called the Darboux integral.

Riemann integration

For a tagged partition (P, t), we define the Riemann sum to be

$$R(f, P, t) = \sum_{j=1}^{n} f(t_j)(x_j - x_{j-1}).$$

Definition 1: A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\delta > 0$ such that

$$|R(f, P, t) - R| < \epsilon,$$

whenever $\|P\| < \delta$. In this case R is called the Riemann integral of the function f on the interval $[a, b]$.

Definition 2: A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists a partition P such that for every tagged refinement (P', t') of P ~~with~~ $\cancel{\|P'\| < \delta}$,

$$|R(f, P', t') - R| < \epsilon.$$

(Recall that the struck out portion was initially part of the definition above, but later I pointed out that it was not necessary.)

Exercise 3

3. For the function $f(x) = 3x^2$ and the partition

$$P_n = \left\{ 0 < \frac{1}{n} < \frac{2}{n} < \ldots < \frac{n-1}{n} < 1 \right\}$$

of $[0, 1]$ find the lower sum, $L(f, P_n)$, upper sum, $U(f, P_n)$. Compute $\sup_n L(f, P_n)$ and $\inf_n U(f, P_n)$.

Solution:

$$L(f, P_n) = \sum_{i=0}^{n-1} 3\frac{i^2}{n^2}\frac{1}{n} = 3\frac{1}{n^3}\frac{n(n-1)(2n-1)}{6}$$

So

$$L(f, P_n) = \frac{2n^2 - 3n + 1}{2n^2} \quad \text{and} \quad U(f, P_n) = \frac{2n^2 + 3n + 1}{2n^2}$$

and

$$\sup_n L(f, P_n) = 1 \quad \text{and} \quad \inf_n U(f, P_n) = 1.$$

Exercise 4

Evaluate $\lim_{n\to\infty}\sum_{i=1}^{n}\frac{n}{i^2+n^2}$ by identifying it as a Riemann sum for a certain continuous function on a certain interval and with respect to a certain partition.

Solution: We observe that

$$\sum_{i=1}^{n}\frac{n}{i^2+n^2}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(i/n)^2+1}.$$

Thus, the given sum is the Riemann sum for the function $\dfrac{1}{x^2+1}$ over the interval $[0,1]$ with respect to the partition $0<\dfrac{1}{n}<\dfrac{2}{n}<\cdots<\dfrac{n-1}{n}<1.$

Since the function $1/(1+x^2)$ is continuous on $[0,1]$, it is Riemann integrable.

Hence the limit of the given sum is $\displaystyle\int_0^1\frac{1}{x^2+1}dx=\pi/4.$

The Fundamental Theorem of Calculus

If f is continuous in $[a, b]$ and differentiable in (a, b), and we define

$$F(x) = \int_a^x f(t)dt,$$

then $F'(x) = f(x)$. This is (part I of) the Fundamental Theorem of Calculus (FTC).

Suppose $G(x) = \int_{p(x)}^{q(x)} f(t)dt$, where the range of $p(x)$ and $q(x)$ lies in $[a, b]$. We can write

$$G(x) = \int_a^{q(x)} f(t)dt - \int_a^{p(x)} f(t)dt = G_1(x) + G_2(x).$$

We can write $G_1(x) = (F \circ q)(x)$ and $G_2(x) = (F \circ p)(x)$. Using the Chain rule and the FTC, we see that

$$G'(x) = f(q(x))q'(x) - f(p(x))p'(x).$$

More exercises:

1. Write the n-th term of the Taylor series around 0 for
(a) $\sin \pi x$, (b) $\cosh x$, (c) $(1 - x^2)^{1/2}$, (d) $(1 + x)^{1/4}$.

2. For the functions $f(x)$ and the point a given below write down the Taylor series around a.
(a) $f(x) = \sin x, a = \pi/2$, (b) $f(x) = x - x^3, a = -2$,
(c) $f(x) = \sqrt{x}, a = 16$, (d) $f(x) = \ln x, a = 2$.

3. In 2. (d) above, write down an expression for $R_n(x)$. Show that $\lim_{n\to\infty} R_n(x) = 0$ if $x \in (2, 4)$. Is there a larger interval in which the Taylor series converges?

4. Let $P_n(x)$ denote the Taylor polynomial of degree n for the function e^x around the point 0. Find an n so that $P_n(-1)$ calculates e^{-1} to five decimal place accuracy?

5. Use 1. (c) and term by term integration of power series to get the Taylor series for $\arcsin x$.

Still more exercises

1. For $f(x) = x$, the partition P_n, and the tags $T_n = \{1/2n, 3/2n, \dots, 2n - 1/2n\}$, write down $R(f, P_n, T_n)$.

2. Evaluate $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{i^k}{n^{k+1}}$ by identifying it as a Riemann sum for a certain continuous function on a certain interval and with respect to a certain partition.

3. Find the derivative of $h(x) = \int_0^{\sqrt{x}} \frac{z^2}{z^4+1} dz$.

4. Find the derivative of $g(x) = \int_0^{\tan x} (\sqrt{t + \sqrt{t}}) dt$.

5. True or false:

$$\left(\frac{\pi}{22}\right)\cos\left(\frac{\pi}{22}\right) + \left(\frac{2\pi}{11}\right)\cos\left(\frac{5\pi}{22}\right) + \left(\frac{2\pi}{11}\right)\cos\left(\frac{9\pi}{22}\right) +$$
$$\left(\frac{\pi}{22}\right)\cos\left(\frac{5\pi}{11}\right) < \left(\frac{\pi}{26}\right) + \left(\frac{3\pi}{13}\right)\cos\left(\frac{\pi}{26}\right) + \left(\frac{3\pi}{13}\right)\cos\left(\frac{7\pi}{26}\right)$$