Contents

1 Sequences and series **1**
- 1.1 Review and properties of sequences . . . 1
- 1.2 Cauchy sequences . . . 4
- 1.3 Series and convergence tests . . . 5

2 Continuous functions **9**
- 2.1 Basic properties. . . . 9
- 2.2 Extreme value theorem . . . 9
- 2.3 Intermediate value theorem . . . 10

3 Differentiation **11**
- 3.1 Basic properties . . . 11
- 3.2 Taylor's theorem . . . 13

4 Integration **15**

5 Power series **20**
- 5.1 Basic properties (Review) . . . 20
- 5.2 Exponential and trigonometric functions . . . 21

Note. I try to keep these notes as self contained as possible, but if I missed something you can likely find it on level 4, 6.1 or 6.2 on this website (which have the function of being A level proof documents). The main purpose of this will be to turn the "yeah obviously, just look at it" proofs that were fine in earlier levels into more rigorous proofs.

1 Sequences and series

1.1 Review and properties of sequences

Definition. A sequence is an ennumerated list of real numbers. We can have sequences on any set not just $\mathbb{R}$.

We have learned a million times now (Numbers and sets, level 4, level 6) what it means for a sequence to converge to a value.

We say a sequence goes to infinity if its reciprocal goes to 0, equivalently if for all L the terms are eventually larger in absolute value than L. When we say "eventually", we mean after N terms where N is an integer depending only on our threshold ε or L.

Note that when we use an epsilon threshold we can replace it by any non-zero positive multiple of ε and it won't make a difference.

Example. $x_n = \frac{1}{n}$ converges to 0 (See numbers and sets)

$x_n = \frac{1}{2^n}$ also converges to 0 (as for all terms $0 < \frac{1}{2^n} < \frac{1}{n}$, alternatively set $N = \max\left(1, 1 + log_2\left(\frac{1}{\varepsilon}\right)\right)$ so we know that for all ε we will be within ε of 0 after N terms.)

$x_n = i * n$ diverges to infinity, for any threshold L it clears the circle of radius L centered at the origin after (any integer greater than L which exists, see numbers and sets) terms.

$x_n = \sin(n)$ oscillates around and does not converge or go to infinity

Same for $x_n = (-1)^n$

We now have some propositions that are obvious enough that we have assumed them to be true in the past.

Proposition. Limits are unique so we can talk about "the" limit.

Proof. Suppose $x_n \to L_1$ and $x_n \to L_2$ for a contradiction, then x_n cannot simaltaneously get to and stay within $\left|\frac{L_1+L_2}{2}\right|$ of both limits, so for any epsilon less than $\left|\frac{L_1+L_2}{2}\right|$ it does not work unless $L_1 = L_2$. In particular, we use the triangle inequality: $|L_1 - L_2| \leq |L_1 - x_n| + |x_n - L_2| \leq 2\varepsilon$, and since this is true for every ε no matter how small we indeed must have $L_1 = L_2$.

□

Imagine this as a picture: A sequence cannot get to and stay within two different disjoint bands.

Proposition. Let $x_n,\ y_n$ be real sequences, then if $x_n \leq y_n$ for every n and we have $x_n \to x$ and $y_n \to y$ then $x \leq y$. In particular, if $x_n \leq y_n \leq z_n$ then $x \leq y \leq z$ if $z_n \to z$. This is sort of not obvious because if we replace this theorem with all strict inequalities it is false (You can construct a sequence strictly smaller than another but they converge to the same limit).

Proof. We just need to show that y is not smaller than x, which is true because $y_n - x_n$ has all non-negative terms so if it converged to -L then for $\varepsilon < \frac{L}{2}$ we have a problem.

□

Proposition. Convergence of complex sequences is equivalent to convergence of their real and imaginary parts.

Proof. If the complex sequence converges then $|z| = \sqrt{Re(z)^2 + Im(z)^2} < \varepsilon$ so $|Re\,(z)| < \varepsilon$. Also, if the real and imaginary parts are less than ε then

$$|z| = \sqrt{Re(z)^2 + Im(z)^2} \leq \sqrt{2\varepsilon^2} = \sqrt{2}\varepsilon$$

□

Proposition. If $x_n \to x$ and $y_n \to y$ then $x_n + y_n \to x + y$

Proof. Eventually $|x_n - x| < \varepsilon,\ |y_n - y| < \varepsilon$ after $\max\left(N_x\left(\varepsilon\right), N_y\left(\varepsilon\right)\right)$ terms, so by the triangle inequality, $|x_n + y_n - x - y| < 2\varepsilon$.

□

Proposition. If $x_n \to x$ and $y_n \to y$ then $x_n y_n \to xy$

Proof. Eventually,

$$|x_n y_n - xy| \leq |x_n y_n - x_n y| + |x_n y - xy| \leq x_n\varepsilon + y\varepsilon = \varepsilon(x_n + y)$$

however we need to show that x_n has some bound M. This is true because eventually $x_n < x+\varepsilon$ so it is eventually bounded after say N terms, and then it is bounded by $\max(x_1, x_2, \ldots, x_N)$. Bounded meaning there is an M such that all terms are less than M in absolute value.

□

[Lecture 1 ends]

Proposition. If $x_n \to x$ and x_n is never 0 and x is not 0 then $\frac{1}{x_n} \to \frac{1}{x}$

Proof.

$$\left|\frac{1}{x_n} - \frac{1}{x}\right| = \left|\frac{x - x_n}{x x_n}\right|$$

Evenetually $|x x_n| \geq \frac{|x|^2}{2}$ (take any $\varepsilon < \frac{|x|}{2}$) and $|x x_n| < \varepsilon$ so $\left|\frac{1}{x_n} - \frac{1}{x}\right| \leq \frac{2\varepsilon}{x^2}$, but x is constant so done.

□

Definition. Any sequence which is increasing or decreasing is called monotone

Proposition. Monotone bounded sequences are convergent

Proof. We proved this in numbers and sets (also in level 4). The proof is very simple, just consider the supremum of the sequence (or the infimum if it is decreasing).

□

Proposition. Every sequence (not necessarily bounded) has a monotone subsequence

Proof. We've proven this in level 4. The idea was to consider peaks (values which are never exceeded again), and note that there are either infinitely many (in which case we're done) or an infinitely non-increasing sequence is forced to exist.

□

Definition. Let $n_1 < n_2 < \dots$ be a sequence of natural numbers, then x_{n_k} is a subsequence of x_n. This definition is obvious though.

Proposition. Every subsequence of a convergent sequence is convergent with the same limit as the main sequence

Proof. If the main limit is L, then since $n_k > k$ as n is an increasing sequence of natural numbers, then for each ε if there is an N such that the main sequence stays within ε of L, then after that for all k greater than that N, x_{n_k} is x_j for some $j > N$ so it is also within ε of L.

□

Proposition. (nested interval property) Let I_n be a sequence of closed intervals in $\mathbb{R}$ that are each contained within the previous one. Then set $I_n = [a_n, b_n]$. Then if $a_n - b_n \to 0$ then the intersection of all I_n's will be exactly one point.

Proof. Since a_n is a monotone sequence bounded above by b_1(by nestedness) it converges to some a that is its least upper bound. And $b_n \to b$. But then a=b by additive properties of limits and $a_n - b_n \to 0$.

For all n, $b_n \geq a \geq a_n$ and so the interval contains a, thus the intersection of all these intervals contains a. It contains no other points because, say, if it contained $a + \varepsilon$ for some non-zero epsilon then by definition of convergence we eventually have $b_n \leq a + \frac{\varepsilon}{2}$ which is a contradiction.

□

We will now prove the Bolzano Weierstrass theorem, which says that every bounded sequence has a convergent subsequence. We proved it in level 4 but we will prove it in a different way here.

Proof. Let M be a bound for a bounded sequence. Let $I_1 = [-M,\ M]$, then this contains every point of our sequence, so $a_1 = -M,\ b_1 = M$. Now take $c = \frac{a_1+b_1}{2} = 0$. Then if I split I_1 into $[a_1, c]\,,\ [c,\ b_1]$ at least one of these halves will have infinitely many terms, and we will call this half I_2. Do this repeatedly to generate a sequence of nested intervals I_n which has a unique common point by the previous lemma.

Construct your sequence by always picking the k'th element of the sequence that is contained in I_k to ensure it is a subsequence (each one will be strctly later in the main sequence than the previous one: the k+1'th one in the k+1'th band occurs at at least the k+1'th position in the k'th band), then it converges to this unique common point. For example, $(-1)^n$ is bounded and not convergent but the subsequence with every other term is convergent.

□

Note that we just know we have at least one convergent subsequence, it is almost never unique.

[Lecture 2 ends]

1.2 Cauchy sequences

Definition: A Cauchy sequence is a sequence where the elements are getting closer to eachother. Ie, for each ε there exists an N such that for all m>n>N, $|a_n - a_m| < \varepsilon$.

