Transcribed and edited by WOO

# Part IA
## Cambridge Mathematical Tripos

Lee, Junwoo

Trinity College, Cambridge

2024 Summer

# Contents

# Preface

These are what I studied during the 2024 long vacation for transfer into part IB mathematical tripos. I guess nobody can tell if it was valuable, or if this was a right decision. Nevertheless, I do not want to study beam bending. Since it is more mathematics-oriented document, I will use the usual conventions, such as definition, lemma, etc.

# CHAPTER 1

# Numbers and Sets

Lecture given by Professor Julia Wolf, Michaelmas Term 2023.

## 1.1 Elementary Number Theory

### 1.1.1 Natural Number

Intuitively speaking, the natural numbers consists of

$$1, 1+1, 1+1+1, \ldots$$

But we cannot be sure if the list above contains all of the natural numbers, or if there are no duplicate numbes. Hence, for more rigorous understanding, we assume the natural numbers, $\mathbb{N}$, is a set containing "1", with an operation "+" satisfying

(1) $\forall n \in \mathbb{N}$, $n + 1 \neq 1$.

(2) $\forall m, n \in \mathbb{N}$, if $m \neq n$, then $m + 1 \neq n + 1$.

(3) For any property $P(n)$, if $P(1)$ is true and $\forall n \in \mathbb{N}$, $P(n) \Rightarrow P(n+1)$, then $P(n)$ is true for all natural numbers.

(1), (2), and (3) are known as the *Peano axioms*, and (3) is called the *induction axiom*. (1) and (2) capture the idea that any two natural numbers are distinct; (3) captures our intuition that the list is complete.[1] Now we can write 2 for $1+1$, 3 for $1+1+1$ etc. and we can define "$+k$" for any natural number $k$:

**Definition 1.1.1.** For every natural number $n$,

$$n + (k+1) = (n+k) + 1$$

It can be thought from induction, taking $P(k) =$ "$+k$ is defined". Similarly, we can define multiplication, powers etc. satisfying

(1) $\forall a, b$, $a + b = b + a$ ($+$ is commutative).

(2) $\forall a, b$, $ab = ba$ ($\cdot$ is commutative).

(3) $\forall a, b, c$, $a + (b + c) = (a + b) + c$ ($+$ is associative).

(4) $\forall a, b, c$, $a(bc) = (ab)c$ ($\cdot$ is associative).

[1] Take $P(n) =$ "$n$ is on this list".

(5) $\forall a, b, c,\ a(b+c) = ab + ac$ (multiplication is distributive over addition).

We define "$a < b$" if $a + c = b$ for some $c \in \mathbb{N}$, satisfying

(6) $\forall a, b, c,\ a < b \Rightarrow a + c < b + c$.

(7) $\forall a, b, c,\ a < b \Rightarrow ac < bc$.

(8) $\forall a, b, c,\ a < b$ and $b < c \Rightarrow a < c$.

(9) $\forall a,\ \neg(a < a)$.

The induction axiom is also known as the *weak principle of induction* (WPI). Equivalent form of this is the strong principle of induction (SPI):

**Theorem 1.1.1.** (*Strong Principle of Induction*) If

(1) $P(1)$ holds, and

(2) $\forall n \in \mathbb{N}$, we have $(P(m)\ \forall m \leqslant n) \Rightarrow P(n+1)$,

then $P(n)$ holds $\forall n \in \mathbb{N}$.

Clearly, SPI $\Rightarrow$ WPI. To see that WPI $\Rightarrow$ SPI, apply the former to

$$Q(n) = \text{“}P(m) \text{ holds } \forall m \leqslant n\text{”}$$

**Theorem 1.1.2.** (*Well-Ordering Principle* (WOP)) If $P(n)$ holds for some $n \in \mathbb{N}$, then there is a least $n \in \mathbb{N}$ such that $P(n)$ holds. "*Every non-empty subset of* $\mathbb{N}$ *has a minimal element*."

**Theorem 1.1.3.** SPI is equivalent to WOP.

*Proof.* First we show that WOP implies SPI. We assume (1) and (2) of SPI, and show that $P(n)$ holds $\forall n \in \mathbb{N}$, using WOP. Suppose, on the contrary, that $P(n)$ is not true for all $n \in \mathbb{N}$. Then,

$$C = \{n \in \mathbb{N} \mid P(n) \text{ is false}\} \neq \varnothing$$

By well-ordering principle, $C$ has a minimal element, say $m$. Now $\forall k < m$, $k \notin C$ (by minimality of $m$), so $P(k)$ holds $\forall k < m$. But by (2) of strong principle of induction, $P(m)$ holds, contradicting $m \in C$. Hence SPI holds.

To show that SPI implies WOP, suppose there is no least $n \in \mathbb{N}$ such that $P(n)$ holds. We want to show that $P(n)$ does not hold for any $n \in \mathbb{N}$, using SPI. Consider

$$Q(n) = \neg P(n)$$

Certainly $P(1)$ is false,[2] so $Q(1)$ holds. Given $n \in \mathbb{N}$, suppose that $Q(k)$ is true $\forall k < n$. Then $P(k)$ is false $\forall k < n$. So $P(n)$ is false as otherwise $n$ would be the minimal element for which $P$ holds. Hence $Q(n)$ is true, and (2) of SPI is satisfied, so $Q(n)$ holds for all $n \in \mathbb{N}$. Thus $P(n)$ is false $\forall n \in \mathbb{N}$. ■

Well-ordering principle enables us to prove $P(n)$ is true $\forall n \in N$ as follows: if not, there is a minimal counterexample, and we may try and derive a contradiction.

### 1.1.2 Integer

The integer set, $\mathbb{Z}$, consist of all symbols $n$, $-n$, where $n \in \mathbb{N}$, along with 0. We can also define $+$ and $\cdots$ etc. on $\mathbb{Z}$ from $\mathbb{N}$, and check that the usual rules (1) - (5) of arithmetic hold. We also have

(10) $\forall a \in \mathbb{Z},\ a + 0 = 0$ (identity for $+$).

(11) $\forall a \in \mathbb{Z},\ \exists b \in \mathbb{Z}$ such that $a + b = 0$ (inverses for $+$). Here, $b = -a$.

Similarly define $a < b$ if $a + c = b$ for some $c \in \mathbb{N}$. Rules (6), (8), and (9) continue to hold, but (7) must be modified:

(7) $\forall a, b, c \in \mathbb{Z},\ a < b$ and $c > 0 \Rightarrow ac < bc$.

[2] Else 1 would be the minimal element.

### 1.1.3 Rational Number

The rational number set, $\mathbb{Q}$, consist of all expressions $a/b$, where $a$, $b$ are integers, $b \neq 0$, and $a/b$ and $c/d$ are regarded as the same if $ad = bc$. Define

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$$

and one can check it does not matters how we wrote $a/b$ or $c/d$ (commutivity). We similarly define multiplication, and

$$\frac{a}{b} < \frac{c}{d}$$

where $b, d > 0$ if $ad < bc$. All the previous rules apply to rational numbers, but, furthermore,

(12) $\forall a \in \mathbb{Q}$, $a \neq 0$, $\exists b$ such that $ab = 1$ (inverse for $\cdot$).

Finally, note that

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q}$$

### 1.1.4 Primes

**Definition 1.1.2.** A natural number $n \geqslant 2$ is *prime* if its only factors are $\pm 1$ and $\pm n$. If $n \geqslant 2$ is not prime, then it is *composite*.

**Proposition 1.1.4.** Every natural number $n \geqslant 2$ can be written as a product of primes.

*Proof.* It is true for $n = 2$. Let $n > 2$ and suppose that claim holds up to and including $n - 1$. If $n$ is prime, done. If $n$ is composite, then $n = ab$ for some $1 < a, b < n$.

By the induction hypothesis, we have

$$a = p_1 p_2 \cdots p_k$$
$$b = q_1 q_2 \cdots q_l$$

for some primes $p_1, \ldots, p_k, q_1, \ldots, q_l$. Hence

$$n = ab = p_1 \cdots p_k q_1 \cdots q_l$$

is a product of primes. Proof is complete by induction. ■

**Theorem 1.1.5.** There are infinitely many primes.

*Proof.* (Euclid, 300 B.C.) Suppose there are finitely many primes, say $p_1, \ldots, p_k$. Let $N = p_1 \cdots p_k + 1$. Then $p_1 \nmid N$, else

$$p_1 | N - p_1 \cdots p_k = 1$$

Likewise, none of $p_2, p_3, \ldots, p_k$ divide $N$, contradicting the fact that $N$ can be written as a product of primes (Proposition 1.1.4). ■

Can a number have more than one representation as a product of primes? Our proof of Proposition 1.1.4 does not give uniqueness.

**Definition 1.1.3.** Given $a, b \in \mathbb{N}$, a natural number $c$ is the *highest common factor* (hcf) or *greatest common divisor* (gcd) of $a$ and $b$ if

(1) $c|a$ and $c|b$;

(2) $d|a$ and $d|b \Rightarrow d|c$.

We write $c = \text{hcf}(a, b)$, or $c = \gcd(a, b)$, or $c = (a, b)$.

For example, the factors of 12 are $1, 2, 3, 4, 6, 12$, and those of 18 are $1, 2, 3, 6, 9, 18$. So the common factors are $1, 2, 3, 6$, hence $\gcd(12, 18) = 6$. But observe that if $a$ and $b$ had common factors $1, 2, 3, 4, 6$, then $a$ and $b$ would have no gcd according to Definition 1.1.3 (2). We will need to show that $\gcd(a, b)$ always exists.

**Proposition 1.1.6.** (*Division Algorithm*) Let $n, k \in \mathbb{N}$. Then we can write

$$n = qk + r$$

where $q$ and $r$ are integers with $0 \leqslant r \leqslant k - 1$.

*Proof.* It is true for $n = 1$. Suppose $n \geqslant 2$ and statement holds for $n - 1$, i.e.

$$n - 1 = qk + r$$

for some $q, r \in \mathbb{Z}$, $0 \leqslant r \leqslant k - 1$.

- If $r < k - 1$, then $n = (n-1) + 1 = qk + (r + 1)$.
- If $r = k - 1$, then $n = (n-1) + 1 = qk + (k-1) + 1 = (q+1)k$.

■

Note that $q$ and $r$ obtained by the division algorithm are unique: if $n = qk + r = q'k + r'$, then

$$(q - q')k = r' - r$$

is an integer smaller than $k$ and larger than $-k$ so $q = q'$ and $r = r'$.

We now introduce the *Euclid's Algorithm*. We break down $a$ and $b$ until $r_{n+1} = 0$ as follows;

$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n \\ r_{n-1} &= q_{n+1} r_n + r_{n+1} \end{aligned}$$

and the algorithm returns $r_n$. N.b. the algorithm terminates in $n < b$ steps, since

$$b > r_1 > r_2 > \ldots > r_n > 0$$

**Theorem 1.1.7.** The output of Euclid's algorithm with input $a$, $b$ is $\gcd(a, b)$.

*Proof.* We show (1) and (2) from Definition 1.1.3.

(1) Have $r_n / r_{n-1}$ (as $r_{n+1} = 0$), so $r_n / r_{n-2}$ and $r_n / r_i$ $\forall i = 1, 2, \ldots, n-1$ by induction. Hence $r_n / b$ and $r_n / a$.

(2) Given $d$ such that $d | a$ and $d | b$, have $d | r_1$, so $d | r_2$ and $d | r_i$ $\forall i = 1, 2, \ldots, n$ by induction.

■

By the division algorithm and Euclid's algorithm, we have found out that the gcd of any two natural numbers always exists, and is unique. For example, consider Euclid's algorithm with 87 and 52 as inputs:

$$\begin{aligned} 87 &= 1 \cdot 52 + 35 \\ 52 &= 1 \cdot 35 + 17 \\ 35 &= 2 \cdot 17 + 1 \\ 17 &= 17 \cdot 1 \end{aligned}$$

so $\gcd(87, 52) = 1$.

**Definition 1.1.4.** When $\gcd(a, b) = 1$, we say that $a$ and $b$ are *coprime*.

Observe that we can reverse the Euclid's algorithm:

$$\begin{aligned}1 &= 35 - 2 \cdot 17\\ &= 35 - 2(52 - 1 \cdot 35)\\ &= -2 \cdot 52 + 3 \cdot 35\\ &= -2 \cdot 52 + 3(87 - 1 \cdot 52)\\ &= -5 \cdot 52 + 3 \cdot 87\end{aligned}$$

**Theorem 1.1.8.** $\forall a, b \in \mathbb{N}$, $\exists x, y \in \mathbb{Z}$ such that

$$xa + yb = \gcd(a, b)$$

*"We can write* $\gcd(a,b)$ *as a linear combination of a and b."*

*Proof.* (*Method 1*) Run Euclid's algorithm with input $a$, $b$ to obtain an output $r_n$. At step $n$, have

$$r_n = xr_{n-1} + yr_{n-2}$$

for some $x, y \in \mathbb{Z}$. But from step $n-1$, we see that $r_{n-1}$ is expressible as

$$r_{n-1} = xr_{n-2} + yr_{n-3}$$

for some $x, y \in \mathbb{Z}$, whence

$$r_n = xr_{n-2} + yr_{n-3}$$

for some $x, y \in \mathbb{Z}$. Continuing inductively, we have $\forall i = 2, \ldots, n-1$,

$$r_n = xr_i + yr_{i-1}$$

for some $x, y \in \mathbb{Z}$. Thus

$$r_n = xa + yb$$

for some $x, y \in \mathbb{Z}$, by steps one and two. ■

*Remark* 1.1.1. Euclid's algorithm not only proves the existence of $x, y \in \mathbb{Z}$, but gives a quick way to find them.

*Proof.* (*Method 2*) Let $g$ be the least positive linear combination of $a$ and $b$, i.e. the least positive integer of the form $xa + yb$ for some $x, y \in \mathbb{Z}$. We shall show that $g = \gcd(a, b)$ (Definition 1.1.3).

To see (2), observe that given $d$ such that $d|a$ and $d|b$, we have

$$d | ax + by \quad \forall x, y \in \mathbb{Z}$$

so in particular, $d|g$.

To see (1), suppose that $g \nmid a$. Then we can write

$$a = qg + r$$

for some $q, r \in \mathbb{Z}$ with $0 < r < g$. Hence

$$r = a - qg = a - q(xa + yb)$$

is also a positive linear combination of $a$ and $b$, and strictly smaller than $g$, contradicting the definition of $g$. Therefore, $g|a$ and by the same argument, $g|b$. ■

*Remark* 1.1.2. (*Method 2*) tells us that $\gcd(a,b)$ exists and is a linear combination of $a$ and $b$, but gives no way to find $\gcd(a,b)$ or the coefficients $x, y \in \mathbb{Z}$.

With help of Theorem 1.1.8, we can tell

$$160x + 72y = 33$$

does not have any integer solution, but

$$87x + 52y = 33$$

does. Let's formalise this.

**Corollary 1.1.8.1.** (*Bézout's Theorem*) Let $a, b \in \mathbb{Z}$. Then the equation

$$ax + by = c$$

has a solution in integers $x, y$ if and only if

$$\gcd(a,b)|c$$

*Proof.* Let $g = \gcd(a,b)$. Suppose there are $x, y \in \mathbb{Z}$ such that

$$ax + by = c$$

Then, since $g|a$ and $g|b$, $g|c$.

Conversely, suppose $g|c$. But Theorem 1.1.8 implies that there exist $x, y \in \mathbb{Z}$ such that

$$g = ax + by$$

But then

$$c = \frac{c}{g}g = \frac{c}{g}(ax + by) = a\left(x\frac{c}{g}\right) + b\left(y\frac{c}{g}\right)$$

■

**Proposition 1.1.9.** If $p$ is a prime and $p|ab$, then $p|a$ or $p|b$.

*Proof.* Suppose $p|ab$ but $p \nmid a$. We claim $p|b$.

Since $p$ is prime and $p \nmid a$, $\gcd(a,p) = 1$. Thus by Theorem 1.1.8, there exists $x, y \in \mathbb{Z}$ such that $xp + ya = 1$. It follows that $xpb + yab = b$, whence $b$ is a multiple of $p$ (as each of $p$ and $ab$ is). ■

*Remark* 1.1.3.

(1) Similarly,

$$p|a_1 a_2 \cdots a_n \Rightarrow p|a_i$$

for some $i = 1, 2, \ldots, n$. Indeed, Proposition 1.1.4 tells us that if $p|a_1 a_2 \cdots a_n$, the $p|a_1$ or $p|a_2 \cdots a_n$, so we may conclude by induction on the number of terms in the product.

(2) We do need $p$ to be prime.

**Theorem 1.1.10.** (*Fundamental Theorem of Arithmetic*) Every natural number $n \geqslant 2$ is expressible as a product primes, uniquely up to reordering.

*Proof.* Existence of factorisation follows from Proposition 1.1.4. We can show uniqueness by induction. Clearly it is unique for $n = 2$. Given $n \geqslant 2$, suppose

$$n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$$

where $p_i, q_j$ are all prime. Want to show that $k = l$, and, after reordering, $p_i = q_i \ \forall i = 1, \ldots, k$. We have

$$p_1 | n = q_1 q_2 \cdots q_l$$

so by Proposition 1.1.9,

$$p_1 | q_i$$

for some $i$. Relabelling the $q_i$, we may assume that $p_1 | q_1$. Since $q_1$ is prime, $p_1 = q_1$, so

$$\frac{n}{p_1} = p_2 \cdots p_k = q_2 \cdots q_l < n$$

By the induction hypothesis, $k = l$, and, after reordering, $p_2 = q_2, \ldots, p_k = q_k$. ■

*Remark* 1.1.4. There are arithmetical systems (permitting addition and multiplication) in which factorisation is not unique.

For example, consider $\mathbb{Z}[\sqrt{-3}]$, meaning all complex numbers of the form $x + y\sqrt{-3} = x + y\sqrt{3}i$ where $x, y \in \mathbb{Z}$. We can add and multiply two elements of $\mathbb{Z}[\sqrt{-3}]$ to get another element of $\mathbb{Z}[\sqrt{-3}]$, e.g.

$$(1 + \sqrt{-3}) + (q - \sqrt{-3}) = 2$$

$$(1 + \sqrt{-3})(q - \sqrt{-3}) = 4$$

In $\mathbb{Z}[\sqrt{-3}]$, we can define what it means to be a prime, and both $1 + \sqrt{-3}$ and $1 - \sqrt{-3}$ happen to be prime in this sense. But, we can also write $4 = 2 \cdot 2$, so factorisation is not unique.

Let's take a look at some applications of the fundamental theorem of arithmetic.

(1) What are the factors of $n = 2^3 3^7 11$? – All numbers of the form $2^a 3^b 11^c$ where $0 \leqslant a \leqslant 3$, $0 \leqslant b \leqslant 7$, $0 \leqslant c \leqslant 1$. There are no others: if, for example, $7|n$, then we would have a factorisation of $n$ involving 7, contradicting uniqueness. More generally, the factors of $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ are precisely the numbers of the form

$$p_1^{b_1} p_2^{b_2} \cdots p_k^{b_k}$$

with $0 \leqslant b_i \leqslant a_i$ $\forall i = 1, 2, \ldots, k$.

(2) What are common factors of $2^3 \cdot 3^7 \cdot 5 \cdot 11^3$ and $2^4 \cdot 3^2 \cdot 11 \cdot 13$? – All numbers of the form $2^3 3^b 11^c$ where $0 \leqslant a \leqslant 3$, $0 \leqslant b \leqslant 2$, $0 \leqslant c \leqslant 1$. Thus the gcd is $2^3 \cdot 3^2 \cdot 11$. In general, the gcd of $p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ and $p_1^{b_1} p_2^{b_2} \cdots p_k^{b_k}$, where $a_i, b_i \geqslant 0$, is

$$p_1^{\min\{a_1,b_1\}} \cdots p_k^{\min\{a_k,b_k\}}$$

(3) What are the common multiples of the two numbers in (2)? – All numbers of the form $2^a 3^b 5^c 11^d 13^2$ where $a \geqslant 4$, $b \geqslant 7$, $c \geqslant 1$, $d \geqslant 3$, $e \geqslant 1$, times any integer. Hence $2^4 \cdot 3^7 \cdot \cdot 5 \cdot 11^3 \cdot 13$ is a common multiple, and any other common multiple is a multiple of it. We say that it is the *least common multiple* (lcm) of the two numbers. In general, the lcm of $p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ and $p_1^{b_1} p_2^{b_2} \cdots p_k^{b_k}$, where $a_i, b_i \geqslant 0$, is

$$p_1^{\max\{a_1,b_1\}} \cdots p_k^{\max\{a_k,b_k\}}$$

Since $\min\{a_i, b_i\} + \max a_i, b_i = a_i + b_i$, we have

$$\gcd(x, y) \cdot \operatorname{lcm}(x, y) = xy$$

for any $x, y$.

(4) Another proof of Theorem 1.1.5, due Erdös (1930): let $p_1, p_2, \ldots, p_k$ be all the primes. Any number which is a product of jest these primes is of the form $(*)$:

$$p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k} = m^2 p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}$$

where $i_l = 0$ or 1. Let $M \in \mathbb{N}$. If a number is equal or less than $M$ is of the form $(*)$, then $m \leqslant \sqrt{M}$. So there are at most $\sqrt{M} 2^k$ numbers of the form $(*)$ that are equal or less than $M$. If $M > \sqrt{M} 2^k$, i.e. if $M > 4^k$, then there must be a number equal or less than $M$ which is not of the form $(*)$. But this number must have a prime factor not amongst the $p_1, \ldots, p_k$. ■

Proof by Euclid tells us that $k^{\text{th}}$ prime is less than $2^{2^k}$, while the proof by Erdös tells us $k^{\text{th}}$ prime is less than $4^k$. In fact, we know $k^{\text{th}}$ prime is $\sim k \ln k$ (*Prime Number Theorem*).

### 1.1.5 Modular Arithmetic

Let $n \geqslant 2$ be a natural number. Then the *integer modulo $n$*, written as $\mathbb{Z}_n$ or $\mathbb{Z}/n\mathbb{Z}$ consist of the integers, with two regarded as the same if they differ by a multiple of $n$. If $x$ and $y$ are the same in $\mathbb{Z}_n$, we write

$$x \equiv y \mod n$$

or $x \equiv y \quad (n)$ or $x = y$ in $\mathbb{Z}_n$. Thus

$$\begin{aligned} x \equiv y \mod n &\Leftrightarrow n | x - y \\ &\Leftrightarrow x = y + kn \text{ for some } k \in \mathbb{Z} \end{aligned}$$

N.b. if $a \equiv a' \mod n$ and $b \equiv b' \mod n$, then

$$n | (a - a') + (b - b') = (a + b) - (a' + b')$$

so $a + b \equiv a' + b' \mod n$. Similarly,

$$n | (a - a')b + a'(b - b') = ab - a'b'$$

so $ab = a'b' \mod n$. Hence we can do arithmetic on modulo $n$ with inheriting the usual rules of arithmetic in $\mathbb{Z}$.

**Example 1.1.1.** Does $2a^2 + 3b^3 = 1$ have a solution with $a, b \in \mathbb{Z}$?
If there is a solution, then $2a^2 \equiv 1 \mod 3$, but $2 \cdot 0^2 \equiv 0$, $2 \cdot 1^2 \equiv 2$, $2 \cdot 2^2 \equiv 2 \mod 3$ so there is no solution.

