

# **Math remark**

## **fondation for analysis and probability**

X

ElegantL<sup>A</sup>T<sub>E</sub>X Program

*Update: April 8, 2025*

# Contents

<b>1</b>	<b>Genenralisation of several inequalities</b>	<b>3</b>
1.1	Jense,Hölder,Minkeoski's inequalities . . . . .	3
1.2	Markov, Tchebychev, Cantelli's inequalities . . . . .	5
<b>2</b>	<b>Classic anaylsis theory and technic</b>	<b>7</b>
2.1	Power series and analytic function . . . . .	7
2.2	Anayltic continuation . . . . .	13
2.3	Limit superior and limit inferior . . . . .	16
2.4	Theorem of transfert: algebra of the random variable . . . . .	18
<b>3</b>	<b>Complex analysis</b>	<b>21</b>
3.1	Holomorphic Function . . . . .	21
<b>4</b>	<b>Functional Anaylsis</b>	<b>25</b>
4.1	Hilbert Space . . . . .	25
4.1.1	prerequisite about inner product . . . . .	26
4.1.2	Hilbert base . . . . .	30
<b>5</b>	<b>probaility</b>	<b>31</b>
5.1	Independence . . . . .	31
5.2	Convergence . . . . .	32
5.3	Law of large number . . . . .	33

# 1 Genenralisation of several inequalities

## 1.1 Jense,Hölder,Minkeoski's inequalities

The description of the Jense inequality depends on the properties of the convex function, it is a strong inequalities which can be applied in many places. We should pay a little attention to the outline of the section:

$$\text{Jense} \Rightarrow \text{Hölder} \Rightarrow \text{Minkeoski}$$

### Theorem 1.1 (Jense)

Suppose  $\varphi : I \rightarrow \mathbb{R}$  is a convex function defined on an interval,  $(X, \mathcal{E}, \mu)$  is a probability space and  $f \in L^1(X)$  with  $\text{im} f \subset I$ , then  $\int_X f d\mu \in I$  and  $\varphi \circ f$  is integrable such that

$$\varphi\left(\int_X f d\mu\right) \leq \int_X \varphi \circ f d\mu$$

The equality holds iff  $f$  is constant almost everywhere.

**Proof.** ■

**Remark** There are many other forms of the Jense's inequality, we take some exemples.

- **Finite forms:** the motivation of the inequality comes from the definition of the convex function, i.e. the real-valued function satisfies

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

for any defined  $x, y$  and  $t \in [0, 1]$ , here a question about the distribution of the weights appears, which is the core of the convex function. we can generalize the inequalities by applying the weights to the  $n$  different points in interval such that  $\sum_i w_i = 1$ , then we can conclude the inequality:

$$f\left(\sum_i w_i x_i\right) \leq \sum_i w_i f(x_i)$$

notice that the defined domaine usually is a convex set, which ensures the effectivity of  $f(\sum_i w_i x_i)$ .

- **Expectation:** By a simple change of the notation, the Jense inequality in a probaility space can be wriiten as the form:

$$\varphi(E[X]) \leq E[\varphi(X)]$$

Applying a classic convex function  $t \mapsto t^2$  we can get the important inequality in proba-bility:

$$E^2[X] \leq E[X^2]$$

- **Concave:** some function like  $t \mapsto \ln t$  is a concave function, the Jense's inequality can be just changed the order of the inequality. The reason is simple, if  $f$  is a concave function, then  $-f$  will be a convex function.

The classic proof of the Hölder's inequality covers the inequality of Young:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

for any  $a, b \geq 0$  and  $p, q > 1$  such that  $1/p + 1/q = 1$ . The complete proof can be found in [Rudin 1 Ex 6.10]. I don't choose the proof here for a comparison of strength of the different inequality. Here the proof is elegant and given by Mon.Mardare in TD, and a similar proof via Jense can be found in [SU TD1 EX15-16].

### Theorem 1.2 (Hölder)

Suppose that  $p, q > 1$  and  $1/p + 1/q = 1$ , for any two mesurable functions  $f, g : (X, \mathcal{E}, \mu) \rightarrow \mathbb{C}$ , we have

$$\int_X |fg| d\mu \leq \left( \int_X |f|^p d\mu \right)^{1/p} \left( \int_X |g|^q d\mu \right)^{1/q}$$

the equality holds iff  $|f| = c|g|$  almost everywhere (u-p.p) for some constant  $c$ .

**Proof.** Let  $u = \frac{|f|}{\|f\|_p}$  and  $v = \frac{|g|}{\|g\|_q}$ , then notice that  $\|u\|_p = \|v\|_q = 1$ . We know  $t \mapsto \ln t$  is a concave function on  $(0, +\infty)$ , so we can estimate by Jense's inequality

$$\begin{aligned} \ln(uv) &= \ln(u + v) = \frac{1}{p} \ln u^p + \frac{1}{q} \ln v^q \\ &\leq \ln\left(\frac{1}{p} u^p + \frac{1}{q} v^q\right) \end{aligned}$$

by the monotone of the function, we have  $uv \leq \frac{1}{p} u^p + \frac{1}{q} v^q$ , hence we can conclude

$$\begin{aligned} \frac{\int_X |fg| d\mu}{\left( \int_X |f|^p d\mu \right)^{1/p} \left( \int_X |g|^q d\mu \right)^{1/q}} &= \int_X uv d\mu \\ &\leq \frac{1}{p} \int_X u^p d\mu + \frac{1}{q} \int_X v^q d\mu \\ &= \frac{1}{p} \|u\|_p^p + \frac{1}{q} \|v\|_q^q = 1 \end{aligned}$$

finally, we discuss the equality. By Jense, we know the equality holds iff  $u = v$ . Notice that if  $u = 0$  or  $v = 0$  almost everywhere, then the inequality can be reduced to  $0 \leq 0$ , it holds, otherwise we can get that  $|f| = \frac{\|f\|_p}{\|g\|_q} |g|$ , so we finish the proof. ■

### Corollary 1.2.1 (Cauchy-Schwartz)

For any two square integrable function  $f, g \in L^2_{\mathbb{C}}(\mathbb{R}^n)$ , we have

$$\left| \int_{\mathbb{R}^n} f \bar{g} dl \right|^2 \leq \int_{\mathbb{R}^n} |f|^2 dl \cdot \int_{\mathbb{R}^n} |g|^2 dl$$

The equality holds iff  $|f| = c|g|$  almost everywhere (u-p.p) for some constant  $c$ .

**Proof.** Although it is the special case of Hölder when  $p = q = 2$ , but usually in a Hilbert space we have the beautiful form as following

$$| \langle u, v \rangle | \leq \|u\| \|v\|$$

the inequality has a good geometric intuition, and the proof of it is very beautiful and elementary. Notice  $\langle u, v \rangle \in \mathbb{C}$ , so there exists  $z \in \mathbb{C}$  such that  $|z| = 1$  and  $z \langle u, v \rangle = | \langle u, v \rangle |$ , and we let  $p(t) = \langle tzu + v, tzu + v \rangle$  defined on  $\mathbb{R}$ , then

$$\begin{aligned} p(t) &= t^2 z \bar{z} \langle u, u \rangle + tz \langle u, v \rangle + t \bar{z} \langle v, u \rangle + \langle v, v \rangle \\ &= t^2 z \bar{z} \langle u, u \rangle + tz \langle u, v \rangle + t \overline{z \langle u, v \rangle} + \langle v, v \rangle \\ &= t^2 \|u\|^2 + 2 | \langle u, v \rangle | t + \|v\|^2 \end{aligned}$$

so  $p(t)$  can be arranged to be a quadratic polynomial with respect to real value  $t$ , and  $p(t) = \|tzu + v\|^2 \geq 0$ , so we have sufficient and necessary condition that

$$\Delta = 4 | \langle u, v \rangle |^2 - 4 \|u\|^2 \|v\|^2 \geq 0$$

which is the inequality we hope to get, and  $\Delta = 0$  happens iff the polynomial satisfies  $p(t) = (t\|u\| + \|v\|)^2 = 0$ . ■

### Theorem 1.3 (Minkowski)

For any  $p \geq 1$ , suppose that  $f, g : (X, \mathcal{E}, \mu) \rightarrow \mathbb{C}$  are two measurable functions, then we have

$$\left( \int_X |f + g|^p d\mu \right)^{1/p} \leq \left( \int_X |f|^p d\mu \right)^{1/p} + \left( \int_X |g|^p d\mu \right)^{1/p}$$

The equality holds iff  $|f| = c|g|$  almost everywhere (u.p.p) for some constant  $c$ .

**Proof.** If ■

Review some basic inequalities (discrete)...