Ok, for real numbers these obviously converge in the real numbers (since we go into an arbitrarily small band which means convergence) but we will prove that it is in fact equivalent to convergence in $\mathbb{R}$ but not in general number systems.

Example. $a_n = \frac{1}{n}$ is Cauchy because for any ε then for n large enough $\frac{1}{n} - \frac{1}{m} < \frac{1}{n} < \varepsilon$.

$a_n = (-1)^n$ is not Cauchy because I can always force the difference to be 2.

Define a sequence by pi like (3, 3.1, 3.14, 3.141, 3.1415, ...), then this is cauchy in $\mathbb{Q}$ but does not converge in $\mathbb{Q}$, it only converges in $\mathbb{R}$.

Proposition. If a sequence is Cauchy then it is bounded

Proof. Fix $\varepsilon = 1$, then there is an N such that for all m>N, $|a_m - a_N| < 1$, so the sequence is eventually bounded by something independent of m and therefore bounded.

□

Proposition. A complex sequence is cauchy if and only if its real and imaginary parts are cauchy

Proof. For any ε there is an N such that for $m>n>N$ we have

$$|Im\,(x_m) - Im\,(x_n)| < \frac{\varepsilon}{\sqrt{2}}, |Re\,(x_m) - Re\,(x_n)| < \frac{\varepsilon}{\sqrt{2}}$$

and therefore

$$|x_m - x_n| = \sqrt{|Im\,(x_m) - Im\,(x_n)|^2 + |Re\,(x_m) - Re\,(x_n)|^2} < \varepsilon$$

□

Proposition. If a sequence converges it is cauchy

Proof. For each ε there is N depending on ε such that for $n, m>N$

$$|x_n - x_m| \leq |x_n - x| + |x - x_m| < 2\varepsilon$$

by convergence so it is cauchy.

□

Proposition. Real cauchy sequences converge (And by a previous proposition this implies that so do complex cauchy sequences)

Proof. Find a subsequence x_{n_k} which converges by the Bolzano-Weierstrass theorem since the sequence is bounded. Let x be the limit of that subsequence. Fix $\varepsilon \to 0$. Then let k be so large that for $m > n \geq n_k$ we have that $|x_n - x_m| < \varepsilon$ and also $|x_{n_k} - x| < \varepsilon$. Now for each m that is that large, by the triangle inequality, setting $n = n_k$, $|x - x_m| \leq |x_n - x_m| + |x - x_n| < 2\varepsilon$.

□

1.3 Series and convergence tests

Definition. Define an infinite sum series as the limit of its partial sums when it exists, ie $\sum_{n=1}^{\infty} a_n = \lim_{k\to\infty} \sum_{n=1}^{k} a_n$. The sum converges if this limit exists.

[Lecture 3 ends]

Example. $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ converges because

$$\sum_{n=1}^{k} \frac{1}{n(n+1)} = \sum_{n=1}^{k} \frac{1}{n} - \frac{1}{n+1} = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots + \frac{1}{k} - \frac{1}{k+1} = 1 - \frac{1}{k+1} \to 1$$

So the series converges to 1.

Example. $\sum_{n=1}^{\infty} n$ clearly does not converge, it definitely does not converge to $-\frac{1}{12}$.

Now we will do some totally trivial and obvious propositions. In fact we have seen these in level 4, so this is just to refresh your memory.

Proposition. Fix some complex numbers A and B. If I have two series $\sum a_n, \sum b_n$ that converge then we have that $\sum Aa_n + Bb_n$ converges

Proof. This is directly from basic properties of convergent sequences.

□

Proposition. If two sums agree after some point then they either both converge or neither converge.

Proof. Eventually the partial sums will just differ by a constant so it is clear their convergence behavior will be the same.

□

Proposition. A sum converges only if the terms approach 0, but this is not a sufficient condition (eg the harmonic series $\sum \frac{1}{n}$ diverges).

Proof. If the terms do not approach 0 there is an ε>0 such that the terms are larger in absolute value than ε for infinitely many n. But $|a_n| > \varepsilon$ implies that $|s_n - s_{n-1}| > \varepsilon$, where s_n is the sequence of partial sums. This implies that s_n is not cauchy since that happens for infinitely many n, completing the proof.

□

Proposition. Suppose all terms in a series are positive, then if they are bounded above by a convergent series that series is convergent.

Proof. The partial sums are bounded above by the sum of the convergent series in question so the result follows from the monotone convergence theorem (In this course when I say MCT I mean the sequence one not the integral one).

□

I'm pretty sure we did all this in numbers and sets. I remember we used this to show $\sum \frac{1}{n^2}$ converges and now we're doing it again for some reason.

Proposition. (Root test) If we are summing non-negative terms a_n, then look at $\sqrt[n]{a_n}$ and suppose it converges to a limit a. Then if a<1 the series converges and if a>1 the series diverges. The idea is we are comparing it to a geometric series. We can't say much if a=1.

Proof. If a>1 then there is an N such that $a_n^{\frac{1}{n}} > 1$ whenever n>N. This implies that $a_n > 1$ and the sum of this clearly diverges.

If a<1 fix an r such that a<r<1, then eventually $\sqrt[n]{a_n} < r$, so we can eventually bound a_n by a geometric series with common ratio r, so apply the previous proposition.

□

Example. For $\sum 2^{-n}$ this confirms convergence. For $\sum 4^n$ this confirms divergence. a=1/2 and a=4 respectively in these cases. For $\sum \frac{1}{n}$ and $\sum \frac{1}{n^2}$ the test is inconclusive.

Proposition. (Ratio test) If $\frac{a_{n+1}}{a_n} \to a$ as $n \to \infty$ this converges if $|a| < 1$ and diverges if $|a| > 1$

Proof. See level 4 for the proof of this as well as the proof that the above test works with absolute values and non-positive terms.

□

Example. $\sum \frac{n}{2^n}$ converges as the ratio between consecutive terms approaches $\frac{1}{2}$.

[Lecture 4 ends]

Proposition: If $f(x)$ is a decreasing non-negative function then $\sum f(n)$ converges if and only if $\int_1^\infty f(x)\, dx$ exists and is finite

Proof. The reason why this proposition is true is by looking at areas by looking at rectangles and noting that the difference (area of the green rectangles in the image below) is bounded above by $f(1)$ since the function is decreasing. Figure 1 shows why the proposition must be true.

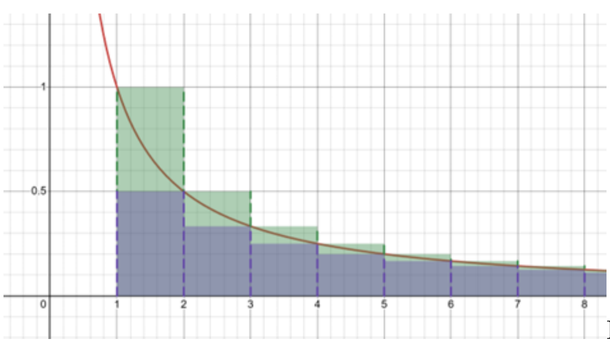

Figure 1

If the sum converges then the sum of the purples converge. The sum of the greens converge by monotone convergence theorem so the integral converges. Conversely if the integral converges subtract the sum of the green to get the sum of the purple which therefore also converges.

□

Example. By the integral test we have that $\sum \frac{1}{n^s}$ converges if and only if s>1.

Example. $\sum \frac{1}{nlog(n)}$ diverges because its integral is $log(log(n))$ which goes to infinity (Very!) slowly.

Proposition. (Cauchy condensation test) Let a_n be a decreasing non-negative sequence. Then $\sum a_n$ converges if and only if $\sum 2^n a_{2^n}$ converges.

Proof. Since the sum is decreasing we have the following.

$$a_1 + a_2 + (a_3 + a_4) + (a_5 + a_6 + a_7 + a_8) + \dots \geq a_1 + a_2 + 2a_4 + 4a_8 + \dots$$

So if the left hand side is finite so is the right hand side.

On the other hand

$$a_1 + (a_2 + a_3) + (a_4 + a_5 + a_6 + a_7) + \dots \leq a_1 + 2a_2 + 4a_4 + 8a_8 + \dots$$

In the end it is clear that both are finite if the other one is finite so that completes the proof.

You can use this to show that $\sum \frac{1}{nlog^2(n)}$ converges, for example.

□

Proposition. (Alternating series test) Suppose that a_n is decreasing and tends to 0, then $\sum (-1)^n a_n$ converges. For example, $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \dots$ converges. In fact this converges to $ln(2)$ by Taylor series and Abel's theorem which is proven later in these notes.