**⁎ Solution of congruences**

**Example 1.1.2.** Solve $7n \equiv 2 \mod 10$.
We note $3 \cdot 7 \equiv 1 \mod 10$ so $3 \cdot 7n \equiv 3 \cdot 2 \mod 10$ and $n \equiv 6 \mod 10$.

As shown in the example above, given $a, b \in \mathbb{Z}$, we say that $b$ is an *inverse of $a$ modulo $n$* if

$$ab \equiv 1 \mod n$$

We say that $a$ is *invertible modulo $n$* or that $a$ is a *unit modulo $n$*, if it has an inverse. For example, in $\mathbb{Z}_{10}$, 3 is an inverse of 7, and both 3 and 7 are units modulo 10; but 4 is not a unit modulo 10 since $4n \not\equiv 1 \mod 10$ for all $n \in \mathbb{Z}$.

*Remark* 1.1.5. If $a$ is a unit modulo $n$, then

(1) its inverse is unique. Suppose $\exists b, b'$ such that

$$ab \equiv ab' \equiv 1 \mod n$$

then

$$b \equiv bab \equiv bab' \equiv b' \mod n$$

(2) we can write $a^{-1}$ for its inverse.

(3) if $ab \equiv ac \mod n$, then $b \equiv c \mod n$.

However, this is not true in general, e.g.

$$4 \cdot 3 \equiv 4 \cdot 8 \mod 10$$

but $3 \not\equiv 8 \mod 10$.

**Proposition 1.1.11.** Let $p$ be prime. Then every $a \neq 0 \mod p$ is a unit modulo $p$.

*Proof.* From $\gcd(a, p) = 1$, $\exists x, y \in \mathbb{Z}$ such that $ax + py = 1$ (Corollary 1.1.8.1). Hence $ax = 1 - py$, so

$$ax \equiv 1 \mod p$$

for some $x \in \mathbb{Z}$. ■

**Proposition 1.1.12.** Let $n \geqslant 2$. Then $a$ is a unit modulo $n$ if and only if $\gcd(a,n) = 1$.

*Proof.*

$$\begin{aligned} \gcd(a,n) = 1 &\Leftrightarrow ax + ny = 1 \text{ for some } x,y \in \mathbb{z} \\ &\Leftrightarrow ax = 1 - ny : \text{ for some } x,y \in \mathbb{z} \\ &\Leftrightarrow ax \equiv 1 \mod n : \text{ for some } x,y \in \mathbb{z} \end{aligned}$$

**Corollary 1.1.12.1.** If $\gcd(a,n) = 1$, then the congruence

$$ax \equiv b \mod n$$

has a unique solution. In particular, if $\gcd(a,n) = 1$, then there is a unique inverse of $a$ modulo $n$.

For cases like $ax \equiv b \mod n$ with $\gcd(a,n) = d \neq 1$, the solution may not exist. $n|ax - b$ so $d|ax - b$ and thus $d|b$ for solution to exist. If $d|b$, write $n = dn'$, $a = da'$, $b = db'$. Then,

$$\begin{aligned} ax \equiv b \mod n &\Leftrightarrow ax - b = kn \text{ for some } k \in \mathbb{Z} \\ &\Leftrightarrow da'x - db' = kdn' : \text{ for some } k \in \mathbb{Z} \\ &\Leftrightarrow a'x - b' = kn' : \text{ for some } k \in \mathbb{Z} \\ &\Leftrightarrow a'x \equiv b' \mod n' \end{aligned}$$

Since $\gcd(a', n') = 1$, $a'x \equiv b' \mod n'$ has a unique solution.

**Example 1.1.3.** Solve $7x \equiv 4 \mod 30$.
We have $\gcd(7,30) = 1$, so by Euclid's algorithm, $13 \cdot 7 - 3 \cdot 30 = 1$. Hence $13 \cdot 7 \equiv 1 \mod 30$, whence

$$x \equiv 13 \cdot 4 \equiv 22 \mod 30$$

**Example 1.1.4.** Solve $10x \equiv 12 \mod 34$.
This is equivalent with $5x \equiv 6 \mod 17$, and we can now solve as the example above to obtain $x \equiv 6 \cdot 7 \equiv 8 \mod 17$.

### ∗ Simultaneous congruences

Note

$$x \equiv 5 \mod 12 \Rightarrow \begin{cases} x \equiv 2 \mod 3 \\ x \equiv 1 \mod 4 \end{cases}$$

However, we cannot be sure if the converse is true, i.e. does $x \equiv 2 \mod 3$ and $x \equiv 1 \mod 4$ imply that $x \equiv 5 \mod 12$? It turns out that it is true for the case above, but it is not generally true – see $x \equiv 1 \mod 4$ and $x \equiv 2 \mod 6$.

**Theorem 1.1.13.** (*The Chinese Remainder Theorem*) Let $m, n$ be coprime, and $a, b \in \mathbb{Z}$. Then there is a unique solution modulo $mn$ to the simultaneous congruences

$$x \equiv a \mod m \text{ and } x \equiv b \mod n$$

That is, there is a solution $x$ to

$$\begin{cases} x \equiv a \mod m \\ x \equiv b \mod n \end{cases}$$

and $y$ is a solution if and only if $x \equiv y \mod mn$.

*Proof.* We first prove existence. Since $\gcd(m,n) = 1$, $\exists s, t \in \mathbb{Z}$ with $sm + tn = 1$. Note

$$sm \equiv 1 \mod n, \quad sm \equiv 0 \mod m$$

and

$$tn \equiv 1 \mod m, \quad tn \equiv 0 \mod n$$

Hence

$$\begin{aligned} x = a(tn) + b(sm) &\equiv a \mod m \\ &\equiv b \mod n \end{aligned}$$

Next, for uniqueness, suppose $y$ is also a solution, i.e. $y \equiv a \mod m$ and $y \equiv b \mod n$. This implies

$$\begin{aligned} y \equiv x \mod m \text{ and } y \equiv x \mod n &\Leftrightarrow m|y-x \text{ and } n|y-x \\ &\Leftrightarrow mn|y-x \text{ since } \gcd(m,n)=1 \\ &\Leftrightarrow y \equiv x \mod mn \end{aligned}$$

■

*Remark* 1.1.6. Theorem 1.1.13 can be extended, by induction, to more than two moduli: if $m_1, m_2, \ldots, m_k$ are pairwise coprime, then $\forall a_1, a_2, \ldots, a_k \in \mathbb{Z}$, $\exists x \in \mathbb{Z}$ such that

$$\begin{gathered} x \equiv a_1 \mod m_1 \\ x \equiv a_2 \mod m_2 \\ \vdots \\ x \equiv a_k \mod m_k \end{gathered}$$

**Definition 1.1.5.** We define the *Euler totient function*, $\varphi(m)$, as the number of integers $a$ with $1 \leqslant a \leqslant m$ such that $\gcd(a,m) = 1$, i.e. $\varphi(m)$ is the number of units modulo $m$. Additionally, we define $\varphi(1) = 1$.

For example, when $p$ is prime, $\varphi(p) = p-1$, and $\varphi(p^2) = p^2 - p$. When $p, q$ are distinct primes,

$$\varphi(p,q) = pq - p - q + 1$$

Next, let's take a look at behavior of a power of an integer modulo $p$.

**Theorem 1.1.14.** (*Fermat's Little Theorem*) Let $p$ be a prime. Then $a^p \equiv a \mod p$ for all $a \in \mathbb{Z}$. Equivalently,

$$a^{p-1} \equiv 1 \mod p$$

for all $a \not\equiv 0 \mod p$.

*Proof.* If $a \not\equiv 0 \mod p$, then $a$ is a unit mod $p$. Thus $ax \equiv ay \mod p$ if and only if $x \equiv y \mod p$. Hence the numbers $a, 2a, \ldots, (p-1)a$ are pairwise incongruent (distinct) modulo $p$, and $\not\equiv 0$ mod $p$, so they are $1, 2, 3, \ldots, p-1$ in some order. Hence

$$a \cdot 2a \cdots (p-1)a \equiv 1 \cdot 2 \cdots (p-1) \mod p$$

or

$$a^{p-1}(p-1)! \equiv (p-1)! \mod p$$

But $(p-1)!$ is a unit mod $p$ (since it is a product of units), so we can cancel it to obtain $a^{p-1} \equiv 1$ mod $p$. ■

**Theorem 1.1.15.** (*Fermat-Euler Theorem*) Let $\gcd(a,m) = 1$. Then $a^{\varphi(m)} \equiv 1 \mod m$.

*Proof.* Let

$$U = \{x \in \mathbb{Z} \mid 0 < x < m, \gcd(x,m) = 1\}$$

■

# CHAPTER 2

## Groups

Lecture by Professor Claude Warnick. Michaelmas term 2023.

# CHAPTER 3

## Analysis I

Lecture by Professor Claude Warnick. Lent term 2024.

## 3.1 Limits and Convergence

### 3.1.1 Sequences

Consider a sequence of real numbers

$$a_1, a_2, a_3, \dots$$

$(a_n)$ where $a_n \in \mathbb{R}$. We start by defining *limit*.

**Definition 3.1.1.** We say $a_n \to a$ as $n \to \infty$ for some $a \in \mathbb{R}$ if: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|a_n - a| < \epsilon \quad \forall n \geqslant N$$

If $a_n \leqslant a_{n+1}$ $\forall n$ we say $(a_n)$ is *increasing*; and *decreasing, strictly increasing, strictly decreasing* for $a_n \geqslant a_{n+1}$, $a_n > a_{n+1}$, $a_n < a_{n+1}$ respectively. In all above cases, $(a_n)$ is *monotone*.

**Theorem 3.1.1.** (*Monotone Convergence Theorem*) An increasing sequence of real numbers which is bounded above converges (i.e. it has a limit).

In other words, monotone convergence theorem says if $a_n \in \mathbb{R}$ $(n \geqslant 1)$, $A \in \mathbb{R}$ with

$$a_1 \geqslant a_2 \geqslant a_3 \geqslant \cdots \text{ and } a_n \geqslant A \quad \forall n$$

there exists $a \in \mathbb{R}$ such that $a_n \to a$ as $n \to \infty$. Equivalently, decreasing sequence bounded below has a limit. Theorem 3.1.1 is also equivalent with Theorem 3.1.2

**Theorem 3.1.2.** (*Least Upper Bound Axiom*) Every non-empty set bounded above has a *supremum*.

**Definition 3.1.2.** (Supremum) If $S \subset \mathbb{R}, S \neq \varnothing$ we say that

$$\sup S = K$$

if

(1) $x \leqslant K \quad \forall x \in S$ ($K$ is upper bound for $S$).

(2) Given $\epsilon > 0$, $\exists x \in S$ such that $x > K - \epsilon$ ($K$ is least upper bound).

Note that similar definition can be made for greatest lower bound and infimum. If $\sup S \in S$, we say $\sup S$ is the *maximum* of $S$, i.e. $\sup S = \max S$.

**Lemma 3.1.3.**

(1) The limit is unique, i.e. if $a_n \to a$ and $a_n \to b$ as $n \to \infty$ then $a = b$.

(2) Subsequences converge to the same limit, i.e. if $a_n \to a$ as $n \to \infty$ and $n_1 < n_2 < \cdots$, then $a_{n_j} \to a$ as $j \to \infty$.

(3) If $a_n = c \quad \forall n$, then $a_n \to c$ as $n \to \infty$.

(4) If $a_n \to a$ and $b_n \to b$ then $a_n + b_n \to a + b$.

(5) If $a_n \to a$ and $b_n \to b$ then $a_n b_n \to ab$.

(6) If $a_n \to a$, $a_n \neq 0$ and $a \neq 0$, then $1/a_n \to 1/a$.

(7) If $a_n \leqslant A \quad \forall n$ and $a_n \to a$, then $a \leqslant A$.

(8) If $a_n \to a$ and $c_n \to a$ as $n \to \infty$ and $a_n \leqslant b_n \leqslant c_n$, then $b_n \to a$.

*Proof.* We will only prove 1, 2, and 5 here.

(1) For any $\epsilon > 0$, we can find $N_1(\epsilon)$ and $N_2(\epsilon)$ such that

$$n \geqslant N_1(\epsilon) \Rightarrow |a_n - a| < \epsilon$$

$$n \geqslant N_2(\epsilon) \Rightarrow |a_n - b| < \epsilon$$

If $n \geqslant \max\{N_1(\epsilon), N_2(\epsilon)\}$, then

$$\begin{aligned} 0 \leqslant |b-a| &= |b - a_n + a_n - a| \\ &\leqslant |b - a_n| + |a_n - a| < 2\epsilon \end{aligned}$$

This implies |b-a|=0, otherwise we can set $\epsilon = |b-a|/3$ to find

$$0 \leqslant |b-a| < \frac{2}{3}|b-a|$$

■

(2) Since $n_j < n_{j+1} \Rightarrow n_{j+1} \geqslant n_j + 1$, by induction, we have $n_j \geqslant j$. As $a_n \to a$ as $n \to \infty$, given $\epsilon > 0$ there exists $N = N(\epsilon)$ such that

$$n \geqslant N(\epsilon) \Rightarrow |a_n - a| < \epsilon$$

So if $j \geqslant N(\epsilon)$, then $n_j \geqslant j \geqslant N(\epsilon)$ and

$$|a_{n_j} - a| < \epsilon$$

■

(3) As $a_n \to a$, $b_n \to b$ as $n \to \infty$, for any $\epsilon > 0$, we can find $N_1(\epsilon)$ and $N_2(\epsilon)$ such that

$$n \geqslant N_1(\epsilon) \Rightarrow |a_n - a| < \epsilon$$

$$n \geqslant N_2(\epsilon) \Rightarrow |b_n - b| < \epsilon$$

Now

$$\begin{aligned} |a_n b_n - ab| &= |a_n b_n - a_n b + a_n b - ab| \\ &\leqslant |a_n b_n - a_n b| + |a_n b - ab| \\ &= |a_n||b_n - b| + |b||a_n - a| \end{aligned}$$

If $n \geqslant N_1(1)$ then $|a_n - a| \leqslant 1$ and hence

$$|a_n| = |a_n - a + a| \leqslant |a_n - a| + |a| \leqslant 1 + |a|$$

Thus if $n \geqslant N_3(\epsilon) = \max\{N_1(1), N_1(\epsilon), N_2(\epsilon)\}$ then

$$|a_n b_n - ab| < (1 + |a|)\epsilon + |b|\epsilon = (1 + |a| + |b|)\epsilon$$

which can be made as small as we like.

■

**Lemma 3.1.4.**

(1) $1/n \to 0$ as $n \to \infty$.

*Proof.* $(1/n)$ is a decreasing sequence, bounded below by 0, so by monotone convergence theorem it has a limit $L$. Now,

$$\frac{1}{2n} = \left(\frac{1}{2}\right)\left(\frac{1}{n}\right) \to \frac{1}{2}L$$

(Lemma 3.1.3 (3), (5)). But $(1/2n)$ is a subsequence of $(1/n)$, so by Lemma 3.1.3 (2),

$$\frac{1}{2n} \to L$$

Then, by Lemma 3.1.3 (1),

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

and $L = 0$. ■

(2) If $|x| < 1$, $x^n \to 0$ as $n \to \infty$.

*Proof.* Suppose $0 \leqslant x < 1$. Then $(x^n)$ is a decreasing sequence bounded below by 0, and it converges by monotone convergence theorem to a limit $L$. Now,

$$x^{n+1} = xx^n \to xL$$

(Lemma 3.1.3 (5)). However, $(x^{n+1})$ is a subsequence of $(x^n)$ so by Lemma 3.1.3 (2), $x^{n+1} \to L$; and consequently by Lemma 3.1.3 (1), $L = xL \Rightarrow L = 0$.

Suppose $-1 < x < 1$. Then,

$$-|x|^n \leqslant x^n \leqslant |x|^n$$

We know $|x|^n \to 0$ and $-|x|^n \to 0$ from above, so $x^n \to 0$ by Lemma 3.1.3 (8). ■

Note that when we talk about a sequence converging, we mean to a *finite* limit, while there is a notion of *tending to infinity*: $a_n \to \infty$ if $\forall M \in \mathbb{R}\ \exists N = N(M)$ such that $n \geqslant N \Rightarrow a_n > M$.

Meanwhile, we can see that our definition of convergence still works if $a_n, a \in \mathbb{C}$ (Definition 3.1.1). Furthermore, Lemma 3.1.3 (1) to (6) all apply for complex sequences, but (7), (8), and Theorem 3.1.1 use the *order relation* so do not carry over directly.

**Lemma 3.1.5.** If $(z_n)$ is a complex sequence, then $z_n \to z$ if and only if $\mathrm{Re}(z_n) \to \mathrm{Re}(z)$ and $\mathrm{Im}(z_n) \to \mathrm{Im}(z)$.

*Proof.* Note if $\omega \in \mathbb{C}$,

$$\max\{|\mathrm{Re}(\omega)|, |\mathrm{Im}(\omega)|\} \leqslant |\omega| \leqslant |\mathrm{Re}(\omega)| + |\mathrm{Im}(\omega)|$$

($\Rightarrow$) Suppose $z_n \to z$. Then, $\forall \epsilon > 0$ there exists $N = N(\epsilon)$ such that

$$\begin{aligned} n \geqslant N \Rightarrow |z_n - z| < \epsilon \Rightarrow & |\mathrm{Re}(z_n) - \mathrm{Re}(z)| < \epsilon \\ & \text{and } |\mathrm{Im}(z_n) - \mathrm{Im}(z)| < \epsilon \end{aligned}$$

Therefore $\mathrm{Re}(z_n) \to \mathrm{Re}(z)$ and $\mathrm{Im}(z_n) \to \mathrm{Im}(z)$.

($\Leftarrow$) Suppose $\mathrm{Re}(z_n) \to \mathrm{Re}(z)$ and $\mathrm{Im}(z_n) \to \mathrm{Im}(z)$. Then $\forall \epsilon > 0\ \exists N_1 = N_1(\epsilon), N_2 = N_2(\epsilon)$ such that

$$n \geqslant N_1 \Rightarrow |\mathrm{Re}(z_n) - \mathrm{Re}(z)| < \epsilon$$
$$n \geqslant N_2 \Rightarrow |\mathrm{Im}(z_n) - \mathrm{Im}(z)| < \epsilon$$

So if $n \geqslant \max\{N_1(\epsilon), N_2(\epsilon)\} = N_3(\epsilon)$, then

$$|z_n - z| \leqslant |\mathrm{Re}(z_n) - \mathrm{Re}(z)| + |\mathrm{Im}(z_n) - \mathrm{Im}(z)| < 2\epsilon$$

and therefore $z_n \to z$. ∎

**Theorem 3.1.6.** (*Bolzano-Weierstrass Theorem*) Every bounded sequence in $\mathbb{R}$ has a convergent subsequence, i.e. if $x_n \in \mathbb{R}$ and there exists $K > 0$ such that $|x_n| \leqslant K \quad \forall n$, we can find $n_1 < n_2 < n_3 < \cdots$ and $x \in \mathbb{R}$ such that $x_{n_j} \to x$ as $j \to \infty$. N.b. we do not assert uniqueness, e.g. if $x_n = (-1)^n$, $x_{2n} \to 1$ but $x_{2n+1} \to -1$.

*Proof.* Set $a_1 = -K$, $b_1 = K$ so that all terms of $(x_n)$ lie in $[a_1, b_1]$. Let $c = (a_1 + b_1)/2$. Either

(1) There are infinitely many terms of $(x_n)$ in $[a_1, c]$ or

(2) There are infinitely many terms of $(x_n)$ in $[c, b_1]$.

If (1) holds, we set $a_2 = a_1$, $b_2 = c$; or if (1) does not hold, we set $a_2 = c$, $b_2 = b_1$. Continue inductively to construct $a_k$, $b_k$ such that infinitely many terms of $(x_n)$ lie in $[a_k, b_k]$, with

$$b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k)$$

and

$$a_k \leqslant a_{k+1} < b_{k+1} \leqslant b_k$$

By construction, $(a_k)$ is increasing and bounded above by $b_1$. So, from the monotone convergence theorem,

$$a_k \to a \in [a_1, b_1] \text{ as } k \to \infty$$

Similarly, since $(b_k)$ is decreasing and bounded below by $a_1$,

$$b_k \to b \in [a_1, b_1] \text{ as } k \to \infty$$

But

$$b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k) \Rightarrow b - a = \frac{1}{2}(b - a)$$

and $b = a$. Now construct $n_j$ as follows.

- Pick $n_1 = 1$.
- Pick $n_j$ to be the smallest integer greater than $n_{j-1}$ such that $x_{n_j} \in [a_j, b_j]$.

We can do this as $[a_j, b_j]$ contains infinitely many terms of $(x_n)$. Viz, $a_j \leqslant x_{n_j} \leqslant b_j$ and by Lemma 3.1.3 (8), $x_{n_j} \to a$. ∎

**Definition 3.1.3.** A sequence $(a_n)$ of real numbers is a *Cauchy sequence* (Cauchy) if for all $\epsilon > 0$ there exists $N = N(\epsilon)$ such that $n, m \geqslant N \Rightarrow |a_n - a_m| < \epsilon$.

**Lemma 3.1.7.** Every convergent sequence is a Cauchy sequence.

*Proof.* If $a_n \to a$, then given $\epsilon > 0\ \exists N = N(\epsilon)$ such that if $n \geqslant N$, $|a_n - a| < \epsilon$. But if $m, n \geqslant N$,

$$|a_n - a_m| = |a_n - a + a - a_m| \leqslant |a_n - a| + |a_m - a| < 2\epsilon$$

∎

**Theorem 3.1.8.** Every Cauchy sequence is convergent.

*Proof.* Suppose $(a_n)$ is Cauchy. First we show $(a_n)$ is bounded. Since $(a_n)$ is Cauchy, given $\epsilon > 0$ $\exists N = N(\epsilon)$ such that

$$|a_n - a_m| < \epsilon \quad \forall n, m \geqslant N(\epsilon)$$

So if $n, m \geqslant N(1)$, then $|a_n - a_m| < 1$. Set $m = N(1)$ to see

$$|a_n| = |a_n - a_{N(1)} + a_{N(1)}| \leqslant |a_n - a_{N(1)}| + |a_{N(1)}| \leqslant 1 + |a_{N(1)}| \quad \forall n \geqslant N(1)$$

Thus if $K = 1 + \max\{|a_1|, \dots, |a_{N(1)}|\}$,

$$|a_n| \leqslant K \quad \forall n$$

Now, by Bolzano-Weierstrass theorem (Theorem 3.1.6), there must be $a \in \mathbb{R}$ and $n_1 < n_2 < n_3 < \cdots$ such that $a_{n_j} \to a$ as $j \to \infty$. We claim, in fact, $a_n \to a$ as $n \to \infty$. To see this, write

$$|a_n - a| = |a_n - a_{n_j} + a_{n_j} - a| \leqslant |a_n - a_{n_j}| + |a_{n_j} - a|$$

and fix $\epsilon > 0$. Since $(a_n)$ is Cauchy, $\exists N_1(\epsilon)$ such that if $n, n_j \geqslant N_1(\epsilon)$, $|a_n - a_{n_j}| < \epsilon$.

