

Part IA - Analysis I

Lectured by W. T. Gowers

Lent 2015

Limits and convergence

Sequences and series in \mathbb{R} and \mathbb{C} . Sums, products and quotients. Absolute convergence; absolute convergence implies convergence. The Bolzano-Weierstrass theorem and applications (the General Principle of Convergence). Comparison and ratio tests, alternating series test. [6]

Continuity

Continuity of real- and complex-valued functions defined on subsets of \mathbb{R} and \mathbb{C} . The intermediate value theorem. A continuous function on a closed bounded interval is bounded and attains its bounds. [3]

Differentiability

Differentiability of functions from \mathbb{R} to \mathbb{R} . Derivative of sums and products. The chain rule. Derivative of the inverse function. Rolle's theorem; the mean value theorem. One-dimensional version of the inverse function theorem. Taylor's theorem from \mathbb{R} to \mathbb{R} ; Lagrange's form of the remainder. Complex differentiation. [5]

Power series

Complex power series and radius of convergence. Exponential, trigonometric and hyperbolic functions, and relations between them. *Direct proof of the differentiability of a power series within its circle of convergence*. [4]

Integration

Definition and basic properties of the Riemann integral. A non-integrable function. Integrability of monotonic functions. Integrability of piecewise-continuous functions. The fundamental theorem of calculus. Differentiation of indefinite integrals. Integration by parts. The integral form of the remainder in Taylor's theorem. Improper integrals. [6]

Contents

1	The real number system	3
2	Convergence of sequences	5
2.1	Sums, products and quotients	5
2.2	Monotone-sequences property	8
2.3	Cauchy sequences	10
2.4	Limit supremum and infimum	11
3	Convergence of infinite sums	13
3.1	Absolute convergence	15
3.2	Convergence tests	17
3.3	Complex versions	19
4	Continuous functions	21
4.1	Continuous induction*	24
5	Differentiability	27
5.1	Limits	27
5.2	Differentiation	27
5.3	Differentiation theorems	31
5.4	Complex differentiation	35
6	Complex power series	36
6.1	Exponential and trigonometric functions	37
6.2	Differentiating power series	41
6.3	Hyperbolic trigonometric functions	43
7	The Riemann Integral	44
7.1	Improper integrals	59

1 The real number system

One can define real numbers as “decimals” consisting of infinitely many digits. However, while this is legitimate, it is not a convenient definition to work with. Instead, we define the real numbers to be “an ordered field with the least upper bound property”. Then we can show that the decimals form “an ordered field with the least upper bound property” if we really want to.

Definition (Field). A *field* is a set \mathbb{F} with two binary operations $+$ and \times that satisfies all the familiar properties satisfied by addition and multiplication in \mathbb{Q} , namely

- (i) Associativity: $\forall a, b, c \in \mathbb{F}, a + (b + c) = (a + b) + c$ and $a \times (b \times c) = (a \times b) \times c$
- (ii) Commutativity: $\forall a, b \in \mathbb{F}, a + b = b + a$ and $a \times b = b \times a$
- (iii) Identity: $\exists 0, 1 \in \mathbb{F}$ such that $\forall a, a + 0 = a$ and $a \times 1 = a$.
- (iv) Inverses: $\forall a \in \mathbb{F}, \exists (-a) \in \mathbb{F}$ such that $a + (-a) = 0$. If $a \neq 0$, then $\exists a^{-1}$ such that $a \times a^{-1} = 1$.
- (v) Distributivity: $\forall a, b, c \in \mathbb{F}, a \times (b + c) = (a \times b) + (a \times c)$

Example. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, integers mod p , $\{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$ are all fields.

Definition (Totally ordered set). An *(totally) ordered set* is a set X with a relation $<$ that satisfies

- (i) Transitivity: if $x, y, z \in X$, $x < y$ and $y < z$, then $x < z$
- (ii) Trichotomy: if $x, y \in X$, exactly one of $x < y, x = y, y < x$ holds

We call it a *totally* ordered set, as opposed to simply an ordered set, when we want to emphasize the order is total, ie. every pair of elements are related to each other, as opposed to a *partial order*, where “exactly one” in trichotomy is replaced with “at most one”.

Definition (Ordered field). An *ordered field* is a field \mathbb{F} with a relation $<$ that makes \mathbb{F} into an ordered set such that

- (i) if $x, y, z \in \mathbb{F}$ and $x < y$, then $x + z < y + z$
- (ii) if $x, y, z \in \mathbb{F}$, $x < y$ and $z > 0$, then $xz < yz$

Lemma. Let \mathbb{F} be an ordered field and $x \in \mathbb{F}$. Then $x^2 \geq 0$.

Proof. By trichotomy, either $x < 0$, $x = 0$ or $x > 0$. If $x = 0$, then $x^2 = 0$. So $x^2 \geq 0$. If $x > 0$, then $x^2 > 0 \times x = 0$. If $x < 0$, then $x - x < 0 - x$. So $0 < -x$. But then $x^2 = (-x)^2 > 0$. \square

Definition (Least upper bound). Let X be an ordered set and let $A \subseteq X$. An *upper bound* for A is an element $x \in X$ such that $\forall a \in A (a \leq x)$. If A has an upper bound, then we say that A is *bounded above*.

An upper bound x for A is a *least upper bound* or *supremum* if nothing smaller than x is an upper bound. That is, we need

- (i) $\forall a \in A (a \leq x)$
- (ii) $\forall y < x (\exists a \in A (a \geq y))$

We usually write $\sup A$ for the supremum of A when it exists. If $\sup A \in A$, then we call it $\max A$, the maximum of A .

Example. Let $X = \mathbb{Q}$. Then the supremum of $(0, 1)$ is 1. The set $\{x : x^2 < 2\}$ is bounded above by 2, but has no supremum (even though $\sqrt{2}$ seems like a supremum, we are in \mathbb{Q} and $\sqrt{2}$ is non-existent!).

$\max[0, 1] = 1$ but $(0, 1)$ has no maximum because the supremum is not in $(0, 1)$.

We can think of the supremum as a point we can get arbitrarily close to in the set but cannot pass through.

Definition (Least upper bound property). An ordered set X has the *least upper bound property* if every non-empty subset of X that is bounded above has a supremum.

Obvious modifications give rise to definitions of lower bound, greatest lower bound (or infimum) etc. It is simple to check that an ordered field with the least upper bound property has the greatest lower bound property.

Definition (Real numbers). The *real numbers* is an ordered field with the least upper bound property.

Of course, it is *very* important to show that this field is unique and there is at least one field that satisfies this property! But we are not doing that in this course.

Lemma (Archimedean property v1)). Let \mathbb{F} be an ordered field with the least upper bound property. Then the set $\{1, 2, 3, \dots\}$ is not bounded above. (Note that these need not refer to natural numbers. We simply define 1 to be the multiplicative identity, $2 = 1 + 1$, $3 = 1 + 2$ etc.)

Proof. If it is bounded above, then it has a supremum x . But then $x - 1$ is not an upper bound. So we can find $n \in \{1, 2, 3, \dots\}$ such that $n > x - 1$. But then $n + 1 > x$, but x is supposed to be an upper bound. \square

While the Archimedean property seems to be trivially true for all ordered fields even if they are not bounded above, actually there are ordered fields in which the integers are not bounded above.

For example, consider the field of rational functions, ie. functions in the form $\frac{P(x)}{Q(x)}$ with $P(x), Q(x)$ being polynomials. We order two functions $\frac{P(x)}{Q(x)}, \frac{R(x)}{S(x)}$ as follows: these two functions intersect only finitely many times because $P(x)S(x) = R(x)Q(x)$ has only finitely many roots. After the last intersection, the function whose value is greater counts as the greater function. It can be checked that these form an ordered field.

In this field, the integers are the constant functions $1, 2, 3, \dots$, but it is bounded above since the function x is greater than all of them.

2 Convergence of sequences

Definition (Sequence). A *sequence* is, formally, a function $a : \mathbb{N} \rightarrow \mathbb{R}$ (or \mathbb{C}). Usually (ie. always), we write a_n instead of $a(n)$. Instead of a , we usually write it as (a_n) , $(a_n)_1^\infty$ or $(a_n)_{n=1}^\infty$ to indicate it is a sequence.

Definition (Convergence of sequence). Let (a_n) be a sequence and $\ell \in \mathbb{R}$. Then a_n *converges to ℓ* , *tends to ℓ* , or $a_n \rightarrow \ell$, if

$$\forall \varepsilon > 0 \exists N \forall n \geq N : |a_n - \ell| < \varepsilon.$$

One can think of $\exists N : \forall n \geq N$ as saying “eventually always”, or as “from some point on”. So the definition means, if $a_n \rightarrow \ell$, then given any ε , there eventually, everything in the sequence is within ε of ℓ .

Lemma (Archimedean property v2). $1/n \rightarrow 0$.

Proof. Let $\varepsilon > 0$. We want to find an N such that $|1/N - 0| = 1/N < \varepsilon$. So **pick N** such that $N > 1/\varepsilon$. This exists such an N by the Archimedean property v1. Then **for all $n > N$** , we have $0 < 1/n \leq 1/N < \varepsilon$. So **$|1/n - 0| \rightarrow \varepsilon$** . \square

Note that the red parts correspond to the *definition* of a sequence. This is generally how we prove convergence from first principles.

Definition (Bounded sequence). A sequence (a_n) is *bounded*

$$\exists C \forall n : |a_n| \leq C.$$

A sequence is *eventually bounded* if

$$\exists C \exists N \forall n \geq N : |a_n| \leq C.$$

The definition of an *eventually bounded* sequence seems a bit daft. Clearly every eventually bounded sequence is bounded! Indeed it is:

Lemma. Every eventually bounded sequence is bounded.

Proof. Let C and N be such that $\forall n \geq N |a_n| \leq C$. Then $\forall n \in \mathbb{N}$, $|a_n| \leq \max\{|a_1|, \dots, |a_{N-1}|, C\}$. \square

The proof is rather trivial. However, most of the time it is simpler to prove that a sequence is eventually bounded, and this lemma saves us from writing that long line every time.

2.1 Sums, products and quotients

Here we prove the things that we think are obviously true, eg. sums and products of convergent sequences are convergent.

Lemma (Sums of sequences). If $a_n \rightarrow a$ and $b_n \rightarrow b$, then $a_n + b_n \rightarrow a + b$.

Proof. Let $\varepsilon > 0$. We want to find a clever N such that for all $n \geq N$, $|a_n + b_n - (a + b)| < \varepsilon$. Intuitively, we know that a_n is very close to a and b_n is very close to b . So their sum must be very close to $a + b$.

Formally, since $a_n \rightarrow a$ and $b_n \rightarrow b$, we can find N_1, N_2 such that $\forall n \geq N_1$, $|a_n - a| < \varepsilon/2$ and $\forall n \geq N_2$, $|b_n - b| < \varepsilon/2$.

Now let $N = \max\{N_1, N_2\}$. Then by the triangle inequality, when $n \geq N$,

$$|(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b| < \varepsilon.$$

□

Lemma (Scalar multiplication of sequences). Let $a_n \rightarrow a$ and $\lambda \in \mathbb{R}$. Then $\lambda a_n \rightarrow \lambda a$.

Proof. If $\lambda = 0$, then the result is trivial.

Otherwise, let $\varepsilon > 0$. Then $\exists N$ such that $\forall n \geq N$, $|a_n - a| < \varepsilon/|\lambda|$. So $|\lambda a_n - \lambda a| < \varepsilon$. □

Lemma. Let (a_n) be bounded $b_n \rightarrow 0$. Then $a_n b_n \rightarrow 0$.

Proof. Let $C \neq 0$ be such that $\forall n : |a_n| \leq C$. Let $\varepsilon > 0$. Then $\exists N$ such that $\forall n \geq N$, $|b_n| < \varepsilon/C$. Then $|a_n b_n| < \varepsilon$. □

Lemma. Every convergent sequence is bounded.

Proof. Let $a_n \rightarrow l$. Then there is an N such that $\forall n \geq N$, $|a_n - l| \leq 1$. So $|a_n| \leq |l| + 1$. So a_n is eventually bounded, and therefore bounded. □

Lemma (Product of sequences). Let $a_n \rightarrow a$ and $b_n \rightarrow b$. Then $a_n b_n \rightarrow ab$.

Proof. Let $c_n = a_n - a$ and $d_n = b_n - b$. Then $a_n b_n = (a + c_n)(b + d_n) = ab + ad_n + bc_n + c_n d_n$.

But by “sum of sequences”, $c_n \rightarrow 0$ and $d_n \rightarrow 0$. So $ad_n \rightarrow 0$ and $bc_n \rightarrow 0$. Since c_n is bounded, $c_n d_n \rightarrow 0$. Hence by sum of sequences, $a_n b_n \rightarrow ab$. □

The proof can be discovered as follows: We know that a_n and b_n get close to a and b . So we can write a_n as a plus some error term, and similarly for b . We realize that the error term of the product is $ad_n + bc_n + c_n d_n$. Then we prove lemmas to show that each of the error terms tend to 0.

On second thought, we only need to consider the error terms of a_n , and let $c_n = a_n - a$. Then $(a + c_n)b_n \rightarrow ab$ and thus $(a_n b_n) \rightarrow ab$.

Proof. (alternative) Observe that $a_n b_n - ab = (a_n - a)b_n + (b_n - b)a$. We know that $a_n - a \rightarrow 0$ and $b_n - b \rightarrow 0$. Since (b_n) is bounded, so $(a_n - a)b_n + (b_n - b)a \rightarrow 0$. So $a_n b_n \rightarrow ab$. □

Note that we no longer write “Let $\varepsilon > 0$ ”. In the beginning, we have no lemmas proven. So we must prove everything from first principles and use the definition. However, after we have proven the lemmas, we can simply use them instead of using first principles. This is similar to in calculus, where we use first principles to prove the product rule and chain rule, then no longer use first principles afterwards.

Lemma (Quotient of sequences). Let (a_n) be a sequence such that $\forall n \neq 0$. Suppose that $a_n \rightarrow a$ and $a \neq 0$. Then $1/a_n \rightarrow 1/a$.

Proof. We have

$$\frac{1}{a_n} - \frac{1}{a} = \frac{a - a_n}{aa_n}.$$

We want to show that this $\rightarrow 0$. Since $a - a_n \rightarrow 0$, we have to show that $1/(aa_n)$ is bounded.

Since $a_n \rightarrow a$, $\exists N$ such that $\forall n \geq N$, $|a_n - a| \leq a/2$. Then $\forall n \geq N$, $|a_n| \geq |a|/2$. Then $|1/(a_n a)| \leq 2/|a|^2$. So $1/(a_n a)$ is bounded. So $(a - a_n)/(aa_n) \rightarrow 0$ and the result follows. \square

Corollary. If $a_n \rightarrow a$, $b_n \rightarrow b$, $b_n, b \neq 0$. Then $a_n/b_n = a/b$.

Proof. We know that $1/b_n \rightarrow 1/b$. So the result follows by the product rule. \square

Lemma (Sandwich rule). Let (a_n) and (b_n) be sequences that both converge to a limit x . Suppose that $a_n \leq c_n \leq b_n$ for every n . Then $c_n \rightarrow x$.

Proof. Let $\varepsilon > 0$. We can find N such that $\forall n \geq N$, $|a_n - x| < \varepsilon$ and $|b_n - x| < \varepsilon$.

The $\forall n \geq N$, we have $x - \varepsilon < a_n \leq c_n \leq b_n < x + \varepsilon$. So $|c_n - x| < \varepsilon$. \square

Example. $1/2^n \rightarrow 0$. For every n , $n < 2^n$. So $0 < 1/2^n < 1/n$. The result follows from the sandwich rule.

Example. We want to show that

$$\frac{n^2 + 3}{(n + 5)(2n - 1)} \rightarrow \frac{1}{2}.$$

Write

$$\frac{n^2 + 3}{(n + 5)(2n - 1)} = \frac{1 + 3/n^2}{(1 + 5/n)(2 - 1/n)} \rightarrow \frac{1}{2},$$

by sum rule, sandwich rule, Archimedean property, product rule and quotient rule.

Example. Let $k \in \mathbb{N}$ and let $\delta > 0$. Then

$$\frac{n^k}{(1 + \delta)^n} \rightarrow 0.$$

This can be summarized as “exponential growth beats polynomial growth eventually”.

By the binomial theorem,

$$(1 + \delta)^n \geq \binom{n}{k+1} \delta^{k+1}.$$

Then if $n \geq 2k$,

$$\binom{n}{k+1} = \frac{n(n-1) \cdots (n-k)}{(k+1)!} \geq \frac{(n/2)^{k+1}}{(k+1)!}.$$

So for sufficiently large n ,

$$\frac{n^k}{(1 + \delta)^n} \leq \frac{n^k 2^{k+1} (k+1)!}{n^{k+1} \delta^{k+1}} = \frac{2^{k+1} (k+1)!}{\delta^{k+1}} \cdot \frac{1}{n} \rightarrow 0.$$

2.2 Monotone-sequences property

Definition (Monotone sequence). A sequence (a_n) is *increasing* if $a_n \leq a_{n+1}$ for all n .

It is *strictly increasing* if $a_n < a_{n+1}$ for all n . (*Strictly*) *decreasing* sequences are defined analogously.

A sequence is (*strictly*) *monotone* if it is (strictly) increasing or (strictly) decreasing.

Definition (Monotone sequences property). An ordered field has the *monotone sequences property* if every increasing sequence that is bounded above converges.

We want to show that the monotone sequences property is equivalent to the least upper bound property.

Lemma. Least upper bound property \Rightarrow monotone-sequences property.

Proof. Let (a_n) be an increasing sequence and let C an upper bound for (a_n) . Then C is an upper bound for the set $\{a_n : n \in \mathbb{N}\}$. By the least upper bound property, it has a supremum s . We want to show that this is the limit of (a_n) .

Let $\varepsilon > 0$. Since $s = \sup\{a_n : n \in \mathbb{N}\}$, there exists an N such that $a_N > s - \varepsilon$. Then since (a_n) is increasing, $\forall n \geq N$, we have $s - \varepsilon < a_N \leq a_n \leq s$. So $|a_n - s| < \varepsilon$. \square

Before showing the other way implication, we first prove the following lemma:

Lemma. Monotone-sequences property \Rightarrow Archimedean property.

Proof. We prove version 2, ie. that $1/n \rightarrow 0$.

Since $1/n > 0$ and is decreasing, by MSP, it converges. Let δ be the limit. We must have $\delta \geq 0$, since if $\delta < 0$, then there would exist n with $3\delta/2 < 1/n < \delta/2 < 0$. Contradiction.