Ok, this is obviously true by this figure 2

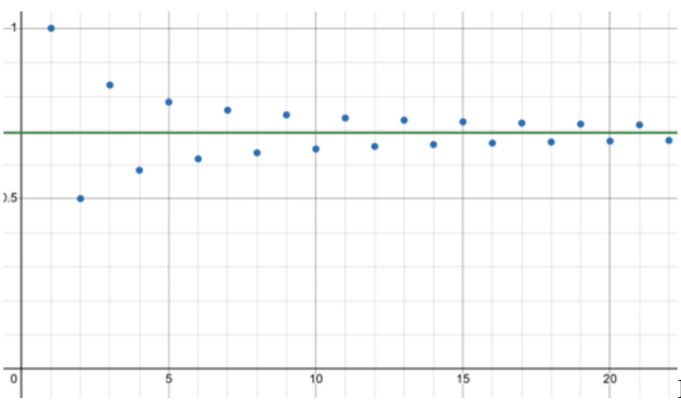

Figure 2

Now lets prove it.

[Lecture 5 ends]

Proof. Define $S_n = \sum_{r=1}^{n} (-1)^{1+r} a_r$

Note that $S_{2n} = (a_1 - a_2) + (a_3 - a_4) + (a_5 - a_6) + \dots + (a_{2n-1} - a_{2n})$ is increasing. Similarly

$S_{2n+1} = a_1 - (a_2 - a_3) - (a_4 - a_5) - (a_6 - a_7) - \dots - (a_{2n} - a_{2n+1})$ is decreasing.

These are monotone and we show they are bounded as follows. S_{2n+1} is bounded above by a_1 and S_{2n} is bounded above by S_{2n+1} since the 2n+1'th term is positive and thus it is bounded above by a_1. Similarly both are bounded below by 0. Hence it is clear they both converge by the monotone convergence theorem and that they converge to the same limit since their differences converge to 0.

□

Proposition. (Dirichlet test) Let b_n be decreasing and let a_n be a sequence such that its partial sums make a bounded sequence, then the series $\sum a_n b_n$ converges.

Proof. Let $S_n = \sum_{j=1}^{n} a_j$ and we say $S_0 = 0$. Now we do some algebra trickery:

$$\sum_{j=m}^{n} a_j b_j = b_n\left(s_n - s_{n-1}\right) + b_{n-1}\left(s_{n-1} - s_{n-2}\right) + \cdots + b_m\left(s_m - s_{m-1}\right)$$

$$= s_n b_n - s_{m-1} b_m - s_m b_{m+1} + s_{m+1} b_{m+1} - s_{m+1} b_{m+2} + \cdots + s_{n-1} b_{n-1} - s_{n-1} b_n$$

$$= s_n b_n - s_{m-1} b_m + \sum_{j=m}^{n-1} s_j\left(b_j - b_{j-1}\right)$$

Now since s is bounded, we can safely write the following. Note that it is not always valid to split lim(a+b) into lim(a)+lim(b) as the left may exist but the right may be, for example, infinity plus minus infinity, but in this case by boundedness we are fine as the non-sum limits exist. We apply the above identity to m=1:

$$\lim_{n\to\infty} \sum_{j=1}^{n} a_j b_j = \lim_{n\to\infty} s_n b_n + \lim_{n\to\infty} \sum_{j=1}^{n} s_j\left(b_j - b_{j+1}\right)$$

Now

$$\sum_{j=1}^{n} |s_j|\,|\left(b_j - b_{j+1}\right)| \le M \sum_{j=1}^{n} |\left(b_j - b_{j+1}\right)| = M \sum_{j=1}^{n} \left(b_j - b_{j+1}\right) = M b_1$$

By the method of differences/telescoping sum, where M is a bound for s, and so the sequence is absolutely convergent. Then since $\lim_{n\to\infty} s_n b_n$ is 0 it follows that $\lim_{n\to\infty} \sum_{j=1}^{n} a_j b_j$ exists as required.

□

A sum $\sum a_n$ is said to be absolutely convergent if $\sum |a_n|$ converges.

We proved this in level 4, but in an effort to make these notes as self contained as possible, I will sketch how the proof went.

Sketch of proof. We let b_n be a rearranging of the a_n's and let N be a threshold so that $\sum_{>N} |a_n| < \varepsilon$ and M be so large so that $b_{1\ldots M}$ contains $a_{1\ldots N}$, then it follows that $\sum_{n=1}^{M} b_n$ is within ε of $\sum a_n$ by the infinite sum variant of the triangle inequality.

Proposition. Absolute convergence implies convergence.

Proof. Let $s_k = \sum_{n=1}^{k} a_n$ and $r_k = \sum_{n=1}^{k} |a_n|$. The r_k's converge and are thus Cauchy, so for any N and M large enough we have $\sum_{n=N}^{M} |a_n| < \varepsilon$ for each fixed ε. Therefore by the triangle inequality $\left|\sum_{n=N}^{M} a_n\right| < \varepsilon$ for large enough N and M, so the s sequence is Cauchy and hence convergent. A general principle based on this is that you can use the triangle inequality for infinite sums.

□

Definition. We say a sequence that is not absolutely convergent is conditionally convergent.

Oh I just realized the example where rearranging terms fails in level 4 + the proof it works for absolute convergent series combined means that the $1 - \frac{1}{2} + \frac{1}{3} + \ldots$ is conditionally convergent which gives another proof that the harmonic series diverges so now we know 3 proofs (Log bound, powers of 2 proof, and this).

We can say something even stronger: If a series of real numbers is conditionally convergent then there is a rearrangement that can make it converge to *any* value. The idea is that such a series would have its positive and negative terms each add to infinity, so we could use these terms to make it oscillate around any value and eventually get and stay as close as we want to that value.

2 Continuous functions

We have discussed in levels 3-7 properties of limits of functions and continuous functions

2.1 Basic properties.

Definition. If we have a subset of the complex numbers then an isolated point is a point in this set with no other points in the set in some neighbourhood around it. An accumulation point is a point that is not an isolated point, or is a limit of sequence of points in the set, such as the boundary of the unit circle being accumulation points of its interior. Equivalently, if for any δ there is a point in the set within δ of that point it is an accumulation point.

[Lecture 6 ends]

Recall that if $f{:}\mathbb{C} \to \mathbb{C}$ then f is continuous at a point x in the domain if for all $\varepsilon > 0$ there is a $\delta > 0$ such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$, and that we also say that if this holds then $f(y)$ tends to $f(x)$ as y tends to x, since if there is a sequence tending to x and this holds then f(that sequence) tends to $f(x)$. Similarly the sequential statement does not hold if $f(y)$ does not approach $f(x)$ as y approaches x, so these two ideas are equivalent.

We say it goes to infinity if its reciprocal goes to 0, but now we are talking about accumulation points of the domain since a function cannot be infinity in the domain.

Note that a function is trivially continuous at an isolated point in the domain since $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$ is trivial for small enough δ – by isolation if the ball of radius δ around x only contains x then $|f(x) - f(y)|$ is automatically 0 and therefore $< \varepsilon$.

Now using our theorems about sequence limits we have that sums or products of continuous functions are continuous, as well as quotients when we are not dividing by 0 near the point in question. It therefore follows that polynomials are continuous since $f(x) = x$ is clearly continuous.

[Lecture 7 ends]

Note that $f(z) \to y$ as $z \to a$ if and only if for every sequence $z_n \to a$, $f(z_n) \to y$. The proof is because if there is a sequence such that $z_n \to a$, $f(z_n)$ does not $\to y$, then that means that for some ε there are points z within any δ of a such that $f(z)$ fails to be within ε of y. Also, if $f(z)$ does not $\to y$ as $z \to a$ then there is an ε such that for every δ we can find a z so that have $|z - a| < \delta$ but $|f(z) - f(a)| > \varepsilon$. Now pick a sequence $\delta_n \to 0$ and pick such a z_n for each δ_n, now clearly $z_n \to a$ but $f(z_n)$ does not approach $f(a)$. Applying this to the case f(a)=y gives that continuity and sequential continuity (ie, saying a function is continuous at a if for every sequence $z_n \to a$ we have $f(z_n) \to f(a)$) are equivalent.

Example. $\sin\left(\frac{1}{x}\right)$ is continuous everywhere but 0 where it is not continuous because on any neighbourhood about 0 it takes every value between -1 and 1.

Proposition. If $g(x)$ is continuous at x and $f(y)$ is continuous at $y = g(x)$ then $f(g(x))$ is continuous at x.

Proof. For all ε we can find a η such that

$$|y - g(x)| < \eta \Rightarrow |f(y) - f(g(x))| < \varepsilon$$

and then for this η we can find a δ such that

$$|z - x| < \delta \Longrightarrow |g(z) - g(x)| < \eta$$

which in turn implies that $|f(g(z)) - f(g(x))| < \varepsilon$. We have done this argument in earlier levels but using words like "arbitrarily close" or "sufficiently close" and now we are doing it symbolically.

□

2.2 Extreme value theorem

Theorem. (Extreme value theorem) Any continuous function defined on a closed and bounded set attains its maximum, so it is actually a maximum and not just a supremum

Proof. I proved this in level 6.2. However, to keep this self contained I will sketch the proof.