Since $a_{n_j} \to a$, $\exists N_2(\epsilon)$ such that if $j \geqslant N_2(\epsilon)$, $|a_{n_j} - a| < \epsilon$. Fix $j$ such that $j \geqslant N_2(\epsilon)$ and $n_j \geqslant N_1(\epsilon)$. Then we have, for any $n \geqslant N_1(\epsilon)$,

$$|a_n - a| < 2\epsilon$$

i.e. $a_n \to a$. ∎

We have shown a real sequence converges if and only if it is Cauchy (Lemma 3.1.7 and Theorem 3.1.8). This is called the *Cauchy's general principle of convergence*.

**Corollary 3.1.8.1.** A complex sequence converges if and only if it is Cauchy.

*Proof.* From estimate at start of proof of Lemma 3.1.5, it follows that $(z_n)$ is Cauchy if and only if $(\mathrm{Re}(z_n))$ and $(\mathrm{Im}(z_n))$ are both Cauchy. ∎

### 3.1.2 Series and Convergence Tests

**Definition 3.1.4.** Suppose $a_j \in \mathbb{R}$( or $\mathbb{C}$). Define sequence of *partial sums* $(S_N)$ where

$$S_N = \sum_{j=1}^{N} a_j$$

Then, if $S_N \to S$ as $N \to \infty$, we write

$$S = \sum_{j=1}^{\infty} a_j$$

and say $\sum_{j=1}^{\infty} a_j$ converges. Conversely, if $S_N$ does not converge, we say $\sum_{j=1}^{\infty} a_j$ diverges.

*Remark* 3.1.1. Any problem on series is a problem about sequences, and vice versa.

**Lemma 3.1.9.**

(1) If $\sum_{j=1}^{\infty} a_j$ and $\sum_{j=1}^{\infty} b_j$ converge, then so does

$$\sum_{j=1}^{\infty} (\lambda a_j + \mu b_j)$$

for $\lambda, \mu \in \mathbb{R}$( or $\mathbb{C}$); and

$$\sum_{j=1}^{\infty} (\lambda a_j + \mu b_j) = \lambda \sum_{j=1}^{\infty} a_j + \mu \sum_{j=1}^{\infty} b_j$$

(2) Suppose there exists $N$ such that $a_j = b_j \ \forall j \geqslant N$. Then either $\sum_{j=1}^{\infty} a_j$ and $\sum_{j=1}^{\infty} b_j$ both converge, or both diverge.

*Proof.* Suppose $n \geqslant N$. Then,

$$S_n = \sum_{j=1}^{N-1} a_j + \sum_{j=N}^{n} a_j$$

and

$$\begin{aligned} D_n &= \sum_{j=1}^{N-1} b_j + \sum_{j=N}^{n} b_j \\ &= S_n + \left( \sum_{j=1}^{N-1} b_j - \sum_{j=1}^{N-1} a_j \right) \\ &= S_n + k \end{aligned}$$

where $k$ finite constant. Thus, $D_n$ converges if and only if $S_n$ converges. ■

An important example is the *geometric series*. Suppose $x \in \mathbb{R}$ and let $a_j = x^{j-1}$ for $j \geqslant 1$. Then,

$$S_n = \sum_{j=1}^{n} x^{j-1} = 1 + x + x^2 + \cdots + x^{n-1}$$

and

$$S_n = \begin{cases} \frac{1-x^n}{1-x} & x \neq 1 \\ n & x = 1 \end{cases}$$

We can check the convergence of $S_n$ for different values of $x$:

- if $|x| < 1$, $x^n \to 0$ and $S_n \to 1/(1-x)$.
- if $x \geqslant 1$, write $x = 1 + \delta$ where $\delta \geqslant 0$. By expansion,

  $$x^n = 1 + \delta n + \cdots + \delta^n \geqslant 1 + \delta n$$

  Hence, $x^n \to \infty$ and $S_n \to \infty$ for $x > 1$.
- if $x = 1$, trivially $S_n = n \to \infty$.
- if $x < -1$, $x^n = (-1)^n |x|^n$ does not converge as $|x|^n \to \infty$. $|S_n| \to \infty$ but $S_n$ may alternate in sign.
- if $x = -1$,

  $$S_n = \begin{cases} 1 & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

  and hence it does not converge.

Therefore, we may conclude that $\sum_{j=1}^{\infty} x^{j-1}$ converges if and only if $|x| < 1$.

Let's take a further look into convergence tests. A simple, but useful result is the $n^{th}$ *term test*.

**Lemma 3.1.10.** If $\sum_{j=1}^{\infty} a_j$ converges, then $a_j \to 0$ as $j \to \infty$.

*Proof.* If $S_n = \sum_{j=1}^{n} a_j$, $a_n = S_n - S_{n-1}$ but

$$S_n \to S \Rightarrow S_{n-1} \to S$$

so $a_n \to 0$. ■

Note that converse of Lemma 3.1.10 is false. For example, consider $\sum_{j=1}^{\infty} 1/j$, i.e. *harmonic series*. If we let $S_n = \sum_{j=1}^{n} 1/j$,

$$\begin{aligned} S_{2n} &= S_n + \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} \\ &\geqslant S_n + \frac{1}{2n} + \frac{1}{2n+1} + \cdots + \frac{1}{2n} \end{aligned}$$

Thus $S_{2n} \geqslant S_n + 1/2$ and $S_{2n} - S_n \geqslant 1/2$. If $S_n \to S$, then $S_{2n} \to S$ and $0 \geqslant 1/2$, which is a contradiction. Therefore, $S_n = \sum_{j=1}^{n} 1/j$ diverges.

Moreover, Lemma 3.1.10 is often used to show a series does not converge.

### ⁎ Series with non-negative terms

We consider, for the time being, series whose terms satisfy

$$a_j \geqslant 0 \quad \forall j$$

**Theorem 3.1.11.** (*Comparison Test*) Suppose $0 \leqslant b_n \leqslant a_n \quad \forall n$. Then if $\sum_{n=1}^{\infty} a_n$ converges, so does $\sum_{n=1}^{\infty} b_n$.

*Proof.* Let $S_N = \sum_{n=1}^{N} a_n$ and $D_N = \sum_{n=1}^{N} b_n$. Both $S_N$ and $D_N$ are monotonically increasing and $S_N \to S$. But

$$b_n \leqslant a_n \Rightarrow D_N \leqslant S_N \leqslant S$$

So, since $(D_N)$ is increasing and bounded above, $D_N$ converges (Theorem 3.1.1). ■

For example of the comparison test, consider $\sum_{n=1}^{\infty} 1/n^2$. We have

$$0 \leqslant \frac{1}{n^2} < \frac{1}{n(n-1)} = \underbrace{\frac{1}{n-1} - \frac{1}{n}}_{a_n}$$

for $n \geqslant 2$. Then,

$$\sum_{n=2}^{N} a_n = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{N-1} - \frac{1}{N}\right) = 1 - \frac{1}{N} \to 1$$

as $N \to \infty$. Hence, since $\sum_{n=2}^{\infty} 1/(n-1)n$ converges, $\sum_{n=2}^{\infty} 1/n^2$ and consequently $\sum_{n=1}^{\infty} 1/n^2$ converges by comparison test.

**Theorem 3.1.12.** (*Root Test*) Suppose $a_n \geqslant 0$ and $a_n^{1/n} \to a$ as $n \to \infty$. Then if $a < 1$, $\sum a_n$ converges; else if $a > 1$, $\sum a_n$ diverges. If $a = 1$, test is inconclusive.

*Proof.* If $a < 1$, choose $r$ such that $a < r < 1$. From Definition 3.1.1, $\exists N$ such that $\forall n \geqslant N$[1]

$$a_n^{1/n} < r \Rightarrow a_n < r^n$$

Since $r < 1$, the series $\sum r^n$ converges, so by comparison test and Lemma 3.1.9 (2), we have that $\sum a_n$ converges.

If $a > 1$, $\exists N$ such that $\forall n \geqslant N$ $a_n^{1/n} > 1 \Rightarrow a_n > 1$. Hence, $\sum a_n$ diverges (Lemma 3.1.10). ■

**Theorem 3.1.13.** (*Ratio Test*) Suppose $a_n > 0$ and $a_{n+1}/a_n \to l$. If $l < 1$ $\sum a_n$ converge; else if $l > 1$ $\sum a_n$ diverge. If $l = 1$, test is inconclusive.

[1] Choose $\epsilon = r - a > 0$.

*Proof.* Suppose $l < 1$ and choose $r$ with $l < r < 1$. Then from Definition 3.1.1, $\exists N$ such that $\forall n \geqslant N$ $a_{n+1}/a_n < r$. Hence, for $n > N$,

$$a_n = \frac{a_n}{a_{n-1}}\frac{a_{n-1}}{a_{n-2}}\cdots\frac{a_{N+1}}{a_N}a_N < a_N r^{n-N} \Rightarrow a_n < kr^n$$

where $k$ is independent of $n$. But since $\sum kr^n$ converges, so does $\sum a_n$ (Lemma 3.1.9 (2)).

If $l > 1$, pick $r > 1$ such that $1 < r < l$. Then, $\exists N$ such that $\forall n \geqslant N$ $a_{n+1}/a_n > r$ and

$$a_n = \frac{a_n}{a_{n-1}}\frac{a_{n-1}}{a_{n-2}}\cdots\frac{a_{N+1}}{a_N}a_N > a_N r^{n-N}$$

But $r^{n-N} \to \infty$ as $n \to \infty$. Thus $a_n \to \infty$ and $\sum a_n$ diverges. ■

Here are some examples for the tests above.

- $\sum(n/2^n)$. Can show the covergence by both ratio and root test.
- $\sum 1/n$ diverges but ratio test and root test both yields 1 (inconclusive).
- $\sum 1/n^2$ converges but ratio test and root test both yields 1 (inconclusive).

*Remark* 3.1.2. To show $n^{1/n} \to 1$, write $n^{1/n} = 1 + \delta_n$ ($\delta_n > 0$). Then, binomial expansion gives

$$n = (1+\delta_n)^n > \frac{n(n-1)}{2}\delta_n^2$$

and

$$0 < \delta_n^2 < \frac{2}{n+1} \Rightarrow \delta_n = 0$$

**Theorem 3.1.14.** (*Cauchy's Condensation Test*) Let $(a_n)$ be a decreasing sequence of positive terms. Then $\sum_{n=1}^{\infty} a_n$ converges if and only if

$$\sum_{n=1}^{\infty} 2^n a_{2^n}$$

converges.

*Proof.* Since $a_n$ is decreasing, note that

$$a_{2^k} \leqslant a_{2^{k-1}+i} \leqslant a_{2^{k-1}}$$

for $1 \leqslant i \leqslant 2^{k-1}$. Let us prove each direction seperately.

($\Rightarrow$) Suppose $\sum a_n$ converges. But

$$\begin{aligned} 2^{k-1}a_{2^k} &= a_{2^k} + \cdots + a_{2^k} \\ &\leqslant a_{2^{k-1}+1} + a_{2^{k-1}+2} + \cdots + a_{2^{k-1}+2^{k-1}} \\ &= \sum_{n=2^{k-1}+1}^{2^k} a_n \end{aligned}$$

Hence,

$$\sum_{k=1}^{K} 2^{k-1}a_{2^k} \leqslant \sum_{k=1}^{K}\left(\sum_{n=2^{k-1}+1}^{2^k} a_n\right) = \sum_{n=2}^{2^K} a_n \leqslant \sum_{n=1}^{\infty} a_n$$

and $\sum_{k=1}^{K} 2^k a_{2^k}$ converges since it is increasing in $K$ and bounded above.

($\Leftarrow$) Suppose $\sum 2^n a_{2^n}$ converges. Then,

$$\begin{aligned} \sum_{n=2^{k-1}+1}^{2^k} a_n &= a_{2^{k-1}+1} + a_{2^{k-1}+2} + \cdots + a_{2^{k-1}+2^{k-1}} \\ &\leqslant a_{2^{k-1}} + \cdots + a_{2^{k-1}} \\ &= 2^{k-1}a_{2^{k-1}} \end{aligned}$$

This implies

$$\sum_{n=2}^{2^K} a_n = \sum_{k=1}^{K}\left(\sum_{n=2^{k-1}+1}^{2^k} a_n\right) \leqslant \sum_{k=1}^{K} 2^{k-1}a_{2^{k-1}} \leqslant \sum_{n=1}^{\infty} 2^{k-1}a_{2^{k-1}}$$

Thus for any $N$, if we pick $K$ such that $2^K > N$,

$$\sum_{n=2}^{N} \leqslant \sum_{n=2}^{2^K} a_n \leqslant \sum_{n=1}^{\infty} 2^{k-1}a_{2^{k-1}}$$

Similarly, by monotone convergence theorem, $\sum_{n=2}^{N} a_n$ converges. ■

**Corollary 3.1.14.1.** (*p-series test*) $\sum_{n\geqslant 1}(1/n^p)$ (for $p > 0$) converges if and only if $p > 1$.

*Proof.* $a_n = 1/n^p$ is a decreasing sequence of positive numbers, since

$$\frac{n}{n+1} < 1 \Rightarrow \left(\frac{n}{n+1}\right)^p < 1 \Rightarrow \frac{1}{(n+1)^p} < \frac{1}{n^p}$$

Meanwhile,

$$2^n a_{2^n} = 2^n\left(\frac{1}{2^n}\right)^p = 2^{n-np} = \left(2^{1-p}\right)^n = r^n$$

We know that $\sum r^n$ converges if and only if $r < 1$, which is equivalent to $p > 1$. Hence, by Cauchy's condensation test, $\sum_{n\geqslant 1}(1/n^p)$ (for $p > 0$) converges if and only if $p > 1$. ■

### ∗ Alternating series

**Theorem 3.1.15.** (*Alternating Series Test*) Suppose $(a_n)$ is a decreasing sequence, tending to 0 as $n \to \infty$. Then, the series

$$\sum_{n=1}^{\infty}(-1)^{n+1}a_n$$

converges.

*Proof.* Let

$$S_n = \sum_{k=1}^{n}(-1)^{k+1}a_k$$

Then we can find $S_{2n}$ is increasing, since

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

Similarly,

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

gives

$$S_{2n+1} \leqslant S_{2n-1} \leqslant \cdots \leqslant S_3 \leqslant S_1$$

Furthermore, $S_{2n+1} = S_{2n} + a_{2n+1} \leqslant S_{2n}$, so

$$S_2 \leqslant S_4 \leqslant \cdots \leqslant S_{2n} \leqslant S_{2n+1} \leqslant S_{2n-1} \leqslant \cdots \leqslant S_3 \leqslant S_1$$

See that $(S_{2n})$ is increasing, bounded above by $S_1$; so $S_{2n} \to S$; and $(S_{2n+1})$ is decreasing, bounded below by $S_2$, so $S_{2n+1} \to \tilde{S}$. But

$$S_{2n+1} = S_{2n} + a_{2n+1}$$

implies $\tilde{S} = S$. Thus, given $\epsilon > 0$ there exists $N_1, N_2$ such that

$$|S_{2n} - S| < \epsilon \quad \forall n \geqslant N_1$$

$$|S_{2n+1} - S| < \epsilon \quad \forall n \geqslant N_2$$

Therefore, if we choose $n \geqslant 2\max\{N_1, N_2\} + 1 = N$, $|S_n - S| < \epsilon \Rightarrow S_n \to S$. ■

One may easily see that $\sum(-1)^{n+1}/n$ converges using the alternating series test.

### ⁎ Absolute and conditional convergence

**Definition 3.1.5.** Let $a_n \in \mathbb{C}$. If $\sum_{n=1}^{\infty} |a_n|$ converges, we say $\sum_{n=1}^{\infty} a_n$ is *absolutely convergent*.

Note that since $|a_n| \geqslant 0$, previous tests can be applied to show absolute convergence.

**Theorem 3.1.16.** If $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, then it is convergent.

*Proof.* Let

$$S_n = \sum_{k=1}^{n} a_k, \quad \bar{S}_n = \sum_{k=1}^{n} |a_k|$$

Suppose that $n \leqslant m$. Then

$$|S_m - S_n| = \left| \sum_{k=n+1}^{m} a_k \right| \leqslant \sum_{k=n+1}^{m} |a_k| = |\bar{S}_m - \bar{S}_n|$$

$(\bar{S}_n)$ is convergent, hence Cauchy (Lemma 3.1.7), so given $\epsilon > 0$ $\exists N$ such that $\forall m \geqslant n \geqslant N$ we have

$$|\bar{S}_m - \bar{S}_n| < \epsilon \Rightarrow |S_m - S_n| < \epsilon$$

Thus $(S_n)$ is Cauchy and it converges (Theorem 3.1.8). ■

For example of absolute convergence, consider $\sum_{n=1}^{\infty} a_n$ where $z \in \mathbb{C}$ and $a_n = (z^n/2^n)$. $|a_n| = (|z|/2)^n$ so $\sum |a_n|$ converges if and only if $|z|/2 < 1$. This gives absolute convergence for $|z| < 2$. If $|z| \geqslant 2$, $|a_n| \geqslant 1$ and there is no absolute convergence. Finally, note that coverse of Theorem 3.1.16 is false.

**Definition 3.1.6.** If $\sum a_n$ converges, but $\sum |a_n|$ does not, we say series is *conditionally convergent*.

N.b. order of terms matters. If rearranged, the sum may change. Consider the rearranged sequences below.

(1) $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$

(2) $1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} - \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \cdots$

If $S_n$ is partial sum of (1), $T_n$ partial sum of (2), then

$$S_n \to S > 0$$
$$T_n \to \frac{3S}{2} \neq S$$

**Definition 3.1.7.** Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a bijection, then $(a'_n)$ with $a'_n = a_{\sigma(n)}$ is a *rearrangement* of $(a_n)$.

**Theorem 3.1.17.** If $a'_n = a_{\sigma(n)}$ is a rearrangement of $(a_n)$, and $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, then

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a'_n$$

*Proof.* Fix $\epsilon > 0$. Since $\sum_{n=1}^{\infty} |a_n|$ converges, $\exists N$ such that

$$\sum_{n=N+1}^{\infty} |a_n| < \epsilon$$

Pick $M$ such that $\sigma^{-1}(k) < M$ for $k = 1, \ldots, N$. Then if $m \geqslant M$,

$$\sum_{n=1}^{\infty} a_n - \sum_{n=1}^{m} a_{\sigma(n)} = \sum_{n \in K_m} a_n$$

where $K_m = \{N+1, N+2, \dots\} \setminus \{\text{finitely many points}\}$. So

$$\left|\sum_{n=1}^{\infty} a_n - \sum_{n=1}^{m} a'_n\right| \leqslant \sum_{n \in K_m} |a_n| \leqslant \sum_{n \geqslant N+1} |a_n| < \epsilon$$

i.e. $\sum_{n=1}^{m} a'_n \to \sum_{n=1}^{\infty} a_n$ as $m \to \infty$. ■

## 3.2 Continuity

### 3.2.1 Continuity of a Function

Suppose $\mathrm{E} \subseteq \mathbb{C}$ is non-empty, and $f : \mathrm{E} \to \mathbb{C}$ is any function (includes case where $\mathrm{E} \subseteq \mathbb{R}$, $f : \mathrm{E} \to \mathbb{R}$).

**Definition 3.2.1.** $f$ is *continuous* at $a \in \mathrm{E}$ if: given $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that

$$|f(z) - f(a)| < \epsilon$$

for all $z \in \mathrm{E}$ with $|z - a| < \delta$, i.e. "*Points close to a in the domain are mapped close to $f(a)$ in range.*"

We say $f$ is continuous on E if $f$ is continuous at $a$ for all $a \in \mathrm{E}$.

**Theorem 3.2.1.** $f : \mathrm{E} \to \mathbb{C}$ is continuous at $a \in \mathrm{E}$ if and only if

$$f(z_n) \to f(a)$$

for all sequences $(z_n)$ with $z_n \in \mathrm{E}$, $z_n \to a$.

*Proof.* ($\Rightarrow$) Suppose $f$ is continuous at $a$, and $(z_n)$ is a sequence $z_n \in \mathrm{E}$ with $z_n \to a$. Let $\epsilon > 0$. Then $\exists \delta > 0$ such that

$$z \in \mathrm{E}, |z - a| < \delta \Rightarrow |f(z) - f(a)| < \epsilon$$

Since $z_n \to a$, there exists $N$ such that

$$n \geqslant N \Rightarrow |z_n - a| < \delta \Rightarrow |f(z) - f(a)| < \epsilon$$

Therefore $f(z_n) \to f(a)$.

($\Leftarrow$) Suppose $f$ is not continuous at $a$. Then, $\exists \epsilon > 0$ such that $\forall \delta > 0$, $\exists z \in \mathrm{E}$ with

$$|z - a| < \delta \text{ and } |f(z) - f(a)| \geqslant \epsilon$$

apply this with

$$\delta = 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots$$

to find for each $n$ and $z_n \in \mathrm{E}$ with

$$|z_n - a| < \frac{1}{n} \text{ and } |f(z_n) - f(a)| \geqslant \epsilon > 0$$

Then $z_n \to a$ as $n \to \infty$ but $f(z_n) \nrightarrow f(a)$, which is a contradiction. Hence $f$ is continuous at $a$. ■

This means we could alternatively take our definition of continuity as:

$$f(z_n) \to f(a)$$

for all sequences $(z_n)$ with $z_n \in \mathrm{E}$, $z_n \to a$.

**Lemma 3.2.2.** Suppose $f, g : \mathrm{E} \to \mathbb{C}$ are continuous at $a \in \mathrm{E}$. Then so are the functions

(1) $f(z) + g(z)$,

(2) $f(z)g(z)$,

(3) $\lambda g(z)$ (constant $\lambda \in \mathbb{C}$),

(4) $1/f(z)$ provided $f(z) \neq 0\ \forall z \in \mathrm{E}$.

*Proof.* Using the sequential characterisation of continuity, this follows easily from analogous results for sequences (Lemma 3.1.3). E.g. if $z_n \to a$,

$$f(z_n) \to f(a) \text{ and } g(z) \to g(a) \Rightarrow f(z_n) + g(z_n) \to f(a) + g(a)$$

etc. ■

Note that Lemma 3.2.2 can also be directly proved from Definition 3.2.1. We now show that composition of continuous functions is continuous.

**Theorem 3.2.3.** Suppose $A, B \subseteq \mathbb{C}$, $f : A \to B$, $g : B \to C$ and that $f$ is continuous at $a \in A$, $g$ is continuous at $f(a) \in B$. Then $g \circ f : A \to \mathbb{C}$ is continuous at $a$.

*Proof.* Suppose $(z_n)$ is any sequence with $z_n \in A$, $z_n \to a$. Then by continuity of $f$ at $a$,

$$f(z_n) \to f(a)$$

By continuity of $g$ at $f(a)$,

$$g(f(z_n)) \to g(f(a)) \Leftrightarrow g \circ f(z_n) \to g \circ f(a)$$

Hence $g \circ f$ is continuous at $a$. ■

Here are some examples of continuous (and discontinuous) functions.

(1) $f(z) = z$ is continuous at all points of $\mathbb{C}$.

(2) By Lemma 3.2.2 and Example (1) above, any polynomial in $z$ is continuous at all points of $\mathbb{C}$.

(3) $f(z) = |z|$ is continuous at all points of $\mathbb{C}$.[2]

(4) $f : \mathbb{R} \to \mathbb{R}$,

$$x \mapsto \begin{cases} 0 & x < 0 \\ 1 & x \geqslant 0 \end{cases}$$

is not continuous at $x = 0$.