## 1.2 Markov, Tchebychev, Cantelli's inequalities

This section covers some basic inequalities in properties, They are always very useful when estimation. And in this section, we always use  $(\Omega, \mathcal{T}, P)$  to denote a probability space

### Theorem 1.4 (Markov)

Suppose that  $Y : \Omega \rightarrow \mathbb{R}^+$  is a random variable, then for any  $t > 0$ , we have

$$P(Y \geq t) \leq \frac{E(Y)}{t}$$

**Proof.** We notice that  $t \cdot \mathbb{1}_{\{Y \geq t\}} \leq Y$ , it is easily to prove. If  $x \in \{Y \geq t\}$ , then  $t \cdot \mathbb{1}_{\{Y \geq t\}}(x) = t \leq Y(x)$ , otherwise  $t \cdot \mathbb{1}_{\{Y \geq t\}}(x) = 0 \leq Y(x)$ , so by monotone of the expectation

$$E[t \cdot \mathbb{1}_{\{Y \geq t\}}] = tE[\mathbb{1}_{\{Y \geq t\}}] = tP(Y \geq t)$$

■

### Theorem 1.5 (Tchebychev)

Suppose that  $Y : \Omega \rightarrow \mathbb{R}$  is a random variable in  $L^2(\Omega)$  (square integrable), then for any  $t > 0$  we have

$$P(|Y - E[Y]| \geq t) \leq \frac{V(Y)}{t^2}$$

**Proof.** We apply Markov to prove it, we can estimate

$$\begin{aligned} P(|Y - E[Y]| \geq t) &= P((Y - E[Y])^2 \geq t^2) \\ &\leq \frac{E[(Y - E[Y])^2]}{t^2} \\ &= \frac{V(Y)}{t^2} \end{aligned}$$

■

### Theorem 1.6 (Cantelli)

Suppose that  $Y : \Omega \rightarrow \mathbb{R}$  is a random variable in  $L^2(\Omega)$  (square integrable), then for any  $t > 0$  we have

$$\begin{aligned} P(Y - E[Y] \geq t) &\leq \frac{V(Y)}{V(Y) + t^2} \\ P(|Y - E[Y]| \geq t) &\leq \frac{2V(Y)}{V(Y) + t^2} \end{aligned}$$

**Proof.** The proof of the inequality need to use inequality: for any  $a, b > 0$ ,

$$P(X \geq a) \leq \frac{E[(X + b)^2]}{(a + b)^2} \quad (1)$$

We write  $P(X \geq a) = E[\mathbb{1}_{\{X \geq a\}}]$ , then we just need to prove  $\mathbb{1}_{\{X \geq a\}} \leq (\frac{X+b}{a+b})^2$ . For any  $w \in \{X \geq a\}$ , we have  $(\frac{X+b}{a+b})^2(w) \geq (\frac{a+b}{a+b})^2 = 1 = \mathbb{1}_{\{X \geq a\}}(w)$ , otherwise will conclude  $(\frac{X+b}{a+b})^2 \geq 0$ , so the inequality 1 is valid.

For proove the Cantelli, we let  $Z = \frac{Y - E[Y]}{\sigma(Y)}$ , then by above inequality

$$\begin{aligned} P(Y \geq t) &\leq \frac{E[(Y + b)^2]}{(t + b)^2} \\ &= \frac{E[Y^2] + 2bE[Y] + b^2E[1]}{(t + b)^2} \\ &= \frac{1 + b^2}{(t + b)^2} \end{aligned}$$

We can put  $b = 1/t$ , then  $\frac{1+b^2}{(t+b)^2} = \frac{1}{1+t^2}$ , and then rewrite the inequality we get  $P(Y - E[Y] \geq t) \leq \frac{V(Y)}{V(Y)+t^2}$ . and notice that

$$P(Y - E[Y] \leq -t) = P((-Y) - E[-Y] \geq t)$$

and  $V(-Y) = V(Y)$  implies the same upper bound, so we finish the proof. ■

The importance of the three inequalities is the estimation of the upper bound of some probability, if we can know the the expected value, we can get an upper bound by Markov; if we also know the variance, a more precise upper bound can be given by Tchebychev and Cantelli.

Cantelli's inequality is a work based on the Tchebychev's inequality, if we rewrite the Tchebychev's inequality by take  $t = \sigma(Y)t$ , then we have

$$P(|X - \mu| \leq \sigma t) \geq \frac{1}{t^2}$$

Which always appears in the some statistical object like confidence interval, but it is a two-side interval estimation. If we rewrite Cantelli's inequality similarly:

$$P(X - \mu \leq \sigma t) \geq \frac{1}{1 + t^2}$$

$$P(|X - \mu| \leq \sigma t) \geq \frac{2}{1 + t^2}$$

We can find that Cantelli's inequality can do one-side estimation and it provides a worse bound then Tchebychev's when  $t \geq \sigma$ .

One appliaction of the Tchebychev is to prove the law of large theorem, suppose that  $X_1, X_2, \dots, X_n$  are the sequence of random variable with same distribution, that means

$$E\left[\frac{X_1 + \dots + X_n}{n}\right] = \mu \quad V\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

then using the Tchebychev's inequality we have

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq c\right) \leq \frac{\sigma^2}{nc^2}$$

Hence as  $n \rightarrow \infty$ , then the probability that the average of the random variable is not the exception  $\mu$  will be zero, that what law of large number want to say: when the number of sample is enough big, then the mean value of the sample is nearly the real mean value.

## 2 Classic analysis theory and technic

### 2.1 Power series and analytic function

This section conatains Three parts:

- overview for power series and the important result
- real analytic function and Abel's theorem
- the superiority of the holomorphic function

Power series is an special case of the series of the function, it gives by the following sequence:

$$f_n : D \rightarrow \mathbb{C}, \quad z \mapsto a_n z^n$$

more general, we can write it as the form  $a_n(z - z_0)^n$ , but notice that we can just change variable by  $u = z - z_0$ , it is a translation, so clearly a diffeomorphism. In above statement  $D$  denote the domain where the series will be well-defined, i.e. convergent, so here needs a clear definition.

### Lemma 2.1 (Abel's Lemma)

Given a complex-valued sequence  $(a_n)_{n \in \mathbb{N}}$ , if there exists  $z_0 \in \mathbb{C}$  such that  $(a_n z_0^n)_{n \in \mathbb{N}}$  is bounded, then we can conclude that

-If  $|z| < |z_0|$ , then power series  $\sum_{n \in \mathbb{N}} a_n z^n$  converge.

-If  $|z| < |z_0|$ , then power series  $\sum_{n \in \mathbb{N}} a_n z^n$  diverge.

**Proof.** The bounded sequence implies a real number  $M$  such that  $a_n z_0^n \leq M$ , so

$$|a_n z^n| = |a_n z_0^n \cdot \frac{z^n}{z_0^n}| \leq M \left| \frac{z}{z_0} \right|^n$$

so clearly the power series converge if  $|z| \leq |z_0|$  by M-test.

If  $|z| > |z_0|$ , then  $(a_n z^n)_{n \in \mathbb{N}}$  is not bounded, so immediately it can't converge to zero, so the power series diverge. ■

The lemma inspire us that the domain is nearly a disk, although we can not determine the convergence of the boundary, but we can formally define the **radius of convergence** of a power series  $\sum_{n \in \mathbb{N}} a_n z^n$  is given by

$$R = \sup\{r \geq 0 \mid (a_n r^n)_{n \in \mathbb{N}} \text{ is bounded}\}$$

and we call  $D = D(0, R)$  the **domain of convergence**. A more simple form of the radius of convergence is given by Cauchy and d'Alembert:

$$1/R = \limsup |a_n|^{1/n} = \lim \left| \frac{a_{n+1}}{a_n} \right|$$

and we convented here  $1/0 = \infty$  and  $1/\infty = 0$ .

The convergence of the power series provides a good case of function, analytic function, which is the theme of the next section. We put function

$$f : \overline{D(0, r)} \rightarrow \mathbb{C}, \quad dz \mapsto \sum_{n \in \mathbb{N}} a_n z^n$$



where  $r \in (0, R)$ , and we suppose here  $R > 0$ . The function will be well-defined and continuous since  $z \mapsto a_n z^n$  is always uniformly continuous on a compact set, so we can conclude that  $f$  is well-defined and continuous on domain convergence when we take  $r \rightarrow R$ . We will take about the derivable of the power series, it is a important property.

### Theorem 2.2 (Derivable)

The power series  $f(z) = \sum_{n \in \mathbb{N}} a_n z^n$  defined on the domaine  $D(0, R)$  with raduis  $R > 0$ , then the derivative of  $f$  is also a power serieis with the same raduis of convergence and has the form

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

**Proof.** ■

An immediate result is that power series is a holomorphic function, and another result is an inversion of the process, that means if we know a function definitely is converged by some power series in some domaine, then we can write each coefficient in terms of the derivative of the function, usually in calculus we call it the Taylor's Formula.

### Corollary 2.2.1 (Taylor's series)

If  $f(z)$  is the sum of the power series  $\sum_{n \in \mathbb{N}} a_n z^n$  on the domaine  $D(0, R)$  with  $R > 0$ , then we have

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

**Proof.** Manipulate above theorem to  $f$  k times we get

$$f^{(k)}(x) = \sum_{n \in \mathbb{N}} n(n-1) \cdots (n-k+1) a_n z^{n-k}$$

and we put  $z = 0$ , then we get  $n! a_n = f^{(k)}(0)$ , that is what we want. ■

we can also conclude the **analytic continuation** about the power series, although it can be got direct from holomorphic function, but the proof here is easy and it gives us the connection.

**Corollary 2.2.2** Suppose that  $\sum_{n \in \mathbb{N}} a_n z^n$  and  $\sum_{n \in \mathbb{N}} b_n z^n$  are two power series with the positive raduis  $R_1$  and  $R_2$ , if there exists an voisinage of zero  $V$  such that  $\sum_{n \in \mathbb{N}} a_n z^n = \sum_{n \in \mathbb{N}} b_n z^n$ , then  $a_n = b_n$  for any  $n \in \mathbb{N}$ .

**Proof.** We Let  $f(z) = \sum_{n \in \mathbb{N}} a_n z^n$  and  $g(z) = \sum_{n \in \mathbb{N}} b_n z^n$  on domaine, then we have

$$0 = (f - g)(z) = \sum_{n \in \mathbb{N}} (a_n - b_n) z^n$$

on the subset of the domain  $V$ , so we applying above theorem we can get that

$$a_n - b_n = n!(f - g)(0) = 0$$

Which is the result we get, ■

After some basic description of power series, we can define the **analytic function**,

**Definition 2.1** Let  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ ,  $U$  is a open set in  $\mathbb{K}$ , a function  $f : U \rightarrow \mathbb{K}$  is an analytic function if for any  $z_0 \in U$ , the function can be converged by a power series in a small voisinage of  $z_0$ .

formally speaking, for any  $z_0 \in U$ , there exists  $r > 0$  and a sequence  $(a_n)_{n \in \mathbb{N}}$  such that in the domaine  $D(z_0, r) \subset U$  we have

$$f(z) = \sum_{n \in \mathbb{N}} a_n (z - z_0)^n$$

**Remark** a power series defined on a domain is clearly an analytic function, the choice of  $r$  can be formulated by  $\min\{|z_0|, R - |z_0|\}$ , where  $R$  is the raduis of convergence.