We need to suppose it is unbounded then construct a sequence of points where the function is $> n$ at a_n and get a contradiction by noting that there is a convergent subsequence, its limit is in the domain by closure, and by continuity f must be infinite at that point.

We show it attains its maximum by finding a_n such that instead of the function being $> n$ it is $> M - \frac{1}{n}$ where M is its supremum.

□

Proposition. The image of closed and bounded sets under continuous complex functions is closed and bounded. This actually implies the extreme value theorem as the closed image has the maximum and the maximum is in the image hence attained by definition.

A closed set contains its supremum because otherwise its complement would contain its supremum and by openness contain a ball around it that means something strictly smaller than the supremum is an upper bound for the set which is a contradiction.

[Lecture 8 ends]

Proof. The "bounded" part follows from the extreme value theorem.

As a reminder closed (as in complement is open) is equivalent to closed (as in contains its limit points) because if the complement is open then a limit point not contained in the set has to have an open ball of stuff not in the set around it so it is not a limit point, and if all the limit points are in the set then any point not in the set has to have a ball of points not in the set around it, otherwise it would be a limit point.

Now we do the closed part. Take y_n in f(x) and suppose its limit is y, then we want to show that y belongs to f(x). The sequence of pre-images has a convergent subsequence that converges to some x, and then by continuity $f(x) = y$.

□

2.3 Intermediate value theorem

Theorem. If $f : [a, b] \to \mathbb{R}$ is continuous then given y between $f(a)$ and $f(b)$ there exists a c between a and b such that $f(c) = y$. So continuous functions map closed intervals to closed intervals, not just closed sets.

This one is obvious, and in fact we have secretly used before, but it's not too obvious since it is false over $\mathbb{Q}$ for example, but if you have a sense of how the real numbers work then it is indeed obvious.

Proof. If f is constant then this is trivially true. If not then we can take f(a) to be the minimum and f(b) to be the maximum since otherwise we could just take the minimum and maximum and work with those points. Set $S = \{x \in [a, b], \ f(x) \leq y\}$. Then a is in S and S is non-empty and S is bounded above by b so S has a supremum d. Now $f(d) \leq y$ because if $f(d) > y$ then $y - f(d)$ is positive, and we can call the difference ε. Now by continuity there is a δ such that for z within δ of d $y - f(z)$ is positive. But then this contradicts d being the supremum, because $d - \delta$ is an upper bound.

On the other hand, suppose that $f(d) < y$, then for the same reason, if $y - f(d) = -\varepsilon$ then for some δ we have that $d + \delta$ is in S so d is not an upper bound for S, which is another contradiction since we supposed d was the supremum so we are done.

□

Example. Now we know that $\sqrt{2}$ exists because x^2 hits 2 somewhere on $[1, 2]$. Ok but we already knew that $\sqrt{2}$ exists anyway but sometimess we like to be *unnecessarily* pedantic about stuff. In fact all n'th roots and logs and stuff exist.

Definition. A function $[a,b] \to \mathbb{R}$ is called monotone if it is increasing or decreasing, meaning that if $x_1 \leq x_2$ then $f(x_1) \leq f(x_2)$. We say a function is strictly increasing if $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$.

Proposition. Strictly increasing or strictly decreasing continuous functions have continuous strictly monotone inverses.

Proof. By the IVT each point has a pre-image and by strictly-monotone-ness it is unique, so there is an inverse. The hard (but kind of obvious if you think about it) part is to show that it is actually continuous and strictly monotone.

[Lecture 9 ends]

Take f to be strictly increasing because we can do a symmetric argument for the other case.

Lets show that its inverse is strictly increasing. If f^{-1} is not strictly increasing then there exist points such that $a < b$ but $f^{-1}(a) \geq f^{-1}(b)$ which contradicts f being strictly increasing. Doing f to both sides gives a contradiction.

Fix $y_0 \in [c,d]$ where we consider f as a function from $[a,b] \to [c,d]$. Then we have 2 cases. Lets say that x_0 is the pre-image of y_0.

Case 1: $x_0 \in (a,b)$, then for any ε there is an η such that for some $0 < \eta \leq \varepsilon$ we have it small enough that $[x_0 - \eta, x_0 + \eta] \subseteq [a,b]$ by openness.

Now $x_0 - \eta < x_0 < x_0 + \eta$ so $f(x_0 - \eta) < y_0 < f(x_0 + \eta)$ as f is strictly increasing. So if we take δ to be $\min\{f(x_0 + \eta) - y_0, y_0 - f(x_0 - \eta)\}$ then indeed we get that $|y - y_0| < \delta$ implies that $\left|f^{-1}(y) - x_0\right| < \eta \leq \varepsilon$ so we have continuity so done.

Case 2: $x_0 = a$ or $x_0 = b$. By symmetry, pick $x_0 = a$.

We just apply case 1 to the function that has had its domain extended to the left of a as something like $f(x) = \begin{cases} x + c - a \text{ if } x < a \\ \text{normal } f(x) \text{ otherwise} \end{cases}$ and show that this is continuous at a by step 1 and thus normal f is continuous there too.

□

3 Differentiation

3.1 Basic properties

Differentiation is the same limit that we already know from A levels, and we recall that it has to exist from the left and from the right for a function to be differentiable there, and from all directions to be complex differentiable – the complex conjugate is not complex differentiable because from the i direction it looks like -x and from the real direction it looks like +x so the limit is -1 from some directions and 1 from others. But we will do more stuff on complex differentiable functions in the complex analysis course.

We cannot define a derivative at an isolated point in a domain, for there to be a limit there has to be points arbitrarily close.

Basic properties like the product rule are easy to derive by messing with the limits - we did this in earlier levels so I leave it as an exercise.

[Lecture 10 ends]

Proposition. We have the chain rule, which says

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

Proof. We proved this in Level 4. It is easy to give a "proof" using limits that has a divide by 0 issue. The idea to get around this was to mess with the function

$$d(y) := \begin{cases} \frac{g(y)-g(f(x))}{y-f(x)} & y \neq f(x) \\ g'(f(x)) & y = f(x) \end{cases}$$

and split into cases $g\left(f\left(x+h\right)\right) = g\left(f\left(x\right)\right)$ and $g\left(f\left(x+h\right)\right) \neq g\left(f\left(x\right)\right)$

□

Remember that the derivative equivalently says

$$f\left(a+h\right) = f\left(a\right) + f'\left(a\right)h + o\left(h\right)$$

Example. $x * \sin\left(\frac{1}{x}\right)$ is differentiable everywhere but 0 where $\lim_{h\to 0}\frac{f(h)}{h} = \lim_{h\to 0}\sin\left(\frac{1}{h}\right)$ which is not a limit that exists.

Here's another obvious one.

Proposition. Differentiability implies continuity

Proof.

$$\lim_{x\to a} f(x) = \lim_{h\to 0} f\left(a+h\right) = \lim_{h\to 0} f\left(a\right) + hf'\left(a\right) + o(h) = 0$$

□

Proposition. Rolle's theorem, ie if f is continuous on $[a,b]$ and differentiable on (a,b) and $f(a) = f(b)$ then $f'(c) = 0$ for some $c \in (a,b)$

Proof. We did this in level 6.2, but again I will sketch the proof to keep these notes self contained. If f is constant the theorem is trivial, and otherwise the extreme value theorem asserts the existence of either an interior minimum or maximum, at which point messing with limits implies the derivative is 0.

□

Theorem. Mean value theorem, if f is continuous on $[a,b]$ and differentiable on (a,b) then $f'(c) = \frac{f(b)-f(a)}{b-a}$ for some $c \in (a,b)$

Proof. Apply rolle's theorem to $g\left(x\right) = f\left(x\right) - \frac{f(b)-f(a)}{b-a}x$

□

[Lecture 11 ends]

A corollary of the mean value theorem is the already known result that if the derivative is non-negative everywhere on an interval the function is increasing on the interval - if it were decreasing the derivative would be negative at some point. We also get that zero derivative on an interval (NOT an arbitrary subset of the real numbers) implies our function is constant.

Theorem. (1D inverse function theorem) If we have a function $f : [a,b] \to \mathbb{R}$ continuous and differentiable on (a,b) and $f'\left(x\right) > 0$ everywhere on (a,b), then f is a bijection and furthermore f has an inverse which is continuous and differentiable except possibly at its end points.

In Analysis II (And Level 8.2) we will prove a generalization of the inverse function theorem to higher dimensions.

We note that the derivative of our inverse function is the reciprocal of the derivative of the original function at the corresponding point – of course, we will show this.

Proof. The bijection part and continuous part follows from a previous theorem because f is strictly increasing by a previous proposition. Therefore we just need to prove the differentiability part.