If $\delta > 0$, then we can find N such that $1/N < 2\delta$. But then for all $n \geq 4N$, we have $1/n \leq 1/(4N) < \delta/2$. Contradiction. Therefore $\delta = 0$. \square

Lemma. Monotone sequences property \Rightarrow least upper bound property.

Proof. Let A be a non-empty set that's bounded above. Pick u_0, v_0 such that u_0 is not an upper bound for A and v_0 is an upper bound. Now do a repeated bisection: having chosen u_n and v_n such that u_n is not an upper bound and v_n is, if $(u_n + v_n)/2$ is an upper bound, then let $u_{n+1} = u_n$, $v_{n+1} = (u_n + v_n)/2$. Otherwise, let $u_{n+1} = (u_n + v_n)/2$, $v_{n+1} = v_n$.

Then $u_0 \leq u_1 \leq u_2 \leq \dots$ and $v_0 \geq v_1 \geq v_2 \geq \dots$. Then

$$v_n - u_n = \frac{v_0 - u_0}{2^n} \rightarrow 0.$$

Note that here we used the Archimedean property since to prove $1/2^n \rightarrow 0$, we sandwich it with $1/n$. But to show $1/n \rightarrow 0$, we need the Archimedean property.

By the monotone sequences property, $u_n \rightarrow s$ (since (u_n) is bounded above by v_0). Since $v_n - u_n \rightarrow 0$, $v_n \rightarrow s$. We now show that $s = \sup A$.

If s is not an upper bound, then there exists $a \in A$ such that $a > s$. Since $v_n \rightarrow s$, then there exists m such that $v_m < a$, contradicting the fact that v_m is an upper bound.

To show it is the *least* upper bound, let $t < s$. Then since $u_n \rightarrow s$, we can find m such that $u_m > t$. So t is not an upper bound. Therefore s is the least upper bound. \square

Lemma. Let (a_n) be a sequence and suppose that $a_n \rightarrow a$. If $\forall n, a_n \leq x$, then $a \leq x$.

Proof. If $a > x$, then set $\epsilon = a - x$. Then we can find N such that $a_N > x$. Contradiction. \square

Note: this can be used to streamline some of our earlier arguments, eg. the monotone sequences property implies $1/n \rightarrow 0$.

Lemma. A sequence can have at most 1 limit.

Proof. Let (a_n) be a sequence, and suppose $a_n \rightarrow x$ and $a_n \rightarrow y$. Let $\epsilon > 0$ and pick N such that $\forall n \geq N, |a_n - x| < \epsilon/2$ and $|a_n - y| < \epsilon/2$. Then $|x - y| \leq |x - a_N| + |a_N - y| < \epsilon/2 + \epsilon/2 = \epsilon$. Since ϵ was arbitrary, x must equal y . \square

Definition (Subsequence). Let (a_n) be a sequence. A *subsequence* of (a_n) is a sequence of the form a_{n_1}, a_{n_2}, \dots , where $n_1 < n_2 < \dots$.

Example. $1, 1/4, 1/9, 1/16, \dots$ is a subsequence of $1, 1/2, 1/3, \dots$.

Lemma (Nested intervals property). Let \mathbb{F} be an ordered field with the monotone sequences property. Let $I_1 \supseteq I_2 \supseteq \dots$ be closed bounded non-empty intervals. Then $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$.

Proof. Let $T_n = [a_n, b_n]$ for each n . Then $a_1 \leq a_2 \leq \dots$ and $b_1 \geq b_2 \geq \dots$. For each $n, a_n \leq b_n \leq b_1$. So the sequence a_n is bounded above. So by the monotone sequences property, it has a limit a . For each $n, a_n \leq a$, since if ever we had $a_n > a$, then $\forall m \geq n, a_m \geq a_n \Rightarrow a > a$, which is a contradiction.

Also, for each fixed n , we have that $\forall m \geq n, a_m \leq b_m \leq b_n$. So $a \leq b_n$. Thus, for all $n, a_n \leq a \leq b_n \Rightarrow a \in I_n$. So $a \in \bigcap_{n=1}^{\infty} I_n$. \square

We can use this to prove that the reals are uncountable:

Proposition. \mathbb{R} is uncountable.

Proof. Suppose the contrary. Let x_1, x_2, \dots be a list of all real numbers. Find an interval that does not contain x_1 . Within that interval, find an interval that does not contain x_2 . Continue *ad infinitum*. Then the intersection of all these intervals is non-empty, but the elements in the intersection are not in the list. Contradiction. \square

Theorem (Bolzano-Weierstrass theorem). Let \mathbb{F} be an ordered field with the monotone sequences property (ie. $\mathbb{F} = \mathbb{R}$).

Then every bounded sequence has a convergent subsequence.

Proof. Let u_0 and v_0 be a lower and upper bound, respectively, for a sequence $(a_n)_{n=1}^{\infty}$. By repeated bisection, we can find a sequence of intervals $[u_0, v_0] \supseteq [u_1, v_1] \supseteq [u_2, v_2] \supseteq \dots$ such that $v_n - u_n = (v_0 - u_0)/2^n$, and such that each $[u_n, v_n]$ contains infinitely many terms of (a_n) .

By the nested intervals property, $\bigcap_{n=1}^{\infty} [u_n, v_n] = \emptyset$. Let x belong to the intersection. Now pick a subsequence a_{n_1}, a_{n_2}, \dots such that $a_{n_k} \in [u_k, v_k]$. We can do this because $[u_k, v_k]$ contains infinitely many a_n , and we have only picked finitely many of them. We will show that $a_{n_k} \rightarrow x$.

Let $\varepsilon > 0$. By the Archimedean property, we can find K such that $v_K - u_K = (v_0 - u_0)/2^K \leq \varepsilon$. This implies that $[u_K, v_K] \subseteq (x - \varepsilon, x + \varepsilon)$, since $x \in [u_K, v_K]$. Then $\forall k \geq K$, $a_{n_k} \in [u_k, v_k] \subseteq [u_K, v_K] \subseteq (x - \varepsilon, x + \varepsilon)$. So $|a_{n_k} - x| < \varepsilon$. \square

2.3 Cauchy sequences

Definition (Cauchy sequence). A sequence (a_n) is *Cauchy* if

$$\forall \varepsilon > 0 \exists N \forall p, q \geq N : |a_p - a_q| < \varepsilon.$$

Roughly, a sequence is Cauchy if all terms are eventually close to each other (as opposed to close to a limit).

Lemma. Every convergent sequence is Cauchy.

Proof. Let $a_n \rightarrow a$. Let $\varepsilon > 0$. Then $\exists N$ such that $\forall n \geq N$, $|a_n - a| < \varepsilon/2$. Then $\forall p, q \geq N$, $|a_p - a_q| \leq |a_p - a| + |a - a_q| < \varepsilon/2 + \varepsilon/2 = \varepsilon$. \square

Lemma. Let (a_n) be a Cauchy sequence with a subsequence (a_{n_k}) that converges to a . Then $a_n \rightarrow a$.

Proof. Let $\varepsilon > 0$. Pick N such that $\forall p, q \geq N$, $|a_p - a_q| < \varepsilon/2$. Then pick K such that $n_K \geq N$ and $|a_{n_K} - a| < \varepsilon/2$.

Then $\forall n \geq N$, we have

$$|a_n - a| \leq |a_n - a_{n_K}| + |a_{n_K} - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

\square

Theorem (The general principle of convergence). Let \mathbb{F} be an ordered field with the monotone-sequence property. Then every Cauchy sequence of \mathbb{F} converges.

Proof. Let (a_n) be a Cauchy sequence. Then it is eventually bounded, since $\exists N$, $\forall n \geq N$, $|a_n - a_N| \leq 1$ by the Cauchy condition. So it is bounded. Hence by Bolzano-Weierstrass, it has a convergent subsequence. Then (a_n) converges to the same limit. \square

Note that the definition of a convergent sequence is

$$\exists l \forall \varepsilon > 0 \exists N \forall n \geq N |a_n - l| < \varepsilon,$$

while that of Cauchy convergence is

$$\forall \varepsilon > 0 \exists N \forall p, q \geq N |a_p - a_q| < \varepsilon.$$

In the first definition, l quantifies over all real numbers, which is uncountable. However, in the second definition, we only have to quantify over natural numbers, which is countable (by the Archimedean property, we only have to consider the cases $\varepsilon = 1/n$).

Since they are equivalent, the second definition is preferred from a logic-simplicity point of view.

Lemma. Let \mathbb{F} be an ordered field with the Archimedean property such that every Cauchy sequence converges. The \mathbb{F} satisfies the monotone sequences property.

Proof. Instead of showing that every bounded sequence converges, and is hence Cauchy, We will show the equivalent statement that every increasing non-Cauchy sequence is not bounded above.

Let (a_n) be an increasing sequence. If (a_n) is not Cauchy, then

$$\exists \varepsilon > 0 \forall N \exists p, q > N : |a_p - a_q| \geq \varepsilon.$$

wlog let $p > q$. Then

$$a_p \geq a_q + \varepsilon \geq a_N + \varepsilon.$$

So for any N , we can find a $p > N$ such that

$$a_p \geq a_N + \varepsilon.$$

Then we can construct a subsequence a_{n_1}, a_{n_2}, \dots such that

$$a_{n_{k+1}} \geq a_{n_k} + \varepsilon.$$

Therefore

$$a_{n_k} \geq a_{n_1} + (k-1)\varepsilon.$$

So by the Archimedean property, (a_{n_k}) , and hence (a_n) , is unbounded. \square

Definition (Complete ordered field). An ordered field (with an appropriate metric) in which every Cauchy sequence converges is called *complete*.

Note: We can have complete ordered field that are not the reals, if they do not have the Archimedean property. For example, we can take the rational functions as before, then take the Cauchy completion of it (ie. add all the limits we need). Then it is already too large to be the reals (it doesn't have the Archimedean property) but is a complete ordered field.

2.4 Limit supremum and infimum

Definition (Limit supremum/infimum). Let (a_n) be a bounded sequence. We define the *limit supremum* as

$$\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \left(\sup_{m \geq n} a_m \right).$$

To see that this exists, set $b_n = \sup_{m \geq n} a_m$. Then (b_n) is decreasing since we are taking the supremum of fewer and fewer things, and is bounded below by any lower bound for (a_n) since $b_n \geq a_n$. So it converges.

Similarly, we define the *limit infimum* as

$$\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \left(\inf_{m \geq n} a_m \right).$$

Example. Take the sequence

$$2, -1, 1\frac{1}{2}, -\frac{1}{2}, 1\frac{1}{3}, -\frac{1}{3}, \dots$$

Then the limit supremum is 1 and the limit infimum is 0.

Lemma. Let (a_n) be a sequence. The following two statements are equivalent:

- $a_n \rightarrow a$
- $\limsup a_n = \liminf a_n = a$.

Proof. If $a_n \rightarrow a$, then let $\varepsilon > 0$. Then

$$\exists n \forall m \geq n : a - \varepsilon \leq a_m \leq a + \varepsilon.$$

It follows that

$$a - \varepsilon \leq \inf_{m \geq n} a_m \leq \sup_{m \geq n} a_m \leq a + \varepsilon.$$

Since ε was arbitrary, it follows that

$$\liminf a_n = \limsup a_n = a.$$

Conversely, if $\liminf a_n = \limsup a_n = a$, then let $\varepsilon > 0$. Then we can find n such that

$$\inf_{m \geq n} a_m > a - \varepsilon \text{ and } \sup_{m \geq n} a_m < a + \varepsilon.$$

It follows that $\forall m \geq n$, we have $|a_m - a| < \varepsilon$. □

3 Convergence of infinite sums

Definition (Convergence of infinite sums and partial sums). Let (a_n) be a real sequence. For each N , define

$$S_N = \sum_{n=1}^N a_n.$$

If the sequence (S_N) converges to some limit s , then we say that

$$\sum_{n=1}^{\infty} a_n = s,$$

and we say that the series $\sum_{n=1}^{\infty} a_n$ *converges*.

We call S_N the N th *partial sum*.

Lemma. If $\sum_{n=1}^{\infty} a_n$ converges. Then $a_n \rightarrow 0$.

Proof. Let $\sum_{n=1}^{\infty} a_n = s$. Then $S_N \rightarrow s$ and $S_{N-1} \rightarrow s$. Then $a_N = S_N - S_{N-1} \rightarrow 0$. □

However, **THE CONVERSE IS FALSE!!!**

Example (Harmonic series). If $a_n = 1/n$, then $a_n \rightarrow 0$ but $\sum a_n = \infty$.

We can prove this as follows:

$$S_{2^n} - S_{2^{n-1}} = \frac{1}{2^{n-1} + 1} + \cdots + \frac{1}{2^n} \geq \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

Therefore $S_{2^n} \geq S_1 + n/2$. So the partial sums are unbounded.

Example (Geometric series). Let $|\rho| < 1$. Then

$$\sum_{n=0}^{\infty} \rho^n = \frac{1}{1 - \rho}.$$

We can prove this by considering the partial sums:

$$\sum_{n=0}^{\infty} \rho^n = \frac{1 - \rho^{N+1}}{1 - \rho}.$$

But $\rho^{N+1} \rightarrow 0$. So this tends to $1/(1 - \rho)$.

Example. $\sum_{n=2}^{\infty} \frac{1}{n(n-1)}$ converges. This is since

$$\frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}.$$

So

$$\sum_{n=2}^N \frac{1}{n(n-1)} = 1 - \frac{1}{N} \rightarrow 1.$$

Lemma. Suppose that $a_n \geq 0$ for every n and the partial sums S_n are bounded above. Then $\sum_{n=1}^{\infty} a_n$ converges.

Proof. The sequence (S_n) is increasing and bounded above. So the result follows from the monotone sequences property. \square

Lemma (Comparison test). Let (a_n) and (b_n) be non-negative sequences, and suppose that $\exists C, N$ such that $\forall n \geq N$, $a_n \leq Cb_n$. Then if $\sum b_n$ converges, then so does $\sum a_n$.

Proof. Let $M > N$. Also for each R , let $S_R = \sum_{n=1}^R a_n$ and $T_R = \sum_{n=1}^R b_n$. We want S_R to be bounded above.

$$S_M - S_N = \sum_{n=N+1}^M a_n \leq C \sum_{n=N+1}^M b_n \leq C \sum_{n=N+1}^{\infty} b_n.$$

So $\forall M \geq N$, $S_M \leq S_N + C \sum_{n=N+1}^{\infty} b_n$. Since the S_M are increasing and bounded, it must converge.

Note: N is fixed from the very beginning in the statement of the lemma. \square

Example.

(i) $\sum \frac{1}{n2^n}$ converges, since $\sum \frac{1}{2^n}$ converges.

(ii) $\sum \frac{n}{2^n}$ converges.

If $n \geq 4$, then $n \leq 2^{n/2}$. That's because $4 = 2^{4/2}$ and for $n \geq 4$, $(n+1)/n < \sqrt{2}$, so when we increase n , we multiply the right side by a greater number by the left. Hence by the comparison test, it is sufficient to show that $\sum 2^{n/2}/2^n = \sum 2^{-n/2}$ converges, which it does (geometric series).

(iii) $\sum 1/\sqrt{n}$ diverges, since $1/\sqrt{n} \geq 1/n$. So if it converged, then so would $\sum 1/n$, but $\sum 1/n$ diverges.

(iv) $\sum 1/n^2$ converges, since for $n \geq 2$, $1/n^2 \leq 1/[n(n-1)]$, and we have proven that the latter converges.

(v) Consider $\sum_{n=1}^{\infty} \frac{n+5}{n^3 - 7n^2/2}$. We show this converges by noting

$$n^3 - \frac{7n^2}{2} = n^2 \left(n - \frac{7}{2} \right).$$

So if $n \geq 8$, then

$$n^3 - \frac{7n^2}{2} \geq \frac{n^3}{2}.$$

Also, $n+5 \leq 2n$. So

$$\frac{n+5}{n^3 - 7n^2/2} \leq 4/n^2.$$

So it converges by the comparison test.

(vi) If $\alpha > 1$, then $\sum 1/n^\alpha$ converges.

Let $S_N = \sum_{n=1}^N 1/n^\alpha$. Then

$$\begin{aligned} S_{2^n} - S_{2^{n-1}} &= \frac{1}{(2^{n-1} + 1)^\alpha} + \cdots + \frac{1}{(2^n)^\alpha} \\ &\leq \frac{2^{n-1}}{(2^{n-1})^\alpha} \\ &= (2^{n-1})^{1-\alpha} \\ &= (2^{1-\alpha})^{n-1}. \end{aligned}$$

But $2^{1-\alpha} < 1$. So

$$S_{2^n} = (S_{2^n} - S_{2^{n-1}}) + (S_{2^{n-1}} - S_{2^{n-2}}) + \cdots (S_2 - S_1) + S_1$$

and is bounded above by comparison with the geometric series $1 + 2^{1-\alpha} + (2^{1-\alpha})^2 + \cdots$

3.1 Absolute convergence

Definition (Absolute convergence). A series $\sum a_n$ *converges absolutely* if the series $\sum |a_n|$ converges.

Example. The series $\sum_{n+1} \frac{(-1)^{n=1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ converges, but not absolutely.

To see the convergence, note that

$$a_{2n-1} + a_{2n} = \frac{1}{2n-1} - \frac{1}{2n} = \frac{1}{2n(2n-1)}.$$

It is easy to compare with $1/n^2$ to get that the partial sums S_{2n} converges. But $S_{2n+1} - S_{2n} = 1/(2n+1) \rightarrow 0$, so the S_{2n+1} converge to the same limit.

Then it does not converge absolutely, because the sum of the absolute values is the harmonic series.

Lemma. Let $\sum a_n$ converge absolutely. Then $\sum a_n$ converges.

Proof. We know that $\sum |a_n|$ converges. Let $S_n = \sum_{n=1}^N a_n$ and $T_N = \sum_{n=1}^N |a_n|$.

We know two ways to show random sequences converge, without knowing what they converge to, namely Monotone sequences and Cauchy sequences. Since S_n is not monotone, we shall try Cauchy sequences.

