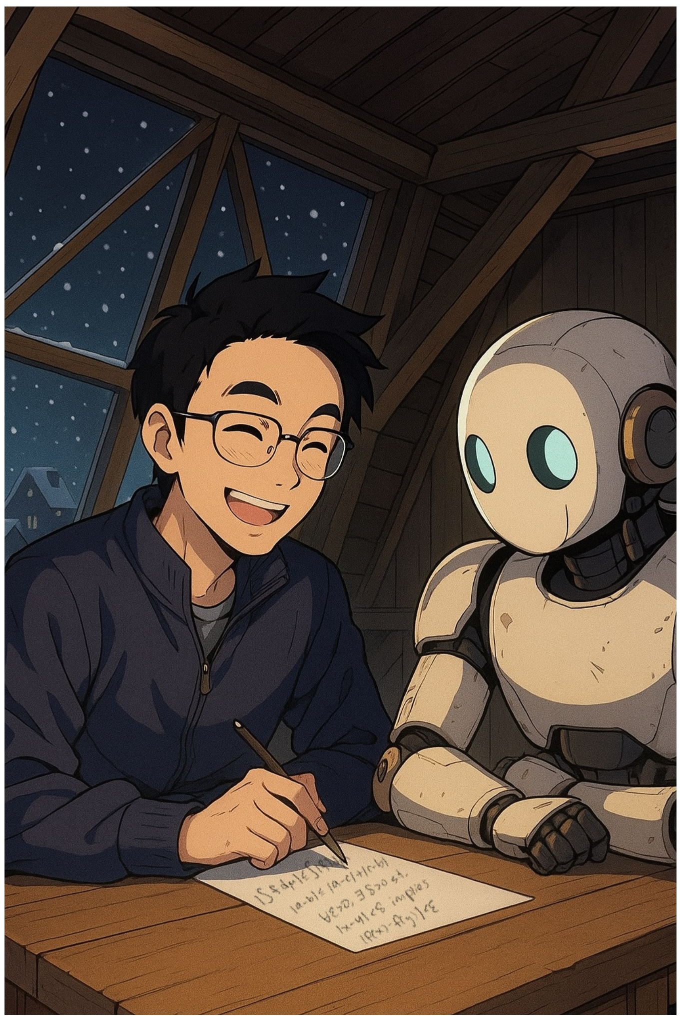
NOTES ON MATHEMATICAL ANALYSIS

WRITTEN BY

TIANYI WANG



PUBLISHED

IN THE WILD

Contents

Preface

This note is based on Professor Francis Su's lectures at Harvey Mudd College and Prof. Qiguang Bai's lectures at NTCU. The text is *Principles of Mathematical Analysis* by Walter Rudin. The note will not be a boring repetition of the text, but more about supplementary information that professor mentions in the lecture. This note, to some degree, reflects the process of my self learning, in which I recall important results we get from both lecture and text, and I use verbal explanation to demonstrate my own understanding. I also added a couple of diagrams that I believe is helpful for you to understand important ideas in the proofs. Now that I have time to enjoy this amazing subject, not only do I invite you to join this exciting journey, but I also wish you the very best for taking your own steps towards the greatest intellectual adventure of human history.

Tianyi
January 2nd, 2020

PS: This note is officially finished now [1].

Tianyi
August 1st, 2021

[1] I was awake for the whole night!

Chapter 1

The Real and Complex Number Systems

God made the integers, all else is the work of man.

-Leopold Kronecker

1.1 Boring Stuff

We will first start by introducting sets.

Definition 1.1.1. If A is any set, we write $x \in A$ to indicate that x is a member. If x is not a member, we write $x \notin A$.

Definition 1.1.2. The set which contains no element is called the empty set. If a set contains at least one element, then the set is nonempty.

Definition 1.1.3. If A and B are two sets, and every element of A is an element of B, we say A is a subset of B. We write $A \subset B$. $A \subset A$ is true for every set A.

Definition 1.1.4. If $A \subset B$ and there is an element in B that is not in A, then A is a proper set of B. If $A \subset B$ and $B \subset A$, we write $A = B$.

Definition 1.1.5. If we have two sets A and B, their union is defined by

$$A \cup B = \{x : x \in A \quad \text{or} \quad x \in B\}$$

Their intersection is defined by

$$A \cap B = \{x : x \in A \quad and \quad x \in B\}$$

We first define the product of two sets.

Definition 1.1.6. The product of two sets A and B is defined as

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

Definition 1.1.7. A relation R is a subset of $A \times B$. If $(a, b) \in R$, write aRb.

Here is an example of the definition above. Define P as the set of all people. Then A "is an ancestor of" is a relation on $P \times P$. I may ask, is the pair (Tom, Jerry) in the relation A? In other words, is Tom an ancestory of Jerry? The answer may depends. But you get the idea. Now we move to equivalent relations.

Definition 1.1.8. An equivalence relation R on a set S is a relation on $S \times S$ such that the following statement holds true
(i) aRa
(ii) $aRb \Rightarrow bRa$
(iii) aRb and $bRc \Rightarrow aRc$.

The first law is called reflexive condition, the second one is called symmetry, and the third one called transitivity. We denote equivalence relation with "$\sim$".

Now we will try to convince ourselves that a function is actually a relation.

Proposition 1.1.1. *A function F from A to B is a relation such that if*

$$aFb \quad and \quad aFb' \Rightarrow b = b'$$

A more conventional way to state this proposition is the following: Consider a function $F : A \to B$, if $F(a) = b$ and $F(a) = b'$, then $b = b'$. Good.

Now we talk about the construction of rational numbers.

Definition 1.1.9. We define $\mathbb{Q}$ to be a relation on $\mathbb{Z}\times(\mathbb{Z}-0)$, where $\mathbb{Z}-0$ is the set of all integers without the element 0. More specifically, $\mathbb{Q}$ is the collection of equivalent classes

$$\mathbb{Q} = \left\{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\right\}$$

where $\frac{p}{q}$ is the class of ordered pairs (p, q) such that $(p, q) \sim (m, n)$ if $pn = qm$ and $q, n \neq 0$.

We now look more carefully at Definition 1.1.9. We constructed the relation "$\sim$", but now we need to prove that it is equivalent. Why? Say if you have two ordered pairs, (1,2) and (2,4), the corresponding $\frac{m}{n}$ should be equivalent: they are both $\frac{m}{n}$. Therefore, we agree that we need to check whether the relation "$\sim$" is an equivalence relation.

Now, according to the definition of equivalence relation, we need to check three things. First, check if $(p, q) \sim (p, q)$. According to our definition, since $pq = qp$, and $q \neq 0$, this is true. Second, check if $(p, q) \sim (m, n)$ implies $(m, n) \sim (p, q)$. Since $pn = qm$ implies $mq = pn$, this is also true. Finally, a more interesting one. Does $(p, q) \sim (m, n)$ and $(m, n) \sim (a, b)$ implies $(p, q) \sim (a, b)$? To check that we use a law called cancellation on $\mathbb{Z}$.

Corollary 1.1.1. ***Cancellation on $\mathbb{Z}$.***
Consider $a, b, c \in \mathbb{Z}$. If $ab = ac$ and $a \neq 0$, then $b = c$.

Continue our proof. Now we already know that $b, q \neq 0$. We multiply b to both sides of equation $pn = qm$, giving $pnb = qmb$. We multiply q to both sides of $mb = an$, giving $qmb = anq$. Therefore, $pnb = anb$. Using the cancellation law on $\mathbb{Z}$, we have $pb = qa$. Q.E.D.

1.2 Properties of $\mathbb{Q}$

Now that we defined $\mathbb{Q}$, we start to define operations on $\mathbb{Q}$ such as addition and multiplication.

First of all, what is a good definition of addition on $\mathbb{Q}$? Let's say we have $\frac{a}{b}$ and $\frac{c}{d}$. Can we define

$$\frac{a}{b} + \frac{c}{d} = \frac{a+c}{b+d} \quad ?$$

Well, you can't stop me from doing that, but there is a problem. Let's say $\frac{a}{b} = \frac{1}{2}$, and $\frac{c}{d} = \frac{3}{4}$. Then we have

$$\frac{1}{2} + \frac{3}{4} = \frac{4}{6}.$$

But what if we pick $\frac{2}{4}$, which is the same class of $\frac{1}{2}$, to add $\frac{3}{4}$? Then we get

$$\frac{2}{4} + \frac{3}{4} = \frac{5}{4}.$$

The problem is, I shouldn't be getting diffrent class like $\frac{4}{6}$ and $\frac{5}{4}$ on the right hand side by adding the same class on the left hand side! The addition, therefore, is *not well defined*. A *well defined* definition should be independent of the representative I choose. What about the following definition?

$$\frac{a}{b} + \frac{c}{d} = \frac{0}{1} \quad ?$$

The result on the right hand side does not depend on the representative I choose. Therefore, it *is* well defined. However, this definition is boring because it doesn't depend on anything! We should then look for more useful definition instead of this because it takes us nowhere.

Now the good definition:

Definition 1.2.1. Addition on $\mathbb{Q}$
The addition on $\mathbb{Q}$ is defined to be

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

where $bd \neq 0$.

But even with the good definition, we are not done yet. We need to prove that it is well defined. So it does not depend on the representative I choose. Mathematically speaking, if $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$, then $(ad+bc, bd) \sim (a'd'+b'c', b'd')$. Let's prove this.

Proof. We know that $ab' = a'b$ and $cd' = c'd$. Then

$$\begin{aligned}(ad+bc)b'd' &= adb'd' + bcb'd' \\ &= a'bdd' + c'db'b \\ &= bd(a'd' + b'c').\end{aligned}$$

This means that indeed $(ad+bc, bd) \sim (a'd'+b'c', b'd')$. The proof is completed. □

Definition 1.2.2. multiplication on $\mathbb{Q}$
The multiplication on $\mathbb{Q}$ is defined to be

$$\frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$

where $bd \neq 0$.

We quickly prove that this definition is well defined.

Proof. We want to prove that if $(a,b) \sim (a',b')$ and $(c,d) \sim (c',d')$, then $(ac,bd) \sim (a'c',b'd')$. So we know that $ab' = a'b$ and $cd' = c'd$. Let take a look at $acb'd'$.

$$acb'd' = ba'cd' = ba'dc' = a'c'bd$$

Therefore,$(ac,bd) \sim (a'c',b'd')$. □

To some extend, $\mathbb{Q}$ extends on $\mathbb{Z}$. You can check that

$$\left\{\frac{n}{1} : n \in \mathbb{Z}\right\}$$

which is a subset of $\mathbb{Q}$, behaves just like $\mathbb{Z}$.

1.3 Ordered Set

We know discuss the order on set $\mathbb{Z}$. First we define what is an order.

Definition 1.3.1. Let S be a set. An order on S is a relation denoted by $<$ that satisfies the following two properties.
(i)If $x, y \in S$, then one and only one of the following is true

$$x < y \quad x = y \quad y < x$$

(ii) If $x, y, z \in S$, if $x < y$ and $y < z$, then $x < z$.

Remark 1.3.1. (i)It is often convenient to write $y > x$ instead of $x < y$.
(ii)$x \leq y$ means either $x < y$ or $x = y$ is true. In other words, $x \leq y$ is the negation of $x > y$.

Definition 1.3.2. An ordered set is a set S in which an order is defined.

For example, $\mathbb{Q}$ is an ordered set if $r < s$ means that $s - r$ is a positive rational number. A more rigorous definition is given below.

Definition 1.3.3. In $\mathbb{Q}$, say $\frac{m}{n}$ a positive rational number if $mn > 0$.

We just made a definition! We should check whether it is well defined. This means that $(m,n) \sim (p,q)$ should imply $pq > 0$.

Proof. We know that $pn = mq$. We multiply both side by qm. This gives $pqnm = mmqq = (mq)^2 > 0$. And since $mn > 0$, it must be true that $pq > 0$. □

1.4 The Incompleteness of $\mathbb{Q}$

We have discussed quite a lot to complete definitions of set $\mathbb{Q}$. And we do see that $\mathbb{Q}$ can solve certain problems that $\mathbb{Z}$ cannot. For example, we have 3 cakes and we want to distribute them equally to 5 people. Naturally we want such a number x that satisfies $3x = 5$. Obviously $x \notin \mathbb{Z}$. But we can see that $x \in \mathbb{Q}$ and $x = \frac{5}{3}$.

However, such a set $\mathbb{Q}$ does not solve all equations. For example, the solution of the equation $x^2 = 2$ is not in $\mathbb{Q}$, which we will show.

Proposition 1.4.1. *No number $x \in \mathbb{Q}$ satisfies the equation $x^2 = 2$.*

Proof. We prove this proposition by contradiction. Assuming that there is an $x \in \mathbb{Q}$ that satisfies the equation. Write $x = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and share no common divisors. Substitute $\frac{p}{q}$ into the equation we get $(\frac{p}{q})^2 = 2$, hence

$$p^2 = 2q^2 \tag{1.1}$$

This means p is even, because if p were odd, p^2 would have been odd. Then write $p = 2m$ where $m \in \mathbb{Z}$. Sustitute $p = 2m$ back into Eqn.(1.1) We have

$$4m^2 = 2q^2 \tag{1.2}$$

Therefore q must be even. But then p, q have common divisor 2, which contradicts our assumption. Therefore there is no such $x = \frac{p}{q}$ that satisfies $x^2 = 2$. □

We showed that there is no such number $x \in \mathbb{Q}$ that satisfies $x^2 = 2$. But we will also show, through this example, that there are certain gaps in $\mathbb{Q}$ that needs to be filled.

Proposition 1.4.2. *Let A be the set of all positive rationals p such that $p^2 < 2$. Let B be the set of all positive rationals p such that $p^2 > 2$. Then A contains no largest number and B contains no smallest.*

Proof. We will prove our claim by showing for every $p \in A$ we can find a rational $q \in A$ such that $p < q$. And for every $p \in B$ we can find a rational $q \in B$ such that $q < p$.
We associate with each rational p the number

$$q = p - \frac{p^2 - 2}{p + 2} = \frac{2p + 2}{p + 2}. \tag{1.3}$$

Then,

$$q^2 - 2 = \frac{2\left(p^2 - 2\right)}{(p + 2)^2}. \tag{1.4}$$

If $p \in A$ then $p^2 - 2 < 0$, so Eqn.(1.3) says $q > p$. And Eqn.(1.4) says $q^2 - 2 < 0$, so $q \in A$. If $p \in B$ then $p^2 - 2 > 0$, so Eqn.(1.3) says $q < p$. And Eqn.(1.4) says $q^2 - 2 > 0$, so $q \in B$. □

The discussion above shows that the rational number system has certain gaps despite the fact that between every two rationals there is another (say $r < s$, we can always create a new number between them: $r < (r+s)/2 < s$). The real number system fill these gaps. This is the motivation of why we develop the real number system and why they play such fundamental role in real analysis.

Remark 1.4.1. The proof of Proposition 1.4.2 is utterly elegant. But how did we figure out we should make the association of Eqn.(1.3) in the first place? First of all it is clear that our goal is to show that for every $p \in A$ we can find a rational $q \in A$ such that $p < q$. And for every $p \in B$ we can find a rational $q \in B$ such that $q < p$.

Let's analyze everything we have more carefully. First discuss the case where $p \in A$. We want such a q that satisfies $p - q < 0$ and $q^2 - 2 < 0$(so that $q \in A$). Looking at the form $p - q < 0$, we naturally want to construct something like $q = p - X$, where X is some number smaller than zero. We first try something like $X = \frac{p^2-2}{x}$, where $x > 0$, since we are guaranteed that $p^2 - 2 < 0$. The advantage of such construction, as you will see, is that it also works when $p \in B$.

But then what is x? We have used condition $p^2 - 2 < 0$. But there is one more condition(restriction): $q^2 - 2 < 0$. So naturally we want to square our expresion of q and find $q^2 - 2$. Now

$$q = p - \frac{p^2-2}{x} = \frac{xp - p^2 - 2}{x}.$$

If we want to square this expression, there will be a p^4, which is troublesome. So we want to cancel the p^2 upstairs by cleverly choosing x. We try $x = p + 2 > 0$. Then

$$q = p - \frac{p^2-2}{p+2} = \frac{2p+2}{p+2}.$$

Hence,

$$q^2 - 2 = \frac{2\left(p^2-2\right)}{(p+2)^2}.$$

So we arrived at such an association between q and p because the algebra turned out to be nice and it satisfies the conditions we want.

1.5 Fields

Definition 1.5.1. A field is a set F with two operations, called addition and multiplication, which satisfies the following axioms.

(A)Axioms for Addition

(A1) If $x \in F$ and $y \in F$, then their sum $x + y$ is in F.
(A2) Addition is commutative: $x + y = y + x$ for all $x, y \in F$.
(A3) Addition is associative: $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$.
(A4) F contains an element 0 such that $0 + x = x$ for every $x \in F$.
(A5) To every $x \in F$ corresponds an element $-x \in F$ such that $x + (-x) = 0$

(M)Axioms for multiplication

(M1) If $x \in F$ and $y \in F$, then their product xy is in F .
(M2) Muttiplication is commutative: $xy = yx$ for all $x, y \in F$.
(M3) Muttipication is associative: $(xy)z = x(yz)$ for all $x, y, z \in F$.
(M4) F contains an element $1 \neq 0$ such that $1x = x$ for every $x \in F$.
(M4) If $x \in F$ and $x \neq 0$ then there exists an element $1/x \in F$ such that $x \cdot (1/x) = 1$

(D)The Distributive Law

$$x(y + z) = xy + xz$$

holds for all $x, y, z \in F$.

Definition 1.5.2. An ordered field is a field F which is also an ordered set such that
(i) $x + y < x + z$ if $x, y, z \in F$ and $y < z$
(ii) $xy > 0$ if $x, y \in F$ and $x > 0$ and $y > 0$.

Proposition 1.5.1. $\mathbb{Q}$ *is a field that satisfies A(1-5),M(1-5),and D.*

Proposition 1.5.2. $\mathbb{Q}$ *is an ordered field.*

1.6 Upper and Lower Bounds

We showed in section 1.4 in Proposition 1.4.2 that the set A that includes all numbers p such that $p^2 < 2$ contains no biggest number. And the set B that includes all numbers p such that $p^2 > 2$ contains no smallest number. Now we define rigorously what the "biggest and smallest number in a set" means.

Definition 1.6.1. Consider $E \subset S$ where S is ordered. If there exists $\beta \in S$, such that for all $x \in E$ we have $x \leq \beta$, then β is an upper bound for E. We say E is bounded above.

Similarly, we have a definition for lower bound.

Definition 1.6.2. Consider $E \subset S$ where S is ordered. If there exists $\beta \in S$, such that for all $x \in E$ we have $x \geq \beta$, then β is an lower bound for E. We say E is bounded below.

Exercise 1.6.1. Prove that $\frac{3}{2}$ is an upper bound of $A = \left\{x : x^2 < 2, x \in \mathbb{Q}\right\}$.

Solution 1.6.1. We prove by contradiction. Suppose $\frac{3}{2}$ is not an upper bound for A, then there exists some x such that $x > \frac{3}{2}$. Then $x^2 > \frac{9}{4} > 2$, we have reached a contradiction.

Now we define the least upper bound.

Definition 1.6.3. If there exists $\alpha \in S$ such that
(i) α is an upper bound of E
(ii) If $\gamma < \alpha$ then γ is not an upper bound for E
Then α is called the least upper bound for E. We write

$$\alpha = \sup E$$

Exercise 1.6.2. Let $S = \mathbb{Q}$. Let $E = \left\{\frac{1}{2}, 1, 2\right\}$. Prove that 2 is the least upper bound for E.

Solution 1.6.2. To prove that 2 is the least upper bound, first prove that 2 is an upper bound. Since $2 \in \mathbb{Q}$ and $2 \geq x, \forall x \in E$, 2 is an upper bound. Second, we prove that 2 is the smallest upper bound. Assume the contrary, then there exists some upper bound $\gamma < 2$, but then $\gamma > x, \forall x \in E$ must be false. Therefore, we have a contradiction. So 2 must be the smallest upper bound.

Exercise 1.6.3. Let $S = \mathbb{Q}$. Let $E = \mathbb{Q}^-$, which is the set of all negative rationals. Prove that $\sup E = 0$.

Solution 1.6.3. We see right away that 0 is an upper bound of E. Now we prove it is the smallest upper bound. Suppose $\gamma < 0$, then we can always construct $\frac{\gamma}{2}$, which belongs to $\mathbb{Q}$ and is bigger than γ. So γ is not an upper bound for E. Therefore $\sup E = 0$.

Exercise 1.6.4. Let $S = \mathbb{Q}$. Let $E = \mathbb{Q}$. Prove that $\sup E$ does not exist.

Solution 1.6.4. According to Definition 1.6.3, the condition (i) is not satisfied. So there is no least upper bound. We say then E is unbounded above.

Exercise 1.6.5. Let $S = \mathbb{Q}$. Let $E = A$. Prove that $\sup A$ does not exist.

Solution 1.6.5. See proof of Proposition 1.4.2 in section 1.4. Note that $\sup A$ does not exist even if A is bounded.

We can define the greatest lower bound in similar mannar.

Definition 1.6.4. If there exists $\alpha \in S$ such that
(i) α is an lower bound of E
(ii) If $\gamma > \alpha$ then γ is not an lower bound for E
Then α is called the greatest lower bound for E. We write

$$\alpha = \inf E$$

1.7 The Construction of $\mathbb{R}$

Please forget everything you have learnt at school; for you have not learnt it.

Edmund Landau

We now start constructing $\mathbb{R}$ only in terms of rationals. We will construct it in such a way that it satisfies the *existence theorem*, which is the core part of this chapter. We first present the theorem.

Theorem 1.7.1. *There exists an ordered field $\mathbb{R}$ which has the least-upper-bound property. Moreover, $\mathbb{R}$ contains $\mathbb{Q}$ as a subfield.*

Now we define the "least-upper-bound property" mentioned above.

Definition 1.7.1. A set S has the least-upper-bound property if every nonempty subset of S that has an upper bound also has a least upper bound.

So a set that satisfies the least-upper-bound property will always have a least upper bound if bounded. This means that we won't encounter the problem of set A we seen before.

1.7.1 Dedekind Cut

We have seen the problem with $\mathbb{Q}$. Some subsets of $\mathbb{Q}$ have no least upper bounds. For example, the set $A := \left\{p : p^2 < 2 \text{ and } p > 0\right\}$ has no least upper bound – you can always find a number $r \in A$ such that $r > p$. Therefore, we cannot express "$\sqrt{2}$" that we know of using only $\mathbb{Q}$ field since there are certain "gaps" in $\mathbb{Q}$.

Therefore, we are motivated to consturct a more complete field, which we call $\mathbb{R}$ field. The idea is to construct $\mathbb{R}$ using only $\mathbb{Q}$, the same way as we constructed $\mathbb{Q}$ using only $\mathbb{Z}$.

But then how do we express "$\sqrt{2}$" using only $\mathbb{Q}$? well, we know that there are many elements in $\mathbb{Q}$ that are approaching "$\sqrt{2}$". For example, $\frac{7}{5}, \frac{41}{29}, \frac{239}{169}, \frac{1393}{985}$... We can then the set that contains these approaching numbers to represent "$\sqrt{2}$". In other words, let

$$\sqrt{2} := \left\{p^2 < 2 \text{ or } p < 0, p \in \mathbb{Q}\right\}$$

This set is illustrated in Figure 1.1. We "cut" the axis into two pieces. Such a cut is called a Dedekind cut, which we will call a "cut" from now on.

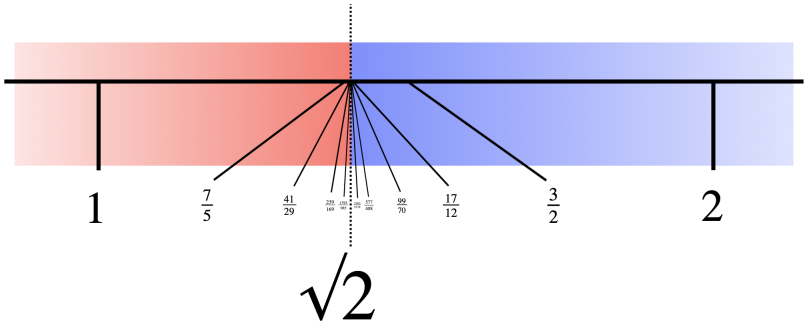


Figure 1.1: Dedekind Cut corresponding to $\sqrt{2}$

A cut, by definition, is a set that has the following three properties: (i) It cannot be empty or the whole $\mathbb{Q}$; (ii), It is closed downward, meaning it should contain everything to the left of the "cutting edge"; (iii) It contains no biggest number. We then put those words into rigorous mathematical definition.

Definition 1.7.2. A cut is any set $\alpha \subset \mathbb{Q}$ with the following three properties:
(I) $\alpha \neq \emptyset, \mathbb{Q}$;
(II) If $p \in \alpha, q \in \mathbb{Q}$, and $q < p$, then $q \in \alpha$;
(III) If $p \in \alpha$, then $p < r$ for some $r \in \alpha$.

Corollary 1.7.1. *(II) in Definition 1.7.2 implies two facts, which will be used freely:*
If $p \in \alpha$ and $q \notin \alpha$, then $p < q$.
If $r \notin \alpha$ and $r < s$ then $s \notin \alpha$.

Proof. (1) Assume the opposite. If $p \in \alpha$ and $p \geq q$, then by (II) $q \in \alpha$. We reached a contradiction. Hence, $p < q$. (2) Assume the opposite. If $s \in \alpha$ and since $r < s$, by (II) we must have $r \in \alpha$. Contradiction. Hence, $s \notin \alpha$. □

Now we begin to define the real system $\mathbb{R}$.

Definition 1.7.3. The member of $\mathbb{R}$ will be certain subset of $\mathbb{Q}$ called cuts.

$$\mathbb{R} := \{\alpha : \alpha \text{ is a cut}\}.$$

In other words, the set of all cuts is defined to be $\mathbb{R}$. From now on we use English letter $p, q, r...$ to represent elements in $\mathbb{Q}$ and greek letter α, β, γ to represent cuts.

1.7.2 Orders of $\mathbb{R}$

Now that we defined $\mathbb{R}$, we want to give it an order.

Definition 1.7.4. (Orders of $\mathbb{R}$). Consider two cuts, $\alpha, \beta \in \mathbb{R}$. We define $\alpha < \beta$ to mean α is a proper set of β.

Now we need to check whether our definition is consistent with Definition 1.3.1. If $\alpha < \beta$ and $\beta < \gamma$ then it is clear that $\alpha < \gamma$. It is also true that at most one of the three relations

$$\alpha < \beta, \quad \alpha = \beta, \quad \beta < \alpha$$

can hold for any pair α, β. Here's proof.

Proof. Assume the first two relations fail. Then α is not a subset of β. Hence there is a $p \in \alpha$ with $p \notin \beta$. If $q \in \beta$, it follows that $q < p$ since $p \notin \beta$ (Corollary 1.7.1). Hence $q \in \alpha$ by (II). Thus $\beta \subset \alpha$. Since $\beta \neq \alpha$, we conclude: $\beta > \alpha$. □

Remark 1.7.1. One thing to point out is that if, for example, the first relation fails, then the second relation $\alpha = \beta$ must fail. Because if $\alpha = \beta$, since $\beta \subset \beta$, we must have $\alpha < \beta$.

Thus now $\mathbb{R}$ is an ordered set.

1.7.3 Define Addition

Now that $\mathbb{R}$ is an ordered set, Let us define the arithmetic on $\mathbb{R}$.

Definition 1.7.5. Take $\alpha, \beta \in \mathbb{R}$. The addition on $\mathbb{R}$ is defined to be:

$$\alpha + \beta := \{p : p = r + s \quad \text{where} \quad r \in \alpha, s \in \beta\}$$

We will now check this definition satisfies the field axioms for addition (A1)-(A5).

(A1) We have to show that $\alpha + \beta$ is a cut. It is clear that $\alpha + \beta \neq \emptyset$. Take $r' \notin \alpha$ and $s' \notin \beta$, it follows then $r' + s' > r + s$ for all $r \in \alpha, s \in \beta$. Therefore, property (I) is met.

For (II), pick $p \in \alpha + \beta$. Then $p = r + s$ with $r \in \alpha, s \in \beta$. If $q < p$, then $q - s < r$. So $q \in \alpha$. But $q = (q - s) + s \in \alpha + \beta$. Thus, (II) holds.

Now go to (III). Choose $t \in \alpha$ such that $t > r$. Then $p = r + s < t + s$ and $t + s \in \alpha$. Thus (III) holds.

(A2) Since addition itself is commutative, it follows $\alpha + \beta = \beta + \alpha$.

(A3) This follows from the associative law in $\mathbb{Q}$.

(A4) *We define our zero element* 0^* *to be the set of all negative rational numbers.* Then if $r \in \alpha$ and $s \in 0^*$, $r + s < r$. Therefore $r + s \in \alpha$. So $\alpha + 0^* \subset \alpha$.

To obtain the opposite, we pick $p, r \in \alpha$ and $r > p$. Then $p - r \in 0^*$, and $p = r + (p - r) \in \alpha + 0^*$. Thus $\alpha + 0^* \subset \alpha$. We then conclude that $\alpha + 0^* = \alpha$.

(A5) Fix $\alpha \in \mathbb{R}$. We define the inverse element β as

$$\beta := \{p : \exists r > 0 \text{ s.t.} - p - r \notin \alpha\}$$

This basically means that β is the set of numbers such that its negative reflex minus a little bit (by an amount of r) fails to be in α.

We first show that $\beta \in \mathbb{R}$. if $s \notin \alpha$ and $p = -s - 1$, then $-p - 1 \notin \alpha$. Hence $p \in \beta$. So $\beta \neq \emptyset$. If $q \in \alpha$, then $q > q - r \in \alpha$. So $-q \notin \beta$. Therefore $\beta \neq \mathbb{Q}$. (I) is done. Pick $p \in \beta$ and $r > 0$. So $-p - r \notin \alpha$. If $q < p$, then $-q - r > -p - r$. Hence $-q - r \notin \alpha$. So $q \in \beta$. (II) holds. Put $t = p + (r/2)$, so $t > p$. And $-t - (r/2) = -p - r \notin \alpha$. So $t \in \beta$. (III)holds. We have proved that β is a cut, or equivalently, $\beta \in \mathbb{R}$.

We then show $\alpha + \beta = 0^*$. If $r \in \alpha, s \in \beta$, then $-s - t \notin \alpha$ for some $t > 0$. But $-s > -s - t$. So $-s \notin \alpha$. Hence $r < -s$, or equivalently $r + s < 0$. Thus $\alpha + \beta \subset 0^*$.

To prove the opposite, pick $v \in 0^*$. Put $w = -v/2$. Then $w > 0$, and there is an integer n such that $nw \in \alpha$ but $(n + 1)w \notin \alpha$ (According to the fact that $\mathbb{Q}$ has the archimedean property, which will be proved later). Put $p = -(n + 2)w$. Then since $-p - w = (n + 1)w \notin \alpha$, and $v = nw + p \in \alpha + \beta$. Thus $0^* \subset \alpha + \beta$.

We then conclude that $\alpha + \beta = 0^*$. β will be denoted by $-\alpha$.

1.7.4 Define Multiplication on $\mathbb{R}^+$

We now define multiplication on $\mathbb{R}$. Multiplication is much more complicated than addition since products of negative rationals are positive. For this reason we first restrict ourselves to $\mathbb{R}^+$, the set of all $\alpha \in \mathbb{R}$ with $\alpha > 0^*$.

Definition 1.7.6. (Multiplication on $\mathbb{R}^+$) Take $\alpha, \beta \in \mathbb{R}^+$. The multiplication on $\mathbb{R}^+$ is defined to be:

$$\alpha\beta := \{p : p \leq rs \quad \text{for some} \quad r \in \alpha, s \in \beta, r, s > 0\}$$

We now need to prove that this definition satisfies (M1)-(M5). Rudin thinks we are smart enough to prove this by ourselves and therefore omitted the proof. So we'll prove it to Rudin.

(M1) (I) It is immediately obvious that $\alpha\beta \neq \emptyset, \mathbb{Q}$. (II) If $p \in \alpha\beta$, take $q \in \mathbb{Q}$ and $q < p$. Then $p \leq rs$ for some positive $r \in \alpha$ and $s \in \beta$. Then $q < p \leq rs$. It follows immediately $q \in \alpha\beta$. (III) Choose $t \in \alpha$ and $t > r$. Let $k = ts$. Then $p \leq rs < ts = k$. But $k = ts \in \alpha\beta$. Done.

(M2),(M3) follow immediately from the commutative and associative law of multiplication on $\mathbb{Q}$.

(M4) We define the identity element 1^* to be

$$1^* := \{q : q < 1\}$$

It is obvious that 1^* is a cut. So $1^* \in \mathbb{R}$. We now prove $1^*\alpha = \alpha$.

First, take $x \in 1^*\alpha$. Then $x \leq qp$ for some $q \in 1^*, p \in \alpha$. But $x \leq qp < p$ and $p \in \alpha$, thus

$x \in \alpha$. Then we have $1^*\alpha \subset \alpha$.

Conversely, take $p \in \alpha$, there exists $q \in \alpha$ and $q > p$. Then $\frac{p}{q} < 1$. So $\frac{p}{q} \in 1^*$. $p \leq q\frac{p}{q}$. So $p \in 1^*\alpha$. This implies $\alpha \subset 1^*\alpha$.

Hence, we conclude $1^*\alpha = \alpha$.

(M5) We define the multiplicative inverse of α to be

$$\beta := \left\{ p : \exists r > 0 \text{ s.t. } \frac{1}{p} - r \notin \alpha \right\}.$$

You can sense that this definition is very similar to the additive inverse. We first prove β is a cut.

(I) It is clear that $\beta \neq \emptyset$. If we choose p such that $\frac{1}{p} \in \alpha$, then $\frac{1}{p} - r \in \alpha$. So $p \notin \beta$. Therefore $\beta \neq \mathbb{Q}$. (II) If $p \in \beta$, then there is some $r > 0$ such that $\frac{1}{p} - r \notin \alpha$. Take $q < p$. Then $\frac{1}{q} > \frac{1}{p}$. Then $\frac{1}{q} - r > \frac{1}{p} - r \notin \alpha$. So $\frac{1}{q} - r \notin \alpha$. Hence $q \in \beta$. (III) If $p \in \beta$, then there is some $r > 0$ such that $\frac{1}{p} - r \notin \alpha$. Take $t = \frac{2p}{2-rp}$. Then $\frac{1}{t} = \frac{1}{p} - \frac{1}{2}$. So $t > p$. Now $\frac{1}{t} - \frac{r}{2} = \frac{1}{p} - r \notin \alpha$. So $t \in \beta$. Done.

Now we prove that $\alpha\beta = 1^*$. First, if $p \in \alpha\beta$, then $p \leq ts$ for some positive $t \in \alpha, s \in \beta$. Since $s \in \beta$, there is some $r > 0$ such that $\frac{1}{s} - r \notin \alpha$. Since $t \in \alpha$ and $\frac{1}{s} - r \notin \alpha$, it must follow then $t < \frac{1}{s} - r$. So $t < \frac{1}{s}$. Thus, $ts < \frac{1}{s}s = 1$. So $ts \in 1^*$. Since 1^* is a cut and $ts \in 1^*$ and $p \leq ts, \alpha\beta \subset 1^*$.

To reach the opposite conclusion, take $0 < t < 1$. So $t \in 1^*$. Take $p \in \alpha$. Since $0 < t < 1$, $0 < 1 - t < 1$. Then $(1-t)p < p$. Hence $(1-t)p \in \alpha$. According to the archimedean property of $\mathbb{Q}$, there is an positive integer n such that $n(1-t)p \in \alpha$ and $(n+1)(1-t)p \notin \alpha$. Call $p_1 = n(1-t)p$ and $p_2 = (n+1)(1-t)p$. Since $p_2 \notin \alpha$ and $p \in \alpha$, $p_2 > p$. So $p_2(1-t) > p(1-t) = p_2 - p_1$. Therefore, $p_2 < \frac{p_1}{t}$. Since $p_2 \notin \alpha$ and $p_2 < \frac{p_1}{t}$, it must follow then $\frac{p_1}{t} \notin \alpha$. Thus, $\frac{p_1}{t} - (\frac{p_1}{t} - p_2) \notin \alpha$. This implies that $\frac{t}{p_2} \in \beta$. Finally, $t \leq p_1\frac{t}{p_1} \in \alpha\beta$. We conclude then $1^* \subset \alpha\beta$.

We can then be certain that $\alpha\beta = 1^*$.

(D) Trivial.

1.7.5 Multiplication on $\mathbb{R}$

We can complete the definition of multiplication by setting $\alpha 0^* = 0^*\alpha = 0^*$ and by setting

$$\alpha\beta - \begin{cases} (-\alpha)(-\beta) & \text{if } \alpha < 0^*, \beta < 0^* \\ [(-\alpha)\beta] & \text{if } \alpha < 0^*, \beta > 0^* \\ -[\alpha \cdot (-\beta)] & \text{if } \alpha > 0^*, \beta < 0^* \end{cases}$$

We now need to redo the proof of (M1)-(M5) and (D). By using the identity $\gamma = -(-\gamma)$, this is trivial. So we omit the proof.

1.7.6 Least Upper Bound Property on $\mathbb{R}$

Having prove that $\mathbb{R}$ is an ordered field, we now need to prove that $\mathbb{R}$ satisfies the least upper bound property. We leave the proof to Rudin. But here is the basic idea.

We want to prove that if you take some nonempty set $A \in \mathbb{R}$ (A is the set of cuts), $\sup A$ exists. Imagine $A = \{\alpha_1, \alpha_2, ...\alpha_n\}$ where all the elements are cuts containing some rationals. Then what is the least thing that contains all the stuff in A? It should be the union of all the cuts, intuitively. So we define

$$\gamma := \alpha_1 \cup \alpha_2 \cup ... \cup \alpha_n$$

and we prove that $\gamma = \sup A$.

1.8 The Real Field

We worked hard to construct $\mathbb{R}$. In previous section, we proved that $\mathbb{R}$ ***is an ordered field with least upper bound property.***

But everything is so NOT obvious. You may object, or at least question, can the real number field, defined as the set of all cuts, express something that $\mathbb{Q}$ fails to express?

Let us look back at the stage when we were defining Dedekind cuts. We say we can express "$\sqrt{2}$" using rationals by assigning it to

$$\gamma := \left\{p^2 < 2 \,\text{or}\, p < 0, p \in \mathbb{Q}\right\}$$

Then by our definition of multiplication on $\mathbb{R}$ (without loss of generality, we only consider the multiplication on $\mathbb{R}^+$),

$$\gamma^2 = \{p : p \leq rs \quad \text{for some positive} \quad r, s \in \gamma\}$$

We will prove that $\gamma^2 = 2^*$, where

$$2^* := \{p : p < 2 \quad \text{or} \quad p < 0\}$$

Proof. We start by verifying 2^* is a cut, we leave the work to you. We now want to prove that $\gamma^2 = 2^*$.

First, if $p \in \gamma^2$, then $p \leq rs$ for some positive $r, s \in \gamma$. Since $p \leq rs$, then $p^2 \leq (rs)^2 = r^2 s^2 < 2 \cdot 2 = 4$. Therefore $p < 2$. We get $\gamma^2 \subset 2^*$.

If $p \in 2^*$ and $p < 0$, since $rs > 0$, it follows immediately that $p \in \gamma^2$. If $0 < p < 2$, then $p^2 < 2 \cdot 2$. Therefore, $p^2 < r^2 s^2$ for some positive $r, s \in \gamma$. Thus $p < rs$. So $p \in \gamma^2$. We then conclude $2^* \subset \gamma^2$.

Hence the proof is finished. □

We will also point out the fact that *any two ordered field with the least upper bound property are isomorphic*. This means that $\mathbb{R}$ is the only field that has such property. If we are able to construct another ordered field with the least upper bound property, then this field must also be $\mathbb{R}$.

Previously in our presentation of defining $\mathbb{R}$, we said the addition and multiplication are defined because of the archimedean property of $\mathbb{Q}$. Now we define what archimedean property is.

Definition 1.8.1. If $x, y \in \mathbb{Q}$ and $x > 0$, and if there exists a positive integer n such that

$$nx > y$$

then we say $\mathbb{Q}$ has the archimedean property.

Now we prove that $\mathbb{Q}$ has the archimedean property.

Proof. First, observe that if $x > y$, the proof is trivial. Now if $x < y$, let $x = \frac{a}{b}, y = \frac{c}{d}$, where $a, b, c, d \in \mathbb{N}^+$. Take $n = bc + 1$. Since $a, d \geq 1$, then $ad \geq 1$. Therefore, $ad(bc + 1) > bc$. We have $nad > bc$. Therefore, $n\frac{a}{d} > \frac{b}{c}$. □

And you may already have guessed that $\mathbb{R}$ also has archimedean property. We know introduce two property of $\mathbb{R}$.

Theorem 1.8.1. *(a)If $x, y \in \mathbb{R}$ and $x > 0$, and if there exists a positive integer n such that*

$$nx > y.$$

(b) If $x, y \in \mathbb{R}$ and $x < y$, then there exists $p \in \mathbb{Q}$ such that $x < p < y$.

Part (a) is the archimedean property of $\mathbb{R}$. Part (b) says $\mathbb{Q}$ is *dense* in $\mathbb{R}$.

Proof. (a) Let $A = \{nx : n \in \mathbb{Z}^+\}$. Assume that (a) is false, then y would be an upper bound for A. But then A has a least upper bound in $\mathbb{R}$. Put $\alpha = \sup A$. Since $x > 0$, $\alpha - x < \alpha$. So $\alpha - x$ is not an upper bound of A. Hence $\alpha - x < mx$ for some positive integer m. But then $\alpha < (m + 1)x \in A$, which is impossible since α is an upper bound of A. Contradiction.

(b) First let's go back to (a). If we put $y = 1$, this means there exists $n \in \mathbb{Z}^+$ such that $\frac{1}{n} < x$ for $x > 0$.

Since $x < y$, we have $y - x > 0$. Choose n such that $\frac{1}{n} < y - x$. But from (a) we also know that the multiples of $\frac{1}{n}$ is unbounded. We choose first multiple $\frac{m}{n}$ such that $\frac{m}{n} > x$.

Then I claim $\frac{m}{n} < y$ is also true. If not, using two inequalities we have

$$\frac{m}{n} > y \qquad \frac{m-1}{n} < x$$

We will get $\frac{1}{n} > y - x$, a contradiction. Therefore, there exists a rational $p = \frac{m}{n}$ such that $x < p < y$.

□

1.9 Euclidean Spaces and the Complex Field

Definition 1.9.1. For each positive integer k, let $\mathbb{R}^k$ be the set of all ordered k-tuples

$$\mathbf{x} = (x_1, x_2, \ldots, x_k)$$

where $x_1, x_2, ..., x_k \in \mathbb{R}$, called the coordinates of $\mathbf{x}$. The elements of $\mathbb{R}^k$ are called points, or vectors.

Definition 1.9.2. (Arithmetic) If $\mathbf{x} = (x_1, x_2, \ldots, x_k)$, $\mathbf{y} = (y_1, y_2, ..., y_k)$ and $\alpha \in \mathbb{R}$, then we define addition to be

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, ..., x_k + y_k)\,.$$

Multiplication of a vector and a scaler is defined to be

$$\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, ..., \alpha x_k)\,.$$

So that $\mathbf{x} + \mathbf{y} \in \mathbb{R}^k$ and $\alpha\mathbf{x} \in \mathbb{R}^k$. These two operations are commutative, associative, and distributive. Therefore $\mathbb{R}^k$ is a vector space over the real field. We define the zero element to be

$$\mathbf{0} := (0, 0, ..., 0)$$

Definition 1.9.3. (Inner Product) We define the inner product of $\mathbf{x}, \mathbf{y}$ to be

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{k} x_i y_i$$

Corollary 1.9.1. *The norm of* $\mathbf{x}$*, also known as the magnitude, can be found by*

$$|\mathbf{x}| = (\mathbf{x} \cdot \mathbf{x})^{1/2} = \left(\sum_{i=1}^{k} x_i^2\right)^{1/2}$$

Now we start talking about the complex number.

Definition 1.9.4. A complex number is an ordered pair of real numbers (a, b), where $a, b \in \mathbb{R}$. We denote the set of complex numbers by $\mathbb{C}$.

Definition 1.9.5. (Arithmetic on $\mathbb{C}$) Let $x = (a, b)$, $y = (c, d)$ be two complex numbers. We write $x = y$ if and only if $a = c$ and $b = d$. We define

$$x + y = (a + c, b + d) \qquad xy = (ac - bd, ad + bc).$$

Theorem 1.9.1. *The definitions of arithmetic turn* $\mathbb{C}$ *into a field, with* $(0, 0)$ *and (1,0) in the role of* $\mathbf{0}$ *and* $\mathbf{1}$.

Proof. See Rudin pg13. □

Definition 1.9.6. $i = (0, 1)$.

Theorem 1.9.2. $i^2 = -1$.

Proof. $i^2 = (0,1)(0,1) = (-1,0) = -1$. □

Theorem 1.9.3. *If $a, b \in \mathbb{R}$, then $(a,b) = a + bi$.*

Definition 1.9.7. If $a, b \in \mathbb{R}$, and $z = a + bi$, then the complex number $\bar{z} = a - bi$ is the conjugate of z. We write

$$a = Re(z) \qquad b = Im(z).$$

Theorem 1.9.4.

(a) $|z| > 0$ *unless* $z = 0$, $|0| = 0$
(b) $|\bar{z}| = |z|$
(c) $|zw| = |z||w|$
(d) $|\operatorname{Re} z| \le |z|$
(e) $|z + w| \le |z| + |w|$

Proof. The proof of (a)-(d) we leave to Rudin. To prove (e), note that $\bar{z}w$ is the conjugate of $z\bar{w}$. So $z\bar{w} + \bar{z}w = 2Re(z\bar{w})$. Hence

$$\begin{aligned}
|z + w|^2 = (z + w)(\bar{z} + \bar{w}) &= z\bar{z} + z\bar{w} + \bar{z}w + w\bar{w} \\
&= |z|^2 + 2\operatorname{Re}(z\bar{w}) + |w|^2 \\
&\le |z|^2 + 2|z\bar{w}| + |w|^2 \\
&= |z|^2 + 2|z||w| + |w|^2 = (|z| + |w|)^2
\end{aligned}$$

Now (e) follows by taking the square root. □

Definition 1.9.8. If $x_1, x_2, \ldots, x_n \in \mathbb{C}$, we write

$$x_1 + x_2 + \cdots + x_n = \sum_{j=1}^{n} x_j.$$

We conclude our discussion of complex numbers with an important inequality, called the Schwarz inequality.

Theorem 1.9.5. *(Schwarz inequality) If $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are complex numbers, then*

$$\left| \sum_{j=1}^{n} a_j \bar{b_j} \right|^2 \le \sum_{j=1}^{n} \left|a_j\right|^2 \sum_{j=1}^{n} \left|b_j\right|^2$$

Proof. Put $A = \sum |a_j|^2$, $B = \sum |b_j|^2$, $C = \sum a_j \bar{b_j}$, where j runs from 1 to n. If $B = 0$, the proof is trivial. Assume therefore that $B > 0$. By Thm.1.9.4 we have

$$\begin{aligned}
\sum \left|Ba_j - Cb_j\right|^2 &= \sum \left(Ba_j - Cb_j\right)\left(B\bar{a}_j - \overline{Cb}_j\right) \\
&= B^2 \sum \left|a_j\right|^2 - B\bar{C} \sum a_j \bar{b}_j - BC \sum \bar{a}_j b_j + |C|^2 \sum \left|b_j\right|^2 \\
&= B^2 A - B|C|^2 \\
&= B\left(AB - |C|^2\right)
\end{aligned}$$

But since $\sum \left|Ba_j - Cb_j\right|^2 \geq 0$, We have

$$B(AB - |C|^2) \geq 0.$$

Since $B > 0$, we have

$$AB - |C|^2 \geq 0,$$

which concludes the proof. □

Remark 1.9.1. Note that if we let $Im(a_i) = Im(b_i) = 0$, in other words $a_i, b_i \in \mathbb{R}$ for all $i \in \mathbb{Z}^+$, then the first term on the right hand side $\sum a_j \bar{b_j} = \sum a_j b_j$, which can be seen as $\mathbf{a} \cdot \mathbf{b}$ where $\mathbf{a} = (a_1, a_2, ...a_n)$ and $\mathbf{b} = (b_1, b_2, ..., b_n)$. Schwarz inequality then implies that

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|.$$

Now that we know the Schwarz inequality, we say a little bit more about Euclidean spaces.

Theorem 1.9.6.

(a) $|\mathbf{x}| \geq 0$
(b) $|\mathbf{x}| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$
(c) $|\alpha\mathbf{x}| = |\alpha||\mathbf{x}|$;
(d) $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$
(e) $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$
(f) $|\mathbf{x} - \mathbf{z}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{z}|$

Proof. (a)-(c) are obvious. (d) is immediate from the Schwarz inequality. By (d) we have

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) \\ &= \mathbf{x} \cdot \mathbf{x} + 2\mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{y} \\ &\leq |\mathbf{x}|^2 + 2|\mathbf{x}||\mathbf{y}| + |\mathbf{y}|^2 \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2 \end{aligned}$$

so (e) is proven. (f) follows from (e) if we make the following substitution:

$$\mathbf{x} \to \mathbf{x} - \mathbf{y} \qquad \mathbf{y} \to \mathbf{y} - \mathbf{z}$$

□

Remark 1.9.2. $\mathbb{R}^2$ is usually called the plane, or the complex plane.

Chapter 2

Basic Topology

Point set topology is a disease from which human race will soon recover.

-Henri Poincare

2.1 Induction

The idea of induction can go back to Peano's Axioms, which we didn't focus much of our attention on. Here, we will use the well ordering principle of the natural numbers introduce the principle of induction. Recall that the set of natural number is

$$\mathbb{N} = \{1, 2, 3, 4...\}$$

Axiom 2.1.1. *(The Well Ordering Property)* $\mathbb{N}$ *is well-ordered. Every subset of* $\mathbb{N}$ *has a least element.*

Now we introduce the principle of induction.

Theorem 2.1.1. *(Principle of Induction) Let* S *be a subset of* $\mathbb{N}$ *such that:*
(i) $1 \in S$,
(ii) if $k \in S$*, then* $k + 1 \in S$,
then $S = \mathbb{N}$.

We will prove that the well ordering principle (WOP) implies the principle of induction (POI).

Proof. We will prove by contradiction. Suppose S has the properties in Thm.2.1.1 and $S \neq \mathbb{N}$. Consider $A = \mathbb{N} \setminus S$. Then according to Ax.2.1.1, there is a least element in A, which we will call n. Then $n - 1 \notin A$ since n is already the least element. This implies that $n - 1 \in S$. But according to (ii), the successor of $n - 1$ should be in S. So $n \in S$. We have reached a contradiction. Therefore, $S = \mathbb{N}$. □

From now on we will sometimes use induction in our proofs. Proof by induction goes as follows: Let $P(n)$ be *statement* indexed by $n \in \mathbb{N}$. To show that $P(n)$ is true for all n, we show:

(1) $P(1)$ is true. Note that this is often known as the base case;
(2) If $P(n)$ is true, then $P(n+1)$ is also true. This is often known as the inductive step.
Then by POI $P(n)$ is true for all n.

Remark 2.1.1. Sometimes it is very difficult to show that $P(n) \to P(n+1)$ is true. Then a conventional trick is to show that $\neg P(n+1) \to \neg P(n)$. This is called proof by contraposition. It is based on the fact that $\neg P(n+1) \to \neg P(n)$ has the same value as $P(n) \to P(n+1)$ on the truth table, as shown by Fig.(2.1) below.

p	q	$\neg p$	$\neg q$	$p \to q$	$\neg q \to \neg p$
T	T	F	F	T	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Figure 2.1: the truth table of $p \to q$ and $\neg q \to \neg p$.

We will soon be using POI and contraposition in the next section, where we focus more on the properties of sets, and the relations between two sets.

2.2 Functions, Countable Sets, Infinite Sets

Definition 2.2.1. Consider two sets A and B. Suppose that for each element $x \in A$, there is associated, in some manner, an element of B, which we denote $f(x)$. Then f is said to be a function from A to B (or a mapping of A into B). Write

$$f : A \to B$$

We call A the domain of f and B the **codomain** of f. The elements $f(x)$ are called the values of f. We define the **range** of f to be:

$$\{f(x) : x \in A\}$$

In other words, the range is the set of all values of f.

Before we move on to the next definition, we shall first introduce one notation. Consider a *set* C. We define

$$f(C) := \{f(x) : x \in C\}$$

We call $f(C)$ the image of C.

Definition 2.2.2. Consider $f : A \to B$.
(i) When $f(A) = B$, we say f is onto (a surjection).
(ii) When $f(x) = f(y)$ implies $x = y$, say f is 1-1 (an injection).
(iii) When f is 1-1 and onto, call f a bijection.

(i) says that the codomain of f is equal to its range. (ii) says that f only sends one point to one point. (iii) says f is a bijection if and only if both (i) and (ii) are satisfied. Fig.(2.2)illustrate the difference.

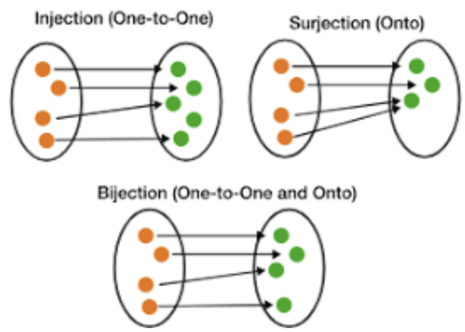


Figure 2.2: Illustrations

Definition 2.2.3. If there exists a 1-1 mapping of A onto B, we say that A and B are in "1-1 correspondence". In other words, A and B are equivalent, and we write $A \sim B$. It has the following properties:
(i) $A \sim A$ (reflexive);
(ii) If $A \sim B$, then $B \sim A$ (symmetric);
(iii) If $A \sim B$ and $B \sim C$, then $A \sim C$ (transitive).
Any relation with these three properties is called an equivalence relation.

Definition 2.2.4. For any positive integer n, let J_n be the set whose elements are the integers 1,2,...n, in other words

$$J_n := \{1, 2, ..., n\}.$$

Define $J_0 := \emptyset$.

Definition 2.2.5. For any set A, we say:
(a) A is finite if $A \sim J_n$ for some n.
(b) A is infinite if A is not finite.
(c) A is countable if $A \sim \mathbb{N}$.
(d) A is uncountable if A is either finite nor countable.
(e) A is at most countable if A is finite or countable.

Example 2.2.1. $A = \{2, 3, 4, 5...\}$ is countable. We can find a 1-1 correspondence between $\mathbb{N}$ and A. The function $f(n) = n + 1$, where $f : \mathbb{N} \to A$.

Exercise 2.2.1. Show that $\{1, 2, 3..., k-1, k+1, k+2, ...\}$ is countable.

Solution 2.2.1. We can assign a function that is 1-1 and onto by considering

$$f(n) = \begin{cases} n & (n < k) \\ n+1 & (n \geq k) \end{cases}$$

Theorem 2.2.1. *$\mathbb{N}$ is infinite.*

Proof. (by induction) We want to show that there is no n such that $\mathbb{N} \sim J_n$. We first show that the base case is true. It is clear that $\mathbb{N} \nsim J_1$ because $\mathbb{N} \setminus J_1 \neq \emptyset$. Then we want to show that if $\mathbb{N} \nsim J_n$, then $\mathbb{N} \nsim J_{n+1}$.

We will prove the inductive step by contraposition. More specifically, we want to prove that if $\mathbb{N} \sim J_{n+1}$, then $\mathbb{N} \sim J_n$. If $\mathbb{N} \sim J_{n+1} = 1, 2, 3..., n+1$, let's call the mapping $h : J_{n+1} \to \mathbb{N}$ (Note that $\sim$ is reflexive). Then if we take out that one additionally element, $\mathbb{N} \setminus \{h(n+1)\} \sim J_n$ must hold. But from Exercise(2.2.1) we already showed that $\mathbb{N} \setminus \{h(n+1)\} \sim \mathbb{N}$. We call this function h'. Therefore $J_n \sim \mathbb{N}$ must be true (illustrated by the following diagram).

$$\begin{tikzcd}
\mathbb{N} \setminus \{h(n+1)\} \arrow[r, "h'"] & J_n \\
\mathbb{N} \arrow[u, "f"] \arrow[ur, "f^{-1} \circ h'"'] &
\end{tikzcd}$$

As shown in the diagram, we just compose the two functions f^{-1} and h'. Therefore we have shown that if $J_{n+1} \sim \mathbb{N}$, then $J_n \sim \mathbb{N}$ (or reverse the two sets since $\sim$ is reflexive). But this is equivalent to saying that if $\mathbb{N} \nsim J_n$ then $\mathbb{N} \nsim J_{n+1}$. Therefore, by induction, we have shown that there is no such n that makes $\mathbb{N} \sim J_n$. Hence, $\mathbb{N}$ is infinite.

□

We can also see that a sequence is countable. For example, define the sequence X to be:

$$X := \{x_1, x_2, ..., x_n, ...\}.$$

We can then assign the bijective mapping to be

$$f(n) = x_n.$$

In other words, if we can list something in a set, then it is countable.

Theorem 2.2.2. *Every infinite subset E of a countable set A is countable.*

Proof. Since A is countable, we can list the elements in A. Let's say, for convinience, $A = \{x_1, x_2, ...x_n, ...\}$. If we can list the things in E, then E is countable. We can let

$$n_1 = \inf\{i : x_i \in E\}.$$

We know that this infimum exists because of the WOP. Then we can let

$$n_2 = \inf\{i : x_i \in E, i > n_1\}.$$

In general, we can let

$$n_k = \inf\{i : x_i \in E, i > n_{k-1}\}.$$

Then we can conclude that

$$E = \{x_{n_1}, x_{n_2}, ...x_{n_k}...\}$$

And since we can list E, we conclude that E is countable. □

Exercise 2.2.2. Prove that $\mathbb{Z}$ is countable.

Solution 2.2.2. See Rudin pg25, Example 2.5.

2.3 Cantor Diagonalization

Theorem 2.3.1. *$\mathbb{Q}$ is countable.*

Proof. Since we know that $\mathbb{Z}$ is countable, then there exists a bijection $u : \mathbb{N} \to \mathbb{Z}$. It is then sufficient to find a bijection $v : \mathbb{Z} \to \mathbb{Q}$ since $u \circ v : \mathbb{N} \to \mathbb{Q}$ is a bijection. You can refer to Rudin pg 29 Thm 2.12, where he introduces the diagonal counting. The idea is to list the rationals in the following way

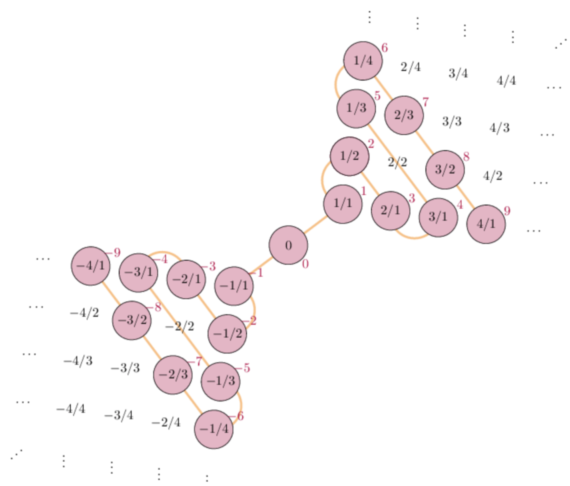


Figure 2.3: Illustrations of such a bijection v.