Let $y \in (f(a), f(b))$ and let x be its unique pre-image. Given h such that $y+h$ is still inside the open interval $(f(a), f(b))$ take k to be a number such that $y+h = f(x+k)$ so $k = f^{-1}(y+h) - x$. Now we want to consider $\frac{(f^{-1}(y+h)-x)}{h}$ because if this approaches a limit as h goes to 0 that is the derivative we want. We get $\frac{k}{f(x+k)-f(x)}$ which as h goes to 0 (and therefore k goes to 0 by continuity) we approach the reciprocal of f'(x), as desired.

□

Example. This means that stuff like arcsin or log or roots are differentiable, without being like "obviously just look at it".

Proposition. (Cauchy mean value theorem) Let f,g be functions taking real values continuous on [a,b] and differentiable (a,b). Then there exists a c in (a,b) such that

$$g'(c)(f(b) - f(a)) = f'(c)(g(b) - g(a))$$

Proof. Try the function

$$h(x) = [g(x) - g(a)][f(b) - f(a)] - [g(b) - g(a)][f(x) - f(a)]$$

Now

$$h'(x) = g'(x)(f(b) - f(a)) - f'(x)(g(b) - g(a))$$

This is 0 at some point because $h(a) = h(b)$ and Rolle's theorem is true.

□

Now we will state L'hopital's rule

Suppose that the following hold:

1. $f(a) = g(a) = 0$
2. $g(x)$ is not 0 if x is close to c
3. f and g are differentiable in the vicinity of a and $\lim_{x \to c} \frac{f'(x)}{g'(x)}$ exists and is equal to l

Then $\frac{f(x)}{g(x)} \to l$ as $x \to a$.

In the lecture, the proof of this was left as an exercise. However, the proof was done in level 6.2.

[Lecture 12 ends]

Note that derivatives don't have to be continuous, for example the derivative of $x^2 \sin\left(\frac{1}{x}\right)$ exists at all non-zero points, is 0 at $x = 0$ and oscillates wildly near $x = 0$ so it is not continuous at 0.

3.2 Taylor's theorem

Definition. We say a function f is smooth if it is infinitely differentiable.

Recall generalized Rolle's theorem from level 6.2, which says

Let f be continuous on $[a, b]$ and n-1 times continuously differentiable, but also n times differentiable on the open interval (a,b), and suppose also that

$$f(a) = f'(a) = f''(a) = \cdots = f^{(n-1)}(a) = f(b) = 0$$

then there is some x in (a,b) such that $f^{(n)}(x) = 0$.

The proof was an easy induction using normal Rolle's theorem.

Theorem. If f on $[a, a+h]$ is continuous and also n-1 times continuously differentiable and on $(a, a+h)$ it is n times differentiable, then the error of the n'th order taylor polynomial about a at a+h by generalized Rolle's theorem $\frac{1}{n!}\, h^n\, f^{(n)}(a+ch)$ for some c between 0 and 1.

Proof. Again, since we proved this in Level 6.2, we will sketch the proof. The result follows easily from the Generalized Rolle's theorem, by subtracting a polynomial cleverly to make the conditions hold.

□

However we can also get the Cauchy form of remainder for Taylor's theorem:

Proof. We get (if we expand about $a+h$ instead of a) the remainder term is as follows (supposing a=0 as we can shift):

$$g(t) = f(h) - \sum_{k=0}^{n-1} \frac{f^{(k)}(t)}{k!}(h-t)^k$$

Note that $g(h) = 0$

$$g'(t) = -\sum_{k=0}^{n-1} \frac{f^{(k+1)}(t)}{k!}(h-t)^k + \sum_{k=0}^{n-1} \frac{f^{(k)}(t)}{(k-1)!}(h-t)^{k-1}$$

Note that most of these terms cancel so we get

$$-\frac{f^{(n)}(t)}{(n-1)!}(h-t)^{n-1}$$

By product and chain rules.

Now set $\phi_p(t) = g(t) - \left(\frac{h-t}{h}\right)^p g(0)$ so $\phi_p(0) = \phi_p(h) = 0$ and p is one of 1,2,3,. . . ,n.

Now by Rolle's theorem there is some C between 0 and 1 such that

$$0 = \phi_p'(Ch) = g'(Ch) + \frac{p(1-C)^{p-1}}{h} g(0).$$

But then this means that

$$g(0) = -\frac{hg'(Ch)}{p(1-C)^{p-1}} = \frac{h^n}{p(n-1)!}(1-C)^{n-p} f^{(n)}(Ch)$$

where I plugged in the definition of g and simplified.

Since $g(0)$ is the remainder of our taylor series at h, we call this the cauchy form of remainder. This gives us another version of Taylor's theorem alternative from the 2 we had in level 6 technical results.

Now for smooth functions, we know that the error is $O(h^n)$ but we do not know that it goes to 0 as n goes to infinity.

However, if $f^{(n)}$ does not grow faster than n! anywhere on $[a, a+h]$ then the error term $\frac{1}{n!}\, h^n\, f^{(n)}(a+ch)$ does go to 0. This is a more standard proof that taylor series work than the differential equation + differentiability of power series argument we gave in level 4 that is literally easier and more intuitive.

□

[Lecture 13 ends]

Example. Suppose we have x^q for q rational but not an integer, which is smooth on $(0, \infty)$.

Now suppose we taylor expand it about $x = 1$, then the remainder after n-1 terms of the polynomial will be of the form

$\begin{pmatrix} q \\ n \end{pmatrix} (1+tx)^{q-n} x^n$ for some t between 0 and 1 since that is the n'th derivative.

Here $\begin{pmatrix} q \\ n \end{pmatrix}$ simply means $\frac{q(q-1)...(q-n+1)}{n!}$

We deduce the generalized binomial theorem because $\left| \begin{pmatrix} q \\ n \end{pmatrix} (1+tx)^{q-n} x^n \right| \approx \left| \begin{pmatrix} q \\ n \end{pmatrix} x^n \right|$ which goes to 0 provided $|x| < 1$.

We can use the cauchy remainder: We have $(1-t)^{n-1} n \begin{pmatrix} q \\ n \end{pmatrix} (1+tx)^{q-n} x^n$

We can write this as $q \begin{pmatrix} q-1 \\ n-1 \end{pmatrix} \left(\frac{1-t}{1+tx}\right)^{n-1} (1+tx)^{q-1} h^n$, and we observe that $\left(\frac{1-t}{1+tx}\right)^{n-1} \to 0$ whenever x is between -1 and 1 and t is between 0 and 1, since the stuff in the brackets is less than 1. We have some constants and we are left with $\begin{pmatrix} q-1 \\ n-1 \end{pmatrix} h^n$ which is just $\begin{pmatrix} q-1 \\ n-1 \end{pmatrix}$ but we have a term going to 0 exponentially fast so we are fine.

By doing it this way we see that the error gets small exponentially and faster for smaller x.

As usual we say a function is analytic if it locally has a taylor series converging to the function.

4 Integration

Recall from level 4 that the Riemann integral is defined if the lower estimates and upper estimates for the area can get within ε of eachother for every ε, and that this always exists for continuous functions on closed bounded intervals.

However, this is not the standard definition, despite the fact that it is easy to show that this is equivalent to the standard definition, which is as follows.

Let f be any bounded function on $[a, b]$, then $\underline{\int_a^b} f(t)dt$ is the supremum over all partitions of $[a, b]$ of the integrals of sums of rectangles that underestimate the function at each partition. This is called the power integral. We similarly define the upper integral as $\overline{\int_a^b} f(t)dt$. f is said to be Riemann integrable if these integrals are equal.

Proposition. This is equivalent to the ε definition

Proof. Fix $\varepsilon > 0$. If f is integrable according to the definition introduced above, then there exist partitions D_1 and D_2 such that if we denote $U_{D_1}$and L_{D_2} as the upper and lower integrals with these partitions respectively, then

$$U_{D_1} < \int f(x)dx + \frac{\varepsilon}{2}$$

$$L_{D_2} < \int f(x)dx - \frac{\varepsilon}{2}$$

Let D be any partition that contains the end points of both D_1 and D_2, then since refining a partition makes you get closer to the actual integral (clearly),

$$U_D - L_D \leq U_{D_1} - L_{D_2} < \varepsilon$$

Conversely, if the above inequality holds, then this implies that $\inf U_D - \sup L_D < \varepsilon$ for every ε so they are equal, since clearly any upper estimate is larger than any lower estimate so the difference is actually always positive.

□

Example. If f is 0 at irrational numbers and 1 at rational numbers it is not riemann integrable because any interval will have some 0 and some 1 so the lower and upper estimates will differ by the size of the interval and will not get arbitrarily small. However this f is Lebesgue integrable.

[Lecture 14 ends]

If we refine a partition we will get a better estimate as if we do a lower estimate then refine the partition then clearly the lower estimate will be at least as good. A refinement formally means that it contains the previous partition.

