MA 105 : Calculus

Lecture Notes (upto Mid-Sem.) for D1 and D4

Sandip Singh

Department of Mathematics

Autumn 2023, IIT Bombay, Mumbai

Table of contents:

Aims of the course

First, welcome to IIT Bombay.

- To briefly review the calculus of functions of one variable and to teach the calculus of functions of several variables.

For details about the syllabus, tutorials, assignments, quizzes, exams and procedures for evaluation please refer to the course booklet which is available on moodle: https://moodle.iitb.ac.in

The emphasis of this course will be on the underlying ideas and methods rather than intricate problem solving (though there will be some of that too). The aim is to get you to think about calculus, in particular, and mathematics in general.

Syllabus

- Convergence of sequences and series, power series.
- Review of limits, continuity, differentiability.
- Mean value theorem, Taylor's theorem, maxima and minima.
- Riemann integrals, fundamental theorem of calculus, improper integrals, applications to area, volume.
- Partial derivatives, gradient and directional derivatives, chain rule, maxima and minima, Lagrange multipliers.
- Double and triple integration, Jacobians and change of variables formula.
- Parametrization of curves and surfaces, vector fields, line and surface integrals.
- Divergence and curl, theorems of Green, Gauss, and Stokes.

Sequences

Definition: A sequence in a set X is a function $f : \mathbb{N} \to X$, that is, a function from the set of 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 occasion to consider sequences of functions sometimes. In later mathematics courses X may be the set of complex numbers $\mathbb{C}$, vector spaces (whatever those maybe), the set of continuous functions on an interval $\mathcal{C}([a, b])$ or other sets of functions.

Rather than denoting a sequence by a function, it is often customary to describe a sequence by listing the first few elements

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

or, more generally by describing the n^{th} term a_n.

Note that we write a_n rather than $a(n)$. 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 f is just the identity function).
2. $a_n = \frac{1}{n}$ (here we can take $X = \mathbb{Q} \subset \mathbb{R}$ if we want, where $\mathbb{Q}$ denotes the set of rational numbers, or we can take $X = \mathbb{R}$ itself).
3. $a_n = \sin\left(\frac{1}{n}\right)$ (here the values taken by a_n are irrational numbers, so it is best to take $X = \mathbb{R}$).
4. $a_n = \frac{n!}{n^n}$.
5. $a_n = n^{1/n}$.
6. $s_n = \sum_{i=0}^{n} r^i$, for some r such that $0 \leq r < 1$.
7. $a_n = \left(n^2, \frac{1}{n}\right)$ (here $X = \mathbb{R}^2$ or $X = \mathbb{Q}^2$).
8. $f_n(x) = \cos nx$ (here X is the space of continuous functions on any interval $[a, b]$ or even on $\mathbb{R}$).
9. $s_n(x) = \sum_{j=0}^{n} \frac{x^j}{j!}$, or writing it out $s_n(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$. Once again X is a space of functions, for instance the space of continuous functions on $\mathbb{R}$.

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 ($a_n = n$) is a monotonically increasing sequence, Example 2 ($a_n = 1/n$) is a monotonically decreasing sequence, while Example 3 ($a_n = \sin\left(\frac{1}{n}\right)$) is also monotonically decreasing. How about Examples 4 and 5?

In Example 4 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 5 ($a_n = n^{1/n}$), 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 monotonically decreasing sequence, that is, the sequence becomes monotonically decreasing after some stage. One can similarly define eventually monotonically increasing sequences.

For any fixed non-negative value of r, Example 6 ($s_n = \sum_{j=0}^{n} r^j$) gives a monotonically increasing sequence, while for any fixed non-negative value of x, the sequence $s_n(x) = \sum_{j=0}^{n} \frac{x^j}{j!}$ in Example 9 also gives a monotonically increasing sequence.

Preliminaries

While all of you are familiar with limits, most of you have probably not worked with a rigorous definition. We will be more interested in limits of functions (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 rigorous 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, if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - \ell| < \epsilon$$

whenever $n > N$.

This is what we mean when we write

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

Equivalently, we say that the sequence $\{a_n\}_{n=1}^{\infty}$ converges to a limit 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

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. Recall that for $x > 0$, x^α is defined as $e^{\alpha \log x}$.
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,

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

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.

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

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

(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 that 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 $|a_n - \ell| < 1$ for all $n > N$.

It follows from $|a_n| = |(a_n - \ell) + \ell|$ and the triangle inequality that $|a_n| \leq |a_n - \ell| + |\ell| < 1 + |\ell|$ for all $n > N$. Now, let

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

and let $M = \max\{M_1, |\ell| + 1\}$. Then, $M > 0$ and $|a_n| \leq M$ for all $n \in \mathbb{N}$. $\square$

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 $\lim_{n\to\infty} a_n = \ell_1$ and $\lim_{n\to\infty} b_n = \ell_2$. We need to show that $\lim_{n\to\infty} a_n b_n = \ell_1 \ell_2$.

Let $\epsilon > 0$ be an arbitrary real number. We need to show that there exists $N \in \mathbb{N}$ such that $|a_n b_n - \ell_1 \ell_2| < \epsilon$, whenever $n > N$. Notice that

$$\begin{aligned} |a_n b_n - \ell_1 \ell_2| &= |a_n b_n - a_n \ell_2 + a_n \ell_2 - \ell_1 \ell_2| \\ &= |a_n (b_n - \ell_2) + (a_n - \ell_1) \ell_2| \\ &\leq |a_n||b_n - \ell_2| + |a_n - \ell_1||\ell_2|, \end{aligned}$$

where the last inequality follows from the triangle inequality.

So in order to guarantee that the left hand side is less than ϵ, 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 on right hand side 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 $M > 0$ such that $|a_n| \leq M$ for all $n \in \mathbb{N}$.

Given the positive real numbers $\frac{\epsilon}{2|\ell_2|+1}$ and $\frac{\epsilon}{2M}$, there exist N_1 and N_2 such that

$$|a_n - \ell_1| < \frac{\epsilon}{2|\ell_2| + 1}, \quad n \geq N_1 \quad \text{and} \quad |b_n - \ell_2| < \frac{\epsilon}{2M}, \quad n \geq N_2.$$

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

$$|a_n||b_n - \ell_2| \leq M \cdot \frac{\epsilon}{2M} = \frac{\epsilon}{2} \quad \text{and} \quad |a_n - \ell_1||\ell_2| \leq \frac{\epsilon}{2|\ell_2| + 1} \cdot |\ell_2| < \frac{\epsilon}{2}.$$

Now, it follows that

$$|a_n b_n - \ell_1 \ell_2| \leq |a_n||b_n - \ell_2| + |a_n - \ell_1||\ell_2| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

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 $a_n \leq M$ (resp. $a_n \geq m$) for some $M \in \mathbb{R}$ (resp. $m \in \mathbb{R}$).

A sequence that is bounded both above and below is obviously bounded (maximum of $|m|$ and $|M|$ works as a bound for $|a_n|$).

Theorem 3: A monotonically 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 $a_n < M_1$ for all n, then $M \leq M_1$. In other words, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $a_N > M - \epsilon$.

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

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

Since a_n is positive for all n,

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

On the other hand,

$$\frac{1}{2}\left(a_n + \frac{2}{a_n}\right) \ge \sqrt{2} \quad \text{(Why is this true?} \quad A.M. \ge G.M.\text{)}$$

so $a_{n+1} \ge \sqrt{2}$ for all $n \ge 1$ 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.

More remarks on limits

Exercise 1. What do you think is the limit of the above sequence (Refer to the supplement to Tutorial 1)?

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.

Theorem 3 can be proved by using the fact that the set of real numbers has the least upper bound property: every nonempty subset of real numbers having an upper bound has the least upper bound. You are now encouraged to prove Theorem 3 using the ϵ - N definition of convergence.

The proof of the least upper bound property of the set of real numbers more or less involves understanding what a real number is. It is related to the notion of Cauchy sequences (again, refer to the supplement to Tutorial 1).

Cauchy sequences

As we saw last time, 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, things are slightly better since we only need to bound the sequence.

There is another very useful notion which allows us to decide whether the sequence converges by looking only at the terms 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 $m, n > N$.

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

Cauchy sequences: Some Remarks

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

One can easily check the converse:

Theorem 5: Every convergent sequence is Cauchy.

Remark 2: 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, subsets 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) ($a_1 = 3/2$ and $a_{n+1} = (a_n + 2/a_n)/2$) 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 $\mathbb{R}$ and more remarks on limits

A set in which every Cauchy sequence converges is called a complete set. Thus, Theorem 4 is sometimes rewritten as

Theorem 4': The set of real numbers is complete.

An important remark: If we change finitely many terms of a sequence, it does not affect the convergence and boundedness properties of a sequence.

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.

Series

Given a sequence a_n of real numbers, we can construct a new sequence, namely the sequence of partial sums s_n:

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

More precisely, we have the sequence

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

which is called the n-th partial sum of the series $\sum_{k=1}^{\infty} a_k$.

For example, we can define $a_n = r^{n-1}$, for some $r \in \mathbb{R}$ and in this case the series $\sum_{k=1}^{\infty} a_k$ is the geometric progression $\sum_{k=0}^{\infty} r^k$ for which the n-th partial sum $s_n = \sum_{k=0}^{n-1} r^k$.

We say that the series $\sum_{k=1}^{\infty} a_k$ converges if the sequence of the corresponding n-th partial sum converges.

When does the series $\sum_{k=0}^{\infty} r^k$ converge?

Infinite series - a rigorous treatment

Let us recall what we mean when we write, for $|r| < 1$,

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

Another way of writing the same expression is

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

The precise meaning is the following. Form the partial sums

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

These partial sums $s_1, s_2, \ldots s_n, \ldots$ form a sequence and by $\sum_{k=1}^{\infty} ar^{k-1} = a/(1-r)$, 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

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

if n is chosen large enough. The case $a = 0$ is trivial, so we assume $a \neq 0$. Since $\lim_{n\to\infty} r^n = 0$ (as $|r| < 1$), there exists $N \in \mathbb{N}$ such that $|r|^n < (1-r)\epsilon/|a|$ for all $n > N$, so for this N, if $n > N$,

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

This shows that the geometric series converges to the given expression. $\square$

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 $| \quad |$ 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 $a(n) = (a(n)_1, a(n)_2)$ in $\mathbb{R}^2$ is said to converge to a point $\ell = (\ell_1, \ell_2)$ (in $\mathbb{R}^2$) if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\sqrt{(a(n)_1 - \ell_1)^2 + (a(n)_2 - \ell_2)^2} < \epsilon$$

whenever $n > N$. A similar definition can be made in $\mathbb{R}^3$ using the distance function on $\mathbb{R}^3$.

Theorems 2 (the Sandwich Theorem) and 3 (about monotonic sequences) don't really make sense for $\mathbb{R}^2$ or $\mathbb{R}^3$ because there is no ordering on these sets, that is, it doesn't really make sense to ask if one point on the plane or in space is less than the other.

The completeness of other spaces

Theorem 4, however, makes perfect sense - one can define Cauchy sequences in $\mathbb{R}^2$ and $\mathbb{R}^3$ exactly as before, using the distance functions - and indeed, remains valid in $\mathbb{R}^2$ and $\mathbb{R}^3$.

So $\mathbb{R}^2$ and $\mathbb{R}^3$ are complete sets too (but $\mathbb{Q}^2$ and $\mathbb{Q}^3$ are not).