We follow the worm back and forth "counting" the rational numbers, skipping any numbers that are not simplified fractions. We will leave the details to you. □

The following lemma can be extracted from above without much effort.

Lemma 2.3.1. *If A is countable, then $A \times A$ is countable.*

Theorem 2.3.2. *$\mathbb{R}$ is uncountable.*

Proof. Assume the opposite, that we can make a somewhat 1-1 correspondence between $\mathbb{N}$ and $\mathbb{R}$. Like the following one:

$$1 \to 0.12345678...$$
$$2 \to 0.31415926...$$
$$3 \to 0.14142323...$$
$$4 \to 0.77777776...$$
$$5 \to 0.45643245...$$
$$...$$

Note that we are using decimal expansions of real numbers (see Rudin pg11 1.22 if you have questions). Suppose that we can list all real numbers using this spreadsheet, we want to prove that there exists some number x^* that is not on the list. First, take x^* to be a real number whose n-th digit does not agree with the n-th digit of the n-th number. For example, $x^* = 0.77717....$ x^* cannot be the first number since the first digit of the first number is 1, not 7. It cannot be the second number since the second digit of the second number is 1 not 7. It cannot be the third number... Therefore, we have created some $x^* \in \mathbb{R}$ but is not on the list. So there is no way we can list $\mathbb{R}$. Hence $\mathbb{R}$ is not countable. $\square$

Remark 2.3.1. Note that this proof does not prove that $\mathbb{Q}$ is uncountable. Yes, you can suppose that we can list all rationals on the list, and you can create a x^* that is not on the list, but there is no guarantee that x^* is rational. Therefore is proof fails for $\mathbb{Q}$.

Theorem 2.3.3. *Let $\{E_n\}$, $n = 1, 2, 3, ...$, be a sequence of countable sets, then their union, denoted by*

$$S = \bigcup_{n=1}^{\infty} E_n,$$

is also countable.

Proof. We want to show that we can list S as a sequence. This is not hard to do. For each set E_n, we can list its element as follows:

$$E_n = \{x_{n1}, x_{n2}, x_{n3}...\}$$

Then we can list S in the following two dimensional array and go in the direction of arrow, shown in the following:

$$\begin{array}{ccccc}
x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & \cdots\cdots \\
x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & \cdots\cdots \\
x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & \cdots\cdots \\
x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & \cdots\cdots \\
\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots
\end{array}$$

In such way we can hit all the element in S. Therefore, we can list S. And if S contains some repetitive elements, we will simply use the fact that *the infinite subset of a countable set is also countable* (which was proved previously) to justify our argument. This means that there exists a subset $T \subset \mathbb{Z}^+$ such that $S \sim T$, which implies that S is *at most countable*. Since E_1 is infinite and $E_1 \subset S$, then S is infinite, so S must be countable. $\square$

Definition 2.3.1. Give set A, the power set, denoted by 2^A, is the set of all subsets of A.

Example 2.3.1. If $A = \{a, b, c, d\}$, then say we have

$$\begin{aligned} D &= \{a, b, c\}\,; \\ E &= \{b, c\}\,; \\ F &= \emptyset. \end{aligned}$$

which are all subsets of A. So $D, E, F \in 2^A$.

Remark 2.3.2. If there are n elements in A, then it is very easy to show that there are 2^n elements in the power set of A, which is why we write 2^A. The idea of the proof is demonstrated in the following remark.

Remark 2.3.3. Note that we can express set D, E, F in binary numbers, indicating whether they contain certain elements of A. specifically,

$$\begin{aligned} A &= \{a, b, c, d\} \\ D &= 1, 1, 1, 0; \\ E &= 0, 1, 1, 0; \\ F &= 0, 0, 0, 0. \end{aligned}$$

Such idea is the key to prove the following theorem.

Theorem 2.3.4. *(Cantor's Theorem) For any set A, we have $A \nsim 2^A$.*

We frist present a clever idea of the proof. We want to apply argument similar to the diagonal prove that says $\mathbb{R}$ is uncountable. Suppose the contrary, then there exists a bijection $f : A \to 2^A$. This function f takes some $x \in A$ and maps it to $f(x) \in 2^A$. Probably the following way:

$$\begin{aligned} x_1 &\to 1, 1, 1, 0, ... \\ x_2 &\to 0, 1, 1, 0, ... \\ x_3 &\to 0, 0, 0, 0, ... \\ x_4 &\to 0, 1, 0, 1, ... \end{aligned}$$

We create a set $P \in 2^A$ by taking the diagonal:

$$P = 1, 1, 0, 1.$$

And we create P^* by turning every 1 in P to 0 and turning every 0 in P to 1:

$$P^* = 0, 0, 1, 0$$

Now we can see that $P^* \in 2^A$ but it cannot be on the list. Therefore, we conclude that $A \nsim 2^A$.

Cantor constructed its proof from this idea but using a more rigorous and nonintuitive construction.

Proof. (by contradiction). Suppose $A \sim 2^A$. Then, there exists a bijection $f : A \to 2^A$. This function f takes some $x \in A$ and maps it to $f(x) \in 2^A$. Consider the set

$$B := \{a : a \notin f(a)\}.$$

Since B contains some elements $a \in A$, it is a subset of A, therefore $B \in 2^A$. According to the assumption, $B = f(x)$ for some $x \in A$. Note that either $x \in B$ or $x \notin B$ is true. We prove our assumption wrong by showing that we will reach a contradiction on both cases.

Suppose $x \in B$, then $x \notin f(x) = B$. Contradiction.

Suppose $x \notin B$, then $x \notin f(x)$. But this implies $x \in B$. Contradiction again.

Therefore, our assumption that $B = f(x)$ for some $x \in A$ must be wrong. Hence, no such bijection f exists. We conclude $A \nsim 2^A$. □

2.4 Metric Spaces

Definition 2.4.1. A set X, whose elements we shall call points, is said to be a metric space if with any two points $p, q \in X$ there is associated a real number $d(p,q)$, called the distance from p to q, such that
(i) $d(p,q) > 0$ if $p \neq q$; $d(p,q) = 0$ if and only if $p = q$;
(ii) $d(p,q) = d(q,p)$;
(iii) $d(p,q) \leq d(p,r) + d(r,q)$, for any $r \in X$.

Any function with these three properties is called a distance function, or a metric.

Example 2.4.1. The distance in $\mathbb{R}^k$ is defined by

$$d(\mathbf{x},\mathbf{y}) = |\mathbf{x} - \mathbf{y}| \qquad (\mathbf{x},\mathbf{y} \in \mathbb{R}^k)$$

From Thm.(1.9.6), we can see that the three properties are satisfied.

Definition 2.4.2. A neighborhood of p is a set $N_r(p)$ consisting of all q such that $d(p,q) < r$ for some $r > 0$. The number r is called the radius of $N_r(p)$.

Example 2.4.2. In $\mathbb{R}$, the neighborhoods are segments. In $\mathbb{R}^2$, the neighborhoods are circles.

Definition 2.4.3. A point p is a *limit point* of the set E if every neighborhood of p contains a point $q \neq p$ such that $q \in E$.

Theorem 2.4.1. *If p is a limit point of a set E, then every neighborhood contains infinitely many points of E.*

Proof. Suppose there is a neighborhood N of p that contains only finitely many points distinct from p, for which we denote $q_1, q_2, ..., q_n$. Put

$$r = \min_{1 \leq m \leq n} d(p, q_m).$$

Then clearly the neighborhood $N_r(p)$ contains no point q of E such that $q \neq p$. A contradiction. □

Corollary 2.4.1. *A finite point set has no limit points.*

We will select some other definitions from Rudin's book, which you can refer to Definition 2.18 in the book.

Definition 2.4.4. E is closed if every limit point of E is a point of E (E *contains all its limit points*).

Definition 2.4.5. A point p is an interior point of E if there is a neighborhood N of p such that $N \subset E$.

Definition 2.4.6. E is open if every point of E is an interior point of E.

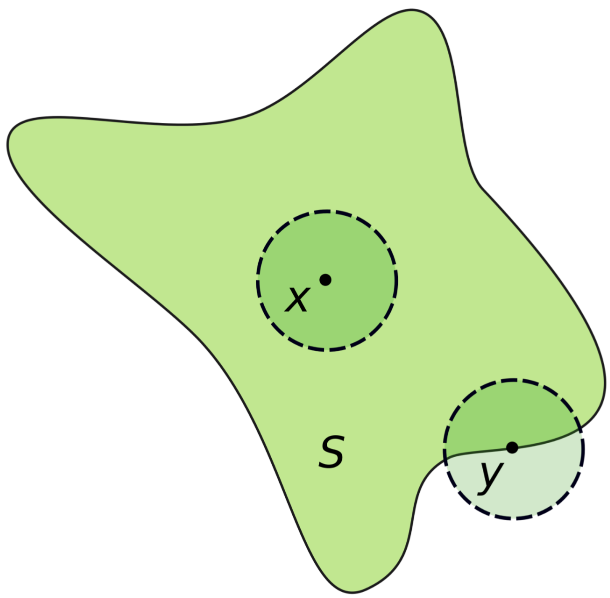


Figure 2.4: S is a closed set since y is not an interior point of S, it is a limit point of S. However, x is an interior point of S.

Theorem 2.4.2. *Every neighborhood is an open set.*

Proof. Consider a neighborhood $E = N_r(p)$. Let q be any point in E. Then there is a positive real number h such that

$$d(p, q) = r - h.$$

For all points s such that $d(q, s) < h$, we have then

$$d(p, s) \leq d(p, q) + d(q, s) \leq r - h + h = r,$$

so that $s \in E$. Thus q is an interior point of E. $\square$

Exercise 2.4.1. A *closed interval* $A = [a, b]$, is defined as the following

$$[a, b] := \{x : a \leq x \leq b\}.$$

Prove that A is closed.

Solution 2.4.1. We claim that a is a limit point of A. Since for every neighborhood $a - \varepsilon < x < a + \varepsilon$ we can find a point $q \in A$. Similar argument applies to point b, and all $x \in A$. Since every limit points of A is contained in A, we conclude that A is closed.

Exercise 2.4.2. Is a single point $\{p\}$ closed?

Solution 2.4.2. Yes. Let $X = \{p\}$. Is every limit point of X contained in X? Since there are no limit points of X, this statement is true. Therefore the set of a single point is closed.

Exercise 2.4.3. We define an *open interval* $A = (a, b)$ to be

$$A = \{x : a < x < b\}.$$

Prove that A is an open set.

Solution 2.4.3. Let q be any point in A. There exists $\varepsilon > 0$ such that if $q - \varepsilon < x < q + \varepsilon$, then $x \in A$ for all x. Therefore there exists such neighborhood $N_\varepsilon(q)$. Hence A is open.

Definition 2.4.7. if X is a metric space and $E \subset X$, and if E' denotes the set of all limit points of E in X, then the closure of E, denoted by $\bar{E}$, is defined as

$$\bar{E} = E \cup E'.$$

Theorem 2.4.3. *$\bar{E}$ is closed.*

Proof. **(My proof)** We want to show that if we take any limit point of $\bar{E}$, say q, then $q \in \bar{E}$. Since $\bar{E} = E \cup E'$, we want to show that either $q \in E$ or q is a limit point of E.
First, observe that if $q \in E$, then $q \in \bar{E}$, and we are done.
If $q \notin E$ and q is a limit point of $\bar{E}$, then

$$\exists e \in \bar{E}, \text{ such that } e \in N_\varepsilon(q), \forall \varepsilon > 0.$$

(i) If $e \in E$, then q is a limit point of E, done.
(ii) If $e \in E'$, then e is a limit point of E since E' is the set of all limit points in E. This implies that

$$\exists r \in \bar{E}, \text{ such that } r \in N_\epsilon(e), \forall \epsilon > 0.$$

But notice that in a metric space X we have

$$d(q, r) \leq d(q, e) + d(e, r) < \varepsilon + \epsilon$$

This would mean that there exists a point $r \in E$ such that $r \in N_{\varepsilon+\epsilon}(q)$, for all $\varepsilon + \epsilon > 0$. Therefore, q is a limit point of E. □

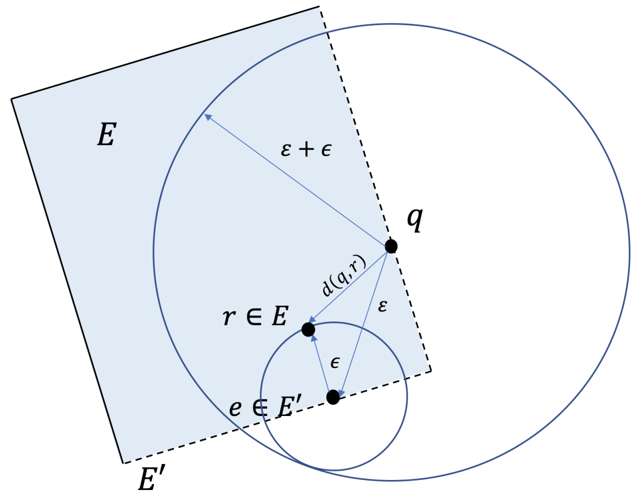


Figure 2.5: A diagram demonstrating the case when $e \in E'$. We can find a neighborhood of q that contains entirely the neighborhood of e.

Rudin's book presents another way of proving Thm.(2.4.3) by introducing the complement of a set, which we will now demonstrate.

Definition 2.4.8. The complement of E, which we denote as E^c, is defined as

$$E^c = \{x : x \notin E\}.$$

Lemma 2.4.1. *A set E is open if and only if its complement is closed.*

Proof. We want to show that E^c is closed implies E is open, and E is open implies E^c is closed. We start by assuming E^c is closed. Choose $x \in E$. Then $x \notin E^c$ and x is not a limit point of E^c (since a closed set contains all its limit points). Hence, there exists a neighborhood N of x such that it doesn't include elements in E^c, or $E^c \cap N = \emptyset$. Then $N \subset E$. Thus x is an interior point of E. So E is open.
Next, assume that E is open. Let x be a limit point of E^c. Then every neighborhood of x comtains a point of E^c. So x is not an interior point of E. But since E is open, if x is not an interior point of E, we have $x \notin E$. This means $x \in E^c$.

□

Corollary 2.4.2. *A set E is closed if and only if its complement is open.*

Now we can use Corollary 2.4.2 to prove the theorem above, which we restate here.

Theorem 2.4.4. *$\bar{E}$ is closed.*

Proof. **(Rudin's proof)** Take $p \in \bar{E}^c$. In other words, $p \notin E$ and p is not a limit point of E. We want to show that p is an interior point of $\bar{E}^c$ ($\bar{E}^c$ is open). Since p is not a limit point of E, there is a neighborhood N such that $N \cap E = \emptyset$. We claim that there also exists a N' such that $N' \cap \bar{E} = \emptyset$.

Suppose $N \cap E' \neq \emptyset$. We know that N cannot intersect E. Therefore the only possible way is for N to be "tangent to" E. Therefore there exists a neighborhood that is a subset of $\bar{E}^c$. So $\bar{E}^c$ is open. Hence, $\bar{E}$ is closed. □

Finally, we will present the proof of Lemma 2.4.1 by professor Francis Su in his lecture, which is even more concise.

Proof. **(Professor Su's proof)** We will turn the theorem into a series of equivalent statements. First of all, if E is open, it means that all points $x \in E$ are interior points of E. This means that for all $x \in E$, there exists some neighborhood $N \subset E$. But $N \subset E$ also means that N is completely disjoined with E^c (In other words, $N \cap E^c = \emptyset$). Therefore, *x cannot be the limit points of E^c for all $x \in E$*. Hence, E^c contains all its limit points and is therefore closed. □

Theorem 2.4.5. *E is closed if and only if $E = \bar{E}$.*

Proof. If $E = \bar{E}$, then Thm.(2.4.4) implies that E is closed. If E is closed, then $E' \subset E$. By definition, we know that $\bar{E} = E$. □

Theorem 2.4.6. *If $E \subset F$ and F is closed, then $\bar{E} \subset F$.*

Proof. If p is a limit point of E, then it is also a limit point of F, and F contains all such points since F is closed. So $p \in F$. This means $E' \subset F$. So $\bar{E} \subset F$. Yay. □

Remark 2.4.1. What this theorem says is that $\bar{E}$ is the smallest closed set that contains E.

Theorem 2.4.7. *Let $\{E_\alpha\}$ be a (finite or infinite) collection of sets E_α, then*

$$\left(\bigcup_\alpha E_\alpha\right)^c = \bigcap_\alpha E_\alpha^c.$$

Proof. Take any $x \in \left(\bigcup_\alpha E_\alpha\right)^c$, then $x \notin E_\alpha$ for any E_α. This means, equivalently, that $x \in E_\alpha^c$ for all $\alpha \in A$. Therefore, x is in the intersection of all E_α^c. □

There are several relations regarding the union (intersection) of open (closed) sets worth noting, which we will now present.

Theorem 2.4.8. *(a) For any collection $\{G_\alpha\}$ of open sets, $\bigcup_\alpha G_\alpha$ is open.*
(b) For any collection $\{F_\alpha\}$ of closed sets, $\bigcap_\alpha F_\alpha$ is closed.
(c) For any finite collection $G_1, ..., G_n$ of open sets, $\bigcap_{i=1}^n G_i$ is open.
(d) For any finite collection $F_1, ..., F_n$ of closed sets, $\bigcup_{i=1}^n F_i$ is closed.

Proof. (a) Put $G = \bigcup_\alpha G_\alpha$. If $x \in G$, then $x \in G_\alpha$ for some α. Since x is an interior point of G_α, it is also an interior point of G. Therefore, G is open. This proves (a).

(b) By Thm.(2.4.7), we have

$$\left(\bigcap_\alpha F_\alpha\right)^c = \bigcup_\alpha F_\alpha^c. \tag{2.1}$$

and F_α^c is open, by Lemma 2.4.1. Hence (a) implies that Eq.(2.1) is open. Hence $\bigcap_\alpha F_\alpha$ is closed. This proves (b).

(c) Put $H = \bigcap_{i=1}^{n} G_i$. For any $x \in H$, there exists neighborhood N_i of x with radius r_i, such that $N_i \subset G_i$ for $i = 1, 2, ..., n$. Put

$$r = \min(r_1, ..., r_n),$$

and let N be the neighborhood of x with radius r. Then $N \subset G_i$ for $i = 1, 2, ..., n$. Hence, $N \subset H$, and H is therefore open.

(d) By taking the complement, we have

$$\left(\bigcup_{i=1}^{n} F_i\right)^c = \bigcap_{i=1}^{n} F_i^c.$$

Since F_i is closed, then F_i^c is open for $i = 1, 2, ...n$. From (c) we know that the intersection of finite open sets is open, so $\left(\bigcup_i F_i\right)^c$ is open. Hence $\bigcup_i F_i$ is closed. □

Remark 2.4.2. We need to say something more about (c),(d) in Thm.(2.4.8). Note that these two statement only holds for finitely many sets. For infinitely many sets, we claim:

(c) Take an infinite collection of open sets, $\{G_\alpha\}$. Their intersection $\bigcap_\alpha G_\alpha$ can be closed.
(d) Take an infinite collection of closed sets, $\{F_\alpha\}$. Their union $\bigcup_\alpha F_\alpha$ can be open.

For (c), consider the following example. Let

$$G_i = \left(-\frac{1}{i}, \frac{1}{i}\right).$$

So G_i are segments between (-1,1). We can see that

$$\bigcap_{i=1}^{\infty} G_i = \{0\},$$

which is closed.

Similarly for (d), consider

$$F_i = \left[-1 + \frac{1}{i}, 1 - \frac{1}{i}\right],$$

which is closed. The union of infinitely many of them is

$$\bigcup_{i=1}^{\infty} F_i = (-1, 1),$$

which is open. The reason why argument in (c) where we take the minimum of a series of radii fails in the infinite case is that the sequence of infinitely many radii doesn't necessarily have a minimum.

2.5 Compact Sets

Now we shall move to the topic on compactness. Compactness is a concept that may look unmotivated when you first encounter it. So it is worthwhile to provide some motivation for defining the concept of compactness.

We have been talking quite a lot about finite set. As you may know, finite sets have a lot of great properties, some of them are listed below.

- They are small sets. So they are bounded.
- They are closed sets.
- In $\mathbb{R}$, they contain their sup and inf.
- When operating finite sets, the operation will end.
- Finite sets have no limit points.

As you can see, there are many great properties of finite sets. However, in analysis, we are dealing with infinity (infinite sets). And as you will see later, *compact sets are the next best thing to being finite.*

Definition 2.5.1. An *open cover* of a set E in X is a collection $\{G_\alpha\}$ of open subsets of X such that $E \subset \bigcup_\alpha G_\alpha$.

Definition 2.5.2. The subcover of $\{G_\alpha\}$ is a subcollection of $\{G_\alpha\}$ that still covers E. Often denoted by $\{G_{\alpha_\gamma}\}$.

Example 2.5.1. In $\mathbb{R}$, the set $\left[\frac{1}{2}, 1\right)$ has cover $\{V_n\}_{n=3}^{\infty}$, where $V_n = \left(\frac{1}{n}, 1 - \frac{1}{n}\right)$. Another cover may be $\{(0, 2)\}$. Another possible cover is $\{W_x\}_{x \in [\frac{1}{2}, 1)}$, where $W_x = N_{1/10}(x)$. Note that in the first and the third example, you can throw some sets away but still cover the set $\left[\frac{1}{2}, 1\right)$. In fact, for the third example, you can throw many things away so that you have a finite subcover (see Su's lecture 11).

Example 2.5.2. $[0, 1]$ in $\mathbb{R}$ has the cover $\{V_n\} \cup \{W_0, W_1\}$. The notation is consistent with the previous exaple.

Now we move to the definition of compactness.

Definition 2.5.3. A set K is compact in X if every open cover of K contains a *finite* subcover.

So to say that K is not compact means there exists a open cover of K in X with no finite subcover.

Example 2.5.3. $\left[\frac{1}{2}, 1\right)$ is not compact because there is a open cover with no finite subcover. The cover $\{V_n\}_{n=3}^{\infty}$ has no finite subcover. Since $n \to \infty$ is required for $\{V_n\}$ to cover all the points in the interval.

Example 2.5.4. $\mathbb{Z}$ in $\mathbb{R}$ is not compact. To prove this, *all we need to do is find a open cover that has no finite subcover*. For example, you can find an open cover $\{N_\varepsilon(x) : x \in \mathbb{Z}, \varepsilon > 0\}$, which are balls covering individual integers. You cannot even remove a single one of them, otherwise $\mathbb{Z}$ cannot be covered.

Theorem 2.5.1. *Finite sets are compact.*

Proof. Consider an open cover $\{G_\alpha\}$ covering $x_1, x_2, ..., x_n$. For all x_i, we can choose G_{α_i} that covers x_i. Then $\{G_{\alpha_i}\}_{i=1}^n$ covers the set. □

Definition 2.5.4. A set K is bounded if $K \subset N_r(x)$ for some $x \in X$.

Theorem 2.5.2. *Compact sets are bounded.*

Proof. Take $\{N_1(x) : x \in K\}$, the balls of radius 1 centering around all points in K. It is clear that the set of all such balls is an open cover of K. Since K is compact, there must be a finite subcover, which we call $\{N_1(x_i)\}_{i=1}^n$. Now we need to find the number R such that $K \subset N_R(x_i)$. We let

$$R = \max_{1 \le i,j \le n} \left\{ d(x_i, x_j) \right\}.$$

So we consider all possible pair distances between two sample points x_i, x_j we chose. Since it is a finite subcover, then there are finitely many $x_1, x_2, ...x_n$. Therefore, the maximum R exists.

Now, we claim that any two points $a, b \in K$ cannot be more than $R + 2$ away from each other. Since $a, b \in K$, then there must be balls from the subcover, say $N_1(x_i)$ and $N_1(x_j)$ that covers points a, b. Consider the diagram below.

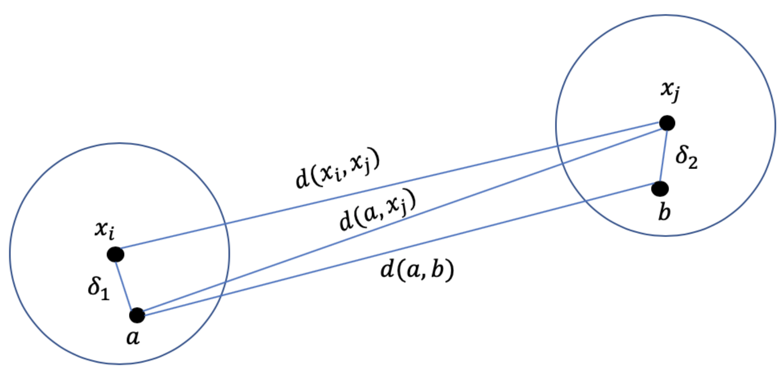


Figure 2.6: We consider the case where x_i, x_j, a, b forms a quadrilateral. The case where they form two triangles can be proved by a similar manner.

The quadrilateral with verticies x_i, x_j, a, b is in the metric space X. So the triangular inequality holds.

$$\begin{aligned} d(a, b) &\le d(a, x_j) + \delta_2 \\ &\le d(x_i, x_j) + \delta_1 + \delta_2 \\ &\le R + 1 + 1 = R + 2. \end{aligned}$$

Therefore, any two points $a, b \in K$ can be no further than $R + 2$. Then it is obvious that $N_{R+2}(x_i)$ contains K, or equivalently, $K \subset N_{R+2}(x_i)$. This concludes the proof. $\square$

Remark 2.5.1. To put it in more intuitive terms, Thm.2.5.2 is simply saying that if a set is compact, then it can't be too large. In some sense, this is the analogy of "finiteness" when we are talking about compact sets.

Now we move to the concept of relative open sets. As you see from Su's lecture, the concept of "openness" depends on which metric space you are in. Professor Su gave a very good example in his lecture 11.

Example 2.5.5. The set $(0, 1)$ is open in $\mathbb{R}$ but not open in $\mathbb{R}^2$.

Therefore, it is necessary to define the concept of relative open sets.

Definition 2.5.5. A set U is open in Y (*often called relative to* Y) if every point of U is an interior point of U. Here interior means using neighborhoods in Y.

Theorem 2.5.3. *Suppose $E \subset Y \subset X$, then E is open in Y if and only if $E = Y \cap G$ for some G that is open in X.*

Proof. **(By Professor Su)** Suppose G is open in X. Suppose $N_r(x) \subset G$, then $N_r(x) \cap Y$ is a neighborhood of x in Y (like the semicircle in the previous example). Since there exists such a neighborhood, x is an interior point of E.

Now suppose that E is open in Y. Then every point x has a neighborhood $N_{r_x}(x) \subset Y \cap E$. Then

$$G = \bigcup_{x \in E} N_{r_x}(x) \quad \text{in } X$$

is open, since by definition, every point x now is an interior point. And we can check immediately that $E = Y \cap G$. $\square$

Proof. **(by Rudin)** Suppose E is open relative to Y. To each point $p \in E$ there is a positive number r_p such that the conditions $d(p, q) < r_p$ and $q \in Y$ imply that $q \in E$. Let

$$V_p = \left\{ q \in X : d(p, q) < r_p \right\}.$$

We define

$$G = \bigcup_{p \in E} V_p.$$

Then it is clear that G is an open subset of X, by Thm.2.4.2 and Thm.2.4.8.

Now we prove that $E = Y \cap G$. Since $p \in V_p$ for all $p \in E$, we have $E \subset G \cap Y$. By our choice of V_p, we have $V_p \cap Y \subset E$ for every $p \in E$. Hence $G \cap Y \subset E$. Thus $E = G \cap Y$. We are halfway through.

Conversely, if G is open in X and $E = G \cap Y$, every $p \in E$ has a neighborhood $V_p \subset G$. Then $V_p \cap Y \subset E$, so E is open relative to Y. $\square$

2.6 Relationship of Compact Sets to Closed Sets

Previously we discussed the concept of relative openness. As you previously saw, the concept of openness depends on which metric space you are in. However, the concept of compactness doesn't depend on which metric space you are in, we now prove this theorem.

Theorem 2.6.1. *Suppose $K \subset Y \subset X$, then K is compact in X if and only if K is compact in Y.*

Proof. (i) Suppose K is compact in Y. We consider open cover $\{U_\alpha\}$ of K in X. Let $V_\alpha = U_\alpha \cap Y$. Then $\{V_\alpha\}$ covers K in Y. By compactness of K in Y, there exists a finite subcover in Y, which we call $\{V_{\alpha_1}, ..., V_{\alpha_n}\}$. Then $\{U_{\alpha_1}, ..., U_{\alpha_n}\}$ covers K in X, as desired.

(ii) Next, Suppose K is compact in X. Consider open covers $\{V_\alpha\}$ of K in Y. By Theorem 2.5, there exists U_α such that $U_\alpha \cap Y = V_\alpha$. So $\{U_\alpha\}$ covers K in X. By compactness in X, we have $\{U_{\alpha_1}, ..., U_{\alpha_n}\}$ covers K in X. Hence $\{V_{\alpha_1}, ..., V_{\alpha_n}\}$ covers K in Y. Therefore, K is compact in Y. □

Remark 2.6.1. As you can see, compactness is an intrinsic property of the set. It does not depend on which metric space it is in.

Now we shall prove an important fact: compact sets are closed.

Theorem 2.6.2. *Compact sets in a metric space are closed.*

Proof. Let K be a compact subset of a metric space X. We shall prove that K^c is open. Equivalently, we consider any point $p \in X$ and $p \notin K$. We want to prove that p has a neighborhood that does not intersect K.

For all $q \in K$, let $V_q = N_{r/2}(q)$, and let $U_q = N_{r/2}(p)$, where $r < \frac{1}{2}d(p,q)$. Then we notice that $\{V_q\}$ covers K. By compactness of K, there exists a finite subcover $\{V_{q_1}, ..., V_{q_n}\}$. In other words, Since K is compact, there are finitely many points $q_1, q_2, ...q_n$ such that

$$K \subset V_{q_1} \cup V_{q_2} \cup ... \cup V_{q_n}.$$

Next, we let

$$W = U_{q_1} \cap U_{q_2} ... \cap U_{q_n}.$$

We can see that W is a ball around p whose radius is $\min d(p, q_i)/2$. We claim that $W \cap V_{q_i} = \emptyset$ for all i, since $W \subset U_{q_i}$ and $U_{q_i} \cap V_{q_i} = \emptyset$. Therefore, p is an interior point of K^c. Hence K^c is open, then K is closed. □

Remark 2.6.2. In the proof, we utilized the concept of compactness and we conclude that there exists *finitely* many points $q_1, q_2, ..., q_n$ such that

$$K \subset \{V_{q_i}\}_{i=1}^{n}.$$

This is why we say compact is the next best thing to being *finite*. Because we can always make connections of compactness to finiteness (finite subcover).

Example 2.6.1. Now recall the last section when we say $(0, 1)$ is not compact. Last time we proved this statement by showing that not every open cover of $(0, 1)$ has a finite subcover. Now we can do it more quickly. Since it is not closed, it is not compact.

Example 2.6.2. Closed sets are not necessarily compact! For example, $\mathbb{R}$ in $\mathbb{R}$ is closed, but not compact since it is not bounded.

Theorem 2.6.3. *A closed subset B of a compact set K is also compact.*

Proof. Let $\{U_\alpha\}$ be a open cover of B. Notice that B^c is open since B is closed. So $\{U_\alpha\} \cup B^c$ is an open cover of K. By compactness of K, there exists a finite subcover of K, which is $\{U_{\alpha_1}, ..., U_{\alpha_n}, B^c\}$. However, since we know that B^c doesn't cover B, or $B^c \cap B = \emptyset$, so $\{U_{\alpha_1}, ..., U_{\alpha_n}\}$ covers B and is finite. Therefore, B is compact. $\square$

Corollary 2.6.1. *If F is closed and K is compact, then $F \cap K$ is compact.*

Proof. K compact means K is closed, then $F \cap K$ is a closed subset of the compact set K. Hence it is compact. $\square$

We conclude this section by setting up the preparation for the next section. We introduce a theorem that will be useful for our following discussion, which by itself is very interesting.

Theorem 2.6.4. *Let $I_n = [a_n, b_n]$ be a closed interval on $\mathbb{R}$, and let $\{I_n\}$ be a sequence of such intervals in $\mathbb{R}$, such that $I_{n+1} \subset I_n$, where $n \in \mathbb{N}$. Then $\bigcap_1^\infty I_n$ is not empty.*

Proof. Let $E = \{a_1, a_2, ...\}$ be the set of all a_n. Then E is nonempty and bounded by b_1. According to the least upper bound property of $\mathbb{R}$, we know that $x = \sup E$ exists.

If m, n are positive integers, then

$$a_n \leq a_{m+n} \leq b_{m+n} \leq b_n$$

by definition. It is clear that $x \leq b_m$ since b_m is an upper bound of E. Then we have $a_m \leq x \leq b_m$ for all $m \in \mathbb{N}$. Since $x \in I_m$ for $m = 1, 2, 3...$, we conclude the proof. $\square$

Professor Su used Thm.2.6.4 to prove that $\mathbb{R}$ is uncountable again. This proof is very elegant and worth mentioning.

Proof. Suppose $\mathbb{R}$ is countable. We list the elements of $\mathbb{R}$. For instance, $\mathbb{R} = \{x_1, x_2, x_3, x_4...\}$. Choose some closed interval I_1 that misses x_1. Then, choose I_2 such that $I_2 \subset I_1$ and I_2 misses x_2. Choose some interval I_3 such that $I_3 \subset I_2$ and I_3 misses x_3...... (Illustrated in Fig.2.7). From Thm.2.6.4, we know that $\bigcap_1^\infty I_n$ is not empty, so there is some element in our union that is not on the list $\{x_1, x_2, x_3, x_4...\}$. Hence $\mathbb{R}$ is uncountable. $\square$

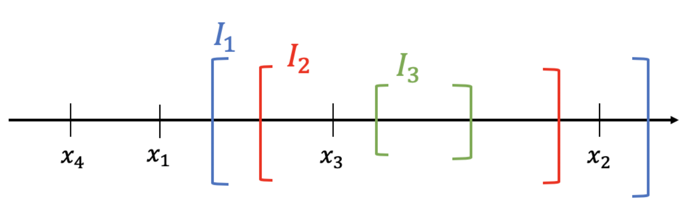


Figure 2.7: The construction can be like this picture.

Definition 2.6.1. Let $a_1 \leq b_1, ..., a_k \leq b_k$ be real numbers, then the set of all points $\mathbf{x} = (x_1, x_2, ..., x_k)$ in $\mathbb{R}^k$ such that $a_i \leq x_i \leq b_i$ for $i = 1, ..., k$ is a k-cell. So a k-cell is

$$[a_1, b_1] \times [a_2, b_2] \times ... \times [a_k, b_k] .$$

Corollary 2.6.2. *Let k be a positive integer. If $\{I_n\}$ is a sequence of k-cells such that $I_n \supset I_{n+1}$, for $n = 1, 2, 3...$, then $\bigcap_1^\infty I_n$ is not empty.*

Proof. We leave the detailed proof to Rudin (pg.39). The key idea is that since Thm.2.6.4 is true on every axis of $\mathbb{R}^k$, then there exists a point $\mathbf{x}^* = (x_1^*, x_2^*, ..., x_k^*)$ that is in I_n for all $n = 1, 2, 3...$ □

2.7 Compactness and Heine-Borel Theorem

This is Professor Su's last lecture on compactness. Today we are going to show that the closed intervals are compact, and closed and bounded sets in $\mathbb{R}^k$ is compact, which is the Heine Borel Theorem. But first we recall what have we have done right now. And let's reflect on why *being compact is the next best thing of being finite.*

- Compact sets are bounded (Thm.2.5.2).
- Compactness is an intrinsic property of a set (Thm.2.6.1). Compare this to open set, which is dependent on your metric space.
- Compact sets are closed (Thm.2.6.2).

Compare these bullets to that of a finite set, which is also bounded and closed.

Theorem 2.7.1. *A closed interval $[a, b]$ in $\mathbb{R}$ is compact.*

Proof. (by contradiction) Suppose $[a, b]$ is not compact. This means that there is an open cover $\{G_\alpha\}$ that has no finite subcover. Put $c_1 = (a + b)/2$. Then it is obvious that $\{G_\alpha\}$ covers both $[a, c_1]$ and $[c_1, b]$. We claim that *at least one* of these two intervals has no finite subcover(if both have finite subcover, I can just take the union of their finite subcover, therefore $[a, b]$ has a finite subcover, which is inconsistent with our assumption).

Without loss of generality, let $I_1 = [a, c_1]$ be the interval with no finite subcover. Put $c_2 = (a + c_1)/2$. Then again at least one of $[a, c_2]$ and $[c_2, c_1]$ has no finite subcover. We call that interval I_2. repeat the process and we have a sequence of intervals

$$I_1 \supset I_2 \supset I_3 \supset I_4...$$

We make two observation here about these intervals:

(i) Their lengths are decreasing by half;

(ii) None of them has a finite subcover.

Now, according to Thm.2.6.4, there exists a point $x \in I_n$ for all n. We know that there exists some cover in G_α that covers this point x, which we call G_{α_x}. Then since G_{α_x} is open, x is an interior point of G_{α_x}. Hence, there exists a neighborhood $N_\varepsilon(x) \subset G_{\alpha_x}$ for some $\varepsilon > 0$. By observation (i), we know that I_n is decreasing by half. Therefore, eventually there will be some $I_n \subset N_\varepsilon(x) \subset G_{\alpha_x}$ for sufficiently big n. Therefore, I_n has a finite subcover, which contradicts observation (ii).

Due to this contradiction, we are forced to conclude that our assumption that $[a, b]$ is not compact cannot be correct. □

Remark 2.7.1. This proof is beautiful! And surprisingly, it is very nontrivial to show that $[a, b]$ is compact – You may think since it is closed, its compactness is self-explanatory. For the picture of this proof, see Su's lecture. It is very helpful for you to totally understand this proof.

What we did above is to prove a weaker version of the next theorem, which we will now prove. The basic idea is similar.

Theorem 2.7.2. *Every k-cell is compact.*

Proof. Let I be a k-cell, consisting all points such that $\mathbf{x} = (x_1, x_2, ..., x_k)$ such that $a_j \leq x_j \leq b_j$ for $1 \leq j \leq k$. Put

$$\delta = \left\{ \sum_1^k (b_j - a_j)^2 \right\}^{1/2}.$$

Then $|\mathbf{x} - \mathbf{y}| < \delta$ if $\mathbf{x}, \mathbf{y} \in I$.

We again prove by contradiction. Suppose there exists an open cover $\{G_\alpha\}$ of I which contains no finite subcover. Again put

$$c_j = \frac{a_j + b_j}{2}.$$

Then we devide I into 2^k k-cells, and at least one of them has no finite subcover, which we call I_1. Repeat the process.

By Cor.2.6.2, we know there exists a point $\mathbf{x}^*$ which lies in every I_n. Then for some α, $\mathbf{x}^* \in G_\alpha$. Since G_α is open, there exists $\varepsilon > 0$ such that $|\mathbf{y} - \mathbf{x}^*| < \varepsilon$ implies that $\mathbf{y} \in G_\alpha$. For sufficiently large n such that $2^{-n}\delta < \varepsilon$, we have $I_n \subset G_\alpha$. A contradiction, as desired. □

Now we move to the key theorem of this section.

Theorem 2.7.3. ***(weaker Heine-Borel Theorem)*** *In $\mathbb{R}$, K is compact if and only if K is closed and bounded.*

Proof. First assume that K is compact, we already showed that K must be closed (Thm.2.6.2) and bounded (Thm.2.5.2).

Then assume that K is bounded and closed. Since K is bounded, it lives in some sufficiently big balls $(-r, r)$ for some $r > 0$. Then it is obvious that $K \subset [-r, r]$. But since $[-r, r]$ is compact, K is a closed subset of a compact set and thus is compact. □

Remark 2.7.2. Note that the statement "If K is closed and bounded, then it is compact" is *not* true in arbitrary metric spaces. The converse is always true in any metric spaces.

Now we prove the stronger version of Heine-Borel Theorem. The idea is almost the same.

Theorem 2.7.4. ***(Heine-Borel Theorem in*** $\mathbb{R}^k$***)*** *If* $E \subset \mathbb{R}^k$*, then* E *is compact if and only if* E *is closed and bounded.*

Proof. Suppose E is compact, proved already. Now suppose E is closed and bounded. Then there exists a big enough ball N such that $N \supset E$. Then there is a sufficiently large k-cell I such that $I \supset E$ for some $I \subset \mathbb{R}^k$. Since E is a closed subset of a compact set I, it is also compact. □

Now we are going to enter some really hard proofs! Put on your thinking cap, like they say.

Theorem 2.7.5. *In* $\mathbb{R}^k$*, a set* K *is compact if and only if every infinite subset* E *of* K *has a limit point in* K*.*

Proof. **(Warning: Long and Hard)** First, assume that K is compact. We will prove by contradiction. If no point in K were a limit point of E, then for all $q \in K$, its neighborhood V_q either contains no point of E ($q \notin E$) or one point of E (namely, contains only q if $q \in E$). Since for all $q \in K$, V_q contains at most one point of E, it is clear that no finite subcover of $\{V_q\}$ covers E (since E contains infinitely many points) and therefore K. But this contradicts the assumption that K is compact. Therefore, every finite subset E of K has a limit point in K.

Conversely, we assume that every infinite subset E of K has a limit point in K. We'll show that K is bounded and closed, therefore compact (Thm.2.7.4).

(i) Prove by contradiction. Suppose K is not bounded. Then K contains a infinite subset (sequence)

$$E = \{\mathbf{x_n} : |\mathbf{x_n}| > n, n \in \mathbb{N}\}.$$

Claim. *E has no limit points in* $\mathbb{R}^k$*, and therefore in* K*, which contradicts the assumption.*

PROOF. Suppose $q \in \mathbb{R}^k$ is a limit point of E. Then q must be bounded between two neighborhoods of the origin O, say with radii n and $n+1$. Then by the definition of a limit point, q has a neighborhood $N_r(q)$ that contains *infinitely* many points of E for all $r > 0$. However, given a fixed $r > 0$, there are only finitely many points between $N_{n-r}(O) \cap N_{n+1+r}(O)$ (shown in Fig.2.8).

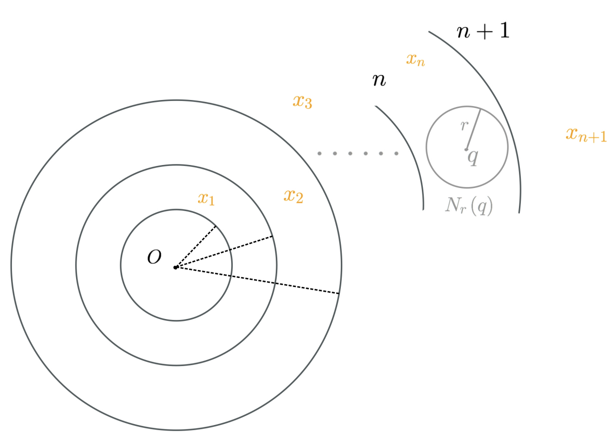


Figure 2.8: $N_r(q)$ only contains finite points in E.

According to Thm.2.4.1, q cannot be a limit point of E. We reached the desired contradiction. Therefore E is bounded. Q.E.D.

(ii) Suppose K is not closed. Then there exists $\mathbf{x_0} \notin K$ such that $\mathbf{x_0}$ is a limit point of K. By definition of the limit points of K, there exists a point $\mathbf{x_n} \in K$ such that $|\mathbf{x_n} - \mathbf{x_0}| < r$ for all $r > 0$. We can let $r = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$. Here is an example.

EXAMPLE. Take $n = 100$ and $r = \frac{1}{100}$, there exists a point $\mathbf{x_{100}} \in K$ such that $|\mathbf{x_{100}} - \mathbf{x_0}| < \frac{1}{100}$ since $\mathbf{x_0}$ is a limit point of K.

Now we define E to be an infinite subset of K by

$$E = \left\{\mathbf{x_n} : |\mathbf{x_n} - \mathbf{x_0}| < \frac{1}{n}, n \in \mathbb{N}\right\}.$$

By our hypothesis, E has a limit point $\mathbf{x}_\infty \in K$.

Claim. $\mathbf{x}_\infty = \mathbf{x_0}$. *Hence* $\mathbf{x_0} \in K$. *This contradicts our assumption that* $\mathbf{x_0} \notin K$ *and hence* K *is closed.*

PROOF. Since $\mathbf{x}_\infty$ is a limit point of E, there exists $m \geq n$ such that $|\mathbf{x_m} - \mathbf{x}_\infty| < \frac{1}{n}$ for all $n \in \mathbb{N}$. On the other hand

$$|\mathbf{x_m} - \mathbf{x_0}| < \frac{1}{m} \leq \frac{1}{n}$$

by our definition of the sequence E, Hence the triangle inequality holds.

$$\begin{aligned}|\mathbf{x_0} - \mathbf{x}_\infty| &\leq |\mathbf{x_m} - \mathbf{x_0}| + |\mathbf{x_m} - \mathbf{x}_\infty| \\ &\leq \frac{1}{n} + \frac{1}{n} = \frac{2}{n}\end{aligned}$$

This holds for all $n \in \mathbb{N}$, even for sufficiently large n. Therefore, we are forced to conclude that $|\mathbf{x_0} - \mathbf{x}_\infty| = 0$ and $\mathbf{x_0} = \mathbf{x}_\infty$. Q.E.D.

Since K is both bounded and closed in $\mathbb{R}^k$, Thm.2.7.4 says that K is compact. And finally, we are done. □

Remark 2.7.3. This theorem is true for *any* metric space X, not just $\mathbb{R}^k$. However, to prove that it is true for any metric space is even harder than proving it is true in $\mathbb{R}^k$. Therefore, we omit the proof for the general case.

Now, since we proved two theorem with "if and only if", we simply collect them altogether like Rudin does.

Theorem 2.7.6. *If a set E in $\mathbb{R}^k$ has one of the following three properties, then it has the other two as well:*
(a) E is closed and bounded.
(b) E is compact.
(c) Every infinite subset of E has a limit point in E.

Proof. We have shown that (a) $\iff$ (b) $\iff$ (c). So we are done. Another way is that you can prove that (a) implies (b), (b) implies (c), and (c) implies (a). It wouldn't be much easier than what we just did. You can refer to Rudin for this. □

Corollary 2.7.1. ***(Bolzano-Weierstrass Theorem)*** *Every bounded infinite subset of $\mathbb{R}^k$ has a limit point in $\mathbb{R}^k$.*

Proof. Being bounded, the set E is a subset of a k-cell $I \subset \mathbb{R}^k$. By Thm.2.7.2, I is compact, and so E has a limit point in I by Thm.2.7.5. □

Now we present a theorem which Professor.Su showed in the very end of his lecture. This is a generalization of Thm.2.6.4.

Theorem 2.7.7. ***(Cantor, Finite Intersection Property)*** *Let $\{K_\alpha\}$ be a collection of compact subsets in X. If any finite subcover of $\{K_\alpha\}$ has nonempty intersection, then $\bigcap_\alpha K_\alpha$ is nonempty.*

Proof. Fix a member of K of $\{K_\alpha\}$ and put $G_\alpha = K_\alpha^c$. We prove by contradiction. That is, assume that no point of K belongs to every K_α. Then every point of K is in some G_α. Then $\{G_\alpha\}$ form an open cover of K. By compactness of K, there is a finite subcover $\{G_{\alpha_1}, ..., G_{\alpha_n}\}$ such that

$$K \subset G_{\alpha_1} \cup G_{\alpha_2} \cup ... \cup G_{\alpha_n}.$$

Since for any point $q \in K$ we have $q \in K_{\alpha_i}^c$ for some $1 \leq i \leq n$, then $q \notin K_{\alpha_i}$ for that particular i. Therefore,

$$K \cap K_{\alpha_1} \cap K_{\alpha_2} \cap ... \cap K_{\alpha_n} = \emptyset.$$

Hence we reached a contradiction, as desired. □

Corollary 2.7.2. *If $\{K_n\}$ us a sequence of nonempty compact sets such that $K_n \supset K_{n+1}$ for $n = 1, 2, 3...$ then $\bigcap_1^\infty K_n \neq \emptyset$.*

We conclude this section by one final theorem.

Theorem 2.7.8. ***(Baby Tychonohoff Theorem)*** *Given two compact sets A, B, their product $A \times B$ is also compact.*

Proof. Okay too hard... □

2.8 Perfect Sets, Cantor Set, Connected Sets

2.8.1 Perfect Sets

Definition 2.8.1. A set E is *perfect* if E is closed and every point of E is a limit point of E.

Theorem 2.8.1. *Let P be a nonempty perfect set in $\mathbb{R}^k$. Then P is uncountable.*

Proof. Since P has limit points, P must be infinite. We prove by contradiction. Suppose P is countable, and denote the points of P by $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \ldots$ We will construct a sequence $\{V_n\}$ of neighborhoods as follows:

Let V_1 be any neighborhood of $\mathbf{x_1}$. If V_1 consists of all $\mathbf{y} \in \mathbb{R}^k$ such that $|\mathbf{y} - \mathbf{x_1}| < r$, the closure $\bar{V}_1$ of V_1 is the set of all $\mathbf{y} \in \mathbb{R}^k$ such that $|\mathbf{y} - \mathbf{x_1}| \leq r$.

Suppose V_n has already been constructed, so that $V_n \cap P \neq \emptyset$. Since every point of P is a limit point of P, there exists a neighborhood V_{n+1} such that:

(1) $\bar{V}_{n+1} \subset V_n$;

(2) $\mathbf{x_n} \notin \bar{V}_{n+1}$;

(3) $V_{n+1} \cap P \neq \emptyset$.

By (3), V_{n+1} satisfies the induction hypothesis, and the construction can proceed.

Put $K_n = \bar{V}_n \cap P$. Since $\bar{V}_n$ is closed and bounded by V_{n+1}, $\bar{V}_n$ is compact. Since $\mathbf{x_n} \notin K_{n+1}$, no point of P lies in $\bigcap_1^\infty K_n$. But since $K_n \subset P$, it can only contain points in P. Therefore, we are forced to conclude that $\bigcap_1^\infty K_n$ is empty.

But each $K_n \neq \emptyset$ (since $V_n \cap P \neq \emptyset$ so by definition $K_n = \bar{V}_n \cap P \neq \emptyset$) according to (3). We also have $K_n \supset K_{n+1}$ since

$$\bar{V}_{n+1} \subset V_n \Rightarrow \bar{V}_{n+1} \subset \bar{V}_n \Rightarrow \bar{V}_{n+1} \cap P \subset \bar{V}_n \cap P \Rightarrow K_{n+1} \subset K_n.$$

This contradicts Cor.2.7.2. □

Corollary 2.8.1. *Every interval $[a, b]$ where $a < b$ is uncountable. In particular, the set of all real numbers is uncountable.*

2.8.2 The Cantor Set

Definition 2.8.2. The Cantor ternary set is constructed by iteratively remove the middle third from a set of line segments. One starts by deleting the open middle third $\left(\frac{1}{3}, \frac{2}{3}\right)$ from the interval $[0, 1]$. leaving two line segments

$$E_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

Next, the open middle third of each of these remaining segment is deleted, leaving four line segments

$$E_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

The process continues infinitely. We define the Cantor ternary set to be

$$P = \bigcap_{n=1}^{\infty} E_n.$$

Figure 2.9: The construction of a Cantor set.

We immediately observe two property of the Cantor set P:
(a) $E_1 \supset E_2 \supset E_3 \supset ...$;
(b) E_n is the union of 2^n intervals, each with length 3^{-n}.

Since P is clearly a closed subset of the compact set $[0, 1]$, it is also compact. And Thm.2.7.7 shows that P is not empty. Besides these two properties, there are some other interesting properties of P.

Theorem 2.8.2. *P has a length of zero, which means that it contains no segments (intervals).*

Proof. We will prove this by showing that the complement of P relative to $[0, 1]$ has the length 1.

From the construction of P (Fig.2.9), we notice that at the n-th step, we are removing 2^{n-1} intervals, all of which are of length $\frac{1}{3^n}$. Therefore, the sum of the length of all intervals removed is

$$\sum_{n=1}^{\infty} 2^{n-1}\left(\frac{1}{3^n}\right) = \frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n = \frac{1}{3}\left(\frac{1}{1-\frac{2}{3}}\right) = 1$$

□

Remark 2.8.1. This theorem is actually saying that the Lebesgue measure of P is 0.

Theorem 2.8.3. *P is perfect.*

Proof. We show that any point $x \in P$ is a limit point of P.

Let $x \in P$, and let S be any segment containing x (this serves as neighborhood). Let I_n be that interval of E_n which contains x. Choose large enough n such that $I_n \subset S$. Let x_n be an end point of I_n such that $x_n \neq x$.

It follows from the construction of P that $x_n \in P$ (see Fig.2.9). Since $x_n \in S$, we conclude that x is a limit point of P. Therefore P is perfect. □

2.8.3 Connected Sets

Definition 2.8.3. Two subsets A and B of a metric space X is said to be *seperated* if both $A \cap \bar{B}$ and $\bar{A} \cap B$ are empty, i.e., if no point of A lies in the closure of B and no point of B lies in the colsure of A.

Definition 2.8.4. A set $E \subset X$ is said to be *connected* if E is *not* a union of two nonempty seperated sets.

Exercise 2.8.1. In $\mathbb{R}^2$, prove that the set $E = \{(x, y) : x, y \in \mathbb{Q}\}$ is not connected.

Solution 2.8.1. Consider the seperation $x = \pi$. There is no point of E that is on the line $x = \pi$. It is clear that this line splits E into two subsets of E whose union is E. Any point p on the left side of the line can't be a limit point of the right side because there is a ball around p that completely misses the line $x = \pi$, and thus the right hand side. Same for the right side.

Lemma 2.8.1. *Let E be a nonempty set of real numbers which is bounded above. Let $y = \sup E$. Then $y \in \bar{E}$. Hence $y \in E$ if E is closed.*

Proof. If $y \in E$ then $y \in \bar{E}$, then we are done. Now, assume that $y \notin E$. For every $h > 0$ there exists then a point $x \in E$ such that $y - h < x < y$. Otherwise $y - h$ would be an upper bound of E. Thus y is a limit point of E and hence in $\bar{E}$. □

Theorem 2.8.4. $[a, b]$ *is connected.*

Proof. (By contradiction) If not, there is a seperation A and B, with $a \in A$ without loss of generality.

Consider the rightmost point s of A, also known as $s = \sup A$. Then $s \in \bar{A}$ by Lemma 2.8.1. Therefore, $s \notin B$, and hence $s \in A$ because we only have two seperations. But if $s \in A$, then $s \notin \bar{B}$ by the definition of a seperated set. Then there exists a $(s - \varepsilon, s + \varepsilon)$ that contains no points in B. But since A, B are the only sets, then this interval $(s - \varepsilon, s + \varepsilon)$ must be in A. But this contradicts the assumption that $s = \sup A$. □

Chapter 3

Numerical Sequences and Series

An equation means nothing to me unless it expresses a thought of God.

-Srinivasa Ramanujan

3.1 Convergent Sequences

Definition 3.1.1. A sequence $\{p_n\}$ in X is a function $f : \mathbb{N} \to X$ that maps n to a point p_n in X.

Definition 3.1.2. A sequence $\{p_n\}$ in a metric space X is said to *converge* if there is a point $p \in X$ with the following property: for every $\varepsilon > 0$ there is an integer N such that $n \geq N$ implies that $d(p_n, p) < \varepsilon$. Here d denotes the distance in X.

Remark 3.1.1. In this case we say $\{p_n\}$ converges to p, or p is a limit of $\{p_n\}$. Write $p_n \to p$ or

$$\lim_{n\to\infty} p_n = p.$$

If $\{p_n\}$ does not converge, it is said to diverge.

Definition 3.1.3. When we are discussing the sequence $\{p_n\}$ of *real numbers*, we define

$$d(p_n, p) = |p_n - p|.$$

Exercise 3.1.1. In $\mathbb{R}$, define the sequence

$$p_n = \frac{n+1}{n}.$$

Show that $p_n \to 1$.

HINT. We want to show that given an $\varepsilon > 0$, there is a integer N that works. Notice that

$$\left|\frac{n+1}{n} - 1\right| = \left|\frac{1}{n}\right| < \varepsilon$$

holds if
$$n > \frac{1}{\varepsilon}$$
for positive n and ε. Therefore, we can consider any $N > \frac{1}{\varepsilon}$. Also remember that N needs to be an integer.

Solution 3.1.1. Given $\varepsilon > 0$, choose
$$N = \lceil \frac{1}{\varepsilon} \rceil + 1.$$
Then for $n \geq N$ we have $n > \frac{1}{\varepsilon}$. Hence $\frac{1}{n} < \varepsilon$. So
$$|p_n - p| = \left| \frac{n+1}{n} - 1 \right| = \left| \frac{1}{n} \right| < \varepsilon. \qquad \blacksquare$$

Theorem 3.1.1. *If $p \in X$ and $q \in X$, and if $\{p_n\}$ converges to p and q, then $p = q$.*

Proof. (By contradiction) Assume $p_n \to p$ and $p_n \to q$. Let $\varepsilon = d(p,q)$. Then there exists N_p such that $n \geq N_p$ implies $d(p_n, p) < \frac{\varepsilon}{2}$. Also there exists N_q such that $n \geq N_q$ implies that $d(p_n, q) < \frac{\varepsilon}{2}$. Let
$$N = \max \left\{ N_p, N_q \right\}.$$
Then $n \geq N$ implies
$$\varepsilon = d(p,q) \leq d(p, p_n) + d(p_n, q) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$
So this says $\varepsilon < \varepsilon$. A contradiction. $\square$

Theorem 3.1.2. *If $\{p_n\}$ converges, then $\{p_n\}$ is bounded.*

Proof. Suppose $p_n \to p$, then there is an integer N such that $n > N$ implies $d(p_n, p) < 1$ (Choose a ball with radius 1). Put
$$r = \max \{1, d(p_1, p), ..., d(p_N, p)\}.$$
Then $d(p_n, p) < r$ for $n = 1, 2, 3, ...$ $\square$

Theorem 3.1.3. *If $E \subset X$ and if p is a limit point of E, then there is a sequence $\{p_n\}$ in E such that $p = \lim_{n\to\infty} p_n$.*

Proof. For each positive integer n, there is a point $p_n \in E$ such that $d(p_n, p) < 1/n$. Then by incessantly choosing balls with radius $\frac{1}{n}$ for $n = 1, 2, 3...$ you have a sequence $\{p_n\}$. Given $\varepsilon > 0$, choose N so that $N\varepsilon > 1$. If $n > N$, it follows that $d(p_n, p) < \varepsilon$. Hence $p_n \to p$. $\square$

Theorem 3.1.4. *$\{p_n\}$ converges to $p \in X$ if and only if every neighborhood of p contains p_n for all but finitely many n.*

Proof. Suppose $p_n \to p$. Let V be a neighborhood of p. For some $\varepsilon > 0$ the condition

$$d(p,q) < \varepsilon \text{ and } q \in X \Rightarrow q \in V.$$

corresponding to this ε, there exists an N such that

$$n \geq N \Rightarrow d(p_n, p) < \varepsilon.$$

And since $p_n \in X$, we have

$$n \geq N \Rightarrow p_n \in V.$$

Conversely, assume every neighborhood of p contains all but finitely many p_n. Fix $\varepsilon > 0$ and let

$$V = \{q \in X : d(p,q) < \varepsilon\}.$$

By our assumption, there exists N corresponding to this V such that

$$n \geq N \Rightarrow p_n \in V.$$

Thus $d(p_n, p) < \varepsilon$ if $n \geq N$. Hence $p_n \to p$. □

Theorem 3.1.5. ***(Limit Rules)*** *Suppose $\{s_n\}, \{t_n\}$ are sequences of complex number, and $\lim_{n\to\infty} s_n = s$, $\lim_{n\to\infty} t_n = t$. Then*
(a) $\lim_{n\to\infty}(s_n + t_n) = s + t$;
(b) $\lim_{n\to\infty} cs_n = cs$, $\lim_{n\to\infty}(c + s_n) = c + s$ for any number c;
(c) $\lim_{n\to\infty} s_n t_n = st$;
(d) $\lim_{n\to\infty} \frac{1}{s_n} = \frac{1}{s}$ provided $s_n \neq 0$ and $s \neq 0$.