The classic discussion about the analytic function is that whether analytic is a smooth function ( $C^\infty$ ), and conversely a smooth function is a analytic function? In the time of Abel, analysis is focused on  $\mathbb{R}$ , and a smooth function is usually considered to do expansion by Talyor's Formula, so the question in  $\mathbb{R}$  is equivalent to the question that for each smooth function whether its Taylor seriees converges to itself or not. In conclusion, the relation show as folloing in  $\mathbb{R}$

$$\text{analytic function} \subset C^\infty \text{smooth function}$$

### Proposition 2.3 (infinitely differentiable)

For any analytic function defined on an open set of  $\mathbb{K}$ , it is infinitely differentiable and its derivative is still an analytic function.

**Proof.** We let  $f : U \rightarrow \mathbb{K}$ , and take  $a \in U$ , the analytic function implies a positive  $r > 0$  and a seauqnce  $(a_n)$ , such that  $f(z) = \sum a_n (z - a)^n$  for any  $z \in D(a, r)$ , then applying theorem 2.2 we can conclude that

$$f^{(k)} = k!a_k$$

for any  $k \in \mathbb{N}$ , hence it is infinitely differentiable. What's more, for any  $z \in D(a, r)$ ,  $f'(z) = \sum n a_n (z - a)^{n-1}$  by the same theorem, which gurantee that the derivative is still an analytic function. ■

Now we will give an example that a real-value function is smooth but not analytic to finish our conclusion.

**Example 2.1** we consider the following real function

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} e^{-1/x^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Notice that  $x \mapsto e^{-1/x^2}$  is a  $C^\infty$  function for any  $x \neq 0$ , and we can prove by recurrence that its  $n$ -times derivative has the form

$$f^{(n)}(x) = \frac{P_n(x)}{x^{3n}} e^{-1/x^2}$$

where  $P_n \in \mathbb{R}[x]$ . Then calculate the limite by changing  $u = 1/x$ , we know

$$\lim_{x \rightarrow 0} \frac{e^{-1/x^2}}{x^{3n}} = \lim_{u \rightarrow \infty} \frac{u^{3n}}{e^{u^2}} = 0$$

so for any  $n \in \mathbb{N}$ ,  $f^{(n)}(0)$  tends to zero, so we extend the function by adding  $f^{(n)}(0) = 0$  such that  $f \in C^\infty(\mathbb{R})$ .

Suppose that  $f$  is analytics at  $x = 0$ , then we can write the Talyor's series at the voinsinage of zero

$$f(x) = \sum_{n \in \mathbb{N}} \frac{f^{(n)}(0)}{n!} x^n = 0$$

which means the function equals to null in the voisinage, clearly absurd since the series is not convergent to itself. Hence,  $f$  can not be analytic at zero although it is a smooth function.

Another work of Abel is about the convergence on the boundary of domaine of convergence, we give it here the version in  $\mathbb{R}$ .

**Theorem 2.4 (Abel's Theorem)**

Let  $f(z) = \sum_{n \in \mathbb{N}} a_n z^n$  be a power seriees defined on a domaine  $(-R, R)$  and  $R > 0$ . If the power series converges at  $z = R$ (or  $z = -R$ ), then the  $f$  is continous (one-sid) at  $z = R$ (or  $z = -R$ ). Or simply, the theorem is equivalent to porve that

$$\lim_{r \rightarrow 1^-} \sum_{n \in \mathbb{N}} a_n r^n = \sum_{n \in \mathbb{N}} a_n$$

if  $\sum_{n \in \mathbb{N}} a_n$  is a covergent series.

**Proof.** the rigorous proof needs to use the **summation by parts formula**:

$$\sum_{n=M}^N a_n b_n = a_N B_N - a_M B_{M-1} - \sum_{n=M}^{N-1} (a_{n+1} - a_n) B_n$$

where  $a_n$  and  $b_n$  are all finite, and  $B_n = \sum_{k=1}^n b_k$  with convention  $B_0 = 0$ . ■

Finally i will talk about holomorphic function. Loosely speaking, holomorphic function is just the differentiable functions in  $\mathbb{C}$ , why do we not just call them "a differentiable function" and give them a special name? The important reason is the crazy regularity of the complex-valued function. By **Cauchy's integral formula**, for any open set containing an open disc  $D$ , if a function  $f$  is holomorphic in the open set, then for any  $z \in D$

$$f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z} d\zeta$$

then by caculation

$$\frac{f(z+h) - f(z)}{h} = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z)(\zeta - z - h)} d\zeta$$

By the continuity of the parametric integration at  $h = 0$ , we can conclude that  $f$  has derivative and  $f'(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z)^2} d\zeta$ , hence we can prove by recurrence that  $f$  is infinitely times differentiable and with the form

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z)^{n+1}} d\zeta \quad (2)$$

That is an amazing result and also a fundamental theorem for complex analysis, so clearly, any holomorphic function is a  $C^\infty$  function, also the most important points is that the high-order derivative only depends on the original function. We recall the theorem 2.2, then we can complete the connection between analytic function and a smooth function in  $\mathbb{C}$ :

$$\text{analytic function} \iff \text{holomorphic function}$$

### Theorem 2.5 (Stein 4.4)

If  $f$  is holomorphic function in an open set  $\Omega \subset \mathbb{C}$ , then  $f$  is an analytic function in  $\Omega$ .

**Proof.** By definition, we just need to verify that in any  $z_0 \in \Omega$ , there exists a voisinage such that the Talyor's sereies  $\sum_{n \in \mathbb{N}} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n$  converges to  $f$ .

Too see that we notice in Cauchy integration formula

$$\frac{1}{\zeta - z} = \frac{1}{\zeta - z_0} \frac{1}{1 - u}$$

by taking  $u = \frac{z - z_0}{\zeta - z_0}$ , then for any  $|u| < 1$ , we can rewrite formula as the form

$$f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z_0} \sum_{n \in \mathbb{N}} u^n d\zeta = \sum_{n \in \mathbb{N}} \frac{(z - z_0)^n}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta$$

Here interchange  $\int / \sum$  can be done since the holomorphic function is continous. and take equation 2 into above equality then we can get the result we hope. ■

**Remark** That means for any holomorphic function, we can do expansion casually.

## 2.2 Analytic continuation

Notice the corollary 2.2.2 we conclude in the previous section, it is an important properties, it claims that the local value of the power series determines the global value. In general, two analytic function  $f$  and  $g$ , if they are equal in a local voinsinage  $V$ , then for any  $x \in V$ , there exists a open ball  $B(x, r)$  such that their power series expanasions are equal. The most important things is that  $V$  is not compact, so our open cover  $B(x, r)$  will bring the properties by extending, so  $f$  and  $g$  are actually equal!

The statement above is a little abstract, it is better to draw a picture from a voinsinage  $V$ , then see that if we want to let  $f = g$  we just need to consider the point nearly the  $\partial V$ ...

In this section we will prove the analytic continuation, and three other important theorems in complex anaylsis, although they have various methods to porve, but here our inspiration is continuation:

Analytic continuation  $\implies$  Isolated zero  $\implies$  Open mapping  $\implies$  Maximum modulus

### Theorem 2.6 (analytic continuation)

Given an **open and connect** set  $U$  in complex plane, and  $f, g : U \rightarrow \mathbb{C}$  are two holomorphic (analytic) functions. If they are equal in a voinsinage of  $U$ , then  $f = g$  in whole  $U$ .

**Proof.** We let  $S = \{z \in U \mid f^{(n)}(z) = g^{(n)}(z) \text{ for any } n \in \mathbb{N}\}$ , and clearly  $S = \bigcup_{n \in \mathbb{N}} \{z \in U \mid f^{(n)}(z) = g^{(n)}(z)\}$ . It is easy to verify that  $S$  is the intersection of infinite closed sets, so  $S$  is closed.

$S$  is also open. We take any  $z_0 \in S$ , we notice that  $S$  is a subset of the voinsinage where  $f$  and  $g$  are equal, then by corollary 2.2.2, it implies that they have the same seires in a voinsinage of  $z_0$  contained in  $S$ , then clearly  $f^{(n)}(z) = a_n = b_n = g^{(n)}(z)$  in the voinsinage, so that  $S$  must be open. What's more  $S$  is not empty, and  $U$  is connect, so  $U = S$ . ■

**Remark** We notice that the proof is based on the topological property of the domain  $U$ , so the properties of holomorphic is speacial.

One directly result of analytic continuation is the isolated zero theorem, it claims that the zero of a holomorphic function must be discrete:

**Theorem 2.7 (isolated zero)** Suppose that  $U$  is an **open and connect set** in complex plane, and  $f \in H(U)$  is a non-zero holomorphic function, then

- If  $z \in U$  such that  $f(z_0) = 0$ , then  $f$  can be expanded in a neighborhood  $\Omega \subset U$ :

$$f(z) = (z - z_0)^n g(z)$$

where  $g$  is a non-vanishing holomorphic function in  $\Omega$ .

- Any zero  $z_0$  of  $f$ , there exists a neighborhood such that  $z_0$  is the unique zero in the neighborhood.

- The set of all zeros of  $f$  is discrete.