Example. $f(x) = x$ on [0,1] is riemann integrable because if we take a partition like $\left(0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right)$ then

Lower sum is $\sum_{i=0}^{n-1} \frac{1}{n} * \frac{i}{n}$ and upper sum is $\sum_{i=1}^{n} \frac{1}{n} * \frac{i}{n}$. So we have $\frac{1}{2} \pm \frac{1}{2n}$ as our sums, and so we see that we have an integrable function with integral $\frac{1}{2}$.

The function on [0,1] that is 0 up to and including $\frac{1}{2}$ and then 1 after because for each ε we can take the partition $\left(0, \frac{1}{2} - \frac{\varepsilon}{2}, \frac{1}{2} + \frac{\varepsilon}{2}, 1\right)$ and the difference between the lower and upper estimates will be less than ε.

Proposition. A function is riemann integrable if it is continuous on a closed bounded interval

Proof. We proved this in level 4. However, the idea was we could suppose it is false and exhibit such an ε where it fails and use the Bolzano weierstrass theorem to get that it cannot be continuous at said point a contradiction.

□

Proposition. An increasing function on a closed bounded interval is riemann integrable

Proof. If we take the partition $\left(0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right)$ then we have to multiply the gap size by the differences but the sum of the differences is f(b)-f(a) so the total difference is bounded by $\frac{f(b)-f(a)}{n}$ which goes to 0 as n goes to infinity.

□

[Lecture 15 ends]

We know that continuous functions are integrable, and it immediately follows that piecewise continuous functions are integrable by first partitioning into the pieces. However, there is a technical detail.

Note that if f and g are bounded above by M integrable and disagree except on a finite set then the integrals are the same as they can be made as close to eachother as we want by making a partition that isolates the "bad points" with a small enough interval around them in the partition, specifically make the width at most $\frac{\varepsilon}{2Mk}$ where k is the number of bad points, so that the difference between the upper and lower sums in these intervals is at most $\frac{\varepsilon}{k}$ and there are at most k of them so the total difference is at most ε. We now know that piecewise continuous implies integrable (as it no longer matters what f is at the points of discontinuity).

We can formally prove stuff like that sums and constant multiples of integrable functions are integrable, here is how you might do that:

If f and g are integrable take a partition that makes the difference less than $\frac{\varepsilon}{2}$ and take the least common partition and the result follows.

If f is integrable and we want Nf to be integrable take a partition that makes the difference less than $\frac{\varepsilon}{N}$ then the result follows.

[Lecture 16 ends]

Integrable functions can get bad. As an example, consider the following function:

$$f : [0,1] \to \mathbb{R}$$

$$\begin{cases} 0 \text{ if } f \text{ is irrational} \\ \frac{1}{q} \text{ if } f \text{ is } \frac{p}{q} \text{ when fully simplified} \end{cases}$$

Figure 3 shows a graph of this function.

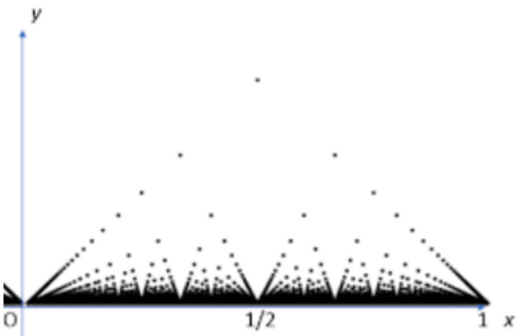

Figure 3

Note that the lower integral is always 0, so we want to show that we can make the upper integral as small as we want by making a clever partition.

Fix $\varepsilon > 0$. Now for all finitely many points that are greater than $\frac{\varepsilon}{2}$ in size, say there are M of them, cover them each by an interval of length $< \frac{\varepsilon}{2M}$. Now in these parts of the partition the upper sum will contribute at most $\frac{\varepsilon}{2}$, but this is also the case in the rest of the parts of the partition, so we have total less than ε.

The underlying reason is that the function is continuous at every point that is not a rational number and that bounded functions on closed bounded intervals are integrable if and only if their discontinuity set has "measure 0". In the analysis lemmas document we will see what this means and why it is true.

[Lecture 17 ends]

Lemma. For all functions f on any interval I,

$$\sup_I |f| - \inf_I |f| \leq \sup_I f - \inf_I f$$

Proof. If f is always positive or always negative this is immediate. Otherwise, if $\inf f < 0 < \sup f$ then we have that

$$\sup_I |f| - \inf_I |f| \leq \sup_I |f| \leq \max\left\{\sup_I (f), \sup_I (-f)\right\} \leq \sup_I (f) + \sup_I (-f) = \sup_I f - \inf_I f$$

□

Lemma.

$$\sup_I f^2 - \inf_I f^2 \leq 2\sup_I |f| \left(\sup_I(f) - \inf_I (f)\right)$$

Proof.

$$f(x)^2 - f(y)^2 = (f(x) + f(y))\ (f(x) - f(y)) \leq 2\sup|f| (\sup(f) - \inf(f))$$

So take supremums on both sides.

□

Lemma. $\left|\int f\right| \leq \int |f|$ if f is integrable.

Proof. Note that by 2 lemmas ago, it follows that if f is integrable, then so is $|f|$. It then follows that since f is between $-|f|$ and f, $-\int |f| \leq \int f$ and $\int -f \leq \int |f|$.

Note that this generalizes to complex integration – the idea is to write the integral as $Re^{i\theta}$, but we will do it in the complex analysis course.

□

Lemma. If f and g are integrable and bounded then so is fg

Proof. Write $fg = \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right]$. Then by 2 lemmas ago we have on each part of a partition $U\left(f^2, P\right) - L\left(f^2 - P\right) \leq 2\text{sup}\,|f|\,[U(f,P) - L(f,P)]$. Clearly, f+g and f-g are integrable, so just applying this to $f = f+g$ and $f = f-g$ we get the result.

□

Proposition. Let $f : [a,b] \to \mathbb{R}$ be riemann integrable and bounded, and set $F(x) = \int_a^x f(t)\,dt$, then F is continuous on [a,b].

Proof. Note that by 2 lemmas ago,

$$|F(x+h) - F(x)| \leq \int_x^{x+h} |f(t)|\,dt \leq h * \sup_{[a,b]} |f|$$

then this goes to 0 as h goes to 0.

□

Theorem. (Fundamental theorem of calculus) Let $f : [a,b] \to \mathbb{R}$ be continuous (and therefore riemann integrable and bounded), then F as above has derivative f. This is the fundamental theorem of calculus, which we convinced you of in levels 3 and 4 but now we will fully formalize it.

Proof. Set $\varepsilon(h) = \frac{F(x+h) - F(x) - hf(x)}{|h|}$, then we want to show that $\varepsilon(h) \to 0$ as $h \to 0$.

We have a bound for the numerator:

$$|F(x+h) - F(x) - hf(x)| = \left| \int_x^{x+h} f(t)dt - hf(x) \right| = \left| \int_x^{x+h} f(t) - f(x)dt \right|$$

$$\leq \int_x^{x+h} |f(t) - f(x)|\,dt \leq h * \sup_{t \in [0,h]} |f(x+t) - f(x)|$$

and by continuity of f it follows that $\varepsilon(h) \to 0$ as $h \to 0$ as required. (Symmetric argument if h is negative)

□

Remark. We need continuity otherwise we could have a jump function and integrate it to get something like an absolute value function which is not differentiable everywhere.

We also know that the integral of a derivative is the original function (up to a constant) when the derivative is continuous.

Theorem. (Fundamental theorem of calculus again) Let f on [a,b] be riemann integrable and suppose that it has an antiderivative F.

Then $F(b) - F(a) = \int_a^b f(t)\,dt$

Proof. For all $\varepsilon > 0$ there exists a partition P of [a,b] such that the upper and lower sums differ by at most ε. By the mean value theorem, for each j, there is a t_j between x_{j-1} and x_j such that $F(x_j) - F(x_{j-1}) = f(t_j)(x_j - x_{j-1})$. But the sum of everything on the left hand side is $F(b) - F(a)$ so it follows that $\sum_{j=1}^{n} f(t_j)(x_j - x_{j-1}) = F(b) - F(a)$. We know that this is somewhere between the lower sum and the upper sum since t_j is between the point where f achieves its supremum and infimum. Because this is true for all partitions, the result follows.

□

Note that it is not automatic that a derivative is riemann integrable, but constructing a non-integrable derivative is difficult and requires results from Level 8.2. The idea is to use a fat cantor set to make the discontinuity set of the derivative have non-zero measure.

[Lecture 18 ends]

As an example, we can now prove that if f and g are continuously differentiable then integration by parts works, ie $\int_a^b f'g = f(b)g(b) - f(a)g(a) - \int_a^b fg'$. We know how to prove this because we have derived integration by parts before.