Finally, to emphasis that only the notion of distance matters for such questions we can define a distance function on $X = \mathcal{C}([a, b])$, the set of continuous functions from $[a, b]$ to $\mathbb{R}$, as follows:

$$\text{dist}(f, g) = \sup_{x \in [a,b]} |f(x) - g(x)|.$$

Then, Cauchy and convergent sequences in X can be defined as before, and we can prove (maybe next semester) that X is complete.

Remark: We **do not** define the distance between the continuous functions f and g as $\inf_{x \in [a,b]} |f(x) - g(x)|$ to make sure that the distance between two **distinct** functions is **nonzero** and it is zero only when they are identical.

The rigorous definition of a limit of a function

Since we have already defined the limit of a sequence rigorously, 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

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

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

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

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

This is just the rigorous 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 at a point 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 sequences and can be proved in almost exactly the same way.

If $\lim_{x\to x_0} f(x) = \ell_1$ and $\lim_{x\to x_0} g(x) = \ell_2$, then

1. $\lim_{x\to x_0} f(x) \pm g(x) = \ell_1 \pm \ell_2$.
2. $\lim_{x\to x_0} f(x)g(x) = \ell_1\ell_2$.
3. $\lim_{x\to x_0} f(x)/g(x) = \ell_1/\ell_2$, provided $\ell_2 \neq 0$

As before, implicit in the formulæ is the fact that 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 $\lim_{x \to x_0} f(x) + g(x) = \ell_1 + \ell_2$. Let $\epsilon > 0$ be arbitrary. Since $\lim_{x \to x_0} f(x) = \ell_1$ and $\lim_{x \to x_0} g(x) = \ell_2$, there exist δ_1, δ_2 in $\mathbb{R}$ such that

$$|f(x) - \ell_1| < \frac{\epsilon}{2} \quad (\text{for } 0 < |x - x_0| < \delta_1)$$

$$\text{and} \quad |g(x) - \ell_2| < \frac{\epsilon}{2} \quad (\text{for } 0 < |x - x_0| < \delta_2).$$

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

$$\begin{aligned} |f(x) + g(x) - (\ell_1 + \ell_2)| &= |f(x) - \ell_1 + g(x) - \ell_2| \\ &\leq |f(x) - \ell_1| + |g(x) - \ell_2| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

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 $f(x) \to \ell_1$, $g(x) \to \ell_2$ and $h(x) \to \ell_3$ 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

$$\ell_1 \leq \ell_2 \leq \ell_3.$$

As before, we have a second version.

Theorem 6: Suppose $\lim_{x \to x_0} f(x) = \lim_{x \to x_0} h(x) = \ell$ 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

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

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 $|g(x)| \leq M$ 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 $|g(x)| \leq M$.

Let $\epsilon > 0$. We would like to show that there exists a $\delta > 0$ such that

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

if $0 < |x - c| < \delta$.

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

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 2: 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 a half-open interval $I_1 = (a, a + \eta)$ such that $f(x)$ is bounded on I_1. Similarly if $c = b$, there exists a half-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 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 $(c - \eta, c + \eta)$. Now apply part (ii) to this interval to deduce that (iii) is true.

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.

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 $M \in \mathbb{R}$ such that

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

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

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

or, alternatively, $f(x) \to \ell$ as $x \to \infty$ (resp. as $x \to -\infty$).

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 as a number ℓ such that for all $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - \ell| < \epsilon$ whenever $|x - c| < \delta$ and $x \in (a, c)$ (that is, $x \in (a, b)$ and $c - \delta < x < c$).

Our notation for this is $\lim_{x \to c-} f(x) = \ell$, 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 (if I mention the history right away you will be able to get the solution by googling!), but feel free to use any method you like.

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 discontinuous 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 off 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 nonzero.

Continuity of other familiar functions

What are the other (continuous) functions we know? How about the trigonometric 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.

Note that $\lim_{x\to 0} \sin x = 0 = \sin 0$, which can be proved easily by using the inequality $|\sin x| \leq |x|$, for all $x \in [-\pi/2, \pi/2]$

and by using the formula $\sin(a+h) = \sin a \cos h + \cos a \sin h$ and $\lim_{x\to a} \sin x = \lim_{h\to 0} \sin(a+h) = \sin a \cos 0 + \cos a \sin 0 = \sin a$, the continuity of the function $\sin x$ at any point $a \in \mathbb{R}$ can be shown (here we have used the continuity of $\cos x$ at 0).

How can we show that $\cos x$ is continuous at each $a \in \mathbb{R}$?

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 trigonometric 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 $\lim_{x \to x_0} \sqrt{x} = \sqrt{x_0}$, that is, we need to show that $|\sqrt{x} - \sqrt{x_0}| < \epsilon$ whenever $0 < |x - x_0| < \delta$ for some δ. First assume that $x_0 \neq 0$. Then

$$|\sqrt{x} - \sqrt{x_0}| = \left| \frac{x - x_0}{\sqrt{x} + \sqrt{x_0}} \right| \leq \frac{|x - x_0|}{\sqrt{x_0}}.$$

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

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

which is what we needed to prove. When $x_0 = 0$, I leave the proof as an 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 $c \in [a, b]$ 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 $y = u$ with u between $f(a)$ and $f(b)$.

The IVT in a picture

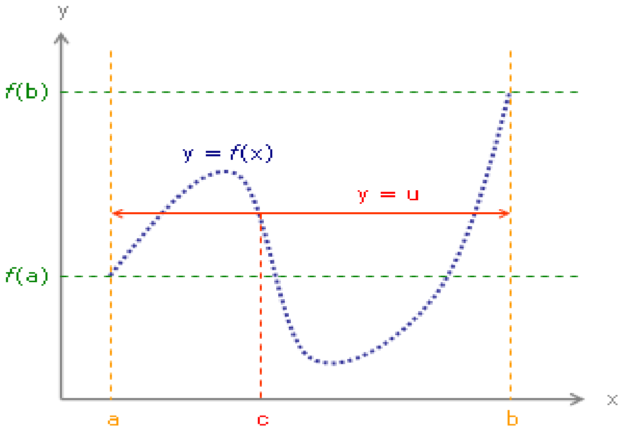


Image created by Enoch Lau see
http://en.wikipedia.org/wiki/File:Intermediatevaluetheorem.png
(Creative Commons Attribution-Share Alike 3.0 Unported license).

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 $a_n > 0$. By using the fact that $\lim_{x \to \pm\infty}(P(x)/x^n) = a_n$, it is easy to see that if we take $x = b > 0$ large enough, $P(b)$ will be positive, and 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 $P(x_0) = 0$. $\square$

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 and bounded intervals

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

Theorem 11: A continuous function on a closed and 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.

Again, we will not prove this, 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 its minimum, although, it is bounded below.

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 $c \in (a, b)$? (Exercise 1.11.(i). Is the statement $\lim_{x\to c} f(x) = 0 \Rightarrow \lim_{x\to c} f(x)g(x) = 0$ true?).

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 $f(0) = 0$. The question asks if this function is continuous at $x = 0$.
How about $x \neq 0$? Why is $f(x)$ continuous? Because it is a composition of the sin function and a rational function $1/x$. Since both of these are continuous when $x \neq 0$, so is $f(x)$.

Let us look at the sequence of points $x_n = \frac{2}{(2n+1)\pi}$. 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.

Sequential continuity

The preceding example showed that in order to demonstrate that a function, say $f(x)$, 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? The following theorem answers the question affirmatively.

Theorem 12: A function $f(x)$ is continuous at a point a if and only if for every sequence x_n converging to a, the sequence $f(x_n)$ converges to $f(a)$.

A function that satisfies the property that for every sequence $x_n \to a$, $\lim_{n\to\infty} f(x_n) = f(a)$ 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 harder to prove.

Proof of Theorem 12

Proof. Let $I \subseteq \mathbb{R}$ and $f : I \longrightarrow \mathbb{R}$ be a function. Let $a \in I$.

$(\Rightarrow)$. Let f be continuous at a. That is, for a given $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$.

Since the sequence x_n converges to a, for the above $\delta > 0$, there exists $N \in \mathbb{N}$ such that $|x_n - a| < \delta$ whenever $n \geq N$.

Hence $|f(x_n) - f(a)| < \epsilon$ whenever $n \geq N$.

$(\Leftarrow)$. We will show this part by using the method of contradiction.

For, let if possible, the function f is NOT continuous at a, that is, there exists $\epsilon > 0$ such that for all $\delta > 0$, there exists $x_\delta \in I$ with $|x_\delta - a| < \delta$ and $|f(x_\delta) - f(a)| \geq \epsilon$.

Now, we find a sequence x_n converging to a, for which, $f(x_n)$ does not converge to $f(a)$.

Proof continued...

For $n \in \mathbb{N}$ and for the same ϵ, if we take $\delta_n = \frac{1}{n}$ then there exists $x_n \in I$ such that $|x_n - a| < \frac{1}{n}$ and $|f(x_n) - f(a)| \geq \epsilon$.

It is now clear that the sequence x_n converges to a but for the ϵ above, there does not exist $N \in \mathbb{N}$ such that $|f(x_n) - f(x)| < \epsilon$, whenever $n \geq N$, that is, the sequence $f(x_n)$ does not converge to $f(a)$ which is a contradiction.

Hence the function f is continuous. $\square$

Remark: Theorem 12 (continuity is same as sequential continuity) goes through without any problems even when the range and/or domain of the function are/is in $\mathbb{R}^2$ or $\mathbb{R}^3$. Exactly the same proof works in this case. Note that we have not yet defined the continuity of functions having range and/or domain in $\mathbb{R}^2$ or $\mathbb{R}^3$. You can try defining it and proving the above theorem in this case.

Differentiability: The definition

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.

The slope of the tangent

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.

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:

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

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

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: By the Sandwich theorem

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

($\lim_{h\to 0} |g(h)| = 0 \iff \lim_{h\to 0} g(h) = 0$). $\square$

Note: Functions that satisfy the property above for $\alpha > 0$ (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 derivatives of the sums, differences, products and quotients of differentiable functions. You should try and remember how to prove these.

You should also recall the chain rule (for $F(x) = f(g(x))$, $F'(x) = f'(g(x))g'(x)$) 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.

Exercise: Show that the differentiable functions are continuous.

Use the simple observation that

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

and then apply the limit rule.

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 and 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 is also differentiable at x_0, we can reason as follows.

We know that $f(x_0 + h) - f(x_0) \leq 0$ for every $h \in \mathbb{R}$ such that $x_0 + h \in X$.

Hence, we see (one half of the Sandwich Theorem!) that when $h > 0$,

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

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 preceding argument is purely local. Before explaining what this means, we give the following definition.

Definition: Let $f : X \to \mathbb{R}$ be a function and x_0 be in X. Suppose there is a 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 $x_0 \in X$, $f'(x_0) = 0$.

Proof: Exercise.

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 and 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. $\square$

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

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

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

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

Proof: Apply Rolle's Theorem to the function

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

$\square$

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. □

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

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 $[c, d]$ (not only on $[c, d]$ but also on (a, b) and differentiable at all the points in (a, b)) and hence, by Theorem 11 must attain its supremum (also the infimum but we will consider only the supremum).

Since $g'(c) = u - f'(c) > 0$ (as g is differentiable at all the points in (a, b), it is also differentiable at c and d) and by the definition of the derivative of a function, for $\epsilon = \frac{g'(c)}{2} > 0$, $\exists \delta > 0$ such that

$$\left| \frac{g(c + h) - g(c)}{h} - g'(c) \right| < \epsilon$$

whenever $|h| < \delta$.

Proof continued...

That is,

$$g'(c) - \epsilon < \frac{g(c+h) - g(c)}{h} < g'(c) + \epsilon$$

whenever $|h| < \delta$.