**Proof.** If  $z_0$  is a zero, then there exists a expansion of power series near the  $z_0$  with

$$f(z) = \sum_{k \in \mathbb{N}} a_k (z - z_0)^k, \quad a_0 = 0$$

So we just take  $n = \min\{i \in \mathbb{N} | a_i \neq 0\}$ , then

$$f(z) = (z - z_0)^n \sum_{k \in \mathbb{N}} a_{n+k} z^k$$

So we just take  $g(z) = \sum_{k \in \mathbb{N}} a_{n+k} z^k$ . It is well-defined since in a small reigon near  $z_0$  and does not contain  $z_0$  we have surely  $g(z) = \frac{f(z)}{(z-z_0)^n}$ , then by analytic continuation  $g$  is holomorphic on a samll neighborhood with  $z_0$  in it, and  $g(z_0) = a_n \neq 0$ . Hence by continuity, there exists a positive  $r > 0$  such that

$$||f(z_0)| - |a_n|| \leq |f(z_0) - a_n| < |a_n|/2$$

for any  $z \in B(z_0, r)$ , which implies  $|f(z_0)| > |a_n|/2 > 0$  in the reigon, so it is non-vanishing. Just take the intersection of the previous neighborhood and  $B(z_0, r)$ , then we get the  $\Omega$  we hope.

The second statement is immediate from the first statement. For the third, if there exists a limit point  $a$  in the set of zeros, then for any small neighborhood with  $a$  in it, it will conatins infinite other zeros, which contradicts with the second statement. ■

With the help of the isolated zero theorem, we can generalize the analytic continuation, locally two holomorphic function are just equal in a non-discrete set:

**Theorem 2.8 (strong analytic continuation)** *Given an **open and connect** set  $U$  in complex plane, and  $f, g : U \rightarrow \mathbb{C}$  are two holomorphic (analytic) functions. If they are equal in a set  $S \subset U$ , and  $U$  admet an accumulation point, then  $f = g$  in  $U$ .*

**Proof.** Just take  $f - g$  as the new holomorphic function in  $U$ , notice it will be zero in  $S$  and  $S$  is not discrete, so  $f - g$  must be zero by isolated zero. ■

**Remark** An direct result if two holomorphic functions are equal in a non-constant convergent sequence, then they will behave equal globally.

The other two result is the maximum modulus principle and open mapping theorem.

### Theorem 2.9 (open mapping)

A non-constant holomorphic function defined on **an open and connect set** must be an open map.

**Proof.** Take any  $z_0 \in U$ , and we will prove that  $f(z_0) \in f(U)$  must be an interior point. By isolated zero theorem, there exists  $B = B(z_0, r) \subset U$  such that in it we have

$$f(z) - f(z_0) = (z - z_0)^n g(z)$$

so  $f$  can be written as the composition of the following holomorphic functions:

- $u : z \mapsto (z - z_0)g^{1/n}(z)$  from  $B$  to  $\mathbb{C}$ .
- $v : z \mapsto z^n$  from  $\mathbb{C}$  to  $\mathbb{C}$ .
- $t : z \mapsto z + f(z_0)$  from  $\mathbb{C}$  to  $\mathbb{C}$ .

so  $f = t \circ v \circ u$ , we just need to verify the three maps are open, it can be proved via **the local inversion theorem: if the derivative is not zero at some point, then locally there exists a diffeomorphism, so the image of the point will be an interior.**

For  $u$  we can compute  $u'(z) = g^{1/n}(z) + \frac{(z-z_0)}{n}g'(z)g^{\frac{1-n}{n}}(z)$ , clearly  $u'(z_0) = g^{1/n}(z_0) \neq 0$ , and  $u$  is holomorphic so class of  $C^1$ , so locally there exists a diffeomorphism between neighborhoods  $U_{z_0}$  and  $V_{u(z_0)}$ , so  $u(z_0)$  is a interior point of  $f(B)$ .

For  $v$  the proof is similar if we take  $z_0 \neq 0$  since  $v'(z) = nz^{n-1}$ . In the case of  $z_0 = 0$ , we take a enough samll positive  $r > 0$ , then  $|z| < r$  implies  $|z|^n < r^n$ , so  $B(v(0), r) \subset B(0, r^n) = v(B(0, r))$ , so  $v$  is open. Finally,  $t$  is clearly open. Thus after composition,  $f$  is open. ■

### Theorem 2.10 (maximum modulus)

If  $f$  in a non-constant holomorphic function on **an open and connect set**  $U$ , then  $|f|$  can not attain a maximum in  $U$ .

-Weakly, if  $K \subset U$  is a compact set, then  $|f|_K$  attain the maximum in the boundary:

$$\sup_{z \in K} |f(z)| \leq \sup_{\partial K} |f(z)|$$

**Proof.** Assume that  $f$  did attain a maximum in  $z_0 \in U$ , then we take a small voinsinage of  $x_0$ , called  $V$ , by open mapping theorem  $f(V)$  must be open with  $f(z_0)$  in it, so there exists  $r > 0$  such that  $B(f(z_0), r) \subset f(D)$ . We take  $\delta = \text{sign}(\text{Re}(f(z_0)))$ , then  $f(z_0) + \delta \frac{r}{2} \in f(D)$ , we denote it  $f(w_0)$ , then  $|f(w_0)| > |f(z_0)|$  clearly, which contradicts with the assumption.

The weakly statement is immediate since  $|f|$  is continuous, so  $f(K)$  will be compact in  $\mathbb{R}$ , so  $|f|$  must attain a maximum in  $K$ , with the strong statement we know that the maximum can not be in  $\text{int}(K)$ , so in boundary. ■

## 2.3 Limit superior and limit inferior

A direct question from **Bolzano-Weierstrass theorem** is that how to determine the range of the value which the convergent subsequence converges to? Generally, the sequence can be classified into three types: convergent, divergent and oscillation.

We take a real sequence  $(a_n)_{n \in \mathbb{N}}$  and define a set

$$A_k = \{a_{k+p} | p \in \mathbb{N}\}$$

it means a set which contains the terms of sequence except  $a_0, \dots, a_{k-1}$ , then we define two new sequence by

$$v_n = \sup A_n, \quad w_n = \inf A_n$$

Notice that  $A_{k+1} \subset A_k$  for any  $k$ , so clearly we have

$$w_n \leq w_{n+1} \leq a_{n+1} \leq v_{n+1} \leq v_n$$

for any  $n$ . Here  $w_n$  is an increasing sequence and  $v_n$  is a decreasing sequence, **so highly likely they will be convergent**. If  $(a_n)_{n \in \mathbb{N}}$  is a convergent sequence, clearly as  $n \rightarrow \infty$ ,  $v_n$  and  $w_n$  will converge to the limit of  $a_n$ , but if  $(a_n)_{n \in \mathbb{N}}$  is oscillated, the convergence will change. See the example below:

**Example 2.2** We define a sequence as following

$$a_n = \begin{cases} \frac{1}{n} & n \equiv 0 \pmod{3} \\ \frac{1}{n} + 1 & n \equiv 1 \pmod{3} \\ \frac{1}{n} + 2 & n \equiv 2 \pmod{3} \end{cases}$$

we calculate the first 6 terms of  $v_n$  and  $w_n$

$$w_1 = w_2 = \dots = w_6 = 0$$

$$v_1 = v_2 = 3/2, v_3 = v_4 = v_5 = 1/5 + 2, v_6 = 1/7 + 2$$

so we can conclude with a little work that  $w_n$  converges to 0, and  $v_n$  converges to 2.

In above example, we easily guess that all possible values the subsequence will converge to are 0, 1, 2, so  $w_n$  and  $v_n$  seem to give the supremum and infimum respectively, hence we



formally define the limit superior and limit inferior of a sequence  $(a_n)_{n \in \mathbb{N}}$  by

$$\limsup a_n := \inf_{k \geq 0} (\sup_{n \geq k} A_n)$$

$$\liminf a_n := \sup_{k \geq 0} (\inf_{n \geq k} A_n)$$

we conclude the properties as folloing

**Proposition 2.11** *For a general real number  $\mathbb{R} \cup \{\pm\infty\}$ , we define  $(a_n)_{n \in \mathbb{N}}$  is a sequence of  $\mathbb{R}$ , then for any subsequence  $(a_{n_k})_{k \in \mathbb{N}}$  which converges to  $L$ , then*

$$\liminf a_n \leq L \leq \limsup a_n.$$

*In another words, all adherence points of the sequence will be in  $[\liminf a_n, \limsup a_n]$ .*

**Proof.** ■

Here we give an application in probability theory about sequence of the events.

**Theorem 2.12 (Borel-Cantelli Lemma)** *Suppose  $(A_i)_{i \in \mathbb{N}}$  is a sequence of events in a probability space  $(\Omega, \mathcal{T}, P)$ , then*

- (a) *If  $\sum_{n \in \mathbb{N}} P(A_i) < +\infty$ , then the probability that infinitely many of them occur is zero.*
- (b) *If  $\sum_{n \in \mathbb{N}} P(A_i) = +\infty$  and the events are independent, then the probability that infinitely many of them occur is 1.*

we need to translate the language into the measure theory. here "the probability that infinitely many of them occur" can be equivalent to the statement below: almost all  $w \in \Omega$  will be in infinitely many of the events  $\{A_i\}$ . we use the limit superior and inferior to formalize the statement

$$\limsup A_i = \bigcap_{N \in \mathbb{N}} \bigcup_{i \geq N} A_i$$

here we fistly prove that

$$\limsup A_i = \{w \in \Omega | w \text{ is in infinitely many of events}\}$$

for any  $w \in \limsup A_i$ ,  $w \in \bigcup_{i \geq 0} A_i$ , so there exists  $x_0 \in \mathbb{N}$  such that  $w \in A_{x_0}$ . Then we consider  $w \in \bigcup_{i \geq x_0+1} A_i$ , so there exists  $x_1 \geq x_0 + 1$  such that  $w \in A_{x_1}$ , ..., by recurrence we can conclude a strictly increasing map of integer  $n \mapsto x_n$ , then we can conclude that  $w \in \bigcap_{n \in \mathbb{N}} A_{x_n}$ .