Proof. (a) Given $\varepsilon > 0$, there exists integers N_1, N_2 such that

$$n \geq N_1 \Rightarrow |s - s_n| < \frac{\varepsilon}{2},$$

$$n \geq N_2 \Rightarrow |t - t_n| < \frac{\varepsilon}{2}.$$

Take $N = \max(N_1, N_2)$, then $n \geq N$ implies

$$|(s_n + t_n) - (s + t)| \leq |s_n - s| + |t_n - t| < \varepsilon.$$

Here we just make use of Thm.1.9.6 (f) and let $\mathbf{x} = s_n + t_n$, $\mathbf{y} = t_n - s$, $\mathbf{z} = s + t$. This proves (a). ■

(b) Trivial.■

(c) Here we use a clever identity

$$s_n t_n - st = (s_n - s)(t_n - t) + s(t_n - t) + t(s_n - s). \tag{3.1}$$

Given $\varepsilon > 0$, there exists integers N_1, N_2 such that

$$n \geq N_1 \Rightarrow |s - s_n| < \sqrt{\varepsilon},$$
$$n \geq N_2 \Rightarrow |t - t_n| < \sqrt{\varepsilon}.$$

Take $N = \max(N_1, N_2)$, then $n \geq N$ implies

$$|(s_n - s)(t_n - t)| < \varepsilon,$$

so that

$$\lim_{n\to\infty} (s_n - s)(t_n - t) = 0.$$

We now apply (a) and (b) to Eqn.3.1 and we get

$$\lim_{n\to\infty} (s_n t_n - st) = 0. \qquad \blacksquare$$

(d) Choose an integer m such that

$$n \geq m \Rightarrow |s_n - s| < \frac{1}{2}|s| \Rightarrow -\frac{1}{2}|s| + s < s_n < \frac{1}{2}|s| + s.$$

If $s = -|s|$, then $|s_n| > 3|s|/2$. If $s = |s|$, then $|s_n| > |s|/2$. Either way, you can verify that

$$|s_n| > \frac{1}{2}|s|.$$

Given $\varepsilon > 0$, there is an $N > m$ such that

$$n \geq N \Rightarrow \left|\frac{1}{s_n} - \frac{1}{s}\right| = \left|\frac{s_n - s}{s_n s}\right| < \frac{2}{|s|^2}|s_n - s| < \varepsilon.$$

$\square$

Now we discuss the limit operation in $\mathbb{R}^k$.

Theorem 3.1.6. *Suppose* $\mathbf{x_n} \in \mathbb{R}^k (n = 1, 2, 3...)$ *and* $\mathbf{x_n} = (\alpha_{1,n}, \ldots, \alpha_{k,n})$. *Then* $\mathbf{x_n}$ *converges to* $\mathbf{x} = (\alpha_1, ..., \alpha_k)$ *if and only if*

$$\lim_{n\to\infty} \alpha_{j,n} = \alpha_j \quad (1 \leq j \leq k).$$

Proof. First, assume that $\mathbf{x_n} \to \mathbf{x}$. Then by the definition of the norm (distance in $\mathbb{R}^k$) we have

$$\begin{aligned} |\mathbf{x_n} - \mathbf{x}| &= \left(\sum_{j=1}^{k} (\alpha_{j,n} - \alpha_j)^2\right)^{1/2} \\ &\geq \left[(\alpha_{j,n} - \alpha_j)^2\right]^{1/2} \\ &\geq |\alpha_{j,n} - \alpha_j|. \end{aligned}$$

Hence $|\alpha_{j,n} - \alpha_j| \le |\mathbf{x_n} - \mathbf{x}|$ immediately shows that $\lim_{n\to\infty} \alpha_{j,n} = \alpha_j$.

Next, if $\lim_{n\to\infty} \alpha_{j,n} = \alpha_j$ holds, then to each $\varepsilon > 0$ there corresponds an integer N such that $n \ge N$ implies

$$\left|\alpha_{j,n} - \alpha_j\right| < \frac{\varepsilon}{\sqrt{k}} \quad (1 \le j \le k).$$

Hence $n \ge N$ implies

$$|\mathbf{x}_n - \mathbf{x}| = \left\{\sum_{j=1}^{k} \left|\alpha_{j,n} - \alpha_j\right|^2\right\}^{1/2} < \varepsilon.$$

□

Theorem 3.1.7. *Suppose $\{\mathbf{x_n}\}, \{\mathbf{y_n}\}$ are sequences in $\mathbb{R}^k$, $\{\beta_n\}$ is a sequence of real numbers, and $\mathbf{x_n} \to \mathbf{x}, \mathbf{y_n} \to \mathbf{y}, \beta_n \to \beta$, then*

$$\lim_{n\to\infty} (\mathbf{x_n} + \mathbf{y_n}) = \mathbf{x} + \mathbf{y}, \quad \lim_{n\to\infty} \mathbf{x_n} \cdot \mathbf{y_n} = \mathbf{x} \cdot \mathbf{y}, \quad \lim_{n\to\infty} \beta_n \mathbf{x_n} = \beta \mathbf{x}.$$

Proof. This follows from Thm.3.1.5 and Thm.3.1.6. □

3.2 Subsequences

Definition 3.2.1. Given a sequence $\{p_n\}$, consider a sequence $\{n_k\}$ of positive ibntegers such that $n_1 < n_2 < n_3 < \ldots$. Then the sequence $\{p_{n_i}\}$ is called a subsequence of $\{p_n\}$. If $\{p_{n_i}\}$ converges, its limit is called a subsequential limit of $\{p_n\}$.

Theorem 3.2.1. *$\{p_n\}$ converges to p if and only if every subsequence of $\{p_n\}$ converges to p.*

Proof. trivial. □

Example 3.2.1. *A divergent sequence may have a convergent subsequence.* Consider the following sequence:

$$\left\{1, \pi, \frac{1}{2}, \pi, \frac{1}{3}, \pi, \frac{1}{4}, \pi \ldots\right\}.$$

This sequence is divergent, but it has a convergent subsequence, namely

$$\left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4} \ldots\right\}.$$

Example 3.2.2. *Not every sequence contains a convergent subsequence*. For example, the set of natural numbers $\mathbb{N}$ doesn't contain any convergent subsequence.

Example 3.2.3. *Not every bounded sequence contains a convergent subsequence.* For general metric spaces this is not true. Suppose my metric space is $\mathbb{Q}$, with the usual metric. The sequence

$$\{3, 3.1, 3.14, 3.141, 3.1415, 3.14159 \ldots\}$$

does not converge in $\mathbb{Q}$, and its subsequence does not converge in $\mathbb{Q}$. In $\mathbb{R}$, it will converge to π!

Definition 3.2.2. A metric space X is *sequentially compact* if every sequence in X has a convergent subsequence.

Theorem 3.2.2. *If $\{p_n\}$ is a sequence in a compact metric space X, then some subsequence of $\{p_n\}$ converges to a point of X (If X is compact, then X is sequentially compact).*

Proof. Let R be the range of the sequence $\{p_n\}$. If R is finite, then some p in $\{p_n\}$ is achieved infinitely many times. Use the subsequence

$$p_{n_1} = p_{n_2} = ... = p.$$

This sequence evidently converges to p.

If R is infinite, $\{p_n\}$ is a infinite subset of a compact set X. By Thm.2.7.5, since X is compact, R has a limit point p in X. The claim is that p then must be a subsequential limit of $\{p_n\}$. Consider the set

$$\left\{p_k : d(p_k, p) < \frac{1}{k}, k \in \mathbb{N}\right\}.$$

No doubt that this subsequence converges to p.

NOTE. We don't need to worry that k does not increase successively: That is, we don't need to worry that the first term of the subsequence is p_{101} and the second becomes p_{50}. Since every neighborhood of p contains infinitely many point of $\{p_n\}$, you can always find a next term in the subsequence whose index is bigger than the previous one, otherwise the neighborhood won't contain infinitely many points of $\{p_n\}$. □

Theorem 3.2.3. *Let E^* be the set of all subsequential limits of a sequence $\{p_n\}$ in X. Then E^* is a closed subset of X.*

Proof. Let q be a limit point of E^*. We want to show that $q \in E^*$.

Choose n_1 such that $p_{n_1} \neq q$ (If n_1 does not exists, then E^* contains only one point, namely q. So we are done). Put $\delta = d(p_{n_1}, q)$. Suppose $n_1, ..., n_{i-1}$ are chosen.

Since q is a limit point of E^*, there is a point $x \in E^*$ such that

$$d(x, q) < 2^{-i}\delta.$$

Since $x \in E^*$, there is an $n_i > n_{i-1}$ such that

$$d(x, p_{n_i}) < 2^{-i}\delta.$$

Thus

$$d(q, p_{n_i}) \leq d(q, x) + d(x, p_{n_i}) < 2^{-i}\delta + 2^{-i}\delta = 2^{1-i}\delta.$$

Thus $\{p_{n_i}\}$ converges to q. Hence $q \in E^*$. □

3.3 Cauchy Sequences and Complete Spaces

In the previous sections, we have been talking about the convergence of a sequence. But the definition of convergence really implies that we already know the limit of a sequence. Then we can proceed the "smaller than ε" proof. However, if we don't know whether the sequence has a limit, or what is the limit of that sequence, can we say whether it is convergent or not? This idea motivates the definition of a Cauchy sequence. The brief idea behind Cauchy sequence is that *if a sequence is convergent, the successive points in the sequence must be closer and closer.*

Definition 3.3.1. A sequence $\{p_n\}$ in a metric space X is said to be a *Cauchy sequence* if for every $\varepsilon > 0$ there is an integer N such that

$$m, n \geq N \Rightarrow d(p_n, p_m) < \varepsilon.$$

Theorem 3.3.1. *If $\{p_n\}$ converges, then $\{p_n\}$ is Cauchy.*

Proof. Given $\varepsilon > 0$, there is an integer N such that

$$n \geq N \Rightarrow d(p, p_n) < \frac{\varepsilon}{2}.$$

So for this N,

$$m, n \geq N \Rightarrow d(p_n, p_m) \leq d(p_n, p) + d(p, p_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Example 3.3.1. *Cauchy sequences are not necessarily convergent.* For example, consider the same sequence we mentioned last time

$$\{3, 3.1, 3.14, 3.141, 3.1415, 3.14159...\}.$$

This sequence is a Cauchy sequence in $\mathbb{Q}$ but does not converge in $\mathbb{Q}$. Always remember to consider the metric space you are in!

This example actually motivates us to consider when the converse of Thm.3.3.1 is true. Now we create a definition below.

Definition 3.3.2. A metric space X is *complete* if every Cauchy sequence converges to a point of X.

Example 3.3.2. $\mathbb{Q}$ is not complete.

Theorem 3.3.2. *Compact metric spaces are complete.*

Proof. Let $\{x_n\}$ be Cauchy. We want to show that $\{x_n\}$ converges.

By Thm.3.2.2 we know that since X is compact, it is sequentially compact. Therefore exists $\{x_{n_k}\}$ converging to $x \in X$. We claim that the entire sequence $\{x_n\}$ converges to x.

Fix $\varepsilon > 0$. There exists a N_1 such that

$$i, j \geq N \Rightarrow d(x_i, x_j) < \frac{\varepsilon}{2}.$$

Since $\{x_{n_k}\}$ converges to x, there exists a N_2 such that

$$n_k \geq N_2 \Rightarrow d(x_{n_k}, x) < \frac{\varepsilon}{2}.$$

Let

$$N = \max\{N_1, N_2\}.$$

We fix an x_{n_k} such that $n_k \geq N$. Then using the triangle inequality we have

$$n \geq N \Rightarrow d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Corollary 3.3.1. $[0, 1]$ *is complete; k-cells in* $\mathbb{R}^k$ *are complete; A closed subset of a compact set is compact, and therefore complete.*

Example 3.3.3. Cantor set is complete, since it is a closed subset of a compact set $[0, 1]$.

Theorem 3.3.3. $\mathbb{R}^k$ *is complete.*

Proof. We show that *If* $\{x_n\}$ *is Cauchy, it is bounded*. Therefore, $\{x_n\}$ lives in some k-cells. Then by Thm.3.3.2, $\{x_n\}$ is convergent. So $\mathbb{R}^k$ is complete.

Fix $\varepsilon > 0$, there is an integer N such that for all $m, n \geq N$, we have $d(x_m, x_n) < \varepsilon$. Then let

$$R = \max\{d(x_N, x_1), ..., d(x_N, x_{N-1}), \varepsilon\}.$$

Therefore this sequence $\{x_n\}$ is in some ball(neighborhood) $B_R(x_N)$. □

Remark 3.3.1. Since $\mathbb{R}^k$ is complete, then we don't need to know the limit of a sequence in $\mathbb{R}^k$ and we can still tell whether it is convergent. We just have to show it is Cauchy.

Exercise 3.3.1. Consider

$$x_n = 1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n} = \sum_{k=1}^{n} \frac{1}{k}.$$

Does the sequence $\{x_n\}$ converge? Be careful!

Solution 3.3.1. Let $n > m$. Then

$$|x_n - x_m| = \sum_{k=m+1}^{n} \frac{1}{k} \geq \frac{n-m}{n} = 1 - \frac{m}{n}.$$

This is simply because we have $n - m$ terms in the summation and each term is bigger than $1/n$ since $n > m$.

Consider the case when $n = 2m$. Then we have

$$|x_{2n} - x_n| > \frac{1}{2}$$

for all n. Therefore, this sequence has no hope to be Cauchy since the condition fails as soon as you make $\varepsilon < 1/2$. Hence this sequence $\{x_n\}$ *is not convergent*. It just grows really slow!

Exercise 3.3.2. Consider the sequence defined by recurrence relation

$$x_1 = 1 \qquad x_2 = 2 \qquad x_n = \frac{1}{2}(x_{n-1} - x_{n-2}).$$

So it takes average of the previous two terms. Does $\{x_n\}$ converge?

Solution 3.3.2. This example is interesting. At the first glance, we have no idea what the limit is. But it's a sequence of real numbers, so we don't need to know it's limit, we only need to know whether it is Cauchy!

You can draw a picture and see what this sequence looks like and you can very easily verify that

$$|x_n - x_{n-1}| = \frac{1}{2^{n-2}} = 2^{-(n-2)}.$$

It is also easy to show that

$$|x_n - x_m| < |x_m - x_{m-1}| = \frac{1}{2^{m-2}}.$$

Now, fix $\varepsilon > 0$. If we want $|x_n - x_m| < \varepsilon$, we can take

$$N = \lceil \log_2(\frac{1}{\varepsilon}) + 2 \rceil.$$

Then it is easy to verify that $n, m \geq N$ implies $|x_n - x_m| < \varepsilon$. Therefore, the sequence is Cauchy. Hence it is convergent.

Definition 3.3.3. A sequence $\{s_n\}$ of real numbers is said to be monotonically increasing if $s_n \leq s_{n+1}$. It is said to be monotonically decreasing if $s_n \geq s_{n+1}$.

Theorem 3.3.4. *Suppose $\{s_n\}$ is monotonic. Then $\{s_n\}$ converges if and only if it is bounded. More over, it converges to its* sup *if it is monotonically increasing, and to its* inf *if monotonically decreasing.*

Proof. By Thm.3.1.2, we know that if we assume $\{s_n\}$ converges, then it is bounded. Now we prove the other direction.

Suppose $\{s_n\}$ is bounded (decreasing case will be similar). Then let $s = \sup E$, where E is the range of $\{s_n\}$. Then $s_n \leq s$ for all $n = 1, 2, 3...$

For every $\varepsilon > 0$, there is an integer N such that

$$s - \varepsilon < s_N \leq s,$$

cause otherwise $s - \varepsilon$ would be an upper bound of E. Since $\{s_n\}$ incrases,

$$\begin{aligned} n \geq N &\Rightarrow s - \varepsilon < s_n \leq s \\ &\Rightarrow s - \varepsilon < s_n < s + \varepsilon \\ &\Rightarrow |s_n - s| < \varepsilon. \end{aligned}$$

This completes the proof. $\square$

WARNING. The following section is long and difficult. We have one last thing to say about Cauchy sequences, which is the title of the next section.

3.4 Completion of X

Exercise 3.4.1. *(Problem 3.24 in Rudin)* Let X be a metric space.
(a) Call two Cauchy sequences $\{p_n\}, \{q_n\}$ in X *equivalent* if

$$\lim_{n\to\infty} d(p_n, q_n) = 0.$$

Prove that this is an equivalence relation.
(b) Let X^* be the set of all equivalence classes so obtained. If $P \in X^*, Q \in X^*, \{p_n\} \in P, \{q_n\} \in Q$, define

$$\Delta(P, Q) = \lim_{n\to\infty} d(p_n, q_n).$$

Show that the number $\Delta(P, Q)$ is unchanged if $\{p_n\}$ and $\{q_n\}$ are replaced by equivalence sequences, and hence that Δ is a distance function in X^*.
(c) Prove that the resulting metric space X^* is complete.
(d) For each $p \in X$ there is a Cauchy sequence all of whose terms are p; Let P_p be the element of X^* which contains this sequence. Prove that

$$\Delta(P_p, P_q) = d(p, q)$$

for all $p, q \in X$. In other words, the mapping $\varphi(p) = P_p$ is an isometry (i.e. a distance preserving mapping) of X into X^*.
(e) Prove that $\varphi(X)$ is dense in X^*, and that $\varphi(X) = X^*$ if X is complete. By (d), we may identify X and $\varphi(X)$ and thus regard X as embedded in the complete metric space X^*. We call X^* the completion of X.

Solution 3.4.1. (a) It is obvious that $\{p_n\} \sim \{p_n\}$. Since we are in a metric space X, the distance function d is symmetric. Hence it is trivial that $\{p_n\} \sim \{q_n\} \Rightarrow \{q_n\} \sim \{p_n\}$. Finally, if $\{p_n\} \sim \{q_n\}$ and $\{q_n\} \sim \{k_n\}$, then by definition it says

$$n \geq N_1 \Rightarrow |p_n - q_n| < \frac{\varepsilon}{2} \quad \text{and} \quad n \geq N_2 \Rightarrow |q_n - k_n| < \frac{\varepsilon}{2}.$$

Hence by taking $N = \max\{N_1, N_2\}$ and using the triangle inequality we have

$$|p_n - k_n| < \varepsilon \Rightarrow \{p_n\} \sim \{k_n\}. \blacksquare$$

(b) Let $\{p'_n\} \sim \{p_n\}, \{q'_n\} \sim \{q_n\}$. We want to show that

$$\lim_{n\to\infty} d(p_n, q_n) = \lim_{n\to\infty} d(p'_n, q'_n).$$

By triangle inequality we have

$$|p'_n - q'_n| \leq |p'_n - p_n| + |p_n - q'_n| \leq |p'_n - p_n| + |q'_n - q_n| + |p_n - q_n|.$$

Taking the limit on both side we obtain

$$\lim_{n\to\infty} d(p'_n, q'_n) \leq \lim_{n\to\infty} d(p_n, q_n).$$

Using the triangle inequality again we have

$$|p_n - q_n| \le |p_n - p_n'| + |p_n' - q_n| \le |p_n - p_n'| + |q_n - q_n'| + |p_n' - q_n'|.$$

Taking the limit again we have

$$\lim_{n\to\infty} d(p_n, q_n) \le \lim_{n\to\infty} d(p_n', q_n').$$

Hence

$$\lim_{n\to\infty} d(p_n, q_n) = \lim_{n\to\infty} d(p_n', q_n'). \blacksquare$$

(c) Suppose $\{P_k\}$ is a Cauchy sequence in X^*. Choose Cauchy sequence $\{p_{kn}\} \in X$ such that $\{p_{kn}\} \in P_k$. We want to show that $\{P_k\}$ converges to some point $P \in X^*$.

Since $\{p_{kn}\}$ is Cauchy, for each k, choose N_k to be the first integer such that

$$m, n \ge N_k \Rightarrow d(p_{kn}, p_{km}) < 2^{-k}.$$

Let

$$p_k = p_{kN_k}.$$

Observe that for any $n \ge N_k$, we have

$$d(p_k, p_{kn}) < 2^{-k}.$$

Also, for any k, l, n we have

$$d(p_k, p_l) \le d(p_k, p_{kn}) + d(p_{kn}, p_{ln}) + d(p_{ln}, p_l).$$

Hence, by taking n sufficiently large and assuming $k < l$ we have

$$\begin{aligned} d(p_k, p_l) &\le d(p_k, p_{kn}) + d(p_{kn}, p_{ln}) + d(p_{ln}, p_l) \\ &\le 2^{-k} + \Delta(P_k, P_l) + 2^{-k} + 2^{-l} \\ &\le 3 \cdot 2^{-k} + \Delta(P_k, P_l). \end{aligned}$$

Note that in the second step we simply used the fact that for sufficiently large n,

$$|d(p_{kn}, p_{ln}) - \Delta(P_k, P_l)| < 2^{-k}.$$

This is true by the definition of Δ metric function.

CLAIM 1. *$\{p_k\}$ is Cauchy.*

PROOF OF CLAIM 1. Since we have

$$d(p_k, p_l) < 3 \cdot 2^{-k} + \Delta(P_k, P_l)$$

and since $\{P_k\}$ is Cauchy, then there exists N_2 such that

$$k, l \ge N_2 \Rightarrow \Delta(P_k, P_l) < \frac{\varepsilon}{2}.$$

It is obvious that there is a N_3 such that

$$k \geq N_3 \Rightarrow 3 \cdot 2^{-k} < \frac{\varepsilon}{2}.$$

Take $k > \max\{N_2, N_3\}$ and we have

$$d(p_k, p_l) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, $\{p_k\}$ is Cauchy. ■

Now, let $P \in X^*$ be the element that contains the sequence $\{p_k\}$.

CLAIM 2. *$P_k \to P$ in X^*.*

PROOF OF CLAIM 2.

$$\begin{aligned}
\Delta(P_k, P) &= \lim_{n\to\infty}(d(p_{kn}, p_n)) \\
&\leq \lim_{n\to\infty}(d(p_{kn}, p_k) + d(p_k, p_n)) \\
&\leq 2^{-k} + \lim_{n\to\infty} \sup \Delta(P_k, P_n) + 3 \cdot 2^{-k}
\end{aligned}$$

Now, given $\varepsilon > 0$, we want $2^{-k+2} < \varepsilon/2$ and $\Delta(P_k, P_n) < \varepsilon/2$. For the former, take

$$k > N_1 = 3 - \lfloor \log_2(\varepsilon) \rfloor.$$

You can easily verify that the first inequality is satisfied. For the later, since we are taking the limit as $n \to \infty$, n is already sufficiently large. Take $k > N_2$ such that

$$k > N_2 \Rightarrow \Delta(P_k, P_n) < \frac{\varepsilon}{2}.$$

Now we claim that if

$$k > N = \max\{N_1, N_2\} \Rightarrow \Delta(P_k, P) < \varepsilon. \blacksquare$$

Now we know that given any Cauchy sequence $\{P_k\} \in X^*$, it is convergent. Therefore, X^* is complete.

(d)

$$\Delta(P_p, P_q) = \lim_{n\to\infty} d(p_n, q_n) = d(p, q)$$

since $p_n = p, q_n = q$ for all n. Hence $\varphi(p) = P_p$ is an isometric mapping from X to X^*. ■

(e) Let P be any element in X^*. To prove that $\varphi(X)$ is dense in X^*, all we need to do is to find a point $p \in X$ such that $\Delta(P, P_p) < \varepsilon$.

Then set $\varepsilon = \Delta(P_1, P_2)/2$, and take $\bar{P} = P_1 + \Delta(P_1, P_2)/2$ (Suppose $P_1 < P_2$). Then We can find a point P_p such that $\Delta(\bar{P}, P_p) < \varepsilon$. Then we know for any $P_1, P_2 \in X^*$ there will be a $P_p = \varphi(p) \in X$ inbetween.

To do this, let $\{p_n\} \in P$ and choose N such that

$$m, n \geq N \Rightarrow d(p_n, p_m) < \frac{\varepsilon}{2}.$$

Now let

$$p = p_{N+1}.$$

Then

$$\Delta(P, P_p) = \lim_{n\to\infty} d(p_n, p) < \frac{\varepsilon}{2} < \varepsilon.$$

Now suppose X is complete. Then for each $P \in X^*$ and $\{p_n\} \in X$ there exists a $p \in X$ such that $p_n \to p$. Obviously, this p is the same for all the sequences equivalent to $\{p_n\}$, by definition. Since we know that $\{p, p, p...\} \sim \{p_n\}$, then $\varphi(p) = P_p = P = \varphi(\{p_n\})$. Hence $P_p = P$. We conclude that $\varphi(X) = X^*$ when X is complete. ■

This exercise problem answers the following question: *If X is not complete, can it be embedded in one that is*? The answer is yes, and it is stated in the following theorem, which can be extracted without effort from the exercise above.

Theorem 3.4.1. *Every metric space (X, d) has a completion (X^*, d).*

Remark 3.4.1. This theorem says that there is a way to define a big space X^* that has the smaller space X as a subset such that the metric d on the big space is preserved when restricted to the smaller space X.

We go back to the exercise above, and state the following Theorem, which is exactly the same as part (c)-(d).

Theorem 3.4.2. *X^* is complete with X isometrically embedded in X^*.*

This is, in fact, another way to construct $\mathbb{R}$ from $\mathbb{Q}$. For proofs on X^* satisfying the field axioms, and other subtle detailds, you can see the construction demonstrated in *Real Analysis I, Terrence Tao*.

3.5 Upper and Lower Limits

Definition 3.5.1. We write $s_n \to +\infty$ if for every real M there is an integer such that $n \geq N$ implies $s_n \geq M$.

Similarly, We write $s_n \to -\infty$ if for every real M there is an integer such that $n \geq N$ implies $s_n \leq M$.

Definition 3.5.2. Given $\{s_n\}$, let E be the set of all possible subsequential limits that $\{s_n\}$ may has, including possibly $\pm\infty$. Let

$$s^* = \sup E \qquad s_* = \inf E.$$

The numbers s^*, s_* are called *upper* and *lower* limits. We use the notation

$$\lim_{n\to\infty} \sup s_n = s^* \qquad \lim_{n\to\infty} \inf s_n = s_*.$$

Theorem 3.5.1. *It should be immediately obvious that*

$$\lim_{n\to\infty} \inf s_n \leq \lim_{n\to\infty} \sup s_n.$$

Theorem 3.5.2. *For a real-values sequence $\{s_n\}$, $s_n \to s$ if and only if*

$$\lim_{n\to\infty} \sup s_n = \lim_{n\to\infty} \inf s_n = s.$$

Theorem 3.5.3. *Let $\{s_n\}$ be a sequence of real numbers. Let E and s^* have the same meaning as in the definition above. Then s^* has the following two properties:*
(a) $s^* \in E$.
(b) If $x > s^*$, there is an integer N such that $n \geq N \Rightarrow s_n < x$.
(c) s^* is the only number with the properties (a) and (b).
The analogous result is true for s_*.

Proof. (a) If $s^* = +\infty$, then E is not bounded above, and there is a subsequence $\{s_{nk}\}$ such that $s_{nk} \to \infty$.

If s^* is a real number, then at least one subsequential limit exists (The sup of an empty set is $-\infty$). Then by Thm.3.2.3, E is a closed subset of X. Then by Lemma.2.8.1, $s^* \in \bar{E}$. Since E is closed, then $s^* \in E$.

If $s^* = -\infty$, then E contains only one element, namely $-\infty$. Then there is no subsequential limit. So that $s_n \to -\infty$.

(b) We prove by contradiction. Suppose there is a number $x > s^*$ such that $s_n \geq x$ for infinitely many n. Then we can find a s_{n1} such that $s_{n1} \geq x > s^*$. Put $x = s_{n1}$. Again, there is some s_{n2} such that $s_{n2} \geq s_{n_1} \geq x > s^*$. Let $x = s_{n2}$ and repeat this process. Then the subsequence $\{s_{nk}\}$ either has a subsequential limit that is bigger than s^*, or goes to infinitity. Either way, this violates the definition of s^* and we have therefore a contradiction.

(c) To show the uniqueness, suppose there are two numbers p and q that both satisfy properties (a) and (b). Suppose $p < q$. Choose x such that $p < x < q$. Since p satisfies (b), we have $s_n < x$ for all $n \geq N$. But then q cannot be a subsequential limit and therefore cannot sastisfy (a). □

Corollary 3.5.1. *If $s_* > x$, there exists an integer N such that*

$$n \geq N \Rightarrow s_n > x.$$

Proof. This proof is similar to that of (b) in the theorem above. □

3.6 Series and Series of Nonnegative Terms

Definition 3.6.1. Give $\{a_n\}$, we associate this sequence with another sequence $\{s_n\}$, where

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

For s_n we also use the symbolic expression

$$\sum_{n=1}^{\infty} a_n, \tag{3.2}$$

which we call an *infinite series* or just a *series*. If $\{s_n\}$ converges to s, we say the series converges and write

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

It should be understood that *s is the limit of a sequence of sums, and is not obtained simply by addition.* If $\{s_n\}$ diverges, we say the series diverges. Sometimes for convenience of notation, we will consider series of the form

$$\sum_{n=0}^{\infty} a_n. \tag{3.3}$$

When there is no possible ambiguity, we sometimes simply write $\sum a_n$ in place of (3.2) and (3.3).

It is clear that every theorem about sequence can be restated in terms of series, and vice versa. For example, the Cauchy criterion can be restated in the following form:

Theorem 3.6.1. *$\sum a_n$ converges if and only if for every $\varepsilon > 0$ there is an integer N such that*

$$m \geq n \geq N \Rightarrow \left| \sum_{k=n}^{m} a_k \right| \leq \varepsilon. \tag{3.4}$$

Corollary 3.6.1. *In particular, by taking $m = n$, (3.4) becomes*

$$n \geq N \Rightarrow |a_n| \leq \varepsilon.$$

In other words, if $\sum a_n$ converges, then

$$\lim_{n\to\infty} a_n = 0.$$

Remark 3.6.1. The converse of Cor.3.6.1 is not necessarily true. For example, consider the harmonic series

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

This series is divergent even though

$$\left| \frac{1}{n} \right| \leq \varepsilon$$

for large enough n.

Theorem 3.6.2. *A series $\sum a_n$ that has nonnegative terms $a_1, a_2, a_3, \ldots \geq 0$ converges if and only if its partial sums form a bounded sequence.*

Proof. Since the series has nonnegative terms, its partial sums are monotonically increasing, in other words, $s_{n+1} \geq s_n$ for all n. Thm.(3.3.4) says monotonic sequence converges if bounded. Therefore the proof is complete. □

Theorem 3.6.3. ***(The Comparison Test)***
(a) If $|a_n| \leq c_n$ for $n \geq N_0$, where N_0 is some fixed integer, and if $\sum c_n$ converges, then $\sum a_n$ converges.
(b) If $a_n \geq d_n \geq 0$ for $n \geq N_0$, and if $\sum d_n$ diverges, then $\sum a_n$ diverges.

Proof. (a) Given $\varepsilon > 0$, There exists $N \geq N_0$ such that

$$m \geq n \geq N \Rightarrow \sum_{k=n}^{m} c_k \leq \varepsilon.$$

By the Cauchy criterion. Hence

$$\left| \sum_{k=n}^{m} a_k \right| \leq \sum_{k=n}^{m} |a_n| \leq \sum_{k=n}^{m} c_k \leq \varepsilon.$$

(b) This follows immediately from (a). If $\sum a_n$ converges, so must $\sum d_n$. □

The comparison test is very useful. But to use it efficiently, we have to become familiar with a number of series of nonnegative terms whose convergence or divergence is known. The simplest of all is perhaps the geometric series.

Theorem 3.6.4. *If $|x| < 1$, then*

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

If $|x| \geq 1$, the series diverges.

Proof. If $x \neq 1$, we have

$$s_n = \sum_{k=0}^{\infty} x^k = \frac{1-x^{n+1}}{1-x}.$$

This can be easily verfied using the polynomial division. The result follows by taking the limit as $n \to \infty$. You can clearly see that it converges only when $|x| < 1$. If $|x| > 1$, it diverges.

Now putting $x = 1$, then we have

$$\sum_{n=0}^{\infty} x^n = 1 + 1 + 1 + 1 ...$$

which is clearly divergent. □

In many cases, the terms of the series decrease monotonically. The following theorem can help us determine the convergence of a series. What is striking is that *a rather thin subsequence of $\{a_n\}$ determines the convergence of $\sum a_n$.*

Theorem 3.6.5. *Suppose $a_1 \geq a_2 \geq a_3 \geq \ldots \geq 0$. Then the series $\sum_{n=1}^{\infty} a_n$ converges if and only if the series*

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots$$

is convergent.

Proof. Let

$$\begin{aligned} s_n &= a_1 + a_2 + \cdots + a_n \\ t_k &= a_1 + 2a_2 + \cdots + 2^k a_{2^k}. \end{aligned}$$

First, assume that $\{t_k\}$ is convergent, then for $n < 2^k$, we have

$$\begin{aligned} s_n &\leq a_1 + (a_2 + a_3) + \cdots + (a_{2^k} + \cdots + a_{2^{k+1}-1}) \\ &\leq a_1 + 2a_2 + \cdots + 2^k a_{2^k} \\ &= t_k. \end{aligned} \tag{3.5}$$

Hence we have

$$s_n \leq t_k.$$

For sufficiently large k, if $t_k \to t$, we have

$$s_n \leq t_k \leq t$$

But when k is arbitrarily large, then 2^k is also arbitrarily large. Hence $n < 2^k$ implies that n can also be arbitrarily large. Hence for sufficiently large n, s_n is monotonically bounded. Hence $\{s_n\}$ converges. We have completed the first half of our proof.

Conversely, suppose that $\{s_n\}$ is convergent. Then for $n > 2^k$, we have

$$\begin{aligned} s_n &\geq a_1 + a_2 + (a_3 + a_4) + \cdots + (a_{2^{k-1}+1} + \cdots + a_{2^k}) \\ &\geq \tfrac{1}{2} a_1 + a_2 + 2a_4 + \cdots + 2^{k-1} a_{2^k} \\ &= \tfrac{1}{2} t_k \end{aligned} \tag{3.6}$$

So that

$$2s_n \geq t_k.$$

Then for large enough n, we have

$$2s \geq 2s_n \geq t_k.$$

Similarly, $n > 2^k$ implies that k can also be sufficiently large. Then t_k is monotonically bounded, hence convergent. □

Theorem 3.6.6. $\sum \frac{1}{n^p}$ *converges if $p > 1$ and diverges if $p \leq 1$.*

Proof. If $p \leq 0$, it is clearly divergent. If $p > 0$, we can use Thm.3.6.5. Then consider the following series

$$\sum_{k=0}^{\infty} 2^k \cdot \frac{1}{2^{kp}} = \sum_{k=0}^{\infty} 2^{(1-p)k}.$$

This geometric series converges only when $1 - p < 0$. Hence it converges when $p > 1$. □

Corollary 3.6.2. *If $p > 1$,*

$$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$$

converges; if $p \leq i$, the series diverges.

Proof. It is easy to see that $\{1/n \log n\}$ decreases monotonically. For $p \leq 0$, the series is clearly divergent. Now, if $p > 0$, then Thm.3.6.5 applies. This lead us to the series

$$\sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k \left(\log 2^k\right)^p} = \sum_{k=1}^{\infty} \frac{1}{(k \log 2)^p} = \frac{1}{(\log 2)^p} \sum_{k=1}^{\infty} \frac{1}{k^p},$$

and the result follows from Thm.3.6.6 immediately. □

3.7 The Number e

Definition 3.7.1.

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

Note that this series clearly converges because its partial sums are bounded. This is quite obvious since

$$\begin{aligned} s_n &= 1 + 1 + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \cdots + \frac{1}{1 \cdot 2 \cdot \cdot n} \\ &< 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{n-1}} < 3. \end{aligned}$$

Theorem 3.7.1.

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

Proof. Let

$$s_n = \sum_{k=0}^{n} \frac{1}{k!}, \quad t_n = \left(1 + \frac{1}{n}\right)^n.$$

By the binomial theorem, we have

$$\begin{aligned} t_n &= \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{n}\right)^k \\ &= 1 + \left[n \cdot \frac{1}{n}\right] + \left[\frac{n(n-1)}{2!} \left(\frac{1}{n}\right)^2\right] + \left[\frac{n(n-1)(n-2)}{3!} \left(\frac{1}{n}\right)^3\right] + \ldots \\ &= 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots \\ &= \begin{array}{l} 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots \\ \qquad + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{n-1}{n}\right). \end{array} \end{aligned}$$

Hence $t_n \le s_n$, so that

$$\limsup_{n\to\infty} t_n \le e \tag{3.7}$$

Here we take the upper limit simply because we don't know whether t_n has a limit. But we do know that it must have an upper limit. Next, if $n \ge m$, for fixed m we have

$$t_n \ge 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \cdots + \frac{1}{m!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{m-1}{n}\right).$$

Taking the limit as $n \to \infty$, we get

$$\liminf_{n\to\infty} t_n \ge 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{m!}.$$

Same reason why we take the lower limit. We just can't be sure whether there is a limit for t_n.

Now we have

$$s_m \le \liminf_{n\to\infty} t_n.$$

Letting $m \to \infty$, we get

$$e \le \liminf_{n\to\infty} t_n. \tag{3.8}$$

Combining 3.7 and 3.8 we have

$$e \le \liminf_{n\to\infty} t_n \le \limsup_{n\to\infty} t_n \le \limsup_{n\to\infty} s_n = e.$$

Thus we are forced to conclude that

$$\limsup_{n\to\infty} t_n = \liminf_{n\to\infty} t_n = e.$$

This completes the proof. □

We can also estimate how fast $\sum \frac{1}{n}!$ is converging. Let $s_n = \sum_{k=0}^{n} \frac{1}{k!}$, we have

$$\begin{aligned} e - s_n &= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \frac{1}{(n+3)!} + \cdots \\ &< \frac{1}{(n+1)!}\left\{1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \cdots\right\} = \frac{1}{n!n}. \end{aligned}$$

Hence

$$0 < e - s_n < \frac{1}{n!n}. \tag{3.9}$$

This means that this series is converging very quickly to e. The inequality 3.9 is not only useful in numerical computation, but is also of theoretical interests, since it enables us to prove that e is irrational.

Theorem 3.7.2. *e is irrational.*

Proof. Suppose e is rational and write $e = p/q$, where $p, q \in \mathbb{Z}^+$. Then by inequality 3.9 we have

$$0 < q!\left(e - s_q\right) < \frac{1}{q}. \tag{3.10}$$

By our assumption, $q!e$ is an integer. Since

$$q!s_q = q!\left(1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{q!}\right)$$

is an integer, we see that $q!(e - s_q)$ must also be an integer.

Since $q \geq 1$, inequality 3.10 implies that *there must be an integer between 0 and 1*. We have reached a contradiction, as desired. □

3.8 The Final Lecture on Series

This is Professor Su's final lecture on series. Actually, he went through this lecture incredibly fast, covering all the remaining materials on series. As a consequence, he also missed lots of esstential materials. I, on the other hand, feel obligated to slow down and carefully go through theorems in the book.

3.8.1 The Root and Ratio Tests

Theorem 3.8.1. ***(The Root Test)*** *Given* $\sum a_n$, *put* $\alpha = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$. *Then*
(a) If $\alpha < 1$, $\sum a_n$ *converges;*
(b) If $\alpha > 1$, $\sum a_n$ *diverges;*
(c) If $\alpha = 1$, *the test is inconclusive.*

Proof. If $\alpha < 1$, choose β so that $\alpha < \beta < 1$, and an integer N such that

$$n \geq N \Rightarrow \sqrt[n]{|a_n|} < \beta$$

by Thm.3.5.3.That is,

$$n \geq N \Rightarrow |a_n| < \beta^n.$$

Since $0 < \beta < 1$, $\sum \beta^n$ converges, then by comparison test, $\sum a_n$ converges.

If $|a_n| > 1$, then the condition $a_n \to 0$ necessary for $\sum a_n$ to converge does not hold.

If $|a_n| = 1$, consider

$$\sum \frac{1}{n} \qquad \sum \frac{1}{n^2}.$$

The first series diverges while the second converges. So the test is inconclusive. □

Theorem 3.8.2. ***(Ratio Test)***[1] *Given series* $\sum a_n$, *let* $\alpha = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$ *and let* $\gamma = \liminf_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$. *Then*

[1] Rudin states this theorem in a confusing way, so does Professor Su. This theorem is a modified version made by Professor Qiguang Bai in her 14th lecture.

(a) If $\alpha < 1$, $\sum a_n$ converges,
(b) If $\gamma > 1$, $\sum a_n$ diverges,
(c) If $\gamma \le 1 \le \alpha$, the test is inconclusive.

Proof. If condition (a) holds, then we can find $\beta < 1$ and an integer N such that

$$n \ge N \Rightarrow \left|\frac{a_{n+1}}{a_n}\right| < \beta$$

again by Thm.3.5.3. In particular

$$\begin{array}{l} |a_{N+1}| < \beta\,|a_N| \\ |a_{n+2}| < \beta\,|a_{N+1}| < \beta^2\,|a_N| \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \left|a_{N+p}\right| < \beta^p\,|a_N|\,. \end{array}$$

That is

$$|a_n| < |a_N|\,\beta^{-N}\cdot\beta^n$$

for $n \ge N$. Then (a) follows from the comparison test since $\sum \beta^n$ converges.

If $|a_{n+1}| > |a_n|$ for all $n \ge n_0$, it is easy to see that $a_n \to 0$ does not hold, and (b) follows.

For (c), we can refer to $\sum 1/n$ and $\sum 1/n^2$ again. ☐

Example 3.8.1. Consider the following series

$$\frac{1}{2}+\frac{1}{3}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{2^3}+\frac{1}{3^3}+\frac{1}{2^4}+\frac{1}{3^4}+\cdots.$$

You can calculate its upper and lower limits using both root test and ratio test. The result is the following:

$$\begin{array}{l} \liminf_{n\to\infty}\frac{a_{n+1}}{a_n} = \lim_{n\to\infty}\left(\frac{2}{3}\right)^n = 0, \\ \liminf_{n\to\infty}\sqrt[n]{a_n} = \lim_{n\to\infty}\sqrt[2n]{\frac{1}{3^n}} = \frac{1}{\sqrt{3}}, \\ \limsup_{n\to\infty}\sqrt[n]{a_n} = \lim_{n\to\infty}\sqrt[2n]{\frac{1}{2^n}} = \frac{1}{\sqrt{2}}, \\ \limsup_{n\to\infty}\frac{a_{n+1}}{a_n} = \lim_{n\to\infty}\frac{1}{2}\left(\frac{3}{2}\right)^n = +\infty. \end{array}$$

According to the theorems given above, the root test indicates convergence, the ratio test gives no information. generally speaking, the root test is more powerful than the ratio test, whereas the ratio test is easier to calculate. The reason is given in the following theorem.

Theorem 3.8.3. *For any sequence $\{c_n\}$ of positive numbers,*

$$\liminf_{n\to\infty}\frac{c_{n+1}}{c_n} \le \liminf_{n\to\infty}\sqrt[n]{c_n} \le \limsup_{n\to\infty}\sqrt[n]{c_n} \le \limsup_{n\to\infty}\frac{c_{n+1}}{c_n}.$$

Proof. We will first prove that

$$\lim_{n\to\infty}\sup\sqrt[n]{c_n} \le \lim_{n\to\infty}\sup\frac{c_{n+1}}{c_n}.$$

Put

$$\alpha = \lim_{n\to\infty} \sup_{c_n} \frac{c_{n+1}}{c_n}.$$

If $\alpha = +\infty$, then we are done. Suppose α is finite. Choose $\beta > \alpha$. By Thm.3.5.3, there is an integer N such that

$$n \geq N \Rightarrow \frac{c_{n+1}}{c_n} \leq \beta.$$

Similar to the proof of the ratio test, we can get a recurrence relation which breaks down to the following:

$$c_{N+p} \leq \beta^p c_N.$$

Switching the index, we let $N + p := n$, then of course $p := n - N$. So this becomes

$$c_n \leq \beta^{n-N} c_N.$$

Taking the nth root on both sides we get

$$\sqrt[n]{c_n} \leq \beta^{1-N/n} \sqrt[n]{c_N}.$$

The limit on the right hand side exists, and is equal to 1, which you can verify. But on the left hand side we have no clue whether the limit exists. So we take the lim sup. We get

$$\limsup_{n\to\infty} \sqrt[n]{c_n} \leq \beta.$$

But this is true for all $\beta > \alpha$. In other words, if we let $\beta = \alpha + \varepsilon$ for any $\varepsilon > 0$, we get

$$\limsup_{n\to\infty} \sqrt[n]{c_n} \leq \alpha + \varepsilon.$$

By taking the limit again and let $\varepsilon \to 0$ we get

$$\limsup_{n\to\infty} \sqrt[n]{c_n} \leq \alpha.$$

The proof of the second inequality

$$\liminf_{n\to\infty} \frac{c_{n+1}}{c_n} \leq \liminf_{n\to\infty} \sqrt[n]{c_n}$$

is very similar, except this time we use Cor.3.5.1. Combine these two inequalities and this theorem follows. □

Remark 3.8.1. This theorem actually shows us why the root test is more powerful than the ratio test. If you look at this inequality

$$\liminf_{n\to\infty} \frac{c_{n+1}}{c_n} \leq \liminf_{n\to\infty} \sqrt[n]{c_n} \leq \limsup_{n\to\infty} \sqrt[n]{c_n} \leq \limsup_{n\to\infty} \frac{c_{n+1}}{c_n},$$

it basically says *whenever the ratio test shows convergence, the root test does too; whenever the root test is inconclusive, the ratio test is too.*

3.8.2 Power Series

Definition 3.8.1. Given a sequence $\{c_n\}$ of complex numbers, the series

$$\sum_{n=0}^{\infty} c_n z^n$$

is called a *power series*. The number c_n is called the *coefficients* of the series. z is a complex number.

Remark 3.8.2. The definition also holds for real numbers. We are just trying to be general.

Theorem 3.8.4. *Given power series $\sum_{n=0}^{\infty} c_n z^n$, put*

$$\alpha = \limsup_{n\to\infty} \sqrt[n]{|c_n|}, \quad R = \frac{1}{\alpha}.$$

Then the series converges if $|z| < R$ and diverges if $|z| > R$. R is called the radius of convergence.

Proof. Put $a_n = c_n z^n$ and apply the root test, we have

$$\limsup_{n\to\infty} \sqrt[n]{|a_n|} = |z| \limsup_{n\to\infty} \sqrt[n]{|c_n|} = \frac{|z|}{R}.$$

□

3.8.3 Summation by Parts

This is an interesting section since it provides a discrete analoge of integration by parts. Summation by parts actually arises from the observation that *even when $\sum a_n$ and $\sum b_n$ both converge, the series $\sum a_n b_n$ may not converge.* So we need to find an expresion of $\sum a_n b_n$ using a_n, b_n and possibly their sums and differences.

Theorem 3.8.5. ***(Summation by Parts)*** *Given two sequences $\{a_n\}, \{b_n\}$, put*

$$A_n = \sum_{k=0}^{n} a_k$$

if $n \geq 0$; put $A_{-1} = 0$. Then if $0 \leq p \leq q$, we have

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p.$$

Proof. **(Rudin, Algebraic Proof)** We rewrite the left hand side and get

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q} (A_n - A_{n-1}) b_n = \sum_{n=p}^{q} A_n b_n - \sum_{n=p-1}^{q-1} A_n b_{n+1}. \tag{3.11}$$

The expression on the right hand side of the theorem is exactly equal to the right hand side of Eqn.3.11. □

The algebraic proof given by Rudin is straightforward. But it doesn't offer any insight that we should have into this theorem. Now we present Professor Su's geometric proof in his lecture.

Proof. **(Su, Geometric Proof)**

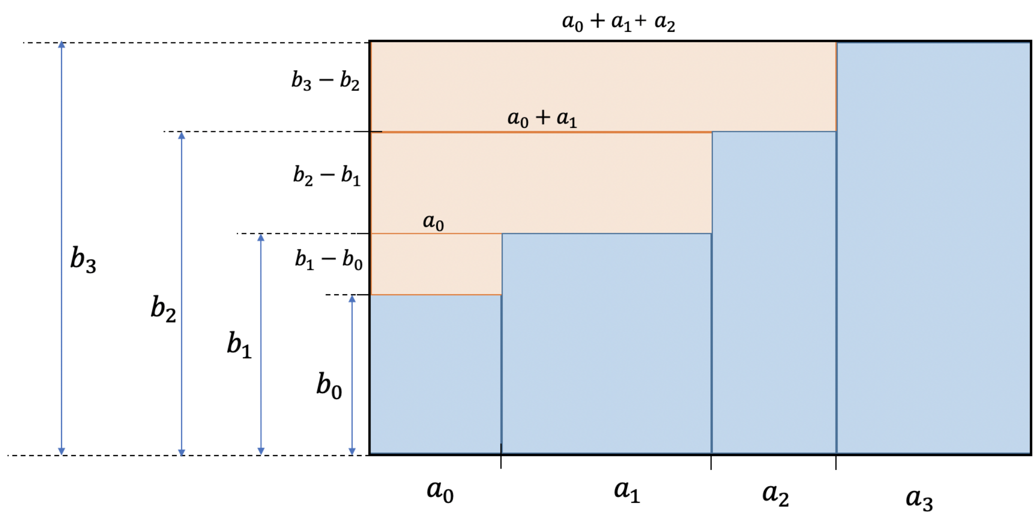


Figure 3.1: A geometric interpretation of summation by parts.

We can define the sequence $\{a_n\}, \{b_n\}$ like the way shown in Fig.3.1. Then $\sum a_n b_n$ is just the sum of areas of the blue rectangles. When this sum of areas becomes hard to compute, we can just use the area of the entire rectangle minus the sum of areas of the orangle rectangles.

If we start from a_0, b_0, and we want to find $\sum_0^3 a_n b_n$, then we can see that

$$\begin{aligned}\sum_0^3 a_n b_n &= \sum_0^2 A_n(b_n - b_{n+1}) + A_3 b_3 - A_{-1} b_0 \\ &= \underbrace{A_3 b_3}_{\text{Whole Rectangle}} \underbrace{- \sum_0^2 A_n(b_{n+1} - b_n)}_{\text{Orange rectangles}}\end{aligned}$$

Note that $A_{-1} b_0 = 0$ since $A_{-1} = 0$. Now we look at cases where we don't start from a_0. Let's start with a_1. Then we have

$$\begin{aligned}\sum_1^3 a_n b_n &= \sum_1^2 A_n(b_n - b_{n+1}) + A_3 b_3 - A_0 b_1 \\ &= \underbrace{A_3 b_3}_{\text{Whole Rectangle}} - \underbrace{\sum_1^2 A_n(b_{n+1} - b_n)}_{\text{Top Two Orange rectangles}} - \underbrace{A_0 b_1}_{\text{Rectangle with Area } A_0 b_1}\end{aligned}$$

Now we gain some geometric intuitions of summation by parts. This should make sense now. □

Remark 3.8.3. Now we have some geometric intuition on the discrete case, it is actually not difficult to generalize it into continous analoge (Though we have not yet defined continuity,differential,and integration).

Suppose you have a function $u(v)$ and you want to find the area under the curve (the shaded shallow blue region in Fig.3.2. However, sometimes the integral

$$\int u(v)dv$$

can be hard to evaluate (we omit the upper and lower limit). However, we can use the same idea we had before, only to generalize it into continuous case.

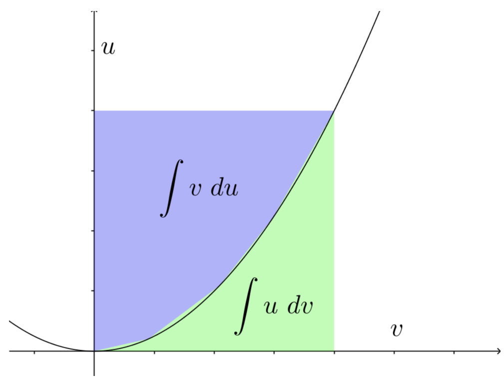


Figure 3.2: A geometric interpretation of integration by parts.

Therefore we have the formula

$$\int udv = uv - \int vdu.$$

In the discrete case, A_n is like the function v, the original function of a_n, which is like dv. And we can see that b_n is like u, so that $b_{n+1} - b_n$ is like du.

Theorem 3.8.6. ***(Dirichlet's Test)*** *Suppose*
(a) the partial sums A_n of $\sum a_n$ form a bounded sequence;

(b) $b_0 \geq b_1 \geq b_2 \geq \ldots$;
(c) $\lim_{n\to\infty} b_n = 0$.
Then $\sum a_n b_n$ *converges.*

Proof. Choose M such that $|A_n| < M$ for all n. Given $\varepsilon > 0$, there is an integer N such that $b_N \leq (\varepsilon/2M)$. For $N \leq p \leq q$, we have

$$\begin{aligned}\left|\sum_{n=p}^{q} a_n b_n\right| &= \left|\sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p\right| \\ &\leq M \left|\sum_{n=p}^{q-1} (b_n - b_{n+1})\right| + M|b_q| + M|b_p| \\ &\leq M \left|\sum_{n=p}^{q-1} (b_n - b_{n+1}) + b_q + b_p\right| = 2M b_p \leq 2M b_N \leq \varepsilon.\end{aligned}$$

Convergence then follows from the Cauchy criterion. Note that in the first inequality we just use $|-A_{p-1}b_p| = |A_{p-1}b_p| < M|b_p|$. □

Theorem 3.8.7. ***(Alternating Series Test)*** *Suppose*
(a) $|c_1| \geq |c_2| \geq |c_3| \geq \ldots$;
(b) $c_{2m-1} \geq 0$ *and* $c_{2m} \leq 0$ *for* $m = 1, 2, 3 \ldots$;
(c) $\lim_{n\to\infty} c_n = 0$.
Then $\sum c_n$ *converges.*

Proof. Apply Thm.3.8.6 and let $a_n = (-1)^{n+1}$ and $b_n = |c_n|$. □

Remark 3.8.4. Series for which (b) holds are called alternating series. Note that a series is an alternating series as long as the signs of consecutive terms alternates. Its odd terms don't necessarily have to be positive and its even term don't necessarily have to be negative. It can also be the other way around.

Theorem 3.8.8. *Suppose the radius of convergence of* $\sum c_n z^n$ *is 1, and suppose* $c_0 \geq c_1 \geq c_2 \ldots$, $\lim c_n = 0$. *Then* $\sum c_n z^n$ *converges at every point on the circle* $|z| = 1$, *except possibly at* $z = 1$.

Proof. Put $a_n = z^n$ and $b_n = c_n$. If $|z| = 1$ and $z \neq 1$, then we have

$$|A_n| = \left|\sum_{m=0}^{n} z^m\right| = \left|\frac{1 - z^{n+1}}{1 - z}\right| \leq \frac{2}{|1 - z|}.$$

So (a) in Thm.3.8.6 is satisfied. It is also obvious that both (b) and (c) are satisfied. The theorem then follows. □

3.8.4 Absolute Convergence

Definition 3.8.2. The series $\sum a_n$ is said to *converge absolutely* if the series $\sum |a_n|$ converges.

Theorem 3.8.9. *If $\sum a_n$ converges absolutely, then $\sum a_n$ converges.*

Proof. The assertion follows from the inequality

$$\left|\sum_{k=n}^{m} a_k\right| \leq \sum_{k=n}^{m} |a_k|,$$

plus the Cauchy criterion. □

Definition 3.8.3. If $\sum a_n$ converges but $\sum |a_n|$ diverges, we say that $\sum a_n$ *converges nonabsolutely* or *conditionally*.

Example 3.8.2. The series

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

converges conditionally.

Remark 3.8.5. The root test and the ratio test are really tests for absolute convergence. They can not give any information on conditionally convergent series.

Example 3.8.3. (From Qiguang Bai's Lecture) Consider the series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\sqrt{n}}.$$

and we let

$$a_n = b_n = \frac{(-1)^{n+1}}{\sqrt{n}}.$$

Both $\sum a_n$ and $\sum b_n$ converges, but $\sum a_n b_n$ diverges. This is because neither $\sum a_n$ nor $\sum b_n$ is absolutely convergent. This may motivate us to replace the condition "convergent" by a stronger condition "absolute convergent" and see if we can get some interesting results. An example of this replacement is given in the next section.

3.8.5 Addition and Multiplication of Series

Theorem 3.8.10. *If $\sum a_n = A$ and $\sum b_n = B$, then $\sum (a_n + b_n) = A + B$ and $\sum c a_n = cA$ for any fixed c.*

Proof. Trivial. Left as exercise. □

Definition 3.8.4. (Cauchy Product) Given $\sum a_n$ and $\sum b_n$, we put

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \quad (n = 0, 1, 2, \dots)$$

and call $\sum c_n$ the *product* of two given series.

Example 3.8.4. The product of two convergent series may actually diverge. See Example 3.49 in Rudin, pg.73.