In particular, for $0 < h < \delta$ such that $c + h \in (c, d)$,

$$g(c+h) - g(c) > h(g'(c) - \epsilon) = h\frac{g'(c)}{2} > 0,$$

that is,

$$g(c+h) > g(c)$$

and hence $g(c)$ cannot be the $\sup_{x \in [c,d]} g(x)$.

Since $g'(d) = u - f'(d) < 0$, for $\epsilon = -\frac{g'(d)}{2} > 0$, $\exists \delta > 0$ such that

$$\left| \frac{g(d+h) - g(d)}{h} - g'(d) \right| < \epsilon$$

whenever $|h| < \delta$.

Proof continued...

That is,

$$g'(d) - \epsilon < \frac{g(d+h) - g(d)}{h} < g'(d) + \epsilon$$

whenever $|h| < \delta$.

In particular, for $-\delta < h < 0$ such that $d + h \in (c, d)$,

$$g(d+h) - g(d) > h(g'(d) + \epsilon) = h\frac{g'(d)}{2} > 0,$$

that is,

$$g(d+h) > g(d)$$

and hence $g(d)$ cannot be the $\sup_{x \in [c,d]} g(x)$.

Thus there exists $x \in (c, d)$ where g attains its supremum, and hence 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 $f'(x_0) = 0$, is 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

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

that is, for a given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$f''(x_0) - \epsilon < \frac{f'(x_0 + h)}{h} < f''(x_0) + \epsilon$$

whenever $|h| < \delta$. Now, by putting $\epsilon = \frac{f''(x_0)}{2}$ in the above inequality we get

$$0 < \frac{f''(x_0)}{2} < \frac{f'(x_0 + h)}{h}$$

whenever $|h| < \delta$. It follows that for $|h| < \delta$,

$$f'(x_0 + h) < 0 \quad \text{if } h < 0, \text{ and } f'(x_0 + h) > 0 \quad \text{if } h > 0.$$

It follows that $f(x)$ is decreasing to the left of x_0 and increasing to the right of x_0 (why?).

The proof of the Second Derivative Test continued...

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 enough to study either the convex or the concave functions.

Examples of concave and convex functions

Here are some examples of convex functions.

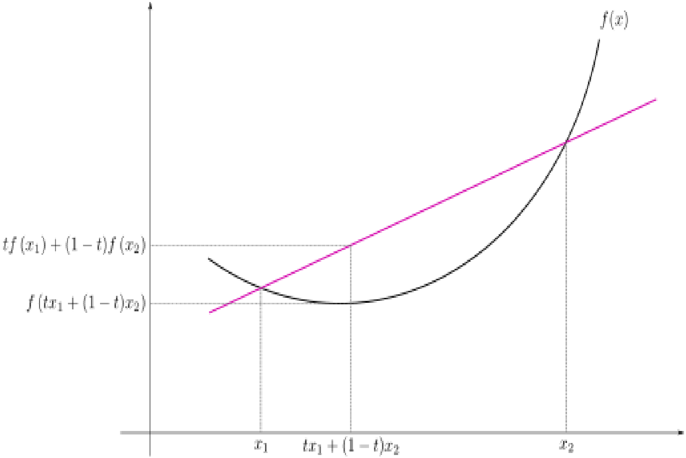
1. $f(x) = x^2$ on $\mathbb{R}$.
2. $f(x) = x^3$ on $[0, \infty)$.
3. $f(x) = e^x$ on $\mathbb{R}$.

Examples of concave functions include

1. $f(x) = -x^2$ on $\mathbb{R}$.
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



Properties of Convex functions

Convex functions have many nice properties. For instance, it is easy to show that convex functions are continuous (do this!) (Hint: Show that, for $x_1 < x_2 < x_3$, $[f(x_2) - f(x_1)]/[x_2 - x_1] \leq [f(x_3) - f(x_1)]/[x_3 - x_1] \leq [f(x_3) - f(x_2)]/[x_3 - x_2]$). More is true.

Exercise 1. Every convex function f (on a bounded interval) is Lipschitz continuous (cf. Exercise 1.16 with $\alpha = 1$), that is, there exists $M > 0$ such that $|f(x+h) - f(x)| \leq M|h|$, for all $x, x+h$ inside the domain of the function f. (Can you think of a convex function which is not Lipschitz continuous? How about the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined as $f(x) = x^2$?; note that this function is Lipschitz continuous on any bounded interval).

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).

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 enough to study either the convex or the concave functions.

Examples of concave and convex functions

Here are some examples of convex functions.

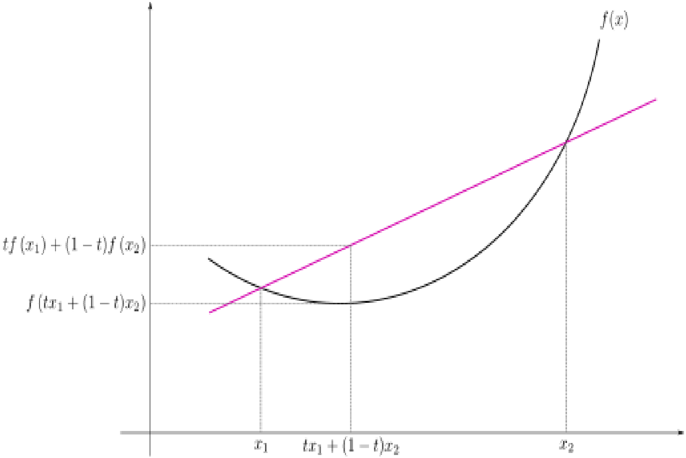
1. $f(x) = x^2$ on $\mathbb{R}$.
2. $f(x) = x^3$ on $[0, \infty)$.
3. $f(x) = e^x$ on $\mathbb{R}$.

Examples of concave functions include

1. $f(x) = -x^2$ on $\mathbb{R}$.
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



Properties of Convex functions

Convex functions have many nice properties. For instance, it is easy to show that convex functions are continuous (do this!) (Hint: Show that, for $x_1 < x_2 < x_3$, $[f(x_2) - f(x_1)]/[x_2 - x_1] \leq [f(x_3) - f(x_1)]/[x_3 - x_1] \leq [f(x_3) - f(x_2)]/[x_3 - x_2]$). More is true.

Exercise 1. Every convex function f (on a bounded interval) is Lipschitz continuous (cf. Exercise 1.16 with $\alpha = 1$), that is, there exists $M > 0$ such that $|f(x+h) - f(x)| \leq M|h|$, for all $x, x+h$ inside the domain of the function f. (Can you think of a convex function which is not Lipschitz continuous? How about the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined as $f(x) = x^2$?; note that this function is Lipschitz continuous on any bounded interval).

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

It follows that 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) of the non-zero derivatives is odd, then we get a sufficient condition.

The Taylor Series: Some Notation

We will first introduce some notation.

The space $\mathcal{C}^k(I)$ will denote the space of k times continuously differentiable functions on an 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 order $0, 1, \ldots, n$ 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

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

Taylor's Theorem

The Taylor polynomials are rigged exactly so that the order n Taylor 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 $f^{(0)}(x_0) = f(x_0)$ by convention.

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

Theorem 19: Let $f \in \mathcal{C}^n(I)$ for some open interval I containing a, and suppose that $f^{(n+1)}$ exists on this interval. Then for each $b \neq a \in I$, there exists c between a and b such that

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

where P_n denotes the Taylor polynomial of order n at a.

It is customary to denote the function $f(b) - P_n(b)$ by $R_n(b)$.

Taylor's Theorem gives us a simple formula for $R_n(b)$. If we can make $R_n(b)$ small, we can approximate our function $f(x)$ by a polynomial.

The proof of Taylor's theorem

Proof: Consider the function

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

Clearly $F(b) = 0$, and

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

Observe that $F(a) = f(b) - P_n(b)$ which need not be zero in general but it is zero if f is a polynomial of degree less or equal to n (as in this case $P_n(b) = f(b)$) and in this case $f^{(n+1)}(c) = 0$ for all c and the theorem gets proved!

We compute $F(a)$ by using the Rolle's Theorem. For, we 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 proved the MVT using Rolle's Theorem), and we see that $g(b) = g(a) = 0$.

Applying Rolle's Theorem we see that there is a c between a and 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

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

This proves what we want. $\square$

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 to $f(b)$ provided we know that the difference $f(b) - P_n(b) = R_n(b)$ 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 $R_n(b) = e^b - P_n(b)$, we obtain

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

since $c \leq b$. (In case when $b < 0$, we get $b \leq c \leq 0$ and $e^c < 1$).

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.

Appendix: The ratio test for the convergence of a series

Theorem: Let $\sum_{k=1}^{\infty} a_k$ be an infinite series and let

$$\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = L.$$

Then, there are three possibilities:

1. If $L < 1$, then the above series is convergent.
2. If $L > 1$, then the above series is divergent.
3. If $L = 1$, then the test is inconclusive.

Proof: Exercise (Hint: 1. The geometric series $\sum_{k=1}^{\infty} a_N r^{k-1}$ is convergent for $|r| < 1$. 2. If $L > 1$, then $\lim_{k\to\infty} a_k \neq 0$. 3. Try to find an example of a convergent series for which $L = 1$. Also, find an example of a divergent series for which $L = 1$).

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, that is, the sequence of the n-th partial sum of the series is a Cauchy sequence. This means that we do not have to guess a value of the limit.

One can use the ratio test too to conclude the convergence of such series!

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.

Computing the values of functions

A calculator or a computer program calculates 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 $|f^{(n)}(x)| \leq 1, \ \forall n \geq 1$.

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 $n \geq 21$.

Computing the values of $\sin x$ for general $x \in \mathbb{R}$

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

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

Hence, we need

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

We know that it is enough to look at $0 \le x \le \pi/2$. Let us be a little careless and allow $x \le 2$ (so we won't get the best possible n, maybe).

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

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

Hence $|R_n(x)| = \frac{|x|^{n+1}}{(n+1)!} < 10^{-16}$ for $n = 31$. $\square$

Smooth functions and Taylor series

Given a smooth function $f(x)$ on on an open interval $I \subseteq \mathbb{R}$, we can write down its associated Taylor polynomials $P_n(x)$ around any point a in I.

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 approximated by Taylor polynomials

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 $x \in I$.

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.

Partitions

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

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

We can think of the points of the partition as dividing the original interval $I = [a, b]$ 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-intervals intersect, they have at most one point in common. Hence, the notation "partition".

Definition: A partition $P' = \{a = x'_0 < x'_1 < \cdots < x'_m = b\}$ 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 P_1 and P_2 have a common refinement $P = P_1 \cup P_2$.

Lower and Upper sums

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

First, we set:

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

Definition: 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 $[a, b]$.

The Darboux integrals

For any partition $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ of $[a, b]$, $m_j \leq M_j$, $\forall\ 1 \leq j \leq n$ and hence

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

Since the function f is bounded on $[a, b]$, there exists $m, M \in \mathbb{R}$ such that

$$m \leq f(x) \leq M$$

for all $x \in [a, b]$ and hence

$$m(b - a) \leq L(f, P) \leq U(f, P) \leq M(b - a)$$

for any partition P of $[a, b]$.

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 $[a, b]$.

The Darboux integrals

Similarly, the upper Darboux integral of f is defined 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.

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 of an interval is a partition that divides the interval into sub-intervals of equal length.

For $[0, 1]$, this is

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

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

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 $L(f, P_n)$ and $U(f, P_n)$ for $f(x) = x$ on $[0, 1]$

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

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

This can be evaluated explicitly:

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

Similarly, we can check that

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

Can we conclude that the Darboux integral is $\frac{1}{2}$ by letting $n \to \infty$? Unfortunately, no, as of now. But we will see soon that the function $f(x) = x$ is Darboux integrable on any finite interval.