(5) $f : \mathbb{R} \to \mathbb{R}$,

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

$\sin x$ is continuous,[3] so if $x \neq 0$ then Theorem 3.2.3 and Lemma 3.2.2 imply $f$ is continuous at $x$. However, $f$ is discontinuous at $x = 0$. Choose $x_n$ such that

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

which imply $f(x_n) = 1$, $x_n \to 0$, $f(0) = 0 \neq \lim_{n\to\infty} f(x_n)$.

(6) $f : \mathbb{R} \to \mathbb{R}$,

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

$f$ is continuous at $x \neq 0$ by same reasons as (5). But $f$ is also continuous at 0 this time. Suppose $x_n \to 0$. Then

$$|f(x_n)| = |x_n| \left|\sin \frac{1}{x_n}\right| \leqslant |x_n|$$

So $|f(x_n)| \leqslant |x_n|$ and $f(x_n) \to 0 = f(0)$.

[2] See reverse triangle inequality.

[3] See later sections for proof.

(7) (*Dirichlet Function*) $f : \mathbb{R} \to \mathbb{R}$,

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

is discontinuous at every point. If $x \in \mathbb{Q}$, take sequence $x_n \notin \mathbb{Q}$, then $x_n \to x$ but

$$f(x_n) = 0 \nrightarrow f(x) = 1$$

If $x \notin \mathbb{Q}$, take $x_n \in \mathbb{Q}$, $x_n \to x$ but

$$f(x_n) = 1 \nrightarrow f(x) = 0$$

### 3.2.2 Limit of a Function

Suppose $E \subseteq \mathbb{C}$, $f : E \to \mathbb{C}$. We want to define

$$\lim_{z \to a} f(z)$$

even when $a$ may not belong to E. For instance, consider $f : \mathbb{C} \backslash \{0\} \to \mathbb{C}$,

$$z \mapsto \sin z / z$$

What is $\lim_{z \to 0} f(z)$? – However, this does not always make sense. If $E = \{0\} \cup [1, 2]$, for $f : E \to \mathbb{R}$ it is impossible to consider $\lim_{x \to 0} f(x)$, because there are no point near 0 except 0 itself.

**Definition 3.2.2.** Suppose $E \subseteq \mathbb{C}$, $a \in \mathbb{C}$. We say that $a$ is a *limit point* of E if for any $\delta > 0$ there exists $z \in E$ such that

$$0 < |z - a| < \delta$$

N.b. $a$ is a limit point of E if and only if there exists a sequence $(z_n)$ such that $z_n \in E$, $z_n \neq a$, and

$$z_n \to a$$

**Definition 3.2.3.** Suppose $E \subseteq C$, $f : E \to \mathbb{C}$ and let $a \in \mathbb{C}$ be a limit point of E. We say

$$\lim_{z \to a} f(z) = l$$

if: given $\epsilon > 0$, $\exists \delta > 0$ such that $\forall z \in E$,

$$0 < |z - a| < \delta \Rightarrow |f(z) - l| < \epsilon$$

**Lemma 3.2.4.** If $f, E, a$ as above, then

$$\lim_{z \to a} f(z) = l$$

if and only if $f(z_n) \to l$ for all sequences $z_n \in E$, $z_n \neq a$, $z_n \to a$.

Observe that the definitions immediately tell us that if $a \in E$ is a limit point then $f$ is continuous at $a$ if and only if

$$\lim_{z \to a} f(z) = f(a)$$

If $a \in E$ is *isolated*, i.e. not a limit point, then $f$ is always continuous at $a$.

Note that limit of functions behave similarly to limits of sequences.

**Lemma 3.2.5.** Suppose $E \subseteq \mathbb{C}$, $a \in \mathbb{C}$ is a limit point, and $f, g : E \to \mathbb{C}$

(1) The limit is unique, i.e. if $f(z) \to A$ and $f(z) \to B$ as $z \to a$, then $A = B$.

(2) If $f(z) \to l$, $g(z) \to k$, then$f(z) + g(z) = l + k$ as $z \to a$;

(3) $f(z)g(z) = lk$ as $z \to a$;

(4) $f(z)/g(z) = l/k$ as $z \to a$ (if $k \neq 0$).

*Proof.* (of (1))

$$|A - B| = |A - f(z) + f(z) - B| \leqslant |A - f(z)| + |f(z) - B|$$

for all $z \in \mathrm{E}$, $z \neq a$. Given $\epsilon > 0$, $\exists \delta_1 > 0$ such that

$$0 < |z - a| < \delta_1, z \in \mathrm{E} \Rightarrow |f(z) - A| < \epsilon,\ z \in \mathrm{E}$$

Also $\exists \delta_2 > 0$ such that

$$0 < |z - a| < \delta_2, z \in \mathrm{E} \Rightarrow |f(z) - B| < \epsilon,\ z \in \mathrm{E}$$

Since $a$ is a limit point, $\exists z \in \mathrm{E}$ such that

$$0 < |z - a| < \min\{\delta_1, \delta_2\} \Rightarrow |A - B| < 2\epsilon$$

Because $\epsilon$ is arbitrary, $A = B$. ■

**Theorem 3.2.6.** (*Intermediate Value Theorem*) Suppose $f : [a,b] \to \mathbb{R}$ is continuous and $f(a) < f(b)$. Then for any $\eta$ with $f(a) < \eta < f(b)$ there exists $c \in [a,b]$ such that $f(c) = \eta$.

*Proof.* Let

$$S = \{x \in [a,b] \mid f(x) < \eta\}$$

Clearly $S$ is not empty as $a \in S$. $S$ is bounded above by $b$. So $S$ has a supremum, $c$ (Theorem 3.1.2). We claim $f(c) = \eta$. From the definition of a supremum (Definition 3.1.2), for each $n \in \mathrm{N}$ there exists $x_n \in S$ with

$$c - \frac{1}{n} \leqslant x_n \leqslant c$$

Hence $x_n \to c$. But $f$ is continuous so $f(x_n) \to f(c)$ and

$$f(x_n) < \eta \Rightarrow f(c) \leqslant \eta$$

In particular, $c \neq b$. Now let

$$\tilde{x}_n = c + \frac{1}{n}$$

For $n$ large enough, $\tilde{x}_n \in [a,b]$, and $\tilde{x}_n \to c$. Furthermore, $f(\tilde{x}_n) \geqslant \eta$ since

$$\tilde{x}_n > c \Rightarrow \tilde{x}_n \notin S$$

Hence

$$f(c) = \lim_{n\to\infty} f(\tilde{x}_n) \geqslant \eta$$

and thus $f(c) = \eta$. ■

This theorem is useful to find roots of functions. For example, we can check existence of $N^{\text{th}}$ roots. Suppose $y > 0$. Consider for $N \in \mathbb{N}$ the function $f : [0, 1+y] \to \mathbb{R}$,

$$x \mapsto x^N$$

Then since

$$(a + y)^N \geqslant 1 + Ny > y$$

we have

$$f(0) < y < f(1+y)$$

and there exists $c \in (0, 1+y)$ such that $f(c) = y$ by intermediate value theorem (IVT). $c$ is a *positive* $N^{th}$ *root* of $y$. In fact, since

$$y_1 < y_2 \Rightarrow f(y_1) < f(y_2)$$

$c$ is unique.

**Lemma 3.2.7.** (*Bounds on Continuous Function*) Suppose $f : [a,b] \to \mathbb{R}$ is continuous. Then there exist $K$ such that

$$|f(x)| \leqslant K \quad \forall x \in [a,b]$$

*Proof.* Suppose not. Then for each $n = 1, 2, \dots$ we can find $x_n \in [a,b]$ with $|f(x_n)| > n$. By Bolzano-Weierstrass theorem (Theorem 3.1.6), $(x_n)$ is a bounded sequence, and hence has a subsequence $x_{n_j} \to x$ for some $x$. Moreover, since

$$a \leqslant x_{n_j} \leqslant b \Rightarrow a \leqslant x \leqslant b$$

so $x \in [a,b]$. But

$$|f(x_{n_j})| > n_j \geqslant j$$

so $f(x_{n_j})$ cannot converge. This contradicts the assumption that $f$ is continuous as $f(x_{n_j}) \nrightarrow f(x)$. ■

**Theorem 3.2.8.** (*Extreme Value Theorem*) Suppose $f : [a,b] \to \mathbb{R}$ is continuous. Then there exist $y, Y \in [a,b]$ such that

$$f(y) \leqslant f(x) \leqslant f(Y) \quad \forall x \in [a,b]$$

*Proof.* The set

$$A = \{f(x) \mid x \in [a,b]\}$$

is bounded by Lemma 3.2.7, so has a supremum $M = \sup A$. From Definition 3.1.2, $M - 1/n$ is not an upper bound of $A$ for $n \in \mathbb{N}$. So $\exists y_n \in A$ with

$$M - \frac{1}{n} \leqslant y_n \leqslant M$$

From definition of $A$, there exists $x_n \in [a,b]$ with

$$y_n = f(x_n) \Rightarrow M - \frac{1}{n} \leqslant f(x_n) \leqslant M$$

$(x_n)$ is a bounded sequence. So by Bolzano-Weierstrass theorem we can pick a subsequence $(x_{n_j})$ such that

$$x_{n_j} \to y$$

for some $y \in [a,b]$. Also $f(x_{n_j}) \to M$, so by continuity of $f$,

$$f(y) = \lim_{j \to \infty} f(x_{n_j}) = M$$

It follows immediately that

$$f(x) \leqslant f(y) \quad \forall x \in [a,b]$$

For lower bound, we can either consider $\inf A$ and argue similarly, or apply result already established to $-f$. ■

Note that, for Lemma 3.2.7 and Theorem 3.2.8, it is crucial that $f$ is defined on a closed bounded interval: e.g. consider

- $f : (0,1] \to \mathbb{R}$, $x \mapsto 1/x$, continuous, but unbounded.
- $f : [0,\infty) \to \mathbb{R}$, $x \mapsto x$, continuous, but unbounded.

### 3.2.3 Inverse Function Theorem

**Definition 3.2.4.** Function $f : [a, b] \to \mathbb{R}$ is

- *increasing* if $a \leqslant x_1 < x_2 \leqslant b \Rightarrow f(x_1) \leqslant f(x_2)$;
- *strictly increasing* if $a \leqslant x_1 < x_2 \leqslant b \Rightarrow f(x_1) < f(x_2)$.

Decreasing and strictly decreasing are defined similarly.

**Theorem 3.2.9.** Suppose $f : [a, b] \to \mathbb{R}$ is continuous and strictly increasing. Let $c = f(a)$, $d = f(b)$. Then

$$f : [a, b] \to [c, d]$$

is a bijection and

$$f^{-1} : [c, d] \to [a, b]$$

is continuous and strictly increasing.

*Proof.* First we note that $f$ is injective. If $x \neq y$, then without loss of generality

$$a \leqslant x < y \leqslant b \Rightarrow f(x) < f(y) \Rightarrow f(x) \neq f(y)$$

Next, $f$ maps into $[c, d]$. If $a \leqslant x \leqslant b$, then

$$c = f(a) \leqslant f(x) \leqslant f(b) = d$$

Also, $f$ is surjective by intermediate value theorem: if $\eta \in [c, d]$ then $\exists x \in [a, b]$ with $f(x) = \eta$.

To show $f^{-1}$ is continuous, fix $z \in (c, d)$ and let $y = f^{-1}(z) \in (a, b)$. Let $\epsilon > 0$ to be sufficiently small that

$$a \leqslant y - \epsilon < y < y + \epsilon \leqslant b \Rightarrow c \leqslant f(y - \epsilon) < f(y) < f(y + \epsilon) \leqslant d$$

Pick $\delta > 0$ such that $(z - \delta, z + \delta) \subset (f(y - \epsilon), f(y + \epsilon))$. Since $f : (y - \epsilon, y + \epsilon) \to (f(y - \epsilon), f(y + \epsilon))$ is bijective, if $|z' - z| < \delta$, then

$$\begin{aligned} z' \in (f(y - \epsilon), f(y + \epsilon)) &\Rightarrow f^{-1}(z') \in (y - \epsilon, y + \epsilon) \\ &\Rightarrow |f^{-1}(z') - f^{-1}(z)| < \epsilon \end{aligned}$$

If $z = c$ or $d$ similar argument works. Therefore $f^{-1}$ is continuous.

To see $f^{-1}$ is increasing, suppose not. Then $\exists c \leqslant z_1 < z_2 \leqslant d$ with $f^{-1}(z_1) \geqslant f^{-1}(z_2)$. But $f$ is strictly increasing; this implies

$$f\left(f^{-1}(z_1)\right) \geqslant f\left(f^{-1}(z_2)\right) \Rightarrow z_1 \geqslant z_2$$

which is a contradiction. Thus $f^{-1}$ is strictly increasing. ∎

## 3.3 Differentiability

In this section, let

$$f : \mathrm{E} \to \mathbb{C}$$

where $\mathrm{E} \subseteq \mathbb{C}$, while mostly E will be a real interval.

**Definition 3.3.1.** Let $x \in \mathrm{E}$ be a limit point. $f$ is said to be *differentiable* at $x$, with derivative $f'(x)$ if

$$\lim_{y \to x} \frac{f(y) - f(x)}{y - x} = f'(x)$$

exists.

If $f$ is differentiable at each $x \in \mathrm{E}$, we say $f$ is differentiable on E. Also, we will assume from now on E has no isolated points (set is either interval or disc).

*Remark* 3.3.1.

(1) Other common notations are
$$\frac{df}{dx}, \frac{dy}{dx}$$
etc.

(2) We can equivalently write
$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

(3) Another ways to phrase definition: let $\epsilon(h) = f(x+h) - f(x) - hf'(x)$; then
$$\lim_{h\to 0} \frac{\epsilon(h)}{h} = 0$$
We say, $f$ has a *best affine approximation* near $x$.

(4) If $f$ is differentiable at $x$, then $f$ is continuous at $x$, since $\epsilon(h) \to 0$ as $h \to 0$ giving $f(x+h) \to f(x)$ as $h \to 0$.

Let's look at some examples.

(1) $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x$ is differentiable at each $x$ and $f'(x) = 1$, i.e.
$$f(x+h) = x+h = f(x) + h \cdot 1 + 0$$
where $1 = f'(x)$ and $0 = \epsilon(h)$.

(2) $f : \mathbb{R} \to \mathbb{R}$, $x \to |x|$ is differentiable for $x \neq 0$. At $x = 0$ consider
$$\frac{f(h)-f(0)}{h} = \frac{|h|}{h}$$
We have
$$\lim_{h_n \to 0+} \frac{f(h)-f(0)}{h} = 1$$
but
$$\lim_{h_n \to 0-} \frac{f(h)-f(0)}{h} = -1$$
hence the limit does not exist and $f$ is not differentiable at $x = 0$.

### 3.3.1 Differentiation of Sums and Products

**Proposition 3.3.1.**

(1) If $f(x) = c \ \forall x \in \mathrm{E}$ then $f$ is differentiable on E with $f'(x) = 0$.

(2) If $f, g$ are differentiable at $x$, so is $f + g$ and
$$(f+g)'(x) = f'(x) + g'(x)$$

(3) If $f, g$ are differentiable at $x$, so is $fg$ and
$$(fg)'(x) = f(x)g'(x) + f'(x)g(x)$$

(4) If $f$ is differentiable at $x$ and $f(y) \neq 0 \ \forall y \in E$ then $1/f$ is differentiable at $x$ and
$$\left(\frac{1}{f}\right)' = -\frac{f'(x)}{f(x)^2}$$

*Proof.*

(1) Can easily see

$$\lim_{h\to 0}\frac{c-c}{h}=0$$

■

(2) By Lemma 3.2.5, and from

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

we have

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

■

(3) Let $\phi(x)=f(x)g(x)$. Then for $h\neq 0$

$$\begin{aligned}\frac{\phi(x+h)-\phi(x)}{h}&=\frac{f(x+h)g(x+h)-f(x)g(x)}{h}\\&=\frac{f(x+h)g(x+h)-f(x+h)g(x)+f(x+h)g(x)-f(x)g(x)}{h}\\&=f(x+h)\left(\frac{g(x+h)-g(x)}{h}\right)+g(x)\left(\frac{f(x+h)-f(x)}{h}\right)\end{aligned}$$

implying

$$\frac{\phi(x+h)-\phi(x)}{h}\to f(x)g'(x)+f'(x)g(x)$$

■

(4) Let $\phi(x)=1/f(x)$. Then,

$$\frac{\phi(x+h)-\phi(x)}{h}=\frac{\frac{1}{f(x+h)}-\frac{1}{f(x)}}{h}=\frac{f(x)-f(x+h)}{h}\frac{1}{f(x)f(x+h)}$$

Again, by Lemma 3.2.5,

$$\frac{\phi(x+h)-\phi(x)}{h}\to-\frac{f'(x)}{f(x)^2}$$

■

Note that (3) and (4) give the quotient rule:

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

For example, consider $f_n(x)=x^n$ where $n\in\mathbb{N}$. Shown that $f_1(x)$ is differentiable and $f_1'(x)=1$, we claim $f_n(x)$ is differentiable with $f_n'(x)=nx^{n-1}$ for all $n$. If the claim is true for $f_{n-1}$, observe that

$$f_n(x)=x^{n-1}x=f_{n-1}(x)f_1(x)$$

By Proposition 3.3.1 (3), $f_n$ is differentiable, and

$$\begin{aligned}f_n'(x)&=f_1(x)f_{n-1}'(x)+f_1'(x)f_{n-1}(x)\\&=x(n-1)x^{n-2}+x^{n-1}=nx^{n-2}\end{aligned}$$

Hence, by induction, $f_n'(x)=nx^{n-1}$ for all $n\in\mathbb{N}$.

Now consider $f_{-n}(x)=x^{-n}$ where $n\in\mathbb{N}$. If $x\neq 0$ $f_{-n}(x)=1/f_n(x)$. So by Proposition 3.3.1 (4), $f_{-n}(x)$ is differentiable and

$$f_{-n}'(x)=-\frac{f_n'(x)}{f_n(x)^2}=-\frac{nx^{n-1}}{x^{2n}}=-nx^{-n-1}$$

Therefore $f_n'(x) = nx^{n-1}$ holds for all $n \in \mathbb{Z}$ (with condition $x \neq 0$ if $n < 0$).

Combining results above with Proposition 3.3.1, we see all polynomial functions are differentiable everywhere, and rational functions $f(x) = p(x)/q(x)$ ($p$, $q$ coprime polynomials) are differentiable everywhere except the roots of $q$.

**Theorem 3.3.2.** (*Chain Rule*) Suppose $U, V \subset \mathbb{C}$ and $f : U \to V$, $g : V \to \mathbb{C}$ are such that $f$ is differentiable at $a \in U$ and $g$ is differentiable at $f(a) \in V$. Then, $g \circ f : U \to \mathbb{C}$ is differentiable at $a$ and

$$(g \circ f)'(a) = g'(f(a))f'(a)$$

*Proof.* Let $f(a) = b$. We can restate the differentiability of $f$ at $a$, $g$ at $b$ as:

$$f(x) = f(a) + (x-a)f'(a) + \epsilon_f(x)(x-a)$$

$$g(y) = g(b) + (y-b)g'(b) + \epsilon_g(y)(y-b)$$

with $\lim_{x\to a} \epsilon_f(x) = \lim_{y\to b} \epsilon_g(y) = 0$. Notice we can set $\epsilon_f(a) = 0$, $\epsilon_g(b) = 0$ to make $\epsilon_f$, $\epsilon_g$ continuous at $a$, $b$ respectively. Now set $y = f(x)$ and write

$$\begin{aligned} g(f(x)) &= g(f(a)) + (f(x) - f(a))g'(f(a)) + \epsilon_g(f(x))(f(x) - f(a)) \\ &= g(f(a)) + \big((x-a)f'(a) + \epsilon_f(x-a)\big)\, g'(f(a)) \\ &\quad + \epsilon_g(f(x))\big((x-a)f'(a) + \epsilon_f(x)(x-a)\big) \end{aligned}$$

Thus

$$\begin{aligned} g(f(x)) &= g(f(a)) + (x-a)f'(a)g'(f(a)) + (x-a)\left[g'(f(a))\epsilon_f(x) + f'(a)\epsilon_g(f(x)) + \epsilon_g(f(x))\epsilon_f(x)\right] \\ &= g(f(a)) + (x-a)f'(a)g'(f(a)) + (x-a)\sigma(x) \end{aligned}$$

where

$$\sigma(x) = g'(f(a))\epsilon_f(x) + f'(a)\epsilon_g(f(x)) + \epsilon_g(f(x))\epsilon_f(x)$$

Note that $\epsilon_f(x) \to 0$ as $x \to a$, and $\epsilon_g(f(x))$ is composition of continuous functions, so $\epsilon_g(f(x)) \to 0$ as $x \to a$. Hence $\sigma(x) \to 0$ as $x \to a$ and therefore $(g \circ f)'(a) = g'(f(a))f'(a)$ from definition of differentiation. ■

*Remark* 3.3.2. Chain rule is often written as

$$\frac{dy}{dx} = \frac{dy}{dt}\frac{dt}{dx}$$

where $y(t(x))$.

### 3.3.2 The Mean Value Theorem

**Theorem 3.3.3.** (*Rolle's Theorem*) Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on $(a,b)$. If $f(a) = f(b)$ then $\exists c \in (a,b)$ such that $f'(c) = 0$.

*Proof.* By Theorem 3.2.8 (EVT), $\exists y, Y \in [a,b]$ such that

$$m = f(y) \leqslant f(x) \leqslant f(Y) = M$$

$\forall x \in [a,b]$. If $m = M = f(a)$ then $f(x) = M$ and $f$ constant, so we can take any $c \in (a,b)$.

Otherwise, either $M > f(a)$ or $m < f(a)$. If $M > f(a)$, then $Y \in (a,b)$. We claim $f'(Y) = 0$. To see this, suppose $h_n \to 0+$. Then

$$\frac{f(Y+h_n) - f(Y)}{h_n} \leqslant 0 \Rightarrow \lim_{h\to 0} \frac{f(Y+h) - f(Y)}{h} \leqslant 0$$

If $h_n \to 0-$,

$$\frac{f(Y+h_n) - f(Y)}{h_n} \geqslant 0 \Rightarrow \lim_{h\to 0} \frac{f(Y+h) - f(Y)}{h} \geqslant 0$$

Hence by uniqueness of limit, $f'(Y) = 0$. If $m < f(a)$ a similiar argument shows $y \in (a,b)$ and $f'(y) = 0$. Regardless $\exists c \in (a,b)$ such that $f'(c) = 0$. ■

**Theorem 3.3.4.** (*Mean Value Theorem*) Let $f : [a,b] \to \mathbb{R}$ be a continuous function, differentiable on $(a,b)$. Then there exists $c \in (a,b)$ such that

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

*Proof.* Consider

$$\phi(x) = (x-a)(f(b)-f(a)) - (b-a)(f(x)-f(a))$$

Clearly, $\phi$ satisfies conditions of Rolle's theorem, so $\exists c \in (a,b)$ such that

$$\phi'(c) = 0 \Rightarrow 0 = (f(b)-f(a)) - (b-a)f'(c)$$

■

If $f$ is differentiable on an open interval $I$ containing $a$, another way to state mean value theorem (MVT) is: given $h$ such that $a+h \in I$, $\exists \Theta = \Theta(h) \in (0,1)$ such that

$$f(a+h) = f(a) + hf'(a+\Theta h)$$

The mean value theorem allows us to export local information about the derivative to global properties of the function.