If $p > q$, then

$$|S_p - S_q| = \left| \sum_{q=1}^p a_n \right| \leq \sum_{n=q+1}^p |a_n| = T_p - T_q.$$

But the sequence T_p converges. So $\forall \varepsilon > 0$, we can find N such that $\forall p > q \geq N$, $T_p - T_q < \varepsilon$, which implies $|S_p - S_q| < \varepsilon$. \square

Definition (Unconditional convergence). A series $\sum a_n$ *converges unconditionally* if the series $\sum_{n=1}^\infty a_{\pi(n)}$ converges for every bijection $\pi : \mathbb{N} \rightarrow \mathbb{N}$, ie. no matter how we re-order the elements of a_n , the sum still converges.

Theorem. If $\sum a_n$ converges absolutely, then it converges unconditionally.

Proof. Let $S_n = \sum_{n=1}^N a_{\pi(n)}$. Then if $p > q$,

$$|S_p - S_q| = \left| \sum_{n=q+1}^p a_{\pi(n)} \right| \leq \sum_{n=q+1}^{\infty} |a_{\pi(n)}|.$$

Let $\varepsilon > 0$. Since $\sum |a_n|$ converges, pick M such that $\sum_{n=M+1}^{\infty} |a_n| < \varepsilon$.

Pick N large enough that $\{1, \dots, M\} \subseteq \{\pi(1), \dots, \pi(N)\}$.

Then if $n > N$, we have $\pi(n) > M$. Therefore if $p > q \geq N$, then

$$|S_p - S_q| \leq \sum_{n=q+1}^p |a_{\pi(n)}| \leq \sum_{n=M+1}^{\infty} |a_n| < \varepsilon.$$

Therefore the sequence of partial sums is Cauchy. \square

Theorem. If $\sum a_n$ converges unconditionally, then it converges absolutely.

Proof. We will prove the contrapositive: if it doesn't converge absolutely, it doesn't converge unconditionally.

Suppose that $\sum |a_n| = \infty$. Let (b_n) be the subsequence of non-negative terms of a_n , and c_n be the subsequence of negative terms. Then $\sum b_n$ and $\sum c_n$ cannot both converge, or else $\sum |a_n|$ converges.

wlog, $\sum b_n = \infty$. Now construct a sequence $0 = n_0 < n_1 < n_2 < \dots$ such that $\forall k$,

$$b_{n_{k-1}+1} + b_{n_{k-1}+2} + \dots + b_{n_k} + c_k \geq 1,$$

This is possible because the b_n are unbounded and we can get it as large as we want.

Let π be the rearrangement

$$b_1, b_2, \dots, b_{n_1}, c_1, b_{n_1+1}, \dots, b_{n_2}, c_2, b_{n_2+1}, \dots, b_{n_3}, c_3, \dots$$

So the sum up to c_k is at least k . So the partial sums tend to infinity. \square

We can prove an even stronger result:

Lemma. Let $\sum a_n$ be a series that converges absolutely. Then for any bijection $\pi : \mathbb{N} \rightarrow \mathbb{N}$,

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_{\pi(n)}.$$

Proof. Let $\varepsilon > 0$. We know that both $\sum |a_n|$ and $\sum |a_{\pi(n)}|$ converge. So let M be such that $\sum_{n>M} |a_n| < \frac{\varepsilon}{2}$ and $\sum_{n>M} |a_{\pi(n)}| < \frac{\varepsilon}{2}$.

Now N be large enough such that

$$\{1, \dots, M\} \subseteq \{\pi(1), \dots, \pi(N)\},$$

and

$$\{\pi(1), \dots, \pi(M)\} \subseteq \{1, \dots, N\}.$$

Then for every $n \geq N$,

$$\left| \sum_{n=1}^N a_n - \sum_{n=1}^N a_{\pi(n)} \right| \leq \sum_{n=M+1}^N |a_n| + \sum_{n=M+1}^N |a_{\pi(n)}| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

We have the first inequality since given our choice of M and N , the first M terms of the $\sum a_n$ and $\sum a_{\pi(n)}$ sums are cancelled by some term in the huge sum.

So $\forall n \geq N$, the partial sums differ by at most ε . So $|\sum a_n - \sum a_{\pi(n)}| \leq \varepsilon$. \square

3.2 Convergence tests

Lemma (Alternating sequence test). Let (a_n) be a decreasing sequence of non-negative reals, and suppose that $a_n \rightarrow 0$. Then $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ converges, ie. $(a_1 - a_2 + a_3 - a_4 + \dots)$ converges).

Proof. Let $S_N = \sum_{n=1}^N (-1)^{n+1} a_n$. Then

$$S_{2n} = (a_1 - a_2) + (a_3 - a_4) + \dots + (a_{2n-1} - a_{2n}) \geq 0,$$

and (S_{2n}) is an increasing sequence.

Also,

$$S_{2n+1} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \dots - (a_{2n} - a_{2n+1}),$$

and (S_{2n+1}) is a decreasing sequence. Also $S_{2n+1} - S_{2n} = a_{2n+1} \geq 0$.

It follows from the monotone sequences property that (S_{2n}) and (S_{2n+1}) converge (they are bounded since $0 \leq S_{2n} \leq S_{2n+1} \leq a_1$).

Since $S_{2n+1} - S_{2n} = a_{2n+1} \rightarrow 0$, they converge to the same limit. \square

Example.

$$1 - \frac{1}{\sqrt[3]{2}} + \frac{1}{\sqrt[3]{3}} - \frac{1}{\sqrt[3]{4}} + \dots \text{converges.}$$

Lemma (Ratio test v1). We have three versions:

(i) If $\exists c < 1$ such that

$$\frac{|a_{n+1}|}{|a_n|} \leq c,$$

for all n , then $\sum a_n$ converges.

(ii) If $\exists c < 1$ and $\exists N$ such that

$$\forall n \geq N : \frac{|a_{n+1}|}{|a_n|} \leq c,$$

then $\sum a_n$ converges. Note that just because the ratio is always less than 1 does not mean it converges. It has to be always less than a fixed number c . Otherwise the test will say that $\sum 1/n$ converges.

(iii) If $\exists \rho \in (-1, 1)$ such that

$$\frac{a_{n+1}}{a_n} \rightarrow \rho,$$

then $\sum a_n$ converges. Note that we have the *open* interval $(-1, 1)$. If $\frac{|a_{n+1}|}{|a_n|} \rightarrow 1$, then the test is inconclusive!

Proof.

- (i) $|a_n| \leq c^{n-1}|a_1|$. Since $\sum c^n$ converges, so does $\sum |a_n|$ by comparison test. So $\sum a_n$ converges absolutely, so it converges.
- (ii) For all $k \geq 0$, we have $|a_{N+k}| \leq c^k |a_N|$. So the series $\sum |a_{N+k}|$ converges, and therefore so does $\sum |a_k|$.
- (iii) If $\frac{a_{n+1}}{a_n} \rightarrow \rho$, then $\frac{|a_{n+1}|}{|a_n|} \rightarrow |\rho|$. So (setting $\varepsilon = (1 - |\rho|)/2$) there exists N such that $\forall n \geq N$, $\frac{|a_{n+1}|}{|a_n|} \leq \frac{1+|\rho|}{2} < 1$. So the result follows from (ii). \square

Example. If $|b| < 1$, then $\sum nb^n$ converges, since

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)a^{n+1}}{na^n} = \left(1 + \frac{1}{n}\right)a \rightarrow a < 1.$$

So it converges.

Note: we can evaluate this directly by evaluating $\sum_{i=0}^{\infty} \sum_{n=i}^{\infty} a^n$.

The following two tests were taught at the end of the course, but are included here for the sake of completeness.

Theorem (Condensation test). Let (a_n) be a decreasing non-negative sequence. Then $\sum_{n=1}^{\infty} a_n < \infty$ if and only if

$$\sum_{k=1}^{\infty} 2^k a_{2^k} < \infty.$$

Proof. This is basically the proof that $\sum \frac{1}{n}$ diverges and $\sum \frac{1}{n^\alpha}$ converges for $\alpha < 1$ but written in a more general way.

We have

$$\begin{aligned} & a_1 + a_2 + (a_3 + a_4) + (a_5 + \cdots + a_8) + (a_9 + \cdots + a_{16}) + \cdots \\ & \geq a_1 + a_2 + 2a_4 + 4a_8 + 8a_{16} + \cdots \end{aligned}$$

So if $\sum 2^k a_{2^k}$ diverges, $\sum a_n$ diverges.

To prove the other way round, simply group as

$$\begin{aligned} & a_1 + (a_2 + a_3) + (a_4 + \cdots + a_7) + \cdots \\ & \leq a_1 + 2a_2 + 4a_4 + \cdots \end{aligned}$$

\square

Example. If $a_n = \frac{1}{n}$, then $2^k a_{2^k} = 1$. So $\sum_{k=1}^{\infty} 2^k a_{2^k} = \infty$. So $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$.

After we formally define integrals, we will prove the integral test:

Theorem (Integral test). Let $f : [1, \infty) \rightarrow \mathbb{R}$ be a decreasing non-negative function. Then $\sum_{n=1}^{\infty} f(n)$ converges iff $\int_1^{\infty} f(x) dx < \infty$.

3.3 Complex versions

Most definitions in the course so far carry over unchanged to the complex numbers. eg., $z_n \rightarrow z$ iff $\forall \varepsilon > 0 \exists N \forall n \geq N : |z_n - z| < \varepsilon$.

Two exceptions are least upper bound and monotone sequences, because the complex numbers do not have an ordering! (It cannot be made into an ordered field because the square of every number in an ordered field has to be positive) Fortunately, Cauchy sequences still work.

We can prove most analogues of the theorems so far by looking at the real and imaginary parts.

Example. Let (z_n) be a Cauchy sequence in \mathbb{C} . Let $z_n = x_n + iy_n$. Then (x_n) and (y_n) are Cauchy. So they converge, from which it follows that $z_n = x_n + iy_n$ converges.

Also, the Bolzano-Weierstrass theorem still holds: If (z_n) is bounded, let $z_n = x_n + iy_n$, then (x_n) and (y_n) are bounded. Then find a subsequence (x_{n_k}) that converges. Then find a subsequence of (y_{n_k}) that converges.

Then nested-intervals property has a “nested-box” property as a complex analogue.

Finally, the proof that absolutely convergent sequences converge still works. It follows that the ratio test still works.

Example. If $|z| < 1$, then $\sum nz^n$ converges. Proof is the same as above.

This is a useful technique which we will use to prove Abel’s test.

Proposition. Let (a_n) and (b_n) be two sequences, and for each n , let S_n be the partial sum $\sum_{k=1}^n a_k$. Then

$$\sum_{n=M+1}^N a_n b_n = S_N b_N - S_M b_{M+1} + \sum_{n=M+1}^{N-1} S_n (b_n - b_{n+1}).$$

This is a discrete analogue of integration by parts.

Proof.

$$\begin{aligned} \sum_{n=M+1}^N a_n b_n &= \sum_{n=M+1}^N (S_n - S_{n-1}) b_n \\ &= \sum_{n=M+1}^N S_n b_n - \sum_{n=M}^{N-1} S_n b_{n+1} \\ &= S_N b_N - S_M b_{M+1} + \sum_{n=M+1}^{N-1} S_n (b_n - b_{n+1}) \end{aligned}$$

□

It is often more helpful to remember the proof and then derive it when needed, instead of remembering the result itself.

Lemma (Abel’s test). Let $a_1 \geq a_2 \geq \dots \geq 0$, and suppose that $a_n \rightarrow 0$. Let $z \in \mathbb{C}$ such that $|z| = 1$ and $z \neq 1$. Then $\sum a_n z^n$ converges.

Proof. Let $S_n = \sum_{k=1}^n z^k = \frac{z-z^{n+1}}{1-z}$. Then $|S_n| \leq \frac{2}{|1-z|}$. By partial summation,

$$\begin{aligned} \sum_{n=M+1}^N z^n \cdot a_n &= S_N a_N - S_M a_{M+1} + \sum_{n=M+1}^{N-1} S_n (a_n - a_{n+1}) \\ \left| \sum_{n=M+1}^N z^n \cdot a_n \right| &\leq \frac{2}{|1-z|} \left(a_N + a_{M+1} + \sum_{n=M+1}^{N-1} (a_n - a_{n+1}) \right) \\ &= \frac{2}{|1-z|} (a_N + a_{M+1} + a_{M+1} - a_N) \\ &= \frac{4a_{M+1}}{|1-z|} \end{aligned}$$

So if $T_N = \sum_{n=1}^N a_n z^n$, then we have that $|T_N - T_M| \leq \frac{4|a_{M+1}|}{|1-z|}$, which tends to 0 as $M \rightarrow \infty$. It follows that T_N is Cauchy. So $\sum a_n z^n$ converges. \square

Example. The series $\sum z^n/n$ converges if $|z| < 1$ or if $|z| = 1$ and $z = 1$, and it diverges if $z = 1$ or $|z| > 1$.

The cases $|z| < 1$ and $|z| > 1$ are trivial from the ratio test, and Abel's test is required for the $|z| = 1$ cases.

4 Continuous functions

Definition (Continuous function). Let $A \subseteq \mathbb{R}$, $a \in A$, and $f : A \rightarrow \mathbb{R}$. Then f is *continuous at a* if

$$\forall \varepsilon > 0 \exists \delta > 0 \forall y \in A : |y - a| < \delta \Rightarrow |f(y) - f(a)| < \varepsilon.$$

f is *continuous* if it is continuous at each every $a \in A$. That is

$$\forall a \in A \forall \varepsilon > 0 \exists \delta > 0 \forall y \in A : |y - a| < \delta \Rightarrow |f(y) - f(a)| < \varepsilon.$$

Intuitively, f is continuous at a if we can obtain $f(a)$ as accurately as we wish by using more accurate values of a (the definition says that if we want to approximate $f(a)$ by $f(y)$ to within accuracy ε , we just have to get our y to within δ of a for some δ).

For example, given the function

$$f(x) = \begin{cases} 0 & x \leq \pi \\ 1 & x > \pi \end{cases}.$$

Suppose we have a computer program that computes this function (which we don't know the formula for), and we want to know what $f(\pi)$ is. Since we cannot input π (it has infinitely many digits), we can try 3, and it gives 0. Then we try 3.14, and it gives 0 again. If we try 3.1416, it gives 1 (since $\pi = 3.1415926 \dots < 3.1416$). We keep giving more and more digits of π , but the result keeps oscillating between 0 and 1. We have no hope of what $f(\pi)$ might be, even approximately. So this f is discontinuous at π .

However, if we have the function $g(x) = x^2$, then we *can* find the (approximate) value of $g(\pi)$. We can first try $g(3)$ and obtain 9. Then we can try $g(3.14) = 9.8596$, $g(3.1416) = 9.86965056$ etc. We can keep trying and obtain more and more accurate values of $g(\pi)$. So g is continuous at π .

Example.

- Constant functions are continuous.
- The function $f(x) = x$ is continuous (take $\delta = \varepsilon$).

Lemma. The following two statements are equivalent for a function $f : A \rightarrow \mathbb{R}$.

- f is continuous
- If (a_n) is a sequence in A with $a_n \rightarrow a$, then $f(a_n) \rightarrow f(a)$.

Proof. (i) \Rightarrow (ii) Let $\varepsilon > 0$ Since f is continuous at a ,

$$\exists \delta > 0 \forall y \in A : |y - a| < \delta \Rightarrow |f(y) - f(a)| < \varepsilon.$$

We want N such that $\forall n \geq N$, $|f(a_n) - f(a)| < \varepsilon$. It is therefore enough to find N such that $\forall n \geq N$, $|a_n - a| < \delta$. Since $a_n \rightarrow a$, such an N exists.

(ii) \Rightarrow (i) We prove the contrapositive: Suppose f is not continuous at a . Then

$$\exists \varepsilon > 0 \forall \delta > 0 \exists y \in A : |y - a| < \delta \text{ and } |f(y) - f(a)| \geq \varepsilon.$$

For each n , we can therefore pick $a_n \in A$ such that $|a_n - a| < \frac{1}{n}$ and $|f(a_n) - f(a)| \geq \varepsilon$. But then $a_n \rightarrow a$ (by Archimedean property), but $f(a_n) \not\rightarrow f(a)$. \square

Example.

(i) Let $f(x) = \begin{cases} -1 & x < 0 \\ 1 & x \geq 0 \end{cases}$. Then f is not continuous because $-\frac{1}{n} \rightarrow 0$ but $f(-\frac{1}{n}) \rightarrow -1 \neq f(0)$.

(ii) Let $f : \mathbb{Q} \rightarrow \mathbb{R}$ with

$$f(x) = \begin{cases} 1 & x^2 > 2 \\ 0 & x^2 < 2 \end{cases}$$

Then f is continuous. For every $a \in \mathbb{Q}$, we can find an interval about a on which f is constant. So f is continuous at a .

(iii) Let

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

Then f is discontinuous. For example, let $a_n = 1/[(2n + 0.5)\pi]$. Then $a_n \rightarrow 0$ and $f(a_n) \rightarrow 1 \neq f(0)$.

Lemma. Let $A \subseteq \mathbb{R}$ and $f, g : A \rightarrow \mathbb{R}$ be continuous functions. Then

- (i) $f + g$ is continuous
- (ii) fg is continuous
- (iii) if g never vanishes, then f/g is continuous.

Proof.

(i) Let $a \in A$ and let (a_n) be a sequence in A with $a_n \rightarrow a$. Then

$$(f + g)(a_n) = f(a_n) + g(a_n).$$

But $f(a_n) \rightarrow f(a)$ and $g(a_n) \rightarrow g(a)$. So

$$f(a_n) + g(a_n) \rightarrow f(a) + g(a) = (f + g)(a).$$

With exactly the same technique for (ii) and (iii). □

With this lemma, from the fact that constant functions and $f(x) = x$ are continuous, we know that all polynomials are continuous. Similarly, rational functions $P(x)/Q(x)$ are continuous except when $Q(x) = 0$.

Lemma. Let $f : A \rightarrow B$ and $g : B \rightarrow \mathbb{R}$. Then if f and g are continuous, $g \circ f : A \rightarrow \mathbb{R}$ is continuous.

Proof. We offer two proofs:

- (i) Let (a_n) be a sequence in A with $a_n \rightarrow a \in A$. Then $f(a_n) \rightarrow f(a)$ since f is continuous. Then $g(f(a_n)) \rightarrow g(f(a))$ since g is continuous. So $g \circ f$ is continuous.

- (ii) Let $a \in A$ and $\varepsilon > 0$. Since g is continuous at $f(a)$, there exists $\eta > 0$ such that $\forall z \in B$, $|z - f(a)| < \eta \Rightarrow |g(z) - g(f(a))| < \varepsilon$.