Theorem 3.8.11. *Suppose*
(a) $\sum_{n=0}^{\infty} a_n$ *converges absolutely;*
(b) $\sum_{n=0}^{\infty} a_n = A$*;*
(c) $\sum_{n=0}^{\infty} b_n = B$*;*
(d) $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ *for* $n = 0, 1, 2...$
Then

$$\sum_{n=0}^{\infty} c_n = AB.$$

Proof. Put

$$A_n = \sum_{k=0}^{n} a_k, \quad B_n = \sum_{k=0}^{n} b_k, \quad C_n = \sum_{k=0}^{n} c_k, \quad \beta_n = B_n - B.$$

Then

$$\begin{aligned} C_n &= a_0 b_0 + (a_0 b_1 + a_1 b_0) + \cdots + (a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0) \\ &= a_0 B_n + a_1 B_{n-1} + \cdots + a_n B_0 \\ &= a_0 (B + \beta_n) + a_1 (B + \beta_{n-1}) + \cdots + a_n (B + \beta_0) \\ &= A_n B + a_0 \beta_n + a_1 \beta_{n-1} + \cdots + a_n \beta_0. \end{aligned}$$

Put

$$\gamma_n = a_0 \beta_n + a_1 \beta_{n-1} + \cdots + a_n \beta_0.$$

We wish to show that $C_n \to AB$. Since $A_n B \to AB$, it suffices to show that

$$\lim_{n\to\infty} \gamma_n = 0. \tag{3.12}$$

Put

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

By (c) we know that $\beta_n \to 0$. We also know that $a_n \to 0$ by (b). So a natural way we can think of is to cut γ_n into halves, for the former half we use the condition $a_n \to 0$ and for the latter half we use $\beta \to 0$. Now, given $\varepsilon > 0$, choose N such that $|\beta_n| \leq \varepsilon$ for $n \geq N$, in which case

$$\begin{aligned} |\gamma_n| &\leq |\beta_0 a_n + \cdots + \beta_N a_{n-N}| + |\beta_{N+1} a_{n-N-1} + \cdots + \beta_n a_0| \\ &\leq |\beta_0 a_n + \cdots + \beta_N a_{n-N}| + \varepsilon\alpha. \end{aligned}$$

Keeping N fixed and let $n \to \infty$, we get

$$\limsup_{n\to\infty} |\gamma_n| \leq \varepsilon\alpha.$$

Since ε is arbitrary, Eqn.3.12 follows, and we are done. □

3.8.6 Rearrangements

Professor Su only briefly mentioned about this subsection (I mean really brief, like 5 min), which I don't appreciate. On the other hand, Professor Bai's lecture gives a much more detailed introduction to rearrangement, one of the most important topics on infinite series. This lecture note is largely excerpted from her lecture[2].

Definition 3.8.5. Let $\{k_n\}$, $n = 1, 2, 3...$, be a sequence in which every positive integer appears once and only once. That is, $k_n : \mathbb{N} \to \mathbb{N}$ is a 1-1 and onto map. Put

$$a'_n = a_{k_n}. \qquad (n = 1, 2, 3...)$$

We say that $\sum a'_n$ is a *rearrangement* of $\sum a_n$.

Theorem 3.8.12. *Let $\sum a_n$ be a series of real numbers which converges but not absolutely converges. Suppose*

$$-\infty \le \alpha \le \beta \le \infty.$$

Then there exists a rearrangement $\sum a'_n$ with partial sums s'_n such that

$$\liminf_{n\to\infty} s'_n = \alpha, \quad \limsup_{n\to\infty} s'_n = \beta. \tag{3.13}$$

Remark 3.8.6. We can put $\alpha = \beta$, and let them be any real numbers. Then 3.13 says $\sum a'_n$ converges. What's more, it can converge to any real numbers you want! So rearrangement will cause significant changes of the final result. Given a convergent series, its rearrangement may not converge. Even if it converges, it can converge possibly to anything! So it may not be the original result.

Proof. Let

$$p_n = \frac{|a_n| + a_n}{2}, \quad q_n = \frac{|a_n| - a_n}{2} \quad (n = 1, 2, 3, \ldots).$$

You can see that $p_n = a_n$ if $a_n > 0$, and $p_n = 0$ if $a_n < 0$. Similarly, you can see that $q_n = 0$ if $a_n > 0$, and $q_n = |a_n|$ if $a_n < 0$.

Another observation you can make is that by taking summation on both sides of p_n, q_n, and given that $\sum |a_n|$ diverges and $\sum a_n$ converges, both $\sum p_n$ and $\sum q_n$ must be divergent.

Now let $P_1, P_2, P_3, ...$ denote the nonnegative terms of $\sum a_n$, in the order in which they occur. And let $Q_1, Q_2, Q_3, ...$ be the absolute values of the negative terms of $\sum a_n$, also in their original order.

Then the series $\sum P_n, \sum Q_n$ only differ from $\sum p_n, \sum q_n$ by zero terms, and are therefore divergent.

We shall construct sequences $\{m_n\}, \{k_n\}$ such that the series

$$\begin{aligned} P_1 + \cdots + P_{m_1} - Q_1 - \cdots - Q_{k_1} + P_{m_1+1} + \cdots \\ + P_{m_2} - Q_{k_1+1} - \cdots - Q_{k_2} + \cdots, \end{aligned} \tag{3.14}$$

[2] Actually really good lectures! And much slower than Su's, so you'll get much more details on how some constructions in the proofs are motivated.

which is clearly a rearrangement of $\sum a_n$, satisfies 3.13.

Choose real-valued sequences $\{\alpha_n\}, \{\beta_n\}$ such that $\alpha_n \to \alpha, \beta_n \to \beta$, and additionally,$\alpha_n < \beta_n, \beta_1 > 0$.

Let m_1, k_1 be the **smallest** integers such that

$$P_1 + \cdots + P_{m_1} > \beta_1,$$
$$P_1 + \cdots + P_{m_1} - Q_1 - \cdots - Q_{k_1} < \alpha_1.$$

Let m_2, k_2 be the **smallest** integer such that

$$P_1 + \cdots + P_{m_1} - Q_1 - \cdots - Q_{k_1} + P_{m_1+1} + \cdots + P_{m_2} > \beta_2,$$
$$P_1 + \cdots + P_{m_1} - Q_1 - \cdots - Q_{k_1} + P_{m_1+1} + \cdots + P_{m_2} - Q_{k_1+1} - \cdots - Q_{k_2} < \alpha_2.$$

We repeat this process. This is possible since both $\sum P_n$ and $\sum Q_n$ diverges. So, if the result is too big since we added too many Ps, then we can make it small enough by subtracting enough Qs from it.

If x_n, y_n denote the partial sums of 3.14 whose last terms are P_{m_n}, $-Q_{k_n}$, and since we choose the smallest integer P_{m_n}, Q_{k_n} such that $x_n > \beta_n, y_n < \alpha_n$, we must have

$$|x_n - \beta_n| \leq P_{m_n}, \quad |y_n - \alpha_n| \leq Q_{k_n}.$$

Since $P_n \to 0$ and $Q_n \to 0$ as $n \to \infty$, we see that $x_n \to \beta$ and $y_n \to \alpha$. Note that $n \to \infty$ implies $m_n, k_n \to \infty$.

Finally, there is no number less than α or greater than β can be a subsequential limit of the partial sums of 3.14. $\square$

Remark 3.8.7. (by Mikhail D, MathStackExchange). A picture is worth a thousand words.

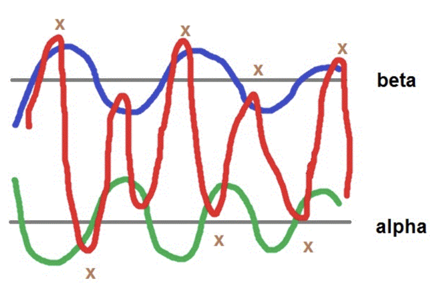


Figure 3.3: An illustration.

Curves in blue and in green are sequences β_n, α_n. They converge to β, α respectively. Curve in red is our Expression 3.14. The way we painted this picture shows a few things about Expression 3.14.

The first terms of it are positive, so the reason why $\beta_1 > 0$.

Next, $x_n > \beta_n$ and $y_n < \alpha_n$. so the red curve sticks out a bit "outside" sequences $\{\beta_n\}, \{\alpha_n\}$, by one last positive term and negative term $P_{m_n}, -Q_{k_n}$.

Differences between 3.14 and these sequences become smaller because $P_n, Q_n \to 0$. It is precisely for this reason it becomes finally "clear" that Expression 3.14 cannot converge to any number greater than β or smaller than α.

Theorem 3.8.13. *If $\sum a_n$ is a series which converges absolutely to s. Then every rearrangement of $\sum a_n$ converges to s.*

Proof. Let $\sum a'_n = \sum a_{k_n}$ be a rearrangement and let

$$s_n = \sum_{k=0}^{n} a_k \qquad s'_n = \sum_{k=0}^{n} a'_k.$$

Given $\varepsilon > 0$, since $\sum |a_n|$ converges, there exists an N such that

$$m \geq n \geq N \Rightarrow \sum_{k=n}^{m} |a_k| < \frac{\varepsilon}{2}$$

and

$$n \geq N \Rightarrow |s_n - s| < \frac{\varepsilon}{2}.$$

Now, choose p so large such that

$$\{1, 2, 3..., N\} \subset \left\{k_1, k_2, ..., k_p\right\}. \qquad \text{Note that } p \geq N$$

Then if $n \geq p$, $s_n - s'_n$ contains no terms a_k where $k < N$. Thus

$$|s_n - s'_n| \leq \sum_{k=l}^{m} |a_k| < \frac{\varepsilon}{2}. \qquad (m \geq l \geq p)$$

Hence

$$|s'_n - s| \leq |s'_n - s_n| + |s_n - s| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for all $n \geq p$. This completes the proof. □

Chapter 4

Continuity

The apex of mathematical achievement occurs when two or more fields which were thought to be entirely unrelated turn out to be closely intertwined.

-Gian-Carlo Rota, Indiscrete Thoughts

4.1 Limits of Functions, Continuity

Definition 4.1.1. Let X, Y be metric spaces; suppose $E \subset X$, $f : E \to Y$, and p is a limit point of E. We write $f(x) \to q$ as $x \to p$, or

$$\lim_{x \to p} f(x) = q$$

if there is a point $q \in Y$ with the following property: For every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$0 < d_X(x, p) < \delta \Rightarrow d_Y(f(x), q) < \varepsilon$$

for all $x \in E$. Here d_X, d_Y refer to the distances in X and Y.

Remark 4.1.1. Note that in our definition, we say $0 < d_X(x, p) < \delta$. Since the distance is, by default, bigger than zero, this implies that $x \neq p$. In other words, the limit of $f(x)$ as $x \to p$ has nothing to do with the value of f at point p. *Nor do we require p to be a point in E.*

Theorem 4.1.1. *Let X, Y, E, f, p have the same meaning as in the definition above. Then* $\lim_{x \to p} f(x) = q$ *if and only if* $\lim_{n \to \infty} f(p_n) = q$ *for every sequence* $\{p_n\}$ *in* E *such that* $p_n \neq p, \lim_{n \to \infty} p_n = p$.

Proof. Suppose $f(x) \to q$ as $x \to p$. Choose $\{p_n\}$ in E such that the conditions $p_n \neq p, \lim_{n \to \infty} p_n = p$ are satisfied. Then there exists $\delta > 0$ such that $d_Y(f(x), q) < \varepsilon$ if $x \in E$ and $d_X(x, p) < \delta$. Also, there exists N such that $n > N$ implies $d_X(p_n, p) < \delta$. Thus, for $n > N$, we have $d_Y(f(p_n), q) < \varepsilon$. The first half of the proof is complete.

Conversely, we prove the contrapositive. That is, suppose $f(x) \nrightarrow q$ as $x \to p$. Then there exists some $\varepsilon > 0$ such that for every $\delta > 0$ there exists a point $x \in E$ such that $d_Y(f(x), q) \geq \varepsilon$ but $0 < d_X(x, p) < \delta$. Taking $\delta_n = 1/n$, we find a sequence of such x that makes $f(p_n) \to q$ false. $\square$

Corollary 4.1.1. *If f has a limit at p, this limit is unique.*

Proof. We use Thm.4.1.1. Since the limit of a sequence is unique (we proved this in the first section of Chapter 3), then so is the limit of a function. □

Theorem 4.1.2. *Suppose $E \subset X$ and p is a limit point of E. Let f, g be complex functions on E and*

$$\lim_{x\to p} f(x) = A, \quad \lim_{x\to p} g(x) = B.$$

Then
(a) $\lim_{x\to p}(f+g)(x) = A + B$;
(b) $\lim_{x\to p}(fg)(x) = AB$;
(c) $\lim_{x\to p}\left(\frac{f}{g}\right)(x) = \frac{A}{B}$, *if* $B \neq 0$.

Proof. Given Thm.4.1.1, this theorem follows immediately from Thm.3.1.5. □

Theorem 4.1.3. *If $\mathbf{f}, \mathbf{g} : E \to \mathbb{R}^k$, then (a) in Thm.4.1.2 is still true, and (b) becomes*

$$\lim_{x\to p}(\mathbf{f}\cdot\mathbf{g})(x) = \mathbf{A}\cdot\mathbf{B}.$$

Proof. Given Thm.4.1.1 and Thm.3.1.7, the theorem immediately follows. □

Definition 4.1.2. Suppose X, Y are metric spaces, $E \subset X$, $p \in E$, and $f : E \to Y$. Then f is *continuous* at p if for every $varepsilon > 0$ there exists a $\delta > 0$ such that

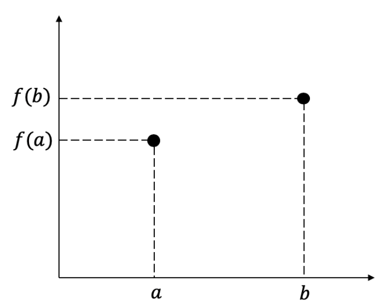
$$0 < d_X(x,p) < \delta \Rightarrow d_Y(f(x), f(p)) < \varepsilon$$

for all $x \in E$.

If f is continous at every point of E, then f is said to be *continous on E*.

Remark 4.1.2. Compare this with the definition of limit of a function, we see that f *has to be defined at point p in order to be continous at p.*

Example 4.1.1. Consider the function f defined by the figure below.



Let $f : E \to Y$ where $E = \{a, b\}$. This function is continuous, though it may not look like a continuous function. Since there are only two points in the set, without loss of generality, consider the point b. The condition

$$d_X(x,b) < \delta \Rightarrow d_Y(f(x), f(b)) < \varepsilon$$

is always satisfied even if you choose very small ε and δ, then the set of points $x \in E$ such that $d_X(x, b) < \delta$ can only be b itself. Therefore $d_Y(f(x), f(b)) = 0 < \varepsilon$ for any $\varepsilon > 0$.

Theorem 4.1.4. *f is continuous at p if and only if* $\lim_{x\to p} f(x) = f(p)$.

Proof. Trivial. □

Theorem 4.1.5. *If f, g are continuous functions on a metric space X, then $f + g, fg, f/g$ are all continous on X (for the last case $g \neq 0$).*

Proof. Follows from Thm.4.1.2 and Thm.4.1.4. □

Theorem 4.1.6. *(a) Let $f_1, ... f_k$ be real functions on a metric space X, and let $\mathbf{f} : X \to \mathbb{R}^k$ be defined by*

$$\mathbf{f}(x) = (f_1(x), \dots, f_k(x)) \quad (x \in X);$$

Then $\mathbf{f}$ is continous if and only if each f_i is continuous.
(b) If $\mathbf{f}, \mathbf{g}$ are continous mappings from X into $\mathbb{R}^k$, then $\mathbf{f} + \mathbf{g}$ and $\mathbf{f} \cdot \mathbf{g}$ are continuous on X.

Proof. The proof is actually very similar to that of Thm.3.1.6. Part (a) follows from the inequality

$$\left|f_j(x) - f_j(y)\right| \le |\mathbf{f}(x) - \mathbf{f}(y)| = \left\{\sum_{i=1}^{k} |f_i(x) - f_i(y)|^2\right\}^{\frac{1}{2}}.$$

Part (b) follows from (a) and Thm.4.1.5. □

Theorem 4.1.7. *Suppose X, Y, Z are metric spaces, $E \subset X$, $f : E \to Y$, g maps the range of f, $f(E)$, into Z. And the mapping $h : E \to Z$ defined by*

$$h(x) = g(f(x)) \quad (x \in E).$$

If f is continuous at a point $p \in E$ and g is continuous at the point $f(p)$, then h is continuous at p.

Proof. Let $\varepsilon > 0$ be given. Since g is continuous at $f(p)$, there exists $\eta > 0$ such that

$$d_Z(g(y), g(f(p))) < \varepsilon \text{ if } d_Y(y, f(p)) < \eta \text{ and } y \in f(E).$$

Since f is continuous at p, then there exists $\delta > 0$ such that

$$d_Y(f(x), f(p)) < \eta \text{ if } d_X(x, p) < \delta \text{ and } x \in E.$$

It then follows that

$$d_Z(h(x), h(p)) = d_Z(g(f(x)), g(f(p))) < \varepsilon$$

If $d_X(x, p) < \delta$ and $x \in E$. Thus h is continuous at p. □

Remark 4.1.3. This theorem says that a continuous function of a continuous function is continuous. This function h is called the composite function of f and g. The notation $h = g \circ f$ is frequently used in this context.

4.2 Operations on Inverse Mappings

Before we dive into the topological interpretation of continuity, it is first necessary for us to gain familiarity with inverse mapping and its operations, as they will be very useful in the proofs of the next section.

We first present the definition of a image and an inverse mapping.

Definition 4.2.1. Given a mapping $f : X \to Y$, Consider a set $A \subset X$. The image of A is defined as

$$f(A) := \{f(x) : x \in A\}.$$

Definition 4.2.2. Given a mapping $f : X \to Y$, an inverse mapping $f^{-1} : Y \to X$. Consider a set $C \in Y$. Then the inverse image of C is defined as

$$f^{-1}(C) := \{x \in X : f(x) \in C\}.$$

Now using these two definitions, we present the operations on an inverse mapping[1].

NOTE. For all the following theorems in this section, we assume that $f : X \to Y$, $A, B \subset X$, and $C, D \subset Y$.

Theorem 4.2.1. *If $A \subset B$, then $f(A) \subset f(B)$.*

Proof. Let $y \in f(A)$. Then exists $a \in A$ such that $f(a) = y$. Since $a \in A \subset B$, we have $a \in B$. Then $f(a) \in f(B)$. Hence, $y \in f(B)$. Done. □

Theorem 4.2.2. $f^{-1}(f(A)) \supset A$.

Proof. Let $a \in A$. Then $f(a) \in f(A)$ by definition. Hence, $a \in f^{-1}(f(A))$. Done. □

Theorem 4.2.3. *If $C \subset D$, then $f^{-1}(C) \subset f^{-1}(D)$.*

Proof. Let $x \in f^{-1}(C)$. Then by definition we have $f(x) \in C$. Hence $f(x) \in D$. This implies that $x \in f^{-1}(D)$. Done. □

Theorem 4.2.4. $f(f^{-1}(C)) \subset C$.

Proof. Let $y \in f(f^{-1}(C))$. Then by definition there exists $x \in f^{-1}(C)$ such that $f(x) = y$. Since $x \in f^{-1}(C)$, then by definition $y = f(x) \in C$. Done. □

Theorem 4.2.5. $f(A \cup B) = f(A) \cup f(B)$.

Proof. Let $y \in f(A \cup B)$. Then exists $x \in A \cup B$ such that $f(x) = y$. Without loss of generality, let $x \in A$. Then $y = f(x) \in f(A) \subset f(A) \cup f(B)$. Hence $f(A \cup B) \subset f(A) \cup f(B)$.

Conversely, let $y \in f(A) \cup f(B)$. Without loss of generality, let $y \in f(B)$. Then exists $x \in B$ such that $f(x) = y$. Hence $x \in A \cup B$. This implies that $y = f(x) \in f(A \cup B)$. So we have $f(A \cup B) \supset f(A) \cup f(B)$. □

[1] the proofs are excerpted from the solution to In Class HW4 of MATH4803, *Foundations of Mathematical Proof* by Prof.Heitsch.

Theorem 4.2.6. $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$.

Proof. Let $x \in f^{-1}(C \cup D)$. Then $f(x) \in C \cup D$. Suppose $f(x) \in C$. Then $x \in f^{-1}(C) \subset f^{-1}(C) \cup f^{-1}(D)$. Hence we have $f^{-1}(C \cup D) \subset f^{-1}(C) \cup f^{-1}(D)$.

Conversely, let $x \in f^{-1}(C) \cup f^{-1}(D)$. Supose $x \in f^{-1}(D)$. Then $f(x) \in f(D) \subset C \cup D$. Thus $x \in f^{-1}(C \cup D)$. This implies that $x \in f^{-1}(C) \cup f^{-1}(D)$. Hence $f^{-1}(C \cup D) \supset f^{-1}(C) \cup f^{-1}(D)$. □

Theorem 4.2.7. $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.

Proof. Let $x \in f^{-1}(C \cap D)$. Then $f(x) \in C \cap D$. Since $f(x) \in C$ and $f(x) \in D$, we have $x \in f^{-1}(C) \cap f^{-1}(D)$. Hence $f^{-1}(C \cap D) \subset f^{-1}(C) \cap f^{-1}(D)$.

Conversely, let $x \in f^{-1}(C) \cap f^{-1}(D)$. Then we have $f(x) \in C$ and $f(x) \in D$. This leads to $x \in f^{-1}(C \cap D)$. Hence $f^{-1}(C \cap D) \supset f^{-1}(C) \cap f^{-1}(D)$. □

The above theorems are treu for any function f. Now we recall the definition of 1-1 and onto functions, and conclude this section with two theorems.

Theorem 4.2.8. *f is 1-1 if and only if $f^{-1}(f(A)) = A$ for all $A \subset X$.*

Proof. Suppose f is 1-1. We know that $f^{-1}(f(A)) \supset A$ for any function f. Let $x \in f^{-1}(f(A))$. Then $f(x) = y \in f(A)$. Since $y = f(A)$, there is an $a \in A$ such that $f(a) = y$. Since $f(x) = y = f(a)$ and f is 1-1, then $x = a$. Hence $x \in A$. Thus we have $f^{-1}(f(A)) \subset A$ and $f^{-1}(f(A)) \supset A$. We conclude that $f^{-1}(f(A)) = A$.

Conversely, assume that $f^{-1}(f(A)) = A$ for all $A \subset X$ but f is not 1-1. That is, there exists $x, x' \in X$ such that $f(x) = f(x')$ but $x \neq x'$. Let $A = \{x\}$. Then $f(A) = \{f(x)\}$. Then we have $\{x, x'\} \subset f^{-1}(f(A))$. But since $A = \{x\}$ and doesn't contain x', $A \neq f^{-1}(f(A))$. A contradiction, as desired. □

Theorem 4.2.9. *f is onto if and only if $f(f^{-1}(C)) = C$ for all $C \subset Y$.*

Proof. Suppose f is onto. We know that $f(f^{-1}(C)) \subset C$ is true for all f. Let $y \in C$. Since f is onto, there must exists $x \in X$ such that $f(x) = y \in C$, since all $y \in Y$ is achieved by f if f is onto. Hence $x \in f^{-1}(C)$. Hence, $y = f(x) \in f(f^{-1}(C))$. We have then $f(f^{-1}(C)) = C$.

Conversely, assume $f(f^{-1}(C)) = C$ for all $C \subset Y$ but f is not onto. That is, there exists $y \in Y$ such that for all $x \in X$, $y \neq f(x)$. Let $C = \{y\}$, then $f^{-1}(C) = \emptyset$. Then $f(f^{-1}(C)) = \emptyset \neq C$ since $y \in C$. A contradiction, as desired. □

4.3 Topology of Continuous Mappings

As mentioned before, we have understood continuity using limit perspective and the subsequential perspective. However, a third perspective, that is topological perspective, could be surprisingly useful. You will see why shortly.

4.3.1 Continuity and Openness

Theorem 4.3.1. *A mapping $f : X \to Y$ is continuous if and only if $f^{-1}(V)$ is open in X for every open set V in Y.*

Proof. First, let f be a continuous mapping. We want to show that every point of $f^{-1}(V)$ is an interior point of $f^{-1}(V)$. So, choose an open set $V \subset Y$. Suppose $p \in X$ and $f(p) \in V$. Since V is open, there exists $\varepsilon > 0$ such that $N_\varepsilon(f(p)) \subset V$. Since f is continuous, then there exists $\delta > 0$ such that

$$f(N_\delta(p)) \subset N_\varepsilon(f(p)) \subset V.$$

Hence,

$$N_\delta(p) \subset f^{-1}(f(N_\delta(p))) \subset f^{-1}(V)$$

by the definition of the inverse mapping. Hence p is an interior point of $f^{-1}(V)$ and $f^{-1}(V)$ is open.

Now suppose the converse is true. That is, suppose $f^{-1}(V)$ is an open set in X for every open set $V \subset Y$. Fix $p \in X$ and $\varepsilon > 0$. Choose our open set in Y to be a ball B around $f(p)$ with radius ε. Then $f^{-1}(B)$ is open. Note that by the definition of an inverse mapping, $p \in f^{-1}(B)$. Then since p is an interior point, there is a ball around p with radius δ such that $N_\delta(p) \subset f^{-1}(V)$. Thus we found the desired δ such that

$$x \in N_\delta(p) \Rightarrow d_Y(f(x), f(p)) < \varepsilon.$$

So f is continuous. □

Corollary 4.3.1. *A mapping $f : X \to Y$ is continuous if and only if $f^{-1}(C)$ is closed in X for every closed set C in Y.*

Proof. Suppose f is continuous. Then we use

$$f^{-1}(E^c) = \left[f^{-1}(E)\right]^c \qquad (E \subset Y).$$

Suppose E is closed, then E^c is open. Then the right hand side $\left[f^{-1}(E)\right]^c$ is open. Then $f^{-1}(E)$ is closed.

Now suppose E closed and $f^{-1}(E)$ closed. Then using the equation above, $f^{-1}(E^c)$ is open. We also know that E^c is open. So f is a continuous mapping. □

Using Thm.4.3.1, we can prove Thm.4.1.7 very easily. We state this theorem again.

Theorem 4.3.2. *Suppose*

$$X \xrightarrow{f} Y \xrightarrow{g} Z,$$

and f, g are continuous. Then $g \circ f$ is also continuous.

Proof. Given U open in Z, then $g^{-1}(U)$ is open in Y since g is continuous. Then $f^{-1}(g^{-1}(U))$ is open in X since f is continuous. But $f^{-1}(g^{-1}(U))$ is really just $(g \circ f)^{-1}(U)$, hence $g \circ f$ must also be a continuous mapping. Done. □

4.3.2 Continuity and Compactness

Theorem 4.3.3. *Suppose f is a continuous mapping of a compact metric space X into a metric space Y. Then $f(X)$ is compact.*

Proof. Choose cover $\{V_\alpha\}$ of $f(X)$. Then let $U_\alpha = f^{-1}(V_\alpha)$. Since X is compact, then there exists a finite subcover $\{U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n}\}$. Then it is clear that the corresponding cover

$$\{V_{\alpha_1}, V_{\alpha_2}, ..., V_{\alpha_n}\}$$

covers $f(X)$. Therefore, $f(X)$ is compact. □

Remark 4.3.1. The theorem above tells us that if f is continuous mapping, while it is possible to map the open interval (a, b) onto $\mathbb{R}$, it is impossible to mape a closed interval $[a, b]$ onto the entire $\mathbb{R}$ since the mapping of a compact set is also compact because of the continuity of f. Compact sets are "small" in some sence. Hence if f is continuous, then if it takes in a "small" input (closed interval), the output must also be relatively small.

Corollary 4.3.2. *Let continuous mapping $f : X \to \mathbb{R}^k$ where X is a compact set, the image $f(X)$ must be closed and bounded.*

Proof. In $\mathbb{R}^k$ compact set is closed and bounded. □

Corollary 4.3.3. *Let continuous mapping $f : X \to \mathbb{R}$, where X is a compact set, then f achieves its maximum and minimum.*

Proof. The image of f, which is $f(X)$ is compact in $\mathbb{R}$, so it is closed and bounded. Being bounded means that $\sup f$ and $\inf f$ exist. Since $f(X)$ is closed, by Lemma 2.8.1, $\sup f \in f(X)$. Similarly, $\inf f \in f(X)$. □

Theorem 4.3.4. *Suppose f is a continous 1-1 mapping of a compact metric space X onto a metric space Y. Then the inverse mapping f^{-1} defined on Y is a continous mapping of Y onto X.*

Proof. Apply Thm.4.3.1 to the inverse mapping. Since we have

$$f^{-1} : f(X) \to X,$$

Then we want to show that $f(V)$ *is an open set in Y for every open set $V \subset X$*. We fix such a set V in X.

The complement $V^c \subset X$ is closed. Hence it is a closed subset of a compact set. Hence V^c is compact. Hence $f(V^c)$ is a compact subset of Y by Thm.4.3.3. Hence $f(V^c)$ is closed in Y by Thm.2.6.2. Since. Since f is 1-1 and onto, we have

$$f(V) = [f(V^c)]^c.$$

Therefore, $f(V)$ is open. □

4.4 Continuity and Connectedness

Theorem 4.4.1. *Consider a continous mapping $f : X \to Y$. Let $E \subset X$. Then $f(E)$ is connected.*

Proof. (Contradiction) Suppose E is connected but $f(E)$ is not. Let $f(E) = A \cap B$ be a seperation. Then let

$$\begin{aligned} U_1 &= f^{-1}(A) \cap E, \\ U_2 &= f^{-1}(B) \cap E. \end{aligned}$$

We claim that $U_1, U_2 \neq \emptyset$. Since $A, B \neq \emptyset$, there exists $a \in A, b \in B$. Then $a, b \in f(E)$. Hence there are $\alpha, \beta \in E$ such that $f(\alpha) = a$, $f(\beta) = b$. But $a \in A$ implies that $\alpha \in f^{-1}(A)$. $b \in B$ implies that $\beta \in f^{-1}(B)$. Hence both U_1, U_2 are nonempty.

Also, we claim $U_1 \cup U_2 = E$. Since

$$\begin{aligned} U_1 \cup U_2 &= (f^{-1}(A) \cap E) \cup (f^{-1}(B) \cap E) \\ &= (f^{-1}(A) \cup f^{-1}(B)) \cap E \\ &= f^{-1}(A \cup B) \cap E \\ &= f^{-1}(f(E)) \cap E \\ &= E. \end{aligned}$$

We note that

$$\begin{aligned} \overline{U_1} &= \overline{f^{-1}(A) \cap E} \\ &= \overline{f^{-1}(A)} \\ &\subset \overline{f^{-1}(\overline{A})} = f^{-1}(\overline{A}). \end{aligned}$$

The last step is true because f is continous and $\overline{A}$ is closed.

Then,

$$\begin{aligned} \overline{U_1} \cap U_2 &\subset f^{-1}(\overline{A}) \cap (f^{-1}(B) \cap E) \\ &\subset f^{-1}(\overline{A}) \cap f^{-1}(B) \\ &= f^{-1}(\overline{A} \cap B) = f^{-1}(\emptyset) = \emptyset. \end{aligned}$$

Hence E is seperated. But this is a contradiction. □

The fact that a continuous mapping preserves connectedness means a lot. It directly leads to the intermediate value theorem, which we will prove shortly. But first we need to prove a lemma.

Lemma 4.4.1. *a set $E \subset \mathbb{R}$ is connected if and only if for all $x, y \in E$, $x < z < y$ implies $z \in E$.*

Proof. We first prove the forward direction. Suppose the contrary, that E is connected but there exist $x, y \in E$ such that $x < z < y$ but $z \notin E$. Then consider

$$A = E \cap (-\infty, z) \qquad B = E \cap (z, \infty).$$

Since $x \in A$, $y \in B$, we know A, B are not empty. We claim that $E = A \cup B$ since

$$\begin{aligned} A \cup B &= (E \cap (-\infty, z)) \cup (E \cap (z, \infty)) \\ &= E \cap ((-\infty, z) \cup (z, \infty)) \\ &= E \cap \mathbb{R} \setminus z = E. \end{aligned}$$

Now,

$$\begin{aligned} \overline{A} \cap B &= ((-\infty, z] \cap E) \cap ((z, \infty) \cap E) \\ &= ((-\infty, z] \cap (z, \infty)) \cap E \\ &= \emptyset \cap E = \emptyset. \end{aligned}$$

Hence E is not connected. A contradiction.

Now we prove the backward direction by proving its contrapositive. That is, if E is not connected, we want to show that there exists $x, y \in E$ such that $x < z < y$ but $z \notin E$.

Let $x \in A$, $y \in B$. Without loss of generality, let $x < y$. Consider

$$S = A \cap [x, y].$$

S is evidently nonempty. Since S is bounded above by y, by the completenesso of $\mathbb{R}$, its suprimum exists. Let $z = \sup S$. Then by Thm.2.8.1, $z \in \overline{S}$. Notice that

$$\overline{S} = \overline{A \cap [x, y]} \subset \overline{A} \cap [x, y].$$

This is easy to verify.

Since $z \in \bar{A}$, then $z \notin B$. So we have $x \leq z < y$.

If $z \notin A$, it follows that $x < z < y$ and $z \notin E$.

If $z \in A$, then $z \notin \overline{B}$. So there exists $r > 0$ such that $N_r(z) \cap B = \emptyset$. Consider $z_1 = z + r/2$. Then $z_1 \notin A$ since $z_1 > z = \sup A$. $z_1 \in N_r(z)$ so $z_1 \notin B$ but $x < z_1 < y$. □

Theorem 4.4.2. ***(Intermediate Value Theorem)*** *Consider a continous mapping $f : [a, b] \to \mathbb{R}$. If $f(a) < f(b)$ and there exists c such that $f(a) < c < f(b)$, then there exists a point $x \in (a, b)$ such that $f(x) = c$.*

Proof. Since $[a, b]$ is connected by Thm.2.8.4, Then by Thm.4.4.1, we have $f([a, b])$ connected. Now this theorem follows if we apply Lem.4.4.1. □

4.5 Uniform Continuity

Definition 4.5.1. Let $f : X \to Y$. We say f is *uniformly continuous* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\forall p, q \in X,\ d_X(p, q) < \delta \Rightarrow d_Y(f(p), f(q)) < \varepsilon.$$

Remark 4.5.1. Compare this definition with continuity. Uniform continuity is a property of a function on a set, whereas continuity is about one point. It would be meaningless to say f is uniformly continuous on some point.

Most importantly, in the definition of continuity, δ depends on p and ε. But for uniform continuity, given ε, we want δ works *for all points p, q of X*.

Every uniformly continuous functions is continuous.

Example 4.5.1. Qiguang Bai's lecture offers some really good examples. $f(x) = x$ is uniformly continuous on $\mathbb{R}$. $f(x) = x^2$ is not uniformly continuous on $(0, 1)$ and on $\mathbb{R}$. But it is uniformly continuous on $[0, 1]$.

The example of $f(x) = x^2$ may inspire us to raise the following conjecture: If $f : X \to Y$ and X is a compact metric space, then maybe f is uniformly continuous. This is actually correct. We will prove this theorem shorly. But first, we will prove a lemma that will be useful.

Lemma 4.5.1. ***(Lebesgue Covering Lemma)*** *If $\{U_\alpha\}$ is a open cover of a compact space X, then there exists $\delta > 0$ such that for all $x \in X$, $N_\delta(x)$ is contained in some U_{α_i}.*

Proof. Since X is compact, there exists a finite subcover $\{U_{\alpha_i}\}_{i=1}^n$. Define the distance between a point $x \in X$ and a set $K \subset X$ to be $\Delta_{x,K}$. This distance is defined as

$$\Delta_{x,K} = \inf_{y \in K} d(x, y).$$

We claim that $\Delta_{x,K}$ is a continuous function of x. Prove: Choose $\delta = \varepsilon$. Then if two points $d(x, x_0) < \delta = \varepsilon$, we have

$$\begin{aligned}
\Delta_{x,K} &= \inf_{y \in K} d(x, y) \\
&\le \inf_{y \in K} (d(x, x_0) + d(x_0, y)) \\
&= d(x, x_0) + \inf_{y \in K} d(x_0, y) \\
&= \varepsilon + \Delta_{x_0,K}.
\end{aligned}$$

Hence we have

$$|\Delta_{x,K} - \Delta_{x_0,K}| < \varepsilon.$$

Since $\Delta_{x,K}, \Delta_{x_0,K} \in \mathbb{R}$, the absolute value is its metric. Hence we proved that $\Delta_{x,K}$ is a continuous function of x.

If $x \in U_{\alpha_i}$, then $\Delta_{x,U^c_{\alpha_i}} > 0$. Otherwise it will be zero. Take

$$\operatorname{avg}(x) = \frac{1}{n} \sum_{i=1}^{n} \Delta_{x,U^c_{\alpha_i}}.$$

It is the average distance of x to all the covers. Since $\operatorname{avg}(x)$ is the finite sum of continuous functions, it is continuous on a compact set X. By Thm.4.3.3, $\operatorname{avg}(x)$ achieves its minimum value δ.

Since $\operatorname{avg}(x) \ge \delta$, it implies that at least one of $\Delta_{x,U^c_{\alpha_i}} \ge \delta$. Then there exists a ball $N_\delta(x) \subset U_{\alpha_i}$. □

Remark 4.5.2. δ is called the Lebesgue number.

Theorem 4.5.1. *Let $f : X \to Y$ be a continuous mapping and X is compact. Then f is uniformly continuous.*

Proof. **(Prof.Su's Proof)** Since f is continuous, for given $x \in X$, there exists δ_x such that

$$d_X(x, p) < \delta_x \Rightarrow d_Y(f(x), f(p)) < \frac{\varepsilon}{2}.$$

Note that $\{N_{\delta_x}(x) : x \in X\}$ is an open cover of X. Since X is compact, there is a finite subcover $\left\{N_{\delta_{x_i}}(x_i) : x \in X\right\}_{i=1}^{n}$. We claim that if there exists a δ such that

$$\forall p, q \in X,\ d_X(p, q) < \delta \Rightarrow p, q \in N_{\delta_i}(x_i) \quad \text{for some } i,$$

Then we have

$$d_Y(f(p), f(q)) \leq d_Y(f(p), f(x)) + d_Y(f(x), f(q)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

And we are done.

By Lem.4.5.1, we know the Lebesgue number δ exists and has such property. Then the theorem follows. □

There is an alternate proof by Rudin, which is much more concise. But the idea is basically the same.

Proof. **(Rudin)** Let $\varepsilon > 0$ be given. Since f is continous, for all $p \in X$ there is a positive number $\phi(p)$ such that

$$q \in X, d_X(p, q) < \phi(p) \quad \text{implies} \quad d_Y(f(p), f(q)) < \frac{\varepsilon}{2}.$$

Let $J(p)$ be the set of all $q \in X$ for which

$$d_X(p, q) < \frac{1}{2}\phi(p).$$

Since $p \in J(p)$, we have $\{J(p)\}_{p \in X}$ an open cover of X. Since X is compact, there exists a finite set of points $p_1, ..., p_n$ in X such that

$$X \subset J(p_1) \cup \cdots \cup J(p_n).$$

Put

$$\delta = \frac{1}{2} \min\left[\phi(p_1), \ldots, \phi(p_n)\right].$$

Then $\delta > 0$. Now let $p, q \in X$ such that $d_X(p, q) < \delta$. Then there exists ain integer m, $1 \leq m \leq n$, such that $p \in J(p_m)$. Hence,

$$d_X(p, p_m) < \frac{1}{2}\phi(p_m),$$

and we also have

$$d_X(q, p_m) \le d_X(p,q) + d_X(p, p_m) < \delta + \frac{1}{2}\phi(p_m) \le \phi(p_m).$$

And by the continuity of f we have

$$d_Y(f(p), f(q)) \le d_Y(f(p), f(p_m)) + d_Y(f(q), f(p_m)) < \varepsilon.$$

This completes the proof. □

4.6 Discontinuities

Definition 4.6.1. We recall the definition of the left limit and the right limit. We use the notation

$$\lim_{t \to x^+} f(t) = f(x+) \qquad \lim_{t \to x^-} f(t) = f(x-).$$

It is clear that the limit of $f(t)$ at x exists if and only if

$$f(x+) = f(x-).$$

Definition 4.6.2. If f is discontinous at x, and both $f(x+)$ and $f(x-)$ exist, then f is said to have a *discontinuity of the first kind*, or a simple discontinuity at x. Otherwise it is of the second kind.

Example 4.6.1. Consider the *Dirichlet function* defined as the following

$$\mathscr{D}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

The Dirichlet function is not continous everywhere.

Proof. For any $c \in \mathbb{R}$ we can find sequence $\{x_n\}$ in $\mathbb{Q}$ and sequence $\{y_n\}$ in $\mathbb{Q}^c$ such that $x_n \to c$ and $y_n \to c$. But $\mathscr{D}(x_n) = 1$ and $\mathscr{D}(y_n) = 0$ for all n. Hence

$$\lim_{n\to\infty} \mathscr{D}(x_n) \neq \lim_{n\to\infty} \mathscr{D}(y_n).$$

This suggest that $\mathscr{D}$ is not continous at c. And since c can be arbitrary, it is not continous everywhere. □

The Dirichlet has the discontinuity of the second kind, since for every half-ball $(p-\delta, p)$ you pick at any $x = p$, the ball will contain both rationals and irrationals. Hence the image

$$\{\mathscr{D}(x) : x \in (p-\delta, p)\}$$

will oscillate between 0 and 1. Hence the condition $d(f(p), f(x)) < \varepsilon$ will not work as soon as you make $\varepsilon < 1$.

Example 4.6.2. Here is a slightly modified version of the Dirichlet function, constructed by Thomae. Consider any rational x, then it can be written as $x = m/n$, where m, n has no common divisors. In other words, $\gcd(m, n) = 1$. Then consider the function

$$\mathscr{T}(x) = \begin{cases} 0 & (x \notin \mathbb{Q}) \\ \frac{1}{n} & \left(x = \frac{m}{n} \in \mathbb{Q}\right) \end{cases}$$

Theorem 4.6.1. *$\mathscr{T}(x)$ is continuous at every irrational point, and discontinous at every rational point. Furthermore, $\mathscr{T}(x)$ has a simple discontinuity at every rational point.*

Proof. First, we prove that $\mathscr{T}(x)$ is continous at every irrational point. Let x be irrational, then for any irrational r, by definition we have

$$|\mathscr{T}(x) - \mathscr{T}(r)| = 0 < \varepsilon$$

for any $\varepsilon > 0$. Hence, we consider the case when r is rational.

Let $\varepsilon > 0$ be given. Choose n such that $1/n < \varepsilon$. Then for $1 \leq j \leq n$ consider a neighborhood $N_{1/j}(x)$ around x. We claim that any such neighborhood contains at most two rationals of the lowest form (no common divisor) having denominator j. Since any two rationals of this kind is

$$\frac{m_j + 1}{j} - \frac{m_j}{j} = \frac{1}{j}$$

apart. And the total interval is $2/j$ away. Hence, if there were three rationals with denominator j in this interval, then x would be rational. Hence we can write

$$\frac{m_j}{j} < x < \frac{m_j + 1}{j}.$$

Then let

$$I_j = \left(\frac{m_j}{j}, \frac{m_j + 1}{j}\right).$$

Now consider the interval

$$I = \bigcap_{j=1}^{n} I_j.$$

By our construction, it is clear that I excludes all rationals of the lowest form with denominator j for all $1 \leq j \leq n$. Also we note that $x \in I$ and I is open since it is the finite intersection of a bunch of open sets. Choose our δ such that

$$N_\delta(x) \subset I.$$

Now, take any rational $r = p/q$ such that $r \in N_\delta(x) \subset I$. Then $q > n$ is necessary, otherwise $q = j$ for some $1 \leq j \leq n$. Hence we have

$$|\mathscr{T}(x) - \mathscr{T}(r)| = \frac{1}{q} < \frac{1}{n} < \varepsilon.$$

Since x is arbitrary, $\mathscr{T}$ is continous at every irrational point.

Now, suppose $r = p/q \in \mathbb{Q}$. Since irrationals [2] are dense in $\mathbb{R}$, so any neighborhood $N_\delta(r)$ will contain $x \notin \mathbb{Q}$. So, no matter how small δ is,

$$|\mathscr{T}(r) - \mathscr{T}(x)| = \frac{1}{q}$$

cannot be arbitrarily small. Hence $\mathscr{T}$ is discontinous at every rational points.

To show that $\mathscr{T}$ has a simple discontinuity at every rational point, we just need to realize that

$$\lim_{x \to y} \mathscr{T}(x) = 0 \quad \forall y \in \mathbb{R}.$$

This is true because if $x \in \mathbb{Q}$ (so we are using rationals to approach y), then the denominator of x will get arbitrarily large as you approach y closer. If $x \notin \mathbb{Q}$, then $|f(x) - 0| = 0 < \varepsilon$ for all $\varepsilon > 0$. Hence the limit of $\mathscr{T}$ exists everywhere. Therefore, its left and right limit must both exist and be equal. This completes the proof. $\square$

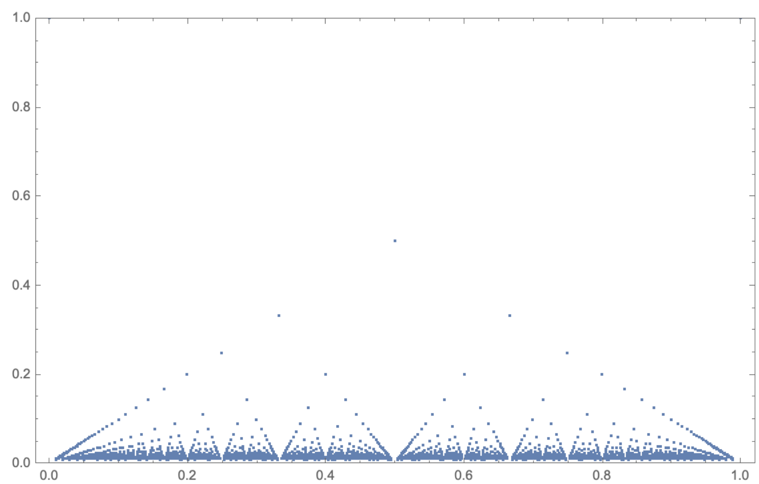


Figure 4.1: Plot $\mathscr{T}(x)$ using MATHEMATICA.

For more examples on discontinous function, you can look at examples in Rudin (pg95).

4.7 Monotonic Functions

We will now mention a few things about monotonic functions, either monotonically increasing or monotonically decreasing.

Definition 4.7.1. Let $f : (a,b) \to \mathbb{R}$. Then f is said to be *monotonically increasing* on (a,b) if $a < x < y < b$ implies $f(x) \leq f(y)$. f is said to be *monotonically decreasing* on (a,b) if $a < x < y < b$ implies $f(x) \geq f(y)$.

[2]Proof: We know that $\mathbb{Q}$ is dense on $\mathbb{R}$, hence we have $a < c < b$ for any $a, b \in \mathbb{R}$ and $c \in \mathbb{Q}$. Then we have $a + \sqrt{2} < c + \sqrt{2} < b + \sqrt{b}$. This is equivalent of saying $d < e < f$ for $d, f \in \mathbb{R}$ and $e \notin \mathbb{Q}$. Done.

Theorem 4.7.1. *Let f be a monotonically increasing function on (a, b), then $f(x+)$ and $f(x-)$ exists at every point of $x \in (a, b)$. More precisely,*

$$\sup_{a<t<x} f(t) = f(x-) \leq f(x) \leq f(x+) = \inf_{x<t<b} f(t).$$

Furthermore, if $a < x < y < b$, then

$$f(x+) \leq f(y-).$$

Proof. By hypothesis, the set of numbers $f(t)$ where $a < t < x$ is bounded above by $f(x)$, and therefore has a sup $f(t) = A$. Evidently, $A \leq f(x)$. We now show that $A = f(x-)$.

Let $\varepsilon > 0$ be given. It follows from the definition of A that there exists $\delta > 0$ such that $a < x - \delta < x$ and

$$A - \varepsilon < f(x - \delta) \leq A.$$

Since f is monotonic, we have

$$f(x - \delta) \leq f(t) \leq A \quad (x - \delta < t < x).$$

Combining the two inequalities above we have

$$|f(t) - A| < \varepsilon \quad (x - \delta < t < x).$$

Hence $f(x-) = A$.

The second half is proved in exactly the same way.

Next, if $a < x < y < b$, we see that

$$f(x+) = \inf_{x<t<b} f(t) = \inf_{x<t<y} f(t).$$

Similarly,

$$f(y-) = \sup_{a<t<y} f(t) = \sup_{x<t<y} f(t).$$

These two equations give

$$f(x+) \leq f(y-).$$

□

Corollary 4.7.1. *Monotonic functions have no discontinuity of the second kind.*

Theorem 4.7.2. *Let $f : (a, b) \to \mathbb{R}$, then the set of points $x \in (a, b)$ at which f is discontinous is at most countable.*

Proof. Suppose f is increasing. Let E be the set of points at which f is discontinous. Then with every point $x \in E$ we associate a rational number $r(x)$ such that

$$f(x-) < r(x) < f(x+).$$

Since $x_1 < x_2$ implies $f(x_1+) \leq f(x_2-)$, we see that $r(x_1) \neq r(x_2)$ if $x_1 \neq x_2$. Hence we have established a 1-1 correspondence between E and a subset of $\mathbb{Q}$, and the later is countable. Hence the theorem follows. □

Chapter 5

Differentiation

This most beautiful system could only proceed from the dominion of an intelligent and powerful being.

-Isaac Newton, The Principia

5.1 The Derivative of a Real Function

Definition 5.1.1. Let $f : [a, b] \to \mathbb{R}$. Define the limit

$$f'(x) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x} \quad (a < t < b, t \neq x).$$

If this limit exists, call $f'(x)$ the derivative of f at x. If f' is defined at point x, say f is *differentiable* at x. If f' is defined at every point of a set $E \subset [a, b]$, then say f is differentiable on E.

Theorem 5.1.1. *Let f be defined on $[a, b]$. If f is differentiable at $x \in [a, b]$, then f is continous at x.*

Proof. We want to prove that $\lim_{t \to x} f(t) - f(x) = 0$. But this is obvious since

$$f(t) - f(x) = \frac{f(t) - f(x)}{t - x} \cdot (t - x) \to f'(x) \quad 0 = 0.$$

□

Remark 5.1.1. The converse of this theorem is not true. We can construct continuous function that is not differentiable at some point. In particular, in the next example we will see a function can be continous everywhere but nowhere differentiable.

Example 5.1.1. Consider the following function, defined as a Fourier series:

$$W(x) = \sum_{n=0}^{\infty} a^n \cos\left(b^n \pi x\right)$$

where $0 < a < 1$ and b is a positive odd integer such that $ab > 1 + \frac{3}{2}\pi$. This function is called the *Weierstrass function*. It is continous everywhere but nowhere differentiable.

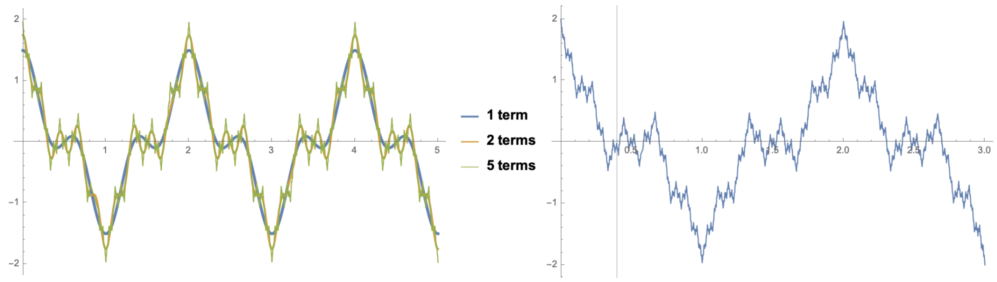


Figure 5.1: Plot $W(x)$ using MATHEMATICA. Here we choose $a = 1/2$ and $b = 3$. On the left are the partial sums of $1, 2, 5$ terms. On the right is the partial sum of 10 terms.

Theorem 5.1.2. *Suppose f and g are defind on interval $[a, b]$ and are differentiable at a point $x \in [a, b]$. Then $f + g$, fg, f/g are differentiable at x, and*

$$\begin{array}{ll} (a) & (f+g)'(x) = f'(x) + g'(x) \\ (b) & (fg)'(x) = f'(x)g(x) + f(x)g'(x) \\ (c) & \left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x) - g'(x)f(x)}{g^2(x)} \end{array}$$

In (c), we assume that $g(x) \neq 0$.

Proof. Basic calculus. Omitted. $\square$

Theorem 5.1.3. *Suppose f is continuous on $[a, b]$ and $f'(x)$ exists at some point $x \in [a, b]$. g is defined on an interval I which contains the range of f, and g is differentiable at the point $f(x)$. If*

$$h(t) = g(f(t)) \quad (a \leq t \leq b)$$

then h is differentiable at x, and

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

Proof. This is the so-called chain rule. Proof omitted. $\square$

Example 5.1.2. There are functions that are differentiable but whose derivative is not continous. Consider

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & (x \neq 0) \\ 0 & (x = 0) \end{cases}$$

Then taking the derivative we obtain

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x} \quad (x \neq 0).$$

At $x = 0$ we obtain, by definition,

$$\left| \frac{f(t) - f(0)}{t - 0} \right| = \left| t \sin \frac{1}{t} \right| \leq |t| \quad (t \neq 0).$$

Letting $t \to 0$ we see that $f'(0) = 0$. But f' is not a continous function since $\cos(1/x)$ does not tend to a limit as $x \to 0$.

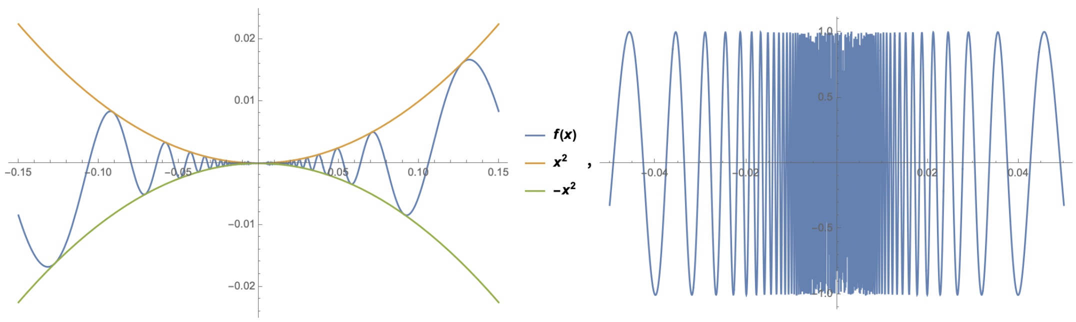


Figure 5.2: Plot $f(x)$ using MATHEMATICA. The left picture is $f(x)$ and the right picture is $f'(x)$.

5.2 Mean Value Theorems

The mean value theorem is very important in calculus because it connects the values of a function and its derivative without using limit.

Definition 5.2.1. Let f be a real function defined on a metric space X. We say that f has a local maximum at a point $p \in X$ if there exists $\delta > 0$ such that $f(q) \leq f(p)$ for all $q \in X$ with $d(p,q) < \delta$.

Theorem 5.2.1. *Let $f : [a,b] \to \mathbb{R}$. If f has a local maximum at $x \in (a,b)$, and if $f'(x)$ exists, then $f'(x) = 0$. Same is true for a local minimum.*

Proof. We just need to consider the derivative as a limit. If $t \to x$ on the left, we have

$$\frac{f(t) - f(x)}{t - x} \geq 0.$$

In the limit case we see that $f'(x) \geq 0$. If $t \to x$ on the right, we have

$$\frac{f(t) - f(x)}{t - x} \leq 0.$$

Hence in the limit case $f'(x) \leq 0$. Therefore, we conclude $f'(x) = 0$. $\square$

Before proving the mean value theorem, we need the following lemma.

Lemma 5.2.1. ***(Rolle's Theorem)*** *Let $f : [a,b] \to \mathbb{R}$ be a continous function which is differentiable in (a,b), and $f(a) = f(b)$. Then there is a point $c \in (a,b)$ at which $f'(c) = 0$.*

Proof. Since f is a continous function defined on $[a,b]$, a compact metric space, we know from Thm.4.3.3 that f must achieve its maximum and minimum on $[a,b]$. There are two cases:
(1) The the two extreme values $f(a) = \max f = \min f = f(b)$, without loss of generality, then f is a constant. Hence $f'(x) = 0$ everywhere. Done.
(2) If f is not a constant, then $\max f \neq \min f$. Suppose, then, without loss of generality, $\max f$ is at one of the endpoint a or b, then $\min f < f(a) = f(b) = \max f$. Hence the minimum cannot be at the endpoint. We conclude then *at least one of the extreme value is in* (a,b). And by Thm.5.2.1, we are done. $\square$

Now we are ready to prove the mean value theorem. We will first look at something called the general mean value theorem, and the mean value theorem immediately follows.

Theorem 5.2.2. ***(General Mean Value Theorem)*** *If $f, g : [a, b] \to \mathbb{R}$ are continous on $[a, b]$ and differentiable in (a, b), then there exists a point $x \in (a, b)$ at which*

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

Proof. Put

$$h(t) = [f(b) - f(a)]g(t) - [g(b) - g(a)]f(t) \quad (a \leq t \leq b).$$

Then h is continous on $[a, b]$ and differentiable in (a, b). Also note that

$$h(a) = f(b)g(a) - f(a)g(b) = h(b).$$

By Rolle's Theorem (Thm.5.2.1), there exists a $x \in (a, b)$ at which

$$h'(x) = [f(b) - f(a)]g'(x) - [g(b) - g(a)]f'(x) = 0.$$

Hence the theorem follows. □

Theorem 5.2.3. ***(Mean Value Theorem)*** *If $f : [a, b] \to \mathbb{R}$ is continous on $[a, b]$ and differentiable in (a, b), then there exists a point $x \in (a, b)$ at which*

$$f(b) - f(a) = (b - a)f'(x).$$

Proof. Take $g(x) = x$ in Thm.5.2.2, and the theorem follows immediately. □

Theorem 5.2.4. *Suppose f is differentiable in (a, b).*
(a) If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotonically increasing.
(b) If $f'(x) = 0$ for all $x \in (a, b)$, then f is a constant.
(c) If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.

Proof. Apply the mean value theorem (Thm.5.2.3) on the equation

$$f(x_2) - f(x_1) = (x_2 - x_1) f'(x).$$

□

5.3 Intermediate Value Theorem for Derivative

Theorem 5.3.1. ***(Intermediate Value Theorem for Derivatives)***
Suppose $f : [a, b] \to \mathbb{R}$ is a differentiable function on $[a, b]$ and suppose $f'(a) < \lambda < f'(b)$. Then there exists a point $c \in (a, b)$ such that $f'(c) = \lambda$.

Proof. Consider the function

$$g(x) = \lambda x - f(x).$$

Then we have

$$g'(x) = \lambda - f'(x).$$

We have

$$g'(a) = \lambda - f'(a) > 0$$
$$g'(b) = \lambda - f'(b) < 0.$$

We know g must achieve its minimum on $[a,b]$ by Thm.4.3.3. By Thm.5.2.1, we know the minimum c cannot be at the endpoints a, b. Hence $c \in (a,b)$. And we have $f'(c) = \lambda$. □

Corollary 5.3.1. *If f is differentiable on $[a,b]$, then f' cannot have any simple discontinuities on $[a,b]$.*

Proof. We prove by contradiction. Suppose $g = f'$ has a simple discontinuity at $c \in [a,b]$. Then this would mean either

$$g(c+) \neq g(c) \quad \text{or} \quad g(c-) \neq g(c)$$

is ture. Without loss of generality, assume that $g(c+) \neq g(c)$. Further, assume that $g(c+) > g(c)$. Then choose λ such that $g(c) < \lambda < g(c+)$. Pick $\varepsilon = g(c+) - \lambda$. Then by the definition of $f(c+)$, there exists $\delta > 0$ such that

$$x - c < \delta \Rightarrow g(x) \in (f(c+) - \varepsilon, f(c+) + \varepsilon).$$

However, this is equivalent as saying

$$x - c < \delta \Rightarrow \lambda < g(x) < g(c+) \qquad \text{for } x \in (c, c+\delta).$$

Then consider the interval $g : [c, c + \delta/2]$, we have

$$g(c) < \lambda < g(c + \frac{\delta}{2}).$$

Hence,

$$g(x) \neq \lambda \ \forall x \in (c, c + \frac{\delta}{2}).$$

But this contradicts Thm.5.3.1, since it suggests that there must be a $x \in (c, c + \delta/2)$ such that $f'(x) = \lambda$. □

5.4 L'Hospital's Rule

[1] Starting from this section, we will be looking at some interesting and quite important applications of the mean value theorem. One of the most famous one is the L'Hospital's rule, which helps us deal with indeterminate limits.