An example of a function that is not Darboux integrable

Here is a function that is not Darboux integrable on $[0, 1]$. 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 $L(f) = 0 \neq 1 = U(f)$.

Useful properties of the Darboux sums

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

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

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.

Therefore, for any two partitions P_1 and P_2, we consider the common refinement $P = P_1 \cup P_2$ of P_1 and P_2, and get

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

It follows that

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

Useful properties of the Darboux sums

Thus, for any partition P, we get

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

By using the above inequalities, one can easily prove the following theorems.

Theorem 1: (a). If $U(f, P) - L(f, P) < \epsilon$ for some partition P and some $\epsilon > 0$, then $U(f, P') - L(f, P') < \epsilon$ for any refinement P' of P and the same $\epsilon > 0$.

(b). If $U(f, P) - L(f, P) < \epsilon$ for the partition $P = \{x_0 < x_1 < \cdots < x_n\}$ and if $s_j, t_j \in [x_{j-1}, x_j]$ are arbitrary points, then $\sum_{j=1}^{n} |f(s_j) - f(t_j)|(x_j - x_{j-1}) < \epsilon$.

Theorem 2: A bounded function $f : [a, b] \to \mathbb{R}$ is Darboux integrable if and only if for every $\epsilon > 0$, there exists a partition P of $[a, b]$ such that

$$U(f, P) - L(f, P) < \epsilon.$$

Exercise: Try proving the above theorems.

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

$$\|P\| = \max_{1 \leq j \leq n} \{|x_j - x_{j-1}|\}.$$

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

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

for any tagged partition (P, t) of $[a, b]$ having $\|P\| < \delta$.

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

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

In this case, R is called the Riemann integral of the function f on the interval $[a, b]$. It is easy to see that, for a given function f on the interval $[a, b]$, the number R is uniquely determined (how?).

Remark: Note that for a given $\epsilon > 0$ and partition P having $\|P\| < \delta$, the above condition should hold for any tagging t of P.

The Riemann integral continued...

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

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

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

Theorem 3: Definition 1 and Definition 2 of the Riemann integral are equivalent.

Theorem 4: The Riemann integral exists if and only if the Darboux integral exists and in this case the two integrals are equal.

Exercise: Try proving the above theorems.

Back to our example

We now show that the function $f(x) = x$ on $[0,1]$ is Riemann integrable using Definition 2.

Let $\epsilon > 0$ be arbitrary. For our fixed partition, we take $P = P_n$ where $n > \frac{1}{2\epsilon}$ is some fixed number.

If (P', t') is any (tagged) refinement of P_n, we have

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

Now, recall that we have already computed

$$L(f, P_n) = \frac{1}{2} - \frac{1}{2n} \quad \text{and} \quad U(f, P_n) = \frac{1}{2} + \frac{1}{2n}.$$

By putting these values in above inequality we get

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

and hence the function $f(x) = x$ on $[0,1]$ is Riemann integrable (by using Definition 2), and its integral is $\frac{1}{2}$. $\square$

Some Remarks

Remark 1: With Theorem 4 in hand, we see that the function $f(x) = x$ is also Darboux integrable $[0, 1]$. In fact, following the above proof one can easily show that the function $f(x) = x$ is Riemann/Darboux integrable on any interval $[a, b]$, for $a, b \in \mathbb{R}$. Does it also follow from Theorem 2?

Remark 2: From now onwards, we will use any of the three definitions (one of the Darboux integral and Definitions 1&2 of the Riemann integral) for computing the Riemann/Darboux integrals as essentially all of them are the same (thanks to Theorems 3&4).

Remark 3: Remember that the partition P (in Theorem 2 and Definition 2) and δ (in Definition 1) we get for a given $\epsilon > 0$ depend on ϵ, that is, P and δ in general get changed when ϵ is changed (you can correlate it with N and δ getting changed with ϵ for limits of sequences and functions).

We have already seen in the proof of the integrability of $f(x) = x$ on $[0, 1]$ (given in the last slide) that for a given $\epsilon > 0$, P is taken as the P_n, for which, $\frac{1}{n} < \epsilon$.

An example

Let us look at the following tutorial problem that is already discussed in the last tutorial class.

Exercise 4.1. Show from the first principle that the function $f : [0, 2] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 \text{ if } x \in [0, 1] \\ 2 \text{ if } x \in (1, 2] \end{cases}$$

is Riemann integrable.

Remark: The solution given below is using the first principle but can you solve this exercise by considering a partition like

$$P_n = \left\{ 0 < \frac{1}{n} < \frac{2}{n} < \cdots < \frac{2n-1}{n} < 2 \right\}$$

of $[0, 2]$ for $n \in \mathbb{N}$, and using Theorem 2?

The solution

Note that the Riemann integrability is same as the Darboux integrability.

Let $P = \{0 = x_0 < x_1 < \cdots < x_{n-1} < x_n = 2\}$ be an arbitrary partition of $[0, 2]$.

The point 1 lies in one of the partitions, say $[x_{i-1}, x_i]$, for some i.

We assume that $1 \neq x_i$ and treat this case first. In this case,

$$\begin{aligned} L(f, P) &= \sum_{j=1}^{i} (x_j - x_{j-1}) + \sum_{j=i+1}^{n} 2(x_j - x_{j-1}) \\ &= (x_i - x_0) + 2(x_n - x_i) \\ &= (x_i - 0) + 2(2 - x_i) \\ &= 4 - x_i, \end{aligned} \tag{3}$$

where x_i is a point in $(1, 2]$.

If $x_i = 1$ for some i, then

$$\begin{aligned} L(f, P) &= \sum_{j=1}^{i+1}(x_j - x_{j-1}) + \sum_{j=i+2}^{n} 2(x_j - x_{j-1}) \\ &= x_{i+1} - x_0 + 2(x_n - x_{i+1}) \\ &= x_{i+1} - 0 + 2(2 - x_{i+1}) \\ &= 4 - x_{i+1}, \end{aligned} \tag{4}$$

where x_{i+1} is a point in $(1, 2]$.

In either case, $\sup_P L(f, P) = L(f) = 3$.

The Upper sums $U(f, P)$ can be treated in exactly the same way. In either of the cases we have treated above we get

$$U(f, P) = 4 - x_{i-1},$$

for a point $x_{i-1} \in [0, 1)$. It follows that $U(f) = \inf_P U(f, P) = 3$.

It follows that $L(f) = U(f) = 3$, which shows that the function is Darboux and hence Riemann integrable, and $\int_0^2 f(x)dx = 3$. $\square$

The main theorem for Riemann integration

The main theorem of Riemann integration is the following:

Theorem 5: 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 particular, continuous functions on closed and bounded intervals are Riemann integrable.

In fact, one can allow even countably many points of 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 5 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).

Properties of the Riemann integral

From the definition of the Riemann integral we can easily prove the following properties. We assume that f and g are Riemann integrable. Then

$$\int_a^b [f(t) + g(t)]dt = \int_a^b f(t)dt + \int_a^b g(t)dt,$$

$$\int_a^b cf(t)dt = c\int_a^b f(t)dt,$$

for any constant $c \in \mathbb{R}$, and finally if $f(t) \leq g(t)$ for all $t \in [a, b]$, then

$$\int_a^b f(t)dt \leq \int_a^b g(t)dt.$$

Implicit in the properties above is the fact that if f and g are Riemann integrable, then so are $f + g$ and cf.

It is not hard to prove either of the properties. One needs only to use the corresponding properties for inf and sup.

Proving the properties of the integral

Observe that on any interval $[x_{j-1}, x_j]$,

$$\inf_{[x_{j-1},x_j]} f + \inf_{[x_{j-1},x_j]} g \leq (f+g)(x) = f(x)+g(x) \leq \sup_{[x_{j-1},x_j]} f + \sup_{[x_{j-1},x_j]} g.$$

Hence

$$\inf_{[x_{j-1},x_j]} f + \inf_{[x_{j-1},x_j]} g \leq \inf_{[x_{j-1},x_j]} (f+g) \leq \sup_{[x_{j-1},x_j]} (f+g) \leq \sup_{[x_{j-1},x_j]} f + \sup_{[x_{j-1},x_j]} g.$$

It follows that for any partitions P_1, P_2 and $P = P_1 \cup P_2$ of $[a, b]$,

$$L(f, P_1) + L(g, P_2) \leq L(f, P) + L(g, P) \leq L(f+g, P) \leq L(f+g)$$

and hence

$$\sup_{P_1} L(f, P_1) + \sup_{P_2} L(g, P_2) \leq L(f+g).$$

That is,

$$L(f) + L(g) \leq L(f+g).$$

Proof continues...

Similarly, for any partitions P_1, P_2 and $P = P_1 \cup P_2$ of $[a, b]$,

$$U(f+g) \leq U(f+g, P) \leq U(f, P) + U(g, P) \leq U(f, P_1) + U(g, P_2)$$

and hence

$$U(f+g) \leq \inf_{P_1} U(f, P_1) + \inf_{P_2} U(g, P_2).$$

That is,

$$U(f+g) \leq U(f) + U(g).$$

It follows from the above two inequalities (for the lower and upper Darboux integrals) that

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

Since f, g are integrable, we get

$$\int_a^b f(t)dt + \int_a^b g(t)dt \leq L(f+g) \leq U(f+g) \leq \int_a^b f(t)dt + \int_a^b g(t)dt$$

Proof continues...

and hence

$$L(f+g) = U(f+g) = \int_a^b f(t)dt + \int_a^b g(t)dt$$

that is, $\int_a^b [f(t)+g(t)]dt$ exits and

$$\int_a^b [f(t)+g(t)]dt = \int_a^b f(t)dt + \int_a^b g(t)dt.$$

Also, observe that

$$\sup_{[x_{j-1},x_j]} (cf) = c \sup_{[x_{j-1},x_j]} f \text{ and } \inf_{[x_{j-1},x_j]} (cf) = c \inf_{[x_{j-1},x_j]} f, \text{ for } c \geq 0.$$

Therefore, for $c \geq 0$ (when $c < 0$, we apply these observations to the function $(-c)f$ and use the fact that $\int_a^b (f+g) = \int_a^b f + \int_a^b g$),

$$\inf_P U(cf, P) = c \inf_P U(f, P), \quad \sup_P L(cf, P) = c \sup_P L(f, P).$$

Proof continues...

It follows that

$$U(cf) = c\int_a^b f(t)dt \quad \text{and} \quad L(cf) = c\int_a^b f(t)dt$$

and hence $L(cf) = U(cf)$, which shows that cf is integrable on $[a, b]$ and

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

In case when $f(t) \geq g(t)$ for all $t \in [a, b]$, we get $f(t) - g(t) \geq 0$ for all $t \in [a, b]$ and hence $L(f - g) \geq 0$ and it follows that

$$\int_a^b [f(t) - g(t)]dt \geq 0$$

that is,

$$\int_a^b f(t)dt \geq \int_a^b g(t)dt. \qquad \square$$

Products of Riemann Integrable Functions

Theorem 6: Let $f : [a, b] \to [m, M]$ be a Riemann integrable function and let $\phi : [m, M] \to \mathbb{R}$ be a continuous function. Then the function $\phi \circ f$ (defined as $\phi \circ f(x) = \phi(f(x))$) is Riemann integrable on $[a, b]$.

The above theorem has the following interesting corollaries.