Since f is continuous at a , $\exists \delta > 0$ such that $\forall y \in A$, $|y - a| < \delta \Rightarrow |f(y) - f(a)| < \eta$. Therefore $|y - a| < \delta \Rightarrow |g(f(y)) - g(f(a))| < \varepsilon$.

□

Theorem. Let $[a, b]$ be a closed interval in \mathbb{R} and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is bounded and attains its bounds, i.e. $f(x) = \sup f$ for some x , and $f(y) = \inf f$ for some y .

Proof. If f is not bounded above, then for each n , we can find $x_n \in [a, b]$ such that $f(x_n) \geq n$ for all n .

By Bolzano-Weierstrass, since $x \in [a, b]$ and is bounded, the sequence (x_n) has a convergent subsequence (x_{n_k}) . Let x be its limit. Then since f is continuous, $f(x_{n_k}) \rightarrow f(x)$. But $f(x_{n_k}) \geq n_k \rightarrow \infty$. So this is a contradiction.

Now let $C = \sup\{f(x) : x \in [a, b]\}$. Then for every n , we can find x_n such that $f(x_n) \geq C - \frac{1}{n}$. So by Bolzano-Weierstrass, (x_n) has a convergent subsequence (x_{n_k}) . Since $C - \frac{1}{n_k} \leq f(x_{n_k}) \leq C$, $f(x_{n_k}) \rightarrow C$. Therefore if $x = \lim x_{n_k}$, then $f(x) = C$.

A similar argument applies if f is unbounded below. □

Theorem (Intermediate value theorem). Let $a < b \in \mathbb{R}$ and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Suppose that $f(a) < 0 < f(b)$. Then there exists an $x \in (a, b)$ such that $f(x) = 0$.

Proof. We have several proofs:

- (i) Let $A = \{x : f(x) < 0\}$ and let $s = \sup A$. We shall show that $f(s) = 0$ (this is similar to the proof that $\sqrt{2}$ exists in Numbers and Sets). If $f(s) < 0$, then setting $\varepsilon = |f(s)|$ in the definition of continuity, we can find $\delta > 0$ such that $\forall y$, $|y - s| < \delta \Rightarrow f(y) < f(s) < 0$. Then $s + \delta/2 \in A$, so s is not an upper bound. Contradiction.

If $f(s) > 0$, by the same argument, we can find $\delta > 0$ such that $\forall y$, $|y - s| < \delta \Rightarrow f(y) > f(s) > 0$. So $s - \delta/2$ is a smaller upper bound.

- (ii) Let $a_0 = a$, $b_0 = b$. By repeated bisection, construct nested intervals $[a_n, b_n]$ such that $b_n - a_n = \frac{b_0 - a_0}{2^n}$ and $f(a_n) < 0 \leq f(b_n)$. Then by the nested intervals property, we can find $x \in \cap_{n=0}^{\infty} [a_n, b_n]$. Since $b_n - a_n \rightarrow 0$, $a_n, b_n \rightarrow x$.

Since $f(a_n) < 0$ for every n , $f(x) \leq 0$. Similarly, since $f(b_n) \geq 0$ for every n , $f(x) \geq 0$. So $f(x) = 0$.

□

Note: It is easy to generalize this to get that, if $f(a) < c < f(b)$, then $\exists x \in (a, b)$ such that $f(x) = c$, by applying the result to $f(x) - c$. Also, we can assume instead that $f(b) < c < f(a)$ and obtain the same result by looking at $-f(x)$.

Corollary. Let $f : [a, b] \rightarrow [c, d]$ be a continuous strictly increasing function with $f(a) = c$, $f(b) = d$. Then f is invertible and its inverse is continuous.

Proof. Since f is strictly increasing, it is an injection (suppose $x \neq y$. wlog, $x < y$. Then $f(x) < f(y)$ and so $f(x) \neq f(y)$). Now let $y \in (c, d)$. By the intermediate value theorem, there exists $x \in (a, b)$ such that $f(x) = y$. So f is a surjection. So it is a bijection and hence invertible.

Let g be the inverse. Let $y \in [c, d]$ and let $\varepsilon > 0$. Let $x = g(y)$. So $f(x) = y$. Let $u = f(x - \varepsilon)$ and $v = f(x + \varepsilon)$ (if $y = c$ or d , make the obvious adjustments). Then $u < y < v$. So we can find $\delta > 0$ such that $(y - \delta, y + \delta) \subseteq (u, v)$. Then $|z - y| < \delta \Rightarrow g(z) \in (x - \varepsilon, x + \varepsilon) \Rightarrow |g(z) - g(y)| < \varepsilon$. \square

With this corollary, we can create more continuous functions, e.g. \sqrt{x} .

4.1 Continuous induction*

Proposition (Continuous induction v1). Let $a < b$ and let $A \subseteq [a, b]$ have the following properties:

- (i) $a \in A$
- (ii) If $x \in A$ and $x \neq b$, then $\exists y \in A$ with $y > x$.
- (iii) If $\forall \varepsilon > 0$, $\exists y \in A : y \in (x - \varepsilon, x]$, then $x \in A$.

Then $b \in A$.

Proof. Since $a \in A$, $A \neq \emptyset$. A is also bounded above by b . So let $s = \sup A$. Then $\forall \varepsilon > 0$, $\exists y \in A$ such that $y > s - \varepsilon$. Therefore, by (iii), $s \in A$.

If $s \neq b$, then by (ii), we can find $y \in A$ such that $y > s$. \square

We can use this to prove the intermediate value theorem.

Proposition (Continuous induction v2). Let $A \subseteq [a, b]$ and suppose that

- (i) $a \in A$
- (ii) If $[a, x] \subseteq A$ and $x \neq b$, then there exists $y > x$ such that $[a, y] \subseteq A$.
- (iii) If $[a, x] \subseteq A$, then $x \in A$.

Then $A = [a, b]$

Proof. We prove that version 1 \Rightarrow version 2. Suppose A satisfies the conditions of v2. Let $A' = \{x \in [a, b] : [a, x] \subseteq A\}$.

Then $a \in A'$. If $x \in A'$ with $x \neq b$, then $[a, x] \subseteq A$. So $\exists y > x$ such that $[a, y] \subseteq A$. So $\exists y > x$ such that $y \in A'$.

If $\forall \varepsilon > 0$, $\exists y \in (x - \varepsilon, x]$ such that $[a, y] \subseteq A$, then $[a, x] \subseteq A$. So by (iii), $[a, x] \subseteq A$, so $x \in A'$. So A' satisfies properties (i) to (iii) of version 1. Therefore $b \in A'$. So $[a, b] \subseteq A$. So $A = [a, b]$. \square

We reprove intermediate value theorem here:

Theorem (Intermediate value theorem). Let $a < b \in \mathbb{R}$ and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Suppose that $f(a) < 0 < f(b)$. Then there exists an $x \in (a, b)$ such that $f(x) = 0$.

Proof. Assume that f is continuous. Suppose $f(a) < 0 < f(b)$. Assume that $\forall x : f(x) \neq 0$, and derive a contradiction.

Let $A = \{x : f(x) < 0\}$. Then $a \in A$. If $x \in A$, then $f(x) < 0$, and by continuity, we can find $\delta > 0$ such that $|y - x| < \delta \Rightarrow f(y) < 0$. So if $x \neq b$, then we can find $y \in A$ such that $y > x$.

We prove the contrapositive of the last condition, i.e.

$$x \notin A \Rightarrow \exists \delta > 0 : \forall y \in A : y \notin (x - \varepsilon, x].$$

If $x \in A$, then $f(x) > 0$ (we assume that f is never zero. If not, we're done). Then by continuity, $\exists \delta > 0$ such that $|y - x| < \delta \Rightarrow f(y) > 0$. So $y \notin A$.

Hence by continuous induction, $b \in A$. Contradiction. \square

Now we prove that continuous functions in closed intervals are bounded.

Theorem. Let $[a, b]$ be a closed interval in \mathbb{R} and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is bounded and attains its bounds, i.e. $f(x) = \sup f$ for some x , and $f(y) = \inf f$ for some y .

Proof. Let $f : [a, b]$ be continuous. Let $A = \{x : f \text{ is bounded on } [a, x]\}$. Then $a \in A$. If $x \in A, x \neq b$, then $\exists \delta > 0$ such that $|y - x| < \delta \Rightarrow |f(y) - f(x)| < 1$. So $\exists y > x$ (eg. take $\min\{x + \delta/2, b\}$) such that f is bounded on $[a, y]$, which implies that $y \in A$.

Now suppose that $\forall \varepsilon > 0, \exists y \in (x - \varepsilon, x]$ such that $y \in A$. Again, we can find $\delta > 0$ such that f is bounded on $(x - \delta, x + \delta)$, and in particular on $(x - \delta, x]$. Pick y such that f is bounded on $[a, y]$ and $y > x - \delta$. Then f is bounded on $[a, x]$. So $x \in A$.

So we are done by continuous induction. \square

Definition (Cover of a set). Let $A \subseteq \mathbb{R}$. A *cover* of A by open intervals is a set $\{I_\gamma : \gamma \in \Gamma\}$ where each I_γ is an open interval and $A \subseteq \bigcup_{\gamma \in \Gamma} I_\gamma$.

A *finite subcover* is a finite subset $\{I_{\gamma_1}, \dots, I_{\gamma_n}\}$ of the cover that is still a cover.

Not every cover has a finite subcover. For example, the cover $\{(\frac{1}{n}, 1) : n \in \mathbb{N}\}$ of $(0, 1)$ has no finite subcover.

Theorem (Heine-Borel*). Every cover of a closed, bounded interval $[a, b]$ by open intervals has a finite subcover. We say closed intervals are *compact* (cf. Metric and topology).

Proof. Let $\{I_\gamma : \gamma \in \Gamma\}$ be a cover of $[a, b]$ by open intervals. Let $A = \{x : [a, x] \text{ can be covered by finitely many of the } I_\gamma\}$.

Then $a \in A$ since a must belong to some I_γ .

If $x \in A$, then pick γ such that $x \in I_\gamma$. The if $x \neq b$, then since I_γ is an open interval, it contains $[x, y]$ for some $y > x$. Then $[a, y]$ can be covered by finitely many I_γ , by taking a finite cover for $[a, x]$ and the I_γ that contains x .

Now suppose that $\forall \varepsilon > 0, \exists y \in A$ such that $y \in (x - \varepsilon, x]$.

Let I_γ be an open interval containing x . Then it contains $(x - \varepsilon, x]$ for some $\varepsilon > 0$. Pick $y \in A$ such that $y \in (x - \varepsilon, x]$. Now combine I_γ with a finite subcover of $[a, y]$ to get a finite subcover of $[a, x]$. So $x \in A$.

Then done by continuous induction. \square

We can use Heine-Borel to prove that continuous functions on $[a, b]$ are bounded.

Theorem. Let $[a, b]$ be a closed interval in \mathbb{R} and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is bounded and attains its bounds, i.e. $f(x) = \sup f$ for some x , and $f(y) = \inf f$ for some y .

Proof. Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then by continuity,

$$\forall x \in [a, b] \exists \delta_x > 0 \forall y : |y - x| < \delta_x \Rightarrow |f(y) - f(x)| < 1.$$

Let $\gamma = [a, b]$ and for each $x \in \gamma$, let $I_x = (x - \delta_x, x + \delta_x)$. So by Heine-Borel, we can find x_1, \dots, x_n such that $[a, b] \subseteq \bigcup_1^n (x_i - \delta_{x_i}, x_i + \delta_{x_i})$.

But f is bounded in each interval $(x_i - \delta_{x_i}, x_i + \delta_{x_i})$ by $|f(x_i)| + 1$. So it is bounded on $[a, b]$ by $\max |f(x_i)| + 1$. \square

5 Differentiability

5.1 Limits

Definition (Limit of functions). Let $A \subseteq \mathbb{R}$ and let $f : A \rightarrow \mathbb{R}$. We say

$$\lim_{x \rightarrow a} f(x) = \ell,$$

if

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x \in A : 0 < |x - a| < \delta \Rightarrow |f(x) - \ell| < \varepsilon.$$

We couldn't care less what happens when $x = a$, hence the strict inequality $0 < |x - a|$. In fact, f doesn't even have to be defined at $x = a$.

Example. Let

$$f(x) = \begin{cases} x & x \neq 2 \\ 3 & x = 2 \end{cases}$$

Then $\lim_{x \rightarrow 2} f(x) = 2$, even though $f(2) = 3$.

Example. Let $f(x) = \frac{\sin x}{x}$. Then $f(0)$ is not defined but $\lim_{x \rightarrow 0} f(x) = 1$.

We will see a proof later after we define what sin means.

Notation. We also write “ $f(x) \rightarrow \ell$ as $x \rightarrow a$ ” for $\lim_{x \rightarrow a} f(x) = \ell$.

We notice that the definition of the limit is suspiciously similar to that of continuity. In fact, if we define

$$g(x) = \begin{cases} f(x) & x \neq a \\ \ell & x = a \end{cases}$$

Then $f(x) \rightarrow \ell$ as $x \rightarrow a$ iff g is continuous at a .

Alternatively, f is continuous at a if $f(x) \rightarrow f(a)$ as $x \rightarrow a$. It follows also that $f(x) \rightarrow \ell$ as $x \rightarrow a$ iff $f(x_n) \rightarrow \ell$ for every sequence (x_n) in A with $x_n \rightarrow a$.

The previous limit theorems of sequences apply here as well

Proposition. If $f(x) \rightarrow \ell$ and $g(x) \rightarrow m$ as $x \rightarrow a$, then $f(x) + g(x) \rightarrow \ell + m$, $f(x)g(x) \rightarrow \ell m$, and $\frac{f(x)}{g(x)} \rightarrow \frac{\ell}{m}$ if g and m don't vanish.

5.2 Differentiation

Definition (Differentiable function). f is *differentiable* at a with derivative λ if

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lambda.$$

Equivalently, if

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} = \lambda.$$

We write $\lambda = f'(a)$.

Here we see why, in the definition of the limit, we say that we don't care what happens when $x = a$. In our definition here, our function is $0/0$ when $x = a$, and we can't make any sense out of what happens when $x = a$.

Alternatively, write the definition of differentiation as

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \epsilon(h),$$

where $\epsilon(h) \rightarrow 0$ as $h \rightarrow 0$. Rearranging, we can deduce that

$$f(x+h) = f(x) + hf'(x) + h\epsilon(h),$$

Note that this equation remains true if we set $\epsilon(0) = 0$ (or anything - say π , but if $\epsilon(0) = 0$, then ϵ is continuous at 0).

Note also that $\frac{h\epsilon(h)}{h} \rightarrow 0$ as $h \rightarrow 0$. We often write $o(h)$ for a function with the property that $\frac{o(h)}{h} \rightarrow 0$ as $h \rightarrow 0$ (cf. Differential Equations). So

Proposition.

$$f(x+h) = f(x) + hf'(x) + o(h).$$

We can interpret this as an approximation of $f(x+h)$:

$$f(x+h) = \underbrace{f(x) + hf'(x)}_{\text{linear approximation}} + \underbrace{o(h)}_{\text{error term}}.$$

And differentiability shows that this is a very good approximation with small $o(h)$ error.

Conversely, we have

Proposition. If $f(x+h) = f(x) + hf'(x) + o(h)$, then f is differentiable at x with derivative $f'(x)$.

Proof.

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \frac{o(h)}{h} \rightarrow f'(x).$$

□

Definition (Multiple derivatives). This is defined recursively: f is $(n+1)$ -times differentiable if it is n -times differentiable and its n -th derivative $f^{(n)}$ is differentiable. We write $f^{(n+1)}$ for the derivative of $f^{(n)}$, ie. then $(n+1)$ th derivative of f .

Informally, we will say f is n -times differentiable if we can differentiate it n times, and the n th derivative is $f^{(n+1)}$.

Lemma (Sum and product rule). Let f, g be differentiable at x . Then $f+g$ and fg are differentiable at x , with

$$\begin{aligned}(f+g)'(x) &= f'(x) + g'(x) \\ (fg)'(x) &= f'(x)g(x) + f(x)g'(x)\end{aligned}$$

Proof.

$$\begin{aligned}
(f+g)(x+h) &= f(x+h) + g(x+h) \\
&= f(x) + hf'(x) + o(h) + g(x) + hg'(x) + o(h) \\
&= (f+g)(x) + h(f'(x) + g'(x)) + o(h) \\
fg(x+h) &= f(x+h)g(x+h) \\
&= [f(x) + hf'(x) + o(h)][g(x) + hg'(x) + o(h)] \\
&= f(x)g(x) + h[f'(x)g(x) + f(x)g'(x)] \\
&\quad + \underbrace{o(h)[g(x) + f(x) + hf'(x) + hg'(x) + o(h)] + h^2 f'(x)g'(x)}_{\text{error term}}
\end{aligned}$$

By limit theorems, the error term is $o(h)$

$$= f(x) + h(f'(x)g(x) + f(x)g'(x)) + o(h).$$

□

Lemma (Chain rule). If f is differentiable at x and g is differentiable at $f(x)$, then $g \circ f$ is differentiable at x with derivative $g'(f(x))f'(x)$.

Proof. In this proof, we use $h\epsilon(h)$ instead of $o(h)$.

$$\begin{aligned}
(g \circ f)(x+h) &= g(f(x+h)) \\
&= g\left[f(x) + \underbrace{hf'(x) + h\epsilon_1(h)}_{\text{the “}h\text{” term}}\right]
\end{aligned}$$

where $\epsilon_1(h) \rightarrow 0$ and $\epsilon_1(0) = 0$

$$\begin{aligned}
&= g(f(x)) + (fg'(x) + h\epsilon_1(h))g'(f(x)) \\
&\quad + (hf'(x) + h\epsilon_1(h))\epsilon_2(hf'(x) + h\epsilon_1(h)) \\
&= g \circ f(x) + hg'(f(x))f'(x) \\
&\quad + \underbrace{h\left[\epsilon_1(h)g'(f(x)) + (f'(x) + \epsilon_1(h))\epsilon_2(hf'(x) + h\epsilon_1(h))\right]}_{\text{error term}}.
\end{aligned}$$

We want to show that the error term is $o(h)$, ie. it divided by h tends to 0 as $h \rightarrow 0$.

But $\epsilon_1(h)g'(f(x)) \rightarrow 0$, $f'(x) + \epsilon_1(h)$ is bounded, and $\epsilon_2(hf'(x) + h\epsilon_1(h)) \rightarrow 0$ because $hf'(x) + h\epsilon_1(h) \rightarrow 0$ and $\epsilon_2(0) = 0$. So our error term is $o(h)$. □

Example.