[1] Starting from this section, all the notes will refer to Qiguang Bai's lectures at NCTU. Su has finished his lectures.

Theorem 5.4.1. *Suppose f, g are real and differentiable in (a, b) and $g'(x) \neq 0$ for all $x \in (a, b)$, where $-\infty \leq a \leq b \leq \infty$. Suppose*

$$\frac{f'(x)}{g'(x)} \to A \text{ as } x \to a.$$

If

$$f(x) \to 0 \text{ and } g(x) \to 0 \text{ as } x \to a,$$

or if

$$g(x) \to +\infty \quad \text{as } x \to a,$$

then

$$\frac{f(x)}{g(x)} \to A \text{ as } x \to a.$$

Proof. Since $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, then both limit exist, so do their left and right limits. Hence the functions cannot have the discontinuity of the first kind (that means they must have removable discontinuity). Hence we assign

$$f(a) = g(a) = 0.$$

This is fine since the limit as $x \to a$ has nothing to do with the behaviors of the functions at a. By our modification, then, f, g are continous at a. Or, in other words: $f, g : [a, b) \to \mathbb{R}$. Now consider $x \in (a, b)$ and $f, g : [a, x] \to \mathbb{R}$. We know that $g'(t) \neq 0$ for all $t \in (a, x)$. Hence we can apply the mean value theorem.

By the general mean value theorem (Thm.5.2.2), we know there exists $c \in (a, x)$ such that

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

Here, c actually depends on x. By our assignment, we know that $f(a) = g(a) = 0$. Hence

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)}.$$

Now, as $x \to a$, we see that $c \to a$. Hence the limit on the right hand side exists and goes to A. Hence the limit on the left hand side exists and also goes to A. This concludes the proof. □

5.5 Taylor's Theorem

Taylor's theorem is another important application of the mean value theorem. Taylor's theorem gives us a polynomial involving f that can actually approximate f very well.

Theorem 5.5.1. *Suppose f is a real function on $[a, b]$, n is a positive integer, $f^{(n-1)}$ is continous on $[a, b]$, $f^{(n)}(t)$ exists for every $t \in (a, b)$. Let α, β be distinct points of $[a, b]$, and define*

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(t - \alpha)^k.$$

Then there exists a point x between α and β such that

$$f(\beta) = P(\beta) + \frac{f^{(n)}(x)}{n!}(\beta - \alpha)^n.$$

Remark 5.5.1. For $n = 1$, this is just the mean value theorem. In general, if we know the bounds on $|f^{(n)}(x)|$, then we can approximate f by a polynomial of degree $n-1$ and we can estimate the error.

Proof. Let M be the number defined by

$$f(\beta) = P(\beta) + M(\beta - \alpha)^n$$

and put

$$g(t) = f(t) - P(t) - M(t-\alpha)^n \quad (a \le t \le b).$$

We have to show that $n!M = f^{(n)}(x)$ for some x between α and β. By taking the derivative of g for n times, we see that

$$g^{(n)}(t) = f^{(n)}(t) - n!M \quad (a < t < b).$$

Hence we want to show that $g^{(n)}(x) = 0$ for some x inbetween α and β.

By simple observation, we notice that

$$g(\alpha) = g'(\alpha) = \cdots = g^{(n-1)}(\alpha) = 0.$$

We also notice that

$$\begin{aligned} g(\beta) &= f(\beta) - P(\beta) - M(\beta - \alpha)^n \\ &= f(\beta) - P(\beta) - \frac{f(\beta) - P(\beta)}{(\beta - \alpha)^n}(\beta - \alpha)^n = 0. \end{aligned}$$

Hence by the mean value theorem (Thm.5.2.3), there exists x_1 inbetween α and β such that $g'(x_1) = 0$.

Since $g'(\alpha) = 0$, again by the mean value theorem, there exists a x_1 inbetween α and x_1 such that $g'(x_2) = 0$... Repeat this process n times, we see that there exists x_n inbetween α and x_{n-1} such that $g'(x_n) = 0$. Since x_n is inbetween α and x_{n-1}, it is also inbetween α and β. $\square$

Chapter 6

The Riemann-Stieltjes Integral

Every feeling is the perception of a truth.

-Gottfried Wilhelm Leibniz

6.1 Definition and Existence of the Integral

Definition 6.1.1. (Riemann Integral) Let $[a, b]$ be a given interval. By a partition P of $[a, b]$ we mean a finite set of points $x_0, x_1, ..., x_n$, where

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

We write

$$\Delta x_i = x_i - x_{i-1} \quad (i = 1, \ldots, n).$$

Now suppose f is a *bounded* real function defined on $[a, b]$. Corresponding to each partition P of $[a, b]$ we put

$$\begin{aligned}
M_i &= \sup f(x) \quad (x_{i-1} \leq x \leq x_i) \\
m_i &= \inf f(x) \quad (x_{i-1} \leq x \leq x_i) \\
U(P, f) &= \sum_{i=1}^{n} M_i \Delta x_i \\
L(P, f) &= \sum_{i=1}^{n} m_i \Delta x_i.
\end{aligned}$$

And finally,

$$\overline{\int_a^b} f dx = \inf U(P, f), \tag{6.1}$$

$$\underline{\int_a^b} f dx = \sup L(P, f), \tag{6.2}$$

where the inf and sup are taken over all partitions P on $[a,b]$. The left of (6.1) and (6.2) are called the upper and lower Riemann integrals of f over $[a,b]$, respectively.

If the upper and lower integrals are equal, we say f is *Riemann-integrable* on $[a,b]$. We write $f \in \mathscr{R}$, where $\mathscr{R}$ denotes the set of Riemann-integrable functions. And we denote the common value of (6.1) and (6.2) by

$$\int_a^b f dx.$$

Since f is bounded, there exist two numbers m, M such that

$$m \leq f(x) \leq M \quad (a \leq x \leq b).$$

Hence, for every P,

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

Hence the numbers $L(P, f)$ and $U(P, f)$ form a bounded set. This shows that *the upper and lower integrals are defined for every bounded function f.*

Definition 6.1.2. (Stieltjes Integral) Let α be a monotonically increasing function on $[a,b]$ (Hence α is bounded on $[a,b]$). Corresponding to each partition P we write

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}).$$

clearly $\Delta\alpha_i \geq 0$. For any real function f that is bounded on $[a,b]$ we put

$$U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta\alpha_i,$$

$$L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta\alpha_i.$$

where M_i, m_i has the same meaning as in the last definition. Define

$$\overline{\int_a^b} f d\alpha = \inf U(P, f, \alpha), \tag{6.3}$$

$$\underline{\int_a^b} f d\alpha = \sup L(P, f, \alpha). \tag{6.4}$$

If the left of (6.3) and (6.4) are equal, we denote their common value by

$$\int_a^b f d\alpha \tag{6.5}$$

or sometimes by

$$\int_a^b f(x) d\alpha(x).$$

This is the Stieltjes integral of f with respect to α over $[a,b]$.

If (6.5) exists, that is, if (6.3) and (6.4) are equal, we say f is integrable with respect to α. Write $f \in \mathscr{R}(\alpha)$.

Remark 6.1.1. By taking $\alpha(x) = x$, we see that the Riemann integeral is a special case of the Stieltjes integral. Hence the Stieltjes integral is a more general definition.

Now we shall investigate the existence of (6.5). From now on, we shall assume that f is real and bounded, and α is monotonically increasing on $[a, b]$. And, when there is no misunderstanding, we will write $\int$ in place of $\int_a^b$.

Definition 6.1.3. The partition P^* is a refinement of P if $P^* \subset P$. Given two partitions, P_1 and P_2, we say that P^* is their common refinement if $P^* = P_1 \cup P_2$.

Theorem 6.1.1. *If P^* is a refinement of P, then*

$$L(P, f, \alpha) \le L\left(P^*, f, \alpha\right)$$

and

$$U\left(P^*, f, \alpha\right) \le U(P, f, \alpha).$$

Proof. First suppose P^* contains just one point more than P, which we call x^*. Suppose $x_{i-1} < x^* < x_i$. Put

$$\begin{array}{ll} w_1 = \inf f(x) & (x_{i-1} \le x \le x^*) \\ w_2 = \inf f(x) & (x^* \le x \le x_i) \end{array}.$$

Clearly $w_1, w_2 \ge m_i$, where

$$m_i = \inf f(x) \quad (x_{i-1} \le x \le x_i) .$$

Hence

$$\begin{aligned} & L\left(P^*, f, \alpha\right) - L(P, f, \alpha) \\ & \quad = w_1 \left[\alpha\left(x^*\right) - \alpha\left(x_{i-1}\right)\right] + w_2 \left[\alpha\left(x_i\right) - \alpha\left(x^*\right)\right] - m_i \left[\alpha\left(x_i\right) - \alpha\left(x_{i-1}\right)\right] \\ & \quad = \left(w_1 - m_j\right) \left[\alpha\left(x^*\right) - \alpha\left(x_{i-1}\right)\right] + \left(w_2 - m_i\right) \left[\alpha\left(x_i\right) - \alpha\left(x^*\right)\right] \ge 0 \end{aligned}$$

If there are k points more in the refinement, repeat this reasoning k times. The proof of the upper limit is similar. □

Theorem 6.1.2.

$$\underline{\int} f d\alpha \le \overline{\int} f d\alpha.$$

Proof. Let P^* be the common refinement of two partitions P_1, P_2. By 6.1, we have

$$L\left(P_1, f, \alpha\right) \le L\left(P^*, f, \alpha\right) \le U\left(P^*, f, \alpha\right) \le U\left(P_2, f, \alpha\right).$$

Hence,

$$L\left(P_1, f, \alpha\right) \le U\left(P_2, f, \alpha\right).$$

If P_2 is fixed and sup is taken over all P_1, we have

$$\underline{\int} f d\alpha \le U\left(P_2, f, \alpha\right).$$

The theorem then follows by taking inf over all P_1. □

Theorem 6.1.3. *$f \in \mathscr{R}(\alpha)$ on $[a, b]$ if and only if for every $\varepsilon > 0$ there exists a partition P such that*

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

Proof. Given $\varepsilon > 0$, suppose $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ is holds for some P. For every P we have

$$L(P, f, \alpha) \leq \underline{\int} f d\alpha \leq \overline{\int} f d\alpha \leq U(P, f, \alpha).$$

Thus the assumption implies

$$0 \leq \overline{\int} f d\alpha - \underline{\int} f d\alpha < \varepsilon.$$

This is true for all $\varepsilon > 0$. Hence $f \in \mathscr{R}(\alpha)$.

Conversely, suppose $f \in \mathscr{R}(\alpha)$, let $\varepsilon > 0$ be given. Then there exists partitions P_1, P_2 such that

$$U\left(P_2, f, \alpha\right) - \int f d\alpha < \frac{\varepsilon}{2},$$

which is true since the integral is the inf U. And something bigger by $\varepsilon/2$ will not be the greatest lower bound and some U will be smaller. Similarly, it also follows that

$$\int f d\alpha - L\left(P_1, f, \alpha\right) < \frac{\varepsilon}{2}.$$

Choose P to be the common refinement of P_1, P_2. Then Thm.6.1 says

$$U(P, f, \alpha) \leq U\left(P_2, f, \alpha\right) < \int f d\alpha + \frac{\varepsilon}{2} < L\left(P_1, f, \alpha\right) + \varepsilon \leq L(P, f, \alpha) + \varepsilon.$$

Hence we conclude

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

□

Theorem 6.1.4. *If $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ holds for some P and some ε, then*
(a) $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$ holds for every refinement of P.
(b) If the P that works is $P = \{x_0, ..., x_n\}$ and if $s_i, t_i \in [x_{i-1}, x_i]$, then

$$\sum_{i=1}^{n} \left|f\left(s_i\right) - f\left(t_i\right)\right| \Delta\alpha_i < \varepsilon.$$

(c) If $f \in \mathscr{R}(\alpha)$ and (b) holds, then

$$\left|\sum_{i=1}^{n} f\left(t_i\right) \Delta\alpha_i - \int_a^b f d\alpha\right| < \varepsilon.$$

Proof. (a) Since $L(P, f, \alpha) \le L(P^*, f, \alpha) \le U(P^*, f, \alpha) \le U(P, f, \alpha)$, the theorem follows.
(b) Since $|f(s_i) - f(t_i)| \le M_i - m_i$, we have

$$\sum_{i=1}^{n} |f(s_i) - f(t_i)| \, \Delta\alpha_i \le U(P, f, \alpha) - L(P, f, \alpha),$$

and the theorem follows
(c) These two inequalities

$$L(P, f, \alpha) \le \sum f(t_i) \, \Delta\alpha_i \le U(P, f, \alpha)$$
$$L(P, f, \alpha) \le \int f d\alpha \le U(P, f, \alpha)$$

proves the theorem immediately. □

Theorem 6.1.5. *If f is continous on $[a, b]$ then $f \in \mathscr{R}(\alpha)$ on $[a, b]$.*

Proof. Given $\varepsilon > 0$, choose η so that

$$[\alpha(b) - \alpha(a)]\eta < \varepsilon,$$

this way we do not worry about the case when $\alpha(b) = \alpha(a)$ is divided by ε. Since $f : [a, b]$ is uniformly continuous, we have

$$|x - t| < \delta \Rightarrow |f(x) - f(t)| < \eta.$$

Let P be the partition such that $\Delta x_i < \delta$ for all i, then $|M_i - m_i| < \eta$ for all i and therefore

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \sum_{i=1}^{n} (M_i - m_i) \, \Delta\alpha_i \\ &\le \eta \sum_{i=1}^{n} \Delta\alpha_i = \eta[\alpha(b) - \alpha(a)] < \varepsilon. \end{aligned}$$

□

Theorem 6.1.6. *If f is monotonic on $[a, b]$ and α is continous and monotonically increasing on $[a, b]$, then $f \in \mathscr{R}(\alpha)$.*

Proof. Given $\varepsilon > 0$, choose the partition to be n equally seperated points such that

$$\Delta\alpha_i = \frac{\alpha(b) - \alpha(a)}{n} < \frac{\varepsilon}{f(b) - f(a)}.$$

Hence we have

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] \\ &= \frac{\alpha(b) - \alpha(a)}{n} \cdot [f(b) - f(a)] < \varepsilon. \end{aligned}$$

Note that since f is monotonically increasing (W.L.O.G), we have

$$M_i = f(x_i) \qquad m_i = f(x_{i-1}).$$

□

Remark 6.1.2. From the above two theorems, we can see that to prove a function f is integrable, our central idea is to prove that

$$U(P, f, \alpha) - L(P, f, \alpha) = \sum_{i=1}^{n} (M_i - m_i)\Delta\alpha_i$$

is small. If f is continous, then we can guarantee that $M_i - m_i$ is small, hence the sum will be small. If f is monotonic and α is continous, then $\Delta\alpha_i$ will be small and hence the sum will be small. In other words, two functions f and α support each other and in a proof of integrability, we will always want to restrict either f or α, and sometimes both.

Theorem 6.1.7. *Suppose f is bounded on $[a, b]$ and has only finitely many discontinuity on $[a, b]$, and α is continous everywhere at which f is discontinous. Then $f \in \mathscr{R}(\alpha)$.*

Proof. Let $\varepsilon > 0$ be given. Let E be the sey of all points on $[a, b]$ where f is discontinous. We can cover E with finitely many intervals $[u_j, v_j]$ such that every point in E lies in the interior of $[u_j, v_j]$ for some j. And choose the length of these intervals such that

$$\sum_j \alpha(v_j) - \alpha(u_j) < \varepsilon,$$

which can be done since α is continuous at these points.

Let

$$I = \bigcup_j (u_j, v_j)$$

and let

$$K = [a, b] \setminus I.$$

We immediately notice that K is compact. Hence f is uniformly continous on K, which means that for $s, t \in K$, we have

$$|s - t| < \delta \Rightarrow |f(s) - f(t)| < \varepsilon.$$

Now, form a partition P as follows. Let the endpoints $u_j, v_j \in P$ for all j but $(u_j, v_j) \cap P = \emptyset$ for all j. If x_{i-1} is not one of u_j, let

$$\Delta x_i = x_i - x_{i-1} < \delta.$$

Notice that

$$M_i - m_i \leq 2M \qquad \text{where } M = \sup|f(x)|.$$

Then we have

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \sum_{\text{pts in } K} (M_i - m_i)\Delta\alpha_i + \sum_{\text{pts in } I} (M_i - m_i)\Delta\alpha_i \\ &= \varepsilon[\alpha(b) - \alpha(a)] + 2M\varepsilon. \end{aligned}$$

Since ε can be arbitrarily small, we conclude that $f \in \mathscr{R}(\alpha)$. □

Theorem 6.1.8. *Suppose $f \in \mathscr{R}(\alpha)$ on $[a,b]$ and $m \leq f \leq M$. Let ϕ be a continous function on $[m,M]$ and let $h(x) = \phi(f(x))$ on $[a,b]$. Then $h \in \mathscr{R}(\alpha)$ on $[a,b]$.*

Proof. Since ϕ is uniformly continuous on $[m,M]$, then for $s,t \in [m,M]$, we have

$$|s-t| < \delta \Rightarrow |\phi(s) - \phi(t)| < \varepsilon.$$

Choose δ such that $\delta < \varepsilon$. Let M_i, m_i be the sup and inf of the function f and let M_i^*, m_i^* be the sup and inf of the function h. Let

$$A = \{i : M_i - m_i < \delta\},$$
$$B = \{i : M_i - m_i \geq \delta\}.$$

For $i \in A$, consider $s,t \in [x_{i-1}, x_i]$, then we know that

$$|f(s) - f(t)| \leq |M_i - m_i| < \delta \Rightarrow |\phi(f(s)) - \phi(f(t))| < \varepsilon \Rightarrow M_i^* - m_i^* < \varepsilon.$$

For $i \in B$, since $f \in \mathscr{R}(\alpha)$, there exists a partition P such that

$$U(P,f,\alpha) - L(P,f,\alpha) < \delta^2.$$

Put $K = \sup_{m \leq t \leq M} |\phi(t)|$. Since $M_i^* - m_i^* \leq 2K$ and $M_i - m_i \geq \delta$, we have

$$\delta \sum_{i \in B} \Delta\alpha_i \leq \sum_{i \in B} (M_i - m_i)\,\Delta\alpha_i < \delta^2.$$

Hence,

$$\sum_{i \in B} \Delta\alpha_i < \delta.$$

Finally,

$$\begin{aligned} U(P,h,\alpha) - L(P,h,\alpha) &= \sum_{i \in A} \left(M_i^* - m_i^*\right)\Delta\alpha_i + \sum_{i \in B} \left(M_i^* - m_i^*\right)\Delta\alpha_i \\ &\leq \varepsilon[\alpha(b) - \alpha(a)] + 2K\delta < \varepsilon[\alpha(b) - \alpha(a) + 2K]. \end{aligned}$$

Since ε is arbitrarily, the theorem follows. □

6.2 Properties of the Integral

In this section we discuss various properties of the Stieltjes integral.

Theorem 6.2.1. $f_1, f_2 \mathscr{R}(\alpha) \Rightarrow f_1 + f_2 \in \mathscr{R}(\alpha)$.

Proof. We first prove that

$$\inf(f_1 + f_2) \geq \inf(f_1) + \inf(f_2).$$

Since $\inf(f_1) \leq f_1$ and $\inf(f_2) \leq f_2$ on $[a,b]$, we have

$$\inf(f_1) + \inf(f_2) \leq f_1 + f_2.$$

Since any lower bound is less than the greatest lower bound, the claim follows. Similarly, we can prove that

$$\sup(f_1 + f_2) \le \sup(f_1) + \sup(f_2).$$

Then let $f = f_1 + f_2$ and for any partition P we have

$$\begin{aligned} L(P, f_1, \alpha) + L(P, f_2, \alpha) &\le L(P, f, \alpha) \\ \le U(P, f, \alpha) &\le U(P, f_1, \alpha) + U(P, f_2, \alpha). \end{aligned}$$

These inequalities hold if we let P to be the common refinement of P_1 and P_2. Hence the inequality implies

$$U(P, f, \alpha) - L(P, f, \alpha) < 2\varepsilon,$$

and the theorem follows. □

Theorem 6.2.2.

$$\int_a^b (f_1 + f_2)\, d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha.$$

Proof. For the same P as that in the last theorem (let P be the common refinement of P_1, P_2 that makes f_1, f_2 integrable), we have

$$U\left(P, f_j, \alpha\right) < \int f_j d\alpha + \varepsilon \quad (j = 1, 2).$$

This implies that

$$U(P, f, \alpha) - L(P, f, \alpha) < 2\varepsilon.$$

□

there are other properties of the integral. Since the proof is similar to these two proofs, we simply omit the proof and directly presents these properties.

Theorem 6.2.3. *(a)* $\int_a^b cf d\alpha = c \int_a^b f d\alpha$.
(b) If $f_1 \le f_2$ *on* $[a, b]$, *then* $\int_a^b f_1 d\alpha \le \int_a^b f_2 d\alpha$.
(c) If $f \in \mathscr{R}(\alpha)$ *on* $[a, b]$ *and if* $a < c < b$, *then* $f \in \mathscr{R}(\alpha)$ *on* $[a, c]$ *and* $[c, b]$ *and* $\int_a^c f d\alpha + \int_c^b f d\alpha = \int_a^b f d\alpha$.
(d) If $f \in \mathscr{R}(\alpha)$ *on* $[a, b]$ *and* $|f(x)| \le M$ *then* $\left|\int_a^b f d\alpha\right| \le M[\alpha(b) - \alpha(a)]$.
(e) If $f \in \mathscr{R}(\alpha_1)$ *and* $f \in \mathscr{R}(\alpha_2)$ *then* $f \in \mathscr{R}(\alpha_1 + \alpha_2)$ *and* $\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2$.
(f) If $f \in \mathscr{R}(\alpha)$ *and c is a positive constant, then* $f \in \mathscr{R}(c\alpha)$ *and* $\int_a^b f d(c\alpha) = c \int_a^b f d\alpha$.

Theorem 6.2.4. *If* $f, g \in \mathscr{R}(\alpha)$ *on* $[a, b]$, *then*
(a) $fg \in \mathscr{R}(\alpha)$;
(b) $|f| \in \mathscr{R}(\alpha)$ *and*

$$\left|\int_a^b f d\alpha\right| \le \int_a^b |f| d\alpha.$$

Proof. Take $\phi(t) = t^2$ and Thm.6.1.8 shows that $f^2 \in \mathscr{R}(\alpha)$ if $f \in \mathscr{R}(\alpha)$. The identity

$$4fg = (f+g)^2 - (f-g)^2$$

proves (a) immediately.
(b) Take $\phi(t) = |t|$ and Thm.6.1.8 shows that $|f| \in \mathscr{R}(\alpha)$ if $f \in \mathscr{R}(\alpha)$. Choose $c = \pm 1$ so that

$$c\int f d\alpha \geq 0.$$

Then

$$\left|\int f d\alpha\right| = c\int f d\alpha = \int c f d\alpha \leq \int |f| d\alpha$$

since $cf \leq |f|$. □

After discussing some basic properties of the Stieltjes integrals, we begin to consider evaluating some simple integrals. We start from the following definition.

Definition 6.2.1. The unit step function is defined by

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

Theorem 6.2.5. *Suppose $a < s < b$ and f is bounded on $[a,b]$. If f is continuous at s and $\alpha(x) = I(x-s)$, then*

$$\int_a^b f d\alpha = f(s).$$

Proof. Let $s \in [x_{i-1}, x_i]$. Then $\Delta\alpha_k = 0$ for all k except when $k = 1$. In this case $\Delta\alpha_i = 1$. Then

$$U(P, f, \alpha) = M_2, \quad L(P, f, \alpha) = m_2.$$

Since f is continous at s, we see that

$$M_2, m_2 \to f(s) \quad \text{when} \quad \Delta\alpha_i \to 0.$$

□

Lemma 6.2.1. *Suppose $c_k \geq 0$ for $k = 1, 2, 3...n$. and $\{s_n\}$ is a sequence of distinc points on $[a,b]$. Let*

$$\alpha(x) = \sum_{k=1}^{n} c_n I(x - s_n).$$

Let f be continuous on $[a,b]$, then

$$\int_a^b f d\alpha = \sum_{k=1}^{n} c_n f(s_n).$$

Proof. Based on Thm.6.2.5 and the properties of the integral, for finite sums, we know

$$\begin{aligned}\int f d\alpha &= \int f d\left(\sum c_n I(x-s_n)\right)\\ &= \sum \int f d\left(c_n I(x-s_n)\right)\\ &= \sum c_n \int f d I(x-s_n)\\ &= \sum c_n f(s_n).\end{aligned}$$

$\square$

Theorem 6.2.6. *Assuming that all the conditions in Lem.6.2.1 holds. Suppose the infinite series* $\sum c_n$ *converges and let*

$$\alpha(x) = \sum_{n=1}^{\infty} c_n I\left(x - s_n\right),$$

then

$$\int_a^b f d\alpha = \sum_{n=1}^{\infty} c_n f\left(s_n\right).$$

Proof. We want to show that

$$\left|\int_a^b f d\alpha - \sum_{i=1}^{N} c_n f\left(s_n\right)\right| < \varepsilon$$

holds for all $\varepsilon > 0$. Let $\varepsilon > 0$ be given. Since $\sum c_n$ is convergent, then there exists an sufficiently large integer such that

$$\sum_{n=N+1}^{\infty} c_n < \frac{\varepsilon}{M}$$

where $M = \sup|f(x)|$. Then let

$$\alpha_1(x) = \sum_{n=1}^{N} c_n I\left(x - s_n\right), \quad \alpha_2(x) = \sum_{N+1}^{\infty} c_n I\left(x - s_n\right).$$

Note that since $I(x - s_n) \leq 1$, we have

$$\alpha_2(x) \sum_{N+1}^{\infty} c_n I\left(x - s_n\right) \leq \sum_{n=N+1}^{\infty} c_n < \frac{\varepsilon}{M}.$$

Also, observe that $\alpha_2(a) = 0$. From Lem.6.2.1, we have

$$\int_a^b f d\alpha_1 = \sum_{i=1}^{N} c_n f\left(s_n\right).$$

Hence,

$$\begin{aligned}\left|\int_a^b f d\alpha - \sum_{i=1}^{N} c_n f\left(s_n\right)\right| &= \left|\int_a^b f d\alpha - \int_a^b f d\alpha_1\right| \\ &= \left|\int_a^b f d\alpha_2\right| \leq M[\alpha(b)-\alpha(a)] < \varepsilon.\end{aligned}$$

Hence we know the series on the left hand side converges to this integral for chosen N. This completes the proof. $\square$

Theorem 6.2.7. *Suppose α increases monotonically and $\alpha' \in \mathscr{R}$ on $[a,b]$. Let f be a bounded real function on $[a,b]$. Then $f \in \mathscr{R}(\alpha)$ if and only if $f\alpha' \in \mathscr{R}$. In that case,*

$$\int_a^b f d\alpha = \int_a^b f(x)\alpha'(x)dx.$$

Proof. Suppose $f \in \mathscr{R}(\alpha)$ and $\alpha' \in \mathscr{R}$. Since $f \in \mathscr{R}(\alpha)$, we know $f \in \mathscr{R}$. Hence by the previous theorem, we know $f\alpha' \in \mathscr{R}$.

Suppose the converse: that is, suppose $f\alpha' \in \mathscr{R}$. We want to show that the upper and lower integrals are of the expressions are equal. Then since $\int f\alpha' dx$ exists, $\int f d\alpha$ must also exist and the theorem follows.

Since $\alpha' \in \mathscr{R}$, we know there exists a partition P such that

$$U\left(P,\alpha'\right) - L\left(P,\alpha'\right) < \varepsilon.$$

The mean value theorem tells us that there exists $t_i \in [x_{i-1}, x_i]$ such that

$$\Delta\alpha_i = \alpha'\left(t_i\right)\Delta x_i.$$

From the theorem in last section, we know that since $\alpha' \in \mathscr{R}$, if $s_i \in [x_{i-1}, x_i]$, then

$$\sum_{i=1}^{n}\left|\alpha'\left(s_i\right) - \alpha'\left(t_i\right)\right|\Delta x_i < \varepsilon.$$

Then we know

$$\begin{aligned}\left|\sum f(s_i)\Delta\alpha_i - \sum f(s_i)\alpha'(s_i)\Delta x_i\right| &= \left|\sum f(s_i)\alpha'(t_i)\Delta x_i - \sum f(s_i)\alpha'(s_i)\Delta x_i\right| \\ &\leq \sum\left|f(s_i)[\alpha'(t_i) - \alpha'(s_i)]\right|\Delta x_i \\ &\leq M\varepsilon\end{aligned}$$

where $M = \sup|f(x)|$ since f is bounded. Hence,

$$\sum f(s_i)\Delta\alpha_i \leq \sum f(s_i)\alpha'(s_i)\Delta x_i + M\varepsilon.$$

Taking the suprimum on both side of the inequality we get

$$U(P, f, \alpha) \leq U'(P, f\alpha') + M\varepsilon.$$

And since

$$\inf U(P, f, \alpha) - \inf U(P, f\alpha') \leq \inf(U(P, f, \alpha) - U(P, f\alpha')) \leq M\varepsilon,$$

we get

$$\left| \overline{\int} f d\alpha - \overline{\int} f\alpha' dx \right| \leq M\varepsilon.$$

Since this is true for any $\varepsilon > 0$, we conclude that

$$\overline{\int} f d\alpha = \overline{\int} f\alpha' dx.$$

Also, since

$$\sum f(s_i)\Delta\alpha_i \leq \sum f(s_i)\alpha'(s_i)\Delta x_i + \varepsilon,$$

we can of course take the infimum on the left hand side:

$$\inf \sum f(s_i)\Delta\alpha_i \leq \sum f(s_i)\alpha'(s_i)\Delta x_i + \varepsilon.$$

Hence we get

$$L(P, f, \alpha) - M\varepsilon \leq \sum f\alpha'(s_i)\Delta x_i.$$

And by the property of the greatest lower bound, we get

$$L(P, f, \alpha) - L(P, f\alpha') \leq M\varepsilon.$$

Hence,

$$L(P, f, \alpha) \leq \sup L(P, f\alpha') + M\varepsilon$$

and we take the suprimum again and get

$$\sup L(P, f, \alpha) \leq \sup L(P, f\alpha') + M\varepsilon.$$

Hence it is also true that

$$\overline{\int} f d\alpha = \overline{\int} f\alpha' dx.$$

And the theorem follows. □

Theorem 6.2.8. ***(Change of Variables)*** *Suppose φ is a strictly increasing continuous function and maps $[A, B]$ onto $[a, b]$. Suppose α is a monotonically increasing function on $[a, b]$ and $f \in \mathscr{R}(\alpha)$ on $[a, b]$. Define β and g on $[A, B]$ by*

$$\beta(y) = \alpha(\varphi(y)), \quad g(y) = f(\varphi(y)).$$

Then $g \in \mathscr{R}(\beta)$ and

$$\int_A^B g d\beta = \int_a^b f d\alpha.$$

$$\begin{tikzcd}
& {[A,B]} \\
{[A,B]} \arrow[ur, "g"] \arrow[r, "\varphi"] \arrow[dr, "\beta"'] & {[a,b]} \arrow[u, "f"'] \arrow[d, "\alpha"'] \\
& {[A,B]}
\end{tikzcd}$$

Proof. Since φ is strictly increasing, given a partition Q on $[A, B]$, its image $P = \varphi(Q) = \{\varphi(y_0), ..., \varphi(y_n)\}$ preserves the order of the partition. We can do some simple calculation as the following:

$$\begin{aligned}
U(Q, g, \beta) &= \sum \sup_{y_i} g(y_i)[\beta(y_i) - \beta(y_{i-1})] \\
&= \sum \sup_{y_i} f(\varphi(y_i))[\alpha(\varphi(y_i)) - \alpha(\varphi(y_{i-1}))] \\
&= \sum \sup_{\varphi_i} f(\varphi_i)[\alpha(\varphi_i) - \alpha(\varphi_{i-1})] \\
&= U(P, f, \alpha).
\end{aligned}$$

Similarly, it is not too hard to get

$$L(Q, g, \beta) = L(P, f, \alpha).$$

Since $f \in \mathscr{R}(\alpha)$, there exists partition P such that

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

Hence we can also get

$$U(Q, g, \beta) - L(Q, g, \beta) < \varepsilon.$$

This implies that $g \in \mathscr{R}(\beta)$ and we get

$$\int_A^B g d\beta = \int_a^b f d\alpha.$$

This concludes the proof. □

Remark 6.2.1. Take $\alpha(x) = x$, then $\beta = \varphi$. Assume that $\varphi' \in \mathscr{R}$ on $[a, b]$. Apply Thm.6.2.7 and we get

$$\int_a^b f(x)dx = \int_A^B f(\varphi(y))\varphi'(y)dy,$$

which we have seen many times when doing an integration that requires change of variables.

6.3 Integration and Differentiation

This is considerably the most important section in the integration. Here we will build the connection between differentiation and integration.

Theorem 6.3.1. *Let $f \in \mathscr{R}$ on $[a,b]$. For $a \leq x \leq b$, put*

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

Then F is continous on $[a,b]$; furthermore, if f is continous at a point x_0 of $[a,b]$, then F is differentiable at x_0, and

$$F'(x_0) = f(x_0).$$

Proof. We first prove that F is continous on $[a,b]$. Since $f \in \mathscr{R}$ and is bounded, for $a \leq t \leq b$, put $M = \sup |f(t)|$. If $a \leq x \leq y \leq b$, then

$$|F(y) - F(x)| = \left| \int_x^y f(t)dt \right| \leq M(y-x).$$

Hence given $\varepsilon > 0$, we see that

$$|y - x| < \frac{\varepsilon}{M} \Rightarrow |F(y) - F(x)| < \varepsilon.$$

This proves the continuity.

Now suppose f is continuous at x_0. Given $\varepsilon > 0$, choose $\delta > 0$ such that

$$|t - x_0| < \delta \Rightarrow |f(t) - f(x_0)| < \varepsilon.$$

Hence, if $x - \delta < s \leq x_0 \leq t < x_0 + \delta$, and $a \leq s \leq t \leq b$, we have

$$\left| \frac{F(t) - F(s)}{t - s} - f\ (x_0) \right| = \left| \frac{1}{t-s} \int_s^t [f(u) - f\ (x_0)]\, du \right| < \varepsilon.$$

It follows that $F'(x_0) = f(x_0)$. □

Theorem 6.3.2. ***(The Fundamental Theorem of Calculus)*** *If $f \in \mathscr{R}$ and if there is a differentiable function F on $[a,b]$ such that $F' = f$, then*

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

Proof. Let $\varepsilon > 0$ be given. Choose a partition P such that

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

The mean value theorem says that there exists $t_i \in [x_{i-1}, x_i]$ such that

$$F\ (x_i) - F\ (x_{i-1}) = f\ (t_i)\, \Delta x_i$$

for $i = 1, 2, ...n$. Thus

$$\sum_{i=1}^{n} f(t_i)\,\Delta x_i = F(b) - F(a).$$

Since $f \in \mathscr{R}$, it follows that

$$\left| F(b) - F(a) - \int_a^b f(x)dx \right| = \left| \sum_{i=1}^{n} f(t_i)\Delta x_i - \int_a^b f(x)dx \right| < \varepsilon.$$

Since ε is arbitrary, this completes the proof. □

6.4 Integration of Vector-Valued Functions

Definition 6.4.1. Let $\mathbf{f} : [a, b] \to \mathbb{R}^k$ and $\mathbf{f} = (f_1, f_2, ..., f_k)$, which are all functions in $\mathscr{R}(\alpha)$. If $f \in \mathscr{R}(\alpha)$, then define

$$\int_a^b \mathbf{f} d\alpha = \left(\int_a^b f_1 d\alpha, \ldots, \int_a^b f_k d\alpha \right).$$

We have some properties of the integration still hold, like the linearity of the integral and the superposition of α_1, α_2. There is also an analogous result from the fundamental theorem of calculus:

Theorem 6.4.1. *If* $\mathbf{f}$ *and* $\mathbf{F}$ *map* $[a, b]$ *into* $\mathbb{R}^k$, *if* $\mathbf{f} \in \mathscr{R}$ *on* $[a, b]$ *and if* $\mathbf{F}' = \mathbf{f}$, *then*

$$\int_a^b \mathbf{f}(t)dt = F(b) - F(a).$$

Here is a rather important theorem that will help us when we are restricting the integrals of vector-valued functions:

Theorem 6.4.2. *If* $\mathbf{f}$ *maps* $[a, b]$ *into* $\mathbb{R}^k$ *and if* $\mathbf{f} \in \mathscr{R}(\alpha)$ *for some monotonically increasing function* α *on* $[a, b]$, *then* $|\mathbf{f}| \in \mathscr{R}(\alpha)$ *and*

$$\left| \int_a^b \mathbf{f} d\alpha \right| \leq \int_a^b |\mathbf{f}| d\alpha.$$

Proof. Since $f_i^2 \in \mathscr{R}(\alpha)$, we know

$$|\mathbf{f}| = \left(f_1^2 + \cdots + f_k^2\right)^{1/2}$$

belongs to $\mathscr{R}(\alpha)$ by the previous theorems. To prove the inequality, put

$$\mathbf{y} = (y_1, ..., y_k) \qquad \text{where } y_i = \int f_i d\alpha.$$

Then we have $\mathbf{y} = \int \mathbf{f} d\alpha$. Notice that

$$|\mathbf{y}|^2 = \sum y_i^2 = \sum y_j \int f_j d\alpha = \int \left(\sum y_j f_j\right) d\alpha.$$

By the Schwarz inequality, we have

$$\sum y_j f_j(t) \leq |\mathbf{y}||\mathbf{f}(t)| \quad (a \leq t \leq b).$$

Hence we have

$$|\mathbf{y}|^2 \leq |\mathbf{y}| \int |\mathbf{f}| d\alpha.$$

If $\mathbf{y} \neq \mathbf{0}$, the theorem follows. Otherwise the theorem is automatically true. □

6.5 Rectifiable Curves

Definition 6.5.1. A continuous mapping $\gamma : [a, b] \to \mathbb{R}^k$ is called a curve in $\mathbb{R}^k$. We associate each partition $P = \{x_0, x_1, ..., x_n\}$ and to each curve γ on $[a, b]$ the number

$$\Lambda(P, \gamma) = \sum_{i=1}^{n} |\gamma(x_i) - \gamma(x_{i-1})|.$$

Define the arclength of γ to be

$$\Lambda(\gamma) = \sup \Lambda(P, \gamma),$$

where the suprimum is taken over all partitions on $[a, b]$. If $\Lambda(\gamma) < \infty$, say the curve is *rectifiable*.

Theorem 6.5.1. *If γ' is continous on $[a, b]$, then γ is rectifiable, and*

$$\Lambda(\gamma) = \int_a^b \left|\gamma'(t)\right| dt.$$

Proof. If $a \leq x_{i-1} < x_i \leq b$, then

$$|\gamma(x_i) - \gamma(x_{i-1})| = \left| \int_{x_{i-1}}^{x_i} \gamma'(t) dt \right| \leq \int_{x_{i-1}}^{x_i} \left|\gamma'(t)\right| dt.$$

Hence,

$$\Lambda(P, \gamma) \leq \int_a^b \left|\gamma'(t)\right| dt$$

for every partition on $[a, b]$. Consequently,

$$\Lambda(\gamma) \leq \int_a^b \left|\gamma'(t)\right| dt.$$

To prove the opposite inequality, let $\varepsilon > 0$ be given. Since γ' is uniformly continous on $[a, b]$, there exists $\delta > 0$ such that

$$\left|\gamma'(s) - \gamma'(t)\right| < \varepsilon \quad \text{if } |s - t| < \delta.$$

Let P be a partition with $\Delta x_i < \delta$ for all i. If $x_{i-1} \le t \le x_i$, it follows from the triangle inequality that

$$\left|\gamma'(t)\right| \le \left|\gamma'(x_i)\right| + \varepsilon.$$

Hence,

$$\begin{aligned}
\int_{x_{i-1}}^{x_i} \left|\gamma'(t)\right| dt &\le \left|\gamma'(x_i)\right| \Delta x_i + \varepsilon \Delta x_i \\
&= \left|\int_{x_{i-1}}^{x_i} \left[\gamma'(t) + \gamma'(x_i) - \gamma'(t)\right] dt\right| + \varepsilon \Delta x_i \\
&\le \left|\int_{x_{i-1}}^{x_i} \gamma'(t) dt\right| + \left|\int_{x_{i-1}}^{x_i} \left[\gamma'(x_i) - \gamma'(t)\right] dt\right| + \varepsilon \Delta x_i \\
&\le |\gamma(x_i) - \gamma(x_{i-1})| + 2\varepsilon \Delta x_i .
\end{aligned}$$

If we add these inequalities, we obtain

$$\begin{aligned}
\int_a^b \left|\gamma'(t)\right| dt &\le \Lambda(P, \gamma) + 2\varepsilon(b - a) \\
&\le \Lambda(\gamma) + 2\varepsilon(b - a).
\end{aligned}$$

Since ε is arbitrary, the theorem follows. □

Chapter 7

Sequences and Series of Functions

"Obvious" is the most dangerous word in mathematics.

-E.T.Bell

The main discussion of this section is based on sequences and series of functions, which is defined as the following:

Definition 7.0.1. Suppose $\{f_n\}, n = 1, 2, 3...,$ is a sequence of functions defined on a set E, and suppose the sequence of numbers $\{f_n(x)\}$ converges for every $x \in E$. We can define a function f by

$$f(x) = \lim_{n\to\infty} f_n(x) \quad (x \in E).$$

We say f *converges to* f *pointwise on* E. And if we define

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad (x \in E),$$

the function f is called the sum of the series $\sum f_n$.

There are several questions arises from this definition. For example, to ask whether the limit of a sequence of continous functions is continuous is the same as asking whether

$$\lim_{t\to x} \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \lim_{t\to x} f_n(t).$$

This is asking whether we can change the limits.

Sometimes we may also want to know whether we can change the order of the differentiation and the limit. That is, whether

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

In physics, if $f = \sum f_n$ and we want to integrate f, one usual trick is to integrate by terms under the summation sign. This is legitimate when

$$\int_a^b f d\alpha = \sum_{n=1}^{\infty} \int_a^b f_n d\alpha.$$

But is this always the case?

Examples in Rudin (pg.144-146) have demonstrated to us clearly that **the general answer to the questions above are no.** However, it will be nice if they were true. Hence, it is natural to ask whether these operations are legitimate under a class of sequences and series of functions. This gives rise to the definition of uniform convergence.

7.1 Uniform Convergence

Definition 7.1.1. We say a sequence of functions $\{f_n\}$ converges uniformly if for every $\varepsilon > 0$ there is an integer N such that

$$n \geq N \Rightarrow |f_n(x) - f(x)| \leq \varepsilon, \ \forall x \in E.$$

Remark 7.1.1. This definition is similar to that of uniform continuity. For a sequence to be convergent, the number N depends on ε and x. However, for the sequence to converge uniformly, we ask that the number N depends only on ε.

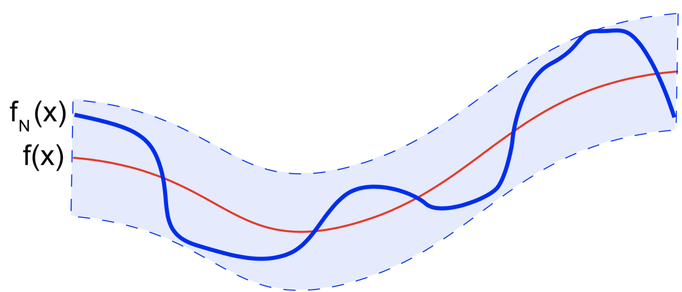


Figure 7.1: For large enough N, the functions $f_n(x)$ for all $n \geq N$ lives in an ε band within the limit $f(x)$.

Definition 7.1.2. We say the series $\sum f_n(x)$ converges uniformly on E is the sequence $\{s_n\}$ of partial sums defined by

$$\sum_{i=1}^{n} f_i(x) = s_n(x)$$

converges uniformly on E.

Theorem 7.1.1. ***(Cauchy criterion for uniform convergence)*** *The sequence of functions $\{f_n\}$, defined on $E \subset \mathbb{R}$, converges uniformly on E if and only if for every $\varepsilon > 0$ there exists an integer N such that*

$$m, n \geq N, x \in E \Rightarrow |f_n(x) - f_m(x)| \leq \varepsilon.$$

Proof. Suppose $\{f_n\}$ converges uniformly on E, then there is an integer N such that

$$n \geq N, x \in E \Rightarrow |f_n(x) - f(x)| \leq \frac{\varepsilon}{2}.$$

Then,

$$|f_n(x) - f_m(x)| \le |f_n(x) - f(x)| + |f(x) - f_m(x)| \le \varepsilon$$

if $m, n \ge N$ and $x \in E$. This completes the first half of the proof.

Conversely, suppose the Cauchy condition holds. By Thm.3.3.1 and the fact that $\mathbb{R}$ is complete, we know that $\{f_n\}$ converges at every x to a limit which we call $f(x)$. Thus $f_n \to f$ on E. Now we prove that the convergence is uniform.

Given $\varepsilon > 0$, there exists N such that

$$|f_{n+k}(x) - f_n(x)| \le \varepsilon, \ \forall n > N, \ k \in \mathbb{N}, \ x \in E.$$

Fix n and let $k \to \infty$ we get

$$|f(x) - f_n(x)| \le \varepsilon, \ \forall n > N, \ x \in E.$$

Hence the function converges uniformly. This completes the proof. □

Theorem 7.1.2. *Suppose $\{f_n\}$ is a sequence of bounded function on E that converges to a bounded function f and suppose*

$$\lim_{n\to\infty} f_n(x) = f(x) \quad (x \in E).$$

Put

$$M_n = \sup_{x\in E} |f_n(x) - f(x)|.$$

Then $f_n \to f$ uniformly on E if and only if $M_n \to 0$ as $n \to \infty$.

Proof. This theorem is an immediate consequence of Thm.7.1.1. We knwo that $f_n \to f$ uniformly if and only if there exist N such that for all $\varepsilon > 0$

$$m, n \ge N, x \in E \Rightarrow |f_n(x) - f_m(x)| \le \varepsilon.$$

Fix n and let $m \to \infty$, we have

$$n \ge N, x \in E \Rightarrow |f_n(x) - f(x)| \le \varepsilon.$$

Then by taking the suprimum we get

$$n \ge N \Rightarrow \sup_{x\in E} |f_n(x) - f(x)| \le \varepsilon.$$

This completes the proof. □

Theorem 7.1.3. *Suppose $\{f_n\}$ is a sequence of functions defined on E and suppose*

$$|f_n(x)| \le M_n \quad (x \in E, n = 1, 2, 3, \dots).$$

Then $\sum f_n$ converges uniformly on E if $\sum M_n$ converges.

Proof. If $\sum M_n$ converges, then for large enough m, n, given $\varepsilon > 0$,

$$\left|\sum_{i=n}^{m} f_i(x)\right| \le \sum_{i=n}^{m} M_i \le \varepsilon \quad (x \in E).$$

Uniform convergence now follows from Thm.7.1.1. □

7.2 Uniform Convergence and Continuity

Theorem 7.2.1. *Suppose $f_n \to f$ uniformly on a set $E \subset \mathbb{R}$ in a metric space. Let x be a limit point of E, and suppose that*

$$\lim_{t\to x} f_n(t) = A_n \quad (n = 1, 2, 3, \dots).$$

Then $\{A_n\}$ converges and

$$\lim_{t\to x} f(t) = \lim_{n\to\infty} A_n.$$

In other words,

$$\lim_{t\to x}\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty}\lim_{t\to x} f_n(t).$$

Proof. Let $\varepsilon > 0$ be given. By the uniform convergence of $\{f_n\}$, there exists an integer N such that

$$m, n \geq N, t \in E \Rightarrow |f_n(t) - f_m(t)| < \varepsilon.$$

Letting $t \to x$ we get

$$m, n \geq N \Rightarrow |A_n - A_m| < \varepsilon.$$

Hence $\{A_n\}$ is Cauchy and converges to a limit, say A.

Next,

$$|f(t) - A| \leq |f(t) - f_n(t)| + |f_n(t) - A_n| + |A_n - A|.$$

We first choose n such that

$$|f(t) - f_n(t)| \leq \frac{\varepsilon}{3} \quad \text{and} \quad |A_n - A| < \frac{\varepsilon}{3}.$$

This is possible for all $t \in E$ given that $f_n \to f$ uniformly. Then, for this n, choose a neighborhood $V_\delta(x)$ such that

$$|f_n(t) - A_n| \leq \frac{\varepsilon}{3}$$

if $t \in V_\delta(x)$. Hence we find that

$$|f(t) - A| \leq \varepsilon \quad \text{for} \quad t \in V_\delta(x).$$

This concludes the proof. □

Corollary 7.2.1. *If $\{f_n\}$ is a sequence of continuous function on E, and if $f_n \to f$ uniformly on E, then f is continuous on E.*

Proof. This is an immediate result from Thm.7.2.1. Since

$$\lim_{t\to x} f_n(t) = f_n(x) = A_n,$$

and since f_n converges uniformly to f, we have

$$\lim_{t\to x} f(t) = \lim_{n\to\infty} A_n = \lim_{n\to\infty} f_n(x) = f(x).$$

Hence f is continuous. □

Theorem 7.2.2. *Suppose K is compact, and*
(a) $\{f_n\}$ is a sequence of continuous function on K,
(b) $\{f_n\}$ converges pointwise to a continuous function f on K,
(c) $f_n(x) \geq f_{n+1}(x)$ for all $x \in K$ for $n = 1, 2, 3, ...$
Then $f_n \to f$ uniformly on K.

Proof. **(Prof.Bai's Proof)** Consider the function $g_n = f_n - f$. We make the following observations:
(i) g_n is continuous,
(ii) $g_n \to 0$ pointwise,
(iii) $g_n \geq g_{n+1}$.
From (iii) we can immediately see that $g_n(x) \geq g_{n+k}(x)$ for all $n, k > 0$ and for all x. Fix n and let $k \to \infty$ we see that $g_n(x) \geq 0$ for all n. Then it suffices to show that $g_n \to 0$ uniformly. Hence, we want to show that given $\varepsilon > 0$, there exists an integer N such that $g_n(x) < \varepsilon$ for all $n \geq N$ and for all $x \in E$. However, from (iii), since $\{g_n\}$ is a decreasing sequence, the goal reduces to finding an integer N such that $g_N(x) < \varepsilon$ for all $x \in E$.

From (ii), we know that for each $x \in E$ there exists an integer N_x such that $g_{N_x}(x) \leq \varepsilon/2$. From (i), we know that there exists a neighborhood $V(x)$ such that

$$t \in V(x) \Rightarrow |g_{N_x}(t) - g_{N_x}(x)| < \frac{\varepsilon}{2}.$$

Using the triangle inequality, we know that

$$\begin{aligned} g_{N_x}(t) &\leq |g_{N_x}(t) - g_{N_x}(x)| + g_{N_x}(x) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

We know that $\{V(x)\}_{x \in K}$ is an open cover of K. Since K is compact, there exists a finite subcover $\{V(x_i)\}_{i=1}^n$ such that

$$K \subset \bigcup_{i=1}^{n} V(x_i).$$

Hence put

$$N = \max_{1 \leq i \leq n} \{N_{x_i}\}.$$

Then, for all $n \geq N$ and for all $t \in K$, there exists i such that $t \in V(x_i)$, hence

$$g_N(t) \leq g_{N_{x_i}}(t) < \varepsilon,$$

as desired. $\square$

Remark 7.2.1. The most important use of compactness is that it turns the local property of a function defined on a compact set into the global property of this function. This proof is a very good example. And here we turn pointwise (local) into uniform convergence (global). In the previous examples we also turn local continuity into uniform continuity.

Proof. **(Rudin's Proof)** Let $g_n(x)$ have the same meaning as in the previous proof. Given $\varepsilon > 0$, define

$$K_n = \{x \in K : g_n(x) \geq \varepsilon\}.$$

Since g_n is continuous and $g^{-1}([\varepsilon, \infty))$ is the inverse image of a closed set (since its complement is open), which is also closed. Hence K_n is closed. And since K_n is a closed subset of a compact set K, it is also compact.

Since $g_n \geq g_{n+1}$, if $x \in K_{n+1}$, then $g_{n+1} \geq \varepsilon$. But since $g_n \geq g_{n+1} \geq \varepsilon$, we know $x \in K_n$. Hence we have

$$K_n \supset K_{n+1} \supset K_{n+2} \ldots$$

From Thm.2.7.7, if $K_n \neq \emptyset$ for all n and $K_n \supset K_{n+1}$ for all n, then we know $\bigcap_n K_n \neq \emptyset$. We claim that $\bigcap_n K_n = \emptyset$ and thus $K_n = \emptyset$ for some n. Hence, for all $x \in K$, there exists $N \in \mathbb{N}$ such that $g_N(x) < \varepsilon$. This will complete the proof, just like what we did in the previous proof.

Since $g_n(x) \to 0$ we see that $x \notin K_n$ if n is sufficiently large. Thus $x \notin \bigcap_n K_n$. Hence $\bigcap_n K_n = \emptyset$, and we are done. □

Definition 7.2.1. If X is a metric space, then $\mathscr{C}(X)$ will denote the set of complex (or real) valued, continuous, bounded functions with domain X. We associate with each $f \in \mathscr{C}(X)$ its supremum norm

$$\|f\| = \sup_{x \in x} |f(x)|.$$

Since f is assumed to be bounded, $||f|| < \infty$. It is not hard to prove the following:
(i) $||f|| = 0$ *only if* $f(x) = 0$ *for every* $x \in X$,
(ii) $||f - g|| = ||g - f||$,
(iii) $\|f + g\| \leq \|f\| + \|g\|$. Note that (iii) is true since

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\| + \|g\|.$$

Hence it is clear that for $f, g \in \mathscr{C}(X)$, we can use supremum norm to be a metric. Thus $\mathscr{C}(X)$ has become a metric space.

We can rephrase Thm.7.1.2 as the following:

Theorem 7.2.3. *A sequence $\{f_n\}$ converges to f with respect to the metric space $\mathscr{C}(X)$ if and only if $f_n \to f$ uniformly on X.*

We can also check that $\mathscr{C}(X)$ is complete.

Theorem 7.2.4. *The above metric makes $\mathscr{C}(X)$ into a complete metric space.*

Proof. Let $\{f_n\}$ be a Cauchy sequence in $\mathscr{C}(X)$. This means that to each $\varepsilon > 0$ corresponds an N such that

$$m, n \geq N \Rightarrow \|f_n - f_m\| < \varepsilon.$$

It follows from Thm.7.1.1 (which can be generalized into any metric space) that there is a function f with domain X to which $\{f_n\}$ converges uniformly. By Thm.7.2.1 we know f is continuous. Moreover, f is bounded since there exists an n such that

$$|f(x) - f_n(x)| < 1, \ \forall x \in X$$

and f_n is bounded.

Thus $f \in \mathscr{C}(X)$ and since $f_n \to f$ uniformly on X, we have $\|f - f_n\| \to 0$ as $n \to 0$. □

7.3 Uniform Convergence and Integration

Theorem 7.3.1. *Let α be monotonically increasing on $[a,b]$. Suppose $f_n \in \mathscr{R}(\alpha)$ on $[a,b]$ for $n = 1,2,3,...$, and suppose $f_n \to f$ uniformly on $[a,b]$. Then $f \in \mathscr{R}(\alpha)$ on $[a,b]$ and*

$$\int_a^b f d\alpha = \lim_{n\to\infty} \int_a^b f_n d\alpha.$$

Proof. Since $f_n \to f$ uniformly, given $\varepsilon > 0$ there exists an N such that

$$n \geq N \Rightarrow |f - f_n| \leq \varepsilon.$$

For this n, we know that

$$f_n - \varepsilon \leq f \leq f_n + \varepsilon.$$

Since $f_n \in \mathscr{R}(\alpha)$, we know the lower integral of $f_n - \varepsilon$ is just its integral. Same for the upper integral of $f_n + \varepsilon$. Hence we have

$$\int_a^b (f_n - \varepsilon)\, d\alpha \leq \underline{\int} f d\alpha \leq \overline{\int} f d\alpha \leq \int_a^b (f_n + \varepsilon)\, d\alpha.$$

Therefore, it is obvious that

$$0 \leq \overline{\int} f d\alpha - \underline{\int} f d\alpha \leq 2\varepsilon[\alpha(b) - \alpha(a)]. \tag{7.1}$$

Since ε is arbitrarily, it is clear that $f \in \mathscr{R}(\alpha)$. Another consequence of Eqn.7.1 is that

$$\left| \int_a^b f d\alpha - \int_a^b f_n d\alpha \right| \leq \varepsilon[\alpha(b) - \alpha(a)].$$

This completes the proof. □

Corollary 7.3.1. *If $f_n \in \mathscr{R}(\alpha)$ on $[a,b]$ and if the series*

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad (a \leq x \leq b),$$

converging uniformly on $[a,b]$, then

$$\int_a^b f d\alpha = \sum_{n=1}^{\infty} \int_a^b f_n d\alpha.$$

Remark 7.3.1. This is the condition for expanding the integrand as a power series and integrating by terms are legal. Very important when doing integrals in physics.

7.4 Uniform Convergence and Differentiation

Theorem 7.4.1. *Suppose $\{f_n\}$ is a sequence of functions differentiable on $[a, b]$ and such that $\{f_n(x_0)\}$ converges for some point $x_0 \in [a, b]$. If $\{f_n'\}$ converges uniformly on $[a, b]$, then*
(i) $\{f_n\}$ converges uniformly to a function f on $[a, b]$,
(ii) $f'(x) = \lim_{n\to\infty} f_n'(x)$ for $x \in [a, b]$.