**Corollary 3.3.4.1.** Let $f : [a,b] \to \mathbb{R}$ continuous, and differentiable on $(a,b)$. Then,

(1) if $f'(x) > 0$ $\forall x \in (a,b)$, $f$ is strictly increasing.

(2) if $f'(x) \geqslant 0$ $\forall x \in (a,b)$, $f$ is increasing.

(3) if $f'(x) = 0$ $\forall x \in (a,b)$, $f$ is constant on $[a,b]$

*Proof.*

(1) Let $a \leqslant x < y \leqslant b$. Hypotheses of mean value theorem apply to $f : [x,y] \to \mathbb{R}$ so $\exists c \in (x,y)$ such that

$$f(y) - f(x) = f'(c)(y-x) > 0$$

(2) Same as above, but use $f'(c) \geqslant 0$.

(3) Pick $x \in (a,b]$. Then, by applying mean value theorem on $[a,x]$, we deduce that there exists $c \in (a,x)$ such that

$$f(x) - f(a) = f'(c)(x-a) = 0$$

and $f$ is constant.

■

Meanwhile, note that the mean value theorem is not necessarily true on general sets: e.g. $f : \mathbb{Q} \to \mathbb{Q}$,

$$x \mapsto \begin{cases} 0 & x^2 < 2 \\ 1 & x^2 > 2 \end{cases}$$

Now we revisit the inverse function theorem.

**Theorem 3.3.5.** Suppose $f : [a,b] \to \mathbb{R}$ is continuous and differentiable on $(a,b)$, with $f'(x) > 0$ $\forall x \in (a,b)$. Let $f(a) = c$, $f(b) = d$. Then,

$$f : [a,b] \to [c,d]$$

is bijective and $f^{-1}$ is differentiable on $(c,d)$ with

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

*Proof.* By Corollary 3.3.4.1 $f$ is strictly increasing on $[a,b]$. By Theorem 3.2.9,

$$f : [a,b] \to [c,d]$$

is bijective and

$$f^{-1} : [c,d] \to [a,b]$$

is continuous and strictly increasing. However, it remains to show $f^{-1}$ is differentiable on $(c,d)$.

Let $y \in (c,d)$ and set $x = f^{-1}(y)$. Given $h$ such that $y+h \in (c,d)$, let $k$ be such that $y+h = f(x+k)$. Then,

$$\frac{f^{-1}(y+h) - f^{-1}(y)}{h} = \frac{x+k-x}{f(x+k)-y} = \frac{k}{f(x+k)-f(x)}$$

Fix $\epsilon > 0$; then by differentiability of $f$ and facts about limits, $\exists \delta > 0$ such that for all $0 < |k| < \delta$ we have

$$\left| \frac{k}{f(x+k)-f(x)} - \frac{1}{f'(x)} \right| < \epsilon$$

Since $f^{-1}$ is continuous, there exists $\delta'$ such that

$$\begin{aligned} 0 < |h| < \delta' &\Rightarrow 0 < |f^{-1}(y+h) - x| < \delta \\ &\Rightarrow 0 < k < \delta \\ &\Rightarrow \left| \frac{f^{-1}(y+h) - f^{-1}(y)}{h} - \frac{1}{f'(x)} \right| < \epsilon \end{aligned}$$

Therefore $f^{-1}$ is differentiable at $y$ and

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

■

With fixed $R > 0$, consider $f : [0,R] \to \mathbb{R}$, $x \mapsto x^n$ for $n \in \mathbb{N}$. $f$ is continuous on $[0,R]$, differentiable on $(0,R)$ with $f'(x) = nx^{n-1} > 0$ $(x \in (0,R))$. Hence by Theorem 3.3.5, $f$ maps $[0,R]$ bijectively onto $[0,R^n]$ and the inverse $g(y) = f^{-1}(y) = y^{1/n}$ is differentiable with

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

More generally, let $h(x) = x^r$ where $r \in \mathbb{Q}$. Writing $r = m/n$ for $m \in \mathbb{Z}$, $n \in \mathbb{N}$ and defining

$$x^r = \left( x^{1/n} \right)^m$$

$h$ is differentiable on $(0,\infty)$ and

$$h'(x) = m \left( x^{1/n} \right)^{m-1} \left( \frac{1}{n} x^{\frac{1}{n}-1} \right) = \frac{m}{n} x^{\frac{m}{n}-1} = rx^{r-1}$$

by chain rule.

**Theorem 3.3.6.** (*Cauchy's Mean Value Theorem*) Let $f,g : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on $(a,b)$. Then there exists $c \in (a,b)$ such that

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

which can be also written as

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

if both sides are well-defined.

*Proof.* Let

$$\phi(x) = (g(x) - g(a))(f(b) - f(a)) - (g(b) - g(a))(f(x) - f(a))$$

$\phi$ is continuous on $[a, b]$, differentiable on $(a, b)$ and $\phi(a) = \phi(b) = 0$. Hence by Rolle's theorem, there exists $c \in (a, b)$ such that

$$\phi'(c) = 0 \Rightarrow g'(c)(f(a) - f(b)) - f'(c)(g(a) - g(b)) = 0$$

■

A famous application of Cauchy's mean value theorem is the *L'Hôpital's rule*. Consider

$$\frac{e^x - 1}{\sin x}$$

as $x \to 0$.[4] With noting that

$$\frac{e^x - 1}{\sin x} = \frac{e^x - e^0}{\sin x - \sin 0}$$

apply Cauchy's mean value theorem with $f(x) = e^x$, $g(x) = \sin x$ on interval $[x, 0]$ or $[0, x]$ to deduce that $\exists \theta_x \in (0, 1)$ such that

$$\frac{e^x - 1}{\sin x} = \frac{e^{\theta_x x}}{\cos(\theta_x x)}$$

Now $0 < |\theta_x x| < |x|$ so as $x \to 0$ $\theta_x x \to 0$. Further, because $e^y / \cos y$ is continuous, we have

$$\frac{e^{\theta_x x}}{\cos(\theta_x x)} \to 1$$

as $x \to 0$. Thus

$$\lim_{x \to 0} \frac{e^x - 1}{\sin x} = 1$$

### 3.3.3 Higher Derivatives and Taylor's Theorem

Suppose $f : (a, b) \to \mathbb{R}$ is differentiable on $(a, b)$. Then we can consider the function

$$f' : (a, b) \to \mathbb{R}, \quad x \mapsto f'(x)$$

If this is differentiable at $x \in (a, b)$ we say $f$ is twice differentiable at $x$ and write

$$(f')'(x) = f''(x) = f^{(2)}(x)$$

We can iterate to define $k$-times differentiability and write

$$f^{(k)}(x) = \left(f^{(k-1)}\right)'(x)$$

Furthermore, function is said to be *smooth* if it is $k$-times differentiable on $(a, b)$ for every $k$. We would like to extend mean value theorem for a function which is more differentiable to incorporate higher derivatives.

**Theorem 3.3.7.** (*Taylor's Theorem with Lagrange Remainder*) Suppose $f$ and its derivatives to order $n - 1$ are continuous on $[a, a + h]$ and $f$ is $n$ times differentiable on $(a, a + h)$. Then there exists $\theta \in (0, 1)$ such that

$$f(a + h) = f(a) + hf'(a) + \frac{h^2}{2!} f''(a) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(a) + \frac{h^n}{n!} f^{(n)}(a + \theta h)$$

[4] We will formally define $e$ and sin later.

N.b. for $n = 1$ Theorem 3.3.7 is mean value theorem, so Theorem 3.3.7 is a $n^{\text{th}}$ order mean value theorem. We say

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

is *Lagrange's form of the remainder*.

Also, considering $g(x) = -f(x)$ and applying Theorem 3.3.7 at $x = -a$, we can show the result also holds if $h < 0$, provided conditions hold on $[a+h, a]$, $(a+h, a)$ etc. – which is able to be assured by stating $f$ is $n$ times differentiable on some interval $(c, d)$ with $[a, a+h] \subset (c, d)$ (or $[a+h, a]$).

*Proof.* (*Method 1*) Observe we can choose $a = 0$ without loss of generality; if we prove the result for $a = 0$, applying it to $g(x) = f(a + x)$ gives general case.

For $0 \leqslant t \leqslant h$ let

$$\phi(t) = f(t) - f(0) - tf'(0) - \cdots - \frac{t^{n-1}}{(n-1)!} f^{(n-1)}(0) - \frac{t^n}{n!} B$$

where we choose $B$ such that $\phi(h) = 0$, i.e.

$$\frac{h^n}{n!} B = f(h) - f(0) - tf'(0) - \cdots - \frac{t^{n-1}}{(n-1)!} f^{(n-1)}(0) \tag{$*$}$$

Now repeatedly apply Rolle's theorem (Theorem 3.3.3) to find another expression for B. First observe $\phi$ and its first $n-1$ derivatives are continuous on $[0, h]$, and $\phi^{(n)}$ exists on $(0, h)$. Also note that $\phi^{(k)}(0) = 0$ for $0 \leqslant k \leqslant n-1$. Then

$$\phi(0) = \phi(h) = 0 \Rightarrow \exists \theta_1 \in (0,1) \text{ such that } \phi'(\theta_1 h) = 0$$

$$\phi'(0) = \phi'(\theta_1 h) = 0 \Rightarrow \exists \theta_2 \in (0,1) \text{ such that } \phi''(\theta_2 \theta_1 h) = 0$$

$$\vdots$$

$$\phi^{(n-1)}(0) = \phi^{(n-1)}(\theta_{n-1} \cdots \theta_1 h) = 0 \Rightarrow \exists \theta_n \in (0,1) \text{ such that } \phi^{(n)}(\theta_n \theta_{n-1} \cdots \theta_1 h) = 0$$

Let $\theta = \theta_1 \theta_2 \cdots \theta_n \in (0, 1)$. Then

$$\phi^{(n)}(\theta h) = 0 \Rightarrow f^{(n)}(\theta h) - B = 0$$

i.e. $B = f^{(n)}(\theta h)$ for some $\theta \in (0,1)$. Rearranging equation $(*)$ gives us the desired result. ■

*Proof.* (*Method 2*) Again, assume $a = 0$ without loss of generality. This time, for $0 \leqslant t \leqslant h$, let

$$F(t) = f(h) - f(t) - (h-t)f'(t) - \frac{(h-t)^2}{2!} f''(t) - \cdots - \frac{(h-t)^{n-1}}{(n-1)!} f^{(n-1)}(t)$$

F is continuous on $[0, h]$ and differentiable on $(0, h)$, and

$$\begin{aligned} F'(t) &= -f'(t) + f'(t) - (h-t)f''(t) + (h-t)f''(t) - \frac{(h-t)^2}{2!} f^{(3)}(t) + \cdots - \frac{(h-t)^{n-1}}{(n-1)!} f^{(n)}(t) \\ &= -\frac{(h-t)^{n-1}}{(n-1)!} f^{(n)}(t) \end{aligned}$$

Now set

$$\phi(t) = F(t) - \left(\frac{h-t}{h}\right)^p F(0)$$

with $1 \leqslant p \leqslant n$, $p \in \mathbb{N}$. Then $\phi(0) = 0 = \phi(h)$. So by Rolle's theorem there exists $\theta \in (0,1)$ such that $\phi'(\theta h) = 0$. But

$$\phi'(\theta h) = F'(\theta h) + \frac{p(1-\theta)^{p-1}}{h} F(0)$$

Hence

$$\begin{aligned}0 = &-\frac{h^{n-1}(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(\theta h)\\ &+\frac{p(1-\theta)^{p-1}}{h}\left(f(h)-f(0)-hf'(0)-\frac{h^2}{2!}f''(0)-\cdots-\frac{h^{n-1}}{(n-1)!}f^{(n-1)}(0)\right)\end{aligned}$$

If $p=n$, we can rearrange to find Taylor's theorem with Lagrange remainder. Otherwise, if $p=1$, we have proved the *Taylor's theorem with Cauchy remainder*. ■

**Theorem 3.3.8.** (*Taylor's Theorem with Cauchy Remainder*) With the same hypotheses as Theorem 3.3.7, we have

$$f(a+h)=f(a)+hf'(a)+\cdots+\frac{h^{n-1}}{(n-1)!}f^{(n-1)}f^{(n-1)}(a)+\tilde{R}_n$$

where

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

for some $\theta\in(0,1)$.

Both versions of Taylor's theorem assert that

$$f(h)=P_{n-1}(h)+R_n(h)$$

where

$$P_{n-1}(h)=\sum_{i=0}^{n-1}\frac{h^i}{i!}f^{(i)}(0)$$

the *Taylor polynomial*, and

$$R_n(h)=\begin{cases}\frac{h^n}{n!}f^{(n)}(a+\theta h) & \text{(Lagrange)}\\ \frac{(1-\tilde{\theta})^{n-1}h^n}{(n-1)!}f^{(n)}(a+\tilde{\theta}h) & \text{(Cauchy)}\end{cases}$$

for some $\theta,\tilde{\theta}\in(0,1)$. Finally, be aware that Taylor's theorem does not say

$$f(h)=\sum_{i=0}^{\infty}\frac{h^i}{i!}f^{(i)}(0)$$

where the right hand side is called the *Taylor series*. In fact this is not always true, even if $f$ is smooth and even if the series converges. To show a function is equal to its Taylor series, we need to show $R_n(h)\to 0$ as $n\to\infty$ for fixed $h$. For example, consider the binomial series $f(x)=(1+x)^r$ where $r\in\mathbb{Q}$. Claim that if $|x|<1$, then

$$(1+x)^r=1+\binom{r}{1}x+\cdots+\binom{r}{n}x^n+\cdots$$

where we define

$$\binom{r}{n}=\frac{r(r-1)\cdots(r-n+1)}{n!}$$

where the series converges absolutely.

*Proof.* Clearly,

$$f^{(n)}(x)=r(r-1)\cdots(r-n+1)(1+x)^{r-n}$$

By Theorem 3.3.7, for $|x|<1$ and $n\geqslant r$,

$$(1+x)^r=1+\binom{r}{1}x+\cdots+\binom{r}{n-1}x^{n-1}+\binom{r}{n}\frac{x^n}{(1+\theta x)^{n-r}}$$

for some $\theta \in (0,1)$, which can depend on both $n$ and $x$. But if $x \geqslant 0$, $(1+\theta x)^{n-r} \geqslant 1$ so

$$0 \leqslant \frac{1}{(1+\theta x)^{n-r}} \leqslant 1 \Rightarrow |R_n(x)| = \left|\binom{r}{n}\frac{x^n}{(1+\theta x)^{n-r}}\right| \leqslant \left|\binom{r}{n}x^n\right|$$

Now observe that $\sum_{n\geqslant 0}\binom{r}{n}x^n$ converges absolutely for $|x|<1$:

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{r-n}{n+1}\right||x| \to |x| < 1 \quad \text{as } n\to\infty$$

So by ratio test and Theorem 3.1.16, $\sum_{n\geqslant 0}\binom{r}{n}x^n$ converges. This implies

$$\binom{r}{n}x^n \to 0 \quad \text{as } n \to \infty$$

Thus, $R_n(x) \to 0$ as $n \to \infty$ for $x \in [0,1)$.

However, if $-1 < x < 0$, the above approach does not work as $(1+\theta x)^{n-1} < 1$. Instead, we can use Cauchy's remainder

$$R_n = \frac{(1-\theta)^{n-1}x^n}{(n-1)!}r(r-1)\cdots(r-n+1)(1+\theta x)^{r-n} = r\binom{r-1}{n-1}\left(\frac{1-\theta}{1+\theta x}\right)^{n-1}(1+\theta x)^{r-1}x^n$$

for some $\theta \in (0,1)$. But $(1-\theta)/(1+\theta x) < 1$ for all $x \in (-1,1)$ as $1+\theta x = 1-\theta+\theta(x+1) \geqslant 1-\theta$. Hence

$$|R_n| \leqslant r|x|\left|\binom{r-1}{n-1}x^{n-1}\right|(1+\theta x)^{r-1}$$

Moreover, $(1+\theta x)^{r-1} \leqslant \max\{(1-|x|)^{r-1},(1+|x|)^{r-1}\}$, and

$$|R_n| \leqslant r|x|\max\{(1-|x|)^{r-1},(1+|x|)^{r-1}\}\left|\binom{r-1}{n-1}x^{n-1}\right| = K_{r,x}\left|\binom{r-1}{n-1}x^{n-1}\right| \to 0$$

as $n \to \infty$ since $K_{r,x}$ is independent of $n$. Therefore $R_n \to 0$ as $n \to \infty$ for $x$ fixed in $-1,1$. ■

## 3.4 Power Series

We want to consider functions defined by power series of the form

$$f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$$

where $a_n, z, z_0 \in \mathbb{C}$. Obviously the series may not converge for every $z \in \mathbb{C}$. First thing to do is to find out the set of points for which it does converge.[5] By translation, we can assume $z_0 = 0$.

**Lemma 3.4.1.** If $\sum_{n=0}^{\infty} a_n z^n$ converges, and $|w| < |z|$, then $\sum_{n=0}^{\infty} a_n w^n$ converges absolutely.

*Proof.* Since $\sum_{n=0}^{\infty} a_n z^n$ converges, $a_n z^n \to 0$. Thus $\exists K > 0$ such that[6]

$$|a_n z^n| \leqslant K \quad \forall n$$

Now consider (note that $z > 0$ to define $w$)

$$|a_n w^n| = |a_n z^n|\left|\frac{w}{z}\right|^n \leqslant K\rho^n$$

where $\rho = |w/z| < 1$ by assumption. Thus $\sum_{n=0}^{\infty}|a_n w^n|$ converges by comparison to the geometric series $\sum_{n=0}^{\infty} K\rho^n$, which converges. ■

We now use Lemma 3.4.1 to show every power series has a well-defined *radius of convergence*.

[5] Note that set is always non-empty as it contains $z_0$.

[6] Recall that convergent series is bounded.

**Theorem 3.4.2.** Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series. Then there exists $R \in [0, \infty]$, the radius of convergence, such that the series converges absolutely for $|z| < R$ and diverges for $|z| > R$.

*Proof.* Let

$$A = \left\{ r \geqslant 0 \;\middle|\; \exists z \in \mathbb{C} \text{ with } |z| = r \text{ such that } \sum_{n=0}^{\infty} a_n z^n \text{ converges} \right\}$$

Clearly, $0 \in A$, and $A$ is non-empty. So let

$$R = \sup A$$

with noting that $\sup A = \infty$ meaning $A$ is unbounded above. From definition of $A$, $\sum a_n z^n$ diverges for $|z| > R$. Suppose $|w| < R$. Then there exists $r \in \mathbb{A}$ with $|w| < r$ so $\exists z \in \mathbb{C}$ with $|z| = r$ and $\sum a_n z^n$ converging. $|w| < |z|$ so by Lemma 3.4.1 $\sum a_n w^n$ converges absolutely. ■

Note that $R = 0$ means $\sum a_n z^n$ converges only for $z = 0$, and $R = \infty$ means $\sum a_n z^n$ converges absolutely for all $z \in \mathbb{C}$. When $0 < R < \infty$ theorem tells us nothing about convergence for $|z| = R$. Now we introduce a lemma that is useful at finding $R$:

**Lemma 3.4.3.** If $|a_{n+1}/a_n| \to l$ as $n \to \infty$, then $R = 1/l$.

*Proof.* We use ratio test. Consider

$$\lim_{n\to\infty} \left| \frac{a_{n+1} z^{n+1}}{a_n z^n} \right| = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| |z| = l|z|$$

so if $l|z| < 1$ we get absolute convergence, and if $l|z| > 1$ we get divergence. ■

Let's take a look at some examples.

- $\sum_{n=0}^{\infty} z^n/n!$ converges absolutely for all $z \in \mathbb{C}$, since

  $$\left| \frac{a_{n+1}}{a_n} \right| = \frac{n!}{(n+1)!} = \frac{1}{n+1} \to 0 = l$$

  and $R = \infty$.
- $\sum_{n=0}^{\infty} n! z^n$ converges only for $z = 0$, since $|a_{n+1}/a_n| = n+1 \to \infty$ for $R = 0$.
- $\sum_{n=0}^{\infty} z^n$ geometric series converges if and only if $|z| < 1$.
- $\sum_{n=0}^{\infty} z^n/n^2$ has radius of convergence $R = 1$; specifically, absolute convergence for $|z| \leqslant 1$ and divergence for $|z| > 1$.
- $\sum_{n=0}^{\infty} z^n/n$ has radius of convergence $R = 1$. For $z = 1$, it diverges (harmonic series). For $|z| = 1$, $z \neq 1$, consider

  $$\begin{aligned}(1-z)\sum_{n=1}^{N} \frac{z^n}{n} &= \sum_{n=1}^{N} \left( \frac{z^n}{n} - \frac{z^{n+1}}{n} \right) \\ &= \sum_{n=1}^{N} \left( \frac{z^{n+1}}{n+1} - \frac{z^{n+1}}{n} \right) + z - \frac{z^{N+1}}{N+1} \\ &= z - z\sum_{n=1}^{N} \frac{1}{n(n+1)} z^n - \frac{z^{N+1}}{N+1}\end{aligned}$$

  If $|z| \leqslant 1$,

  $$\sum_{n=1}^{N} \frac{1}{n(n+1)} z^n$$

  converges absolutely by comparison to $\sum 1/n(n+1)$, and

  $$\left| \frac{z^{N+1}}{N+1} \right| \leqslant \frac{1}{N+1} \to 0$$

  Therefore $\sum_{n=0}^{\infty} z^n/n$ converges for $|z| \leqslant 1$, $z \neq 1$.

We conclude that nothing can be said in general about convergence on $|z| = R$, and need to analyse case by case. We shall see that inside the radius of convergence, power series are very well behaved, and "can treated like polynomials."