- (i) Constant functions are differentiable with derivative 0
- (ii) $f(x) = \lambda x$ is differentiable with derivative λ
- (iii) By induction, x^n is differentiable with derivative nx^{n-1} (product rule)
- (iv) Hence all polynomials are differentiable

Example. Let $f(x) = 1/x$. If $x \neq 0$, then

$$\frac{f(x+h) - f(x)}{h} = \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \frac{\left(\frac{-h}{x(x+h)}\right)}{h} = \frac{-1}{x(x+h)} \rightarrow \frac{-1}{x^2}$$

by limit theorems.

Lemma (Quotient rule). If f and g are differentiable at x , and $g(x) \neq 0$, then f/g is differentiable at x with derivative

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

Proof. First note that $1/g(x) = h(g(x))$ where $h(y) = 1/y$. So $1/g(x)$ is differentiable at x with derivative $\frac{-1}{g(x)^2}g'(x)$ by the chain rule.

By the product rule, f/g is differentiable at x with derivative

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

□

Lemma. If f is differentiable at x , then it is continuous at x .

Proof. As $y \rightarrow x$, $\frac{f(y) - f(x)}{y - x} \rightarrow f'(x)$. Since, $y - x \rightarrow 0$, $f(y) - f(x) \rightarrow 0$ by product theorem of limits. So $f(y) \rightarrow f(x)$. So f is continuous at x . □

Theorem. Let $f : [a, b] \rightarrow [c, d]$ be a differentiable on (a, b) , continuous on $[a, b]$, and strictly increasing. Suppose that $f'(x)$ never vanishes. Suppose further that $f(a) = c$ and $f(b) = d$. Then f has an inverse g and for each $y \in (c, d)$, g is differentiable at y with derivative $1/f'(g(y))$.

In human language, this states that if f is invertible, then the derivative of f^{-1} is $1/f'$.

Note that the conditions will (almost) always require f to be differentiable on *open* interval (a, b) , continuous on *closed* interval $[a, b]$. This is because it doesn't make sense to talk about differentiability at a or b since the definition of $f'(a)$ requires f to be defined on both sides of a . This is the case for all our theorems that require differentiability/continuity.

Proof. g exists by an earlier theorem about inverses of continuous functions.

Let $y, y+k \in (c, d)$. Let $x = g(y)$, $x+h = g(y+k)$. Then $f(x) = y$ and $f(x+h) = y+k$. We want to write $g(y+k)$ as $g(y) + kg'(y) + o(k)$. We have

$$g(y+k) = x+h = g(y) + h. \quad (*)$$

Also

$$y+k = f(x+h) = f(x) + hf'(x) + h\epsilon(h) = y + hf'(x) + h\epsilon(h).$$

So $k = hf'(x) + h\epsilon(h)$. Therefore

$$h = \frac{k}{f'(x) + \epsilon(h)}.$$

As $k \rightarrow 0$, so does h (by continuity of g), and $\epsilon(0) = 0$. So this is

$$h = \frac{k}{f'(x) + \epsilon_1(k)} = \frac{k}{f'(x)} - \frac{k\epsilon_1(k)}{f'(x)(f'(x) + \epsilon_1(k))} = \frac{k}{f'(x)} + k\epsilon_2(k).$$

Since $f'(x) = f'(g(y))$, substituting into (*), we obtain

$$g(y+k) = g(y) + k \frac{1}{f'(g(y))} + \epsilon_2(k).$$

□

Example. Let $f(x) = x^{1/2}$ for $x > 0$. Then f is the inverse of $g(x) = x^2$. So

$$f'(x) = \frac{1}{g'(f(x))} = \frac{1}{2x^{1/2}} = \frac{1}{2}x^{-1/2}.$$

Similarly, we can show that the derivative of $x^{1/q}$ is $\frac{1}{q}x^{1/q-1}$.

Then let's take $x^{p/q} = (x^{1/q})^p$. By the chain rule, its derivative is

$$p(x^{1/q})^{p-1} \cdot \frac{1}{q}x^{1/q-1} = \frac{p}{q}x^{\frac{p-1}{q} + \frac{1}{q} - 1} = \frac{p}{q}x^{\frac{p}{q}-1}.$$

5.3 Differentiation theorems

Theorem (Rolle's theorem). Let f be continuous on a closed interval $[a, b]$ (with $a < b$) and differentiable on (a, b) . Suppose that $f(a) = f(b)$. Then there exists $x \in (a, b)$ such that $f'(x) = 0$.

It is intuitively obvious: if you move up and down, and finally return to the same point, then you must have changed direction some time. Then $f'(x) = 0$ at that time.

Proof. If f is constant, then we're done.

Otherwise, there exists u such that $f(u) \neq f(a)$. wlog, $f(u) > f(a)$. Since f is continuous, it has a maximum, and since $f(u) > f(a) = f(b)$, the maximum is not attained at a or b .

Suppose maximum is attained at $x \in (a, b)$. Then for any $h \neq 0$, we have

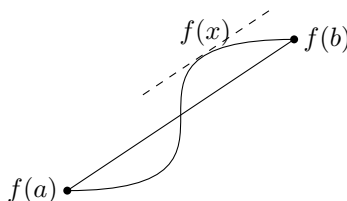
$$\frac{f(x+h) - f(x)}{h} \begin{cases} \leq 0 & h > 0 \\ \geq 0 & h < 0 \end{cases}$$

since $f(x+h) - f(x) \leq 0$ by maximality of $f(x)$. As we take the limit as $h \rightarrow 0$, we know that $f'(x) \leq 0$ and $f'(x) \geq 0$. So $f'(x) = 0$. □

Corollary (Mean value theorem). Let f be continuous on $[a, b]$ ($a < b$), and differentiable on (a, b) . Then there exists $x \in (a, b)$ such that

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

Note that $\frac{f(b)-f(a)}{b-a}$ is the slope of the line joining $f(a)$ and $f(b)$.



Proof. The mean value theorem is often described as “rotate your head and apply Rolle’s”. However, if we actually rotate it, we might end up with a non-function. What we actually want is a shear.

Let

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

Then

$$g(b) - g(a) = f(b) - f(a) - \frac{f(b) - f(a)}{b - a}(b - a) = 0.$$

So by Rolle’s theorem, we can find $x \in (a, b)$ such that $g'(x) = 0$. So

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

as required. \square

Example. Suppose $f'(x) > 0$ for every $x \in (a, b)$. Then for u, v in $[a, b]$, we can find $w \in (u, v)$ such that

$$\frac{f(v) - f(u)}{v - u} = f'(w) > 0.$$

It follows that $f(v) > f(u)$. So f is strictly increasing.

Similarly, if $f'(x) \geq 2$ for every x and $f(0) = 0$, then $f(1) \geq 2$, since we can find $x \in (0, 1)$ such that

$$2 \leq f'(x) = \frac{f(1) - f(0)}{1 - 0} = f(1).$$

Theorem (1D form of inverse function theorem). Let f be a function with continuous derivative on (a, b) .

Let $x \in (a, b)$ and suppose that $f'(x) \neq 0$. Then there is an open interval (u, v) containing x on which f is invertible (as a function from (u, v) to $f((u, v))$). Moreover, if g is the inverse, then $g'(f(z)) = \frac{1}{f'(z)}$ for every $z \in (u, v)$.

This says that if f has a non-zero derivative, then it has an inverse locally and the derivative of the inverse is $1/f'$.

Note that this not only requires f to be differentiable, but the derivative itself also has to be continuous.

Proof. wlog, $f'(x) > 0$. By the continuity, of f' , we can find $\delta > 0$ such that $f'(z) > 0$ for every $z \in (x - \delta, x + \delta)$. By the mean value theorem, f is strictly increasing on $(x - \delta, x + \delta)$, hence injective. Also, f is continuous on $(x - \delta, x + \delta)$ by differentiability.

Then done by the inverse function theorem. \square

Theorem (Higher-order Rolle's theorem). Let f be continuous on $[a, b]$ ($a < b$) and n -times differentiable on an open interval containing $[a, b]$. Suppose that

$$f(a) = f'(a) = f^{(2)}(a) = \dots = f^{(n-1)}(a) = f(b) = 0.$$

Then $\exists x \in (a, b)$ such that $f^{(n)}(x) = 0$.

Proof. Induct on n . The $n = 0$ base case is just Rolle's theorem.

Suppose we have $k < n$ and $x_k \in (a, b)$ such that $f^{(k)}(x_k) = 0$. Since $f^{(k)}(a) = 0$, we can find $x_{k+1} \in (a, x_k)$ such that $f^{(k+1)}(x_{k+1}) = 0$ by Rolle's theorem.

So the result follows by induction. \square

Corollary. Suppose that f and g are both differentiable on an open interval containing $[a, b]$ and that $f^{(k)}(a) = g^{(k)}(a)$ for $k = 0, 1, \dots, n-1$, and also $f(b) = g(b)$. Then there exists $x \in (a, b)$ such that $f^{(n)}(x) = g^{(n)}(x)$.

Proof. Apply generalised Rolle's to $f - g$. \square

Now we shall show that for any f , we can find a polynomial p of degree at most n that satisfies the conditions for g , ie. a p such that $p^{(k)}(a) = f^{(k)}(a)$ for $k = 0, 1, \dots, n-1$ and $p(b) = f(b)$.

A useful ingredient is the observation that if

$$Q_k(x) = \frac{(x-a)^k}{k!},$$

then

$$Q_k^{(j)}(a) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

Therefore, if

$$Q(x) = \sum_{k=0}^{n-1} f^{(k)}(a) Q_k(x),$$

then

$$Q^{(j)}(a) = f^{(j)}(a)$$

for $j = 0, 1, \dots, n-1$. To get $p(b) = f(b)$, we use our n th degree polynomial term:

$$p(x) = Q(x) + \frac{(x-a)^n}{(b-a)^n} (f(b) - Q(b)).$$

Then our final term does not mess up our first $n-1$ derivatives, and $p(b) = f(b)$.

By the previous corollary, we can find $x \in (a, b)$ such that

$$f^{(n)}(x) = p^{(n)}(x).$$

That is,

$$f^{(n)}(x) = \frac{n!}{(b-a)^n} (f(b) - Q(b)).$$

Therefore

$$f(b) = Q(b) + \frac{(b-a)^n}{n!} f^{(n)}(x).$$

Alternatively,

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

Setting $h = b - a$, we can rewrite this as

Theorem (Taylor's theorem).

$$f(a+h) = \underbrace{f(a) + hf'(a) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a)}_{(n-1)\text{-degree approximation to } f \text{ near } a} + \underbrace{\frac{h^n}{n!}f^{(n)}(x)}_{\text{error term}}.$$

for some $x \in (a, a+h)$.

Note: Strictly speaking, we only proved it for the case when $h > 0$, but we can easily show it holds for $h < 0$ too by considering $g(x) = f(-x)$.

Note that the remainder term is *not* necessarily small, but this often gives us the best $(n-1)$ -degree approximation to f near a . For example, if $f^{(n)}$ is bounded by C near a , then

$$\left| \frac{h^n}{n!}f^{(n)}(x) \right| \leq \frac{C}{n!}|h|^n = o(h^{n-1}).$$

Example. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a differentiable function such that $f(0) = 1$ and $f'(x) = f(x)$ for every x (intuitively, we know it is e^x , but that thing doesn't exist!). Then for every x , we have

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

While it seems like we can prove this works by differentiating it and see that $f'(x) = f(x)$, the sum rule only applies for finite sums. We don't know we can differentiate a sum term by term. So we have to use Taylor's theorem.

Since $f'(x) = f(x)$, it follows that all derivatives exist. By Taylor's theorem,

$$f(x) = f(0) + f'(0)x + \frac{f^{(2)}(0)}{2!}x^2 + \cdots + \frac{f^{(n-1)}(0)}{(n-1)!}x^{n-1} + \frac{f^{(n)}(u)}{n!}x^n.$$

for some u between 0 and x . This equals to

$$f(x) = \sum_{k=0}^{n-1} \frac{x^k}{k!} + \frac{f^{(n)}(u)}{n!}x^n.$$

We must show that the remainder term $\frac{f^{(n)}(u)}{n!}x^n \rightarrow 0$ as $n \rightarrow \infty$. (*Note:* x is fixed, but u can depend on n)

But we know that $f^{(u)} = f(u)$, but since f is differentiable, it is continuous, and is bounded on $[0, x]$. Suppose $|f(u)| \leq C$ on $[0, x]$. Then

$$\left| \frac{f^{(n)}(u)}{n!}x^n \right| \leq \frac{C}{n!}|x|^n \rightarrow 0$$

from limit theorems.

5.4 Complex differentiation

Definition (Complex differentiability). Let $f : \mathbb{C} \rightarrow \mathbb{C}$ (or more generally, from $A \subseteq \mathbb{C} \rightarrow \mathbb{C}$), then f is differentiable at z with derivative $f'(z)$ if

$$\lim_{h \rightarrow 0} \frac{f(z+h) - f(z)}{h} \text{ exists and equals } f'(z).$$

Equivalently,

$$f(z+h) = f(z) + hf'(z) + o(h).$$

This is exactly the same definition with real differentiation, but has very different properties!

All the usual rules - chain rule, product rule etc. also apply (with the same proofs). Also the derivatives of polynomials are what you expect.

Example. $f(z) = \bar{z}$ is *not* differentiable.

$$\frac{\overline{z+h} - \bar{z}}{h} = \frac{\bar{h}}{h} = \begin{cases} 1 & h \text{ is real} \\ -1 & h \text{ is purely imaginary} \end{cases}$$

The weirdness is because we often think of \mathbb{C} as \mathbb{R}^2 , but they are not the same. For example, reflection is a linear map in \mathbb{R}^2 , but not in \mathbb{C} . A linear map in \mathbb{C} is something in the form $x \mapsto bx$, which can only be a dilation or rotation, not reflections or other weird things.

Example. $f(z) = |z|$ is also not differentiable. If it were, then $|z|^2$ would be as well (by the product rule). So would $\frac{|z|^2}{z} = \bar{z}$ when $z \neq 0$ by the quotient rule. At $z = 0$, it is certainly not differentiable, since it is not even differentiable on \mathbb{R} .

6 Complex power series

Definition (Complex power series). A *complex power series* is a series of the form

$$\sum_{n=0}^{\infty} a_n z^n.$$

when $z \in \mathbb{C}$ and $a_n \in \mathbb{C}$ for all n . When it converges, it is a function of z .

Lemma. Suppose that $\sum_{n=0}^{\infty} a_n z^n$ converges and $|w| < |z|$, then $\sum a_n w^n$ converges. Moreover, $\sum a_n w^n$ converges absolutely.

Proof. We know that

$$|a_n w^n| = |a_n z^n| \cdot \left| \frac{w}{z} \right|^n.$$

Since $\sum a_n z^n$ converges, the terms $a_n z^n$ are bounded. So pick C such that

$$|a_n z^n| \leq C$$

for every n . Then

$$0 \leq \sum_{n=0}^{\infty} |a_n w^n| \leq \sum_{n=0}^{\infty} C \left| \frac{w}{z} \right|^n,$$

which converges (geometric series). So by the comparison test, $\sum a_n w^n$ converges absolutely. \square

Note: It follows that if $\sum a_n z^n$ does not converge, and $|w| > |z|$, then $\sum a_n w^n$ does not converge.

Now let $R = \sup\{|z| : \sum a_n z^n \text{ converges}\}$ (R may be infinite). If $|z| < R$, then we can find z_0 with $|z_0| \in (|z|, R]$ such that $\sum_{n=0}^{\infty} a_n z_0^n$ converges. So by lemma above, $\sum a_n z^n$ converges. If $|z| > R$, then $\sum a_n z^n$ diverges by definition of R .

Definition (Radius of convergence). The *radius of convergence* of a power series $\sum a_n z^n$ is

$$R = \sup \left\{ |z| : \sum a_n z^n \text{ converges} \right\}.$$

$\{z : |z| < R\}$ is called the *circle of convergence*.¹

If $|z| < R$, then $\sum a_n z^n$ converges. If $|z| > R$, then $\sum a_n z^n$ diverges. When $|z| = R$, the series can converge at some points and not the others.

Example. $\sum_{n=0}^{\infty} z^n$ has radius of convergence of 1. When $|z| = 1$, it diverges (since the terms do not tend to 0).

Example. $\sum_{n=0}^{\infty} \frac{z^n}{n}$ has radius of convergence 1, since the ratio of $(n+1)$ th term to n th is

$$\frac{z^{n+1}/(n+1)}{z^n/n} = z \cdot \frac{n}{n+1} \rightarrow z.$$

¹Note to pedants: yes it is a disc, not a circle

So if $|z| < 1$, then the series converges by the ratio test. If $|z| > 1$, then eventually the terms are increasing in modulus.

If $z = 1$, then it diverges (harmonic series). If $|z| = 1$ and $z \neq 1$, it converges by Abel's test.

Example. The series $\sum_{n=1}^{\infty} \frac{z^n}{n^2}$ converges for $|z| \leq 1$ and diverges for $|z| > 1$.

Lemma. The radius of convergence of a power series $\sum a_n z^n$ is

$$R = \frac{1}{\limsup \sqrt[n]{|a_n|}}.$$

Often $\sqrt[n]{|a_n|}$ converges, so we only have to find the limit.

Proof. Suppose $|z| < 1/\limsup \sqrt[n]{|a_n|}$. Then $|z| \limsup \sqrt[n]{|a_n|} < 1$. Therefore there exists N and $\varepsilon > 0$ such that

$$\sup_{n \geq N} |z| \sqrt[n]{|a_n|} \leq 1 - \frac{1}{\varepsilon}$$

by the definition of \limsup . Therefore

$$|a_n z^n| \leq (1 - \varepsilon)^n$$

for every $n \geq N$, which implies (by comparison with geometric series) that $\sum a_n z^n$ converges absolutely.

On the other hand, if $|z| \limsup \sqrt[n]{|a_n|} > 1$, it follows that $|z| \sqrt[n]{|a_n|} \geq 1$ for infinitely many n . Therefore $|a_n z^n| \geq 1$ for infinitely many n . So $\sum a_n z^n$ does not converge. \square

Example. The radius of convergence of $\frac{z^n}{2^n}$ is 2 because $\sqrt[n]{|a_n|} = \frac{1}{2}$ for every n . So $\limsup \sqrt[n]{|a_n|} = \frac{1}{2}$. So $1/\limsup \sqrt[n]{|a_n|} = 2$.