Again, integration by substitution works: If f is continuous and g is continuously differentiable and g(a)=A and g(b)=B then $\int_a^b f(x)\,dx = \int_A^B f(g(t))\,g'(t)\,dt$. Again the proof is the same – we just use the fundamental theorem of calculus and the chain rule.

Theorem. (Taylor's theorem with integral form of remainder)

Let f be n+1 times differentiable on [a,b] with it's (n+1)'th derivative continuous. Then we have that $f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!}f^{(2)}(a) + \cdots + \frac{(b-a)^n}{n!}f^{(n)}(a) + \int_a^b \frac{(b-t)^n}{n!}f^{(n+1)}(t)\,dt$.

Proof. As in level 6.2, we can do this by induction/integration by parts.

We can alternatively write the remainder at h after the (n-1)'th derivative term as

$$\frac{1}{(n-1)!}\int_0^h (h-u)^{n-1} f^{(n)}(a+u)\,dt$$

I write it this way just because that is how it was written in the lecture.

□

Proposition. Let $f, g : [a,b] \to \mathbb{R}$ be continuous such that g is never 0. Then there exists $c \in [a,b]$ such that we have $\int_a^b f(x)\,g(x)\,dx = f(c)\int_a^b g(x)\,dx$

Proof. Apply the cauchy mean value theorem to $F(x) = \int_a^x f(t)\,g(t)dt,\ G(x) = \int_a^x g(t)\,dt$. We get

$$g(c)\int_a^b f(x)\,g(x)\,dx = [F(b) - F(a)]\,G'(c) = F'(c)\,[G(b) - G(a)] = f(c)\,g(c)\int_a^b g(x)\,dx$$

Cancelling g(c) from both sides since it is not 0 gives the result.

□

Now take the taylor remainder $\frac{1}{(n-1)!}\int_0^h (h-u)^{n-1} f^{(n)}(a+u)\,dt$

And assume that $f^{(n)}$ is continuous.

Rewrite the integral as

$$\frac{h^n}{(n-1)!}\int_0^1 (1-t)^{n-1} f^{(n)}(a+th)\,dt$$

By a simple substitution/scaling. Now apply the previous proposition:

The remainder is

$$\frac{h^n}{(n-1)!}f^{(n)}(a+\theta h)\int_0^1 (1-t)^{n-1}dt$$

For some θ between 0 and 1. But this is just

$$\frac{h^n}{n!}f^{(n)}(a+\theta h)$$

Which is another form of taylor's theorem.

But if we take g=1 and f the rest of the integral we get the remainder is

$$\frac{h^n}{(n-1)!}(1-\theta)^{n-1}f^{(n)}(a+\theta h)$$

For some θ between 0 and 1.

Of course we can do improper integrals by taking a limit or taking multiple limits. We know this from level 5. As an example,

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \int_{-\infty}^{-1} e^{-x^2}dx + \int_1^{\infty} e^{-x^2}dx + \int_{-1}^{1} e^{-x^2}dx$$

The first 2 terms are finite by comparing to $e^{-|x|}$ which we know has a convergent integral.

[Lecture 19 ends]

Note that just like in the example above, we can test for convergence by comparison or by showing that the ratio between one integral with a known convergent integral approaches something finite. For example, $\frac{x}{x^4+1}$ is $< \frac{1}{x^3}$ so its integral from 1 to infinity converges. In fact it is $\frac{\pi}{8}$ if you want to calculate it.

Example. By considering when the limit exists, $\int_0^1 \frac{1}{x^p}dx$ if and only if $p<1$.

Example. If we want to know when $\int_0^{\frac{1}{2}} \frac{1}{x\log^{\mathrm{p}}(x)}dx$, then the substitution $u=\log(x)$ and reducing to the previous example means it converges if and only if p>1.

[Lecture 20 ends]

5 Power series

5.1 Basic properties (Review)

We say a sequence of functions converges pointwise if for every x, $f_n(x) \to f(x)$ as a sequence of real or complex numbers. Series of functions are just the sums of functions.

For the limit of a sequence of continuous functions to be continuous, we essentially need

$$\lim_{y\to x}\lim_{n\to\infty} f_n(y) = \lim_{y\to x} f(y) = f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty}\lim_{y\to x} f(y)$$

As we know from differential equations (well, usually not taught in DEs but we did it because we're cool), if we have uniform convergence we can swap these limits, but swapping limits is not always trivial. It is false in general: For example the limit of

$$f(x) = \begin{cases} nx & x < \frac{1}{n} \\ 1 & \text{Otherwise} \end{cases}$$

on $[0,1]$ is 0 at 0 and 1 everywhere else, which is not continuous but it is a limit of continuous functions.

In fact, it is not even true that

$\lim_{m\to\infty}\lim_{n\to\infty}\frac{m}{m+n}=0$ but $\lim_{n\to\infty}\lim_{m\to\infty}\frac{m}{m+n}=1$

We will focus in this course on power series. Recall that the radius of convergence of a power series $c_0+c_1x+c_2x^2+\ldots$ with $c_i\in\mathbb{C}$ is given by $\frac{1}{\text{limsup}|c_n|^{\frac{1}{n}}}$ and that we have absolute convergence whenever we are strictly inside the radius of convergence.

As an example, if $c_n=n^a$ for any real a, then $|c_n|^{\frac{1}{n}}=n^{\frac{a}{n}}=e^{\frac{a}{n}\log(n)}$. This goes to 1 as n goes to infinity, since n grows faster than log n, or by L'hopital's rule. So the radius of convergence of any such power series is 1.

[Lecture 21 ends]

See levels 4 and 6 for power series properties such as differentiability (which implies continuity) and integrability and all that.

Sketch of proof that power series are differentiable and integrable (See levels 4 and 6.1 for a more complete proof):

First we rewrite $b^n-a^n-n\left(b-a\right)a^{n-1}$ as $\left(b-a\right)^2\left(b^{n-2}+2ab^{n-3}+...+\left(n-1\right)a^{n-2}\right)$ to make the algebra work in future steps, then we show that the derivative and integral would not affect the radius of convergence using the formula for the radius of convergence (meaning that the reverse derivative always exists so they are integrable), then we show that the "algebraic" derivative and the actual derivative go to 0 as the h in the definition of the derivative goes to 0.

[Lecture 22 ends]

5.2 Exponential and trigonometric functions

Let e^x be defined by its power series. Recall from level 4 how we can prove the basic properties $(e^x)'=e^x,\ e^0=1,\ e^{a+b}=e^ae^b$.

We will give an alternative easier proof of $e^{a+b}=e^ae^b$.

Let $f\left(z\right)=e^{a+b-z}e^z$, then $f'\left(z\right)=-e^{a+b-z}e^z+e^{a+b-z}e^z$ by the product and chain rules. Therefore f is constant. Therefore, $e^{a+b-b}e^b=e^{a+b-0}e^0$, so the result follows.

Lets formalize results we already know.

e^x is always positive for real numbers since it cannot cross 0 as if $e^x=0$ then e^{-x} is undefined. But by the intermediate value theorem, if e^x were ever negative, it would have to cross 0. And we know that e^x is strictly increasing – this is immediate from the series definition, and the derivative is always positive.

Note that e^x for real numbers x indeed hits every real number but 0 because $\lim_{x\to-\infty}e^x=\frac{1}{\lim_{x\to\infty}e^x}$, but we know that $e^x\geq 1+x$ so it gets arbitrarily large, so the limit tends to 0.

Now that we know that the exp is strictly increasing, we know it has an inverse, which is the logarithm. We knew this before but now we know know it.

By our inverse functions theorems we rederive that the derivative of log is $\frac{1}{x}$.

We want to use taylor series to show that $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots=\ln\left(2\right)$. However, this is on the edge of the radius of convergence, so we need another lemma. Once we have this, we will know immediately from the arctan series that $1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\cdots=\frac{\pi}{4}$.

Theorem. Suppose we have a power series with radius of convergence R. If $\sum c_nx^n$ converges and $|x|=R$, then the limit as a goes to R along the line arg(a)=arg(x) in the direction away from the origin of $\sum c_na^n$ is $\sum c_nx^n$

Why do we want this: Because we know that taylor series work strictly inside the radius of convergence, but series like $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\cdots=\ln\left(2\right)$ and $1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}-\cdots=\frac{\pi}{4}$ rely on knowing that the taylor series for $\ln(1+x)$ and arctan(x) respectively work at the boundary of their radius of convergence. If we can show that the power series that we get from using the log taylor series evaluated at 0.9, 0.99, 0.999, etc that we know are equal to $\ln(1.9), \ln(1.99), \ln(1.999)$ etc converge to $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots$, which clearly converges itself, then it will mean that $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots$ coincides with $\ln(2)$ (since ln is continuous there), so the taylor series will work at the edge of the radius of convergence whenever it actually converges there and the function we are expressing as a taylor series is continuous there, and we know already that it works inside the radius of convergence. So that is why we want this theorem.