For another hand, if  $w$  in in infinitely many of evnets, then we label all the events where it occurs, and it will be countable, so there exists a correspondence with  $\mathbb{N}$  and we denote it by  $f$ , then  $w \in \bigcap_{n \in \mathbb{N}} A_{f(n)} \subset \limsup A_i$ .

So we can conclude Borel-Cantelli Lemma by the measure theory as following:

- (a)  $\sum_{n \in \mathbb{N}} P(A_i) < +\infty \implies P(\limsup A_i) = 0.$
- (b)  $\sum_{n \in \mathbb{N}} P(A_i) = +\infty \wedge \text{independence} \implies P(\limsup A_i) = 1.$

So finally we finish the proof of theorem:

**Proof.** we notice that

$$w \mapsto \#\{i \in \mathbb{N} | w \in A_i\}$$

is a mesurable function by discrete measure, and then we notice that

$$\begin{aligned} \int_{\Omega} \#\{i \in \mathbb{N} | w \in A_i\} dP &= \int_{\Omega} \sum_{n \in \mathbb{N}} \mathbb{1}_{A_i} dP \\ &= \sum_{n \in \mathbb{N}} P(A_i) \end{aligned}$$

so if  $\sum_{n \in \mathbb{N}} P(A_i) < +\infty$ , then  $\#\{i \in \mathbb{N} | w \in A_i\}$  must be almost finite, i.e. almost all  $w \in \Omega$  occur in finite many of events, so  $\limsup A_i = 0$ .

If  $\sum_{n \in \mathbb{N}} P(A_i) = +\infty$ , by independence we have

$$P(\cup_{i=N}^M A_i^c) = \prod_{i=N}^M (1 - P(A_i)) \leq e^{-\sum_{i=N}^M P(A_i)}$$

as  $M \rightarrow \infty$  we can know that  $P(\cap_{i \geq N} A_i^c) = 0$  for any  $N$ , then by De Morgan's Law

$$P(\limsup A_i) = P((\cup_{N \in \mathbb{N}} \cap_{i \geq N} A_i^c)^c) = 1 - 0 = 1$$

■

## 2.4 Theorem of transfert: algebra of the random variable

We firstly talk about the proof of the theorem of transfert, and then apply it into The probability to get the general form of experience, and finally see some examples of the caculation of random variable via the theorem.

Notice that in this section  $\mathbb{R}^+ = [0, +\infty]$  for genenralisation.

**Theorem 2.13 (Transfert)** Suppose that  $(X, \mathcal{E}, \mu)$  is a measure space and  $(Y, \mathcal{T})$  is a measurable space. Given a measurbale function  $f : X \rightarrow Y$ , then

(1) The function invites a natural measure  $v : \mathcal{T} \rightarrow \mathbb{R}^+$  by

$$v(B) = \mu(f^{-1}(B)) = \mu \circ f^{-1}(B)$$

we call it the image measure of  $f$  such that  $(Y, \mathcal{T}, v)$  is a measure space.

(2) For any  $v$ -integrable (or **postive measurbale**) function  $h : Y \rightarrow \mathbb{C}$ , we have

$$\int_Y h dv = \int_X h \circ f d\mu$$

**Proof.** (1) Just verify. (2) By «technic» lemma, for any positive measurbale function  $h$  there exists  $B_n \in \mathcal{T}$  and  $c_n \in \mathbb{R}^+$  such that  $h = \sum_{n \in \mathbb{N}} c_n 1_{B_n}$ , so

$$\begin{aligned} \int_Y h dv &= \sum_{n \in \mathbb{N}} c_n \int_Y 1_{B_n} dv \\ &= \sum_{n \in \mathbb{N}} c_n v(B_n) \\ &= \sum_{n \in \mathbb{N}} c_n \mu(f^{-1}(B_n)) \\ &= \sum_{n \in \mathbb{N}} c_n \int_X 1_{f^{-1}(B_n)} d\mu \\ &= \sum_{n \in \mathbb{N}} c_n \int_X 1_{B_n} \circ f d\mu \\ &= \int_X \left( \sum_{n \in \mathbb{N}} c_n 1_{B_n} \right) \circ f d\mu = \int_X h \circ f d\mu \end{aligned}$$

Then some easy work to generalize any measurbale function... ■

With the basic theorem we can translate the theorem into the probability, we notice that the image measure is actually the probability distribution of a random variable, hence  $Y = \mathbb{R}$  and  $\mathcal{T} = \mathcal{B}(\mathbb{R})$ . And the intergation of a measure function under a probability measure is just the expectation.

Let us see the example we

**Theorem 2.14 (Transfert in probability)** *Given a probability space  $(\Omega, \mathcal{T}, P)$  and  $P_X$  is a probability distribution of a random variable  $X$ , then for any postive borel function  $g$ , we have*

$$E[g(X)] = \int_{\Omega} g(X) dP = \int_{\mathbb{R}} g dP_X$$

Furthermore if  $X$  is continous with density  $f_X : \mathbb{R} \rightarrow \mathbb{R}$ , then

$$E[g(X)] = \int_{\mathbb{R}} g \cdot f_X dl$$

where  $l$  is the lebesgue measure.

**Corollary 2.14.1** *For a continous random variable  $X$  with density  $f_X$ , the expactation*

$$E[X] = \int_{\mathbb{R}} t f_x(t) dt$$

**Remark** For an abstract probability space, it contains events which can not directly be described by number, and random variable gives a correspondence between the probability of the events and value in  $\mathbb{R}$ , that is important. So theorem of transfert just let us can operate the expectation of a random variable directly from an integration of a real-valued function, so it gives a correspondence in the forms of integration.

Here is some application of theorem of transfert:

**Proposition 2.15 (convolution)** Suppose that  $X, Y$  are two independent random variable, then the distribution of the random variable  $X + Y$  is the convolution of  $P_X$  and  $P_Y$ :

$$P_{X+Y}(A) = \int_{\mathbb{R}} P_X(A - y) dP_Y(y)$$

If  $f_X$  and  $f_Y$  are their density functions, then the density of  $X + Y$  is also the convolution of  $f_X$  and  $f_Y$ :

$$f_{X+Y}(t) = \int_{\mathbb{R}} f_X(t - y) f_Y(y) dy$$

**Proof.** For any  $A \in \mathcal{B}(\mathbb{R})$ , we have

$$\begin{aligned} (P_X * P_Y)(A) &= \int_{\mathbb{R}} P_X(A - y) dP_Y(y) \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} 1_{A-y}(x) dP_X(x) \right) dP_Y(y) \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} 1_A(x + y) dP_X(x) \right) dP_Y(y) \\ &= \int_{\mathbb{R}^2} 1_A(x + y) dP_{(X,Y)}(x, y) \end{aligned}$$

Notice that if we take  $g(x, y) = x + y$ , then clearly  $1_A(x + y) = 1_A \circ g(x, y)$ , so  $(P_X * P_Y)(A) = E[1_A(X + Y)] = P(X + Y \in A) = P_{X+Y}(A)$ . For the density, we take a borel function  $h$ , then

$$\begin{aligned} E[h(X + Y)] &= \int_{\mathbb{R}} h dP_{X+Y} \\ &= \int_{\mathbb{R}} h d(P_X * P_Y) \\ &= \int_{\mathbb{R}^2} h(x + y) f_X(x) f_Y(y) dx dy \\ &= \int_{\mathbb{R}^2} h(z) f_X(z - y) f_Y(y) dy dz \\ &= \int_{\mathbb{R}} h(z) \left( \int_{\mathbb{R}} f_X(z - y) f_Y(y) dy \right) dz \end{aligned}$$

Hence we can conclude our result by the theorem of transfert. ■

## 3 Complex analysis

### 3.1 Holomorphic Function

Complex-valued function is a function defined on  $\mathbb{C}$ , formally it is

$$f : \mathbb{C} \rightarrow \mathbb{C}, \quad z \mapsto f(z)$$

$\mathbb{C}$  can be seen same as  $\mathbb{R}^2$  in view of  $\mathbb{R}$ -vector space, so there exists a correspondence with two real variable functions.

For any  $z$ , there exists  $u, v \in \mathbb{R}$  such that  $f(z) = u + iv$ , then

$$u = \frac{1}{2}(f(z) + \overline{f(z)}), \quad v = \frac{1}{2i}(f(z) - \overline{f(z)})$$

so by taking all  $z$ , we can get two functions from  $\mathbb{C}$  to  $\mathbb{R}$ , and we just identify  $z \in \mathbb{R}$  with  $(x, y) \in \mathbb{R}^2$  by  $z = x + iy$ , then

$$f(z) = f(x + iy) = f(x, y) = u(z) + iv(z) = (u(x, y), v(x, y))$$

Here we should notice that  $\mathbb{R}^2$  equips the norm  $\|\cdot\|_2$  to make it compatible.

Notice that  $\mathbb{C}$  naturally has the norm structure, so it makes sense to define its differentiation directly from the topology:

**Definition 3.1** Let  $\Omega$  be an open set of  $\mathbb{C}$ , a function  $f : \Omega \rightarrow \mathbb{C}$  is called holomorphic at  $z_0 \in \Omega$  if the limit

$$\lim_{h \rightarrow 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

exists. We use  $H(\Omega)$  to denote the set of all holomorphic functions on  $\Omega$ .

**Example 3.1** An important example that a complex-valued function is not holomorphic is  $z \mapsto \bar{z}$ . We observe the function at  $z = 0$

$$\frac{\overline{z+h} - \bar{z}}{h} = \frac{\bar{h}}{h} = \frac{re^{-a}}{re^a} = e^{-2a}$$

so the limit does not exist. By if we rewrite the function in real variables by  $(x, y) \mapsto (x, -y)$ , then clearly the jacobian matrix is  $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ , which means the function is always differentiable in  $\mathbb{R}^2$ , so here difference appears and we can realize that the differentiation of complex-value function is a little different from.