Proof. (i) Given $\varepsilon > 0$, since f_n is convergent at some point $x_0 \in [a, b]$, then there exists an N_1 such that
$$m, n \geq N_1 \Rightarrow |f_n(x_0) - f_m(x_0)| < \frac{\varepsilon}{2}.$$
Also, since $\{f_n'\}$ is uniformly convergent, there exists an N_2 such that
$$m, n \geq N_{\Rightarrow} |f_n'(t) - f_m'(t)| < \frac{\varepsilon}{2(b-a)}$$
for some $t \in [a, b]$. Take
$$N = \max\{N_1, N_2\}.$$
Then, for $m, n \geq N$, using the mean value theorem, we get
$$\begin{aligned}
|f_n(x) - f_m(x)| &= |f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0) + f_n(x_0) - f_m(x_0)| \\
&\leq |f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0)| + |f_n(x_0) - f_m(x_0)| \\
&= |(f_n - f_m)(x) - (f_n - f_m)(x_0)| + |f_n(x_0) - f_m(x_0)| \\
&= |(f_n - f_m)'(t)(x - x_0)| + |f_n(x_0) - f_m(x_0)| \\
&\leq \frac{\varepsilon|x - x_0|}{2(b-a)} + \frac{\varepsilon}{2} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$
This proves (i). We denote
$$f(x) = \lim_{n\to\infty} f_n(x) \quad (a \leq x \leq b).$$

(ii) Fix a point x on $[a, b]$ and define
$$\phi_n(t) = \frac{f_n(t) - f_n(x)}{t - x}, \quad \phi(t) = \frac{f(t) - f(x)}{t - x}.$$
For $a \leq t \leq b$ and $t \neq x$. Then $\lim_{t\to x} \phi_n(t) = f_n'(t)$. Since we want to show that $f'(x) = \lim_{n\to\infty} f_n(x)$, it suffices to show that $\{\phi_n(t)\}$ converges uniformly.

Since we know that
$$|f_n(x) - f_m(x) - f_n(t) + f_m(t)| \leq \frac{|x - t|\varepsilon}{2(b-a)},$$
It is clear that
$$|\phi_n(t) - \phi_m(t)| \leq \frac{\varepsilon}{2(b-a)} \quad (n \geq N, m \geq N).$$
Hence $\{\phi_n\}$ converges uniformly. And applying Thm.7.2.1, the theorem follows. □

Remark 7.4.1. If the continuity of f_n' is assumed in addition to above hypothesis, then a much shorter proof can be based on the fundamental theorem of calculus. Define

$$g_n(x) = \int_{x_0}^{x} f_n'(t)dt = f_n(x) - f_n(x_0).$$

Since $f_n'(t)$ converges uniformly on $[a, b]$, there exists an N such that

$$m, n \geq N \Rightarrow |f_n'(t) - f_m'(t)| < \frac{\varepsilon}{b-a}.$$

Then, we know that

$$|g_n(x) - g_m(x)| \leq \left| \int_{x_0}^{x} (f_n'(t) - f_m'(t))dt \right| \leq \frac{\varepsilon}{b-a}(x - x_0) \leq \varepsilon.$$

This shows that $\{g_n(x)\}$ is uniformly convergent on $[a, b]$. Suppose that $f_n(x_0) \to M$ is a sequence of number, and since

$$g_n(x) + f_n(x_0) = f_n(x),$$

by taking the limit we have

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} g_n(x) + f_n(x_0) = g(x) + M = f(x).$$

Hence $\{f_n(x)\}$ is uniformly continuous on $[a, b]$. Hence the proof will be shorter if we make the condition of the theorem above stronger. ■

Theorem 7.4.2. *There exists a real continuous function on the real line which is nowhere differentiable.*

Proof. Define

$$\varphi(x) = |x| \quad (-1 \leq x \leq 1)$$

and extend the definition of $\varphi(x)$ to all real x by requiring that

$$\varphi(x + 2) = \varphi(x).$$

Thus we have made $\varphi(x)$ into a periodic function with period 2. Then, with a little bit of geoemtry, we see that

$$|\varphi(s) - \varphi(t)| \leq |s - t| \tag{7.2}$$

for all s and t. Thus $\varphi(x)$ is continuous on $\mathbb{R}^1$. Define

$$f(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n \varphi\left(4^n x\right).$$

Since $0 \leq \varphi \leq 1$, Thm.7.1.3 shows that this series converges uniformly on $\mathbb{R}^1$. Also, Thm.7.2.1 shows that f is continuous on $\mathbb{R}^1$.

Now fix a real number x and a positive integer m. Put $\delta_m = \pm\frac{1}{2}\cdot 4^{-m}$, where the sign is chosen so that no integers lies between $4^m x$ and $4^m(x+\delta_m)$. This can be done since $4^m|\delta_m| = 1/2$. Define

$$\gamma_n = \frac{\varphi\left(4^n\left(x+\delta_m\right)\right) - \varphi\left(4^n x\right)}{\delta_m}.$$

When $n > m$, $4^n\delta_m$ is an even integer, so that $\gamma_n = 0$. When $0 \le n \le m$, (7.2) implies that $|\gamma_n| \le 4^n$. Hence $\gamma_n > -4^n$.

Again, with some geometry, it is not hard to show that $|\gamma_m| = 4^m$. Hence,

$$\begin{aligned}
\left|\frac{f(x+\delta_m) - f(x)}{\delta_m}\right| &= \left|\sum_{n=0}^{m}\left(\frac{3}{4}\right)^n \gamma_n\right| \\
&= \left|\left(\frac{3}{4}\right)^m \gamma_m + \sum_{n=0}^{m-1}\left(\frac{3}{4}\right)^n \gamma_n\right| \\
&\ge \left|\left(\frac{3}{4}\right)^m \gamma_m - \sum_{n=0}^{m-1} 3^n\right| \\
&\ge \left|\left(\frac{3}{4}\right)^m \gamma_m\right| - \left|\sum_{n=0}^{m-1} 3^n\right| \\
&\ge 3^m - \sum_{n=0}^{m-1} 3^n = \frac{1}{2}(3^m + 1).
\end{aligned}$$

As $m \to \infty$, $\delta_m \to 0$. It follows that f is not differentiable at x. □

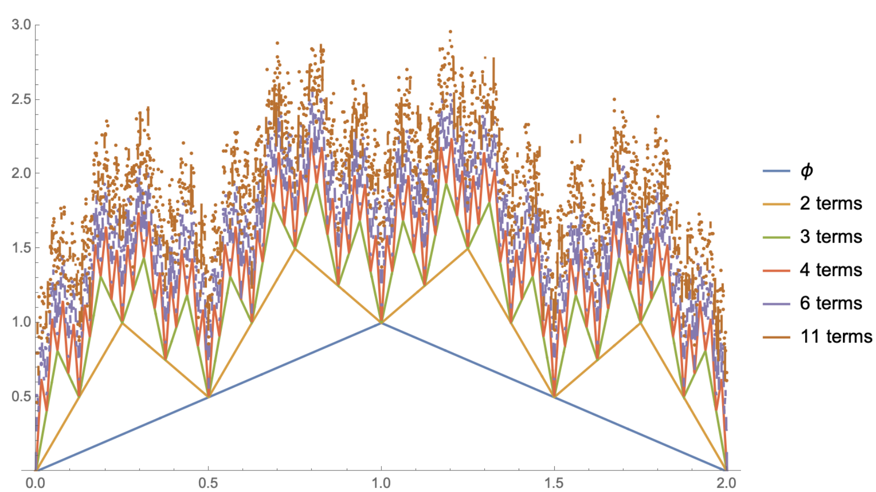


Figure 7.2: A plot of partial sums s_n of f using MATHEMATICA when $n = 1, 2, 3, 5, 10$ respectively, along with the original function $\varphi(x)$. As we add more and more terms in the series, f has more and more not-differentiable points. If we add infinitely many terms, f will have the fractal structure similar to what is shown in the picture.

7.5 Equicontinuous Families of Functions

Recall Thm.3.2.2, which says some subsequence of a sequence in a compact metric space X converges to a point in X. Then a easy corollary of this theorem is the following.

Corollary 7.5.1. *Every bounded sequence in $\mathbb{R}^k$ contains a convergent subsequence.*

Proof. This follows directly from Thm.3.2.2 and the fact that every bounded subset of $\mathbb{R}^k$ lies in a compact subset of $\mathbb{R}^k$ (a k-cell). □

Hence now we are interested in whether something similar is true for sequences of functions in either $\mathbb{R}^1$ or $\mathbb{R}^2$ (real or complex). We first specify what does boundedness mean for a sequence of functions.

Definition 7.5.1. Let $\{f_n\}$ be a sequence of functions defined on a set E. We say $\{f_n\}$ is *piece-wise bounded* on E if the sequence $\{f_n\}$ is bounded for every $x \in E$. That is, there exists a finite-valued funtion $\phi(x)$ defined on E such that

$$|f_n(x)| < \phi(x) \quad (x \in E, n = 1, 2, 3, \dots).$$

We say $\{f_n\}$ is *uniformly bounded* on E if there exists a number M such that

$$|f_n(x)| < M \qquad \forall n \text{ and } \forall x \in E.$$

If $\{f_n\}$ is pointwise bounded on E and E_1 is a countable subset of E, it is always possible to find a subsequence of $\{f_n\}$ such that it converges for every $x \in E_1$ (you'll see). However, even if $\{f_n\}$ is uniformly bounded continuous functions on a compact set E, **there need not exist a subsequence which converges pointwise on** E.

Example 7.5.1. Let

$$f_n(x) = \sin nx \quad (0 \leq x \leq 2\pi, n - 1, 2, 3, \dots).$$

Suppose there exists a sequence $\{n_k\}$ such that $\{\sin n_k x\}$ converges, for every $x \in [0, 2\pi]$. In that case we must have

$$\lim_{k\to\infty} (\sin n_k x - \sin n_{k+1} x) = 0 \quad (0 \leq x \leq 2\pi).$$

Hence,

$$\lim_{k\to\infty} (\sin n_k x - \sin n_{k+1} x)^2 = 0 \quad (0 \leq x \leq 2\pi).$$

We appeal to Lebesgue's theorem (Rudin, Theorem 11.32), then

$$\lim_{k\to\infty} \int_0^{2\pi} (\sin n_k x - \sin n_{k+1} x)^2 \, dx = 0.$$

But calculating this integral by ourself and we can see that

$$\int_0^{2\pi} (\sin n_k x - \sin n_{k+1} x)^2 \, dx = 2\pi.$$

And we reached a contradiction. Hence our assumption is wrong. Even though $|f_n(x)| \leq 1$ for all n and for all $x \in [0, 2\pi]$, there does not exists a convergent subsequence.

Another question is whether every convergent sequence contains a uniformly convergent subsequence. **This need not be true either**, as domonstrated by our next example.

Example 7.5.2. Let

$$f_n(x) = \frac{x^2}{x^2 + (1 - nx)^2} \quad (0 \le x \le 1, n = 1, 2, 3, \ldots).$$

Then $|f_n(x)| \le 1$ for all n and for all x in the domain. Hence this sequence is uniformly bounded on [0, 1]. Also, since

$$\lim_{n\to\infty} f_n(x) = 0 \quad (0 \le x \le 1),$$

$\{f_n(x)\}$ is pointwise convergent. However, notice that

$$f_n\left(\frac{1}{n}\right) = 1 \qquad (n = 1, 2, 3...).$$

Therefore, this sequence is by no way uniform convergent. And no subsequence of it can be uniform convergent.

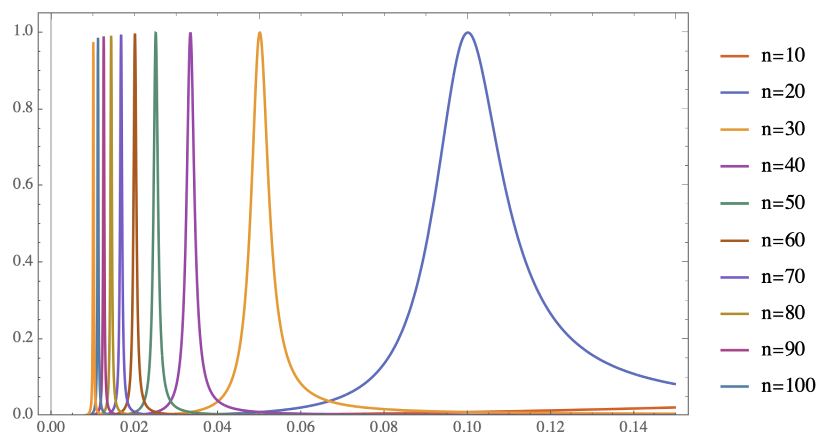


Figure 7.3: The plot of $\{f_n\}$ on $x \in [0, 0.15]$ when $n = 10, 20, 30..., 100$ using MATHEMATICA. As $n \to \infty$, the maximum value 1 gets closer and closer to the origin.

The concept needed in this connection is equicontinuity. It is defined as the following.

Definition 7.5.2. A family $\mathscr{F}$ of functions f defined on a set E in a metric space X is said to be *equicontinuous* on E if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - f(y)| < \varepsilon \qquad \forall d(x, y) < \delta,\ \forall x, y \in E,\ \forall f \in \mathscr{F}.$$

Remark 7.5.1. This definition is even stronger than uniform continuity since the δ in the uniform continuity, though independent of points, are still dependent on the function itself. However, since there can be more than one function in the collection $\mathscr{F}$, the definition requires the chosen δ to work *for all functions in this collection.*

Example 7.5.3. The collection of functions $\{f_n\}$ defined by

$$f_n = (-1)^n$$

are equicontinuous. For all $x, y \in E$, for all $d(x, y) < \delta$, the differences $|f(x) - f(y)| = 0 < \varepsilon$ for all $\varepsilon > 0$.

Example 7.5.4. *Finite collections $\{f_1, f_2, ..., f_n\}$ of uniform continuous functions are equicontinuous.* Given $\varepsilon > 0$, for f_i there exists a δ_i such that

$$|f_i(x) - f_i(y)| < \varepsilon \qquad \forall x, y \in E,\ \forall d(x, y) < \delta_i.$$

Take

$$\delta = \min\{\delta_1, \delta_2, ..., \delta_n\}.$$

Then this δ works for all functions in this collection.

Example 7.5.5. The sequence defined in Example 7.5.2 is not equicontinuous. We want to prove that there exists an n and $\varepsilon > 0$ such that for all $\delta > 0$, $|x - y| < \delta$ but $|f_n(x) - f_n(y)| \geq \varepsilon$.

Given $\delta > 0$, take $\varepsilon = 1/2$. Take $y = 0$ and $x = 1/n$. Choose n such that $1/n < \delta$. Then,

$$|x - y| = \frac{1}{n} < \delta,$$

However,

$$|f_n(x) - f_n(y)| = 1 > \frac{1}{2} = \varepsilon.$$

Theorem 7.5.1. *If $\{f_n\}$ is a pointwise bounded sequence of functions (real or complex) on a countable set E, then $\{f_n\}$ has a subsequence $\{f_{n_k}\}$ such that $\{f_{n_k}(x)\}$ converges for every $x \in E$.*

Proof. Let $\{x_i\}, i = 1, 2, 3...$ be the points of E arranged in a sequence. Since $\{f_n(x_1)\}$ is bounded, there exists a subsequence, which we call $\{f_{1,k}\}$ such that $\{f_{1,k}\}$ converges as $k \to \infty$. It is clear that this sequence S_1 converges at x_1.

Let us now consider the sequence $\{f_{1,k}(x_2)\}$. This sequence is a subsequence of $\{f_n\}$. Since $\{f_n\}$ is pointwise bounded, this sequence is also bounded at x_2. Hence, from this subsequence $\{f_{1,k}(x_2)\}$, we can extract another subsequence $\{f_{2,k}\}$ that is convergent at x_2, which we call S_2. However, since S_2 is a subsequence of S_1, it is also convergent at x_1.

We can repeat this process and consider $\{f_{2,k}(x_3)\}$ and we can gain extract a subsequence from it called S_3, which is convergent at x_1, x_2 and x_3. Repeat again...

Now let us consider the array

$$\begin{array}{ll} S_1 : f_{1,1},\ f_{1,2},\ f_{1,3},\ f_{1,4}... & \text{convergent at } x_1 \\ S_2 : f_{2,1},\ f_{2,2},\ f_{2,3},\ f_{2,4}... & \text{convergent at } x_1, x_2 \\ S_3 : f_{3,1},\ f_{3,2},\ f_{3,3},\ f_{3,4}... & \text{convergent at } x_1, x_2, x_3 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \end{array}$$

We go down the diagonal of this array and consider the sequence

$$S : f_{1,1},\ f_{2,2},\ f_{3,3},\ f_{4,4}...$$

We notice that S is a subsequence of $\{f_n\}$, and based on our construction, it has no repetitive elements. Namely, $f_{1,1} \neq f_{2,2} \neq f_{3,3}$... And as $n \to \infty$, the sequence S, which is $\{f_{n,n}\}$ will converge at every $x_i \in E$. □

Theorem 7.5.2. *If K is a compact metric space, if $f_n \in \mathscr{C}(K)$ for $n = 1, 2, 3...$, and if $\{f_n\}$ converges uniformly on K, then $\{f_n\}$ is equicontinuous on K.*

Proof. Let $\varepsilon > 0$ be given. Since $\{f_n\}$ converges uniformly, it must be uniformly Cauchy by Thm.7.1.1. Hence there exists an integer N such that

$$\|f_n - f_N\| < \frac{\varepsilon}{4} \quad (n > N).$$

Since continous functions are uniformly continuous on compact sets, there is a $\delta > 0$ such that

$$d(x, y) < \delta \Rightarrow |f_i(x) - f_i(y)| < \frac{\varepsilon}{2} \qquad (1 \leq i \leq N).$$

Now we claim that this δ would work for all $n > N$ and $d(x, y) < \delta$. Since

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f_n(y)| < \varepsilon.$$

This concludes the proof. □

Now, it is natural to ask whether the converse of this theorem is true. That is, *will a equicontinuous sequence of functions $\{f_n\}$ on a compact set, where $f_n \in \mathscr{C}(K)$ imply that $\{f_n\}$ is uniformly convergent, or have a uniformly convergent subsequence*. **Not exactly**. However, if we make some slight modification, we will see that this implication is true.

Theorem 7.5.3. *If K is a compact metric space, if $f_n \in \mathscr{C}(K)$ for $n = 1, 2, 3...$, and if $\{f_n\}$ converges pointwise and is equicontinuous on K, then $\{f_n\}$ converges uniformly.*

Proof. At each point $x \in K$, since $\{f_n\}$ converges pointwise, there exists an N such that

$$m, n \geq N \Rightarrow |f_n(x) - f_m(x)| < \frac{\varepsilon}{3}.$$

Also, since $\{f_n\}$ is equicontinuous, we know

$$d(x, y) < \delta \Rightarrow |f_n(x) - f_n(y)| < \frac{\varepsilon}{3} \qquad \forall n.$$

Since K is compact, there exist finite collection $x_1, ..., x_n \in K$ such that

$$K \subset \bigcup_{i=1}^{n} V_\delta(x_i).$$

Thus given any $x \in K$, we know that $x \in V_\delta(x_i)$ for some i. Hence

$$|f_n(x) - f_n(x_i)| < \frac{\varepsilon}{3} \qquad \forall n.$$

Therefore, for all $x \in K$,

$$\begin{aligned}|f_n(x) - f_m(x)| &= |f_n(x) - f_n(x_i) + f_n(x_i) - f_m(x_i) + f_m(x_i) - f_m(x)| \\ &\leq |f_n(x) - f_n(x_i)| + |f_n(x_i) - f_m(x_i)| + |f_m(x_i) - f_m(x)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

□

Definition 7.5.3. E is dense in X if every point of X is a limit point of E, or a point of E (or both).

Lemma 7.5.1. *Let E be a dense subset of K. Then there are finitely many points $x1, ..., x_n \in E$ such that*

$$K \subset \bigcup_{i=1}^{n} V_\delta(x_i). \tag{7.3}$$

Proof. It is clear that for any positive number δ' we have $K \subset \bigcup_{y \in K} V_{\delta'}(y)$. Since K is compact, there exists a finite collection of points $y_1, ..., y_n \in K$ such that

$$K \subset \bigcup_{i=1}^{n} V_{\delta'}(y_i).$$

Since E is a dense subset of K, then the points $y_1, ..., y_n$ are limit points of E. For every $1 \leq i \leq n$ we have $x_i \in V_{\delta'}(y_i)$ for some $x_i \in E$. Let us denote, for each i, that

$$d(x_i, y_i) = \Delta_i.$$

Then consider, for each i, the neighborhood $V_{\delta'+\Delta_i}(x_i)$. It is clear that

$$V_{\delta'}(y_i) \subset V_{\delta'+\Delta_i}(x_i) \qquad (1 \leq i \leq n).$$

Take

$$\delta = \max\left\{\delta' + \Delta_1, ..., \delta' + \Delta_n\right\}.$$

And we found the desired δ. Then (7.3) follows immediately. □

And now we are ready to prove the following theorem.

Theorem 7.5.4. *If K is compact, if $f_n \in \mathscr{C}(K)$ for $n = 1, 2, 3...$, and if $\{f_n\}$ is pointwise bounded and equicontinuous on K, then*
(a) $\{f_n\}$ is uniformly bounded on K,
(b) $\{f_n\}$ contains a uniformly convergent subsequence.

Proof. (a) Let $\varepsilon > 0$ be given and choose $\delta > 0$ in accordance with the definition of equicontinuity such that

$$d(x, y) < \delta \Rightarrow |f_n(x) - f_n(y)| < \varepsilon, \quad \forall n.$$

Since K is compact, there are finitely many points $p_1, ..., p_r \in K$ such that every $x \in K$ corresponds to at least one p_i with $d(x, p_i) < \delta$. Since $\{f_n\}$ is pointwise bounded, there exist $M_i < \infty$ such that $|f_n(p_i)| < M_i$ for all n. If $M = \max\{M_1, ..., M_r\}$, then $|f_n(x)| < M + \varepsilon$ for every $x \in K$. This proves (a). ■

(b) Let E be a countable dense subset of K (Rudin Chapter 2 Exercise 25 shows why such a set E exists). Thm.7.5.1 shows that $\{f_n\}$ has a subsequence $\{f_{n_i}\}$ such that $\{f_{n_i}(x)\}$ converges for every $x \in E$.

Put $f_{n_i} = g_i$ for simplicity. We shall prove that $\{g_i\}$ converges uniformly on K.

Let $\varepsilon > 0$ and pick $\delta > 0$ as in the beginning of the proof. Since E is dense in K, and K is compact, Lemma 7.5.1 says that there are finitely many points $x_1, ..., x_m \in E$ such that

$$K \subset V(x_1, \delta) \cup \cdots \cup V(x_m, \delta).$$

Since $\{g_i(x)\}$ converges at every $x \in E$, there is an integer N such that

$$i, j \geq N \Rightarrow \left|g_i(x_s) - g_j(x_s)\right| < \varepsilon \qquad (1 \leq s \leq m).$$

If $x \in K$, then $x \in V(x_s, \delta)$ for some s so that

$$i, j \geq N \Rightarrow |g_i(x) - g_i(x_s)| < \varepsilon, \qquad \forall i.$$

It then follows that

$$\begin{aligned}\left|g_i(x) - g_j(x)\right| &\leq |g_i(x) - g_i(x_s)| + \left|g_i(x_s) - g_j(x_s)\right| + \left|g_j(x_s) - g_j(x)\right| \\ &< 3\varepsilon.\end{aligned}$$

This completes the proof. □

7.6 The Stone-Weierstrass Theorem

Theorem 7.6.1. *If f is a continuous function on $[a, b]$, there exists a sequence of polynomials P_n such that*

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

uniformly on $[a, b]$. If f is real, then P_n may be taken real.

Before we start the proof, we will have to reduce this problem into a special case, and then we can start.

Although f is defined on $[a, b]$, we can assume, without loss of generality, that $[a, b] = [0, 1]$. We may also assume that $f(0) = f(1) = 0$. The reasons are the following.

First, consider the transformation $y = a + (b - a)x$ and let

$$f(x) = g(y) = g(a + (b - a)x).$$

You can see that

$$f(0) = g(a) \qquad f(1) = g(b).$$

Hence even though $[a, b]$ need not equal to $[0, 1]$, but by assuming that f is defined on $[0, 1]$, we can find another function g defined on $[a, b]$ by considering the transformation above. And it is not hard to see that

$$g(y) = f\left(\frac{y-a}{b-a}\right) \Rightarrow g(x) = f\left(\frac{x-a}{b-a}\right).$$

Hence, if there is a polynomial $P_n(x) \to f(x)$, then we know that

$$P_n\left(\frac{x-a}{b-a}\right) = Q_n(x) \to g(x).$$

This means that if we can find a polynomial P_n approaching f in this special reduced case where $f : [0, 1]$, we can consider the transformation above and then there must also be a polynomial $Q_n(x)$ approaching $g : [a, b]$ for the general case.

Similarly, one can considering reducing the problem and let $f(0) = f(1) = 0$, even though this may not be the case. Consider the transformation

$$g(x) = f(x) - f(0) - x[f(1) - f(0)].$$

We notice that $g(0) = g(1) = 0$. If there is a polynomial $Q_n \to g$, then we can consider the polynomial

$$P_n(x) = Q(x) + f(0) + x[f(1) - f(0)],$$

which must also approach to f. Hence if there exists a polynomial Q_n approaching the reduced function g, there must also be a polynomial P_n approaching the general function f.

Now we are ready for the proof.

Proof. We can assume, without loss of generality, that $[a, b] = [0, 1]$. We may also assume that $f(0) = f(1) = 0$. We define f to be zero for x outside $[0, 1]$. Then f is uniformly continuous on the whole line.

Put

$$Q_n(x) = c_n\left(1 - x^2\right)^n \quad (n = 1, 2, 3, \ldots),$$

where c_n is chosen such that

$$\int_{-1}^{1} Q_n(x)dx = 1 \quad (n = 1, 2, 3, \ldots).$$

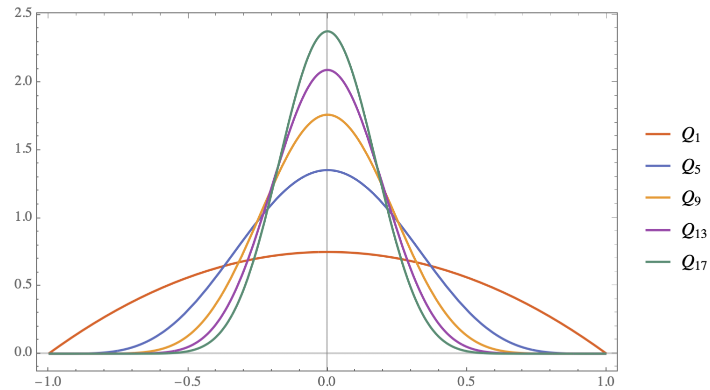


Figure 7.4: The plot of several such polynomials using MATHEMATICA.

We can see that as $n \to \infty$, $Q_n(x) \to 0$ where $x \neq 0$ and $Q_n(x)$ grows larger where $x = 0$. Think of Q_n as some sort of density function. As n grows, more weights concentrate around the origin, but the total mass is constant since the integral of such function is always 1.

We need some information on how fast c_n is growing. We can do a rough estimation using the inequality $(1-x^2)^n \geq 1 - nx^2$. We have

$$\begin{aligned}
\int_{-1}^{1} \left(1-x^2\right)^n dx = 2\int_0^1 \left(1-x^2\right)^n dx &\geq 2\int_0^{1/\sqrt{n}} \left(1-x^2\right)^n dx \\
&\geq 2\int_0^{1/\sqrt{n}} \left(1-nx^2\right) dx \\
&= \frac{4}{3\sqrt{n}} \\
&> \frac{1}{\sqrt{n}}.
\end{aligned}$$

Hence it follows that

$$c_n < \sqrt{n} \tag{7.4}$$

By observing Pic.7.4 and (7.4) we realize that for any $\delta > 0$, we have

$$Q_n(x) \leq \sqrt{n}\left(1-\delta^2\right)^n \quad (\delta \leq |x| \leq 1). \tag{7.5}$$

So that $Q_n \to 0$ uniformly in $\delta \leq |x| \leq 1$ since the term $(1-\delta^2)^n$ decreases at a faster rate than $\sqrt{n}$ increases (you can use L'Hospital rule to verify). Now set

$$P_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)dt \quad (0 \leq x \leq 1).$$

Why is $P_n(x)$ a polynomial of x? We do simple change of variables and let $u = x + t$. Then

$$P_n(x) = \int_{x-1}^{x+1} f(u)Q_n(u-x)du.$$

Notice that $x - 1 < 0$ and $x + 1 > 1$ since $x \in [0, 1]$. However, the function $f(x)$ is only nonzero on the interval $[0, 1]$. Hence the integral is zero outside of this region. Hence

$$P_n(x) = \int_0^1 f(u)Q_n(u-x)du = \int_0^1 f(t)Q_n(t-x)dt.$$

The last integral is a polynomial of x since $Q_n(t-x)$ is a polynomial of x and this is an integral with respect to t, hence we can take the powers of x outside of the integral, and the integration will only determine the coefficients of this polynomial. Hence P_n is a polynomial of x.

Since f is continuous on a compact interval, it is uniformly continuous on $[a, b]$. Given $\varepsilon > 0$, choose $\delta > 0$ such that

$$|y - x| < \delta \Rightarrow |f(y) - f(x)| < \frac{\varepsilon}{2}.$$

Also, f must also achieve its maximum. Set $M = \sup|f(x)|$. Using (7.5) and the fact that $f(x) = \int_{-1}^1 f(x)Q_n(x)dx$, we see that for $0 \le x \le 1$,

$$\begin{aligned}
|P_n(x) - f(x)| &= \left|\int_{-1}^1 [f(x+t) - f(x)]Q_n(t)dt\right| \\
&\le \int_{-1}^1 |f(x+t) - f(x)|Q_n(t)dt \\
&\le 2M\int_{-1}^{-\delta} Q_n(t)dt + \frac{\varepsilon}{2}\int_{-\delta}^{\delta} Q_n(t)dt + 2M\int_0^1 Q_n(t)dt \\
&\le 4M\sqrt{n}\left(1-\delta^2\right)^n + \frac{\varepsilon}{2} \\
&< \varepsilon
\end{aligned}$$

for large enough n such that the first term is smaller than $\varepsilon/2$. This proves the theorem. □

Remark 7.6.1. Let us reconsider why

$$P_n(x) = \int_{-1}^1 f(x+t)Q_n(t)dt \quad (0 \le x \le 1)$$

will be close to $f(x)$ when n is large.

When n is so large, from Pic.7.4, we can see that $Q_n(t) \to 0$ everywhere except where $t = 0$. Hence the integrand is close to zero everywhere except where $t = 0$. The value of $f(x+t)$ when $t = 0$ is $f(x)$. Hence, at the point $t = 0$, when n is so large, we will have

$$P_\infty(x) \approx f(x)\int_{-1}^1 Q_n(t)dt = f(x).$$

This may give some intuition. It also offers us some insight behind this construction.

Stone also generalized Thm.7.6.1, which is first proved by Weierstrass. His generalization is later on called the Stone Weierstrass Theorem. We will encounter this theorem again in real analysis, and Rudin has made a very good prove in the book. Since we only live so long, let us omit the proof and everything whatsoever and conclude this chapter.

Chapter 8

Some Special Functions

There cannot be a language more universal and more simple, more free from errors and obscurities...more worthy to express the invariable relations of all natural things [than mathematics]. [It interprets] all phenomena by the same language, as if to attest the unity and simplicity of the plan of the universe, and to make still more evident that unchangeable order which presides over all natural causes.

-Joseph Fourier

8.1 Power Series

A review of the power series section in Chapter 3 is recommended.

Definition 8.1.1. Functions represented by a series of this form

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n,$$

also known as the *power series*, are called *analytic functions*. If the series converges for $|x-a| < R$ for some $R > 0$, then f is said to be expanded in a power series about the point $x = a$. Usually we take $a = 0$ without loss of generality.

Lemma 8.1.1. *Every power series converges absolutely in the interior of its interval of convergence, by the root test.*

Proof. Let the radius of convergence be R. We can write any $x \in (-R, R)$ in the form of $R - \varepsilon$ for some $0 < \varepsilon < 2R$. Then the claim is that the series $\sum_n c_n (R-\varepsilon)^n$ converges absolutely.

Let $a_n = |c_n (R-\varepsilon)^n|$. Using the root test, we see that

$$\limsup_{n\to\infty} \sqrt[n]{|a_n|} = |R-\varepsilon| \limsup_{n\to\infty} \sqrt[n]{|c_n|} = \frac{|R-\varepsilon|}{R} < 1.$$

This completes the proof. □

Theorem 8.1.1. *Suppose the series*

$$\sum_{n=0}^{\infty} c_n x^n \tag{8.1}$$

converges for $|x| < R$*, and define*

$$f(x) = \sum_{n=0}^{\infty} c_n x^n \quad (|x| < R). \tag{8.2}$$

Then (8.1) converges uniformly on $[-R+\varepsilon, R-\varepsilon]$ *no matter which* $0 < \varepsilon < 2R$ *is chosen. The function* f *is continous and differentiable in* $(-R, R)$*, and*

$$f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1} \quad (|x| < R).$$

Proof. Let $\varepsilon > 0$ be given. For $|x| \leq R - \varepsilon$ we have

$$|c_n x^n| \leq |c_n (R-\varepsilon)^n|.$$

And since $\sum c_n (R-\varepsilon)^n$ converges absolutely in the interior of its interval of convergence by Lem.8.1.1, we know $\sum c_n x^n$ converges uniformly on $[-R+\varepsilon, R-\varepsilon]$ by Thm.7.1.3.

Since

$$\lim_{n\to\infty} \sqrt[n]{n} = 1,$$

we have

$$\limsup_{n\to\infty} \sqrt[n]{n\,|c_n|} = \limsup_{n\to\infty} \sqrt[n]{|c_n|}.$$

So that the series $\sum c_n x^n$ and $\sum n c_n x^{n-1}$ have the same interval (radius) of convergence.

Since $\sum n c_n x^{n-1}$ is a power series, it converges uniformly in $[-R+\varepsilon, R-\varepsilon]$ for every $\varepsilon > 0$. Hence we can change the order of the differentiation and the limit by Thm.7.4.1. Hence

$$f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1} \quad (|x| < R).$$

And the continuity of f follows from the existence of f' by Thm.5.1.1. □

Corollary 8.1.1. *Under the assumptions of the last theorem,* f *has derivatives of all orders in* $(-R, R)$*, which are given by*

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1) c_n x^{n-k} = k! \sum_{n=k}^{\infty} \binom{n}{k} c_n x^{n-k}.$$

In particular,

$$f^{(k)}(0) = k! c_k \quad (k = 0, 1, 2, \ldots).$$

Proof. This corollary follows if we apply Thm.8.1.1 successively to f, f', f''.... Putting $x = 0$ in the first equatiom gives the second equation. □

Remark 8.1.1. Notice that even though a function f may have derivatives of all orders, the series $\sum c_n x^n$, where $c_n = f^{(n)}(0)/n!$, need not converges to $f(x)$ for any $x \neq 0$. The following example demonstrates this point.

Example 8.1.1. Consider the piecewise function defined by

$$f(x) = \begin{cases} e^{-1/x^2} & (x \neq 0) \\ 0 & (x = 0). \end{cases}$$

This function has derivatives of all orders at $x = 0$, but $f^{(n)}(0) = 0$ for all $n = 1, 2, 3....$ Hence its power series expansion converges to f only when $x = 0$.

Theorem 8.1.2. *Suppose $\sum c_n$ converges. Put*

$$f(x) = \sum_{n=0}^{\infty} c_n x^n \quad (-1 < x < 1).$$

Then

$$\lim_{x \to 1} f(x) = \sum_{n=0}^{\infty} c_n.$$

Proof. Let

$$s_n = c_0 + c_1 + ... + c_n \quad \text{and} \quad s_{-1} = 0.$$

Then

$$\begin{aligned} \sum_0^m c_n x^n &= \sum_0^m (s_n - s_{n-1}) x^n \\ &= \sum_0^m s_n x^n - \sum_0^m s_{n-1} x^n \\ &= s_m x^m + \sum_0^{m-1} s_n x^n - \sum_0^{m-1} s_n x^{n+1} \\ &= (1-x) \sum_0^{m-1} s_n x^n + s_m x^m. \end{aligned}$$

For $|x| < 1$, we let $m \to \infty$ and we notice that $s_m x^m \to 0$. Therefore,

$$f(x) = (1-x) \sum_{n=0}^{\infty} s_n x^n.$$

Suppose $s = \lim_{n\to\infty} s_n$. Let $\varepsilon > 0$ be given. Choose N so that

$$n > N \Rightarrow |s - s_n| < \frac{\varepsilon}{2}.$$

Then, since

$$(1-x)\sum_{n=0}^{\infty} x^n = 1 \quad (|x|<1),$$

we obtain

$$|f(x)-s| = \left|(1-x)\sum_{n=0}^{\infty}(s_n - s)\, x^n\right| \le (1-x)\sum_{n=0}^{N}|s_n - s|\,|x|^n + \frac{\varepsilon}{2} \le \varepsilon$$

if $1-\delta < x < 1$ for some suitably chosen $\delta > 0$. This proves the theorem. □

Example 8.1.2. This example shows one beautiful application of Thm.8.1.2. We can find the power series representation of $\arctan(x)$. Since

$$\arctan'(x) = \frac{1}{1+x^2} = \sum_{n=0}^{\infty}(-x^2)^n,$$

for $|x|<1$, then

$$\arctan(x) = \sum_{n=0}^{\infty}\int_0^x (-x^2)^n dx = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{2n+1} \qquad |x|<1.$$

The series $\sum(-1)^n/(2n+1)$ is an alternating series whose magnitude of the successive terms are decreasing. Then by Thm.3.8.7, this series converges. Hence the power series of $\arctan(x)$ is continous at the endpoint $x=1$. Hence we get

$$\arctan(1) = \frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7}...$$

Theorem 8.1.3. *If $\sum a_n, \sum b_n, \sum c_n$ converge to A, B, C respectively, and if $c_n = a_0 b_n + ... + a_n b_0$, then $C = AB$.*

Proof. Define

$$f(x) = \sum_{n=0}^{\infty} a_n x^n, \quad g(x) = \sum_{n=0}^{\infty} b_n x^n, \quad h(x) = \sum_{n=0}^{\infty} c_n x^n$$

for $0 \le x \le 1$. For $x<1$, these series converges absolutely. Hence by Thm.3.8.11, we see that

$$f(x)\cdot g(x) = h(x) \quad (0 \le x < 1).$$

By Thm.8.1.2, we see that

$$f(x) \to A, \quad g(x) \to B, \quad h(x) \to C$$

when $x \to 1$. Hence this imply that $AB = C$. □

Now we ask the question whether the inversion of the order of summation is always legal. We know that it is legal for finite sums, however, for infinite summation, **the answer is no.** Consider the following example:

Example 8.1.3. Consider the double sequence defined by

$$a_{ij} = \begin{cases} 0 & (i < j) \\ -1 & (i = j) \\ 2^{j-i} & (i > j). \end{cases}$$

We can also think of a_{ij} being the number in the ith row and jth column of the array

$$\begin{array}{ccccc} -1 & 0 & 0 & 0 & \cdots \\ \frac{1}{2} & -1 & 0 & 0 & \cdots \\ \frac{1}{4} & \frac{1}{2} & -1 & 0 & \cdots \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{2} & -1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{array}$$

Notice that

$$\begin{aligned} \sum_j \sum_i a_{ij} &= \sum_j a_{1j} + a_{2j} + ... \\ &= (a_{11} + a_{21} + a_{31} + ...) + (a_{12} + a_{22} + a_{32} + ...) + (a_{13} + a_{23} + a_{33} + ...) \\ &= 0 + 0 + 0 + 0... = 0. \end{aligned}$$

whereas

$$\begin{aligned} \sum_i \sum_j a_{ij} &= \sum_i a_{i1} + a_{i2} + a_{i3} + ... \\ &= (a_{11} + a_{12} + a_{13} + ...) + (a_{21} + a_{22} + a_{23} + ...) + ... \\ &= (-1) + (-1 + 1/2) + (-1 + 1/2 + 1/4) + (-1 + 1/2 + 1/4 + 1/8)... \\ &= -1 - \frac{1}{2} - \frac{1}{4} - \frac{1}{8}... \\ &= (-1) \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = -2. \end{aligned}$$

Hence in general,

$$\sum_i \sum_j a_{ij} \neq \sum_j \sum_i a_{ij}.$$

Theorem 8.1.4. *Given a double sequence $\{a_{ij}\}$, suppose that*

$$\sum_{j=1}^{\infty} \left|a_{ij}\right| = b_i \quad (i = 1, 2, 3, \ldots) \tag{8.3}$$

and $\sum b_i$ converges, then

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

Proof. Let E be a countable set consisting of the points $x_0, x_1, x_2 ...$ and suppose that $x_n \to x_0$ as $n \to \infty$. Define

$$f_i\left(x_0\right) = \sum_{j=1}^{\infty} a_{ij} \quad (i = 1, 2, 3, \ldots), \tag{8.4}$$

$$f_i\left(x_n\right) = \sum_{j=1}^{n} a_{ij} \quad (i, n = 1, 2, 3, \ldots), \tag{8.5}$$

$$g(x) = \sum_{i=1}^{\infty} f_i(x) \quad (x \in E). \tag{8.6}$$

According to (8.4), (8.5), and (8.3), we have

$$\lim_{x_n \to x_0} f_i(x_n) = \lim_{n \to \infty} f_i(x_n) = \lim_{n \to \infty} \sum_{j=1}^{n} a_{ij} = f_i(x_0) \leq b_i.$$

Hence each f_i is continous at x_0. Since $|f_i(x)| \leq b_i$ for $x \in E$, then $g(x)$ converges uniformly, so that g is continous at x_0 by Cor.7.2.1. It follows that

$$\begin{aligned}
\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} &= \sum_{i=1}^{\infty} f_i\left(x_0\right) = g\left(x_0\right) = \lim_{n \to \infty} g\left(x_n\right) \\
&= \lim_{n \to \infty} \sum_{i=1}^{\infty} f_i\left(x_n\right) = \lim_{n \to \infty} \sum_{i=1}^{\infty} \sum_{j=1}^{n} a_{ij} \\
&= \lim_{n \to \infty} \sum_{j=1}^{n} \sum_{i=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.
\end{aligned}$$

Here we just switch the sums without proof since it involves rearrangement arguments. □

Theorem 8.1.5. ***(Taylor's Theorem)*** *Suppose $f(x) = \sum c_n x^n$ converging in $|x| < R$. If $-R < a < R$, then f can be expanded in a power series about the point $x = a$ which converges in $|x - a| < R - |a|$, and*

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n \quad (|x - a| < R - |a|).$$

Proof. We have

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} c_n[(x-a)+a]^n \\ &= \sum_{n=0}^{\infty} c_n \sum_{m=0}^{n} \binom{n}{m} a^{n-m}(x-a)^m \\ &= \sum_{m=0}^{\infty} \left[\sum_{n=m}^{\infty} \binom{n}{m} c_n a^{n-m} \right] (x-a)^m. \end{aligned}$$

This is the desired expansion about $x = a$ according to Cor.8.1.1. We need to justify the change which was made in the order of summation. Thm.8.1.4 tells us that this is permissible if

$$\sum_{n=0}^{\infty} \sum_{m=0}^{n} \left| c_n \binom{n}{m} a^{n-m}(x-a)^m \right|$$

converges. But this series is the same as

$$\sum_{n=0}^{\infty} |c_n| \cdot (|x-a| + |a|)^n,$$

which converges since $|x-a| + |a| < R$.

Note that in the last step the upper and lower limit changed when we switch the sum. We can explanin this by defining a double sequence

$$A_{n,m} = \begin{cases} \binom{n}{m} c_n a^{n-m}(x-a)^m & m \leq n \\ 0 & m > n \end{cases}.$$

Hence we are adding all the points under the line $n = m$. When we change the order of the summation and fix an m first, then in order to add everything under the line $n = m$, we have to start from $n = m$ to infinity. Also, m must also go to infinity. □

If two power series converge to the same function in $(-R, R)$, then Cor.8.1.1 tells us that they must have the same coefficients. This conclusion can be deduced, however, from a much weaker hypothesis.

Theorem 8.1.6. *Suppose $\sum a_n x^n, \sum b_n x^n$ converge in the segment $S = (-R, R)$. Let E be the set of all $x \in S$ at which*

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} b_n x^n \tag{8.7}$$

If E has a limit point in S, then $a_n = b_n$ for all n. Hence (8.7) holds for all $x \in S$.

Proof. **(Warning: Hard)** We put

$$f(x) = \sum_{n=0}^{\infty} (a_n - b_n)x^n = \sum_{n=0}^{\infty} c_n x^n \qquad (x \in S).$$

Then $f(x) = 0$ for all $x \in E$. Let

$$A = \text{the set of all limit points of } E \text{ in } S$$

and let $B = S \setminus A$. We notice several facts about A:
(1) $A \neq \emptyset$ by our assumption that E has at least one limit point in S.
(2) $A \cup B = S$ and $A \cap B = \emptyset$.
(3) $f(x) = 0$ for all $x \in A$.
(4) A is closed.

PROOF OF (3): Since every point $x \in A$ is a limit point of E, then for each x there exists a subsequence $\{x_n\}$ in E such that $x_n \to x$. Since f is continous on S, then

$$f(x) = f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} 0 = 0.$$

PROOF OF (4): Let a be a limit point of A, we want to show that $a \in A$. Since a is a limit point of A,

$$\exists\{x_n\} \to a \qquad \text{where } x_n \in A \ \forall n.$$

But since every x_n is a limit point of E,

$$\exists\{x_{n,k}\} \to x_n \qquad \text{where } x_{n,k} \in E \ \forall n, k.$$

Then consider the array

$$\begin{array}{cccccc}
x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & \cdots \longrightarrow & x_1 \\
x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & \cdots \longrightarrow & x_2 \\
x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & \cdots \longrightarrow & x_3 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \big\downarrow \\
\cdots & \cdots & \cdots & \cdots & \cdots & a
\end{array}$$

Hence if we go down the diagonal and construct $\{x_{n,n}\}$, it will converge to a. Hence there exists a sequence in E that converges to a, hence a is a limit point of E. Therefore, $a \in A$, and A is closed. Why will this sequence converge to a?

Since $x_n \to a$, there exists an N_1 such that

$$n \geq N_1 \Rightarrow |x_{N_1} - a| < \frac{\varepsilon}{2}.$$

Also, since $x_{n,n} \to x_n$, we can choose the first $N_2 \geq N_1$ such that

$$|x_{N_2,N_2} - x_{N_2}| < \frac{\varepsilon}{2}.$$

Then take $N = \max\{N_1, N_2\}$. For all $n \geq N$, we have

$$\begin{aligned}|x_{n,n} - a| &\leq |x_{n,n} - x_n| + |x_n - a| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Now, we claim that A **is open.** If A is open, then B is closed. Hence $A \cap \bar{B} = A \cap B = \emptyset$ according to (2). Also, since A is closed according to (4), we have $\bar{A} \cap B = A \cap B = \emptyset$. Hence A, B are seperated. But since $S = (-R, R)$ is connected, it cannot be the union of two seperated set. This means that one of A, B must be empty. Since A is not empty according to (1), this suggests that $B = \emptyset$ and $A = S$. And since $f(x) = 0$ for all $x \in A$ according to (3), $S = A$ implies that (8.7) holds for all $x \in S$, **and the proof is finished.** Therefore, all we need to show is that A is open.

If $x_0 \in A$, then Taylor's theorem says that we can expand f around x_0:

$$f(x) = \sum_{n=0}^{\infty} d_n (x - x_0)^n \quad (|x - x_0| < R - |x_0|).$$

WE CLAIM THAT $d_n = 0$ FOR ALL n. Suppose the contrary, let k be the smallest nonnegative integer such that $d_k \neq 0$. Then

$$f(x) = (x - x_0)^k g(x) \quad (|x - x_0| < R - |x_0|), \tag{8.8}$$

where

$$g(x) = \sum_{m=0}^{\infty} d_{k+m} (x - x_0)^m.$$

Since g is continous at x_0 and $g(x_0) = d_k \neq 0$, there exists a $\delta > 0$ such that

$$|x - x_0| < \delta \Rightarrow g(x) \neq 0. \tag{8.9}$$

According to (8.8), we have

$$\forall 0 < |x - x_0| < \delta \Rightarrow f(x) \neq 0.$$

But if $x_0 \in A$ (x_0 is a limit point of E), there exists a sequence $\{x_n\}$in E that converges to x_0. Hence there exists a N such that

$$n \geq N \Rightarrow |x_n - x_0| < \delta \qquad \text{and } f(x_n) = 0.$$

But (8.9) says that no $x \in (-\delta + x_0, \delta + x_0)$ will satisfy $f(x) = 0$. We reached a contradiction. Hence $d_n = 0$ for all n.

$d_n = 0$ for all n implies that

$$f(x) = 0 \qquad \forall x \in V_{R-|x_0|}(x_0),$$

hence x_0 is an interior point of A. Therefore, A is open. And this concludes the proof. □

Chapter 9

Functions of Several Variables

Beauty is the first test: there is no permanent place in the world for ugly mathematics.

-G.H.Hardy

9.1 Some Linear Algebra

We assume that reader has mastered basic linear algebra. Just some notation: We use captical letters (such at A) to denote linear maps, and bold letter to denote vectors (such as $\mathbf{x}$). The letter c, λ, ect will denote elements in the base field over which the vector space is defined. All vector spaces are assumed to be finite-dimensional.

Proposition 9.1.1. Let A be a linear operator on a finite dimensional vector space X. Then A is injective if and only if A is surjective.

Proof. Use the rank-nullity theorem: $\dim(\ker A) + \dim(\text{range } A) = \dim X$. Then A is injective if and only if $\dim(\ker A) = 0$, which holds if and only if $\dim(\text{range } A) = \dim X$. This proves the claim. $\square$

Definition 9.1.2. Let $L(X, Y)$ be the set of all linear transformations from X to Y. We can add scalar multiples of elements in $L(X, Y)$:

$$(c_1 A_1 + c_2 A_2)\,\mathbf{x} = c_1 A_1 \mathbf{x} + c_2 A_2 \mathbf{x} \quad (\mathbf{x} \in X)$$

and the multiplication is given by the composite

$$(BA)\mathbf{x} = B(A\mathbf{x}) \quad (x \in X).$$

This makes $L(X, Y)$ itself into a vector space.

Definition 9.1.3. For $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, we defined the norm $\|A\|$ of A to be $\|A\| = \sup_{|\mathbf{x}|\leq 1} |A\mathbf{x}|$. Here, $\mathbf{x}$ ranges over all vectors in $\mathbb{R}^n$ with $|\mathbf{x}| \leq 1$.

Proposition 9.1.4. $|A\mathbf{x}| \le \|A\||\mathbf{x}|$.

Proof. By the linearity of A we have

$$|A\mathbf{y}| = |\mathbf{y}|\left|A\left(\frac{\mathbf{y}}{|\mathbf{y}|}\right)\right| \le |\mathbf{y}| \sup_{\mathbf{y}\in\mathbb{R}^n}\left|A\left(\frac{\mathbf{y}}{|\mathbf{y}|}\right)\right| = |\mathbf{y}| \sup_{|\mathbf{x}|\le 1}|A\mathbf{x}| = |\mathbf{y}|\|A\|.$$

This proves the claim. $\square$

Proposition 9.1.5. If $A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$ and c is a scalar, then

$$\|A + B\| \le \|A\| + \|B\| \qquad \|cA\| = |c|\|A\|.$$

Proof. By the last Proposition we have

$$|(A + B)\mathbf{x}| = |A\mathbf{x} + B\mathbf{x}| \le |A\mathbf{x}| + |B\mathbf{x}| \le (\|A\| + \|B\|)|\mathbf{x}|.$$

Now the claim follows by taking the sup over all vectors of length less than or equal to one on both sides. The second part is similar. $\square$

Proposition 9.1.6. If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $B \in L(\mathbb{R}^m, \mathbb{R}^k)$, then we have $\|BA\| \le \|B\|\|A\|$.

Proof. $|(BA)\mathbf{x}| = |B(A\mathbf{x})| \le \|B\||A\mathbf{x}| \le \|B\|\|A\||\mathbf{x}|$. $\square$

Hence now we have defined a metric on the space $L(\mathbb{R}^n, \mathbb{R}^m)$, we can discuss the topology of the space and continuity of maps.

Theorem 9.1.7. *Let Ω be the set of all invertible linear operators on $\mathbb{R}^n$. Then*

(a) If $A \in \Omega$, $B \in L(\mathbb{R}^n)$, and

$$\|B - A\| < \frac{1}{\|A^{-1}\|},$$

then $B \in \Omega$.

(b) Ω is an open set of $L(\mathbb{R}^n)$ and the inversion map $A \mapsto A^{-1}$ is continuous on Ω.

Proof. Put $\|A^{-1}\| = 1/\alpha$ and $\|B - A\| = \beta$. Then by assumptipn we have $\beta < \alpha$. For every $\mathbf{x} \in \mathbb{R}^n$, we have

$$\alpha|\mathbf{x}| = \alpha|A^{-1}A\mathbf{x}| \le \alpha\|A^{-1}\| \cdot |A\mathbf{x}| = |A\mathbf{x}| \le |(A - B)\mathbf{x}| + |B\mathbf{x}| \le \beta|\mathbf{x}| + |B\mathbf{x}|.$$

Therefore we get that

$$(\alpha - \beta)|\mathbf{x}| \le |B\mathbf{x}|.$$

Since $\alpha - \beta > 0$, we have that $B\mathbf{x} \neq 0$ if $\mathbf{x} \neq 0$. Hence B is injective. By Proposition 9.1.1, B is surjective, hence invertible. Therefore, $B \in \Omega$. This proves part (a). For part (b), in order to prove the inversion map $A \mapsto A^{-1}$ is continuous, let α, β be the same quantity defined before. We consider B such that $\|B - A\| < \alpha$. By the equation above, replace $\mathbf{x}$ by $B^{-1}\mathbf{y}$ and we get

$$(\alpha - \beta)|B^{-1}\mathbf{y}| \le |BB^{-1}\mathbf{y}| = |\mathbf{y}|.$$

The identity

$$B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1}$$

gives

$$\|B^{-1} - A^{-1}\| \le \|B^{-1}\|\|A - B\|\|A^{-1}\| \le \frac{\beta}{\alpha(\alpha - \beta)}.$$

Since $\beta \to 0$ as $B \to A$, the proof is complete. $\square$

Remark 9.1.8. Theorem 9.1.7 part (a) actually says that Ω is an open set.

Theorem 9.1.9. *If $A = [a_{ij}]$ is a matrix representation of the linear map A, then*

$$\|A\| \le \left\{ \sum_{i,j} a_{ij}^2 \right\}^{1/2}.$$

Proof. Let $\mathbf{x} = (c_1, \cdots, c_n)$ and $A = [a_{ij}]$. Then we have

$$|A\mathbf{x}|^2 = \sum_i \left(\sum_j a_{ij} c_j \right)^2 \le \sum_i \left(\sum_j a_{ij}^2 \cdot \sum_j c_j^2 \right) = \sum_{i,j} a_{ij}^2 |\mathbf{x}|^2.$$

Now taking square roots on both sides and then take the sup gives the desired result. $\square$

9.2 Differentiation

We now extend our definition of a differentiable function on the real line to a much more general case.

Definition 9.2.1. Let $E \subseteq \mathbb{R}^n$ be an open set. We use boldface $\mathbf{f}$ to denote a vector-valued function. Suppose $\mathbf{f} : E \subseteq \mathbb{R}^n \to \mathbb{R}^m$, and $\mathbf{x} \in E$. If there exists a linear transformation $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that

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

then we say $\mathbf{f}$ is differentiable at $\mathbf{x}$, and we write $\mathbf{f}'(\mathbf{x}) = A$.

It is proven in Rudin Thm 9.12 that if $\mathbf{f}'(\mathbf{x})$ exists at some $\mathbf{x} \in E$, this linear transformation is unique. We notice that if $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, and $\mathbf{f}$ is differentiable on E, then the map $\mathbf{f} : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$. Since at each point $\mathbf{x} \in \mathbb{R}^n$, we have a linear transformation $\mathbf{f}'(\mathbf{x})$. Finally, we can write the definition as the following form

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{h} + \mathbf{r}(\mathbf{h})$$

where $\lim_{h \to 0} \frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|} = 0$. This shows that if $\mathbf{f}$ is differentiable, then it is continous.

Example 9.2.2. If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, then $\mathbf{f}'(\mathbf{x}) = A$ for all $\mathbf{x} \in \mathbb{R}^n$. Since $A(\mathbf{x} + \mathbf{h}) - A\mathbf{x} = A\mathbf{h}$, we are done.

Definition 9.2.3. Let $E \subseteq \mathbb{R}^n$ be an open subset and let $\mathbf{f} : E \to \mathbb{R}^m$. Let $\{\mathbf{e_1}, \cdots, \mathbf{e_n}\}$ and $\{\mathbf{u_1}, \cdots, \mathbf{u_m}\}$ be the standard basis for $\mathbb{R}^n, \mathbb{R}^m$. Then with respect to these basis we can write

$$\mathbf{f} = (f_1, \cdots, f_m)$$

where each $f_i : E \to \mathbb{R}$. Equivalently, $f_i(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{u_i}$. For $\mathbf{x} \in E$, we define the partial derivative of f_i with respect to x_j by

$$\frac{\partial f_i}{\partial x_j} = (D_j f_i)(\mathbf{x}) = \lim_{t \to 0} \frac{f_i(\mathbf{x} + t\mathbf{e_j})}{t}$$

provided the limit exists.

Theorem 9.2.4. *Suppose $\mathbf{f} : E \to \mathbb{R}^m$ as above, and $\mathbf{f}$ is differentiable at $\mathbf{x} \in E$, then the partial derivatives $(D_j f_i)$ exist, and they completely determine the linear transformation $\mathbf{f}'(\mathbf{x})$. More specifically,*

$$\mathbf{f}'(\mathbf{x}) = \begin{bmatrix} D_1 f_1 & \cdots & D_n f_1 \\ \vdots & \cdots & \vdots \\ D_1 f_m & \cdots & D_n f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

The proof is given in the book, and it's not difficult. So we omit it. However, it is worth mentioning some intuition involved. To find $\mathbf{f}'(\mathbf{x})$, we want to find the best linear approximation of the vector $\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x})$. Writting it out in matrix form, we find that

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x} + \mathbf{h}) - f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x} + \mathbf{h}) - f_m(\mathbf{x}) \end{bmatrix} \approx \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}.$$

This is because for each f_i, we have

$$f_i(\mathbf{x} + \mathbf{h}) - f_i(\mathbf{x}) = f_i(x_1 + h_1, \cdots, x_n + h_n) - f_i(x_1, \cdots, x_n) \approx \sum_{k=j}^{n} \frac{\partial f_i}{\partial x_j} h_j$$

in the first order approximation. Even though there is a remainder term $r(\mathbf{x})$, we see that $r(\mathbf{x})/\mathbf{x}$ will quickly decay to zero as $\mathbf{x}$ goes to zero.