**Theorem 3.4.4.** Suppose

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

has radius of convergence $R > 0$. Then $f$ is differentiable at all points $z$ with $|z| < R$, and

$$f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}$$

*Proof.* We start by stating two auxillary lemmas:

**Lemma 3.4.5.** If $\sum_{n=0}^{\infty} a_n z^n$ has radius of convergence $R$, then so do $\sum_{n=1}^{\infty} n a_n z^{n-1}$ and $\sum_{n=2}^{\infty} n(n-1) a_n z^{n-2}$

**Lemma 3.4.6.**

(1) Let $n, r \in \mathbb{N}$. For all $2 \leqslant r \leqslant n$,

$$\binom{n}{r} \leqslant n(n-1)\binom{n-2}{r-2}$$

(2) Let $n \in \mathbb{N}$. For all $z, h \in \mathbb{C}$,

$$|(z+h)^n - z^n - nhz^{n-1}| \leqslant n(n-1)(|z|+|h|)^{n-2}|h|^2$$

Assume, for now, that these results hold. By Lemma 3.4.5, we may define

$$g(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}$$

where $|z| < R$. We need to show

$$I = \frac{f(z+h) - f(z) - hg(z)}{h} \to 0$$

as $h \to 0$. Fix $z$ with $|z| < R$, and assume $|z| + |h| < r < R$ for some $r$. All sums in $I$ converge by Lemma 3.4.5, and

$$I = \frac{1}{h} \sum_{n=0}^{\infty} a_n [(z+h)^n - z^n - hnz^{n-1}]$$

Then by continuity of absolute value function,

$$\begin{aligned} |I| &= \frac{1}{|h|} \left| \lim_{N\to\infty} \sum_{n=0}^{N} a_n [(z+h)^n - z^n - hnz^{n-1}] \right| \\ &= \lim_{N\to\infty} \underbrace{\frac{1}{|h|} \left| \sum_{n=0}^{N} a_n [(z+h)^n - z^n - hnz^{n-1}] \right|}_{I_N} \end{aligned}$$

Now, using the triangle inequality and Lemma 3.4.6 (2),

$$\begin{aligned}
|I_N| &\leqslant \frac{1}{|h|}\sum_{n=0}^{N}|a_n||(z+h)^n - z^n - hnz^{n-1}| \\
&\leqslant \frac{1}{|h|}\sum_{n=0}^{N}|a_n||n(n-1)(|z|+|h|)^{n-2}|h|^2| \\
&\leqslant |h|\sum_{n=0}^{N}n(n-1)|a_n|r^{n-2} \\
&\leqslant |h|\sum_{n=0}^{\infty}n(n-1)|a_n|r^{n-2} = |h|A_r
\end{aligned}$$

where $A_r$ converges by Lemma 3.4.5 plus the fact that $r < R$. We have shown that

$$|I| = \lim_{N\to\infty}|I_N| \leqslant |h|A_r \to 0$$

as $h \to 0$. ∎

Here we proof two auxillary lemmas:

*Proof.* (Lemma 3.4.5) Suppose $0 < |z| < R$, then $\exists r$ such that $|z| < r < R$. We know $\sum_{n=0}^{\infty} a_n r^n$ converges so $a_n r^n \to 0$ as $n \to \infty$ and hence $\exists K$ such that

$$|a_n r^n| \leqslant K \quad \forall n \geqslant 0$$

Thus

$$|na_n z^{n-1}| = \frac{|a_n r^n|}{|z|} n \left|\frac{z}{r}\right|^n \leqslant \frac{K}{|z|} n\rho^n$$

where $K$ is independent of $n$ and $\rho = |z|/r < 1$. But $\sum_{n=1}^{\infty} n\rho^n$ converges by ratio test:

$$\left|\frac{(n+1)\rho^{n+1}}{n\rho^n}\right| = \rho\left(1+\frac{1}{n}\right) \to \rho < 1$$

Therefore, $\sum_{n=1}^{\infty} na_n z^{n-1}$ converges absolutely by comparison.

If $|z| > R$, then

$$|na_n z^{n-1}| \geqslant \frac{1}{|z|}|a_n z^n|$$

If $\sum_{n=1}^{\infty} na_n w^{n-1}$ converges for some $|w| > R$, then taking $|w| > |z| > R$ we have $\sum_{n=1}^{\infty} na_n z^{n-1}$ converging absolutely by Lemma 3.4.1. This implies that $\sum a_n z^n$ converges absolutely (by comparison), contradicting $R$ being radius of convergence of $\sum a_n z^n$. Therefore $\sum_{n=1}^{\infty} na_n w^{n-1}$ diverges, and $\sum_{n=1}^{\infty} na_n z^{n-1}$ has radius of convergence $R$.

For the series $\sum_{n=2}^{\infty} n(n-1)a_n z^{n-2}$, apply result above, starting with $\sum_{n=1}^{\infty} na_n z^{n-1}$. ∎

*Proof.* (Lemma 3.4.6)

(1) Simple manipulation gives

$$\binom{n}{r}\Big/\binom{n-2}{r-2} = \frac{n!}{(n-r)!r!}\frac{(n-r)!(r-2)!}{(n-2)!} = \frac{n(n-1)}{r(r-1)} \geqslant n(n-1)$$

since $r \in \mathbb{N}$, $r \geqslant 2$.

(2) By binomial expansion, (1), and triangle inequality, we have

$$\begin{aligned}
|(z+h)^n - z^n - nhz^{n-1}| &= \left|\sum_{r=2}^{n} \binom{n}{r} z^{n-r} h^r\right| \\
&\leqslant \sum_{r=2}^{n} n(n-1)\binom{n-2}{r-2}|z|^{n-r}|h|^r \\
&= n(n-1)|h|^2 \sum_{r=2}^{n} \binom{n-2}{r-2}|z|^{n-r} \\
&= n(n-1)|h|^2(|z|+|h|)^{n-2}
\end{aligned}$$

■

N.b. Theorem 3.4.4 can be iterated: power series are smooth inside the radius of convergence.

### 3.4.1 The Standard Functions

**⁎ Exponentials and logarithms**

We saw in previous example that $\sum_{n=0}^{\infty} z^n/n!$ has $R = \infty$, i.e. it converges at all $z \in \mathbb{C}$. With this fact, we define the *exponential function* as follows.

**Definition 3.4.1.** (*Exponential Function*) Let $\exp : \mathbb{C} \to \mathbb{C}$,

$$z \mapsto \exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

We immediately have from Theorem 3.4.4 that exp is differentiable and

$$\exp'(z) = \exp(z)$$

Nextly, we claim $\exp(a+b) = \exp(a)\exp(b)$. To show this, we need the following fact: if $F : \mathbb{C} \to \mathbb{C}$ satisfies $F'(z) = 0$ for all $z \in \mathbb{C}$, then $F$ is constant.[7]

*Proof.* Consider $g : \mathbb{R} \to \mathbb{C}$, $g(t) = F(tz)$ for some fixed $z \in \mathbb{C}$. By chain rule,

$$g'(t) = zF'(tz) = 0$$

We can write

$$g(t) = u(t) + iv(t)$$

with $u, v$ real, and then

$$g'(t) = u'(t) + iv'(t)$$

giving

$$u'(t) = v'(t) = 0 \Rightarrow u, v \text{ constant}$$

by Corollary 3.3.4.1. Therefore $F(z) = F(0)$ (put $t = 0$ and $t = 1$). But since $z$ is arbitrary, $F$ is constant. ■

Now for $a, b \in \mathbb{C}$ consider

$$F(z) = \exp(a+b-z)\exp(z)$$

We compute

$$F'(z) = -\exp'(a+b-z)\exp(z) + \exp(a+b-z)\exp'(z) = 0$$

to find out that $F(0) = F(b)$. Hence $\exp(a+b)\exp(0) = \exp(a)\exp(b)$, but

$$\exp(0) = \sum_{n=0}^{\infty} \frac{0^n}{n!} = 1$$

Therefore, $\exp(a+b) = \exp(a)\exp(b)$ for all $a, b \in \mathbb{C}$.

From now on we restrict to $\mathbb{R}$.

[7] Note the difference with Corollary 3.3.4.1 – range is now $\mathbb{C}$.

**Theorem 3.4.7.** Consider $\exp : \mathbb{R} \to \mathbb{R}$.

(1) exp is everywhere differentiable and $\exp'(x) = \exp(x)$.

(2) $\exp(x + y) = \exp(x)\exp(y)$ for all $x, y \in \mathbb{R}$.

(3) $\exp(x) > 0$ for all $x \in \mathbb{R}$.

(4) exp is strictly increasing.

(5) $\exp(x) \to \infty$ as $x \to \infty$; and $\exp \to 0$ as $x \to -\infty$.

(6) $\exp : \mathbb{R} \to (0, \infty)$ is a bijection.

*Proof.* (1) and (2) are done. We prove the remainings.

(3) If $x > 0$, clearly

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \geqslant 1 > 0$$

Also, $\exp(0) = 1$ and $\exp(x - x) = \exp(x)\exp(-x) = 1$. Thus $\exp(-x) > 0$ for all $x > 0$.

(4) $\exp'(x) = \exp(x) > 0$. Hence exp is strictly increasing by Corollary 3.3.4.1.

(5) $\exp(x) \geqslant 1 + x$ for $x \geqslant 0$ so if $x \to \infty$, $\exp(x) \to \infty$. Meanwhile, for $x \geqslant 0$,

$$\exp(-x) = \frac{1}{\exp(x)}$$

so $\exp(-x) \to 0$ as $x \to \infty$, i.e. $\exp(x) \to 0$ as $x \to -\infty$.

(6) Injectivity is immediate from being strictly increasing. For surjectivity, suppose $y \in (0, \infty)$. Then from (4), there exists $a, b \in \mathbb{R}$ such that

$$\exp(a) < y < \exp(b)$$

Applying intermediate value theorem (Theorem 3.2.6) to $\exp : [a, b] \to \mathbb{R}$ gives $\exists x \in \mathbb{R}$ such that $\exp(x) = y$. Hence exp is surjective.

■

Notice that $\exp : (\mathbb{R}, +) \to ((0, \infty), \times)$ is a group isomorphism. And since exp is a bijection, it has an inverse function

$$\ln : (0, \infty) \to \mathbb{R}$$

which is called the *logarithmic function*.

**Theorem 3.4.8.**

(1) $\ln : (0, \infty) \to \mathbb{R}$ is a bijection, and

$$\ln(\exp(x)) = x \quad \forall x \in \mathbb{R},$$

$$\exp(\ln(t)) = t \quad \forall t \in (0, \infty)$$

(2) ln is differentiable and monotone, with

$$\ln'(t) = \frac{1}{t}$$

(3) $\ln(st) = \ln(s) + \ln(t)$ where $s, t > 0$.

(4) $\ln(x) \to \infty$ as $x \to \infty$; and $\ln(x) \to -\infty$ as $x \to 0$.

*Proof.*

(1) Trivial from construction (ln is the inverse of exp).

(2) Inverse function theorem (Theorem 3.3.5) gives that ln is differentiable and

$$\ln'(t) = \frac{1}{\exp'(\ln(t))} = \frac{1}{\exp(\ln t)} = \frac{1}{t}$$

for all $t > 0$.

(3) Because $e$ is an isomorphism, so is its inverse.

■

Now define for $\alpha \in \mathbb{R}$ and $x > 0$,

$$\Gamma_\alpha(x) = \exp(\alpha \ln(x))$$

**Theorem 3.4.9.** Suppose $x, y > 0$ and $\alpha, \beta \in \mathbb{R}$. Then

(1) $\Gamma_\alpha(x, y) = \Gamma_\alpha(x)\Gamma_\alpha(y)$.

(2) $\Gamma_{\alpha+\beta}(x) = \Gamma_\alpha(x)\Gamma_\beta(x)$.

(3) $\Gamma_\alpha(\Gamma_\beta(x)) = \Gamma_{\alpha\beta}(x)$.

(4) $\Gamma_1(x) = x$, $\Gamma_0(x) = 1$.

*Proof.*

(1) $\Gamma_\alpha(xy) = \exp(\alpha \ln(xy)) = \exp(\alpha \ln(x) + \alpha \ln(y)) = \exp(\alpha \ln(x)) \exp(\alpha \ln(y)) = \Gamma_\alpha(x)\Gamma_\alpha(y)$.

(2) $\Gamma_{\alpha+\beta}(x) = \exp((\alpha+\beta)\ln(x)) = \exp(\alpha \ln(x) + \beta \ln(x))$
$= \exp(\alpha \ln(x)) \exp(\beta(\ln(x))) = \Gamma_\alpha(x)\Gamma_\beta(x)$.

(3) $\Gamma_\alpha(\Gamma_\beta(x)) = \exp(\alpha \ln(\exp(\beta \ln(x)))) = \exp(\alpha\beta \ln(x)) = \Gamma_{\alpha\beta}(x)$.

(4) $\Gamma_1(x) = \exp(\ln(x)) = x$, $\Gamma_0(x) = \exp(0) = 1$.

■

Now suppose $p, q \in \mathbb{Z}$, $p, q \geqslant 1$. Then,

$$\Gamma_p(x) = \Gamma_{1+1+\cdots+1}(x) = \Gamma_1(x)\Gamma_1(x)\cdots\Gamma_1(x) = x^p$$

and

$$\Gamma_p(x)\Gamma_{-p}(x) = \Gamma_0(x) = 1 \Rightarrow \Gamma_{-p}(x) = \frac{1}{x^p}$$

Further,

$$\left(\Gamma_{1/q}(x)\right)^q = \Gamma_{1/q}(x)\cdots\Gamma_{1/q}(x) = \Gamma_{1/q+\cdots+1/q}(x) = \Gamma_1(x) = x$$

gives $\Gamma_{1/q}(x) = x^{1/q}$. Finally,

$$\Gamma_{p/q}(x) = \Gamma_{\underbrace{1/q + \cdots + 1/q}_{p \text{ times}}}(x) = \left(\Gamma_{1/q}(x)\right)^p = (x^1/q)^p = x^{p/q}$$

. Thus $\Gamma_\alpha(x)$ agrees with $x^\alpha$ for $x \in \mathbb{Q}$ as previously defined. Hence we can extend the definition of $x^\alpha$ into $x \in \mathbb{R}$ and write

$$x^\alpha = \Gamma_\alpha(x)$$

for $\alpha \in \mathbb{R}$, $x \in (0, \infty)$. We also introduce *Euler's number*

$$e = \exp(1) = \sum_{n=1}^{\infty} \frac{1}{n!}$$

and write exp as a power:

$$\exp(x) = \exp(x \ln e) = \Gamma_x(e) = e^x$$

Now we can generalise as follows.

$$(x^\alpha)' = (e^{\alpha \ln x})' = \frac{\alpha}{x} e^{\alpha \ln x} = \alpha x^{\alpha - 1}$$

for all $x > 0$, $\alpha \in \mathbb{R}$. Meanwhile, for $f(x) = a^x$ ($a > 0$, $x \in \mathbb{R}$),

$$f'(x) = (e^{x \ln a})' = e^{x \ln a} \ln a = a^x \ln a$$

In particular, $f'(x) > 0$ for $a > 1$ and $f'(x) < 0$ for $a < 1$.

**Lemma 3.4.10.** For any $r > 0$ we have

(1) $x^r e^{-x} \to 0$ as $x \to \infty$.

(2) $x^{-r} \ln x \to 0$ as $x \to \infty$.

(3) $x^r \ln x \to 0$ as $x \to 0+$.

*Proof.*

(1) For $x > 0$,

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots > \frac{x^n}{n!}$$

for any $n \in \mathbb{N}$. Pick $n$ such that $n - r \geqslant 1$. Then,

$$0 \leqslant \frac{x^r}{e^x} \leqslant \frac{n!}{x^{n-r}} \leqslant \frac{n!}{x}$$

so $x^r e^- x \to 0$ as $x \to \infty$.

(2) Pick $m$ such that $m \geqslant 2/r$. For $x > 1$,

$$0 \leqslant \frac{x^{1/r}}{e^x} \leqslant \frac{m!}{x^{m-1/r}} \leqslant \frac{m!}{x^{1/r}}$$

Let $x = \ln y$ and write

$$0 \leqslant \frac{(\ln y)^{1/r}}{r} \leqslant \frac{m!}{(\ln y)^{1/r}}$$

giving

$$0 \leqslant \frac{\ln y}{y^r} \leqslant \frac{(m!)^r}{\ln y}$$

But $\ln y \to \infty$ as $y \to \infty$; thus $\ln y / y^r \to 0$.

(3) Set $z = 1/y$ where $y$ is from above. We have

$$0 \leqslant -z^r \ln z \leqslant \frac{(m!)^r}{-\ln z}$$

Since $\ln z \to -\infty$ as $z \to 0+$, $z^r \ln z \to 0$.

∎

$\ast$ **Trigonometric functions**

**Definition 3.4.2.** (*Trigonometric Functions*) We define

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{(2k)!}$$

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{(2k+1)!}$$

We can check that, like the exponential function, both series have infinite radius of convergence by ratio test. Hence,

$$\cos' z = -\sin z \text{ and } \sin' z = \cos z$$

by Theorem 3.4.4. Also observe that

$$\begin{aligned} e^{iz} &= \lim_{N\to\infty} \left( \sum_{k=0}^{2N+1} \frac{(iz)^k}{k!} \right) \\ &= \lim_{N\to\infty} \left( \sum_{k=0}^{N} \frac{(iz)^{2k}}{(2k)!} + \sum_{k=0}^{N} \frac{(iz)^{2k+1}}{(2k+1)!} \right) \\ &= \lim_{N\to\infty} \left( \sum_{k=0}^{N} (-1)^k \frac{z^{2k}}{(2k)!} \right) + i \sum_{k=0}^{N} (-1)^k \frac{z^{2k+1}}{(2k+1)!} \\ &= \cos z + i \sin z \end{aligned}$$

Similarly, $e^{-iz} = \cos z - i \sin z$. Hence,

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz})$$

and

$$\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$$

These formulae give many trigonometric identities, e.g. $\cos z = \cos(-z)$, $\sin(z) = -\sin(-z)$, $\cos(0) = 1$, $\sin(0) = 0$ etc. Employing $e^{a+b} = e^a e^b$, we also find

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

and

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

for all $z, w \in \mathbb{C}$. Furthermore, set $w = -z$ to deduce

$$\cos^2 + \sin^2 = \cos(0) = 1 \tag{$\ast$}$$

Now, if $x \in \mathbb{R}$ then $\sin x, \cos x \in \mathbb{R}$ and $(\ast)$ implies $|\cos x|, |\sin x| \leqslant 1$. N.b. this need not be true away from real axis, e.g. if $y \in \mathbb{R}$,

$$\cos iy = \frac{1}{2}(e^{-y} + e^{y}) \to \infty$$

as $y \to \pm\infty$.

$\ast$ **Periodicity of trigonometric functions**

**Proposition 3.4.11.** There is a smallest positive number $\omega$ (where $\sqrt{2} < \omega/2 < \sqrt{3}$) such that

$$\cos \frac{\omega}{2} = 0$$

*Proof.* Suppose $0 \leqslant x \leqslant 2$.

$$\sin x = \left(x - \frac{x^3}{3!}\right) + \left(\frac{x^5}{5!} - \frac{x^7}{7!}\right) + \cdots + \frac{x^{2n-1}}{(2n-1)!}\left(1 - \frac{x^2}{2n(2n+1)}\right) + \cdots$$

Since each term in parenthesis are positive, $\sin x \geqslant 0$, and if $0 < x \leqslant 2$, $\sin x > 0$. So $\cos' x = -\sin x < 0$ for $0 < x < 2$. Hence $\cos x$ is strictly decreasing in this interval, hence it has at most one root in $[0, 2]$. To complete the proof we show $\cos\sqrt{2} > 0 > \cos\sqrt{3}$, then intermediate valuable theorem implies a root exists in $[\sqrt{2}, \sqrt{3}]$. But

$$\cos\sqrt{2} = 1 - \underbrace{\frac{(\sqrt{2})^2}{2!}}_{=0} + \underbrace{\frac{(\sqrt{2})^4}{4!} - \frac{(\sqrt{2})^6}{6!}}_{>0} + \cdots + \underbrace{\frac{(\sqrt{2})^{2n}}{(2n)!}\left(1 - \frac{2}{(2n+1)(2n+2)}\right)}_{>0} + \cdots$$

so $\cos\sqrt{2} > 0$, and

$$\cos\sqrt{3} = \underbrace{1 - \frac{(\sqrt{3})^2}{2!} + \frac{(\sqrt{3})^4}{4!}}_{-1/8} - \left(\underbrace{\frac{(\sqrt{3})^6}{6!} - \frac{(\sqrt{3})^8}{8!}}_{>0}\right) - \cdots - \frac{(\sqrt{3})^{2n}}{(2n)!}\left(\underbrace{1 - \frac{3}{(2n+1)(2n+2)}}_{>0}\right) - \cdots$$

implies $\cos\sqrt{3}$, completing the proof. ■

**Corollary 3.4.11.1.** $\sin\omega/2{=}1$.

*Proof.* From

$$\sin^2\frac{\omega}{2} + \cos^2\frac{\omega}{2} = 1$$

we have $\sin^2\omega/2 = 1 \Rightarrow \sin\omega/2 = \pm 1$. But $\omega/2 \in (0, 2)$ so $\sin\omega/2 > 0$. Therefore $\sin\omega/2 = 1$. ■

Since we established some properties, we now define $\pi = \omega$. Periodic properties of the trigonometric functions follows.

**Theorem 3.4.12.**

(1) $\sin(z + \pi/2) = \cos z$; $\cos(z + \pi/2) = -\sin z$.

(2) $\sin(z + \pi) = -\sin z$; $\cos(z + \pi) = -\cos z$.

(3) $\sin(z + 2\pi) = \sin z$; $\cos(z + 2\pi) = \cos z$.

*Proof.* (1) Follows from addition formulae plus $\cos(\pi/2) = 0$, $\sin(\pi/2) = 1$. (2) and (3) follow by iterating. ■

It follows that

$$e^{iz+2\pi i} = \cos(z + 2\pi) + i\sin(z + 2\pi) = \cos z + i\sin z = e^{iz}$$

and hence $e^{\omega}$ is periodic with period $2\pi i$. Also

$$e^{i\pi} = \cos\pi + i\sin\pi = -\cos 0 + i(-\sin 0) = -1$$

giving the *Euler's identity*:

$$e^{i\pi} + 1 = 0$$

### ⁎ Trigonometric functions and geometry

Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, define $\mathbf{x} \cdot \mathbf{y}$ as usual. We set $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$. Cauchy-Schwarz shows that

$$|\mathbf{x} \cdot \mathbf{y}| \leqslant \|\mathbf{x}\|\|\mathbf{y}\|$$

Hence for $\mathbf{x} \neq 0$, $\mathbf{y} \neq 0$,

$$-1 \leqslant \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \leqslant 1$$

Thus we can define the angle between $\mathbf{x}$ and $\mathbf{y}$ as the unique $\theta \in [0, \pi]$ with

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$$

$\ast$ **Hyperbolic functions**

**Definition 3.4.3.** (*Hyperbolic Functions*) Define

$$\cosh z = \frac{1}{2}(e^z + e^{-z})$$

$$\sinh z = \frac{1}{2}(e^z - e^{-z})$$

where $z \in \mathbb{C}$.

It follows that

$$\cosh z = \cos(iz) \text{ and } \sinh z = -i\sin(iz)$$

Also,

$$\cosh' z = \sinh z, \quad \sinh' z = \cosh z$$

and

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

## 3.5 Integration

### 3.5.1 Riemann Integral

Informally, we define

$$\int_a^b f(x)dx$$

as the (signed) area under the graph of $f(x)$. To find the area, we approximate graph from above and below by rectangles. If $f$ is 'nice', estimates from above and below will be close for a suitably fine dissection of $[a,b]$. We assume $f : [a,b] \to \mathbb{R}$ is bounded, i.e. $\exists K$ such that $|f(x)| \leqslant K$ for all $x \in [a,b]$.

**Definition 3.5.1.** A *dissection* $\mathcal{D}$ of $[a,b]$ is a finite subset of $[a,b]$ containing the end points $a$, $b$. We write

$$\mathcal{D} = \{x_0, x_1, \ldots, x_n\}$$

with

$$a = x_0 < x_1 < \cdots < x_n = b$$

Associated to a dissection are the *upper* and *lower sums*, given by

$$U(f,\mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \sup_{x \in [x_{j-1}, x_j]} f(x)$$

and

$$L(f,\mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \inf_{x \in [x_{j-1}, x_j]} f(x)$$

respectively. Clearly, for any dissection, $L(f,\mathcal{D}) \leqslant U(f,\mathcal{D})$.