Although the two types of differentiation is not the same thing, is there some connection between them? Let us firstly consider a set of differential operators:

$$\begin{cases} x = \frac{z+\bar{z}}{2} \\ y = \frac{z-\bar{z}}{2i} \end{cases} \implies \begin{cases} \frac{\partial}{\partial z} = \frac{\partial}{\partial x} \frac{\partial x}{\partial z} + \frac{\partial}{\partial y} \frac{\partial y}{\partial z} = \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) \\ \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial x} \frac{\partial x}{\partial \bar{z}} + \frac{\partial}{\partial y} \frac{\partial y}{\partial \bar{z}} = \frac{1}{2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) \end{cases}$$

Suppose  $f$  is a complex-value function holomorphic at 0 with  $f(0) = 0$ , then

$$f(z) = L(z) + o(z) = ax + by + o(z) = \frac{a-ib}{2}z + \frac{a+ib}{2}\bar{z} + o(z)$$

with  $L$  a  $\mathbb{C}$ -linear map with  $a, b \in \mathbb{C}$  and  $z = x + iy$ , then by above operators we know that

$$\frac{f(z)}{z} = \frac{\partial f}{\partial z}(0) + \frac{\partial f}{\partial \bar{z}}(0) \cdot \frac{\bar{z}}{z} + \frac{o(z)}{z}$$

if the limit exists as  $z$  tends to zero, then the term having  $\frac{\bar{z}}{z}$  must vanish, that means  $\frac{\partial f}{\partial \bar{z}} = 0$ , that is the necessary condition that  $f$  is holomorphic at  $z = 0$ .

Furthermore, if we merely view  $\mathbb{C}$  as  $\mathbb{R}^2$  to write function as the form of linear combination by two real-value functions  $f = u + iv$ , then we can conclude the Jacobien matrix

$$J_f(x, y) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

Then  $L(z)$  can be viewed as vector  $(x, y)$  operates on  $J_f(x, y)$ , and notice that  $L(z) = k \cdot z$  form some  $k \in \mathbb{C}$  and  $\mathbb{C}$  is isomorphic to a subspace of  $M_2(\mathbb{R})$  with the form  $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ , so immediately the jacobien matrix has the same form and we get a relation : **Cauchy-Riemann**

**Equation:**

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \end{cases}$$

Let us reflect the procedure above, the different things is that  $L$  must be a  $\mathbb{C}$ -linear map instead of  $\mathbb{R}$ -linear map, which is crucial.

**Lemma 3.1** Let  $V = \text{Lin}(\mathbb{R}^2, \mathbb{C})$  be the set of all linear maps from  $\mathbb{R}^2$  to  $\mathbb{C}$ , then

- (1)  $V$  is a  $\mathbb{R}$ -vector space with  $dx : (x, y) \mapsto x$  and  $dy : (x, y) \mapsto y$  as the base.
- (2)  $dz = dx + idy$  and  $d\bar{z} = dx - idy$  compose another base of  $V$
- (3)  $V$  is a  $\mathbb{C}$ -vector space if and only if for any  $L \in V$  under the base  $dz$  and  $d\bar{z}$ , the coefficient of  $d\bar{z}$  is zero.

**Proof.** (1) For any  $L \in V$ , we have  $L(x, y) = x \cdot L(1, 0) + y \cdot L(0, 1)$  notice that  $dx$  and  $dy$  actually are in  $V$ , and  $dx(x, y) = x$ ,  $dy(x, y) = y$ , then  $L = L(1, 0)dx + L(0, 1)dy$  with  $L(1, 0), L(0, 1) \in \mathbb{C}$ .

(2) we take  $c_1 dz + c_2 d\bar{z} = 0$ , then  $(c_1 + c_2)dx + i(c_1 - c_2)dy = 0$ , which implies  $c_1 = c_2 = 0$ , so  $dz$  and  $d\bar{z}$  are linearly independent.

(3) Take  $f = adz + bd\bar{z} \in V$  and  $u \in \mathbb{C}^2$

$$\begin{aligned} f(iu) - if(u) &= adz(iu) + bd\bar{z}(iu) - iad\bar{z}(u) - ibd\bar{z}(u) \\ &= -2bu_y \end{aligned}$$

The result must be zero since the map is linear over  $\mathbb{C}$ , then  $b = 0$ . ■

Let us consider differential

$$df : U \rightarrow \text{Lin}(\mathbb{R}^2, \mathbb{C})$$

notice that in differential  $L(1, 0) = Df(a, b)(1, 0) = f_x(a, b)$ , so we have that

$$df = f_x dx + f_y dy$$

Now if we expect to write  $df$  as the combination of  $dz$  and  $d\bar{z}$ , so

$$df = \mu dz + \lambda d\bar{z} = (\mu + \lambda)dx + i(\mu - \lambda)dy$$

with calculation we have that

$$\mu = \frac{f_x - if_y}{2}, \quad \lambda = \frac{f_x + if_y}{2}$$

Which implies the differential operators like above.

So we can conclude the fundamental theorem of holomorphic function.

**Theorem 3.2** Suppose that  $f = u + iv$  is a complex function with  $u, v$  two functions of class  $C^1(\mathbb{R}^2, \mathbb{R})$ , then following statements are equivalent.

(1)  $f$  is holomorphic on  $\Omega$ .

(2)  $\frac{\partial f}{\partial \bar{z}} = 0$  on  $\Omega$ .

(3) Cauchy-Riemann Equation is satisfied.

**Proof.** (2)  $f$  is holomorphic, then for any  $z_0 \in \Omega$ ,  $df(z_0)$  must be a  $\mathbb{C}$ -linear map in  $\text{Lin}(\mathbb{R}^2, \mathbb{C})$ , so the coefficient of  $d\bar{z}$  must be zero by above lemma, and notice that the coefficient is just  $\frac{\partial f}{\partial \bar{z}}(z_0)$ .

(3) Just calculate

$$\begin{aligned} 0 &= \frac{\partial f}{\partial \bar{z}} = \frac{1}{2} \left( \frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} \right) \\ &= \frac{1}{2} \left( \frac{\partial u + iv}{\partial x} + i \frac{\partial u + iv}{\partial y} \right) \\ &= \frac{1}{2} \left[ \left( \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right) + i \left( \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \right] \end{aligned}$$
■

**Corollary 3.2.1** *For a holomorphic function  $f = u + iv$ , we have some details:*

*(a) If we see  $f$  as a real-value function  $F(x, y)$ , then for any defined  $z = x + iy$*

$$\det J_F(x, y) = |f'(z)|$$

$$(b) f'(z) = u_x + iv_x = v_y - iu_y$$

**Proof.** The result can be calculated by using C-R equation. ■



## 4 Functional Analysis

### 4.1 Hilbert Space

It is boring to talk about the finite-dimensional normed space, that is because each such space can be seen the same with Euclidean space  $\mathbb{R}^n$  with a certain norm (both in terms of algebraic and topological structures), and functional space is usually a infinite-dimensional space, so our linear algebra will be not enough (for someone only know matrix), that will be the main content in Hilbert space: generalize the topic about inner product in linear algebra, refine the result with some idea of topology.

Here are some two simple inspiration for the next content. The first is a simple result: **Any finite-dimension subspace of a normed vector space is closed.** The statement will be not correct if the dimension is infinite, hence we can speculate that some good properties we hope will still hold if the subspace is closed. The second is about duality and dual space, that is a beautiful and strong structure. In finite-dimensional case, a linear map can be seen as a matrix  $M$  and we often talk about its transpose  $M^T$ , it appears frequently in a number of topics involving orthogonal groups, but when the dimension is infinite, it is not easy to see them as a matrix and not easy to consider its transpose, so we need a equivalent concept: adjoint operator, it refers to inner product.

Moreover, the good properties about orthogonality in finite-dimensional vector space is as following: if  $\{e_1, \dots, e_k\}$  is a orthonormal basis for a vector space  $E$  with a inner product, then for any  $x \in E$  we have a decomposition:

$$x = \langle e_1, x \rangle e_1 + \dots + \langle e_k, x \rangle e_k$$

and its norm will be

$$\|x\|^2 = |\langle e_1, x \rangle|^2 + \dots + |\langle e_k, x \rangle|^2$$

That will be very useful for analysis. For example, in Fourier analysis we usually consider a orthonormal basis  $\{e^{inx}\}_{i \in \mathbb{Z}}$ , but in the time of Fourier mathematicien just know how to expand function as a fourier series:

$$f \approx \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{inx}$$

it is difficult to study the convergence of the series. That is a very classic question, new tool and technique come from algebra gives a very bright methodology. Here i really suggest reader to read some materials: The chapter 6 of [1] it covers complete definition of inner product in finite-dimension space, some technique will be the same in Hilbert space; The chapter 3, section F of [1], it gives a rough introduction about dual space in the finite-dimension, it is a good

motivation and beginning for someone who wants to know duality; First 3 chapters of [2], it gives a clear motivation about Fourier analysis, and show the difficulties about the topic without advanced tool.

### 4.1.1 prerequisite about inner product

**Definition 4.1** An inner product on a  $\mathbb{K}$ -vector space  $V$  is a function:

$$\begin{aligned}\langle \cdot, \cdot \rangle : V \times V &\rightarrow \mathbb{K} \\ (u, v) &\mapsto \langle u, v \rangle\end{aligned}$$

which satisfies following properties:

**Positivity:**  $\langle v, v \rangle \geq 0$  for all  $v \in V$ .

**Definiteness:**  $\langle v, v \rangle = 0$  if and only if  $v = 0$ .

**Additivity in first slot:**  $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$  for all  $u, v, w \in V$ .

**Homogeneity in first slot:**  $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$  for all  $\lambda \in \mathbb{F}$  and all  $u, v \in V$ .