Theorem 9.2.5. *Let $E \subseteq \mathbb{R}^n$ be an open set. Suppose $\mathbf{f} : E \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{f}$ is differentiable at $\mathbf{x_0} \in E$, $\mathbf{g}$ maps an open set containing $\mathbf{f}(E)$ to $\mathbb{R}^k$, and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x_0})$. Then the mapping $\mathbf{F} : E \to \mathbb{R}^k$ defined by*

$$\mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$$

is differentiable at $\mathbf{x_0}$ and

$$\mathbf{F}'(\mathbf{x_0}) = \mathbf{g}'(\mathbf{f}(\mathbf{x_0}))\, \mathbf{f}'(\mathbf{x_0}).$$

There is a somewhat technical proof in the book. Here, we convince the reader that this theorem is true, intuitively. Suppose the first order partial derivatives of all functions exist, then

$$\mathbf{F}'(\mathbf{x}) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial g_k}{\partial x_1} & \cdots & \frac{\partial g_k}{\partial x_n} \end{bmatrix},$$

since $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_m(\mathbf{x}))$ are just functions of $x_1, \cdots, x_n$ in each component, hence $\mathbf{g}(\mathbf{f}(\mathbf{x})) = (g_1(f_1, \cdots, f_n), \cdots, g_n(f_1, \cdots, f_n))$ should also be functions of $x_1, \cdots, x_n$ in each component. By the last theorem, $\mathbf{F}'(x)$ should be completely determined by the partial derivatives. Now, by the chain rule, we have that

$$\frac{\partial g_i}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial g_i}{\partial f_k}\frac{\partial f_k}{\partial x_j} = \sum_{k=1}^{m} (D_k g_i)(\mathbf{f}(\mathbf{x_0}))\frac{\partial f_k}{\partial x_j}.$$

Hence with some familiarity with matrix multiplication, one can see that

$$\mathbf{F}'(\mathbf{x}) = \underbrace{\begin{bmatrix} D_1 g_1 & \cdots & D_m g_1 \\ \vdots & \cdots & \vdots \\ D_1 g_k & \cdots & D_m g_k \end{bmatrix}}_{k\times m} \underbrace{\begin{bmatrix} D_1 f_1 & \cdots & D_n f_1 \\ \vdots & \cdots & \vdots \\ D_1 f_m & \cdots & D_n f_m \end{bmatrix}}_{m\times n} = \mathbf{g}'\left(\mathbf{f}\left(\mathbf{x}_0\right)\right)\mathbf{f}'\left(\mathbf{x}_0\right).$$

Definition 9.2.6. A differentiable map $\mathbf{f} : E \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is called continuously differentiable in E if $\mathbf{f}' : E \to L(\mathbb{R}^n, \mathbb{R}^m)$ is a continuous mapping. That is, for every $\mathbf{x} \in E$ and every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $|\mathbf{x} - \mathbf{y}| < \delta$ and $\mathbf{y} \in E$ we have

$$\|\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\| < \epsilon.$$

We write $\mathbf{f} \in \mathscr{C}'(E)$ for continuous differentiable map on E.

Theorem 9.2.7. *Let $\mathbf{f} : E \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Then $\mathbf{f} \in \mathscr{C}'(E)$ if and only if the partial derivatives $D_j f_i$ exists for all $1 \le i \le m, 1 \le j \le n$.*

Proof. Assume that $\mathbf{f} \in \mathscr{C}'(E)$. Then

$$\left(D_j f_i\right)(\mathbf{x}) = \left(\mathbf{f}'(\mathbf{x})\mathbf{e}_j\right) \cdot \mathbf{u}_i.$$

Hence

$$\left(D_j f_i\right)(\mathbf{y}) - \left(D_j f_i\right)(\mathbf{x}) = \left\{\left[\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\right]\mathbf{e}_j\right\} \cdot \mathbf{u}_i.$$

Since $|\mathbf{u}_i| = |\mathbf{e}_j| = 1$, we have

$$\left|\left(D_j f_i\right)(\mathbf{y}) - \left(D_j f_i\right)(\mathbf{x})\right| \le \left|\left[\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\right]\mathbf{e}_j\right| \le \left\|\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\right\|.$$

Hence $D_j f_i$ is continuous.

Conversely, it suffices to consider the case when $m = 1$ (i.e. $f : E \to \mathbb{R}$). Since if $\mathbf{f} : E \to \mathbb{R}^m$ for some $m > 1$, then each component of $\mathbf{f}$ will be a map from E to $\mathbb{R}$. Hence if the statement holds for $m = 1$, then each component of $\mathbf{f}$ with be $\mathscr{C}'$, thus so is $\mathbf{f}$.

Since E is open, there exists an open ball $B_r(\mathbf{x}) \subseteq E$, and the continuity of $D_j f_i$ shows that r can be chosen such that

$$\left|\left(D_j f\right)(\mathbf{y}) - \left(D_j f\right)(\mathbf{x})\right| < \frac{\epsilon}{n} \quad (\mathbf{y} \in B_r(\mathbf{x}), 1 \le j \le n).$$

Suppose $\mathbf{h} = \sum h_j \mathbf{e}_j$, and $|\mathbf{h}| < r$, put $\mathbf{v}_0 = \mathbf{0}$ and $\mathbf{v}_k = h_1\mathbf{e}_1 + \cdots + h_k\mathbf{e}_k$ for $1 \le k \le n$. Then

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \sum_{j=1}^{n} \left[f\left(\mathbf{x}+\mathbf{v}_j\right) - f\left(\mathbf{x}+\mathbf{v}_{j-1}\right)\right].$$

Since $|\mathbf{v}_k| < r$ for $1 \le k \le n$ and $B_r(\mathbf{x})$ is convex, the segment connecting $\mathbf{x}+\mathbf{v}_{j-1}$ and $\mathbf{x}+\mathbf{v}_j$ lies in $B_r(\mathbf{x})$. Also, since $\mathbf{v}_j = \mathbf{v}_{j-1} + h_j\mathbf{e}_j$, this means that

$$\mathbf{x} + \mathbf{v}_{j-1} + th_j\mathbf{e}_j \in B_r(\mathbf{x}) \qquad t \in [0,1].$$

Then define

$$g(t) = f(\mathbf{x} + \mathbf{v}_{j-1} + th_j\mathbf{e}_j).$$

Then by the mean value theorem there exists $\theta_j \in (0,1)$ such that

$$h_j\left(D_j f\right)\left(\mathbf{x}+\mathbf{v}_{j-1}+\theta_j h_j\mathbf{e}_j\right) = g(1) - g(0) = f\left(\mathbf{x}+\mathbf{v}_j\right) - f\left(\mathbf{x}+\mathbf{v}_{j-1}\right).$$

Since $\mathbf{x} + \mathbf{v}_{j-1} + \theta_j h_j\mathbf{e}_j \in B_r(\mathbf{x})$, the quantity on the LHS differs from $h_j(D_j f)(\mathbf{x})$ by less than $|h_j|\epsilon/n$. Hence it follows that

$$\left|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \sum_{j=1}^{n} h_j\left(D_j f\right)(\mathbf{x})\right| \le \frac{1}{n}\sum_{j=1}^{n}\left|h_j\right|\epsilon \le |\mathbf{h}|\epsilon$$

for all $\mathbf{h}$ such that $|\mathbf{h}| < r$. This shows that f is differentiable at $\mathbf{x}$ and

$$[f'(\mathbf{x})] = \left[(D_1 f)(\mathbf{x}) \quad \cdots \quad (D_n f)(\mathbf{x})\right].$$

Since all partial derivatives are continuous, we have $f'(\mathbf{x})$ continuous. This completes the proof. □

Remark 9.2.8. For the reverse direction of the last theorem, it is tempting to argue as follows:

Conversely, suppose that all partial derivatives exist and are continuous, we have that $\mathbf{f}'(\mathbf{x}) = [D_j f_i(\mathbf{x})]$. *Hence we have*

$$\left\|\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\right\| = \left\|\left[D_j f_i(\mathbf{y}) - D_j f_i(\mathbf{x})\right]\right\| \le \sum_{i,j}\left(D_j f_i(\mathbf{y}) - D_j f_i(\mathbf{x})\right)^2.$$

The last inequality follows from Theorem 9.1.9. Since eact $D_j f_i$ *is continuous, we see that the left hand side can be made arbitrary small provided that* $|\mathbf{y} - \mathbf{x}|$ *is small.*

Unfortunately, this argument is **Wrong**. Theorem 9.2.4 only tells us that if **f** is differentiable and all partial derivative exists, then the derivative is equal to the Jacobian matrix. However, the converse does not hold. A function can have all its partial derivatives, and fails to be differentiable. The example below is a good warning.

Example 9.2.9. Let $f(0,0) = 0$ and

$$f(x, y) = \frac{xy}{x^2 + y^2} \qquad \text{if } (x, y) \neq (0, 0).$$

Then we show that $(D_1 f)(x, y)$ and $(D_2 f)(x, y)$ exist at every point of $\mathbb{R}^2$, but f is not continous (hence not differentiable) at $(0, 0)$.

First, $f(x, y)$ is clearly not continuous, for if we choose $x = y$ and let $(x, y) = (x, x) \to (0, 0)$, then we have

$$|f(x, y) - f(0, 0)| = |f(x, x) - (0, 0)| = \frac{1}{2} \not< \epsilon.$$

Hence $f(x, y)$ is not continuous at $(0, 0)$. However, its partial derivatives exist at every point.

$$\begin{aligned}
(D_1 f)(x, y) &= \lim_{t \to 0} \frac{f(x + t, y) - f(x, y)}{t} \\
&= \lim_{t \to 0} \frac{\frac{(x+t)y}{(x+t)^2 + y^2} - \frac{xy}{x^2 + y^2}}{t} && \text{[of the form 0/0]} \\
&= \lim_{t \to 0} \frac{[(x + t)^2 + y^2]y - (x + t)y \cdot 2(x + t)}{[(x + t)^2 + y^2]^2} && \text{[L'Hospital's Rule]} \\
&= \frac{(x^2 + y^2)y - 2x^2 y}{(x^2 + y^2)^2} \\
&= \frac{y^3 - x^2 y}{(x^2 + y^2)^2}.
\end{aligned}$$

Hence $(D_1 f)(x, y)$ is defined everywhere when $(x, y) \neq (0, 0)$. To calculate $(D_2 f)(x, y)$, we have that

$$(D_2 f)(x, y) = \lim_{t \to 0} \frac{f(x, y + t) - f(x, y)}{t} = \frac{x^3 - y^2 x}{(y^2 + x^2)^2}.$$

We can either do a direct computation, or we observe that $f(x, y) = f(y, x)$, hence our calculation for $(D_1 f)(x, y)$ immediately turns into $(D_2 f)(x, y)$ if we do the substitution $x \mapsto y$ and $y \mapsto x$. Either way, we find that $(D_2 f)(x, y)$ exists if $(x, y) \neq (0, 0)$. Now we calculate the derivative at (0,0). We have

$$(D_1 f)(0, 0) = \lim_{t \to 0} \frac{f(t, 0) - f(0, 0)}{t} = \lim_{t \to 0} \frac{0 - 0}{t} = 0.$$

And similarly

$$(D_2 f)(0, 0) = \lim_{t \to 0} \frac{f(0, t) - f(0, 0)}{t} = \lim_{t \to 0} \frac{0 - 0}{t} = 0.$$

Thus both $(D_1 f)(x, y)$ and $(D_2 f)(x, y)$ exist at every point $(x, y) \in \mathbb{R}^2$. This proves the claim.

We end this section with one last theorem.

Theorem 9.2.10. *Suppose* $\mathbf{f}$ *maps a convex open set* $E \subseteq \mathbb{R}^n$ *to* $\mathbb{R}^m$. *If* $\mathbf{f}$ *is differentiable in* E, *and there is a real number* M *such that* $\|\mathbf{f}'(\mathbf{x})\| \leq M$ *for all* $\mathbf{x} \in E$, *then*

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|$$

for all $\mathbf{a}, \mathbf{b} \in E$.

Proof. Fix $\mathbf{a}, \mathbf{b} \in E$ and define a line segment

$$\gamma(t) = (1-t)\mathbf{a} + t\mathbf{b}$$

for all $t \in \mathbb{R}$. Since E is convex, the if $t \in [0,1]$, we have $\gamma(t) \subseteq E$. Then put

$$\mathbf{g}(t) = \mathbf{f}(\gamma(t)).$$

Taking the derivative by the chain rule gives

$$\mathbf{g}'(t) = \mathbf{f}'(\gamma(t))\gamma'(t) = \mathbf{f}'(\gamma(t))(\mathbf{b} - \mathbf{a}).$$

Hence

$$\left|\mathbf{g}'(t)\right| \leq \left\|\mathbf{f}'(\gamma(t))\right\| |\mathbf{b} - \mathbf{a}| \leq M|\mathbf{b} - \mathbf{a}|.$$

Notice that this is true for all $t \in [0,1]$. By the mean value theorem,

$$|\mathbf{g}(1) - \mathbf{g}(0)| \leq M|\mathbf{b} - \mathbf{a}|.$$

Notice that $\mathbf{g}(0) = \mathbf{f}(\mathbf{a})$ and $\mathbf{g}(1) = \mathbf{f}(\mathbf{b})$. This proves the claim. □

9.3 The Inverse Function Theorem

The key result in Rudin's PMA Chapter 9 is the inverse function theorem. The inverse function theorem says that a continuously differentiable mapping $\mathbf{f}$ is invertible in a neighborhood of any point $\mathbf{x}$ at which the linear transformation $\mathbf{f}(\mathbf{x})$ is invertible.

Let us first study two simple examples, which will give some intuition of what this theorem is saying.

- Suppose that $f : \mathbb{R} \to \mathbb{R}$. Then the derivative f' is a map from $\mathbb{R}$ to $L(\mathbb{R}, \mathbb{R})$, which is the operation of multiplying by the number $f'(x)$. Suppose that $f'(a)$ is invertible for some a (i.e. $f'(a) \neq 0$), then WLOG assume $f'(a) > 0$. Then since f is continuously differentiable (i.e. f' is continuous), then we can choose a neighborhood $B_\delta(a)$, such that $f'(x) > 0$ for all $x \in B_\delta(a)$. Hence f is a strictly increasing function on $B_\delta(a)$. Then of course f is bijective, hence invertible, when restricted to this interval $B_\delta(a)$.

- Suppose now $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ for some $A \in L(\mathbb{R}^n, \mathbb{R}^n)$. Then By Example 9.14 in Rudin, we know that $\mathbf{f}'(\mathbf{x}) = A'(\mathbf{x}) = A$. Hence if $\mathbf{f}'(\mathbf{a})$ is invertible for some $\mathbf{a} \in \mathbb{R}^n$, then A is invertible. Then clearly $\mathbf{f} = A$ is invertible restricted to any open set $\mathbf{a} \in U_\mathbf{a} \subseteq \mathbb{R}^n$.

Now we give the formal description of the inverse function theorem, with full generality. We divide in inverse function theorem into two parts, Thm.9.3.1 and Thm.9.3.2. Our mainly difficulties lie in the proof of Thm.9.3.1, which is quite technical.

Theorem 9.3.1. *Suppose* $\mathbf{f}$ *is a* $\mathscr{C}'$*-mapping of an open set* $E \subseteq \mathbb{R}^n$ *into* $\mathbb{R}^n$*, such that* $\mathbf{f}'(\mathbf{a})$ *is invertible at the point* $\mathbf{a} \in E$*. Then there exists open sets* U *and* V *in* $\mathbb{R}^n$ *such that* $\mathbf{a} \in U, \mathbf{b} \in V$*,* $\mathbf{f}$ *is injective restricted on* U*, and* $\mathbf{f}(U) = V$*.*

Proof. We will divide this proof in two steps. We put $\mathbf{f}'(\mathbf{a}) = A$, and choose λ so that

$$2\lambda\|A^{-1}\| = 1.$$

Since $\mathbf{f}'$ is continuous at $\mathbf{a}$, there is an open ball $U \subseteq E$, centered at $\mathbf{a}$, such that

$$\|\mathbf{f}'(\mathbf{x}) - A\| < \lambda \qquad (\mathbf{x} \in U).$$

- **STEP 1: SHOW $f : U \to \mathbb{R}^n$ IS INJECTIVE AND SURJECTIVE:**
  Take an arbitrary $\mathbf{y} \in \mathbb{R}^n$, for each $\mathbf{y}$ we associate it with a function φ, defined by

  $$\varphi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x})) \quad (\mathbf{x} \in E).$$

  Hence $\varphi_{\mathbf{y}}$ is a map from a neighborhood of $\mathbf{x}$ to a neighborhood of $\mathbf{x}$. We notice that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$ if and only if $\mathbf{x}$ is a fixed point of $\varphi_{\mathbf{y}}$. We show that $\varphi_{\mathbf{y}}$ is a contraction. Notice that by taking the derivative

  $$\varphi_{\mathbf{y}}'(\mathbf{x}) = I - A^{-1}\mathbf{f}'(\mathbf{x}) = A^{-1}\left(A - \mathbf{f}'(\mathbf{x})\right)$$

  Hence

  $$\|\varphi_{\mathbf{y}}'(\mathbf{x})\| \le \|A^{-1}\|\|A - \mathbf{f}'(\mathbf{x})\| = \frac{1}{2\lambda} \cdot \lambda = \frac{1}{2}.$$

  Hence the derivative is uniformly bounded, by Theorem 9.2.10 we know that

  $$|\varphi_{\mathbf{y}}(\mathbf{x_1}) - \varphi_{\mathbf{y}}(\mathbf{x_2})| \le \frac{1}{2}|\mathbf{x_1} - \mathbf{x_2}|.$$

  This shows that $\varphi_{\mathbf{y}}$ is indeed a contraction, and it has at most one fixed point in U. Hence for every $\mathbf{y} \in \mathbb{R}^n$, there is at most one $\mathbf{x} \in U$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$. This shows that $\mathbf{f} : U \to \mathbb{R}^n$ is injective. Notice that we do not know whether the fixed point in U exists for a given $\mathbf{y}$ yet, but we know if the fixed point exists, it is unique. So at most one point $\mathbf{x}$ will be mapped to $\mathbf{y}$ by $\mathbf{f}$.

  Of course, since we know $\mathbf{f} : U \to \mathbb{R}^n$ is injective, if we consider $\mathbf{f} : U \to \mathbf{f}(U)$, then it is also surjective, hence a bijection. Hence we choose $V = \mathbf{f}(U)$. By the claim of the theorem, we need to prove that $V = \mathbf{f}(U)$ is open.

- **STEP 2: SHOW THAT $V = \mathbf{f}(U)$ IS OPEN:**
  We pick any $\mathbf{y_0} \in V$, then since $V = \mathbf{f}(U)$, there exists some $\mathbf{x_0} \in U$ such that $\mathbf{y_0} = \mathbf{f}(\mathbf{x_0})$. We choose a positive number $r > 0$ so small that such that the closure if the ball centered at

$\mathbf{x_0}$ with radius r is still contained in U (i.e. $\overline{B_r(\mathbf{x_0})} \subseteq U$. We will show that if $\mathbf{y} \in B_{\lambda r}(\mathbf{y_0})$, then $\mathbf{y} \in V$, hence V is open.

We fix a $\mathbf{y}$ such that $|\mathbf{y} - \mathbf{y_0}| < \lambda r$. Then for this fixed $\mathbf{y}$, We consider the map $\varphi_{\mathbf{y}}$ defined above, we show that

$$\varphi_{\mathbf{y}} : \overline{B_r(\mathbf{x_0})} \to \overline{B_r(\mathbf{x_0})}$$

is a contraction on a complete metric space, hence we can use the fixed point theorem.

We have

$$\left|\varphi_{\mathbf{y}}(\mathbf{x}_0) - \mathbf{x}_0\right| = \left|A^{-1}(\mathbf{y} - \mathbf{y}_0)\right| < \left\|A^{-1}\right\| \lambda r = \frac{r}{2}.$$

Therefore, if we have $\mathbf{x} \in \overline{B_r(\mathbf{x_0})}$, then it follows from above that

$$\begin{aligned} \left|\varphi_{\mathbf{y}}(\mathbf{x}) - \mathbf{x}_0\right| &\le \left|\varphi_{\mathbf{y}}(\mathbf{x}) - \varphi_{\mathbf{y}}(\mathbf{x}_0)\right| + \left|\varphi_{\mathbf{y}}(\mathbf{x}_0) - \mathbf{x}_0\right| \\ &< \frac{1}{2}|\mathbf{x} - \mathbf{x}_0| + \frac{r}{2} \le r. \end{aligned}$$

This shows that if $\mathbf{x} \in \overline{B_r(\mathbf{x_0})}$, then $\varphi_{\mathbf{y}}(\mathbf{x}) \in \overline{B_r(\mathbf{x_0})}$ as well. Also, since $\overline{B_r(\mathbf{x_0})}$ is closed and bounded in $\mathbb{R}^n$, it is compact, hence complete. Then $\varphi_{\mathbf{y}}$ restricted on $\overline{B_r(\mathbf{x_0})}$ is a contraction on a complete metrix space, hence it has a unique fixed point $\mathbf{x} \in \overline{B_r(\mathbf{x_0})}$ (this $\mathbf{x}$ refers to the unique fixed point). And for this $\mathbf{x}$ we have $\mathbf{f}(\mathbf{x}) = \mathbf{y}$. Hence

$$\mathbf{y} \in \mathbf{f}(\overline{B_r(\mathbf{x_0})}) \subseteq \mathbf{f}(U) = V.$$

Hence this proves that V is open.

Hence, we have proven that $\mathbf{f}$ is injective in U in Step 1. And in Step 2, we proved that $V = \mathbf{f}(U)$ is open. Hence $\mathbf{f} : U \to V$ is an invertible map. This completes the proof. □

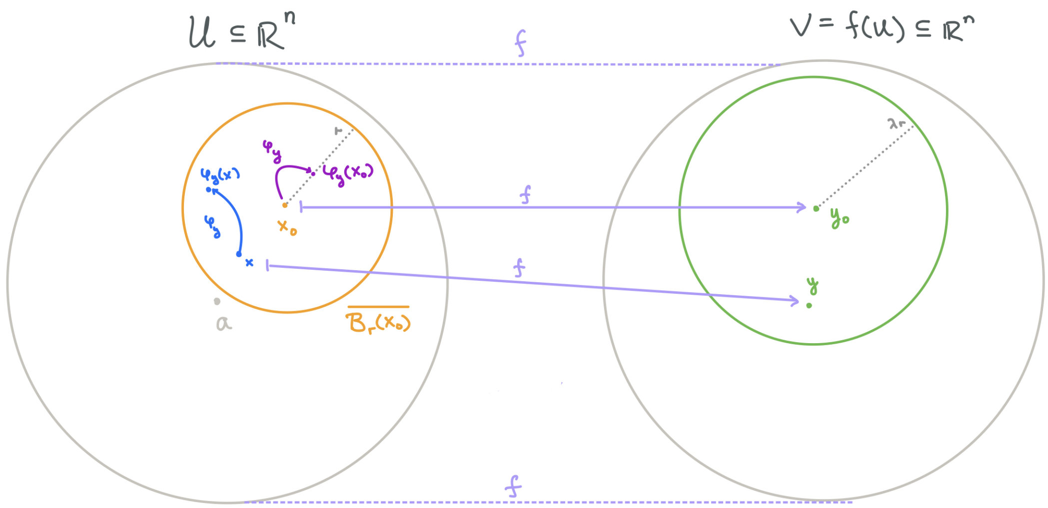
More specifically, if $\mathbf{y} \in V$, how do we find its inverse image? Notice that if we just pick $\mathbf{x_0} \in U$ and its image $\mathbf{y_0} = \mathbf{f}(\mathbf{x_0})$, then then clearly

$$\mathbf{f}(U) = V = \bigcup_{\mathbf{y_0} \in V} B_{\lambda r(\mathbf{y_0})}(\mathbf{y_0}),$$

where the radius r may depend on $\mathbf{y}_0$. Then the corresponding $\overline{B_{r(\mathbf{y_0})}(\mathbf{x_0})}$ will contain a fixed point $\mathbf{x}$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{y}$. Notice that $\varphi_{\mathbf{y}}$ will map $\mathbf{x} \in \overline{B_{r(\mathbf{y_0})}(\mathbf{x_0})}$ implies $\varphi_{\mathbf{y}}(\mathbf{x}) \in \overline{B_{r(\mathbf{y_0})}(\mathbf{x_0})}$, since everything inside $\overline{B_{r(\mathbf{y_0})}(\mathbf{x_0})}$ will be mapped back to $\overline{B_{r(\mathbf{y_0})}(\mathbf{x_0})}$. This can be illustrated by the following picture.

Hence if we consider $\mathbf{y}$ only in the ball of a chosen center $B_{\lambda r(\mathbf{y_0})}(\mathbf{y_0})$, then we can always find its inverse image on the corresponding ball $\overline{B_{r(\mathbf{y_0})}(\mathbf{x_0})}$. Hence, if we let the center of the ball $\mathbf{y_0}$ run over all V, then we can find the inverse image of any $\mathbf{y} \in V$ in U. Notice that this process is more complicated than we thought, even though we know that if $\mathbf{f} : U \to \mathbb{R}^n$ is injective, then $\mathbf{f} : U \to \mathbf{f}(U)$ is bijective, thus invertible.

Now we present the second part of the inverse function theorem.



Theorem 9.3.2. *Suppose* $\mathbf{f}$ *is a* $\mathscr{C}'$*-mapping of an open set* $E \subseteq \mathbb{R}^n$ *into* $\mathbb{R}^n$*, such that* $\mathbf{f}'(\mathbf{a})$ *is invertible at the point* $\mathbf{a} \in E$*. If* $\mathbf{g}$ *is the inverse of* $\mathbf{f}$ *(which exists by Thm.9.3.1), defined in* V *by* $\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x}$ *for* $\mathbf{x} \in U$*, then* $\mathbf{g} \in \mathscr{C}'(V)$*.*

Proof. Pick $\mathbf{y}, \mathbf{y} + \mathbf{k} \in V$. Then there exists $\mathbf{x}, \mathbf{x} + \mathbf{h} \in U$ such that $\mathbf{y} = \mathbf{f}(\mathbf{x})$, and similarly $\mathbf{y} + \mathbf{k} = \mathbf{f}(\mathbf{x} + \mathbf{h})$. Let $\varphi = \varphi_{\mathbf{y}}$ be defined as before, we have

$$\varphi(\mathbf{x} + \mathbf{h}) - \varphi(\mathbf{x}) = \mathbf{h} + A^{-1}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x} + \mathbf{h})] = \mathbf{h} - A^{-1}\mathbf{k}.$$

By the previous result, we know that

$$|\varphi(\mathbf{x} + \mathbf{h}) - \varphi(\mathbf{x})| = \left|\mathbf{h} - A^{-1}\mathbf{k}\right| \leq \frac{1}{2}|\mathbf{h}|.$$

By the triangle inequality we have

$$|\mathbf{h}| - |A^{-1}\mathbf{k}| \leq |\mathbf{h} - A^{-1}\mathbf{k}| \leq \frac{1}{2}|\mathbf{h}|.$$

Therefore,

$$|A^{-1}\mathbf{k}| \geq \frac{1}{2}|\mathbf{h}|.$$

Hence

$$|\mathbf{h}| \leq 2\left\|A^{-1}\right\| |\mathbf{k}| = \lambda^{-1}|\mathbf{k}|.$$

Since we know that

$$|\mathbf{f}'(\mathbf{x}) - A| < \lambda = \frac{1}{2\|A^{-1}\|} \qquad (\mathbf{x} \in U),$$

by Thm.9.1.7 we know $\mathbf{f}'(\mathbf{x})$ has an inverse, say T. Since

$$\mathbf{g}(\mathbf{y} + \mathbf{k}) - \mathbf{g}(\mathbf{y}) - T\mathbf{k} = \mathbf{h} - T\mathbf{k} = -T\left[\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{f}'(\mathbf{x})\mathbf{h}\right],$$

the inequality before implies

$$\frac{|\mathbf{g}(\mathbf{y}+\mathbf{k})-\mathbf{g}(\mathbf{y})-T\mathbf{k}|}{|\mathbf{k}|} \le \frac{\|T\|}{\lambda}\cdot\frac{|\mathbf{f}(\mathbf{x}+\mathbf{h})-\mathbf{f}(\mathbf{x})-\mathbf{f}'(\mathbf{x})\mathbf{h}|}{|\mathbf{h}|}.$$

As $\mathbf{k}\to\mathbf{0}$, the inequality $|\mathbf{h}|\le\lambda^{-1}|\mathbf{k}|$ shows that $\mathbf{h}\to\mathbf{0}$. Hence the RHS of this inequality tends to zero, so does the LHS. This shows that $\mathbf{g}'(\mathbf{y})=T$. But T was chosen to be the inverse of $\mathbf{f}'(\mathbf{x})=\mathbf{f}'(\mathbf{g}(\mathbf{y}))$. Thus

$$\mathbf{g}'(\mathbf{y})=\left\{\mathbf{f}'(\mathbf{g}(\mathbf{y}))\right\}^{-1}\quad(\mathbf{y}\in V).$$

Finally, since $\mathbf{g}$ is continuous (for $\mathbf{g}$ is differentiable), $\mathbf{f}'$ is continuous, and inversion is also continuous by Thm.9.1.7. Thus we see that $\mathbf{g}\in\mathscr{C}'(V)$. □

Remark 9.3.3. Actually, the crucial part is to prove that $\mathbf{g}$ is differentiable. Once we know that, we can take derivatives. Since $\mathbf{g}(\mathbf{f}(\mathbf{x}))=\mathbf{x}=I\mathbf{x}$, by taking derivative we have

$$\mathbf{g}'(\mathbf{y})\mathbf{f}'(\mathbf{x})=I.$$

This gives

$$\mathbf{g}'(\mathbf{y})=\{\mathbf{f}'(\mathbf{x})\}^{-1}.$$

This gives an expression for the derivative of $\mathbf{g}$. And once we know this, since $\mathbf{f}'(\mathbf{x})$ is continous, and $A\mapsto A^{-1}$ is continous, we know that $\mathbf{g}'(\mathbf{y})$ is continous, hence $\mathscr{C}'$.

Corollary 9.3.4. *If* $\mathbf{f}$ *is a* $\mathscr{C}'$*-mapping of an open set* $E\subseteq\mathbb{R}^n$ *into* $\mathbb{R}^n$ *and if* $\mathbf{f}'(\mathbf{x})$ *is invertible for every* $\mathbf{x}\in E$*, then* $\mathbf{f}(W)$ *is an open subset of* $\mathbb{R}^n$ *for every open set* $W\subseteq E$*. In other words,* $\mathbf{f}$ *maps open sets to open sets.*

Proof. This is immediate from Thm.9.3.1. □

9.4 The Implicit Function Theorem

The implicit function theorem is also a central result in the theory of multivariable functions. To put it in the simple term, if f is a $\mathscr{C}'$ mapping in the plane, then locally, the equation $f(x,y)=0$ can be solved for y in terms of x in a neighborhood of any point (a,b) with $f(a,b)=0$ and $\partial f/\partial y\neq 0$. The proof is an (clever) application of the inverse function theorem.

To elaborate on the simple case, consider the function

$$f(x,y)=x^2+y^2-1.$$

Then if $f(x,y)=0$, we know that the intersection of f with the plane at zero is a circle $x^2+y^2=1$. Then locally on this circle, we can either express $y=\pm\sqrt{1-x^2}$ in terms of x or $x=\pm\sqrt{1-y^2}$ in terms of y. Before we give the full description and proof of the theorem, we need some notation.

Definition 9.4.1. If $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, \cdots, y_m) \in \mathbb{R}^m$, Then since $\mathbb{R}^{n+m} \cong \mathbb{R}^n \times \mathbb{R}^m$, we identify $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ as a point of $\mathbb{R}^{n+m}$:

$$(x_1, \cdots, x_n, y_1, \cdots, y_m) \in \mathbb{R}^{n+m}.$$

Every linear transformation $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^m)$ (which is an $m \times (n+m)$ matrix) can be split into a linear transformation $A_x \in L(\mathbb{R}^n, \mathbb{R}^m)$ (an $m \times n$ matrix) and a linear transformation $A_y \in L(\mathbb{R}^m, \mathbb{R}^m)$ (an $m \times m$ matrix). That is, we can split the matrix

$$[A] = [A_x, A_y].$$

Now we present the implicit function theorem.

Theorem 9.4.2 (implicit function theorem)**.** *Let* $\mathbf{f} : E \subseteq \mathbb{R}^{n+m} \to \mathbb{R}^m$ *be a* $\mathscr{C}'$ *mapping, such that* $\mathbf{f}(\mathbf{a}, \mathbf{b}) = \mathbf{0}$ *for some point* $(\mathbf{a}, \mathbf{b}) \in E$. *Put* $A = \mathbf{f}'(\mathbf{a}, \mathbf{b}) = [A_x, A_y]$ *and assume that* A_y *is invertible. Then there exist an open set* $U \subseteq \mathbb{R}^{n+m}$, *with* $(\mathbf{a}, \mathbf{b}) \in U$, *an open set* $W \subseteq \mathbb{R}^n$, *and a* $\mathscr{C}'$ *mapping*

$$\mathbf{g} : W \to \mathbb{R}^m$$

having the following properties:

- $\mathbf{g}(\mathbf{a}) = \mathbf{b}$*;*
- $\mathbf{f}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = 0$ *for all* $\mathbf{x} \in W$*;*
- *If* $(\mathbf{x}, \mathbf{y}) \in U$ *and* $\mathbf{f}(\mathbf{x}, \mathbf{y}) = 0$, *then* $\mathbf{g}(\mathbf{x}) = \mathbf{y}$*;*
- $\mathbf{g}'(\mathbf{a}) = -(A_y)^{-1} A_x$.

Proof. As we said before, the proof requires inverse function theorem. Hence we should look for a map from $\mathbb{R}^{n+m}$ to $\mathbb{R}^{n+m}$. Consider, therefore, the map

$$\begin{aligned} \mathbf{F} : \qquad E &\longrightarrow \mathbb{R}^{n+m} \\ (\mathbf{x}, \mathbf{y}) &\longmapsto (\mathbf{x}, \mathbf{f}(\mathbf{x}, \mathbf{y})). \end{aligned}$$

Hence we have $\mathbf{F}(\mathbf{a}, \mathbf{b}) = (\mathbf{a}, \mathbf{0})$. Now, clearly $\mathbf{F}$ is $\mathscr{C}'$ since all of its components are $\mathscr{C}'$. Thus we can take its derivative, which is given by the block matrix

$$\mathbf{F}'(\mathbf{a}, \mathbf{b}) = \begin{bmatrix} I_{n \times n} & 0 \\ A_x & A_y \end{bmatrix}.$$

To see this, we can write out the component of $\mathbf{F}$. Let

$$\begin{aligned} &\mathbf{F}(x_1, \cdots, x_n, y_1, \cdots, y_m) \\ &= (x_1, \cdots, x_n, f_1(x_1, \cdots, x_n, y_1, \cdots, y_m), \cdots, f_m(x_1, \cdots, x_n, y_1, \cdots, y_m)). \end{aligned}$$

And the rest follows from the definition of jacobian matrix, and A_x, A_y. Clearly the block matrix $\mathbf{F}'(\mathbf{a}, \mathbf{b})$ is invertible, since its determinant is equal to the determinant of the block:

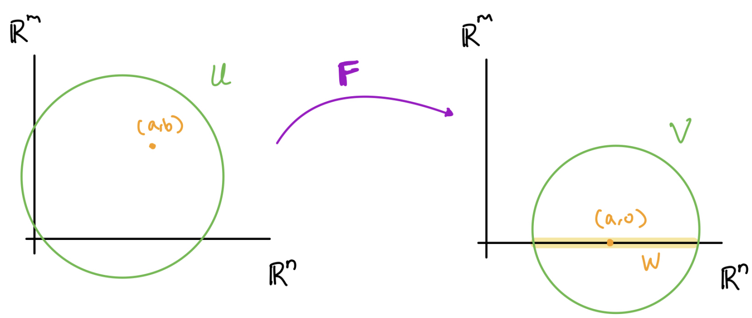
$$\det\big(\mathbf{F}'(\mathbf{a}, \mathbf{b})\big) = \det\big(A_y\big) \neq 0$$

by assumption. Moreover, its inverse is given by

$$[\mathbf{F}'(\mathbf{a},\mathbf{b})]^{-1} = \begin{bmatrix} I_{n\times n} & 0 \\ -(A_y)^{-1}A_x & A_y^{-1} \end{bmatrix}.$$

Again, this is just simple linear algebra. Hence by the inverse function theorem (Theorem 9.3.1), there exists open sets $U \subseteq \mathbb{R}^{n+m}$ containing $(\mathbf{a},\mathbf{b})$, and open set $V \subseteq \mathbb{R}^{n+m}$ containing $\mathbf{F}(\mathbf{a},\mathbf{b}) = (\mathbf{a},\mathbf{0})$, such that $\mathbf{F} : U \to V$ is a bijection with a $\mathscr{C}'$ inverse.

Now that we find an open set $U \subseteq \mathbb{R}^{n+m}$ (good), and an open set $V \subseteq \mathbb{R}^{n+m}$ (not quite, since we want an open set $W \subseteq \mathbb{R}^n$). Now, since in the image V of $\mathbf{F}(\mathbf{x},\mathbf{y}) = (\mathbf{x},\mathbf{f}(\mathbf{x},\mathbf{y}))$, we are only interested in the point such that $\mathbf{f}(\mathbf{x},\mathbf{y}) = 0$, therefore, we should consider the projection of V on $\mathbb{R}^n$. This is shown in the following figure:



Therefore, let our desired open set in $\mathbb{R}^n$ be

$$W = \{\mathbf{x} \in \mathbb{R}^n \,|\, (\mathbf{x},\mathbf{0}) \in V\}.$$

Since W is the projection of an open set V, it is open. For every $\mathbf{x} \in W$, then $(\mathbf{x},\mathbf{0}) \in V$. Hence there exists a unique $\mathbf{y}$ (with $(\mathbf{x},\mathbf{y}) \in \mathbf{U}$) such that $\mathbf{F}(\mathbf{x},\mathbf{y}) = (\mathbf{x},\mathbf{0})$. Hence we can define a function $\mathbf{g} : W \to \mathbb{R}^m$ that associate each $\mathbf{x}$ with its unique corresponding $\mathbf{y}$.

Now we have defined our $\mathbf{g}(\mathbf{x})$. Let us check the properties in the theorem.

- $\mathbf{g}(\mathbf{a}) = \mathbf{b}$ since $\mathbf{F}(\mathbf{a},\mathbf{b}) = (\mathbf{a},\mathbf{0})$;
- By definition $\mathbf{g}(\mathbf{x})$ is the unique vector in $\mathbf{R}^m$ such that $\mathbf{f}(\mathbf{x},\mathbf{g}(\mathbf{x})) = 0$.
- We verify first that $\mathbf{g}$ is $\mathscr{C}'$. Since we can apply the inverse and get $(\mathbf{x},\mathbf{g}(\mathbf{x})) = \mathbf{F}^{-1}(\mathbf{x},\mathbf{0})$, with $\mathbf{F}^{-1}$ being $\mathscr{C}'$ by the inverse function theorem, we have that the RHS is $\mathscr{C}'$. Hence every component on the LHS must be $\mathscr{C}'$. Therefore $\mathbf{g}$ is $\mathscr{C}'$. The third property simply follows from the definition of $\mathbf{g}$.
- Again we have $(\mathbf{x},\mathbf{g}(\mathbf{x})) = \mathbf{F}^{-1}(\mathbf{x},\mathbf{0})$. Thus we can define a function $\Phi(\mathbf{x}) = (\mathbf{x},\mathbf{g}(\mathbf{x}))$ (the LHS). Then its Jacobian matrix is given by

$$\Phi'(\mathbf{a}) = \begin{bmatrix} I_{n\times n} \\ \mathbf{g}'(\mathbf{a}) \end{bmatrix} = \underbrace{\begin{bmatrix} I_{n\times n} & 0 \\ -(A_y)^{-1}A_x & A_y^{-1} \end{bmatrix}}_{\mathbf{F}'^{-1}} \underbrace{\begin{bmatrix} I_{n\times n} \\ 0 \end{bmatrix}}_{[\mathbf{x}\mapsto(\mathbf{x},\mathbf{0})]'} = \begin{bmatrix} I_{n\times n} \\ -(A_y)^{-1}A_x \end{bmatrix},$$

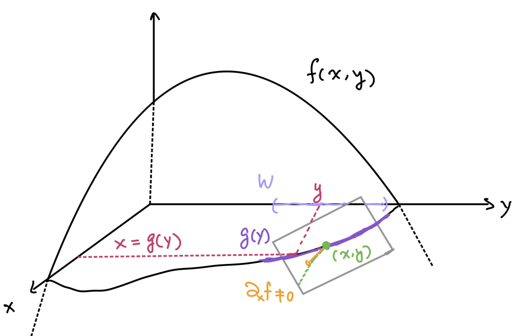
which is equal to the derivative of the RHS. In the second step, we used the fact that $\mathbf{F}^{-1'} = \mathbf{F}'^{-1}$ and $\mathbf{F}'^{-1}$ just follows from the previous work. Hence we have $\mathbf{g}'(\mathbf{a}) = -(A_y)^{-1}A_x$.

This completes the proof of the theorem. □

We can consider what this means in the case when Let $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose $f(x, y) = 0$ at some point (x, y) and the derivative at that point is given by

$$f'(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}.$$

Hence A_x is invertible if and only if $\frac{\partial f}{\partial x} \neq 0$. Thus, we use approximate $f(x, y)$ by its tangent plane $T_{(x,y)}f$ near (x, y), this plane will not be parallel to the $x - z$ plane, i.e. f has no flat tangent in the x direction near the point (x, y). Then locally near (x, y), the level curve of $f(x, y)$ can be viewed as a curve (function) parametrized by y. Hence we can find an open set W such that for all $y \in W$, we have $f(g(y), y) = 0$. This is shown by the following figure.



Remark 9.4.3. In Rudin, the proof is simlar, except that we are looking at functions $\mathbf{f} : E \subseteq \mathbb{R}^{n+m} \to \mathbb{R}^n$ and assuming that A_x is invertible. Hence the theorem becomes the following:

Theorem 9.4.4 (Rudin Version of the inverse function theorem)**.** *Let $\mathbf{f} : E \subseteq \mathbb{R}^{n+m} \to \mathbb{R}^n$ be a $\mathscr{C}'$ mapping, such that $\mathbf{f}(\mathbf{a}, \mathbf{b}) = \mathbf{0}$ for some point $(\mathbf{a}, \mathbf{b}) \in E$. Put $A = \mathbf{f}'(\mathbf{a}, \mathbf{b}) = [A_x, A_y]$ and assume that A_x is invertible. Then there exist an open set $U \subseteq \mathbb{R}^{n+m}$, with $(\mathbf{a}, \mathbf{b}) \in U$, an open set $W \subseteq \mathbb{R}^m$, and a $\mathscr{C}'$ mapping*

$$\mathbf{g} : W \to \mathbb{R}^n$$

having the following properties:

- $\mathbf{g}(\mathbf{b}) = \mathbf{a}$*;*
- $\mathbf{f}(\mathbf{g}(\mathbf{y}), \mathbf{y}) = 0$ *for all* $\mathbf{y} \in W$*;*
- *If* $(\mathbf{x}, \mathbf{y}) \in U$ *and* $\mathbf{f}(\mathbf{x}, \mathbf{y}) = 0$*, then* $\mathbf{g}(\mathbf{y}) = \mathbf{x}$*;*
- $\mathbf{g}'(\mathbf{b}) = -(A_x)^{-1}A_y$.

Example 9.4.5. Consider the equation

$$G(x, y, z) = x^3 + 3y^2 + 4xz^2 - 3z^2y - 1 = 0.$$

Does this equation define a function $z = z(x, y)$ in a neighborhood of $A = (x = 1, y = 1)$ and $B = (x = 1, y = 0)$?

Solution. When $x = 1, y = 1$, we substitute this into the equation and get $1+3+4z^2-3z^2-1 = 0$, or equivalently $3 + z^2 = 0$. Hence no solution of the initial point (x_0, y_0, z_0) in this case. When $x = 1, y = 0$ we get the equation $z^2 = 0$. Hence $z = 0$. Thus the initial point is $(x_0, y_0, z_0) = (1, 0, 0)$. However,

$$\frac{\partial G}{\partial z} = (8xz - 6yz)|_{(1,0,0)} = 0.$$

Hence the implicit function theorem fails in this case. $\square$

Example 9.4.6. Can the following system of equations

$$\begin{aligned} x^2 - y^2 - u^3 + v^2 + 4 &= 0 \\ 2xy + y^2 - 2u^2 + 3v^4 + 8 &= 0 \end{aligned}$$

be solved for u, v in terms of x, y in a neighborhood of $(x, y, u, v) = (2, -1, 2, 1)$?

Solution. Define a function

$$\mathbf{F}(x, y, u, v) = \left(x^2 - y^2 - u^3 + v^2 + 4, 2xy + y^2 - 2u^2 + 3v^4 + 8\right).$$

So that $\mathbf{F}(2, -1, 2, 1) = (0, 0)$. We compute its derivative

$$\mathbf{F}'(x, y, u, v) = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{bmatrix} = \begin{bmatrix} 2x & -2y & -3u^2 & -2v \\ 2y & 2x + 2y & -4u & 12v^3 \end{bmatrix}.$$

Since we are interested in expressing u, v in terms of x, y, we consider the submatrix

$$A = \left[\begin{array}{cc} -3u^2 & 2v \\ -4u & 12v^3 \end{array}\right]\Bigg|_{(2,-1,2,1)} = \left[\begin{array}{cc} -12 & 2 \\ -8 & 12 \end{array}\right].$$

Computation shows that $\det A = -128 \neq 0$. Hence by the implicit function theorem, we can find functions $u(x, y)$ and $v(x, y)$ in a neighborhood of $(x_0, y_0) = (2, -1)$ such that the equation above are satisfied. $\square$

9.5 Determinants

As we previously said, we assume the reader has mastered basic linear algebra. Hence the reader should be familiar with determinant and all its properties. But for the sake of completeness, we mention briefly the definitions, and quickly summarized the key result without proof.

Definition 9.5.1. Let $(j_1, \cdots, j_n)$ be an ordered n-tuple of integers. We define

$$s\left(j_1, \ldots, j_n\right) = \prod_{p<q} \operatorname{sgn}\left(j_q - j_p\right).$$

Here the sgn function is defined to be

$$\operatorname{sgn} x = \begin{cases} 1 & x > 0, \\ 0 & x = 0, \\ -1 & x < 0. \end{cases}$$

Then let A be a linear operator on $\mathbb{R}^n$, with matrix representation $[A]$ with respect to the standard basis. Let $a(i, j)$ denote the entries in the ith row and jth column. The determinant of $[A]$ is defined to be

$$\det[A] = \sum_{(j_1, \cdots, j_n)} s\left(j_1, \ldots, j_n\right) a\left(1, j_1\right) a\left(2, j_2\right) \cdots a\left(n, j_n\right).$$

The sum runs over all the ordered n-tuples.

This way of defining the determinant may seem unfamilar, but we'll quickly do an example to see that it is really the same definition that we are familiar with.

Example 9.5.2. Let A be a linear operator on $\mathbb{R}^2$. Then we consider all the ordered 2-tuples (j_1, j_2) of $\{1, 2\}$. Since if $j_1 = j_2$, by definition $s(j_1, j_2) = 0$, it suffices to consider two cases: $j_1 = 1, j_2 = 2$, then $s(j_1, j_2) = 1$, and $j_1 = 2, j_2 = 1$, then $s(j_1, j_2) = -1$. Hence

$$\det[A] = \sum_{(j_1, j_2)} s\left(j_1, j_2\right) a\left(1, j_1\right) a\left(2, j_2\right) = a(1,1)a(2,2) - a(1,2)a(2,1).$$

Now we will quickly recall some properties of the determinant.

Theorem 9.5.3 (properties of the determinant)**.** *Let A, B be linear operators on $\mathbb{R}^n$, whose matrix with respect to the standard basis is $[A], [B]$. Let I denote the identity (operator) matrix.*

- $\det I = 1$*;*
- $\det(\mathbf{x_1}, \cdots, \mathbf{x_n})$ *is a linear function on each column vector* $\mathbf{x_j}$*;*
- $\det\big(\mathbf{x_1}, \cdots, \mathbf{x_i}, \cdots, \mathbf{x_j}, \cdots, \mathbf{x_n}\big) = -\det\big(\mathbf{x_1}, \cdots, \mathbf{x_j}, \cdots, \mathbf{x_i}, \cdots, \mathbf{x_n}\big)$*;*
- *If* $[A]$ *has two equal columns, then* $\det[A] = 0$*;*
- $\det([B][A]) = \det[B] \det[A]$*;*
- *A is invertible if and only if* $\det[A] \neq 0$*;*
- *Let β be a new basis for $\mathbb{R}^n$, then the determinant is invariant under change of basis. That is,* $\det[A]_\beta = \det[A]$.

The proofs are standard in any linear algebra textbooks. So we omit them. With determinant, we can define an important quantity called the Jacobian.

Definition 9.5.4. If $\mathbf{f} : E \subseteq \mathbb{R}^n \to \mathbb{R}^k$ and $\mathbf{f}$ is differentiable at a point $\mathbf{x} \in E$, the determinant of the linear operator $\mathbf{f}'(\mathbf{x})$ is called the Jacobian of $\mathbf{f}$ at $\mathbf{x}$. In symbols,

$$J_{\mathbf{f}}(\mathbf{x}) = \det \mathbf{f}'(\mathbf{x}) = \frac{\partial(f_1, \cdots, f_n)}{\partial(x_1, \cdots, x_n)}.$$

One confusing part of that we sometimes refer to $\mathbf{f}'(\mathbf{x})$ as Jacobian matrix. Usually, the context is clear.

9.6 Derivatives of Higher Order

The central question in this section is when can we exchange the order of partial derivatives. More specifically, let $f : \mathbb{R}^n \to \mathbb{R}$. We would like to know under what circumstances will the following equality hold:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

To simplify things, we will restrict ourselves to two variable case. But later on we will generalize this result. We first give a result analogous to the mean value theorem in the mutivariable case.

Theorem 9.6.1. *Suppose $f : E \subseteq \mathbb{R}^2 \to \mathbb{R}$, and both $D_1 f$ and $D_{21} f$ exist at every point of E. Suppose $Q \subseteq E$ is a closed rectangle with sides parallel to the coordinate axes, having (a, b) and $(a + h, b + k)$ as opposite vertices ($h, k \neq 0$). Put*

$$\Delta(f, Q) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b).$$

Then there is a point (x, y) in the interior of Q such that

$$\Delta(f, Q) = hk\,(D_{21} f)\,(x, y).$$

Proof. Put

$$u(t) = f(t, b + k) - f(t, b).$$

Then two applications of the mean value theorem show that there exist $x \in (a, a + h)$ and $y \in (b, b + k)$ such that

$$\begin{aligned}\Delta(f, Q) &= u(a + h) - u(a)\\ &= hu'(x)\\ &= h\,[(D_1 f)\,(x, b + k) - (D_1 f)\,(x, b)]\\ &= hk\,(D_{21} f)\,(x, y).\end{aligned}$$

□

Remark 9.6.2. Notice that in the first order approximation, we have that when the rectangle has small dimension, then

$$\begin{aligned}\frac{\Delta(f,Q)}{hk} &= \frac{1}{h}\left(\frac{f(a+h,b+k)-f(a+h,b)}{k} - \frac{f(a,b+k)-f(a,b)}{k}\right)\\ &\approx \frac{(D_2 f)(a+h,b)-(D_2 f)(a,b)}{h}\\ &\approx D_{12} f.\end{aligned}$$

Hence in the following proof, we will take advantage of this observation and show that when h, k are small, then $D_{21} f$ approaches $\Delta(f,Q)/hk$. However, also notice that intuitively why we can exchange derivative: We can write this quantity above in a different way:

$$\begin{aligned}\frac{\Delta(f,Q)}{hk} &= \frac{1}{k}\left(\frac{f(a+h,b+k)-f(a,b+k)}{h} - \frac{f(a+h,b)-f(a,b)}{h}\right)\\ &\approx \frac{(D_1 f)(a,b+k)-(D_2 f)(a,b)}{k}\\ &\approx D_{21} f.\end{aligned}$$

Theorem 9.6.3. *Suppose $f : E \subseteq \mathbb{R}^2 \to \mathbb{R}$, and $D_1 f, D_{21} f, D_2 f$ exist at every point of E, and $D_{21} f$ is continous at some point $(a,b) \in E$. Then $D_{12} f$ exists at (a,b) and*

$$(D_{12} f)(a,b) = (D_{21} f)(a,b).$$

Proof. Put $A = (D_{21} f)(a,b)$. Choose $\epsilon > 0$. If Q is sufficently small then by the continuity of $D_{21} f$ at (a,b), we have

$$|A - (D_{21} f)(x,y)| < \epsilon$$

for all $(x,y) \in Q$. Then by the previous theorem, we have that

$$\left|\frac{\Delta(f,Q)}{hk} - A\right| < \epsilon.$$

Fix h and let $k \to 0$. Then the inequality above inplies

$$\left|\frac{(D_2 f)(a+h,b) - (D_2 f)(a,b)}{h} - A\right| \le \epsilon.$$

Since ϵ is arbitrary, and the inequality holds for any sufficently small $h \neq 0$, it follows from the inequality that $(D_{12} f)(a,b) = A$. This proves the claim. □

Corollary 9.6.4. *If $f \in \mathscr{C}'(E)$, then $D_{12} f = D_{21} f$.*

Definition 9.6.5. Let $f : \mathbb{R}^n \to \mathbb{R}$. By induction, $\mathscr{C}^k(E)$ can be defined as follows for every positive integer k. To say $f \in \mathscr{C}^k(E)$ means that the partial derivatives $D_1 f \cdots, D_n f \in \mathscr{C}^{k-1}(E)$.

Theorem 9.6.6. *Let $f \in \mathscr{C}^k(E)$. Then repeated application of Corollary 9.6.4 shows that the kth order derivative*

$$D_{i_1 i_2 \cdots i_k} f = D_{i_1} D_{i_2} \cdots D_{i_k} f$$

is unchanged if the subscripts $i_1, \cdots, i_k$ are permuted. For instance, if $n \geq 3$, then for every $f \in \mathscr{C}^{(4)}(E)$, we have $D_{1213} f = D_{3112} f$.

Proof. We use induction on k. When $k = 2$, i.e. $f \in \mathscr{C}^{(2)}(E)$, the claim is proven by the corollary after Theorem 9.41. Now suppose that the claim holds for all $1 \leq n \leq k-1$, we want to prove that the result holds for $n = k$ (i.e. $f \in \mathscr{C}^k(E)$). Let $A = \{i_1, \cdots, i_k\}$, then given any $\tau \in S_A$ such that $\tau(i_k) = i_k$, we know from the induction hypothesis ($D_1 f \cdots, D_n f \in \mathscr{C}^{(k-1)}(E)$) that

$$D_{i_1, \cdots, i_k} f = D_{\tau(i_1), \cdots, \tau(i_k)} f. \tag{*}$$

Hence if $\tau = (i_j, i_{j+1})$, where $1 \leq j \leq k-2$, then we know (*) holds. Also, since $f \in \mathscr{C}^k(E)$, clearly $f \in \mathscr{C}^2(E)$. Then by Thm 9.41 and the corollary we know that

$$D_{i_{k-1}, i_k} f = D_{i_k, i_{k-1}} f.$$

That is, (*) also holds if $\tau = (i_{k-1}, i_k)$. Therefore, (*) holds for all adjacent transpositions τ of the form $\tau = (i_j, i_{j+1})$, where $1 \leq j \leq k-1$.

Next, we show that if $\tau = \sigma \circ \sigma'$ is a composition of two adjacent transposition of the form above, then (*) still holds. We introduce some notation. Let $A = \{i_1, \cdots, i_k\}$ and $A' = \sigma'(A) = \{\sigma'(i_1), \cdots, \sigma'(i_k)\}$. Then from above, we know that adjacent transposition will not change the value of the kth order derivatives. Hence

$$D_{\tau(A)} f = D_{\sigma \circ \sigma'(A)} f = D_{\sigma(A')} f = D_{A'} f = D_{\sigma(A)} f = D_A f.$$

However, we know that S_A can be generated by just the sets of all adjacent transpositions [1], i.e.

$$S_A = \langle (i_j, i_{j+1}) | 1 \leq j \leq k-1 \rangle.$$

Hence if $D_{\tau(A)} f = D_A f$ for any adjacent transposition τ, and $D_{\sigma \circ \sigma'(A)} f = D_A f$ for any compositions of adjacent transpositions, we have

$$D_{\tau(A)} f = D_A f$$

for any $\tau \in S_A$. This proves the claim. □

9.7 Integration

Definition 9.7.1. Suppose I^k is a k-cell in $\mathbb{R}^k$ consisting of all $\mathbf{x} = (x_1, \cdots, x_k)$ such that $a_i \leq x_i \leq b_i$ for $i = 1, \cdots, k$. Let I^j be the j-cell in $\mathbb{R}^j$ defined by the first j inequalities. And let f be a real continous (hence uniformly continuous) function on I^k.

[1] See, for example, David Dummit and Richard Foote, *Abstract Algebra*, Section 3.5, Exercise 3

Put $f = f_k$. Define f_{k-1} on I^{k-1} by

$$f_{k-1}(x_1, \ldots, x_{k-1}) = \int_{a_k}^{b_k} f_k(x_1, \ldots, x_{k-1}, x_k)\, dx_k.$$

The uniform continuity of f_k on I^k shows that f_{k-1} is continous on I^{k-1}. Hence we can repeat this process. After k steps we arrive at a number f_0, which we call the integral of f over I^k, denoted by

$$\int_{I^k} f(\mathbf{x}) d\mathbf{x} \quad \text{or} \quad \int_{I^k} f.$$

The next theorem says that the order of integration does not matter:

Theorem 9.7.2. *For every $f \in \mathscr{C}(I^k)$, Denote the integral obtained in the definition above by $L(f)$, and denote the integration carried out by some other order $L'(f)$. Then $L(f) = L'(f)$.*

Proof. If $h(\mathbf{x}) = h_1(x_1) \cdots h_k(x_k)$ where $h_j \in \mathscr{C}([a_j, b_j])$. Then

$$L(h) = \prod_{i=1}^{k} \int_{a_i}^{b_i} h_i(x_i)\, dx_i = L'(h).$$

If $\mathscr{A}$ is the set of all finite sums of such functions h, it follows that $L(g) = L'(g)$ for all $g \in \mathscr{A}$. Also, $\mathscr{A}$ is an algebra of functions on I^k to which the Stone-Weierstrass theorem applies. Put $V = \prod_1^k (b_i - a_i)$. If $f \in \mathscr{C}(I^k)$ and $\epsilon > 0$, there exists $g \in \mathscr{A}$ such that $\|f - g\| < \epsilon/V$. Then $|L(f-g)| < \epsilon$ and $|L'(f-g)| < \epsilon$. Then since

$$L(f) - L'(f) = L(f-g) + L'(g-f),$$

we have

$$|L(f) - L'(f)| < 2\epsilon.$$

□

Definition 9.7.3. The support of a function f on $\mathbb{R}^k$, denoted by Supp f, is the closure of the set of all points $\mathbf{x} \in \mathbb{R}^k$ at which $f(\mathbf{x}) \neq 0$. If f is a continous function with compact support, let I^k be any k-cell which contains Supp f, and define

$$\int_{\mathbb{R}^k} f = \int_{I^k} f.$$

Now, we would like to do some preparation work for proving the change of variable formula on multiple integrals. To do this, we need two major results in advance: primitive mapping decomposition theorem, and partition of unity.