Corollary 1: Let $f, g : [a, b] \to \mathbb{R}$ be bounded functions which are Riemann integrable on $[a, b]$. Then $f \cdot g$, $|f|$ and f^n (for any positive integer n) are Riemann integrable.

Proof: Exercise (Hint: $f \cdot g = \frac{1}{4}[(f+g)^2 - (f-g)^2]$).

Corollary 2: If $f : [a, b] \to \mathbb{R}$ is a Riemann integrable function and $[c, d] \subseteq [a, b]$. Then the function $g : [c, d] \to \mathbb{R}$ defined as $g(x) = f(x)$ for all $x \in [c, d]$ is Riemann integrable.

Proof: Exercise (Hint: First show that the characteristic function $\chi_{[c,d]}$ is integrable on $[a, b]$ and then show that $f \cdot \chi_{[c,d]}$ is integrable on $[a, b]$ by using Corollary 1, and $\int_a^b f(t)\chi_{[c,d]}(t)dt = \int_c^d g(t)dt$.)

Another property of the Riemann Integral

Theorem 7: 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 $P = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$, 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 $[c, b]$.

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

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

By Corollary 2 of Theorem 6, $\int_a^c f(t)dt$ and $\int_c^b f(t)dt$ exist, and hence $L(f)_{[a,c]} = \int_a^c f(t)dt$ and $L(f)_{[c,b]} = \int_c^b f(t)dt$.

Since $L(f) = \int_a^b f(t)dt$, we obtain that

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

The fundamental theorem of calculus: 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 8: 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 on $[a, x]$ for any $x \in [a, b]$ because of Theorem 5 (every continuous function is Riemann integrable).

We show that

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

for $h > 0$. The same way one can prove the case $h < 0$.

By Theorem 7, 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

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

Let $m(h)$ and $M(h)$ be the constant functions given, respectively, by the infimum and supremum of the function f on $[x, x+h]$.

Then, $m(h) \leq f(t) \leq M(h)$, for all $t \in [x, x+h]$, and hence

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

Dividing by h and taking the limit gives

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

We now show that $\lim_{h\to 0} m(h)$ exists and the value of this limit is $f(x)$. By a similar argument one can show that $\lim_{h\to 0} M(h)$ also exists and the value of this limit is also $f(x)$.

For, since $m(h) = \inf_{t\in[x,x+h]} f(t)$, for a given $\epsilon > 0$, there exists $s \in [0, h]$ such that

$$f(x+s) < m(h) + \epsilon/2 \Rightarrow f(x+s) - m(h) < \epsilon/2.$$

Now, we write

$$f(x) - m(h) = f(x) - f(x+s) + f(x+s) - m(h)$$

$$\Rightarrow |f(x) - m(h)| \leq |f(x) - f(x+s)| + |f(x+s) - m(h)|.$$

Since the function f is continuous, there exists $\delta > 0$ such that

$$|f(x+h) - f(x)| < \epsilon/2$$

whenever $|h| < \delta$.

Thus, we obtain that

$$|f(x)-m(h)| \leq |f(x)-f(x+s)|+|f(x+s)-m(h)| < \epsilon/2+\epsilon/2 = \epsilon$$

whenever $|h| < \delta$ (since $s \in [0, h]$) and hence

$$\lim_{h\to 0} m(h) = f(x).$$

Similarly, $\lim_{h\to 0} M(h) = f(x)$. Now, we go back to our inequality at the end of the last to the last slide.

By the Sandwich theorem for limits (use version 2), we see that the limit in the middle exists and is equal to $f(x)$, that is,

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

This proves that $F(x)$ is differentiable on (a, b) and $F'(x) = f(x)$.

How to show that $F(x)$ is continuous on $[a, b]$? Can you show that $F(x)$ is continuous at the end points a, b? This I leave as an exercise. (Hint: $|\int_c^d f(t)dt| \leq \int_c^d |f(t)|dt$) $\square$

Keeping the notation as in the Theorem, we obtain

Corollary: $\int_c^d f(t)dt = \int_a^d f(t)dt - \int_a^c f(t)dt = F(d) - F(c)$ for any two points $c, d \in [a, b]$.

The Fundamental Theorem of Calculus Part II

Theorem 9: 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 $g'(t) = f(t)$. 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 hence is stronger than the corollary we have just stated.

Proof: We can write:

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

where $\{a = x_0 < x_1 < \cdots < x_n = b\}$ is an arbitrary partition of $[a, b]$.

Using the mean value theorem for each of the intervals $I_i = [x_{i-1}, x_i]$, we can write

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

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

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

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

The calculation above is valid for any partition.

Since f is Riemann integrable on $[a, b]$, for a given $\epsilon > 0$ there exists $\delta > 0$ such that

$$\left| \sum_{i=1}^{n} [f(t_i)(x_i - x_{i-1})] - \int_a^b f(t)dt \right| < \epsilon$$

for any tagged partition (P, t) of $[a, b]$ having $\|P\| < \delta$ (using the first definition of Riemann integration).

Now, we construct a partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a, b]$ having $\|P\| < \delta$ and consider the tagging $t = \{c_i : x_{i-1} \leq c_i \leq x_i, 1 \leq i \leq n\}$ (of P) that we get from $(*)$ which we consider for this partition P having $\|P\| < \delta$.

Now, it follows from $(*)$ that

$$\left| g(b) - g(a) - \int_a^b f(t)dt \right| = \left| \sum_{i=1}^{n} [f(c_i)(x_i - x_{i-1})] - \int_a^b f(t)dt \right| < \epsilon.$$

Since the above inequality holds for all positive real number ϵ, we get

$$\left| g(b) - g(a) - \int_a^b f(t)dt \right| = 0,$$

that is,

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

FTC: Applications

Exercise 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.

Solution: Consider $F(x) = \int_a^x f(t)dt,\ x \in \mathbb{R}$. Then $F'(x) = f(x)$. Note that

$$\int_{u(x)}^{v(x)} f(t)dt = \int_a^{v(x)} f(t)dt - \int_a^{u(x)} f(t)dt = F(v(x)) - F(u(x)).$$

Use the Chain rule to see that

$$\begin{aligned} \frac{d}{dx}\int_{u(x)}^{v(x)} f(t)dt &= F'(v(x))v'(x) - F'(u(x))u'(x) \\ &= f(v(x))v'(x) - f(u(x))u'(x). \end{aligned}$$

Using this formula for $G(x) = \int_x^{x+p} f(t)dt$, we see that $G'(x) = 0$ for all $x \in \mathbb{R}$, and hence $G(x)$ is constant. Thus $\int_a^{a+p} f(t)dt$ has the same value for every real number a. $\square$

Mean Value Theorem for Integrals

Theorem 10 (Mean Value Theorem for Integrals): If f is continuous on $[a, b]$, then there exists $c \in [a, b]$ such that

$$\int_a^b f(x)dx = f(c)(b-a).$$

Proof: Since the function

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

is continuous on $[a, b]$ and differentiable on (a, b) with $F'(x) = f(x)$ for all $x \in (a, b)$ (by the Fundamental Theorem of Calculus), there is $c \in (a, b)$ such that

$$f(c) = F'(c) = \frac{F(b) - F(a)}{b-a} = \frac{1}{b-a}\int_a^b f(x)dx$$

(by the Mean Value Theorem).

Thus

$$f(c)(b-a) = \int_a^b f(x)dx. \quad \square$$

The logarithmic function

Definition: The natural logarithmic function is defined on $(0, \infty)$ by

$$\ln x = \int_1^x \frac{1}{t} dt.$$

It is clear that $\ln 1 = 0$, $\ln x > 0$ for $x \in (1, \infty)$, and $\ln x < 0$ for $x \in (0, 1)$.

Theorem 11:

1. $\ln(xy) = \ln x + \ln y$
2. $\ln(\frac{x}{y}) = \ln x - \ln y$
3. $\ln(x^r) = r \ln x$, if r is a rational number.

Proof: (1). Let $f(t) = \ln(ty)$. Then, $f'(t) = \frac{1}{t}$. Therefore, by FTC - Part II, $\ln x = f(x) - f(1)$, that is, $\ln(xy) = \ln x + \ln y$. For (2), put $x = 1/y$ in (1) and get $\ln(1/y) = -\ln y$, and then use (1) again, for the product of x and $1/y$. (3) is clear if $r \in \mathbb{Z}$. Observe that $\ln x = \ln[(x^{1/q})^q] = q \ln(x^{1/q})$ for $q \neq 0 \in \mathbb{Z}$, which shows that $\ln(x^{1/q}) = (1/q) \ln x$. Now, if $r = p/q$ for $p, q \in \mathbb{Z}$, using the first case (for $p \in \mathbb{Z}$) we get $\ln(x^{p/q}) = (p/q) \ln x$. $\square$

The exponential function

Remark: $\ln x$ is increasing and concave (why?). Moreover, by IVT, there exists a number $e > 1$ such that $\ln e = 1$ (as $\exists x \in (1, \infty)$ such that $\ln x > 0$ (why?) and $\ln x^n = n \ln x \to \infty$ as $n \to \infty$).

It follows that $\ln x$ is a strictly increasing function whose range is full of $\mathbb{R}$. Therefore, it is invertible and has an inverse. We denote this by $\exp(x)$.

That is,

$$\exp(x) = y \iff \ln y = x$$

In particular, $\exp(0) = 1, \exp(1) = e$.

Since, $\ln(e^r) = r \ln e = r$, we get $e^r = \exp(r)$, when r is a rational number. Therefore, we define $e^x := \exp(x)$ for any $x \in \mathbb{R}$.

Laws of Exponents:

$e^{x+y} = e^x e^y,\ e^{x-y} = \frac{e^x}{e^y},\ (e^x)^r = e^{rx},$ if r is rational.

Proof: Use the laws of exponents for $\ln x$ (Theorem 11). $\square$

The exponential function

Theorem 12: $\frac{d}{dx}(e^x) = e^x$.

Proof: If f is differentiable with nonzero derivative, then f^{-1} is also differentiable and in this case, using the chain rule we get

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

Thus, $\frac{d}{dx}(e^x) = \frac{1}{(\ln)'(e^x)} = \frac{1}{1/e^x} = e^x$. $\square$

Remark: Now, we can define a^x whenever $a > 0$ and $x \in \mathbb{R}$ as

$$a^x = e^{x \ln a}.$$

Exercise: Show that $\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e$.

Solution: Let $f(x) = \ln x$. Then $f'(x) = 1/x$. Thus, $f'(1) = 1$. But,

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

Thus, by using the sequential criterion for limits, if we consider the sequence $\{1/n\}$ converging to 0, then

$$1 = f'(1) = \lim_{n\to\infty} \frac{f(1+(1/n))}{(1/n)} = \lim_{n\to\infty} \ln\left(1+\frac{1}{n}\right)^n .$$

Since the logarithmic function is continuous, we obtain that

$$\lim_{n\to\infty} \ln\left(1+\frac{1}{n}\right)^n = \ln\left(\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n\right)$$

and hence $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$. $\square$

Taylor series: Recall that if f is a $\mathcal{C}^\infty$ function on $\mathbb{R}$, then the Taylor series expansion of f about a is

$$f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots$$

Taylor Series for e^x

Consider the series $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ for a fixed x. If we choose $N > 2x > 0$, then for all $n > N$,

$$\frac{x^{n+1}}{(n+1)!} = \frac{x^n}{n!}\frac{x}{(n+1)} \leq \frac{x^n}{n!}\frac{x}{N} \leq \frac{x^n}{n!}\frac{1}{2}.$$

Thus, for $m \geq n > N$,

$$s_m - s_n = \frac{x^{n+1}}{(n+1)!} + \cdots + \frac{x^m}{m!} \leq \frac{x^{n+1}}{(n+1)!}\left(1 + \frac{1}{2} + \cdots + \frac{1}{2^{m-n}}\right)$$

and hence

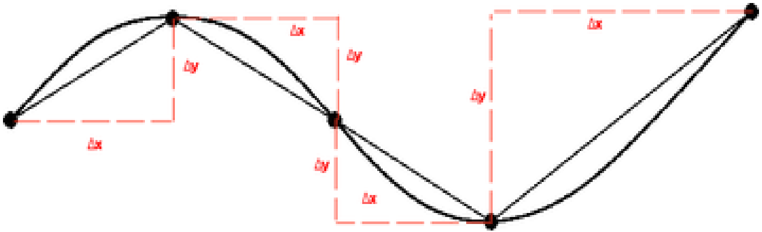
$$s_m - s_n \leq \frac{2x^{n+1}}{(n+1)!} \leq \frac{x^n}{n!}.$$

This shows that the sequence of partial sums of the Taylor series for e^x is Cauchy. Hence the series is convergent.