**Conjugate symmetry:**  $\langle u, v \rangle = \overline{\langle v, u \rangle}$  for all  $u, v \in V$ .

**Remark** When  $\mathbb{K} = \mathbb{R}$ , the inner product is actually a symmetric bilinear form which is positive definite; When  $\mathbb{K} = \mathbb{C}$ , the inner product is actually a Hermitien form which is positive definite.

Inner product has a closed relation with norm, each inner product can define a standard norm as following:

$$\|x\| = \sqrt{\langle x, x \rangle}$$

It is easy to verify that it is a norm (a nice exercise). But for the uniqueness, we will state below:

**Proposition 4.1** Inner product defines a norm uniquely up to homeomorphism. More precisely, if  $\langle \cdot, \cdot \rangle$  is a inner product of a vector space  $V$ , then there exists a natural map  $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that  $f(\langle u, u \rangle)$  defines a norm on  $V$ .

**Proof.** Firstly, for any  $x = \langle u, u \rangle$ , and any  $a \in \mathbb{K}$ , we have  $f(\langle au, au \rangle) = f(|a|^2 x) = |a|f(x)$ . Notice that  $f$  must be continuous since  $\langle u, u \rangle$  defines a norm, and each norm can be seen as a continuous map. Hence we just need to find the possible continuous function  $f$  such that  $f(a^2 x) = af(x)$  for any  $a, x > 0$ , it is not difficult to get that the unique possible is the map with the form  $x \mapsto C\sqrt{x}$ . Notice that  $\|x\| = \sqrt{\langle x, x \rangle}$  is a norm as we state above, so for any other  $\|x\|_c = c\sqrt{\langle x, x \rangle}$  with positive number  $c$ , it is still a norm with simple verification. and notice that

$1/M\|x\|_c \leq \|x\| \leq M\|x\|_c$  for  $M \geq \max(c, 1/c)$ , so the topology they define are equivalent. ■

According to the standard norm above, we can conclude several properties about inner product:

**Proposition 4.2** Suppose that  $\langle \cdot, \cdot \rangle$  is a product of a vector space  $V$ , then

(a) **Cauchy-Schwartz inequality:**  $\langle x, y \rangle \leq \|x\| \cdot \|y\|$ ,  $\forall x, y \in V$

The equality holds if and only if  $x = \lambda y$  with  $\lambda \in \mathbb{K} - \{0\}$ .

(b) **Pythagore's formula:** for any  $x, y \in V$ ,

$$\|x + y\|^2 = \|x\|^2 + 2 \cdot \operatorname{Re}(\langle x, y \rangle) + \|y\|^2$$

(c) **Parallelogram Identity:** for any  $x, y \in V$ ,

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

**Proof.** good exercise. ■

Parallelogram identity needs an attention here, it is important to determine whether a norm can invite an inner product, the proof is a little complicated.

**Theorem 4.3** A normed vector space can have inner product structure if and only if the norm satisfies **Parallelogram identity**.

more precisely, for an normed vector space  $(V, \|\cdot\|)$  satisfying P.I. we can define inner product as following:

-If  $V$  is over  $\mathbb{R}$ :

$$\langle x, y \rangle := \frac{\|x + y\|^2 - \|x - y\|^2}{4}$$

-If  $V$  is over  $\mathbb{C}$ :

$$\langle x, y \rangle := \frac{\|x + y\|^2 - \|x - y\|^2}{4} - i \cdot \frac{\|x + iy\|^2 - \|x - iy\|^2}{4}$$

**Proof.** We just need to prove the necessity. It refers to some algebraic calculation so omit some parts. If  $V$  is over  $\mathbb{R}$ , we prove the homogeneity. We fix  $u \in V$  such that  $\phi_u : x \mapsto \langle x, u \rangle$ . then we need to prove  $\phi_u(kx) = k\phi_u(x)$  for any  $k \in \mathbb{R}$ , it is easy to observe that  $\phi_u$  is a continuous map since  $\|\cdot\|$  is continuous. Moreover, we can prove the additivity:  $\phi_u(x + y) = \phi_u(x) + \phi_u(y)$ , then by the two properties (continuous and additivity),  $\phi_u$  is homogenous automatically.

If  $V$  is over  $\mathbb{C}$ , similarly... ■

Orthogonality is a core concept in inner product, it involves the algebraic treatment of the geometric concept of perpendicularity. here are some vocabulary:

**Definition 4.2**  $\langle \cdot, \cdot \rangle$  is the inner product of a vector space.

**-orthogonal:** we call two vector  $x, y \in V$  is orthogonal if

$$\langle x, y \rangle = 0$$

or equiavlently, it satisfying  $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ .

**-orthonormal list:** a set of vector  $\{e_i\}_{i \in I}$  is a orthonormal list if for any  $m, n \in I$

$$\langle e_m, e_n \rangle = \delta_{m,n}$$

In particular if the list spans the vector space, we call it **orthonormal base**.

**-orthogonal complement:** for a non-empty subset  $U \subset V$ , the orthogonal complement of  $U$  is defined by a vector space(**needs a verification!**) as following:

$$U^\perp := \{u \in V \mid \langle u, v \rangle = 0 \quad \forall v \in U\}$$

There exists a classic method to generate a orthonormal list according to a given linearly independent list or a base, it is called **Gram-Schmidt procedure**.

**Proposition 4.4**

Finally we discuss the orthogonal projection, it refers to the decomposition of the space, which will be fundamental if we have base function.

**Theorem 4.5 (Existence and Uniqueness of projection)**

Suppose that  $F$  is a closed and convex subset of a vector space  $V$  with a inner product  $\langle \cdot, \cdot \rangle$ , that for any  $x \in V$  there exists a unique vector  $x_F$  such that

$$\|x - x_F\| = d(x, F) := \inf_{y \in F} \|x - y\|$$

**Proof.** equivalently, we just need to prove the set  $S = \{\|x - y\| : y \in F\}$  is closed, then clearly a such vector  $x_F$  will exist by **Supremum Property** of the real number.

such vector is unique, we just suppose that there exists another vector  $x' \in F$  such that  $\|x - x'\| = d(x, F)$ , then by parallelogram equality, we will have

$$\|2x - x_F - x'\|^2 + \|x' - x_F\|^2 = 2\|x - x_F\|^2 + 2\|x - x'\|^2 = 4d^2(x, F)$$

then rewrite it we have

$$\|x' - x_F\| = 4d^2(x, F) - 4\|x - \frac{x_F + x'}{2}\|^2 \leq 0$$

So immediately  $x_F = x'$  ■

Notice that the uniqueness depends on the inner product, for the general normed space we can't conclude the uniqueness although the existence still holds. With the theorem it makes sense to give orthogonal projection a formal definition (**Notice that a subspace is indeed convex**):

**Definition 4.3** Give a closed subspace  $F \subset V$ , the (orthogonal) projection of some  $x \in V$  on  $F$  is a vector  $P_F(x)$  satisfies

$$\|x - P_F(x)\| = \min_{y \in F} \|x - y\|$$

or equivalently (**needs a proof**)

$$\langle x - P_F(x), y \rangle = 0 \quad \forall y \in F$$

According to these properties we can conclude the fundamental decomposition according to the orthogonality:

**Theorem 4.6**  $F$  is a closed subspace of vector space  $V$ , then

$$F \oplus F^\perp = V$$

**Proof.** It is immediately if we write any  $x \in V$  as the form:

$$P_F(x) + (x - P_F(x)) = x$$

■

So it makes sense to consider projection map

$$P_F : V \rightarrow F, x \mapsto P_F(x)$$

Here are its properties:

**Proposition 4.7**  $F$  is a closed subspace of vector space  $V$ .

- (a)  $P_F$  is a linear map in  $\mathcal{L}(V)$ .
- (b)  $P_F$  is continuous.
- (c)  $P_F$  is surjective with  $P_F(x) = x$  for any  $x \in F$ .
- (f)  $\text{Im} P_F \oplus \ker P_F = V$
- (d)  $P_F \circ P_F = P_F$
- (e)  $P_F$  is a self-adjoint operator, i.e. for any  $x, y \in V$

$$\langle P_F(x), y \rangle = \langle x, P_F(y) \rangle$$

notice that if  $\dim V < +\infty$  and  $M = [P_F]_{\mathcal{B}}$ , the condition is same with  $\overline{M^T} = M$ .

With the basic knowledge we can now formally define the Hilbert space:

**Definition 4.4**  $(H, \langle \cdot, \cdot \rangle)$  is a **pre-Hilbert space** if  $H$  is an inner product space over  $\mathbb{R}$  or  $\mathbb{C}$ ; if the pre-Hilbert space is complete under the standard norm  $\|x\| = \sqrt{\langle x, x \rangle}$ , then it is a **Hilbert space**.

**Example 4.1** Consider the case when  $H = \mathbb{C}$ , complex numbers form a complete space by the modulus  $\|z\|^2 = z\bar{z}$ , so we can define in it  $\langle x, y \rangle = x\bar{y}$ . And we can generalize it for  $H = \mathbb{C}^n$ , it gives **general Euclidean space** with the inner product:

$$\langle x, y \rangle = x_1\bar{y}_1 + \dots + x_n\bar{y}_n$$

It is the classic case of Hilbert space.