But often it is easier to find the radius of convergence from elementary methods such as the ratio test, eg. for $\sum n^2 z^n$.

6.1 Exponential and trigonometric functions

Definition (Exponential function). The *exponential function* is

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

By the ratio test, this converges on all of \mathbb{C} .

A fundamental property of this function is that

$$e^{z+w} = e^z e^w.$$

Once we have this property, we can say that

Proposition. The derivative of e^z is e^z .

Proof.

$$\begin{aligned}\frac{e^{z+h} - e^z}{h} &= e^z \left(\frac{e^h - 1}{h} \right) \\ &= e^z \left(1 + \frac{h}{2!} + \frac{h^2}{3!} + \cdots \right)\end{aligned}$$

But

$$\left| \frac{h}{2!} + \frac{h^2}{3!} + \cdots \right| \leq \frac{|h|}{2} + \frac{|h|^2}{4} + \frac{|h|^3}{8} + \cdots = \frac{|h|/2}{1 - |h|/2} \rightarrow 0.$$

So

$$\frac{e^{z+h} - e^z}{h} = e^z.$$

□

But we must still prove that $e^{z+w} = e^z e^w$.

Consider two sequences $(a_n), (b_n)$. Their *convolution* is the sequence (c_n) defined by

$$c_n = a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \cdots + a_n b_0.$$

The relevance of this is that if you take

$$\left(\sum_{n=0}^N a_n z^n \right) \left(\sum_{n=0}^N b_n z^n \right) \text{ and } \sum_{n=0}^N c_n z^n,$$

and equate coefficients of z^n , you get

$$c_n = a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \cdots + a_n b_0.$$

Theorem. Let $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ be two absolutely convergent series, and let (c_n) be the convolution of the sequences (a_n) and (b_n) . Then $\sum_{n=0}^{\infty} c_n$ converges (absolutely), and

$$\sum_{n=0}^{\infty} c_n = \left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right).$$

Proof. We first show that a rearrangement of $\sum c_n$ converges absolutely. Hence it converges unconditionally, and we can rearrange it back to $\sum c_n$.

Consider the series

$$(a_0 b_0) + (a_0 b_1 + a_1 b_0) + (a_0 b_2 + a_1 b_1 + a_2 b_0) + \cdots \quad (*)$$

Let

$$S_N = \sum_{n=0}^N a_n, \quad T_N = \sum_{n=0}^N b_n, \quad U_N = \sum_{n=0}^N |a_n|, \quad V_N = \sum_{n=0}^N |b_n|.$$

Also let $S_N \rightarrow S, T_N \rightarrow T, U_N \rightarrow U, V_N \rightarrow V$ (these exist since $\sum a_n$ and $\sum b_n$ converge absolutely).

If we take the modulus of the terms of $(*)$, and consider the first $(N+1)^2$ terms (ie. the first $N+1$ brackets), the sum is $U_N V_N$. Hence the series converges absolutely to UV . Hence $(*)$ converges.

The partial sum up to $(N+1)^2$ of the series (*) itself is $S_N T_N$, which converges to ST . So the whole series converges to ST .

Since it converges absolutely, it converges unconditionally. Now consider a rearrangement:

$$a_0 b_0 + (a_0 b_1 + a_1 b_0) + (a_0 b_2 + a_1 b_1 + a_2 b_0) + \cdots$$

Then this converges to ST as well. But the partial sum of the first $1+2+\cdots+N$ terms is $c_0 + c_1 + \cdots + c_N$. So

$$\sum_{n=0}^N c_n \rightarrow ST = \left(\sum_{n=0}^{\infty} a_n \right) \left(\sum_{n=0}^{\infty} b_n \right).$$

□

Corollary.

$$e^z e^w = e^{z+w}.$$

Proof. By theorem above (and definition of e^z),

$$\begin{aligned} e^z e^w &= \sum_{n=0}^{\infty} \left(1 \cdot \frac{w^n}{n!} + \frac{z}{1!} \frac{w^{n-1}}{(n-1)!} + \frac{z^2}{2!} \frac{w^{n-2}}{(n-2)!} + \cdots + \frac{z^n}{n!} \cdot 1 \right) \\ e^z e^w &= \sum_{n=0}^{\infty} \frac{1}{n!} \left(w^n + \binom{n}{1} z w^{n-1} + \binom{n}{2} z^2 w^{n-2} + \cdots + \binom{n}{n} z^n \right) \\ &= \sum_{n=0}^{\infty} (z+w)^n \text{ by the binomial theorem} \\ &= e^{z+w}. \end{aligned}$$

□

Note: if (c_n) is the convolution of (a_n) and (b_n) , then the convolution of $(a_n z^n)$ and $(b_n z^n)$ is $(c_n z^n)$. Therefore if both $\sum a_n z^n$ and $\sum b_n z^n$ converge absolutely, then their product is $\sum c_n z^n$.

Note that we have now completed the proof that the derivative of e^z is e^z .

Now we define $\sin z$ and $\cos z$:

Definition (Sine and cosine).

$$\begin{aligned} \sin z &= \frac{e^{iz} - e^{-iz}}{2i} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots \\ \cos z &= \frac{e^{iz} + e^{-iz}}{2} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots \end{aligned}$$

We now prove certain basic properties of \sin and \cos , using known properties of e^z .

Proposition.

$$\begin{aligned} \frac{d}{dz} \sin z &= \frac{ie^{iz} + ie^{-iz}}{2i} = \cos z \\ \frac{d}{dz} \cos z &= \frac{ie^{iz} - ie^{-iz}}{2} = -\sin z \\ \sin^2 z + \cos^2 z &= \frac{e^{2iz} + 2 + e^{-2iz}}{4} + \frac{e^{2iz} - 2 + e^{-2iz}}{-4} = 1. \end{aligned}$$

It follows that if x is real, then $|\cos x|$ and $|\sin x|$ are at most 1.

Proposition.

$$\begin{aligned}\cos(z + w) &= \cos z \cos w - \sin z \sin w \\ \sin(z + w) &= \sin z \cos w + \cos z \sin w\end{aligned}$$

Proof.

$$\begin{aligned}\cos z \cos w - \sin z \sin w &= \frac{(e^{iz} + e^{-iz})(e^{iw} + e^{-iw})}{4} + \frac{(e^{iz} - e^{-iz})(e^{iw} - e^{-iw})}{4} \\ &= \frac{e^{i(z+w)} + e^{-i(z+w)}}{2} \\ &= \cos(z + w).\end{aligned}$$

Differentiating both sides wrt z gives

$$-\sin z \cos w - \cos z \sin w = -\sin(z + w).$$

So

$$\sin(z + w) = \sin z \cos w + \cos z \sin w.$$

□

When x is real, we know that $\cos x \leq 1$. Also $\sin 0 = 0$, and $\frac{d}{dx} \sin x = \cos x \leq 1$. So for $x \geq 0$, $\sin x \leq x$, “by the mean value theorem”. Also, $\cos 0 = 1$, and $\frac{d}{dx} \cos x = -\sin x$, which, for $x \geq 0$, is greater than $-x$. From this, it follows that when $x \geq 0$, $\cos x \geq 1 - \frac{x^2}{2}$ (the $1 - \frac{x^2}{2}$ comes from “integrating” $-x$, (or finding a thing whose derivative is $-x$)).

Continuing in this way, we get that for $x \geq 0$, if you take truncate the power series for $\sin x$ or $\cos x$, it will be $\geq \sin x, \cos x$ if you stop at a positive term, and \leq if you stop at a negative term. For example,

$$\sin x \geq x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!}.$$

In particular,

$$\cos 2 \leq 1 - \frac{2^2}{2!} + \frac{2^4}{4!} = 1 - 2 + \frac{2}{3} < 0.$$

Since $\cos 0 = 1$, it follows by the intermediate value theorem that there exists some $x \in (0, 2)$ such that $\cos x = 0$. Since $\cos x \geq 1 - \frac{x^2}{2}$, we can further deduce that $x > 1$.

Definition (Pi). Define the smallest x such that $\cos x = 0$ be $\frac{\pi}{2}$.

Since $\sin^2 z + \cos^2 z = 1$, it follows that $\sin \frac{\pi}{2} = \pm 1$. Since $\cos x > 0$ on $[0, \frac{\pi}{2}]$, $\sin \frac{\pi}{2} \geq 0$ by the mean value theorem. So $\sin \frac{\pi}{2} = 1$.

Thus

Proposition.

$$\begin{aligned}\cos\left(z + \frac{\pi}{2}\right) &= -\sin z \\ \sin\left(z + \frac{\pi}{2}\right) &= \cos z \\ \cos(z + \pi) &= -\cos z \\ \sin(z + \pi) &= -\sin z \\ \cos(z + 2\pi) &= \cos z \\ \sin(z + 2\pi) &= \sin z\end{aligned}$$

Proof.

$$\begin{aligned}\cos\left(z + \frac{\pi}{2}\right) &= \cos z \cos \frac{\pi}{2} - \sin z \sin \frac{\pi}{2} \\ &= -\sin z \sin \frac{\pi}{2} \\ &= -\sin z\end{aligned}$$

and similarly for others. \square

6.2 Differentiating power series

We shall show that inside the circle of convergence, the derivative of $\sum_{n=0}^{\infty} a_n z^n$ is given by the obvious formula $\sum_{n=1}^{\infty} n a_n z^{n-1}$.

We first prove some (seemingly arbitrary and random) lemmas to build up the proof of the above statement. They are done so that the final proof will not be full of tedious algebra.

Lemma. Let a and b be complex numbers. Then

$$b^n - a^n - n(n-a)a^{n-1} = (b-a)^2(b^{n-2} + 2ab^{n-3} + 3a^2b^{n-4} + \cdots + (n-1)a^{n-2}).$$

Proof. If $b = a$, we are done. Otherwise,

$$\frac{b^n - a^n}{b - a} = b^{n-1} + ab^{n-2} + a^2b^{n-3} + \cdots + a^{n-1}.$$

Differentiate both sides with respect to a . Then

$$\frac{-na^{n-1}(b-a) + b^n - a^n}{(b-a)^2} = b^{n-2} + 2ab^{n-3} + \cdots + (n-1)a^{n-2}.$$

Rearranging gives the result.

Alternatively, we can do

$$b^n - a^n = (b-a)(b^{n-1} + ab^{n-2} + \cdots + a^{n-1}).$$

Subtract $n(b-a)a^{n-1}$ to obtain

$$(b-a)[b^{n-1} - a^{n-1} + a(b^{n-2} - a^{n-2}) + a^2(b^{n-3} - a^{n-3}) + \cdots]$$

and simplify. \square

Note: this implies

$$(z+h)^n - z^n - nhz^{n-1} = h^2((z+h)^{n-2} + 2z(z+h)^{n-3} + \cdots + (n-1)z^{n-2}).$$

Lemma. Let $a_n z^n$ have radius of convergence R , and let $|z| < R$. Then $\sum n a_n z^{n-1}$ converges (absolutely).

Proof. Pick r such that $|z| < r < R$. Then $\sum |a_n| r^n$ converges, so the terms $|a_n| r^n$ are bounded above by, say, C . Now

$$\sum |a_n z^{n-1}| = \sum n |a_n| r^{n-1} \left(\frac{|z|}{r}\right)^{n-1} \leq \frac{C}{r} \sum n \left(\frac{|z|}{r}\right)^{n-1}$$

The series $\sum n \left(\frac{|z|}{r}\right)^{n-1}$ converges, by the ratio test. So $\sum n |a_n z^{n-1}|$ converges, by the comparison test. \square

Corollary. Under the same conditions,

$$\sum_{n=2}^{\infty} \binom{n}{2} a_n z^{n-2}$$

converges absolutely.

Proof. Apply Lemma above again and divide by 2. \square

Theorem. Let $\sum a_n z^n$ be a power series with radius of convergence R . For $|z| < R$, let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \text{ and } g(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

Then f is differentiable with derivative g .

Proof. We want $f(z+h) - f(z) - hg(z)$ to be $o(h)$. We have

$$f(z+h) - f(z) - hg(z) = \sum_{n=2}^{\infty} a_n ((z+h)^n - z^n - hn z^{n-1}).$$

We started summing from $n=2$ since the $n=0$ and $n=1$ terms are 0. Using our first lemma, we are left with

$$h^2 \sum_{n=2}^{\infty} a_n ((z+h)^{n-2} + 2z(z+h)^{n-3} + \cdots + (n-1)z^{n-2})$$

We want the huge infinite series to be bounded, and then the whole thing is a bounded thing times h^2 , which is definitely $o(h)$.

Pick r such that $|z| < r < R$. If h is small enough that $|z+h| \leq r$, then the last infinite series is bounded above (in modulus) by

$$\sum_{n=2}^{\infty} |a_n| (r^{n-2} + 2r^{n-2} + \cdots + (n-1)r^{n-2}) = \sum_{n=2}^{\infty} |a_n| \binom{n}{2} r^{n-2},$$

which is bounded. So done. \square

Example. The derivative of

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$

is

$$1 + z + \frac{z^2}{2!} + \cdots = e^z.$$

So we have another proof that of this fact.

Similarly, the derivatives of $\sin z$ and $\cos z$ work out as $\cos z$ and $-\sin z$.

6.3 Hyperbolic trigonometric functions

Definition (Hyperbolic sine and cosine). We define

$$\begin{aligned}\cosh z &= \frac{e^z + e^{-z}}{2} = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \frac{z^6}{6!} + \cdots \\ \sinh z &= \frac{e^z - e^{-z}}{2} = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \frac{z^7}{7!} + \cdots\end{aligned}$$

Either from the definition or from differentiating the power series, we get that

Proposition.

$$\begin{aligned}\frac{d}{dz} \cosh z &= \sinh z \\ \frac{d}{dz} \sinh z &= \cosh z\end{aligned}$$

Also, by definition, we have

Proposition.

$$\begin{aligned}\cosh iz &= \cos z \\ \sinh iz &= i \sin z\end{aligned}$$

Also,

Proposition.

$$\cosh^2 z - \sinh^2 z = 1,$$

7 The Riemann Integral

Definition (Dissections). Let $[a, b]$ be a closed interval. A *dissection* of $[a, b]$ is sequence $a = x_0 < x_1 < x_2 < \cdots < x_n = b$.

Definition (Upper and lower sums). Given a dissection \mathcal{D} , the *upper sum* and *lower sum* are defined by the formulae

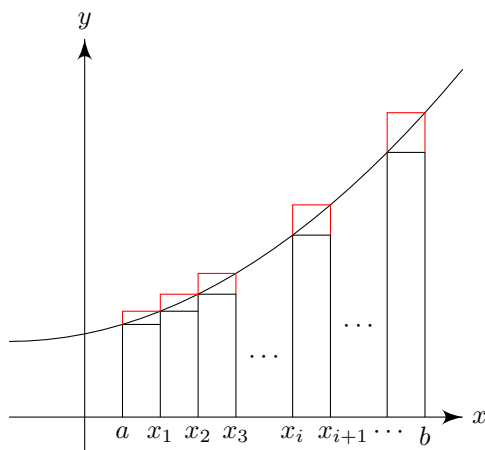
$$S_{\mathcal{D}}(f) = \sum_{i=1}^n (x_i - x_{i-1}) \sup_{x \in [x_{i-1}, x_i]} f(x)$$

$$s_{\mathcal{D}}(f) = \sum_{i=1}^n (x_i - x_{i-1}) \inf_{x \in [x_{i-1}, x_i]} f(x)$$

Sometimes we use the shorthand

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

The upper sum is the total area of the red rectangles, while the lower sum is the total area of the black rectangles:

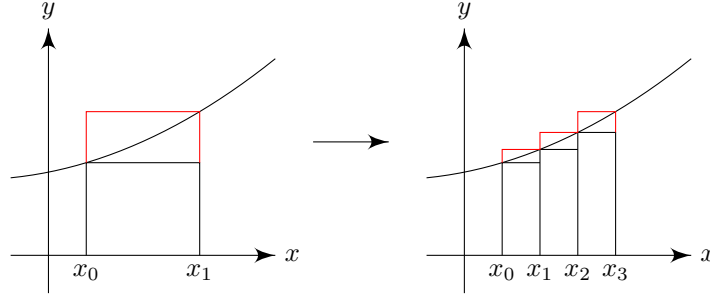


Definition (Refining dissections). If \mathcal{D}_1 and \mathcal{D}_2 are dissections of $[a, b]$, we say that \mathcal{D}_2 *refines* \mathcal{D}_1 if every point of \mathcal{D}_1 is a point of \mathcal{D}_2 .

Lemma. If \mathcal{D}_2 refines \mathcal{D}_1 , then

$$S_{\mathcal{D}_2} f \leq S_{\mathcal{D}_1} f \text{ and } s_{\mathcal{D}_2} f \geq s_{\mathcal{D}_1} f.$$

Using the picture above, this is because if we cut up the dissections into smaller pieces, the red rectangles can only get chopped into shorter pieces and the black rectangles can only get chopped into taller pieces.



Proof. Let \mathcal{D} be $x_0 < x_1 < \dots < x_n$. Let \mathcal{D}_2 be obtained from \mathcal{D}_1 by the addition of one point z . If $z \in (x_{i-1}, x_i)$, then

$$\begin{aligned} S_{\mathcal{D}_2}f - S_{\mathcal{D}_1}f &= \left[(z - x_{i-1}) \sup_{x \in [x_{i-1}, z]} f(x) \right] \\ &\quad + \left[(x_i - z) \sup_{x \in [z, x_i]} f(x) \right] - (x_i - x_{i-1})M_i. \end{aligned}$$

But $\sup_{x \in [x_{i-1}, z]} f(x)$ and $\sup_{x \in [z, x_i]} f(x)$ are both at most M_i . So this is at most $M_i(z - x_{i-1} + x_i - z - (x_i - x_{i-1})) = 0$. So

$$S_{\mathcal{D}_2}f \leq S_{\mathcal{D}_1}f.$$

By induction, the result is true whenever \mathcal{D}_2 refines \mathcal{D}_1 .

A very similar argument shows that $s_{\mathcal{D}_2}f \geq s_{\mathcal{D}_1}f$. \square

Definition (Least common refinement). If \mathcal{D}_1 and \mathcal{D}_2 be dissections of $[a, b]$. Then the least common refinement of \mathcal{D}_1 and \mathcal{D}_2 is the dissection made out of the points of \mathcal{D}_1 and \mathcal{D}_2 .