Does this Taylor series converge to e^x? Yes, as the Taylor's theorem insures that the remainder $R_n(x)$ associated to the function e^x converges to Zero. Therefore $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

Arc length

The picture below and the discussion on the next slide are from Wikipedia (http://en.wikipedia.org/wiki/Arc_length).



See: http://en.wikipedia.org/wiki/File:Arclength-2.png

In the picture above, the curve $y = f(x)$ is being approximated by straight line segments which form the hypotenuses of the right angled triangles shown in the picture.

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, $y_i = f(x_i)$, $\Delta x_i = x_i - x_{i-1}$ and $\Delta y_i = f(x_i) - f(x_{i-1})$ corresponding to a partition $P = \{a = x_0 < \cdots < x_n = b\}$ of $[a, b]$.

The formula for arc length

Now, using the MVT for $y = f(x)$ on $[a, b]$ we get

$$S \sim \sum_{i=1}^{n} \sqrt{1 + \left(\frac{\Delta y_i}{\Delta x_i}\right)^2}\, \Delta x_i = \sum_{i=1}^{n} \sqrt{1 + (y'(t_i))^2}\, \Delta x_i$$

for some $t_i \in [x_{i-1}, x_i]$.

It follows that the arc length of the curve $y = f(x)$ (defined on $[a, b]$) is

$$S := \lim_{||P|| \to 0} \sum_{i=1}^{n} \sqrt{1 + (y'(t_i))^2}\, \Delta x_i = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx,$$

provided this limit exists which is equivalent of saying that the function $f'(x)$ is Riemann integrable on $[a, b]$.

Exercise 4.10. (ii) Find the length of the curve

$$y = \int_0^x \sqrt{\cos 2t}\, dt,\ 0 \le x \le \pi/4.$$

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

$$\int_0^{\frac{\pi}{4}} \sqrt{1 + \cos 2x}\, dx = \sqrt{2} \int_0^{\pi/4} \cos x\, dx = 1.$$

$\square$

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.

Exercise: Show that the graphs of piecewise $\mathcal{C}^1$ functions are rectifiable.

Convergence of Power series

We have already seen the convergence of a specific power series (namely, the Taylor series for e^x). There is a general test we can use to determine if a power series converges.

Theorem 13: Let $\sum_{n=0}^{\infty} a_n(x-b)^n$ be a power series about the point b. If

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \frac{1}{R}$$

for some $R \in \mathbb{R}$, the series converges in the interval $(b-R, b+R)$ to a smooth function. (if the limit is 0, the series converges on the whole real line).

Roughly speaking $|a_n|$ behaves like $1/R^n$ for large n. Hence, the terms in the power series can be bounded by $|(x-b)|^n/R^n$, and this latter (geometric) series converges in $(b-R, b+R)$. This argument can be made precise. Proving that the series is smooth is trickier and we will not get into it.

A convergent Taylor series (or more generally a convergent “power series”) can be differentiated and integrated “term by term”. That is, if

$$f(x) = \sum_{n=0}^{\infty} a_n x^n, \quad \text{then} \quad f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

And similarly,

$$\int_a^b \sum_{n=0}^{\infty} a_n x^n dx = \sum_{n=0}^{\infty} a_n \int_a^b x^n dx.$$

We will not be proving these facts but you can use them below.

Exercise 5: Using Taylor series write down a series for the integral

$$\int \frac{e^x}{x} dx.$$

Solution: We simply integrate term by term to get

$$\log x + x + \frac{x^2}{2 \cdot 2!} + \ldots = \log x + \sum_{k=1}^{\infty} \frac{x^k}{k \cdot k!}.$$

We can obtain Taylor series for the inverse trigonometric functions in this way. Indeed we could define the function $\arcsin x$ in this way:

$$\arcsin x = \int_0^x \frac{1}{\sqrt{1-t^2}}\,dt.$$

Now we can use the binomial theorem for the integrand. Note that the binomial theorem for arbitrary real exponents is an example of Taylor series:

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 + \cdots$$

It is not too hard to prove that the series on the right hand side above converges for $|x| < 1$. Applying the binomial theorem for $\alpha = -1/2$ to the integrand, we get

$$\arcsin x = \int_0^x \left(1 + \frac{1}{2}t^2 - \frac{\frac{1}{2}\cdot\frac{3}{2}}{2!}t^4 + \cdots\right) dt.$$

Integrating this term by term, you should verify that you get the series for $\arcsin x$ that you can derive directly from Taylor series.

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

We will now 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 formulae it makes sense to first identify this subset, which is sometimes called 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) $f(x,y) = \frac{xy}{x^2-y^2}$

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

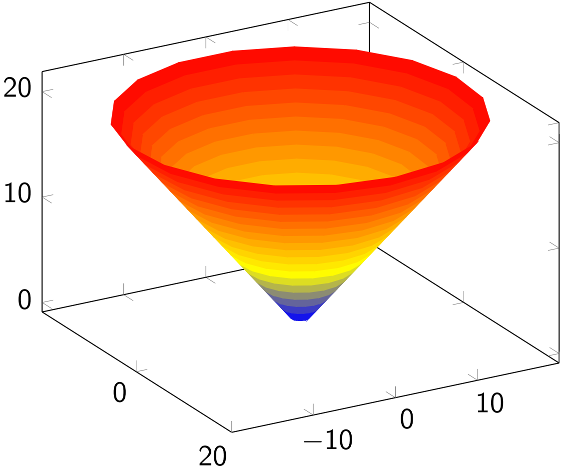
The second thing one should do with a function from $\mathbb{R}^2 \to \mathbb{R}$ is to study its range. This is done in different ways.

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, and in this case the level set is called level curve.

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.

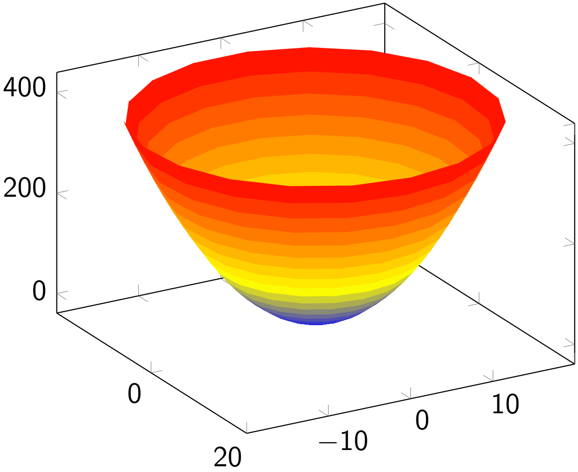
When one plots the $f(x, y) = c$ for some constant c one gets a curve. Such a curve is usually called a contour line (the contour "lives" in the $z = c$ plane).

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 $z = f(0, y)$ 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.

Recall that for $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$, $x - y = (x_1 - y_1, x_2 - y_2)$ and the distance between x and y is given by

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

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

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

whenever $0 < \|x - c\| < \delta$. What set is $\{x \in \mathbb{R}^2 : 0 < \|x - c\| < \delta\}$?

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

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

Exercise 1: Formulate the definition of a limit and of continuity for functions from $\mathbb{R}^3$ to $\mathbb{R}$.

The rules for limits and continuity

Exercise 2: Show that the projection functions $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$ defined as $f_1(x, y) = x$ and $f_2(x, y) = y$ for all $(x, y) \in \mathbb{R}^2$ are continuous functions, using the ϵ, δ definition of continuity (Hint: does $\delta = \epsilon$ work?)

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). In certain cases Exercise 2 may help in determining the continuity of many functions at given points.

Remark: Since the sequential criterion for limit and continuity holds for functions of several variables (see the Remark following the proof of Theorem 12 in slides of Lecture 6), we often use it for proving the nonexistence of limit and discontinuity of functions.

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

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

is continuous at $(0,0)$.

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

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

This shows that f is not continuous at 0.

But does the limit of $f(x,y)$ at $(0,0)$ exist?

Iterated limits

When evaluating a limit of the form $\lim_{(x,y)\to(a,b)} f(x,y)$ one may naturally be tempted to let x go to a first, and then let y go to b.

Does this give the limit in the previous sense?

Exercise 5.5: Let $f : \mathbb{R}^2 \setminus \{(0,0)\} \longrightarrow \mathbb{R}$ be defined as

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

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 does not 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

$$\frac{\partial f}{\partial x_1}(a, b) := \lim_{x_1 \to a} \frac{f((x_1, b)) - f((a, b))}{x_1 - a} = \lim_{t \to 0} \frac{f((a + t, b)) - f((a, b))}{t}.$$

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 x_2:

$$\frac{\partial f}{\partial x_2}(a, b) := \lim_{x_2 \to b} \frac{f((a, x_2)) - f((a, b))}{x_2 - b} = \lim_{t \to 0} \frac{f((a, b + t)) - f((a, b))}{t}.$$

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 defined as

$$\nabla_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} = \lim_{t \to 0} \frac{f((x_1 + tv_1, x_2 + tv_2)) - f((x_1, x_2))}{t}.$$

It measures the rate of change of the function f at x along the path $x + tv$.

Observe that if we take $v = (1, 0)$ in the above definition, we obtain $\partial f / \partial x_1$, while $v = (0, 1)$ yields $\partial f / \partial x_2$.

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. (why?)

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" (why?).

In the section on iterated limits, we studied the following function from Exercise 5.5:

$$f(x,y) = \frac{x^2y^2}{x^2y^2 + (x-y)^2} \quad \text{for} \quad (x,y) \neq (0,0).$$

Let us further set $f(0,0) = 0$. You can check that every directional derivative of f at $(0,0)$ exists and is equal to 0, except along $y = x$ (that is, along the unit vector $v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$) 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 $\lim_{(x,y)\to(0,0)} f(x,y)$ 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 at that point.

Let us go back and examine the notion of differentiability for a function $f(x)$ of one variable.

Suppose f is differentiable at the point x_0. When is a line passing through the point $(x_0, f(x_0))$ on the curve $y = f(x)$ is the tangent line to the curve?

Recall that the equation of a line passing through the point $(x_0, f(x_0))$ and having slope m is

$$y = g(x) = f(x_0) + m(x - x_0).$$

If we consider the difference $f(x) - g(x) = f(x) - f(x_0) - m(x - x_0)$ and write $h = (x - x_0)$, we see that the difference can be rewritten as

$$f(x_0 + h) - f(x_0) - m \cdot h.$$

The above line is the tangent line to the curve $y = f(x)$ at the point $(x_0, f(x_0))$ on the curve if $m = f'(x_0)$, which is equivalent of saying that the above difference

$$f(x_0 + h) - f(x_0) - m \cdot h = o(h) = \varepsilon_1(h)h$$

where $\varepsilon_1(h)$ is a function of h that goes to 0 as h goes to 0, and in this case the function $o(h) = \varepsilon_1(h)h$ is a function of h "that goes to zero faster than h" (that is, $\lim_{h \to 0} \frac{o(h)}{h} = 0$).