**Example 4.2**  $L^p$  space is a Hilbert space only if  $p = 2$ , since its norm must satisfy the parallelogram identity, so we take two disjoint and finite measurable sets  $A$  and  $B$ , then consider  $f = 1_A$  and  $g = 1_B$ , then by identity:

$$|\mu(A)|^{2/p} + |\mu(B)|^{2/p} = |\mu(A) + \mu(B)|^{2/p}$$

If  $p$  is finite, the equality holds only if  $p = 2$ . For  $p = +\infty$ , same identity we have the contradiction:

$$2\|f\|_\infty^2 + 2\|g\|_\infty^2 = 4 \neq 2 = \|f + g\|_\infty^2 + \|f - g\|_\infty^2$$

In the case  $p = 2$  we can define inner product on  $L^2(X, \mathcal{E}, \mu)$  space by

$$\langle f, g \rangle = \int_X f\bar{g}d\mu$$

#### 4.1.2 Hilbert base

## 5 probaility

### 5.1 Independence

**Definition 5.1** Given a probability space  $(\Omega, \mathcal{T}, P)$

- events  $A_1, A_2, \dots \in \mathcal{T}$  are called independent if for any  $i \neq j$ ,

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

- random variables  $X_1, X_2, \dots$  are called independent if for any  $i \neq j$ ,

$$P(X_i \in A; X_j \in B) = P(X_i \in A)P(X_j \in B) \quad \forall A, B \in \mathcal{B}(\mathbb{R})$$

**Remark** Independence is a simple concept in probability but not in mathematics. Notice that in measure product of Borel set, we have  $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ , and equivalently we can rewrite the definition of the independence of two random variable by

$$P((X, Y) \in A \times B) = P(X \in A)P(Y \in B)$$

that means it invites a new probability measure with the properties

$$P_{(X,Y)}(A \times B) = P_X(A)P_Y(B)$$

where  $(X, Y)$  is a natural random variable defined on  $\Omega^2$  by  $(a, b) \mapsto (X(a), X(b))$ . Clearly we can deduced that independence is the trival condition that measure product is well-defined in probability space.

**Theorem 5.1** Two random variable  $X$  and  $Y$  are independent if and only if  $P_{(X,Y)} = P_X \otimes P_Y$ , in another word, they define the same measure in space  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ .

**Proof.** Firstly,  $P_X$  and  $P_Y$  are two natural probability measure on  $\mathbb{R}$ , so from the construction of the measure product,  $P_X \otimes P_Y$  exists. So it is a measure on space  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ . We put  $S = \{A \times B | A, B \in \mathcal{B}(\mathbb{R})\}$ . it is the pi-system on  $\mathbb{R}^2$ . We notice that on  $S$ ,  $P_{(X,Y)} = P_X \otimes P_Y$  if and only if the two random variables are independent, so by the **theorem of unique exitension of the measure** we can know that they are actually same in  $\mathbb{R}^2$ . ■

Several result can be concluded from the theorem, it refers to some probability value with respect two independence variable.

**Corollary 5.1.1** Suppose that  $X, Y$  are two independent random variable in  $(\Omega, \mathcal{T}, P)$ , then

(1) For any two borel functions  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  such that  $f(X), g(Y)$  are  $P$ -integrable,

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

- (2) If  $X$  and  $Y$  are continuous with density  $f_X$  and  $f_Y$ , then random variable  $(X, Y)$  admits a density  $f_X \cdot f_Y$ .
- (3)  $Cov(X, Y) = 0$
- (4)  $Var(X + Y) = Var(X) + Var(Y)$

**Proof.** (1) Using Fubini's theorem

$$\begin{aligned} E[f(X)g(Y)] &= \int_{\mathbb{R}^2} f(a)g(b)dP_X(a)dP_Y(b) \\ &= \left( \int_{\mathbb{R}} f dP_X \right) \cdot \left( \int_{\mathbb{R}} g dP_Y \right) \\ &= E[f(X)]E[g(Y)] \end{aligned}$$

(2) For any  $A, B \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} P_{(X,Y)}(A \times B) &= P_X(A)P_Y(B) \\ &= \left( \int_{\mathbb{R}} f_X(t) \cdot 1_A(t) dt \right) \cdot \left( \int_{\mathbb{R}} f_Y(u) \cdot 1_B(u) du \right) \\ &= \int_{\mathbb{R}^2} 1_{A \times B}(t, u) (f_X(t)f_Y(u)) dl(t, u) \end{aligned}$$

so clearly  $(t, u) \mapsto f_X(t)f_Y(u)$  is the density of  $(X, Y)$  since the class  $A \times B$  generates  $\mathcal{B}(\mathbb{R}^2)$ .

(3) and (4) are immediate after calculation. ■

## 5.2 Convergence

The convergence of the sequence of random variable is related to functional sequence, we can classify different types of convergence, they are very delicate.

**Definition 5.2** Given a sequence of random variable  $(Y_n)_{n \in \mathbb{N}}$  of  $(\Omega, \mathcal{T}, P)$ , and  $p \geq 1$ :

- $(Y_n)_{n \in \mathbb{N}}$  converges to  $Y$  **by norme**  $L^p$  if

$$\|Y_n - Y\|_p \xrightarrow{n \rightarrow \infty} 0$$

- $(Y_n)_{n \in \mathbb{N}}$  converges to  $Y$  **by measure** (probability) if

$$\forall c > 0, \quad P(|Y_n - Y| > c) \xrightarrow{n \rightarrow \infty} 0$$

- $(Y_n)_{n \in \mathbb{N}}$  converges to  $Y$  **almost surely** if

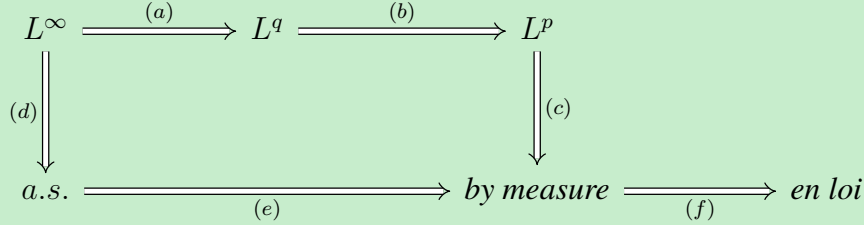
$$\text{for almost every } w \in \Omega, \quad Y_n(w) \xrightarrow{n \rightarrow \infty} Y(w)$$



- $(Y_n)_{n \in \mathbb{N}}$  converges to  $Y$  **en loi** if for any continuous bounded function  $f : \mathbb{R} \rightarrow \mathbb{R}$ :

$$E[f(Y_n)] \xrightarrow{n \rightarrow \infty} E[f(Y)]$$

**Proposition 5.2 (Implication of the convergence)**



where  $1 \leq p \leq q \leq \infty$

**Proof.** (a) ■

### 5.3 Law of large number

**Definition 5.3** A sequence of random variable  $(X_n)_{n \in \mathbb{N}}$  is called *identically distributed* if they share the same distribution, that means for any  $i, j \in \mathbb{N}$ :

$$P_{X_i} = P_{X_j}$$

Usually we use "id." to denote the random variable.

There are a criteria to determine if the random variables are identically distributed after changing variable:

**Proposition 5.3** If  $X_1$  and  $X_2$  are two independent and identically distributed random variable, and  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a positive borel function(or integrable), then  $f(X_1)$  and  $f(X_2)$  are still independent and identically distributed.

**Proof.** **The condition for independence can be that  $f$  is just a borel function.** We just need to verify that for any Borel set  $B$ , there exists a borel set  $A$  such that

$$(f \circ X)^{-1}(B) = X^{-1} \circ f^{-1}(B) = X^{-1}(A)$$

then it is clear. For id. we notic that by the theorem of transfert

$$E[f(X_i)] = \int_{\mathbb{R}} f dP_{X_i}$$

idectical distribution implies  $E[f(X_1)] = E[f(X_2)]$ , then we take  $f = 1_A$  with  $A$  a borel se, immediate we get that  $P_{X_1}(A) = P_{X_2}(A)$  for any borel set  $A$ , so they are equal. ■

We have same result for the independence of two random variable, Now we can conclude the important theorem in probability theory:

**Theorem 5.4 (Strong Law)**

Suppose that  $(X_i)_{i \in \mathbb{N}}$  is a sequence of independent, identically distributed and integrable random variables, then for almost every  $\omega \in \Omega$

$$\frac{S_n(\omega)}{n} = \frac{(X_1 + \dots + X_n)(\omega)}{n} \xrightarrow[n \rightarrow \infty]{} E[X_0]$$

Or we can say the sample mean  $\frac{S_n}{n}$  converges to zero almost surely.

**Corollary 5.4.1** For any borel set  $A$ , we have a almost surely convergence

$$\frac{\#\{i \leq n | X_i(\omega) \in A\}}{n} \xrightarrow[n \rightarrow \infty]{} P(X_0 \in A)$$

**Proof.** Notice that  $\#\{i \leq n | X_i(\omega) \in A\} = \sum_{i=0}^n 1_{\{X_i \in A\}}$ , so we take  $Y_k = 1_{X_k \in A}$  as the new sequence of random variable, then using the theorem we get that the it will converge to  $E[Y_0] = P(X_0 \in A)$ . ■

**Theorem 5.5 (Weak Law)**

Suppose that  $(X_i)_{i \in \mathbb{N}}$  is a sequence of independent, identically distributed and square integrable random variables, then for any  $\epsilon > 0$ ,

$$P(|\frac{S_n}{n} - E(X_0)| > \epsilon) \xrightarrow[n \rightarrow \infty]{} 0$$

Or we can say the sample mean  $\frac{S_n}{n}$  converges to  $E[X_0]$  by probability.

**Proof.** The proof is immediate via the strong law by convergence. square integrable implies integrable by Cauchy-Schwartz inequalit, then the sequence converges almost surely to  $E[X_0]$ , clearly convergent by probability. ■

Here we give a theortical application of law of large number. If  $X_1, X_2, X_3$  are independent, then

## References

- [1] Sheldon Axler. *Linear algebra done right*. Springer Nature, 2024.
- [2] Elias M Stein and Rami Shakarchi. *Fourier analysis: an introduction*. Vol. 1. Princeton University Press, 2011.