Corollary. Let \mathcal{D}_1 and \mathcal{D}_2 be two dissections of $[a, b]$. Then

$$S_{\mathcal{D}_1}f \geq s_{\mathcal{D}_2}f.$$

Proof. Let \mathcal{D} be the least common refinement (or indeed any common refinement). Then by lemma above (and by definition),

$$S_{\mathcal{D}_1}f \geq S_{\mathcal{D}}f \geq s_{\mathcal{D}} \geq s_{\mathcal{D}_2}f.$$

\square

Definition (Upper, lower, and Riemann integral). The *upper integral* is

$$\overline{\int_a^b} f(x) \, dx = \inf_{\mathcal{D}} S_{\mathcal{D}}f.$$

The *lower integral* is

$$\underline{\int_a^b} f(x) \, dx = \sup_{\mathcal{D}} s_{\mathcal{D}}f.$$

If these are equal, then we call their common value the *Riemann integral* of f , and is denoted $\int_a^b f(x) \, dx$.

If this exists, we say f is *Riemann integrable*.

We will later prove the *fundamental theorem of calculus*, which says that integration is the reverse of differentiation. But why don't we simply define integration as anti-differentiation, and prove that it is the area of the curve? There are things that we cannot find (a closed form of) the anti-derivative of, say e^{-x^2} . In these cases, we wouldn't want to say the integral doesn't exist - it surely does according to this definition!

Note: If f is unbounded above in $[a, b]$, then for any dissection \mathcal{D} , there must be some i such that f is unbounded on $[x_{i-1}, x_i]$. So $M_i = \infty$. So $S_{\mathcal{D}}f = \infty$.

Similarly, if f is unbounded below, then $s_{\mathcal{D}}f = -\infty$. So unbounded functions are not Riemann integrable.

Example. Let $f(x) = x$ on $[a, b]$. Intuitively, we know that the integral is $(b^2 - a^2)/2$, and we will show this using the definition above. Let $\mathcal{D} = x_0 < x_1 < \dots < x_n$ be a dissection. Then

$$S_{\mathcal{D}}f = \sum_{i=1}^n (x_i - x_{i-1})x_i$$

We *know* that the integral is $\frac{b^2 - a^2}{2}$. So we put each term of the sum into the form $\frac{x_i^2 - x_{i-1}^2}{2}$ plus some error terms.

$$\begin{aligned} &= \sum_{i=1}^n \left(\frac{x_i^2}{2} - \frac{x_{i-1}^2}{2} + \frac{x_i^2}{2} - x_{i-1}x_i + \frac{x_{i-1}^2}{2} \right) \\ &= \frac{1}{2} \sum_{i=1}^n (x_i^2 - x_{i-1}^2 + (x_i - x_{i-1})^2) \\ &= \frac{1}{2}(b^2 - a^2) + \frac{1}{2} \sum_{i=1}^n (x_i - x_{i-1})^2. \end{aligned}$$

Definition (Mesh). The *mesh* of a dissection \mathcal{D} is $\max_i (x_i - x_{i-1})$.

Then if the mesh is $< \delta$, then

$$\frac{1}{2} \sum_{i=1}^n (x_i - x_{i-1})^2 \leq \frac{\delta}{2} \sum_{i=1}^n (x_i - x_{i-1}) = \frac{\delta}{2}(b - a).$$

So by making δ small enough, we can show that for any $\varepsilon > 0$,

$$\overline{\int_a^b x \, dx} < \frac{1}{2}(b^2 - a^2) + \varepsilon.$$

Similarly,

$$\underline{\int_a^b x \, dx} > \frac{1}{2}(b^2 - a^2) - \varepsilon.$$

So

$$\int_a^b x \, dx = \frac{1}{2}(b^2 - a^2).$$

Example. Define $f : [0, 1] \rightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}.$$

Let $x_0 < x_1 < \dots, x_n$ be a dissection. Then for every i , we have $m_i = 0$ (since there is an irrational in every interval), and $M_i = 1$ (since there is a rational in every interval). So

$$S_{\mathcal{D}}f = \sum_{i=1}^n M_i(x_i - x_{i-1}) = \sum_{i=1}^n (x_i - x_{i-1}) = 1.$$

Similarly, $s_{\mathcal{D}}f = 0$. Since \mathcal{D} was arbitrary, we have

$$\overline{\int_0^1} f(x) \, dx = 1, \quad \underline{\int_0^1} f(x) \, dx = 0.$$

So f is *not* Riemann integrable.

Note: This function is not interesting at all. The whole point of its existence is to show undergraduates that there are some functions that are not integrable!

Note: It is important to say that it is not *Riemann* integrable. There are other notions for integration in which this function is integrable. For example, this function is *Lebesgue-integrable*.

Proposition (Riemann's integrability criterion). This is sometimes known as Cauchy's integrability criterion.

Let $f : [a, b] \rightarrow \mathbb{R}$. Then f is Riemann integrable if and only if for every $\varepsilon > 0$, there exists a dissection \mathcal{D} such that

$$S_{\mathcal{D}} - s_{\mathcal{D}} < \varepsilon.$$

This is rather trivial and expected result, but it is useful since usually it is easier to show this than to use the definition of integrability directly.

Proof. (\Rightarrow) Suppose that f is integrable. Then (by definition of Riemann integrability), there exist \mathcal{D}_1 and \mathcal{D}_2 such that

$$S_{\mathcal{D}_1} < \int_a^b f(x) \, dx + \frac{\varepsilon}{2},$$

and

$$s_{\mathcal{D}_2} > \int_a^b f(x) \, dx - \frac{\varepsilon}{2}.$$

Let \mathcal{D} be a common refinement of \mathcal{D}_1 and \mathcal{D}_2 . Then

$$S_{\mathcal{D}}f - s_{\mathcal{D}}f \leq S_{\mathcal{D}_1}f - s_{\mathcal{D}_2}f < \varepsilon.$$

(\Leftarrow) Conversely, if there exists \mathcal{D} such that

$$S_{\mathcal{D}}f - s_{\mathcal{D}}f < \varepsilon,$$

then

$$\inf S_{\mathcal{D}}f - \sup s_{\mathcal{D}}f < \varepsilon,$$

which is, by definition, that

$$\overline{\int_a^b} f(x) \, dx - \underline{\int_a^b} f(x) \, dx < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this gives us that

$$\overline{\int_a^b} f(x) \, dx = \underline{\int_a^b} f(x) \, dx.$$

So f is integrable. □

We want to prove that integration is linear, ie

$$\int_a^b (\lambda f(x) + \mu g(x)) \, dx = \lambda \int_a^b f(x) \, dx + \mu \int_a^b g(x) \, dx.$$

We do this step by step:

Proposition. Let $f : [a, b] \rightarrow \mathbb{R}$ be integrable, and $\lambda \geq 0$. Then λf is integrable, and

$$\int_a^b \lambda f(x) \, dx = \lambda \int_a^b f(x) \, dx.$$

Proof. Let \mathcal{D} be a dissection of $[a, b]$. Since

$$\sup_{x \in [x_{i-1}, x_i]} \lambda f(x) = \lambda \sup_{x \in [x_{i-1}, x_i]} f(x),$$

and similarly for inf, we have

$$\begin{aligned} S_{\mathcal{D}}(\lambda f) &= \lambda S_{\mathcal{D}}f \\ s_{\mathcal{D}}(\lambda f) &= \lambda s_{\mathcal{D}}f. \end{aligned}$$

So if we choose \mathcal{D} such that $S_{\mathcal{D}}f - s_{\mathcal{D}}f < \varepsilon/\lambda$, then $S_{\mathcal{D}}(\lambda f) - s_{\mathcal{D}}(\lambda f) < \varepsilon$. So the result follows from Riemann's integrability criterion. □

Proposition. Let $f : [a, b] \rightarrow \mathbb{R}$ be integrable. Then $-f$ is integrable, and

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

Proof. Let \mathcal{D} be a dissection. Then

$$\begin{aligned} \sup_{x \in [x_{i-1}, x_i]} -f(x) &= - \inf_{x \in [x_{i-1}, x_i]} f(x) \\ \inf_{x \in [x_{i-1}, x_i]} -f(x) &= - \sup_{x \in [x_{i-1}, x_i]} f(x). \end{aligned}$$

Therefore

$$S_{\mathcal{D}}(-f) = \sum_{i=1}^n (x_i - x_{i-1})(-m_i) = -s_{\mathcal{D}}(f).$$

Similarly,

$$s_{\mathcal{D}}(-f) = -S_{\mathcal{D}}f.$$

So

$$S_{\mathcal{D}}(-f) - s_{\mathcal{D}}(-f) = S_{\mathcal{D}}f - s_{\mathcal{D}}f.$$

Hence if f is integrable, then $-f$ is integrable by the Riemann integrability criterion. \square

Proposition. Let $f, g : [a, b] \rightarrow \mathbb{R}$ be integrable. Then $f + g$ is integrable, and

$$\int_a^b (f(x) + g(x)) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx.$$

Proof. Let \mathcal{D} be a dissection. Then

$$\begin{aligned} S_{\mathcal{D}}(f + g) &= \sum_{i=1}^n (x_i - x_{i-1}) \sup_{x \in [x_{i-1}, x_i]} (f(x) + g(x)) \\ &\leq \sum_{i=1}^n (x_i - x_{i-1}) \left(\sup_{u \in [x_{i-1}, x_i]} f(u) + \sup_{v \in [x_{i-1}, x_i]} g(v) \right) \\ &= S_{\mathcal{D}}f + S_{\mathcal{D}}g \end{aligned}$$

Therefore,

$$\overline{\int_a^b} (f(x) + g(x)) \, dx \leq \overline{\int_a^b} f(x) \, dx + \overline{\int_a^b} g(x) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx.$$

Similarly,

$$\underline{\int_a^b} (f(x) + g(x)) \, dx \geq \underline{\int_a^b} f(x) \, dx + \underline{\int_a^b} g(x) \, dx.$$

So the upper and lower integrals are equal, and the result follows. \square

So we now have that

$$\int_a^b (\lambda f(x) + \mu g(x)) \, dx = \lambda \int_a^b f(x) \, dx + \mu \int_a^b g(x) \, dx.$$

Proposition. Let $f, g : [a, b] \rightarrow \mathbb{R}$ be integrable, and suppose that $f(x) \leq g(x)$ for every x . Then

$$\int_a^b f(x) \, dx \leq \int_a^b g(x) \, dx.$$

Proof. Follows immediately from the definition. \square

Proposition. Let $f : [a, b] \rightarrow \mathbb{R}$ be integrable. Then $|f|$ is integrable.

Proof. Note that we can write

$$\sup_{x \in [x_{i-1}, x_i]} f(x) - \inf_{x \in [x_{i-1}, x_i]} f(x) = \sup_{u, v \in [x_{i-1}, x_i]} |f(u) - f(v)|.$$

Similarly,

$$\sup_{x \in [x_{i-1}, x_i]} |f(x)| - \inf_{x \in [x_{i-1}, x_i]} |f(x)| = \sup_{u, v \in [x_{i-1}, x_i]} ||f(u)| - |f(v)||.$$

For any pair of real numbers, x, y , we have that $||x| - |y|| \leq |x - y|$ by the triangle inequality. Then for any interval $u, v \in [x_{i-1}, x_i]$, we have

$$||f(u)| - |f(v)|| \leq |f(u) - f(v)|.$$

Hence we have

$$\sup_{x \in [x_{i-1}, x_i]} |f(x)| - \inf_{x \in [x_{i-1}, x_i]} |f(x)| \leq \sup_{x \in [x_{i-1}, x_i]} f(x) - \inf_{x \in [x_{i-1}, x_i]} f(x).$$

So for any dissection \mathcal{D} , we have

$$S_{\mathcal{D}}(|f|) - s_{\mathcal{D}}(|f|) \leq S_{\mathcal{D}}(f) - S_{\mathcal{D}}(f).$$

So the result follows from Riemann's integrability criterion. \square

Note: Combining these two propositions, we get that if

$$|f(x) - g(x)| \leq C,$$

for every $x \in [a, b]$, then

$$\left| \int_a^b f(x) \, dx - \int_a^b g(x) \, dx \right| \leq C(b - a).$$

Proposition (Additivity property). Let $f : [a, c] \rightarrow \mathbb{R}$ be integrable, and let $b \in (a, c)$. Then the restrictions of f to $[a, b]$ and $[b, c]$ are Riemann integrable, and

$$\int_a^b f(x) \, dx + \int_b^c f(x) \, dx = \int_a^c f(x) \, dx$$

Similarly, if f is integrable on $[a, b]$ and $[a, c]$, then it is integrable on $[a, c]$ and the above equation also holds.

Proof. Let $\varepsilon > 0$, and let $a = x_0 < x_1 < \dots < x_n = c$ be a dissection of \mathcal{D} of $[a, c]$ such that

$$S_{\mathcal{D}}(f) \leq \int_a^c f(x) \, dx + \varepsilon,$$

and

$$s_{\mathcal{D}}(f) \geq \int_a^c f(x) \, dx - \varepsilon.$$

Let \mathcal{D}' be the dissection made of \mathcal{D} plus the point b . Let \mathcal{D}_1 be the dissection of $[a, b]$ made of points of \mathcal{D}' from a to b , and \mathcal{D}_2 be the dissection of $[b, c]$ made of points of \mathcal{D}' from b to c . Then

$$S_{\mathcal{D}_1}(f) + S_{\mathcal{D}_2}(f) = S_{\mathcal{D}'}(f) \leq S_{\mathcal{D}}(f),$$

and

$$s_{\mathcal{D}_1}(f) + s_{\mathcal{D}_2}(f) = s_{\mathcal{D}'}(f) \geq s_{\mathcal{D}}(f).$$

Since $S_{\mathcal{D}}(f) - s_{\mathcal{D}}(f) < 2\varepsilon$, and both $S_{\mathcal{D}_2}(f) - s_{\mathcal{D}_2}(f)$ and $S_{\mathcal{D}_1}(f) - s_{\mathcal{D}_1}(f)$ are greater than or equal to 0, we have $S_{\mathcal{D}_1}(f) - s_{\mathcal{D}_1}(f)$ and $S_{\mathcal{D}_2}(f) - s_{\mathcal{D}_2}(f)$ are

less than 2ε . Since ε is arbitrary, it follows that the restrictions of f to $[a, b]$ and $[b, c]$ are both Riemann integrable. Furthermore,

$$\int_a^b f(x) \, dx + \int_b^c f(x) \, dx \leq S_{\mathcal{D}_1}(f) + S_{\mathcal{D}_2}(f) = S_{\mathcal{D}'}(f) \leq S_{\mathcal{D}}(f) \leq \int_a^c f(x) \, dx + \varepsilon.$$

Similarly,

$$\int_a^b f(x) \, dx + \int_b^c f(x) \, dx \geq s_{\mathcal{D}_1}(f) + s_{\mathcal{D}_2}(f) = s_{\mathcal{D}'}(f) \geq s_{\mathcal{D}}(f) \geq \int_a^c f(x) \, dx - \varepsilon.$$

Since ε is arbitrary, it follows that

$$\int_a^b f(x) \, dx + \int_b^c f(x) \, dx = \int_a^c f(x) \, dx.$$

The other direction is left as an (easy) exercise. \square

Definition (Uniform continuity). Let $A \subseteq \mathbb{R}$ and let $f : A \rightarrow \mathbb{R}$. Then f is *uniformly continuous* if

$$\forall \varepsilon \exists \delta > 0 \forall x \forall y |x - y| < \delta \Rightarrow |f(x) - f(y)| \leq \varepsilon.$$

Note that this is DIFFERENT from normal continuity. Normal continuity says that at every point x , I can find a δ that works for this point. Uniform continuity says that I can find a δ that works for *any* x .

Example. The function $f(x) = x^2$ is *not* uniformly continuous in \mathbb{R} . We want to show

$$\exists \varepsilon > 0 \forall \delta > 0 \exists x \exists y |x - y| < \delta \text{ and } |f(x) - f(y)| \geq \varepsilon.$$

Pick $\varepsilon = 1$. For any $\delta > 0$, let $x = \frac{1}{\delta}$ and $y = \frac{1}{\delta} + \frac{\delta}{2}$. Then

$$y^2 - x^2 = 1 + \frac{\delta^2}{4} \geq 1.$$

Other continuous but not uniformly continuous functions include $f : (0, \infty) \rightarrow \mathbb{R}$ defined by $x \mapsto \frac{1}{x}$ and $\sin x^2$.

Theorem. Let $a < b$ and let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is uniformly continuous.

Proof. Suppose that f is not uniformly continuous. Then

$$\exists \varepsilon > 0 \forall \delta > 0 \exists x, y : |x - y| \leq \delta \text{ and } |f(x) - f(y)| \geq \varepsilon.$$

Therefore, we can find sequences $(x_n), (y_n)$ such that for every n , we have

$$|x_n - y_n| \leq \frac{1}{n} \text{ and } |f(x_n) - f(y_n)| \geq \varepsilon.$$

Then by Bolzano-Weierstrass theorem, we can find a subsequence (x_{n_k}) converging to some x . Since $|x_{n_k} - y_{n_k}| \leq \frac{1}{n_k}$, $y_{n_k} \rightarrow x$ as well. But $|f(x_{n_k}) - f(y_{n_k})| \geq \varepsilon$ for every k . So $f(x_{n_k})$ and $f(y_{n_k})$ cannot both converge to the same limit. So f is not continuous at x . \square

Theorem. Every continuous function f on a closed bounded interval $[a, b]$ is Riemann integrable.

Proof. Let $\varepsilon > 0$. We want to construct a dissection of $[a, b]$ such that the difference between the supremum and infimum of each interval is small, so that the difference between $s_{\mathcal{D}}$ and $S_{\mathcal{D}}$ is small. So we want f to vary very little within each interval. We can do this by uniform continuity.

Let $\varepsilon > 0$. Since f is continuous, it is uniformly continuous. So we can find $\delta > 0$ such that

$$|f(y) - f(x)| < \frac{\varepsilon}{b-a}$$

whenever $|y - x| < \delta$. Let \mathcal{D} be a dissection of $[a, b]$ of mesh less than δ (ie. the width of each interval is less than δ). Then

$$S_{\mathcal{D}}f - s_{\mathcal{D}}f = \sum_{i=1}^n (x_i - x_{i-1})(M_i - m_i).$$

Since for each i , we have $x_i - x_{i-1} < \delta$, we have $M_i - m_i < \frac{\varepsilon}{b-a}$ for each i . So

$$S_{\mathcal{D}}f - s_{\mathcal{D}}f < \frac{\varepsilon}{b-a} \sum_{i=1}^n (x_i - x_{i-1}) = \varepsilon.$$

So we are done by Riemann's integrability criterion. \square

With this result, we know that a lot of things are integrable, eg. e^{-x^2} .