The preceding idea generalises to two (or more) dimensions.

Let $f(x, y)$ be a function which has both partial derivatives at (x_0, y_0). In the two variable case we will consider the difference of $z = f(x, y)$ (defining a surface in $\mathbb{R}^3$) and $z = g(x, y)$ (defining the tangent plane to the surface $z = f(x, y)$ in $\mathbb{R}^3$).

Let us first determine how to find the equation of the tangent plane to the surface $z = f(x, y)$ at the point $P = (x_0, y_0, z_0)$ on the surface. The tangent plane is the graph of the function

$$z = g(x, y) = z_0 + a(x - x_0) + b(y - y_0)$$

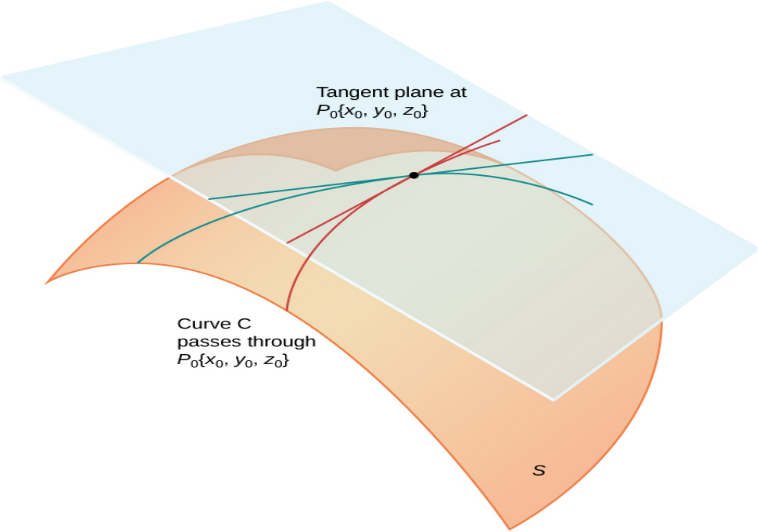
for some $a, b \in \mathbb{R}$.

Now, we determine the values of a and b, for which, the above plane $z = g(x, y)$, passing through the point $P = (x_0, y_0, z_0)$ *on the surface* $z = f(x, y)$, is the tangent plane to the surface.

If we fix the y variable as $y = y_0$ and treat $f(x, y)$ only as a function of x, we get a curve on the surface $z = f(x, y)$.

Similarly, if we treat $g(x, y)$ as function only of x (by fixing $y = y_0$), we obtain a line on the plane $z = g(x, y)$.

The tangent plane in a picture



https://openstax.org/books/calculus-volume-3/pages/4-4-tangent-planes-and-linear-approximations

The tangent line to the curve passing through (x_0, y_0, z_0) must be the same as the line passing through (x_0, y_0, z_0), and, in any event, their slopes (which are given by the derivatives of the curve $z = f(x, y_0)$ at $x = x_0$ and the line $z = g(x, y_0)$ at $x = x_0$, resp.) must be the same.

Since the above derivatives with respect to x are same as the partial derivatives with respect to x, we get

$$\frac{dz}{dx}(x_0) = \frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial g}{\partial x}(x_0, y_0) = a.$$

Arguing in exactly the same way but fixing the x variable as $x = x_0$ 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 the surface $z = f(x, y)$ at the point (x_0, y_0, z_0) (remember that $z_0 = f(x_0, y_0)$) is

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0). \quad \square$$

A remark on tangent plane to the surface $z = f(x, y)$

Remark: Note that the general form of the equation of a plane passing thorough the point $P = (x_0, y_0, z_0)$ is

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

for some $(a, b, c) \neq (0, 0, 0) \in \mathbb{R}^3$.

Since we assumed that the partial derivatives of the function $f(x, y)$ at (x_0, y_0) exist (and hence are finite real numbers), using the way we have determined the values of a and b in the last slides we get that the number c appearing in the above equation of the plane has to be a nonzero real number (as the derivative of the curve $z = f(x, y_0)$ at $x = x_0$ exists, the slope of the tangent line to this curve at (x_0, y_0) cannot be infinite) and then without loss of any generality c can be taken as -1.

It now follows from the above discussion that the general form of the equation of the tangent plane can be taken as

$$z = g(x, y) = z_0 + a(x - x_0) + b(y - y_0)$$

and the rest is given in the last couple of slides.

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 1: 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

$$\lim_{(h,k)\to(0,0)} \frac{\left|f(x_0 + h, y_0 + k) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k\right|}{\|(h, k)\|} = 0.$$

We could rewrite this as

$$\begin{aligned} f((x_0, y_0) + (h, k)) - f(x_0, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - \frac{\partial f}{\partial y}(x_0, y_0)k \\ = \varepsilon(h, k)\|(h, k)\| \end{aligned}$$

where $\varepsilon(h, k)$ is a function that goes to 0 as $(h, k) \to (0, 0)$.

This form of differentiability now looks exactly like the one variable version case (put $o(h, k) = \varepsilon(h, k)\|(h, k)\|$). Can you guess the derivative $f'(x_0, y_0)$ of the function $f(x, y)$ at (x_0, y_0)?

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 a 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 2: The function $f(x, y)$ is said to be differentiable at a point (x_0, y_0) if there exists a matrix denoted $Df(x_0, y_0)$ with the property that

$$f((x_0, y_0)+(h, k))-f(x_0, y_0)-Df(x_0, y_0)\begin{pmatrix} h \\ k \end{pmatrix} = o(h, k) = \varepsilon(h, k)\|(h, k)\|$$

for some function $\varepsilon(h, k)$ that goes to 0 as (h, k) goes to $(0, 0)$.

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 $\mathbb{R}$.

A condition for differentiability

Exercise: Show that a function $f(x, y)$ is differentiable in the sense of Definition 1 if and only if it is differentiable in the sense of Definition 2 with

$$Df(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}.$$

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).

Theorem 1: 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).

Remark: We omit the proof of this theorem. However, we note that a function whose partial derivatives exist and are continuous is said to be of class $\mathcal{C}^1$. The theorem says that every $\mathcal{C}^1$ function is differentiable.

Differentiability $\Rightarrow$ continuity

Theorem: Let U be a subset of $\mathbb{R}^2$ and $(x_0, y_0) \in U$. If $f : U \longrightarrow \mathbb{R}$ is differentiable at (x_0, y_0), then f is continuous at (x_0, y_0).

Proof: Note that for $\epsilon = 1, \exists \delta_1 > 0$ such that

$|f((x_0, y_0) + (h, k)) - f(x_0, y_0)| = \left| \varepsilon(h, k)\|(h, k)\| + Df(x_0, y_0) \begin{pmatrix} h \\ k \end{pmatrix} \right|$

$\leq |\varepsilon(h, k)| \cdot \|(h, k)\| + \|Df(x_0, y_0)\| \cdot \|(h, k)\|$ (property of dot product)

$< (1 + K)\|(h, k)\|$ (since $\varepsilon(h, k)$ tends to 0 as (h, k) tends to $(0, 0)$)

whenever $||(h, k)|| < \delta_1$ (where $K = \|Df(x_0, y_0)\|$).

Therefore, for a given $\epsilon > 0$, if we take $\delta = \min\{\delta_1, \frac{\epsilon}{1+K}\}$, then

$$|f((x_0, y_0) + (h, k)) - f(x_0, y_0)| < \epsilon$$

whenever $||(h, k)|| < \delta$. Hence the differentiable function f is continuous. $\square$

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 at b and c, respectively, while x is varied, we get the partial derivative of the function f with respect to x at the point (a, b, c) as

$$\frac{\partial f}{\partial x}(a, b, c) = \lim_{x \to a} \frac{f(x, b, c) - f(a, b, c)}{x - a} = \lim_{t \to 0} \frac{f(a + t, b, c) - f(a, b, c)}{t}.$$

In a similar way, we can define the partial derivatives

$$\frac{\partial f}{\partial y}(a, b, c) \quad \text{and} \quad \frac{\partial f}{\partial z}(a, b, c).$$

Differentiability in three variables

Exercise 1: Formulate a definition of differentiability for a function of three variables.

Exercise 2: Formulate the analogue of Theorem 1 (of this section) 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 derivative of vector-valued functions

We now define the derivative of a function $f : U \to \mathbb{R}^n$, where U is a subset of $\mathbb{R}^m$.

Recall that we can write $f = (f_1, f_2, \ldots, f_n)$ where $f_j = \pi_j \circ f : U \to \mathbb{R}$ and $\pi_j : \mathbb{R}^n \to \mathbb{R}$ is the projection on the j-th coordinate defined as $(y_1, y_2, \ldots, y_n) \mapsto y_j$.

The function f is said to be differentiable at a point x if there exists an $n \times m$ matrix $Df(x)$ such that

$$\lim_{h \to (0,0,\ldots,0)} \frac{\|f(x+h) - f(x) - Df(x)(h)\|}{\|h\|} = 0$$

where $x = (x_1, x_2, \ldots, x_m)$, $h = (h_1, h_2, \ldots, h_m)$ are vectors in $\mathbb{R}^m$ and $Df(x)(h) = Df(x) \cdot h$ is a vector in $\mathbb{R}^n$ (we are considering h here as a column vector, that is, a matrix of order $m \times 1$).

The matrix $Df(x)$ is usually called the total derivative of f. It is also referred as the Jacobian matrix. What are its entries?

From our experience in the 1×2 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:

$$Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_m}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_m}(x) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial f_n}{\partial x_1}(x) & \frac{\partial f_n}{\partial x_2}(x) & \cdots & \frac{\partial f_n}{\partial x_m}(x) \end{bmatrix}.$$

In the 3×1 case (that is, when $m = 1$, $n = 3$ and $f = (f_1, f_2, f_3) : U(\subseteq \mathbb{R}) \to \mathbb{R}^3$) we get

$$f'(t) = Df(t) = \begin{bmatrix} f_1'(t) \\ f_2'(t) \\ f_3'(t) \end{bmatrix}.$$

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 $\mathbb{R}$ to $\mathbb{R}^3$).

Norm of a matrix

Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

be an $n \times m$ matrix with entries in $\mathbb{R}$. One can define the norm of the matrix A as

$$\|A\| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2}.$$

Just by using the fact that

$$|a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m| \leq \sqrt{\sum_{j=1}^{m} a_{ij}^2} \sqrt{\sum_{j=1}^{m} x_j^2}$$

one can show easily that

$$\|A(x)\| \leq \|A\| \cdot \|x\|$$

for $x = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$.

An Exercise and a Remark

Exercise: Following the proof of the continuity of differentiable scalar fields (and by using the property of the norm of a matrix), show that the differentiable vector-valued functions are also continuous.

Note that the scalar (real) valued functions of multi-variables are also known as scalar fields and vector-valued functions as vector fields.

Remark: Theorem 1 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 $\mathbb{R}^m$!).