Proof. We will suppose that $R = 1$ and that the point we are considering where the power series converges is the point $x = 1$. This is because we can scale and rotate it as necessary afterwards, but this will simplify calculations.

Therefore we are supposing that $\sum_{n=0}^{\infty} c_n$ converges to a value we will call s and that the power series $\sum c_n x^n$ has radius of convergence 1.

Note that $c_n = \sum_{r=0}^{n} c_r - \sum_{r=0}^{n-1} c_r$. This seems like a complicated way of doing things but it will work out.

Now let x be a real number strictly between 0 and 1.

Now $\sum_{n=0}^{N} c_n x^n = c_0 + \sum_{n=1}^{N} c_n x^n = c_0 + \sum_{n=1}^{N} (\sum_{r=0}^{n} c_r - \sum_{r=0}^{n-1} c_r) x^n$. Now we can do a little trick: We hav the following terms in our sum:

$\sum_{r=0}^{1} c_r x$	$-\sum_{r=0}^{0} c_r x$
$\sum_{r=0}^{2} c_r x^2$	$-\sum_{r=0}^{1} c_r x^2$
$\sum_{r=0}^{3} c_r x^3$	$-\sum_{r=0}^{2} c_r x^3$
...	...
$\sum_{r=0}^{N} c_r x^N$	$-\sum_{r=0}^{N-1} c_r x^{N+1}$

Now we will separate out the top-right and bottom-left terms, and simplify the top right term to $c_0 x$. Then the sum of the rest of the terms (add each term to the one directly to the bottom-right of it) is exactly $\sum_{n=1}^{N-1} \sum_{r=0}^{n} c_r \left(x^n - x^{n+1}\right)$. So

$$\sum_{n=0}^{N} c_n x^n = c_0 - c_0 x + \sum_{r=0}^{N} c_r x^N + \sum_{n=1}^{N-1} \sum_{r=0}^{n} c_r \left(x^n - x^{n+1}\right)$$

Therefore

$$\sum_{n=0}^{N} c_n x^n = c_0(1-x) + \sum_{r=0}^{N} c_r x^N + \sum_{n=1}^{N-1} \sum_{r=0}^{n} c_r (1-x) x^n$$

But then we can put the term $c_0(1-x)$ into the sum on the right, ie we get

$$\sum_{n=0}^{N} c_n x^n = \sum_{r=0}^{N} c_r x^N + (1-x) \sum_{n=0}^{N-1} \sum_{r=0}^{n} c_r x^n$$

. By the hypothesis of the theorem, the right hand side converges. Since $\sum_{r=0}^{N} c_r$ converges, it is bounded. Since x^N goes to 0 as N gets large, $\sum_{r=0}^{N} c_r x^N$ goes to 0 since $\sum_{r=0}^{N} c_r$ is bounded. Therefore when we take the limit as N goes to infinity that term vanishes and we get

$$\sum_{n=0}^{\infty} c_n x^n = (1-x) \sum_{n=0}^{\infty} \sum_{r=0}^{n} c_r x^n$$

The sum on the right converges since differs from a convergent sum by something that approaches 0. Now our goal is to show that as x goes to 1 from below, $\sum_{n=0}^{\infty} c_n x^n$ approaches s. Note that since x<1, we can safely say that $(1-x) \sum_{n=0}^{\infty} x^n = 1$ (geometric series or generalized binomial theorem).

Therefore we can subtract s from both sides of the equation we deduced above to get that

$$\sum_{n=0}^{\infty} c_n x^n - s = (1-x) \sum_{n=0}^{\infty} \left(\sum_{r=0}^{n} c_r - s \right) x^n$$

Therefore if we can show that the right hand side tends to 0, we will know that the left hand side tends to 0 so we will be done. Now lets pick a number ε as small as we like then pick an M such that for all $n \geq M$, $|\sum_{r=0}^{n} c_r - s| < \frac{\varepsilon}{2}$, possible by definition of summing to infinity.

Then we will write $\sum_{n=0}^{\infty}\left(\sum_{r=0}^{n} c_r - s\right)x^n$ as $\sum_{n=0}^{M-1}\left(\sum_{r=0}^{n} c_r - s\right)x^n + \sum_{n=M}^{\infty}\left(\sum_{r=0}^{n} c_r - s\right)x^n$

$$\sum_{n=0}^{\infty} c_n x^n - s = (1-x)\sum_{n=0}^{M-1}\left(\sum_{r=0}^{n} c_r - s\right)x^n + \sum_{n=M}^{\infty}\left(\sum_{r=0}^{n} c_r - s\right)x^n$$

So by the triangle inequality

$$\left|\sum_{n=0}^{\infty} c_n x^n - s\right| \leq |1-x|\sum_{n=0}^{M-1}\left|\sum_{r=0}^{n} c_r - s\right||x|^n + \sum_{n=M}^{\infty}\left|\sum_{r=0}^{n} c_r - s\right||x|^n$$

$$\left|\sum_{n=0}^{\infty} c_n x^n - s\right| \leq |1-x|\sum_{n=0}^{M-1}\left|\sum_{r=0}^{n} c_r - s\right||x|^n + \sum_{n=M}^{\infty}\frac{\varepsilon}{2}|x|^n$$

$$\left|\sum_{n=0}^{\infty} c_n x^n - s\right| \leq |1-x|\sum_{n=0}^{M-1}\left|\sum_{r=0}^{n} c_r - s\right||x|^n + |1-x|\,\frac{\varepsilon}{2}\frac{|x|^M}{1-|x|}$$

(last line by geometric series). Since $0 < x < 1$, $|x|^M < 1$, so the last term is ast most $\frac{\varepsilon}{2}$. Since M does not depend on x and x is a positive real number less than 1, the term $|1-x|\sum_{n=0}^{M-1}\left|\sum_{r=0}^{n} c_r - s\right||x|^n$ is at most $|1-x|\sum_{n=0}^{M-1}\left|\sum_{r=0}^{n} c_r - s\right|$, which is just $|1-x|\,C$ for some constant C. Therefore this tends to 0 as x gets close enough to 1, in particular it is eventually less than $\frac{\varepsilon}{2}$. Therefore, for x close enough to 1,

$$\left|\sum_{n=0}^{\infty} c_n x^n - s\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

This means the power series can be made arbitrarily close to s by making x sufficiently close to 1 from the left of 1 on the number line. This completes the proof of the theorem.

□

Now using our stuff about exponentials and logarithms we can define $x^a = e^{a\log(x)}$ and derive the usual properties of exponents.

Recall that by repeated l'hopital's rule we can formalize the "exponentials beat powers" and "powers beat logarithms" ideas.

[Lecture 23 ends]

Recall that $\sin(x) = \frac{e^{ix}-e^{-ix}}{2}$ and $\cos(x) = \frac{e^{ix}+e^{-ix}}{2}$, and that all of the trig identities we know follow from the properties of the exponential and taking real and imaginary parts.

We want to see another way of seeing that these are periodic, ie $\sin(x) = \sin(x+2\pi)$ and similarly for cos.

To do this, we will prove that there is some smallest w such that $\cos\left(\frac{w}{2}\right) = 0$.

We will prove that $\sin\left(\frac{w}{2}\right) = 1$, and that sin and cos are periodic with period 2w. We will prove also that adding w to sin or cos makes the result minus. We will prove also that $\sin\left(x+\frac{w}{2}\right) = \cos(x)$ and that $\cos\left(x+\frac{w}{2}\right) = -\sin(x)$.

Ok so here is how we do this:

We can calculate by numerical calculation and an easy comparison to series that $\cos(2) < 0$ since the first 4 terms of the taylor series give $-\frac{19}{45}$ and summing the absolute value of the remaining terms, ie $\sum_{k=4}^{\infty}\frac{2^{2k}}{(2k)!}$ is bounded above by $\sum_{k=4}^{\infty}\frac{2^{2k}}{8^k}$ which is $\frac{1}{8}$ (since $8^4 < (2*4)!$ And it is true for k>4 by induction) and we know trivially $\cos(0) = 1$. So by the intermediate value theorem, we know that there is some first place cos attains 0 between 0 and 2. We will call this $\frac{w}{2}$.

Now we get $\sin\left(\frac{w}{2}\right) = 1$ since sin was increasing on $\left(0, \frac{w}{2}\right)$ by the derivative. By the addition formula, get get $\sin\left(x+\frac{w}{2}\right) = \cos(x)$ and that $\cos\left(x+\frac{w}{2}\right) = -\sin(x)$. So periodicity follows from repeatedly applying these. We get periodicity of the exponential as well.

Proposition. 2w is the perimeter of the unit circle.

Proof. We want to consider the length of the curve e^{it} as t goes from 0 to 2w. By the arc length formula, we will have $\int_0^{2w} \left|\frac{d}{dt}e^{it}\right| dt = \int_0^{2w} 1dt = 2w$. This is the length because we are **defining** length this way.

□

[Lecture 24 ends]