Theorem. Let $f : [a, b] \rightarrow \mathbb{R}$ be monotone. Then f is Riemann integrable.

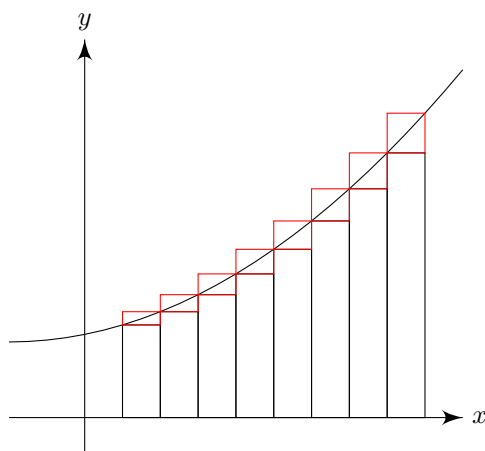
Note that monotone functions need not be "nice". It can even have infinitely many discontinuities. For example, if $f : [0, 1] \rightarrow \mathbb{R}$ maps x to the $1/(\text{first non-zero digit in the binary expansion of } x)$, with $f(0) = 0$.

Proof. let $\varepsilon > 0$. Let \mathcal{D} be a dissection of mesh less than $\frac{\varepsilon}{f(b)-f(a)}$. Then

$$\begin{aligned} S_{\mathcal{D}}f - s_{\mathcal{D}}f &= \sum_{i=1}^n (x_i - x_{i-1})(f(x_i) - f(x_{i-1})) \\ &\leq \frac{\varepsilon}{f(b) - f(a)} \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \\ &= \varepsilon. \end{aligned}$$

\square

Pictorially, we see that the difference between the upper and lower sums is total the area of the red rectangles.



To calculate the total area, we can stack the red areas together to get something of width $\frac{\varepsilon}{f(b)-f(a)}$ and height $f(b) - f(a)$. So the total area is just ε .

Lemma. Let $a < b$ and let f be a bounded function from $[a, b] \rightarrow \mathbb{R}$ that is continuous on (a, b) . Then f is integrable.

An example where this would apply is $\int_0^1 \sin \frac{1}{x}$. It gets nasty near $x = 0$, but its “nastiness” is confined to $x = 0$ only. So as long as its nastiness is sufficiently contained, it would still be integrable.

The idea of the proof is to integrate from a point x_1 very near a up to a point x_{n-1} very close to b . Since f is bounded, the regions $[a, x_1]$ and $[x_{n-1}, b]$ is small enough to not cause trouble.

Proof. Let $\varepsilon > 0$. Suppose that $|f(x)| \leq C$ for every $x \in [a, b]$. Let $x_0 = a$ and pick x_1 such that $x_1 - x_0 < \frac{\varepsilon}{8C}$. Also choose z between x_1 and b such that $b - z < \frac{\varepsilon}{8C}$.

Then f is continuous $[x_1, z]$. Therefore it is integrable on $[x_1, z]$. So we can find a dissection \mathcal{D}' with points $x_1 < x_2 < \dots < x_{n-1} = z$ such that

$$S_{\mathcal{D}'} f - s_{\mathcal{D}'} f < \frac{\varepsilon}{2}.$$

Let \mathcal{D} be the dissection $a = x_0 < x_1 < \dots < x_n = b$. Then

$$S_{\mathcal{D}} f - s_{\mathcal{D}} f < \frac{\varepsilon}{8C} \cdot 2C + \frac{\varepsilon}{2} + \frac{\varepsilon}{8C} \cdot 2C = \varepsilon.$$

So done by Riemann integrability criterion. \square

Example.

$$- f(x) = \begin{cases} \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases} \text{ defined on } [-1, 1] \text{ is integrable.}$$

$$- g(x) = \begin{cases} x & x \leq 1 \\ x^2 + 1 & x > 1 \end{cases} \text{ defined on } [0, 1] \text{ is integrable.}$$

Corollary. Every piecewise continuous and bounded function on $[a, b]$ is integrable.

Proof. Partition $[a, b]$ into intervals I_1, \dots, I_k , on each of which f is (bounded and) continuous. Hence for every I_j with end points x_{j-1}, x_j , f is integrable on $[x_{j-1}, x_j]$ (which may not equal I_j , eg. I_j could be $[x_{j-1}, x_j)$). But then by the additivity property of integration, we get that f is integrable on $[a, b]$ \square

Lemma. Let $f : [a, b] \rightarrow \mathbb{R}$ be Riemann integrable, and for each n , let \mathcal{D}_n be the dissection $a = x_0 < x_1 < \dots < x_n = b$, where $x_i = a + \frac{i(b-a)}{n}$ for each i . Then

$$S_{\mathcal{D}_n} f \rightarrow \int_a^b f(x) \, dx$$

and

$$s_{\mathcal{D}_n} f \rightarrow \int_a^b f(x) \, dx.$$

This shows that we needn't have defined Riemann integration with arbitrary dissections, but only these nice, uniform dissections. However, we opt to define it the general way so that we can easily talk about things like least common refinements.

Proof. Let $\varepsilon > 0$. We need to find an N . The only thing we know is that f is Riemann integrable, so we use it:

Since f is integrable, there is a dissection \mathcal{D} , say $u_0 < u_1 < \dots < u_m$, such that

$$S_{\mathcal{D}} f - \int_a^b f(x) \, dx < \frac{\varepsilon}{2}.$$

We also know that f is bounded. Let C be such that $|f(x)| \leq C$.

Now if we have an arbitrary \mathcal{D}_n , we can make it look like a refinement of \mathcal{D} , except that

For any n , let \mathcal{D}' be the least common refinement of \mathcal{D}_n and \mathcal{D} . Then

$$S_{\mathcal{D}'} f \leq S_{\mathcal{D}} f.$$

Also, the sums $S_{\mathcal{D}_n} f$ and $S_{\mathcal{D}'} f$ are the same, except that at most m of the subintervals $[x_{i-1}, x_i]$ are subdivided in \mathcal{D}' .

For each interval that gets chopped up, the upper sum decreases by at most $\frac{n-a}{n} 2C$. Therefore

$$S_{\mathcal{D}_n} f - S_{\mathcal{D}'} f \leq \frac{b-a}{n} 2C \cdot m.$$

Pick n such that $2Cm(b-a)/n < \frac{\varepsilon}{2}$. Then

$$S_{\mathcal{D}_n} f - S_{\mathcal{D}} f < \frac{\varepsilon}{2}.$$

So

$$S_{\mathcal{D}_n} f - \int_a^b f(x) \, dx < \varepsilon.$$

This is true whenever $n > \frac{4C(b-a)m}{\varepsilon}$. Since we also have $S_{\mathcal{D}_n} f \geq \int_a^b f(x) \, dx$, therefore

$$S_{\mathcal{D}_n} f \rightarrow \int_a^b f(x) \, dx.$$

The proof for lower sums is similar. \square

For convenience, we define the following:

Notation. If $b > a$, we define

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

We now prove that the fundamental theorem of calculus, which says that integration is the reverse of differentiation.

Theorem (Fundamental theorem of calculus, part 1). Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous, and for $x \in [a, b]$, define

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

Then F is differentiable and $F'(x) = f(x)$ for every x .

Proof.

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

Let $\varepsilon > 0$. Since f is continuous, at x , then there exists δ such that $|y - x| < \delta$ implies $|f(y) - f(x)| < \varepsilon$.

If $|h| < \delta$, then

$$\begin{aligned} \left| \frac{1}{h} \int_x^{x+h} f(t) \, dt - f(x) \right| &= \left| \frac{1}{h} \int_x^{x+h} (f(t) - f(x)) \, dt \right| \\ &\leq \frac{1}{|h|} \left| \int_x^{x+h} |f(t) - f(x)| \, dt \right| \\ &\leq \frac{\varepsilon|h|}{|h|} \\ &= \varepsilon. \end{aligned}$$

□

Corollary. If f is continuously differential on $[a, b]$, then

$$\int_a^b f'(t) \, dt = f(b) - f(a).$$

Proof. Let

$$g(x) = \int_a^x f'(t) \, dt.$$

Then

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

Since $g'(x) - f'(x) = 0$, $g(x) - f(x)$ must be a constant function by the mean value theorem. We also know that

$$g(a) = 0 = f(a) - f(a)$$

So we must have $g(x) = f(x) - f(a)$ for every x , and in particular, for $x = b$. □

Theorem (Fundamental theorem of calculus, part 2). Let $F : [a, b] \rightarrow \mathbb{R}$ be a differentiable function with derivative f , and suppose that f is integrable. Then

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

Note that this is a stronger result than the corollary above, since it does not require that f is continuous.

Proof. Let \mathcal{D} be a dissection $x_0 < x_1 < \cdots < x_n$. We want to make use of this dissection. So write

$$F(b) - F(a) = \sum_{i=1}^n (F(x_i) - F(x_{i-1})).$$

For each i , there exists $u_i \in (x_{i-1}, x_i)$ such that $F(x_i) - F(x_{i-1}) = (x_i - x_{i-1})f(u_i)$ by the mean value theorem. So

$$F(b) - F(a) = \sum_{i=1}^n (x_i - x_{i-1})f(u_i).$$

We know that $f(u_i)$ is somewhere between $\sup_{x \in [x_i, x_{i-1}]} f(x)$ and $\inf_{x \in [x_i, x_{i-1}]} f(x)$ by definition. Therefore

$$s_{\mathcal{D}}f \leq F(b) - F(a) \leq S_{\mathcal{D}}f.$$

Since f is integrable and \mathcal{D} was arbitrary, $s_{\mathcal{D}}$ and $S_{\mathcal{D}}f$ can both get arbitrarily close to $\int_a^b f(t) \, dt$. So

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

□

Proposition. Let $f, g : [a, b] \rightarrow \mathbb{R}$ be integrable. Then fg is integrable.

Proof. Let C be such that $|f(x)|, |g(x)| \leq C$ for every $x \in [a, b]$. Write L_i and ℓ_i for the sup and inf of g in $[x_{i-1}, x_i]$. Now let \mathcal{D} be a dissection, and for each i , let u_i and v_i be two points in $[x_{i-1}, x_i]$.

We will pretend that u_i and v_i are the maximum and minimum when we write the proof, but we cannot assert that they are, since fg need not have maxima and minima. We will then note that since our results hold for arbitrary u_i and v_i , it must hold when fg is at its supremum and infimum.

We find what we pretend is the difference between the upper and lower sum:

$$\begin{aligned} & \left| \sum_{i=1}^n (x_i - x_{i-1})(f(v_i)g(v_i) - f(u_i)g(u_i)) \right| \\ &= \left| \sum_{i=1}^n (x_i - x_{i-1})(f(v_i)(g(v_i) - g(u_i)) + (f(v_i) - f(u_i))g(u_i)) \right| \\ &\leq \sum_{i=1}^n (C(L_i - \ell_i) + (M_i - m_i)C) \\ &= C(S_{\mathcal{D}}g - s_{\mathcal{D}}g + S_{\mathcal{D}}f - s_{\mathcal{D}}f). \end{aligned}$$

Since u_i and v_i are arbitrary, it follows that

$$S_{\mathcal{D}}(fg) - s_{\mathcal{D}}(fg) \leq C(S_{\mathcal{D}}f - s_{\mathcal{D}}f + S_{\mathcal{D}}g - s_{\mathcal{D}}g).$$

Since C is fixed, and we can get $S_{\mathcal{D}}f - s_{\mathcal{D}}f$ and $S_{\mathcal{D}}g - s_{\mathcal{D}}g$ arbitrary small (since f and g are integrable), we can get $S_{\mathcal{D}}(fg) - s_{\mathcal{D}}(fg)$ arbitrarily small. So \square

Theorem (Integration by parts). Let $f, g : [a, b] \rightarrow \mathbb{R}$ be integrable such that f' and g' exist. Then

$$\int_a^b f(x)g'(x) \, dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x) \, dx.$$

Proof. Note that $f(x)g'(x) + f'(x)g(x)$ is integrable by sum/product rules. Also,

$$\int_a^b (f(x)g'(x) + f'(x)g(x)) \, dx = \int_a^b (fg)'(x) \, dx = f(b)g(b) - f(a)g(a)$$

by the fundamental theorem of calculus. \square

Theorem (Taylor's theorem with the integral form of the remainder). Let f be $n + 1$ times differentiable on $[a, b]$ with $f^{(n+1)}$ continuous. Then

$$\begin{aligned} f(b) &= f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!}f^{(2)}(a) + \cdots \\ &\quad + \frac{(b-a)^n}{n!}f^{(n)}(a) + \int_a^b \frac{(b-t)^n}{n!}f^{(n+1)}(t) \, dt. \end{aligned}$$

Proof. Induction on n .

When $n = 0$, the theorem says

$$f(b) - f(a) = \int_a^b (b-a)f'(t) \, dt.$$

which is true by the fundamental theorem of calculus.

Now observe that

$$\begin{aligned} \int_a^b \frac{(b-t)^n}{n!}f^{(n+1)}(t) \, dt &= \left[\frac{-(b-t)^{n+1}}{(n+1)!}f^{(n+1)}(t) \right]_a^b \\ &\quad + \int_a^b \frac{(b-t)^{n+1}}{(n+1)!}f^{(n+1)}(t) \, dt \\ &= \frac{(b-a)^{n+1}}{(n+1)!} + \int_a^b \frac{(b-t)^{n-1}}{(n-1)!}f^{(n+2)}(t) \, dt. \end{aligned}$$

So done. \square

Note that the form of the integral remainder is rather weird and unexpected. How could we have come up with it? We might start with the fundamental

theorem of algebra and integrate by parts. The first attempt would be to integrate 1 to t and differentiate $f'(t)$ to $f^{(2)}(t)$. So we have

$$\begin{aligned} f(b) &= f(a) + \int_a^b f'(t) \, dt \\ &= f(a) + [tf'(t)]_a^b - \int_a^b tf^{(2)}(t) \, dt \\ &= f(a) + bf'(b) - af'(a) - \int_a^b tf^{(2)}(t) \, dt \end{aligned}$$

We want something in the form $(b-a)f'(a)$, so we take that out and see what we are left with.

$$= f(a) + (b-a)f'(a) + b(f'(b) - f'(a)) - \int_a^b tf^{(2)}(t) \, dt$$

Then we note that $f'(b) - f'(a) = \int_a^b f^{(2)}(t) \, dt$. So we have

$$= f(a) + (b-a)f'(a) + \int_a^b (b-t)f^{(2)}(t) \, dt.$$

Then we can see that the right thing to integrate is $(b-t)$ and continue to obtain the result.

Theorem (Integration by substitution). Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Let $g : [u, v] \rightarrow \mathbb{R}$ be continuously differentiable, and suppose that $g(u) = a, g(v) = b$, and f is defined everywhere on $g([u, v])$ (and still continuous). Then

$$\int_a^b f(x) \, dx = \int_u^v f(g(t))g'(t) \, dt.$$

Proof. By the fundamental theorem of calculus, f has an anti-derivative F defined on $g([u, v])$. Then

$$\begin{aligned} \int_u^v f(g(t))g'(t) \, dt &= \int_u^v F'(g(t))g'(t) \, dt \\ &= \int_u^v (F \circ g)'(t) \, dt \\ &= F \circ g(v) - F \circ g(u) \\ &= F(b) - F(a) \\ &= \int_a^b f(x) \, dx. \end{aligned}$$

□

We can think of “integration by parts” as what you get by integrating the product rule, and “integration by substitution” as what you get by integrating the chain rule.

7.1 Improper integrals

It is sometimes sensible to talk about integrals of unbounded functions or integrating to infinity. But we have to be careful and write things down nicely.

Definition (Improper integral). Suppose that we have a function $f : [a, b] \rightarrow \mathbb{R}$ such that, for every $\varepsilon > 0$, f is integrable on $[a + \varepsilon, b]$ and $\lim_{\varepsilon \rightarrow 0} \int_{a+\varepsilon}^b f(x) \, dx$ exists. Then we define the improper integral

$$\int_a^b f(x) \, dx \text{ to be } \lim_{\varepsilon \rightarrow 0} \int_{a+\varepsilon}^b f(x) \, dx.$$

even if the Riemann integral does not exist.

We can do similarly for $[a, b - \varepsilon]$, or integral to infinity:

$$\int_a^\infty f(x) \, dx = \lim_{b \rightarrow \infty} \int_a^b f(x) \, dx.$$

when it exists.

Example.

$$\int_\varepsilon^1 x^{-1/2} \, dx = \left[2x^{-1/2} \right]_\varepsilon^1 = 2 - 2\varepsilon^{1/2} \rightarrow 2.$$

So

$$\int_0^1 x^{-1/2} \, dx = 2,$$

even though $x^{-1/2}$ is unbounded on $[0, 1]$.

Note that officially we are required to make $f(x) = x^{-1/2}$ a function with domain $[0, 1]$. So we can assign $f(0) = \pi$, or any number, since it doesn't matter.

Example.

$$\int_1^x \frac{1}{t^2} \, dt = \left[-\frac{1}{t} \right]_1^x = 1 - \frac{1}{x} \rightarrow 1 \text{ as } x \rightarrow \infty$$

by the fundamental theorem of calculus. So

$$\int_1^\infty \frac{1}{x^2} \, dx = 1.$$

Theorem (Integral test). Let $f : [1, \infty] \rightarrow \mathbb{R}$ be a decreasing non-negative function. Then $\sum_{n=1}^\infty f(n)$ converges iff $\int_1^\infty f(x) \, dx < \infty$.

Proof. We have

$$\int_n^{n+1} f(x) \, dx \leq f(n) \leq \int_{n-1}^n f(x) \, dx,$$

since f is decreasing (the right hand inequality is valid only for $n \geq 2$). It follows that

$$\int_1^{N+1} f(x) \, dx \leq \sum_{n=1}^N f(n) \leq \int_1^N f(x) \, dx + f(1)$$

So if the integral exists, then $\sum f(n)$ is increasing and bounded above by $\int_1^\infty f(x) \, dx$, so converges.

If the integral does not exist, then $\int_1^N f(x) \, dx$ is unbounded. Then $\sum_{n=1}^N f(n)$ is unbounded, hence does not converge. \square

Example. Since $\int_1^x \frac{1}{t^2} \, dt < \infty$, it follows that $\sum_{n=1}^\infty \frac{1}{n^2}$ converges.