Rules for the total derivative

Rule 1: Just like in the one variable case, if f and g are differentiable

$$D(f+g)(x) = Df(x) + Dg(x)$$

and

$$D(cf)(x) = cDf(x),\ \forall c \in \mathbb{R}.$$

Rule 2: Somewhat harder, but only because the notation gets more cumbersome, is the Chain rule: Let $U \subseteq \mathbb{R}^m$, $V \subseteq \mathbb{R}^n$, $f : U \to \mathbb{R}^n$ be a function which differentiable at $x_0 \in U$ and $f(U) \subseteq V$. Let $g : V \to \mathbb{R}^\ell$ be a function which is differentiable at $f(x_0)$. Then $g \circ f : U \to \mathbb{R}^\ell$ is differentiable at x_0 and

$$D(g \circ f)(x_0)_{\ell \times m} = Dg(f(x_0))_{\ell \times n} \circ Df(x_0)_{n \times m},$$

where $\circ$ on the right hand side denotes the matrix multiplication.

The Chain Rule: Applications

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 2: With notation as above

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

For a function $w(t) = f(x(t), y(t), z(t))$ in three variables, the chain rule takes the form

$$\frac{dw}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}.$$

Remark: We omit the proof of the above theorem here but if you are interested in seeing that I can upload it on Moodle later.

An application to tangents of curves

A simple example to verify the chain rule: Let $z = f(x, y) = xy$, $x(t) = t^3$ and $y(t) = t^2$. Then $z(t) = t^5$, so $z'(t) = 5t^4$.

On the other hand, using the chain rule we get

$$z'(t) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = y \cdot 3t^2 + x \cdot 2t = 3t^4 + 2t^4 = 5t^4.$$

Example: A continuous mapping $c : I \to \mathbb{R}^n$ on an interval $I \subseteq \mathbb{R}$ is called a path or curve in $\mathbb{R}^n$, $(n = 2, 3)$. The function $c(t)$ will be given by a tuple of functions form.

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 $(g(t), h(t), k(t))$.

We can write

$$c(t) = g(t)\mathbf{i} + h(t)\mathbf{j} + k(t)\mathbf{k}, \quad \text{and if } c'(t_0) = g'(t_0)\mathbf{i} + h'(t_0)\mathbf{j} + k'(t_0)\mathbf{k}$$

exists and is nonzero, it represents the tangent vector to the curve $c(t)$ at the point $c(t_0)$.

For an example, consider the curve $c(t) = (t, \sqrt{1-t^2})$ in $\mathbb{R}^2$ and defined on the interval $[-1, 1]$. Observe that the curve $c(t)$ represents the **upper unit semicircle** centered at the origin.

You can verify easily that whenever $c'(t_0)$ exists and is nonzero, the tangent line to the circle $c(t)$ at the point $c(t_0)$ is the line that passes through the point $c(t_0)$ and is parallel to the tangent vector $c'(t_0)$. □

So far our example has nothing to do with the chain rule. Suppose $z = f(x, y)$ is a surface, and our curve given by $c(t) = (g(t), h(t), f(g(t), h(t))$ lies on the surface $z = f(x, y)$.

Let us compute the tangent vector to the curve at $c(t_0) = (x_0, y_0, z_0)$. 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

$$k'(t_0) = \frac{\partial f}{\partial x}(x_0, y_0)g'(t_0) + \frac{\partial f}{\partial y}(x_0, y_0)h'(t_0).$$

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 to the surface $z = f(x, y)$ at the point (x_0, y_0, z_0) on the surface 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

$$c'(t_0) = \left(g'(t_0), h'(t_0), \frac{\partial f}{\partial x}(x_0, y_0)g'(t_0) + \frac{\partial f}{\partial y}(x_0, y_0)h'(t_0)\right)$$

at the point $c(t_0) = (x_0, y_0, z_0)$ on the curve $c(t)$ lies on the plane, we need only check that its dot product with the normal vector is 0. But this is now clear. $\square$

The Gradient of scalar fields

When viewed as a row vector rather than as a matrix, the derivative matrix of $f : U \to \mathbb{R}$ at a point $(a_1, a_2, \dots, a_m) \in U \subseteq \mathbb{R}^m$ is called the gradient of f at $(a_1, a_2, \dots, a_m)$ and is denoted as $\nabla f(a_1, a_2, \dots, a_m)$. Thus

$$\nabla f(a_1, a_2, \dots, a_m) = \left(\frac{\partial f}{\partial x_1}(a_1, a_2, \dots, a_m), \dots, \frac{\partial f}{\partial x_m}(a_1, a_2, \dots, a_m) \right).$$

For the case $m = 2$, in terms of the coordinate vectors $\mathbf{i}$ and $\mathbf{j}$ the gradient can be written as

$$\nabla f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)\mathbf{i} + \frac{\partial f}{\partial y}(x_0, y_0)\mathbf{j}.$$

As we will see in the next slide, the gradient is related to the directional derivative in the direction v:

$$\nabla_v f = \nabla f \cdot v.$$

Another application of the chain rule: **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 $f(c(t))$ (observe that $\frac{d}{dt} f(c(t))$ at $t = 0$ is same as the directional derivative of f at (x_0, y_0, z_0) in the direction of the vector $v = (v_1, v_2, v_3)$):

$$\frac{d}{dt} f(c(0)) = \frac{\partial f}{\partial x}(x_0, y_0, z_0) v_1 + \frac{\partial f}{\partial y}(x_0, y_0, z_0) v_2 + \frac{\partial f}{\partial z}(x_0, y_0, z_0) v_3$$

and this can be rewritten as

$$\nabla_v f(x_0, y_0, z_0) = \frac{d}{dt} f(c(0)) = \nabla f(x_0, y_0, z_0) \cdot v.$$

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 a (differentiable) curve in $\mathbb{R}^3$. Then, by writing $c(t) = (x(t), y(t), z(t))$ and then using the chain rule for the derivative of $f(c(t))$ we obtain that

$$\frac{d}{dt} f(c(t)) = \nabla f(c(t)) \cdot c'(t).$$

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 assume that $\nabla f(x_0, y_0, z_0) \neq (0, 0, 0)$).

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:

$$S = \{(x, y, z) \mid f(x, y, z) = b\}$$

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 $s_0 \in S$ (if it exists) is a plane that contains the tangent lines at s_0 to all curves passing through s_0 and lying on S.

If $c(t)$ is a curve on the surface S given by $f(x, y, z) = b$, we see that $f(c(t)) = b$, and hence

$$\frac{d}{dt} f(c(t)) = 0.$$

On the other hand, by the chain rule,

$$0 = \frac{d}{dt} f(c(t)) = \nabla f(c(t)) \cdot c'(t).$$

Thus, if $s_0 = c(t_0) = (x_0, y_0, z_0)$ 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 $s_0 = (x_0, y_0, z_0)$.

Hence, if $\nabla f(x_0, y_0, z_0) \neq 0$, then $\nabla f(x_0, y_0, z_0)$ is perpendicular to the tangent plane of S at (x_0, y_0, z_0). How to determine a vector perpendicular to the tangent plane at the point (x_0, y_0, z_0) on the surface S given by $z = f(x, y)$? It is determined by $\nabla(z - f(x, y))$ at the point (x_0, y_0, z_0).

The equation of the tangent plane

Since we know that the gradient of f is normal to the level surface $S = \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = c\}$ (provided the gradient is nonzero), it allows us to write down the equation of the tangent plane of S at the point $s_0 = (x_0, y_0, z_0)$. The equation of this plane is (here f_x, f_y, f_z denote the respective partial derivatives of f)

$$f_x(x_0, y_0, z_0)(x-x_0)+f_y(x_0, y_0, z_0)(y-y_0)+f_z(x_0, y_0, z_0)(z-z_0) = 0.$$

For the curve $f(x, y) = c$, by considering y as a function of x (implicitly) we obtain (by differentiating $f(x, y)$ with respect to x and using the chain rule) that

$$f_x(x_0, y_0) + f_y(x_0, y_0)\frac{dy}{dx} = 0 \Rightarrow \frac{dy}{dx} = -\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}$$

and hence the equation of the tangent line to the curve $f(x, y) = c$ passing through (x_0, y_0) is

$$y - y_0 = -\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}(x - x_0),$$

that is, $f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = 0.$

Gravitational force as gradient of the potential energy

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 for $\sqrt{x^2 + y^2 + z^2}$. This notation is very useful.

For instance, the 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 called the potential function/energy.

Keeping our previous discussion in mind, we know that if

$$V = -\frac{GMm}{r}, \quad \mathbf{F} = -\frac{GMm}{r^3}\cdot\mathbf{r} = -\nabla V.$$

What are the level surfaces of V? Clearly, r must be a constant on these level sets, so the level surfaces are spheres.

Since $\mathbf{F} = -\nabla V$, we see that the gravitational force $\mathbf{F}$ is orthogonal to the sphere and points towards the origin. $\square$

Review - problems involving the gradient

Exercise 1: Find the points on the hyperboloid $x^2 - y^2 + 2z^2 = 1$ where the normal line is parallel to the line that joins the points $(3, -1, 0)$ and $(5, 3, 6)$.

Solution: The hyperboloid is an implicitly defined surface. A normal vector at a point (x_0, y_0, z_0) on the hyperboloid is given by the gradient of the function $x^2 - y^2 + 2z^2$ at (x_0, y_0, z_0):

$$\nabla f(x_0, y_0, z_0) = (2x_0, -2y_0, 4z_0).$$

We require this vector to be parallel to the line joining the points $(3, -1, 0)$ and $(5, 3, 6)$. This line lies in the same direction as the vector $(5 - 3, 3 + 1, 6 - 0) = (2, 4, 6)$. Thus we need only solve the equations

$$(2x_0, -2y_0, 4z_0) = \lambda(2, 4, 6),$$

for some $\lambda \in \mathbb{R}$ such that the point (x_0, y_0, z_0) lies on the hyperboloid. By solving the above equations, we find that $x_0 = \lambda$, $y_0 = -2\lambda$ and $z_0 = (3/2)\lambda$. Substituting x_0, y_0, z_0 in the equation of the hyperboloid yields $\lambda = \pm\sqrt{2/3}$. $\square$

Problems involving the gradient, continued

Exercise 2: Find the directions in which the directional derivative of $f(x, y) = x^2 + \sin xy$ at the point $(1, 0)$ has the value 1.

Solution: We compute ∇f first:

$$\nabla f(x, y) = (2x + y \cos xy, x \cos xy),$$

so at $(1, 0)$ we get, $\nabla f(1, 0) = (2, 1)$.

To find the directional derivative in the direction $v = (v_1, v_2)$ (where v is a unit vector), we simply take the dot product with the gradient:

$$\nabla_v f(1, 0) = \nabla f(1, 0) \cdot v = (2, 1) \cdot (v_1, v_2) = 2v_1 + v_2.$$

This will have value "1" when $2v_1 + v_2 = 1$, subject to $v_1^2 + v_2^2 = 1$, which yields $v_1 = 0, v_2 = 1$ or $v_1 = 4/5, v_2 = -3/5$. $\square$

Review of the gradient

Exercise 3: Find $D_u F(2,2,1)$ where D_u denotes the directional derivative of the function $F(x,y,z) = 3x - 5y + 2z$ and u is the unit vector in the direction of 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 $g(x,y,z) = x^2 + y^2 + z^2 = 9$ at $(2,2,1)$ is given by

$$u = \frac{\nabla g(2,2,1)}{\|\nabla g(2,2,1)\|}.$$

We see that $\nabla g(2,2,1) = (4,4,2)$ so the corresponding unit vector u is $\frac{1}{3}(2,2,1)$.

To get the directional derivative we simply take the dot product of $\nabla F(2,2,1) = (3,-5,2)$ with $u = \frac{1}{3}(2,2,1)$:

$$D_u F(2,2,1) = (3,-5,2) \cdot \frac{1}{3}(2,2,1) = -\frac{2}{3}.$$

$\square$