**Lemma 3.5.1.** Suppose $\mathcal{D}'$,$\mathcal{D}$ are dissections with $\mathcal{D}' \supseteq \mathcal{D}$. Then,

$$L(f,\mathcal{D}) \leqslant L(f,\mathcal{D}') \leqslant U(f,\mathcal{D}') \leqslant U(f,\mathcal{D})$$

We say $\mathcal{D}'$ is a refinement of $\mathcal{D}$.

*Proof.* Let $\mathcal{D} = \{x_0, \ldots, x_n\}$ with $x_0 < x_1 < \cdots < x_n$. First, consider the case where $\mathcal{D}' = \mathcal{D} \cup \{y\}$. Then, $y \in (x_{r-1}, x_r)$ for some $r$. Clearly,

$$\sup_{[x_{r-1},y]} f(x) \leqslant \sup_{[x_{r-1},x_r]} f(x)$$

and

$$\sup_{[y,x_r]} f(x) \leqslant \sup_{[x_{r-1},x_r]} f(x)$$

Combining these,

$$\begin{aligned}(y - x_{r-1}) \sup_{[x_{r-1},y]} f(x) + (x_r - y) \sup_{[y,x_r]} f(x) &\leqslant (y - x_{r-1}) \sup_{[x_{r-1},x_r]} f(x) + (x_r - y) \sup_{[x_{r-1},x_r]} f(x) \\ &= (x_r - x_{r-1}) \sup_{[x_{r-1},x_r]} f(x)\end{aligned}$$

Hence $U(f, \mathcal{D}') \leqslant U(f, \mathcal{D})$. Similarly,

$$\inf_{[x_{r-1},y]} f, \inf_{[y,x_r]} f \geqslant \inf_{[x_{r-1},x_r]} f$$

and $L(f, \mathcal{D}') \geqslant U(f, \mathcal{D})$. Now if $\mathcal{D}'$ has more extra points, we add them one at a time and use the result recursively. ■

**Lemma 3.5.2.** Suppose $\mathcal{D}_1$ and $\mathcal{D}_2$ are two dissections. Then,

$$L(f, \mathcal{D}_1) \leqslant U(f, \mathcal{D}_2)$$

*Proof.* Since $\mathcal{D}_1 \subseteq \mathcal{D}_1 \cup \mathcal{D}_2$ and $\mathcal{D}_2 \subseteq \mathcal{D}_1 \cup \mathcal{D}_2$, we have, from Lemma 3.5.1,

$$L(f, \mathcal{D}_1) \leqslant L(f, \mathcal{D}_1 \cup \mathcal{D}_2) \leqslant U(f, \mathcal{D}_1 \cup \mathcal{D}_2) \leqslant U(f, \mathcal{D}_2)$$

■

Note that if $\mathcal{D}_0 = \{a, b\}$,

$$L(f, \mathcal{D}_0) = (b - a) \inf_{[a,b]} f \geqslant -(b - a)K$$

and

$$U(f, \mathcal{D}_0) = (b - a) \sup_{[a,b]} f \leqslant (b - a)K$$

Since $\mathcal{D}_0 \subseteq \mathcal{D}$ for any $\mathcal{D}$, we deduce

$$\{U(f, \mathcal{D}) \mid \mathcal{D} \text{ dissections}\}$$

and

$$\{L(f, \mathcal{D}) \mid \mathcal{D} \text{ dissections}\}$$

are bounded above by $(b-a)K$, and below by $-(b-a)K$; and both sets are non-empty. We can define the *upper* and *lower integra*l now.

**Definition 3.5.2.** The *upper integral* of $f$ is

$$I^*(f) = \overline{\int_a^b} f(x)dx = \inf_{\mathcal{D}} U(f, \mathcal{D})$$

and the *lower integral* of $f$ is

$$I_*(f) = \underline{\int_a^b} f(x)dx = \sup_{\mathcal{D}} L(f, \mathcal{D})$$

Also notice that from Lemma 3.5.2,

$$L(f,\mathcal{D}_1) \leqslant U(f,\mathcal{D}_2) \Rightarrow L(f,\mathcal{D}_1) \leqslant \inf_{\mathcal{D}_2} U(f,\mathcal{D}_2) \Rightarrow \sup_{\mathcal{D}_1} L(f,\mathcal{D}_1) \leqslant \inf_{\mathcal{D}_2} U(f,\mathcal{D}_2) \Rightarrow I_*(f) \leqslant I^*(f)$$

Hence,

$$(b-a)\inf_{[a,b]} f(x) \leqslant I_*(f) \leqslant I^*(f) \leqslant (b-a)\sup_{[a,b]} f(x)$$

Now we formally define the integral as follows.

**Definition 3.5.3.** A bounded function $f : [a,b] \to \mathbb{R}$ is *Riemann integrable* (integrable) if

$$I^*(f) = I_*(f)$$

Then we define

$$\int_a^b f(x)dx = I^*(f) = I_*(f) = \int_a^b f$$

**Example 3.5.1.** Consider function $f : [0,1] \to \mathbb{R}$, $x \mapsto x$. Let

$$\mathcal{D}_k = \left\{0, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1\right\}$$

(uniform dissection). Then,

$$U(f,\mathcal{D}_k) = \sum_{j=1}^{k} \left(\frac{j}{k} - \frac{(j-1)}{k}\right) \sup_{\left[\frac{j-1}{k}, \frac{j}{k}\right]} x = \sum_{j=1}^{k} \frac{1}{k}\frac{j}{k} = \frac{1}{k^2}\frac{1}{2}k(k+1) = \frac{1}{2} + \frac{1}{2k}$$

Similarly,

$$L(f,\mathcal{D}_k) = \frac{1}{2} - \frac{1}{2k}$$

Therefore,

$$I^*(f) = \inf_{\mathcal{D}} U(f,\mathcal{D}) \leqslant \inf_{\mathcal{D}_k} U(f,\mathcal{D}_k) = \frac{1}{2}$$

and

$$I_*(f) = \sup \mathcal{D} L(f,\mathcal{D}) \geqslant \sup_{\mathcal{D}_k} L(f,\mathcal{D}_k) = \frac{1}{2}$$

But $I_*(f) \leqslant I^*(f)$ so $I_*(f) = I^*(f) = 1/2$. Thus $f$ is integrable and

$$\int_0^1 x dx = \frac{1}{2}$$

**Example 3.5.2.** Consider $f : [0,1] \to \mathbb{R}$,

$$x \mapsto \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

This time, $f$ is not Riemann integrable. For any dissection $\mathcal{D}$,

$$U(f,\mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \sup_{[x_{j-1},x_j]} f = \sum_{j=1}^{n} (x_j - x_{j-1}) = 1$$

and

$$L(f,\mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1}) \inf_{[x_{j-1},x_j]} f = 0$$

Thus, $I_*(f) \neq I^*(f)$ and $f$ is not Riemann integrable.

**Theorem 3.5.3.** (*Cauchy Criterion for Integrability*) A bounded function $f : [a,b] \to \mathbb{R}$ is integrable if and only if for all $\epsilon > 0$ there exists a dissection $\mathcal{D}$ with

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) < \epsilon$$

*Proof.* For any dissection, we have

$$0 \leqslant I^*(f) - I_*(f) \leqslant U(f,\mathcal{D}) - L(f,\mathcal{D}) < \epsilon$$

If the criterion holds, then

$$0 \leqslant I^*(f) - I_*(f) < \epsilon$$

for all $\epsilon > 0$. So $I^*(f) = I_*(f)$ and $f$ is integrable.

Conversely, suppose $f$ is integrable and fix $\epsilon > 0$. It follows from definition of supremum and infimum that there exist $\mathcal{D}_1, \mathcal{D}_2$ with

$$U(f,\mathcal{D}_1) \leqslant I^*(f) + \epsilon/2$$

and

$$L(f,\mathcal{D}_2) \geqslant I_*(f) - \epsilon/2$$

Let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$. Then,

$$\begin{aligned} U(f,\mathcal{D}) &\leqslant U(f,\mathcal{D}_1), \\ L(f,\mathcal{D}) &\geqslant L(f,\mathcal{D}_2) \end{aligned}$$

and we can find

$$\begin{aligned} 0 \leqslant U(f,\mathcal{D}) - L(f,\mathcal{D}) &\leqslant U(f,\mathcal{D}_1) - L(f,\mathcal{D}_2) \\ &\leqslant I^*(f) + \epsilon/2 - I_*(f) + \epsilon/2 = \epsilon \end{aligned}$$

∎

This criterion (Theorem 3.5.3) can be used to show monotone functions and continuous functions are integrable.

Observe that if $f : [a,b] \to \mathbb{R}$ is monotone, it is bounded by $f(a)$ and $f(b)$.

**Theorem 3.5.4.** Suppose $f : [a,b] \to \mathbb{R}$ is monotone. Then $f$ is integrable.

*Proof.* Suppose $f$ is increasing. Then $\sup_{[x_{j-1},x_j]} f = f(x_j)$ and $\inf_{[x_{j-1},x_j]} f = f(x_{j-1})$. Thus for any dissection $\mathcal{D}$,

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) = \sum_{j=1}^{n} (x_j - x_{j-1})[f(x_j) - f(x_{j-1})]$$

Now choose a uniform dissection into $n$ pieces, i.e.

$$\mathcal{D}_n = \left\{ a, a + \frac{(b-a)}{n} + a + \frac{2(b-a)}{n}, \ldots, a + \frac{n-1}{n}(b-a), b \right\}$$

Then,

$$\begin{aligned} U(f,\mathcal{D}_n) - L(f,\mathcal{D}_n) &= \sum_{j=1}^{n} \frac{(b-a)}{n} \left[ f\left(a + \frac{j(b-a)}{b}\right) - f\left(a + \frac{(j-1)}{n}(b-a)\right) \right] \\ &= \frac{(b-a)}{n}(f(b) - f(a)) \end{aligned}$$

Taking $n$ sufficiently large, for any $\epsilon > 0$, we can find $\mathcal{D} = \mathcal{D}_n$ such that

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) < \epsilon$$

Hence $f$ is integrable by Theorem 3.5.3. ∎

If $f : [a,b] \to \mathbb{R}$ is continuous, then $f$ is bounded by the extreme value theorem (Theorem 3.2.8).

**Theorem 3.5.5.** Suppose $f : [a,b] \to \mathbb{R}$ is continuous. Then $f$ is integrable.

*Proof.* We prove the contrapositive, i.e. if $f$ is not integrable, then $f$ is not continuous. Suppose $f$ is not integrable; then by Theorem 3.5.3, there exists $\epsilon_0 > 0$ such that, for all dissections $\mathcal{D}$,

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) > \epsilon_0$$

Since

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) = \sum_{j=1}^{n}(x_j - x_{j-1})[f(x_j) - f(x_{j-1})]$$

it follows that for any dissection there is a $j$ with

$$\sup_{[x_{j-1},x_j]} f - \inf_{[x_{j-1},x_j]} f > \frac{\epsilon_0}{b-a}$$

otherwise

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) \leqslant \sum_{j=1}^{n}(x_j - x_{j-1})\frac{\epsilon_0}{b-a} = \epsilon_0$$

In particular, we can find $y,z \in [x_{j-1},x_j]$ with

$$f(y) - f(z) > \frac{\epsilon_0}{b-a}$$

Now let $\mathcal{D}_n$ be the uniform dissection into $n$ equal intervals:

$$\mathcal{D}_n = \left\{a, a + \frac{(b-a)}{n} + a + \frac{2(b-a)}{n}, \ldots, a + \frac{n-1}{n}(b-a), b\right\}$$

Applying the argument above to $\mathcal{D}_n$, there must exist $y_n$, $z_n$ in the same subinterval so that $|y_n - z_n| < (b-a)/n$, with

$$f(y_n) - f(z_n) > \frac{\epsilon_0}{b-a}$$

$(y_n)$ is bounded by construction, so by Bolzano-Weierstrass theorem we can take a convergent subsequence $y_{n_k} \to \eta$ for some $\eta \in [a,b]$. Moreover, since

$$|y_{n_k} - z_{n_k}| < \frac{b-a}{n_k} \leqslant \frac{b-a}{k}$$

we have $z_{n_k} \to \eta$ also. But

$$f(y_{n_k}) - f(z_{n_k}) > \frac{\epsilon_0}{b-a} > 0$$

implies $f(y_{n_k})$ and $f(z_{n_k})$ cannot both converges to the same limit. So $f$ is not continuous at $\eta$ (Theorem 3.2.1). ■

However, a function not being continuous nor monotone does not suggest that function is not integrable. Between monotone and continuous functions, we have a large class of integrable functions, e.g. *Thomae's function*.[8]

[8] Moreover, this function is continuous at each irrational, and discontinuous at each rational.

### 3.5.2 Elementary Properties of the Integral

Before establishing various properties of the integral, we first give some properties of infimum and supremum.

**Lemma 3.5.6.** Suppose $I = [a, b]$ and $f, g : I \to \mathbb{R}$ are bounded. Then,

(1) If $f(x) \leqslant g(x)$ for all $x \in I$,
$$\sup_I f \leqslant \sup_I g \text{ and } \inf_I f \leqslant \inf_I g$$

(2) $\sup_I(-f) = -\inf_I f$.

(3) If $\lambda > 0$, then
$$\sup_I(\lambda f) = \lambda \sup_I f \text{ and } \inf_I(\lambda f) = \lambda \inf_I f$$

(4)
$$\sup_I(f+g) \leqslant \sup_I f + \sup_I g$$
$$\inf_I(f+g) \geqslant \inf_I f + \inf_I g$$

(5) $\sup_I |f| - \inf_I |f| \leqslant \sup_I f - \inf_I f$.

(6) $\sup_I f^2 - \inf_I f^2 \leqslant 2\sup_I |f|(\sup_I f - \inf_I f)$.

*Proof.* We will only prove (4)-(6).

(4) We have $f(x) \leqslant \sup_I f$, $g(x) \leqslant \sup_I g$ for all $x \in I$. Hence
$$f(x) + g(x) \leqslant \sup_I f + \sup_I g \quad \forall x \in I$$
and
$$\sup_I(f+g) \leqslant \sup_I f + \sup_I g$$
Proof is similar for inf.

(5)
- If $f(x) \geqslant 0$ for all $x$, then $|f| = f$ and
$$\sup_I |f| - \inf_I |f| = \sup_I f - \inf_I f$$
- If $f(x) \leqslant 0$ for all $x$, then $|f| = -f$ and
$$\sup_I |f| - \inf_I |f| = \sup_I(-f) - \inf_I(-f) = -\inf_I f + \sup_I f$$
by (2).
- If $\inf_I f < 0 < \sup_I f$, then
$$\begin{aligned}\sup_I |f| - \inf_I |f| = \sup_I |f| &= \max\{\sup_I f, \sup_I(-f)\}\\ &\leqslant \sup_I f + \sup_I(-f)\\ &= \sup_I f - \inf_I f\end{aligned}$$

(6) From
$$f(x)^2 - f(y)^2 = (f(x)+f(y))(f(x)-f(y)) \leqslant 2\sup_I |f|(\sup_I f - \inf_I f)$$
we take supremum over $x$ and $y$ for the result.

■

**Corollary 3.5.6.1.** For any dissection $\mathcal{D}$ of $[a,b]$, and bounded functions $f,g:[a,b]\to\mathbb{R}$,

(1) If $f(x)\leqslant g(x)$ for all $x\in[a,b]$,

$$U(f,\mathcal{D})\leqslant U(g,\mathcal{D}) \text{ and } L(f,\mathcal{D})\leqslant L(g,\mathcal{D})$$

(2) $L(-f,\mathcal{D})=-U(f,\mathcal{D})$.

(3) If $\lambda>0$,

$$U(\lambda f,\mathcal{D})=\lambda U(f,\mathcal{D}) \text{ and } L(\lambda f,\mathcal{D})=\lambda L(f,\mathcal{D})$$

(4)

$$U(f+g,\mathcal{D})\leqslant U(f,\mathcal{D})+U(g,\mathcal{D})$$
$$L(f+g,\mathcal{D})\leqslant L(f,\mathcal{D})+L(g,\mathcal{D})$$

(5) $U(|f|,\mathcal{D})-L(|f|,\mathcal{D})\leqslant U(f,\mathcal{D})-L(f,\mathcal{D})$.

(6) $U(f^2,\mathcal{D})-L(f^2,\mathcal{D})\leqslant 2\sup_{[a,b]}|f|(U(f,\mathcal{D})-L(f,\mathcal{D}))$.

*Proof.* Recall the definition of upper/lower sum and results follow from Lemma 3.5.6. For (6) we additionally need that

$$\sup_{[x_{j-1},x_j]}|f|\leqslant\sup_{[a,b]}|f|$$

■

**Theorem 3.5.7.** Let $f$, $g$ be bounded and integrable on $[a,b]$. Then,

(1) If $f(x)\leqslant g(x)$ for all $x\in[a,b]$,

$$\int_a^b f(x)dx\leqslant\int_a^b g(x)dx$$

(2) If $\lambda\in\mathbb{R}$, $\lambda f$ is integrable, and

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

(3) $f+g$ is integrable, and

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

(4) $|f|$ is integrable, and

$$\left|\int_a^b f(x)dx\right|\leqslant\int_a^b |f(x)|dx$$

(5) The product $fg$ is integrable.

*Proof.*

(1) For any dissection $\mathcal{D}$, we have

$$L(f,\mathcal{D})\leqslant L(g,\mathcal{D})\leqslant I_*(g)$$

Hence $I_*(f)\leqslant I_*(g)$. But $f$, $g$ integrable, so

$$\int_a^b f(x)dx\leqslant\int_a^b g(x)dx$$

(2) First consider $\lambda > 0$. Since $f$ is integrable, $\forall \epsilon > 0$, $\exists \mathcal{D}$ such that

$$U(f,\mathcal{D}) \leqslant \int_a^b f(x)dx + \epsilon, \quad L(f,\mathcal{D}) \geqslant \int_a^b f(x)dx - \epsilon$$

Then, by Corollary 3.5.6.1 (3),

$$\begin{aligned}\lambda \int_a^b f(x)dx - \lambda\epsilon \leqslant \lambda L(f,\mathcal{D}) = L(\lambda f,\mathcal{D}) \leqslant I_*(\lambda f)& \\ \leqslant I^*(\lambda f) \leqslant U(\lambda f,\mathcal{D}) = \lambda U(f,\mathcal{D}) \leqslant \lambda \int_a^b f(x)dx + \lambda\epsilon&\end{aligned}$$

But $\epsilon$ is arbitrary so

$$I_*(\lambda f) = I^*(\lambda f) = \lambda \int_a^b f(x)dx$$

Now consider $\lambda = -1$, $\mathcal{D}$ as above. Then,

$$\int_a^b f(x)dx - \epsilon \leqslant L(f,\mathcal{D}) = -U(-f,\mathcal{D})$$

and

$$\int_a^b f(x)dx + \epsilon \geqslant U(f,\mathcal{D}) = -L(-f,\mathcal{D})$$

Therefore

$$-\int_a^b f(x)dx - \epsilon \leqslant L(-f,\mathcal{D}) \leqslant I_*(-f) \leqslant I^*(-f) \leqslant U(-f,\mathcal{D}) \leqslant -\int_a^b f(x)dx + \epsilon$$

and

$$I_*(-f) = I^*(-f) = -\int_a^b f(x)dx$$

since $\epsilon$ arbitrary. And we can combine the two results above to get desired for all $\lambda \in \mathbb{R}$.

(3) Since $f, g$ integrable, $\forall \epsilon > 0$ there exist $\mathcal{D}_1, \mathcal{D}_2$ such that

$$\int_a^b f(x)dx - \epsilon \leqslant L(f,\mathcal{D}_1) \leqslant U(f,\mathcal{D}_1) \leqslant \int_a^b f(x)dx + \epsilon$$

$$\int_a^b g(x)dx - \epsilon \leqslant L(g,\mathcal{D}_2) \leqslant U(g,\mathcal{D}_2) \leqslant \int_a^b g(x)dx + \epsilon$$

Now let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$. Then, by Corollary 3.5.6.1 (4),

$$\begin{aligned}\int_a^b f(x)dx + \int_a^b g(x)dx - 2\epsilon &\leqslant L(f,\mathcal{D}_1) + L(g,\mathcal{D}_2) \\ &\leqslant L(f,\mathcal{D}) + L(g,\mathcal{D}) \\ &\leqslant L(f+g,\mathcal{D})\end{aligned}$$

Similarly,

$$\begin{aligned}\int_a^b f(x)dx + \int_a^b g(x)dx + 2\epsilon &\geqslant U(f,\mathcal{D}_1) + U(g,\mathcal{D}_2) \\ &\geqslant U(f,\mathcal{D}) + U(g,\mathcal{D}) \\ &\geqslant U(f+g,\mathcal{D})\end{aligned}$$

Combining two, we obtain

$$\int_a^b f(x)dx + \int_a^b g(x)dx - 2\epsilon \leqslant I_*(f+g) \leqslant I^*(f+g) \leqslant \int_a^b f(x)dx + \int_a^b g(x)dx + 2\epsilon$$

Since $\epsilon$ arbitrary, $f + g$ is integrable and

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

(4) By Theorem 3.5.3 and Corollary 3.5.6.1 (5), $\forall \epsilon > 0\ \exists \mathcal{D}$ such that

$$\epsilon > U(f,\mathcal{D}) - L(f,\mathcal{D}) \geqslant U(|f|,\mathcal{D}) - L(|f|,\mathcal{D})$$

We immediately see that $|f|$ is integrable by Theorem 3.5.3. Also,

$$-|f(x)| \leqslant f(x) \leqslant |f(x)| \quad \forall x$$

so by (1),

$$-\int_a^b |f(x)|dx \leqslant \int_a^b f(x)dx \leqslant \int_a^b |f(x)|dx$$

(5) Noting

$$fg = \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right] \tag{†}$$

it is sufficient to show $f^2$ is integrable. Suppose $|f(x)| \leqslant K\ \forall x \in [a,b]$. Then given $\epsilon > 0$, there exists $\mathcal{D}$ such that

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) < \frac{\epsilon}{2K}$$

By Corollary 3.5.6.1 (6),

$$U(f^2,\mathcal{D}) - L(f^2,\mathcal{D}) \leqslant 2\sup_{[a,b]}|f|(U(f,\mathcal{D}) - L(f,\mathcal{D})) < \epsilon$$

Hence $f^2$ is integrable, and consequently $fg$ is integrable by (†).

■

**Theorem 3.5.8.** (*Additivity of the Integral*) Suppose $f : [a,b] \to \mathbb{R}$ and let $c \in (a,b)$. Then $f$ is integrable if and only if $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable, and

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

*Proof.* First, suppose $f : [a,b] \to \mathbb{R}$ is integrable. Let $\epsilon > 0$. By Theorem 3.5.3, $\exists \mathcal{D}$, a dissection of $[a,b]$, such that

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) < \epsilon$$

By considering $\mathcal{D} \cup \{c\}$ if necessary, we can assume $\mathcal{D} = \{x_0,\ldots,x_n\}$ and $x_l = c$ for some $l$. Let $\mathcal{D}_L = \{x_0,\ldots,x_l\}$ and $\mathcal{D}_R = \{x_l,\ldots,x_n\}$, which are dissections of $[a,c]$ and $[c,b]$ respectively. Then, from

$$L(f,\mathcal{D}) = L(f|_{[a,c]},\mathcal{D}_L) + L(f|_{[c,b]},\mathcal{D}_R)$$
$$U(f,\mathcal{D}) = U(f|_{[a,c]},\mathcal{D}_L) + U(f|_{[c,b]},\mathcal{D}_R)$$

we have

$$\left[U(f|_{[a,c]},\mathcal{D}_R) - L(f|_{[a,c]},\mathcal{D}_R)\right] + \left[U(f|_{[c,b]},\mathcal{D}_R) - L(f|_{[c,b]},\mathcal{D}_R)\right] < \epsilon$$

Both terms in square brackets are non-negative, so

$$U(f|_{[a,c]},\mathcal{D}_R) - L(f|_{[a,c]},\mathcal{D}_R) < \epsilon$$

and

$$U(f|_{[c,b]},\mathcal{D}_R) - L(f|_{[c,b]},\mathcal{D}_R) < \epsilon$$

Therefore $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable.

Conversely, suppose $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable. Theorem 3.5.3 says that $\exists \mathcal{D}_L, \mathcal{D}_R$ such that

$$U(f|_{[a,c]},\mathcal{D}_R) - L(f|_{[a,c]},\mathcal{D}_R) < \epsilon$$
$$U(f|_{[c,b]},\mathcal{D}_R) - L(f|_{[c,b]},\mathcal{D}_R) < \epsilon$$

But if we let $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_R$, we have

$$U(f,\mathcal{D}) - L(f,\mathcal{D}) < 2\epsilon$$

Hence $f$ is integrable.

Finally, we deduce that

$$\int_a^b f(x)dx \geqslant L(f,\mathcal{D}) = L(f|_{[a,c]},\mathcal{D}_L) + L(f|_{[c,b]},\mathcal{D}_R) \geqslant \int_a^c f(x)dx - \epsilon + \int_c^b f(x)dx - \epsilon$$

Similarly,

$$\int_a^b f(x)dx \leqslant U(f,\mathcal{D}) = U(f|_{[a,c]},\mathcal{D}_L) + U(f|_{[c,b]},\mathcal{D}_R) \leqslant \int_a^c f(x)dx + \epsilon + \int_c^b f(x)dx + \epsilon$$

Therefore,

$$\left| \int_a^b f(x)dx - \int_a^c f(x)dx - \int_c^b f(x)dx \right| \leqslant 2\epsilon$$

for arbitrary $\epsilon$, completing the proof. ∎

Finally, we introduce a convention that if $a > b$,

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

and if $a = b$

$$\int_a^a f = 0$$

With this convention, it follows that if $|f| \leqslant K$,

$$\left| \int_a^b f(x)dx \right| \leqslant K|b-a|$$

### 3.5.3 Piecewise Continuous Functions

**Definition 3.5.4.** A bounded function $f : [a,b] \to \mathbb{R}$ is *piecewise continuous* if ther exists a dissection $\mathcal{D}$ of $[a,b]$ such that for each $i = 1,2,\ldots,n$, if we define $f_i : (x_{i-1},x_i) \to \mathbb{R}$, $x \mapsto f(x)$, then $f_i$ is continuous and has a limit as $x \to x_{i-1}$ and $x \to x_i$. I.e. the function $\bar{f}_i : [x_{i-1},x_i] \to \mathbb{R}$ given by

$$\bar{f}_i(x) = \begin{cases} \lim_{x\to x_{i-1}} f_i(x) & x = x_{i-1} \\ f_i(x) & x \in (x_{i-1},x_i) \\ \lim_{x\to x_i} f_i(x) & x = x_{i+1} \end{cases}$$

is continuous (hence integrable).

Consequently, from the observation obove, together with Theorem 3.5.8, piecewise continuous function is integrable. Furthermore,[9] if $f : [a,b] \to \mathbb{R}$ is integrable, and $g : [a,b] \to \mathbb{R}$ is bounded, with $g(x) = f(x)\ \forall x \in [a,b]\backslash\{y_1,\ldots,y_N\}$, then $g$ is integrable and

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

[9] Left as an exercise.

### 3.5.4 The Fundamental Theorem of Calculus

Suppose $f : [a,b] \to \mathbb{R}$ is integrable (hence bounded). Let

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

for $x \in [a,b]$.

**Theorem 3.5.9.** $F$ is continuous.

*Proof.* Suppose $|f| \leqslant K$ and assume $x, x+h \in [a,b]$. Then

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

gives

$$|F(x+h) - F(x)| = \left| \int_x^{x+h} f(t)dt \right| \leqslant K|h| \to 0$$

as $h \to 0$. Therefore, $\lim_{h\to 0}(F(x+h) - F(x)) = 0 \Rightarrow \lim_{h\to 0} F(x+h) = F(x)$ and $F$ is continuous. ■

**Theorem 3.5.10.** (*Fundamental Theorem of Calculus I*, FTC 1) If in addition $f$ is continuous at $x$, then $F$ is differentiable at $x$, and

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

*Proof.* We need to estimate

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| = \left| \frac{1}{h} \int_x^{x+h} f(t)dt - f(x) \right| = \frac{1}{|h|} \left| \int_x^{x+h} (f(t) - f(x))dt \right|$$

Now, given $\epsilon > 0$ $\exists \delta > 0$ such that

$$|t - x| < \delta \Rightarrow |f(t) - f(x)| < \epsilon$$

since $f$ is continuous. Suppose $|h| < \delta$. Then $|f(t) - f(x)| < \epsilon$ for all $t$ between $x$ and $x+h$. Therefore,

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| \leqslant \frac{1}{|h|} \epsilon |h| = \epsilon$$

i.e.

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

■

Note that for Theorem 3.5.10 to hold, the condition $f$ is continuous is necessary. For instance, consider

$$f(x) = \begin{cases} -1 & x \in [-1, 0] \\ 1 & x \in (0, 1] \end{cases}$$

Although $f$ is montone and thus integrable,

$$F(x) = \begin{cases} -x - 1 & x \leqslant 0 \\ x - 1 & x \geqslant 0 \end{cases}$$

i.e. $F(x) = -1 + |x|$ is not differentiable at $x = 0$.

**Corollary 3.5.10.1.** (*Integration is Inverse of Differentiation*) If $f = g'$ is continuous on $[a,b]$, then

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

*Proof.* $F - g$ has zero derivative on $[a,b]$ by Theorem 3.5.10, and $F(a) = 0$. Hence $(F-g)(a) = -g(a)$. So by Corollary 3.3.4.1,

$$(F-g)(x) = -g(a) \quad \forall x$$

and $F(x) = g(x) - g(a)$. ∎

Note that all continuous functions have an *indefinite integral*, or *anti-derivative*:

$$\int f(x)dx = \int_a^x f(t)dt$$

where $a$ arbitrary.

**Theorem 3.5.11.** (*Fundamental Theorem of Calculus II*, FTC 2) Suppose that $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$, differentiable on $(a,b)$ and that $f'$ is integrable. Then

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

N.b. Theorem 3.5.11 is stronger than Corollary 3.5.10.1 since it does not require $f'$ from Theorem 3.5.11 to be continuous.

*Proof.* Let $\mathcal{D}$ be any dissection of $[a,b]$. Applying the mean value theorem (Theorem 3.3.4) to $f$ on $[x_{j-1}, x_j]$, $\exists \xi_j \in (x_{j-1}, x_j)$ such that

$$f'(\xi_j)(x_j - x_{j-1}) = f(x_j) - f(x_{j-1})$$

This gives

$$\sum_{j=1}^{n} f'(\xi_j)(x_j - x_{j-1}) = f(b) - f(a)$$

But

$$\sum_{j=1}^{n}(x_j - x_{j-1}) \inf_{[x_{j-1},x_j]} f' \leqslant \sum_{j=1}^{n}(x_j - x_{j-1})f'(\xi_j) \leqslant \sum_{j=1}^{n}(x_j - x_{j-1}) \sup_{[x_{j-1},x_j]} f'$$

So

$$L(f', \mathcal{D}) \leqslant f(b) - f(a) \leqslant U(f', \mathcal{D}) \Rightarrow I_*(f') \leqslant f(b) - f(a) \leqslant I^*(f')$$

But because $f'$ is integrable,

$$I_*(f') = I^*(f') = \int_a^b f'(x)dx = f(b) - f(a)$$

∎

Following are the important corollarys of fundamental theorem of calculus.

**Corollary 3.5.11.1.** (*Integration by Parts*) Suppose $f'$ and $g'$ exist and are continuous on $[a,b]$. 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.* By the product rule and Corollary 3.5.10.1,

$$\int_a^b f'(x)g(x)dx = \int_a^b ((fg)'(x) - f(x)g'(x))dx = \int_a^b (fg)'(x)dx - \int_a^b f(x)g'(x)dx$$
$$= f(b)g(b) - f(a)g(a) - \int_a^b f(x)g'(x)dx$$

■

**Corollary 3.5.11.2.** (*Integration by Substitution*) Let $g : [\alpha, \beta] \to [a, b]$ with $g(\alpha) = a$, $g(\beta) = b$ and suppose $g'$ exists and is continuous on $[\alpha, \beta]$. Let $f : [a, b] \to \mathbb{R}$ be continuous. Then

$$\int_a^b f(x)dx = \int_\alpha^\beta f(g(t))g'(t)dt$$

*Proof.* Set $F(x) = \int_a^x f(t)dt$ as before. Let $h(t) = F(g(t))$ which is well defined for $t \in [\alpha, \beta]$ as $g$ takes value in $[a, b]$. Then,

$$\begin{aligned}
\int_\alpha^\beta f(g(t))g'(t)dt &= \int_\alpha^\beta F'(g(t))g'(t)dt \\
&= \int_\alpha^\beta h'(t)dt \\
&= h(\beta) - h(\alpha) \\
&= F(g(\beta)) - F(g(\alpha)) \\
&= F(b) - F(a) = \int_a^b f(t)dt
\end{aligned}$$

■

**Theorem 3.5.12.** (*Taylor's Theorem with Integral Remainder*) Suppose $f$ and its first $n$ derivatives exist, and are continuous for $x \in [0, h]$. Then,

$$f(h) = f(0) + hf'(0) + \cdots + \frac{h^{n-1}f^{(n-1)}(0)}{(n-1)!} + R_n$$

where

$$R_n = \frac{h^n}{(n-1)!}\int_0^1 (1-t)^{n-1}f^{(n)}(th)dt$$

*Proof.* By Theorem 3.5.11,

$$\begin{aligned}
f(h) &= f(0) + \int_0^h f'(u)du \\
&= f(0) - \left[(h-u)f'(u)\right]_0^h + \int_0^h (h-u)f''(u)du \\
&= f(0) + hf'(0) + \int_0^h \left(-\frac{d}{du}\frac{(h-u)^2}{2}\right) f''(u)du
\end{aligned}$$

Keep integrating by parts to get

$$f(h) = f(0) + hf'(0) + \cdots + \frac{h^{n-1}f^{(n-1)}(0)}{(n-1)!} + \underbrace{\int_0^h \frac{(h-u)^{n-1}}{(n-1)!}f^{(n)}(u)du}_{R_n}$$

Substitution $u = th$ gives the desired

$$R_n = \frac{h^n}{(n-1)!}\int_0^1 (1-t)^{n-1}f^{(n)}(th)dt$$

■

*Remark* 3.5.1. Note we assume $f^{(n)}$ is continuous but in previous versions of the Taylor's theorem we just assumed $f^{(n)}$ exists. If we make this additional assumption, we can recover Cauchy and Lagrange form of remainder.

**Theorem 3.5.13.** Suppose $f, g : [a, b] \to \mathbb{R}$ are continuous and $g(x) \neq 0$ for all $x \in (a, b)$. Then $\exists c \in (a, b)$ such that

$$\int_a^b f(x)g(x)dx = f(c)\int_a^b g(x)dx$$

N.b. if $g(x) = 1$, we get $\int_a^b f(x)dx = f(c)(b-a)$.

*Proof.* Apply Cauchy's mean value theorem (Theorem 3.3.6) to

$$F(x) = \int_a^x (fg)(x)dx \text{ and } G(x) = \int_a^x g(x)dx$$

to deduce that there exists $c \in (a, b)$ such that

$$(F(b)-F(a))G'(c) = F'(c)(G(b)-G(a)) \Rightarrow \left(\int_a^b (fg)(x)dx\right)g(c) = f(c)g(c)\int_a^b g(x)dx$$

$g(c) \neq 0$ so we can cancel out $g(c)$ to obtain the desired result. ■

Now consider

$$R_n = \frac{h^n}{(n-1)!}\int_0^1 (1-t)^{n-1} f^{(n)}(th)dt$$

To obtain the Cauchy remainder, apply Theorem 3.5.13 with $\tilde{g} = 1$ and $\tilde{f}(t) = (1-t)^{n-1}f^{(n)}(th)$ to get $\exists \theta \in (0,1)$ such that

$$\int_0^1 \tilde{f}(t)dt = \tilde{f}(\theta) \Rightarrow R_n = \frac{h^n}{(n-1)!}(1-\theta)^{n-1}f^{(n)}(\theta h)$$

Alternatively, split integrands as

$$\underbrace{\left((1-t)^{n-1}\right)}_{\tilde{g}}\underbrace{\left(f^{(n)}(th)\right)}_{\tilde{f}}$$

to find

$$\int_0^1 (1-t)^{n-1}f^{(n)}(th)dt = f^{(n)}(\theta h)\int_0^1 (1-t)^{n-1}dt$$

for some $\theta \in (0,1)$, by theorem 3.5.13. But $\int_0^1 (1-t)^{n-1}dt = 1/n$, so we obtain the Lagrange remainder

$$R_n = \frac{h^n}{n!}f^{(n)}(\theta h)$$

Finally, we will point out that it is necessary to include the assumption that $f'$ is integrable in the hypotheses of Theorem 3.5.11. For example, consider $f : [-1, 1] \to \mathbb{R}$, given by

$$f(x) = \begin{cases} |x|^{3/2}\sin\frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

satisfies that $f$ is continuous on $[-1, 1]$, differentiable on $(-1, 1)$, but $f'$ is unbounded hence not integrable.

### 3.5.5 Improper Integrals

The theory of integration we have developed so far applies to bounded functions defined on bounded intervals. For a full treatment for functions whose domain or image is unbounded, we have to wait for the Lebesgue integral.[10] Note that content of this subsection, Riemann integral, is a partial resolution.

**⁎ Integrals on an infinite domain**

Suppose $f : [a, \infty) \to \mathbb{R}$ is integrable on each interval $[a, R]$ $(R > a)$ and

$$\int_a^R f(x)dx \to l$$

as $R \to \infty$. Then we say $\int_a^\infty f(x)dx$ exists, or converges, and

$$\int_a^\infty f(x)dx = l$$

Otherwise, if $\int_a^R f(x)dx$ has no limit, we say $\int_a^\infty f(x)dx$ *diverges*. A similar definition applies to $\int_{-\infty}^a f(x)dx$ for $f : (-\infty, a] \to \mathbb{R}$.

Furthermore, if $\int_{-\infty}^a f(x)dx = l_1$ and $\int_a^\infty f(x)dx = l_2$ we say $\int_{-\infty}^\infty f(x)dx$ exists, and

$$\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^a f(x)dx + \int_a^\infty f(x)dx$$

independent of $a$.[11] Be careful that this is not the same as

$$\int_{-R}^R f(x)dx$$

converging. For instance, $\int_{-R}^R xdx = 0$ but $\int_0^\infty xdx$ does not converge.

**⁎ Integrands with an isolated singularity**

Suppose $f : (b, c] \to \mathbb{R}$ is integrable on each $[b + \delta, c]$ for $0 < \delta < c - b$. Then if

$$\int_{b+\delta}^c f(x)dx \to l$$

as $\delta \to 0$, we say $\int_b^c f(x)dx$ exists or converges and equals $l$; and it is similarly defined for $f : [a, b) \to \mathbb{R}$ integrable on $[a, b - \delta]$ $(0 < \delta < b - a)$.

If $f : [a, c] \backslash \{b\} \to \mathbb{R}$ and

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

exist, we say $\int_a^c f(x)dx$ exists, and

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

**⁎ Properties and examples of improper integral**

For example,

$$\int_1^\infty \frac{dx}{x^k}$$

[10] See Probability and Measure, Part II.
[11] This needs check.

converges if and only if $k > 1$, and

$$\int_0^1 \frac{dx}{x^k}$$

converges if and only if $k < 1$, since ($k \neq 1$)

$$\int_0^R \frac{dx}{x^k} = \left[\frac{x^{1-k}}{1-k}\right]_1^R = \frac{R^{1-k}-1}{1-k}$$

- converges as $R \to \infty$ if $1-k < 0$, not if $1-k > 0$;
- converges as $R \to 0$ if $1-k > 0$, not if $1-k < 0$;

and because if $k = 1$,

$$\int_1^R \frac{dx}{x} = [\ln x]_1^R = \ln R$$

does not converge as $R \to \infty$ or $R \to 0$.

**Lemma 3.5.14.** Suppose $f, g : [a, \infty) \to \mathbb{R}$ are integrable on each $[a, R]$, $R > a$. If $0 \leqslant f(x) \leqslant g(x)$ for all $x \geqslant a$, then

$$\int_a^\infty g(x)dx \text{ converges } \Rightarrow \int_a^\infty f(x)dx \text{ converges}$$

*Proof.* Note that

$$R \mapsto \int_a^R f(x)dx$$

is increasing as $f(x) \geqslant 0$, and

$$\int_a^R f(x)dx \leqslant \int_a^R g(x)dx \leqslant \int_a^\infty g(x)dx$$

Hence $R \mapsto \int_a^R f(x)dx$ is increasing and bounded above. Let

$$l = \sup_{R \geqslant a} \int_a^R f(x)dx$$

and claim

$$l = \lim_{R \to \infty} \int_a^R f(x)dx$$

To see this, let $\epsilon > 0$. Then from definition of supremum, $\exists R_0$ such that

$$\int_a^{R_0} f(x)dx \geqslant l - \epsilon$$

So if $R \geqslant R_0$,

$$l - \epsilon \leqslant \int_a^{R_0} f(x)dx \leqslant \int_a^R f(x)dx \leqslant l$$

■

**Example 3.5.3.** Consider

$$\int_0^\infty e^{-x^2/2}dx$$

For $x > 1$, $e^{-x^2/2} \leqslant e^{-x/2}$, and

$$\int_1^R e^{-x/2}dx = 2(e^{-1/2} - e^{-R/2}) \to 2e^{-1/2}$$

as $R \to \infty$. Hence $\int_1^\infty e^{-x^2/2} dx$ converges, and therefore

$$\int_0^\infty e^{-x^2/2} dx$$

converges because from

$$\int_0^R e^{-x^2/2} dx = \int_0^1 e^{-x^2/2} dx + \int_1^R e^{-x^2/2} dx$$

left hand side converges as $R \to \infty$ if and only if $\int_1^R e^{-x^2/2} dx$ does.

However, be careful that, unlike the situation with series,

$$\int_0^\infty f(x)dx \text{ converges } \not\Rightarrow f(x) \to 0 \text{ as } x \to \infty$$

∗ **Integral test**

**Theorem 3.5.15.** (*Integral Test*) Suppose $f : [1, \infty) \to \mathbb{R}$ is positive and decreasing. Then,

(1) the integral $\int_1^\infty f(x)dx$ converges if and only if the sum $\sum_{n=1}^\infty f(n)$ converges.

(2) as $n \to \infty$,

$$\sum_{r=1}^n f(r) - \int_1^n f(x)dx \to l$$

for some $l$ with $0 \leqslant l \leqslant f(1)$.

Note that (2) of Theorem 3.5.15 implies (1) of that. Also $f$ decreasing from the hypotheses of that implies $f$ integrable on $[1, R]$ $(R > 1)$ by Theorem 3.5.4.

*Proof.* If $n - 1 \leqslant x \leqslant n$,

$$f(n-1) \geqslant f(x) \geqslant f(n) \Rightarrow f(n-1) \geqslant \int_{n-1}^n f(x)dx \geqslant f(n)$$

Adding, we have

$$\sum_{r=1}^{n-1} f(r) \geqslant \int_1^n f(x)dx \geqslant \sum_{r=2}^n f(r)$$

From here, claim (1) is clear since $\int_1^R f(x)dx$ is monotonically increasing, as is $\sum_{r=1}^n f(r)$.

Let

$$\phi(n) = \sum_{r=1}^n f(r) - \int_1^n f(x)dx$$

Then,

$$\phi(n) - \phi(n-1) = f(n) - \int_{n-1}^n f(x)dx \leqslant 0$$

and

$$\phi(n) = f(n) + \sum_{r=1}^{n-1} f(r) - \int_1^n f(x)dx \geqslant f(n) \geqslant 0$$

So $\phi(n)$ is decreasing and bounded below. Hence, from monotone convergence theorem, $\phi(n) \to l$ with $0 \leqslant l \leqslant \phi(1) = f(1)$. ■

**Example 3.5.4.**

(1) $\sum_{n=1}^\infty 1/n^k$ converges if $k > 1$. We saw $\int_1^\infty (dx/x^k)$ converges if $k > 1$. So integral test gives result.

(2) Consider
$$\sum_{n=2}^{\infty} \frac{1}{n \ln n}$$
Let $f(x) = 1/x \ln x$ for $x \geqslant 2$. Then,
$$\int_2^R \frac{dx}{x \ln x} = [\ln(\ln x)]_2^R = \ln(\ln R) - \ln(\ln 2) \to \infty$$
as $R \to \infty$, and the corresponding sum diverges by integral test.

**Corollary 3.5.15.1.** (*Euler-Mascheroni Constant*) As $n \to \infty$,
$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n \to \gamma$$
with $0 \leqslant \gamma \leqslant 1$.[12]

*Proof.* Set $f(x) = 1/x$ in Theorem 3.5.15 (2). ■

### 3.5.6 Lebesgue's Criterion for Riemann Integrability

**Theorem 3.5.16.** (*Lebesgue's Criterion*) If $f : [a, b] \to \mathbb{R}$ is bounded, then $f$ is Riemann integral if and only if
$$\{x \mid f \text{ discontinuous at } x\}$$
has *measure zero*.

**Definition 3.5.5.** For an interval $I$, let $|I|$ denote its length. A subset $A \subset \mathbb{R}$ has *measure zero* if $\forall \epsilon > 0$ there exists a countable collection of intervals $I_j$ such that
$$A \subset \bigcup_{j=1}^{\infty} I_j \text{ and } \sum_{j=1}^{\infty} |I_j| < \epsilon$$
i.e. "can cover $A$ with countable numbers of infinitesimal intervals."

**Lemma 3.5.17.**

(1) If $B$ has measure zero and $A \subset B$, then $A$ has measure zero.

(2) If $A_k$ has measure zero for each $k \in \mathbb{N}$, then $\bigcup_{k \in \mathbb{N}} A_k$ has measure zero.

*Proof.*

(1) Obvious from definition.

(2) Fix $\epsilon > 0$. For each $k$, pick $I_j^k$ $(j = 1, 2, \ldots)$ such that
$$A_k \subset \bigcup_{j=1}^{\infty} I_j^k \text{ and } \sum_{j=1}^{\infty} |I_j^k| < \epsilon 2^{-k}$$
Then consider $\{I_j^h\}_{j,h=1}^{\infty}$.

■

**Will return.**

[12] Question: is $\gamma$ irrational?

## 3.6 Example Sheets

### 3.6.1 Sheet 1