Definition 9.7.4 (Primitive Mappings)**.** Let $\mathbf{G} : E \subseteq \mathbb{R}^n \to \mathbb{R}^n$, and if there is an integer m and a real function g with domain E such that

$$\mathbf{G}(\mathbf{x}) = \sum_{i \neq m} x_i \mathbf{e}_i + g(\mathbf{x})\mathbf{e}_m = (x_1, \cdots, x_{m-1}, g(\mathbf{x}), x_{m+1}, \cdots, x_n) \quad (\mathbf{x} \in E)$$

then we call $\mathbf{G}$ primitive. A primitive mapping is thus one that changes at most one coordinate. Note that the equation above is equivalent to

$$\mathbf{G}(\mathbf{x}) = \mathbf{x} + [g(\mathbf{x}) - x_m]\, \mathbf{e}_m.$$

Remark 9.7.5. Notice that if g is differentiable at some point $\mathbf{a} \in E$, then so is $\mathbf{G}$. And we clearly have

$$J_{\mathbf{G}}(\mathbf{a}) = \det\big[\mathbf{G}'(\mathbf{a})\big] = \det \begin{bmatrix} 1 & 0 & 0 & \cdots & \frac{\partial g}{\partial x_1} & 0 & \cdots & 0 \\ & 1 & & & & & & \\ & & & \ddots & & & & \\ & & & & \frac{\partial g}{\partial x_m} & & & \\ & & & & & \ddots & & \\ & & & & & & & 1 \end{bmatrix} = \frac{\partial g}{\partial x_m}(\mathbf{a}).$$

Definition 9.7.6 (Flip)**.** A linear operator B on $\mathbb{R}^n$ that interchanges some pair of coordinate and leaves the others fixed are called a flip. For example, an filp on $\mathbb{R}^4$ can be

$$B\left(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 + x_4\mathbf{e}_4\right) = x_1\mathbf{e}_1 + x_2\mathbf{e}_4 + x_3\mathbf{e}_3 + x_4\mathbf{e}_2.$$

Or equivalently,

$$B(x_1, x_2, x_3, x_4) = (x_1, x_4, x_3, x_2).$$

Definition 9.7.7 (Projection)**.** Let P_j be operators on $\mathbb{R}^n$, where $j = 0, 1, \cdots, n$, defined by the following: $P_0\mathbf{x} = 0$ and

$$P_m\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_m\mathbf{e}_m = (x_1, \cdots, x_m).$$

Then it is easy to see that $\ker P_m = \operatorname{span}(\mathbf{e}_{m+1}, \cdots, \mathbf{e}_n)$ and range $P_m = \operatorname{span}(\mathbf{e}_1, \cdots, \mathbf{e}_m)$. Also, it is clear that $P_j^2 = P_j$ for all j.

Now we present the first preparatory result:

Theorem 9.7.8 (Primitive Mapping Decomposition)**.** *Let* $\mathbf{F}$ *be a* $\mathscr{C}'$ *mapping of an open set* $E \subseteq \mathbb{R}^n$ *into* $\mathbb{R}^n$*, with* $\mathbf{0} \in E, \mathbf{F}(\mathbf{0}) = \mathbf{0}$*, and* $\mathbf{F}'(\mathbf{0})$ *is invertible. Then there is a neighborhood of* $\mathbf{0}$ *in* $\mathbb{R}^n$ *in which a representation*

$$\mathbf{F}(\mathbf{x}) = B_1 \cdots B_{n-1}\mathbf{G}_n \circ \cdots \circ \mathbf{G}_1(\mathbf{x})$$

is valid. Each $\mathbf{G}_i$ *is a primitive* $\mathscr{C}'$ *mapping in some neighborhood of* $\mathbf{0}$ *with* $\mathbf{G}_i(\mathbf{0}) = 0$ *and* $\mathbf{G}_i'(\mathbf{0})$ *invertible. Each* B_i *is either a flip or the identity.*

Proof. We prove by induction. Put $\mathbf{F} = \mathbf{F}_1$. Assume $1 \leq m \leq n-1$. We make the following induction hypothesis:

- V_m is a neighborhood of $\mathbf{0}$;
- $\mathbf{F}_m \in \mathscr{C}'(V_m)$;
- $\mathbf{F}_m(\mathbf{0}) = \mathbf{0}$;
- $\mathbf{F}_m'(\mathbf{0})$ is invertible;

- $P_{m-1}\mathbf{F}_m(\mathbf{x}) = P_{m-1}x$ for $\mathbf{x} \in V_m$.

These hypothesis obviously holds for the base case $m = 1$. Since $P_{m-1}\mathbf{F}_m(\mathbf{x}) = P_{m-1}x$ (meaning $\mathbf{F}_m(\mathbf{x})$ has the same first $m-1$ coordinates as $\mathbf{x}$), there must exists $\mathscr{C}'$ functions $\alpha_m, \cdots, \alpha_k$ in V_m such that

$$\mathbf{F}_m(\mathbf{x}) = P_{m-1}\mathbf{x} + \sum_{t=m}^{n} \alpha_i(\mathbf{x})\mathbf{e}_i = (x_1, \cdots, x_m - 1, \alpha_m(\mathbf{x}), \cdots, \alpha_n(\mathbf{x})).$$

Then taking the derivative we have

$$\mathbf{F}'_m(\mathbf{x}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \\ 0 & \ddots & 0 & 0 & 0 & \\ & & 1 & & & \\ D_1\alpha_m & \cdots & D_{m-1}\alpha_m & D_m\alpha_m & \cdots & D_n\alpha_m \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ D_1\alpha_n & \cdots & D_{m-1}\alpha_n & D_m\alpha_n & \cdots & D_n\alpha_n \end{bmatrix}$$

Hence we have

$$\mathbf{F}'_m(\mathbf{0})\mathbf{e}_m = \sum_{i=m}^{n} (D_m x_i)(\mathbf{0})\mathbf{e}_i.$$

Since $\mathbf{F}'(\mathbf{0})$ is invertible, the LHS is nonzero, neither is the RHS. Hence there exists a k such that $m \le k \le n$ and $(D_m\alpha_k)(\mathbf{0}) \ne 0$. Let B_m be the filip that interchanges m and k. Define

$$\mathbf{G}_m(\mathbf{x}) = \mathbf{x} + [\alpha_k(\mathbf{x}) - x_m]\,\mathbf{e}_m \quad (\mathbf{x} \in V_m).$$

Hence $\mathbf{G}_m \in \mathscr{C}'(V_m)$, $\mathbf{G}_m$ is primitive, and $\mathbf{G}'_m(\mathbf{0})$ is invertible (Since $(D_m\alpha_k)(\mathbf{0}) \ne 0$). Thus by the inverse function theorem there is an open set U_m with $\mathbf{0} \in U_m \subseteq V_m$, such that $\mathbf{G}_m$ is invertible. Moreover, $\mathbf{G}_m^{-1}$ is continuously differentiable. Hence define

$$\mathbf{F}_{m+1}(\mathbf{y}) = B_m\mathbf{F}_m \circ \mathbf{G}_m^{-1}(\mathbf{y}) \quad (\mathbf{y} \in V_{m+1}). \tag{*}$$

Then we can check that the induction hypothesis (except the last one) still holds for $\mathbf{F}_{m+1}$. Also, for $\mathbf{x} \in U_m$ we have that

$$\begin{aligned} P_m\mathbf{F}_{m+1}(\mathbf{G}_m(\mathbf{x})) &= P_m B_m \mathbf{F}_m(\mathbf{x}) \\ &= P_m\left[P_{m-1}\mathbf{x} + \alpha_k(\mathbf{x})\mathbf{e}_m + \cdots\right] \\ &= P_{m-1}\mathbf{x} + \alpha_k(\mathbf{x})\mathbf{e}_m \\ &= P_m\mathbf{G}_m(\mathbf{x}). \end{aligned}$$

This follows from the definition of $\mathbf{F}_m$. Then we have

$$P_m\mathbf{F}_{m+1}(\mathbf{y}) = P_m\mathbf{y} \quad (\mathbf{y} \in V_{m+1}).$$

Thus all induction hypothesis holds. Notice that $B_m B_m = I$. Hence by squeezing (*) between $B_m^{-1} \circ (\,) \circ G_m$ we find that

$$\mathbf{F}_m(\mathbf{x}) = B_m\mathbf{F}_{m+1}(\mathbf{G}_m(\mathbf{x})) \quad (\mathbf{x} \in U_m).$$

Hence we have

$$\begin{aligned}\mathbf{F}=\mathbf{F}_1 &= B_1\mathbf{F}_2\circ\mathbf{G}_1\\ &= B_1B_2\mathbf{F}_3\circ\mathbf{G}_2\circ\mathbf{G}_1=\cdots\\ &= B_1\cdots B_{n-1}\mathbf{F}_n\circ\mathbf{G}_{n-1}\circ\cdots\circ\mathbf{G}_1\end{aligned}$$

in some neighborhood of $\mathbf{0}$. Since by the definition of $F_n=(x_1,\cdots,x_{n-1},\alpha_n(\mathbf{x}))$, we have $\mathbf{F}_n$ primitive. This proves the claim. $\square$

Remark 9.7.9. The procedure of the proof can be illustrated by the following diagram:

$$\begin{tikzcd}
\mathbf{x}=(x_1,\cdots,x_n) \arrow[r, shift left, "\mathbf{G}_1"] \arrow[rr, bend left, "\mathbf{F}=\mathbf{F}_1"] & (x_1,\cdots,\alpha_k,\cdots,x_n) \arrow[l, shift left, "\mathbf{G}_1^{-1}"] \arrow[r, "B_1\circ\mathbf{F}_2"] & \mathbf{F}(\mathbf{x})=(\alpha_1,\cdots,\alpha_n)
\end{tikzcd}$$

We first start off with a primitive mapping $\mathbf{G}_1$, such that it is invertible in a small neighborhood of zero. Then inside this neighborhood, as the diagram suggests, we have that

$$B_2\circ\mathbf{F}_2=\mathbf{F}_1\circ\mathbf{G}_1^{-1}.$$

Hence this gives us a new map

$$\mathbf{F}_2=B_1\circ\mathbf{F}_1\circ G_1^{-1}.$$

Then we apply the same procedure to $\mathbf{F}_2$, and decompose it into $B_2\circ\mathbf{F}_3$. As we work our way up to $\mathbf{F}(\mathbf{x})$, we are getting a series of primitive maps behind us, and a series of flips composed with $\mathbf{F}_m$ to work with. Then at some point $\mathbf{F}_m$ will become primitive, and we are done.

Our next big theorem is partition of unity.

Theorem 9.7.10 (Partition of Unity). *Suppose K is a compact subset of $\mathbb{R}^n$ and $\{V_\alpha\}$ is an open cover of K. Then there exist functions $\psi_1,\cdots,\psi_s\in\mathscr{C}(\mathbb{R}^n)$ such that*

(a) $0\le\psi_i\le 1$ *for* $1\le i\le s$*;*

(b) Each ψ_i has its support in some V_α;

(c) $\psi_1(\mathbf{x})+\cdots+\psi_s(\mathbf{x})=1$ *for every* $\mathbf{x}\in K$.

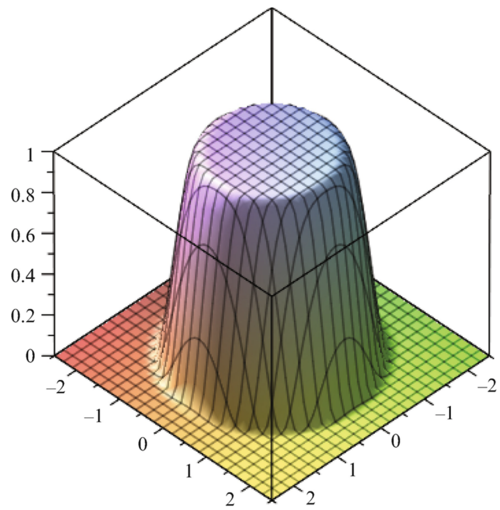
Proof. associate with each $\mathbf{x}\in K$ an index $\alpha(\mathbf{x})$ so that $x\in V_{\alpha(\mathbf{x})}$. Then there are open balls $B(\mathbf{x})$ and $W(\mathbf{x})$, centered at $\mathbf{x}$, with

$$\overline{B(\mathbf{x})}\subset W(\mathbf{x})\subset\overline{W(\mathbf{x})}\subset V_{\alpha(\mathbf{x})}.$$

Since K is compact, there are points $\mathbf{x}_1,\cdots,\mathbf{x}_s$ in K such that

$$K\subset B(\mathbf{x}_1)\cup\cdots\cup B(\mathbf{x}_s).$$

Then there $\varphi_1,\cdots,\varphi_s\in\mathscr{C}(\mathbb{R}^n)$ such that $\varphi_i(\mathbf{x})=1$ on $B(\mathbf{x}_i)$, and $\varphi(\mathbf{x})=0$ outside $W(\mathbf{x}_i)$, and $0\le\varphi_i(\mathbf{x})\le 1$ on $\mathbb{R}^n$.



This is always possible, as the figure above suggests. Then define $\psi_1 = \varphi_1$ and

$$\psi_{i+1} = (1-\varphi_1)\cdots(1-\varphi_i)\,\varphi_{i+1}.$$

We verify that this is our desired set of functions.

(a) Clear by definition.

(b) We have $\mathrm{Supp}(\psi_{i+1}) \subseteq \mathrm{Supp}(\varphi_{i+1}) \subseteq W(\mathbf{x}_{i+1}) \subseteq V_{\alpha(\mathbf{x}_{i+1})}$.

(c) We observe the following:

$$\begin{aligned}
&1-\psi_1 = 1-\varphi_1\\
&1-\psi_1-\psi_2 = (1-\varphi_1)-(1-\varphi_1)\varphi_1 = (1-\varphi_1)(1-\varphi_2)\\
&\vdots\\
&1-\sum_{i=1}^{s}\psi_i(\mathbf{x}) = \prod_{i=1}^{s}(1-\varphi_i(\mathbf{x})).
\end{aligned}$$

This is true for all $\mathbf{x} \in \mathbb{R}^n$. If $\mathbf{x} \in K$, then $\mathbf{x} \in B(\mathbf{x}_i)$ for some i. Hence $\varphi_i(\mathbf{x}) = 1$, and we have

$$1-\sum_{i=1}^{s}\psi_i(\mathbf{x}) = 0.$$

This proves the theorem. □

Now we are ready to prove the change of variable theorem on multiple integrals.

Theorem 9.7.11 (Change of Variable)**.** *Suppose T is a 1-1 $\mathscr{C}'$ mapping of an open set $E \subseteq \mathbb{R}^k$ into $\mathbb{R}^k$ such that $J_T(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in E$. If f is a continuous function on $\mathbb{R}^k$ whose support in compact and lies in $T(E)$, then*

$$\int_{\mathbb{R}^k} f(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^k} f(T(\mathbf{x}))\,|J_T(\mathbf{x})|\,d\mathbf{x}.$$

Remark 9.7.12. We first show that this theorem is true for $k = 1$. Then in this case we have that

$$\int_{\mathbb{R}} f(y)dy = \int_{\mathbb{R}} f(T(x))|T'(x)|dx.$$

Suppose that $T : [x_1, x_2] \to \mathbb{R}$ and $\operatorname{Supp} f \subseteq T([x_1, x_2])$. Then if $T'(x) > 0$, we have $T(x_1) < T(x_2)$. Hence

$$\int_{\mathbb{R}} f(y)dy = \int_{T(x_1)}^{T(x_2)} f(y)dy = \int_{x_1}^{x_2} f(T(x))T'(x)dx = \int_{\mathbb{R}} f(T(x))|T'(x)|dx.$$

If $T'(x) < 0$, then

$$\int_{\mathbb{R}} f(y)dy = \int_{T(x_2)}^{T(x_1)} f(y)dy = -\int_{x_1}^{x_2} f(T(x))T'(x)dx = \int_{\mathbb{R}} f(T(x))|T'(x)|dx.$$

Proof. We divide the proof into several steps.

1. We first prove that the theorem holds for primitive mappings, flips, and translations. Notice that if T is a primitive $\mathscr{C}'$ mapping, then by Remark 9.7.5, we have that

$$\begin{aligned}
\int_{\mathbb{R}^k} f(\mathbf{y})d\mathbf{y} &= \int_{\mathbb{R}^{k-1}} \left(\int_{\mathbb{R}} f(y_1, \cdots, y_m, \cdots, y_k)dy_m \right) dy_1 \cdots dy_k \\
&= \int_{\mathbb{R}^{k-1}} \left(\int_{\mathbb{R}} f(x_1, \cdots, g(\mathbf{x}), \cdots, x_k) \left| \frac{\partial g}{\partial x_m} \right| dx_m \right) dx_1 \cdots dx_k \\
&= \int_{\mathbb{R}^k} f(T(\mathbf{x}))|J_T(\mathbf{x})|d\mathbf{x}.
\end{aligned}$$

by our previous remark on the one dimensional case and the result that order of integration does not matter. The theorem certainly holds for flips, since $|J_B(\mathbf{x})| = 1$ for all $\mathbf{x}$ since B is just a matrix that switches row and $B' = B$. Then WLOG assume our function has two variables.

$$\begin{aligned}
\int_{\mathbb{R}^2} f(T(\mathbf{x}))|J_T(\mathbf{x})|d\mathbf{x} &= \int_E f(T(\mathbf{x}))|J_T(\mathbf{x})|d\mathbf{x} \\
&= \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_2, x_1)dx_1dx_2 \\
&= \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(y_1, y_2)dy_1dy_2 \\
&= \int_{T(E)} f(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^2} f(\mathbf{y})d\mathbf{y}.
\end{aligned}$$

Finally, the theorem holds for translation $T_\epsilon : \mathbf{x} \mapsto \mathbf{x} + \epsilon$ where $\epsilon = (\epsilon_1, \cdots, \epsilon_k)$. Then since $|J_{T_\epsilon}(\mathbf{x})| = 1$ we have that

$$\begin{aligned}
\int_{\mathbb{R}^k} f(T_\epsilon(\mathbf{x}))|J_{T_\epsilon}(\mathbf{x})|d\mathbf{x} &= \int_{a_k}^{b_k}\cdots\int_{a_1}^{b_1} f(x_1+\epsilon_1,\cdots,x_k+\epsilon_k)dx_1\cdots dx_k \\
&= \int_{a_k}^{b_k}\cdots\left(\int_{a_1+\epsilon_1}^{b_1+\epsilon_1} f(y_1,\cdots,x_k+\epsilon_k)dy_1\right)\cdots dx_k \\
&= \int_{a_k+\epsilon_k}^{b_k+\epsilon_k}\cdots\int_{a_1+\epsilon_1}^{b_1+\epsilon_1} f(y_1,\cdots,y_k)dy_1\cdots dy_k \\
&= \int_{T(E)} f(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^k} f(\mathbf{y})d\mathbf{y}.
\end{aligned}$$

2. Next, if the theorem is true for transformations P, Q, then the theorem is true for $S(\mathbf{x}) = P(Q(\mathbf{x}))$. Since

$$\begin{aligned}
\int f(\mathbf{z})d\mathbf{z} &= \int f(P(\mathbf{y}))\,|J_P(\mathbf{y})|\,d\mathbf{y} \\
&= \int f(P(Q(\mathbf{x})))\,|J_P(Q(\mathbf{x}))|\,\left|J_Q(\mathbf{x})\right|\,d\mathbf{x} \\
&= \int f(S(\mathbf{x}))\,|J_S(\mathbf{x})|\,d\mathbf{x}
\end{aligned}$$

where we have used the relation

$$J_P(Q(\mathbf{x}))J_Q(\mathbf{x}) = \det P'(Q(\mathbf{x}))\det Q'(\mathbf{x}) = \det P'(Q(\mathbf{x}))Q'(\mathbf{x}) = \det S'(\mathbf{x}) = J_s(\mathbf{x}).$$

3. Let us define the map U on the set $E - \mathbf{a} = \{\mathbf{x}-\mathbf{a} |\ \mathbf{x} \in E\}$, given by

$$U(\mathbf{y}) = T(\mathbf{a}+\mathbf{y}) - T(\mathbf{a}).$$

Then at every point $\mathbf{y} \in E - \mathbf{a}$, we have a neighborhood of $\mathbf{y}$ such that in that neighborhood U can be decomposed into a bunch of primitive mappings and filps:

$$U(\mathbf{y}) = B_1\cdots B_{k-1}G_k \circ\cdots\circ G_1(\mathbf{y}).$$

Then

$$\begin{aligned}
T(\mathbf{x}) &= T(\mathbf{a}) + (T(\mathbf{a}+(\mathbf{x}-\mathbf{a})) - T(\mathbf{a})) = T(\mathbf{a}) + U(\mathbf{x}-\mathbf{a}) \\
&= T(\mathbf{a}) + B_1\cdots B_{k-1}\mathbf{G}_k\circ\mathbf{G}_{k-1}\circ\cdots\circ\mathbf{G}_1(\mathbf{x}-\mathbf{a}).
\end{aligned}$$

Setting $V = T(U)$, it follows that the theorem holds if the support of f lies in V. Thus *each point* $\mathbf{y} \in T(E)$ *lies in an open set* $V_\mathbf{y} \subseteq T(E)$ *such that the theorem holds for all continuous functions whose support lies in* $V_\mathbf{y}$. Now let f be a continous function with compact support $K \subseteq T(E)$. Since $\{V_\mathbf{y}\}$ covers K, partition of unity says there exists a family of functions ψ_i such that $f = \sum \psi_i f$, where each ψ_i is continuous and has its support in some $V_\mathbf{y}$. Thus the theorem holds for each $\psi_i f$, hence also for their sum f.

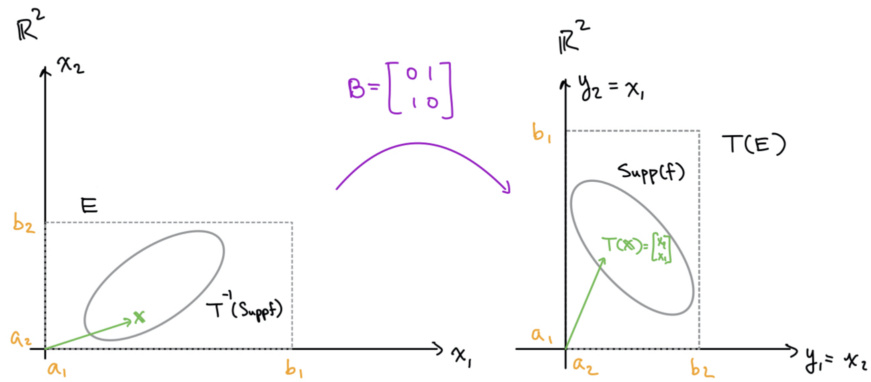
We have proved the theorem. □

Remark 9.7.13. The reader may already noticed that we used the modified notation

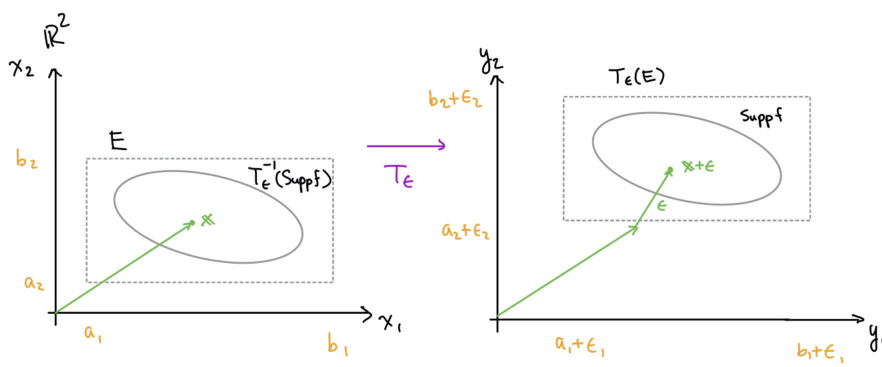
$$\int_{T(E)} f(\mathbf{y}) d\mathbf{y} = \int_E f(T(\mathbf{x})) \left| J_T(\mathbf{x}) \right| d\mathbf{x}$$

in the proof of the theorem. Turns out that we can define integral on a more general domain than just k-cells.

The figure below illustrates why in the step 1 of the proof, the change of variable fourmula holds for flips and translation. Here we take $k = 2$.



$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(T(\mathbf{x})) dx_1 dx_2 = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(\mathbf{y}) dy_1 dy_2.$$



$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(T(\mathbf{x})) dx_1 dx_2 = \int_{a_2+\epsilon_2}^{b_2+\epsilon_2} \int_{a_1+\epsilon_1}^{b_1+\epsilon_1} f(\mathbf{y}) dy_1 dy_2.$$

Notice that in this case, both B and T_ϵ are invertible, hence they are injective on any open sets of $\mathbb{R}^2$. And we choose the open set E such that it is a rectangle and Supp $f \subseteq T(E)$.

Chapter 10

Integration of Differential Forms

We can only see a short distance ahead, but we can see plenty there that needs to be done.

-Alan Turing

For those who previously studied calculus before, it is hard to miss the subtle connections in all of the key fourmulas, as they share great similarity.

- Fundamental Theorem of Calculus:

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

  We know if the set we are integrating on are discrete (i.e. sets of points), then integration degenerates into a discrete sum. Hence on the LHS we are sort of integrating a function f over zero dimension, while on the RHS we are integrating its derivatives over a one dimensional line segment $[a, b]$.

- Fundamental Theorem of Line integrals: Similar to the previous case, but we are integrating the gradient of $f : \mathbb{R}^n \to \mathbb{R}$ over a curve C parametrized by $r : [a, b] \to \mathbb{R}^n$. Then we have

$$f(\mathbf{r}(b)) - f(\mathbf{r}(a)) = \int_C \nabla f \cdot d\mathbf{r}.$$

  We can view ∇f as some sort of derivative.

- Green's Theorem: Here let D be a region with boundary C. Then

$$\int_C P dx + Q dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA.$$

  On the LHS we are inetgrating something over a path C, which is 1 dimensional, and on the RHS we are inetgrating some sort of derivative related to that something, over a 2 dimensional region.

- Divergence Theorem: Here $\mathbf{F}$ is a vector field, E is a volume with boundary S, then

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iiint_E \nabla \cdot \mathbf{F} dV.$$

  On the LHS we are integrating a function $\mathbf{F}$ over a 2 dimensional region S, i.e. a surface integral, while on th RHS we are integrating its derivative $\nabla \cdot \mathbf{F}$ over a 3 dimensional volume.

- Stokes' Theorem: This is a generalization of Green's theorem: Let S be a surface with boundary C. Then

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_S \nabla \times \mathbf{F} \cdot d\mathbf{S}.$$

All these key results have the following similar forms:

$$\int_{k-\mathrm{dim}} (\text{blah}) = \int_{k+1-\mathrm{dim}} (\text{some sort of derivative related to blah}).$$

This is the beginning of a journey – a grand unifield theory of integration in Euclidean space. We will start by defining that "something", which are something called differential forms, and we will study how to integrate differential form. Then, finally, we will establish our guess above and furnish it into the elegant form in the following

$$\int_{\partial\Omega} \omega = \int_\Omega d\omega.$$

This is the generalized stokes' theorem, and all the formulas above will be special cases of this theorem.

10.1 Differential Forms

Definition 10.1.1. Let $\mathbf{v} = (v_1, \cdots, v_n) \in \mathbb{R}^n$. We define n linear maps $dx_i : \mathbb{R}^n \to \mathbb{R}$, where $i = 1, \cdots, n$, given by

$$dx_i(\mathbf{v}) = v_i.$$

Proposition 10.1.2. $dx_1, \cdots, dx_n$ form a basis of the dual vector space $(\mathbb{R}^n)^*$ of $\mathbb{R}^n$.

Proof. Let $\phi : \mathbb{R}^n \to \mathbb{R}$. Then let $a_i = \phi(\mathbf{e}_i)$. Then clearly $\phi = a_1 dx_1 + \cdots + a_n dx_n$. Hence this is a spanning set. Now, if $\phi = c_1 dx_1 + \cdots + c_n dx_n = 0$ is the zero map, then $\phi(\mathbf{e}_1) = c_1 = 0$. And similarly $c_i = 0$ for all i. Hence this is a linearly independent spanning set, i.e. a basis. $\square$

Definition 10.1.3. Let $I = (i_1, \cdots, i_k)$ be an ordered k tuple. Define

$$d\mathbf{x}_I : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}$$

by the following:

$$d\mathbf{x}_I(\mathbf{v}_1,\dots,\mathbf{v}_k) = \begin{vmatrix} dx_{i_1}(\mathbf{v}_1) & \cdots & dx_{i_1}(\mathbf{v}_k) \\ \vdots & \ddots & \vdots \\ dx_{i_k}(\mathbf{v}_1) & \cdots & dx_{i_k}(\mathbf{v}_k) \end{vmatrix}.$$

If $i_1 < \cdots < i_k$, we say I is increasing. It is clear from the definition above that if I has repeated elements, then $d\mathbf{x}_I = 0$.

Proposition 10.1.4. Let $\Lambda^k(\mathbb{R}^n)^*$ denote the vector space of alternating multilinear functions from $(\mathbb{R}^n)^k$ to $\mathbb{R}$. Then the set of $d\mathbf{x}_I$ with increasing ordered k-tuple I forns a basis of this vector space. In particular, $\dim \Lambda^k(\mathbb{R}^n)^* = \binom{n}{k}$. Any $T \in \Lambda^k(\mathbb{R}^n)^*$ can be written as the following linear combination:

$$T = \sum_{I \text{ increasing}} a_I d\mathbf{x}_I$$

where $a_I \in \mathbb{R}$.

Definition 10.1.5. Let $I = (i_1, \cdots, i_k)$ be ordered k-tuples and $J = (j_1, \cdots, j_l)$ be ordered l-tuples. Let $\omega = \sum a_I d\mathbf{x}_I \in \Lambda^k(\mathbb{R}^n)^*$ and $\eta = \sum b_J d\mathbf{x}_J$. we define the so called "wedge product":

$$\wedge : \Lambda^k(\mathbb{R}^n)^* \times \Lambda^l(\mathbb{R}^n)^* \to \Lambda^{k+l}(\mathbb{R}^n)^*$$

by

$$\omega \wedge \eta = \sum (a_I b_J)\, d\mathbf{x}_I \wedge d\mathbf{x}_J = \sum (a_I b_J)\, dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdot \wedge dx_{j_l}.$$

Proposition 10.1.6. The wedge product has the following properties:

1. bilinear: $(\omega + \phi) \wedge \eta = \omega \wedge \eta + \phi \wedge \eta$ and $(c\omega) \wedge \eta = c(\omega \wedge \eta)$.
2. skew-commutative: $\omega \wedge \eta = (-1)^{kl} \eta \wedge \omega$ for $\omega \in \Lambda^k(\mathbb{R}^n)^*$ and $\eta \in \Lambda^l(\mathbb{R}^n)^*$.
3. associative: $(\omega \wedge \eta) \wedge \phi = \omega \wedge (\eta \wedge \phi)$.

Definition 10.1.7. A differential k form on $\mathbb{R}^n$ is an expression

$$\omega = \sum_{k \text{ -tuples } I} f_I(\mathbf{x}) d\mathbf{x}_I = \sum_{I=(i_1,\cdots,i_k)} f_I dx_{i_1} \wedge \cdots \wedge dx_{i_k}$$

for some smooth functions f_I. A 0 form is a smooth function by definition. Notice that at every point $\mathbf{x} \in \mathbb{R}^n$, the differential form varies. Hence we can think of a differential k form on $\mathbb{R}^n$ as map

$$\omega : \mathbb{R}^n \times (\mathbb{R}^n)^k \to \mathbb{R}.$$

So it eats a point $\mathbf{x} \in \mathbb{R}^n$ and the dual vectors $dx_{i_1} \cdots dx_{i_k}$ eats k vectors in $\mathbb{R}^n$ and splits out a number. Notice that we can always reorder each term by the rules in the Proposition above and make every term an increasing $d\mathbf{x}_I$s, hence we may as well define

$$\omega = \sum_{\text{increasing } k \text{ -tuples } I} f_I(\mathbf{x}) d\mathbf{x}_I = \sum_{i_1 < \cdots < i_k} f_I dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

The set of k-forms on $\mathbb{R}^n$ is naturally a vector space, denoted by $A^k(\mathbb{R}^n)$. Actually, this is a module over the ring of smooth functions.

Remark 10.1.8. Hence we have the more notationally intuitive notion:

$$d\mathbf{x}_I(\mathbf{v}_1,\ldots,\mathbf{v}_k)=(dx_{i_1}\wedge\cdots\wedge dx_{i_k})(\mathbf{v}_1,\ldots,\mathbf{v}_k)=\begin{vmatrix} dx_{i_1}(\mathbf{v}_1) & \cdots & dx_{i_1}(\mathbf{v}_k)\\ \vdots & \ddots & \vdots\\ dx_{i_k}(\mathbf{v}_1) & \cdots & dx_{i_k}(\mathbf{v}_k)\end{vmatrix}.$$

Then immediately by the definition of $d\mathbf{x}_I$ and th Proposition above, we have

$$dx_i\wedge dx_i=0 \qquad dx_i\wedge dx_j=-dx_j\wedge dx_i.$$

And since k forms are build up by 1 forms, knowing these two relation will be enough for us to reorder k forms $fd\mathbf{x}_I$, hence also general k form $\omega=\sum_I f_I d\mathbf{x}_I$.

Now we study how to differentiate a k form.

Definition 10.1.9. The differentiation map $d : A^k(\mathbb{R}^n)\to A^{k+1}(\mathbb{R}^n)$ is defined recursively by the following. Let $f\in A^0(\mathbb{R}^n)$ be a 0 form (a smooth function $f(x_1,\cdots,x_n)$), then d maps it to a one form by

$$df=\sum_{j=1}^n\frac{\partial f}{\partial x_j}dx_j.$$

If $\omega=\sum_I f_I(\mathbf{x})d\mathbf{x}_I$ is a k-form, then we define

$$d\omega=\sum_I df_I\wedge d\mathbf{x}_I=\sum_I\sum_{j=1}^n\frac{\partial f_I}{\partial x_j}dx_j\wedge dx_{i_1}\wedge\cdots\wedge dx_{i_k}.$$

Example 10.1.10. Let $\omega\in A^1(\mathbb{R}^2\setminus\{\mathbf{0}\})$ be a 1 form defined by

$$\omega=\frac{-ydx+xdy}{x^2+y^2}.$$

Then

$$\begin{aligned} d\omega&=d\left(-\frac{y}{x^2+y^2}\right)\wedge dx+d\left(\frac{x}{x^2+y^2}\right)\wedge dy\\ &=-\frac{\partial}{\partial y}\left(\frac{y}{x^2+y^2}\right)dy\wedge dx+\frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right)dx\wedge dy\\ &=\frac{\left(x^2+y^2\right)-2y^2}{\left(x^2+y^2\right)^2}dx\wedge dy+\frac{\left(x^2+y^2\right)-2x^2}{\left(x^2+y^2\right)^2}dx\wedge dy=0.\end{aligned}$$

Proposition 10.1.11. Let $U\subseteq\mathbb{R}^n$ be an open subset, let $\omega\in A^k(U)$ and $\eta\in A^l(U)$. Let f be a smooth function. Then

1. When $k=l$ we have $d(\omega+\eta)=d\omega+d\eta$.
2. $d(f\omega)=df\wedge\omega+fd\omega$.
3. $d(\omega\wedge\eta)=d\omega\wedge\eta+(-1)^k\omega\wedge d\eta$.

4. $d(d\omega) = 0$. Notationally $d^2 = 0$.

Proof. (1),(2) are left as exercise. For (3), since d commutes with summation, it suffices to consider the case $\omega = f d\mathbf{x}_I$ and $\eta = g d\mathbf{x}_J$. Then

$$\begin{aligned} d(\omega \wedge \eta) &= d\left(fgd\mathbf{x}_I \wedge d\mathbf{x}_J\right) = d(fg) \wedge d\mathbf{x}_I \wedge d\mathbf{x}_J \\ &= (gdf + fdg) \wedge d\mathbf{x}_1 \wedge d\mathbf{x}_J = gdf \wedge d\mathbf{x}_I \wedge d\mathbf{x}_J + fdg \wedge d\mathbf{x}_I \wedge d\mathbf{x}_J \\ &= (df \wedge d\mathbf{x}_I) \wedge (gd\mathbf{x}_J) + (-1)^k (fd\mathbf{x}_I) \wedge (dg \wedge d\mathbf{x}_J) \\ &= d\omega \wedge \eta + (-1)^k \omega \wedge d\eta. \end{aligned}$$

To prove (4), again suppose $\omega = f d\mathbf{x}_I$, we have

$$d\omega = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} dx_j \wedge d\mathbf{x}_I$$

and

$$d(d\omega) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} dx_i \wedge dx_j \wedge d\mathbf{x}_I = \sum_{i<j} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial^2 f}{\partial x_j \partial x_i} \right) dx_i \wedge dx_j \wedge d\mathbf{x}_I = 0$$

since $D_{ij} f = D_{ji} f$. $\square$

10.2 Pullback

Definition 10.2.1. Let $U \subseteq \mathbb{R}^m$ be open and $\mathbf{g} : U \to \mathbb{R}^n$ be smooth. If $\omega \in A^k(\mathbb{R}^n)$, we define $\mathbf{g}^*\omega \in A^k(U)$, the pullback of ω by $\mathbf{g}$, as follows: The pullback of the function is composition:

$$\mathbf{g}^* f = f \circ g.$$

And to pull back a basis 1 form, if $\mathbf{g}(\mathbf{u}) = \mathbf{x}$, then set

$$\mathbf{g}^* dx_i = dg_i = \sum_{j=1}^{m} \frac{\partial g_i}{\partial u_j} du_j.$$

And finally, let the pullback commutes with wedge products:

$$\mathbf{g}^*\left(dx_{i_1} \wedge \cdots \wedge dx_{i_k}\right) = dg_{i_1} \wedge \cdots \wedge dg_{i_k}, \quad \text{which we can abbreviate as } d\mathbf{g}_I.$$

Lastly, the pullback of a sum is the sum of the pullbacks:

$$\mathbf{g}^*\left(\sum_I f_I d\mathbf{x}_I\right) = \sum_I (f_I \circ \mathbf{g})\, d\mathbf{g}_I = \sum_I (f_I \circ \mathbf{g})\, dg_{i_1} \wedge \cdots \wedge dg_{i_k}.$$

In other words, by pulling back we have changed the domain of the differential form. Before, $\omega \in A^k(\mathbb{R}^m)$ eats vectors in $\mathbb{R}^n$, and now it eats vectors in $\mathbb{R}^m$.

$$\begin{tikzcd} U \subseteq \mathbb{R}^m \arrow[r, "\mathbf{g}"] \arrow[dr, "\mathbf{g}^*\omega=\sum_I f_I(\mathbf{u})d\mathbf{g}_I"'] & \mathbb{R}^n \arrow[d, "\omega(\mathbf{x})=\sum_I f_I(\mathbf{x})d\mathbf{x}_I"] \\ & A^k(\mathbb{R}^k) \end{tikzcd}$$

Example 10.2.2. If $\omega = xdx + ydy$ and $\mathbf{g}(u,v) = (u\cos v, u\sin v)$, which is the transition map from cartesian coordinate to polar coordinate. Then

$$\begin{aligned} \mathbf{g}^*\omega &= (u\cos v)(\cos v du - u\sin v dv) + (u\sin v)(\sin v du + u\cos v dv) \\ &= u\left(\cos^2 v + \sin^2 v\right) du + u^2(-\cos v\sin v + \cos v\sin v)dv = udu. \end{aligned}$$

And

$$\begin{aligned} \mathbf{g}^*(dx\wedge dy) = \mathbf{g}^*dx \wedge \mathbf{g}^*dy &= (\cos v du - u\sin v dv)\wedge(\sin v du + u\cos v dv) \\ &= u\left(\cos^2 v + \sin^2 v\right) du\wedge dv = udu\wedge dv. \end{aligned}$$

Therefore, we have

$$\mathbf{g}^*\left(e^{-(x^2+y^2)}dx\wedge dy\right) = ue^{-u^2}du\wedge dv.$$

Notice that this is the change of variable that we used when we evaluate $\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}dxdy$.

Example 10.2.3. Let $\mathbf{g}: \mathbb{R}^2 \to \mathbb{R}^3$ be given by $\mathbf{g}(u,v) = (u\cos v, u\sin v, v)$. Then

$$\begin{aligned} \mathbf{g}^*dx &= \cos v du - u\sin v dv \\ \mathbf{g}^*dy &= \sin v du + u\cos v dv \\ \mathbf{g}^*dz &= dv. \end{aligned}$$

Then we have

$$\begin{aligned} \mathbf{g}^*(dx\wedge dy) &= udu\wedge dv \\ \mathbf{g}^*(dx\wedge dz) &= \cos v du\wedge dv \\ \mathbf{g}^*(dy\wedge dz) &= \sin v du\wedge dv. \end{aligned}$$

Hence if

$$\omega = \left(x^2+y^2\right)dx\wedge dy + xdx\wedge dz + ydy\wedge dz$$

then

$$\begin{aligned} \mathbf{g}^*\omega &= u^2(udu\wedge dv) + (u\cos v)(\cos v du\wedge dv) + (u\sin v)(\sin v du\wedge dv) \\ &= u\left(u^2+1\right)du\wedge dv. \end{aligned}$$

By the calculation above the following Proposition becomes obvious:

Proposition 10.2.4. We have

$$\mathbf{g}^*\left(dx_{i_1}\wedge\cdots\wedge dx_{i_k}\right) = \sum_{1\le j_1<\cdots<j_k\le m} \begin{vmatrix} \frac{\partial g_{i_1}}{\partial u_{j_1}} & \cdots & \frac{\partial g_1}{\partial u_j} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{I_k}}{\partial u_{j_1}} & \cdots & \frac{\partial g_{i_k}}{\partial u_{j_k}} \end{vmatrix} du_{j_1}\wedge\cdots\wedge du_{j_k}.$$

And if $k = n$ in $\mathbb{R}^n$, and $\mathbf{g} : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ then we have

$$\mathbf{g}^*(dx_1 \wedge \cdots \wedge dx_n) = \frac{\partial(g_1, \cdots, g_n)}{\partial(u_1, \cdots, u_n)} du_1 \wedge \cdots \wedge du_n.$$

Proof. Omitted. □

Proposition 10.2.5. Let $U \subseteq \mathbb{R}^m$ be open and $\mathbf{g} : U \to \mathbb{R}^n$ be smooth. If $\omega \in A^k(\mathbb{R}^n)$ then

$$\mathbf{g}^*(d\omega) = d\left(\mathbf{g}^*\omega\right).$$

Proof. If $k = 0$, then

$$\begin{aligned} d\left(\mathbf{g}^* f\right) = d(f \circ \mathbf{g}) &= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \circ \mathbf{g} \right) \frac{\partial g_i}{\partial u_j} \right) du_j \\ &= \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \circ \mathbf{g} \right) \left(\sum_{j=1}^{m} \frac{\partial g_i}{\partial u_j} du_j \right) = \sum_{i=1}^{n} \mathbf{g}^* \left(\frac{\partial f}{\partial x_i} \right) \mathbf{g}^* dx_i \\ &= \mathbf{g}^* \left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i \right) = \mathbf{g}^*(df). \end{aligned}$$

Since d and pullback are linear, it suffices to prove the result for $\omega = f d\mathbf{x}_I$. We have

$$\begin{aligned} \mathbf{g}^*\left(d\left(f d\mathbf{x}_I\right)\right) &= \mathbf{g}^*\left(df \wedge d\mathbf{x}_I\right) = \mathbf{g}^*(df) \wedge \mathbf{g}^*\left(d\mathbf{x}_I\right) = \mathbf{g}^*(df) \wedge d\mathbf{g}_I \\ &= d\left(\mathbf{g}^* f\right) \wedge d\mathbf{g}_I = d\left(\left(\mathbf{g}^* f\right) d\mathbf{g}_I\right) = d\left(\mathbf{g}^*\left(f d\mathbf{x}_I\right)\right). \end{aligned}$$

□

Now we define the integral of an n form:

Definition 10.2.6. Given an n form $\omega = f(\mathbf{x}) dx_1 \wedge \cdots \wedge dx_n$ on a region $\Omega \subseteq \mathbb{R}^n$, then define

$$\int_\Omega \omega = \int_\Omega f dV.$$

It is very important that the n form ω must be written as a function multiple of the standard n form $dx_1 \wedge \cdots \wedge dx_n$.

In the special case of $\det(D\mathbf{g}) > 0$, we have the following restatement of the change of variable theorem:

Proposition 10.2.7. Let $\Omega \in \mathbb{R}^n$ be a region and $\mathbf{g} : \Omega \to \mathbb{R}^n$ be smooth and injective, with $\det(D\mathbf{g}) > 0$ (we say $\mathbf{g}$ is orientation preserving). Then for any n form $\omega = f(\mathbf{x}) dx_1 \wedge \cdots \wedge dx_n$ on $\mathbf{g}(\Omega)$ we have

$$\int_{\mathbf{g}(\Omega)} \omega = \int_\Omega \mathbf{g}^*\omega.$$

10.3 Manifold in $\mathbb{R}^n$, Orientation

Definition 10.3.1. An orientation on $\mathbb{R}^k$ is a choice of $+$ or $-$ for each ordered bases of $\mathbb{R}^k$. Two bases $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ and $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ give the same orientation if the linear map T taking one basis to another has positive determinant ($\det T > 0$).

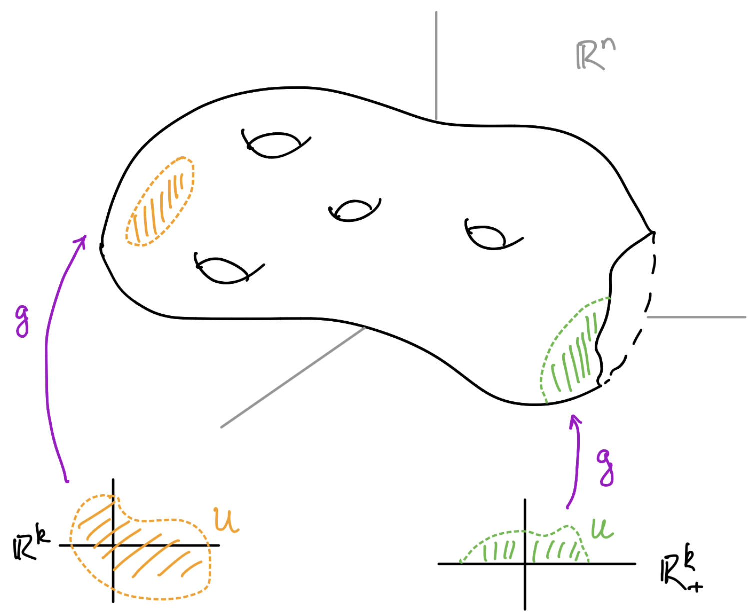
Example 10.3.2. Consider the standard orientation $\{\mathbf{e}_1, \mathbf{e}_2\}$ in $\mathbb{R}^2$. Now consider the basis $\mathbf{v}_1 = (2, 2), \mathbf{v}_2 = (2, -1)$ in $\mathbb{R}^2$. Then the corresponding matrix sending the standard basis to $\mathbf{v}_1, \mathbf{v}_2$ is

$$\begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix},$$

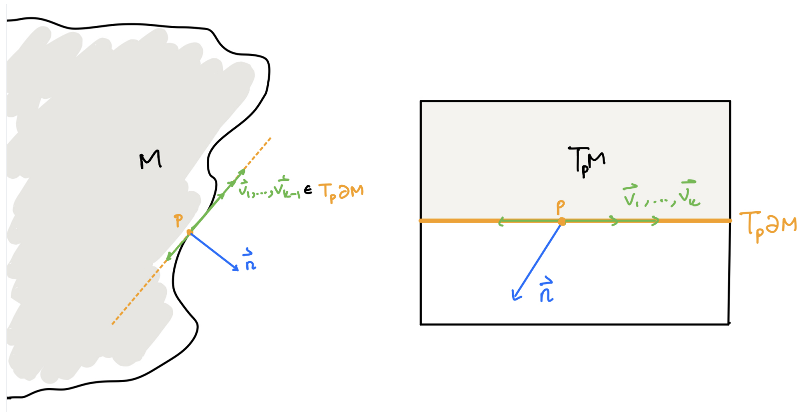
which has negative determinant. Hence they give different orientation of $\mathbb{R}^2$. This is consistent with the right-hand rule, since $\mathbf{e}_1 \times \mathbf{e}_2$ is pointing outward whereas $\mathbf{v}_1 \times \mathbf{v}_2$ points inward.

Definition 10.3.3. A subset $M \subseteq \mathbb{R}^n$ is called a k-dimensional manifold with boundary if for each pint $\mathbf{p} \in M$, there is an open set $W \subseteq \mathbb{R}^n$, an open set $U \subseteq \mathbb{R}^k$, and a homemorphism $\mathbf{g} : U \to \mathbb{R}^n$ such that

- $\mathbf{g}(U) = W \cap M$;
- U is an open subset of either $\mathbb{R}^k$ or of $\mathbb{R}^k_+ = \{\mathbf{u} \in \mathbb{R}^k : u_k \geq 0\}$.



An orientation on a manifold M is a continuous assignment of an orientation to all the tangent spaces $T_\mathbf{p}M$. Since $T_\mathbf{p}M$ is a k-dimensional vector space, whose orientation is determined by the definition in the very beginning. Usually, we use the **standard orientation** on $\mathbb{R}^k$, given by the standard basis, which we call positive. Given an orientation on M, there is an induced orientation on the boundary of M, denoted by ∂M, which is a $k - 1$-dimensional manifold. The induced orientation is defined by the following: Let $\mathbf{p} \in \partial M$. Given a basis $\mathbf{v}_1, \cdots, \mathbf{v}_{k-1} \in T_{\mathbf{p}}\partial M$, this basis defines an orientation for ∂M. We say they define a positive orientation (basis) for $T_{\mathbf{p}}\partial M$ if and only if with $\mathbf{n}$ the outward pointing normal (pointing away from M), the ordered basis $\{\mathbf{n}, \mathbf{v}_1, \cdots, \mathbf{v}_{k-1}\}$ defined a positive basis for $T_{\mathbf{p}}M$.



Example 10.3.4. In this example we consider the orientation of $\partial\mathbb{R}^k_+$. The positive orientation of $T_{\mathbf{p}}\mathbb{R}^k_+$ is given by the standard basis $\{\mathbf{e}_1, \cdots, \mathbf{e}_k\}$. Then the outward pointing normal of $\partial\mathbb{R}^k_+$ away from $\mathbb{R}^k_+$ is given by $-\mathbf{e}_k$. At the same time, $B = \{\mathbf{e}_1, \cdots, \mathbf{e}_{k-1}\}$ clearly consitutes a basis for $\partial\mathbb{R}^k_+$. The orientation of B coincides with the orientation of $\mathbb{R}^{k-1}$. Moreover, it coincides with the boundary orientation if and only if

$$B' = \{-\mathbf{e}_k, \mathbf{e}_1, \cdots, \mathbf{e}_{k-1}\}$$

is a positive basis for $T_{\mathbf{p}}\mathbb{R}^k_+ = \{\mathbf{e}_1, \cdots, \mathbf{e}_k\}$. We see that

$$\det(-\mathbf{e}_k, \mathbf{e}_1, \cdots, \mathbf{e}_{k-1}) = (-1)^k \det(-\mathbf{e}_k, \mathbf{e}_1, \cdots, \mathbf{e}_k).$$

Hence we conclude that

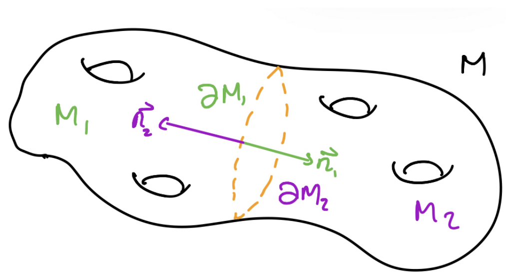
$$\partial\mathbb{R}^k_+ = (-1)^k\mathbb{R}^{k-1}.$$

10.4 Stokes' Theorem

Theorem 10.4.1. *Let M be a compact, oriented k-dimensional manifold with boundary, and ∂M is provided with the boundary orientation (defined in the last section). Let $\omega \in A^{k-1}(M)$. Then*

$$\int_{\partial M} \omega = \int_M d\omega.$$

Remark 10.4.2. Notice that if M is a compact k-manifold (with no boundary), then $\partial M = \emptyset$ and by Stokes's Theorem $\int_M d\omega = 0$. To see this, suppose M is a manifold without boundary, then we can slice M into two pieces, $M = M_1 \cup M_2$, which shares common boundary, $\partial M_1 = \partial M_2$. But then they have the opposite boundary orientation.



Hence by Stokes's theorem,

$$\int_M d\omega = \int_{M_1} d\omega + \int_{M_2} d\omega = \int_{\partial M_1} \omega + \int_{\partial M_2} \omega = 0.$$

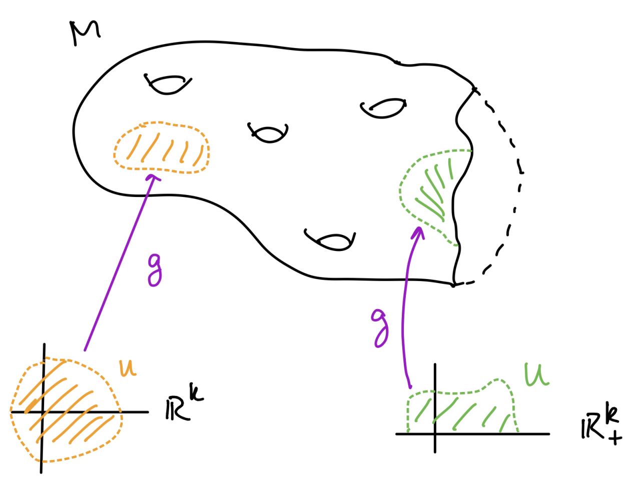
Proof of the Stokes' Theorem. Since M is a manifold, there exists $\mathbf{g}_i : U_i \supseteq \mathbb{R}^k \to M$ where $i = 1, \cdots, n$ such that together, $\mathbf{g}_i(U_i)$ covers the entire M. By using the partition of unity, we know there exists $\rho_1, \cdots, \rho_n$ such that Supp $\rho_i \subseteq U_i$. Moreover,

$$\omega = \sum_{i=1}^{n} \rho_i \omega = \omega_1 + \cdots + \omega_n.$$

Moreover, since M is oriented, we can take each $\mathbf{g}_i$ to be orientation preserving. Since both sides of the equation of Stokes's thconeorem are linearly in ω, we reduce the proof to the case that ω is zero outside of a compact subset of a single orientation preserving coordinate chart, $\mathbf{g} : U \to \mathbb{R}^n$, where U is open either in $\mathbb{R}^k$ or $\mathbb{R}^k_+$. Then we have

$$\int_M d\omega = \int_{\mathbf{g}(U)} d\omega = \int_U \mathbf{g}^*(d\omega) = \int_U d\left(\mathbf{g}^*\omega\right).$$

Then there are two cases given below. We need to check both of them.



Either way, consider the pullback of ω by $\mathbf{g}$, which is a $k-1$ form, which we can write as

$$\mathbf{g}^*\omega = \sum_{i=1}^{k} f_i(\mathbf{x}) dx_1 \wedge \cdots \wedge \widehat{dx_j} \wedge \cdots \wedge dx_k$$

where $\widehat{dx_j}$ indicates that dx_j is omitted. Since $\int_M d\omega = \int_U d\,(\mathbf{g}^*\omega)$, we would like to consider $d(\mathbf{g}^*\omega)$.

$$\begin{aligned} d\,(\mathbf{g}^*\omega) &= \sum_{i=1}^{k} \frac{\partial f_i}{\partial x_i} dx_i \wedge dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k \\ &= \sum_{i=1}^{k} \left((-1)^{i-1} \frac{\partial f_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_i \wedge \cdots \wedge dx_k. \end{aligned}$$

If we are in the first case, i.e. U is open in $\mathbb{R}^k$, then $\omega = 0$ on ∂M, so we need to prove $\int_M d\omega = \int_U d\,(\mathbf{g}^*\omega) = 0$. Since $\omega = 0$ outside of U, we can choose a rectangle $R = [a_1, b_1] \times \cdots \times [a_k, b_k]$ containing U. Hence we may integrate over this rectangle and we will get $\int_U d(\mathbf{g}^*\omega) = \int_R d(\mathbf{g}^*\omega)$. We have

$$\begin{aligned} \int_U d\,(\mathbf{g}^*\omega) &= \int_R \sum_{i=1}^{k} \left((-1)^{i-1} \frac{\partial f_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_i \wedge \cdots \wedge dx_k \\ &= \sum_{i=1}^{k} (-1)^{i-1} \int_R \frac{\partial f_i}{\partial x_i} dx_1 dx_2 \cdots dx_k \\ &= \sum_{i=1}^{k} (-1)^{i-1} \int_{a_k}^{b_k} \ldots \int_{a_1}^{b_1} \left(\int_{a_i}^{b_i} \frac{\partial f_i}{\partial x_i} dx_i \right) dx_1 \cdots \widehat{dx_i} \cdots dx_k \\ &= \sum_{i=1}^{k} (-1)^{i-1} \int_{a_k}^{b_k} \cdots \int_{a_1}^{b_1} [f_i(x_1, ..., b_i, ..., x_k) - f_i(x_1, ..., a_i, ..., x_k)] dx_1 \cdots \widehat{dx_i} \cdots dx_k \\ &= 0. \end{aligned}$$

The last equality follows from the fact that $f_i = 0$ everywhere on the boundary of R. This proves the first case.

Now we consider the other case, where U is open in $\mathbb{R}^k_+$. Once again, we extend U to a rectangle $R = [a_1, b_1] \times \cdots \times [a_{k-1}, b_{k-1}] \times [0, b_k]$ containing U. Now we have

$$
\begin{aligned}
\int_M d\omega &= \int_U d\left(\mathbf{g}^*\omega\right) = \int_R \sum_{i=1}^{k} \left((-1)^{i-1} \frac{\partial f_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_i \wedge \cdots \wedge dx_k \\
&= \sum_{i=1}^{k} (-1)^{i-1} \int_R \frac{\partial f_i}{\partial x_i} dx_1 dx_2 \cdots dx_k \\
&= \sum_{i=1}^{k} (-1)^{i-1} \int_{a_k}^{b_k} \ldots \int_{a_1}^{b_1} \left(\int_{a_i}^{b_i} \frac{\partial f_i}{\partial x_i} dx_i \right) dx_1 \cdots \widehat{dx_i} \ldots dx_k \\
&= \sum_{i=1}^{k} (-1)^{i-1} \int_{a_k}^{b_k} \ldots \int_{a_1}^{b_1} [f_i(x_1, ..., b_i, ..., x_k) - f_i(x_1, ..., a_i, ..., x_k)] dx_1 \cdots \widehat{dx_i} \cdots dx_k \\
&= (-1)^{k-1} \int_{a_{k-1}}^{b_{k-1}} \cdots \int_{a_1}^{b_1} [f_k(x_1, ..., b_k) - f_k(x_1, ..., 0)] dx_1 \cdots dx_{k-1} \\
&= (-1)^k \int_{\partial R} f_k(x_1, ..., 0) dx_1 \cdots dx_{k-1} \\
&= \int_{U \cap \partial \mathbb{R}^k_+} \mathbf{g}^*\omega = \int_{\partial M} \omega.
\end{aligned}
$$

Notice that in the last step, we have turned an integration over the boundary ∂R in $\mathbb{R}^{k-1}$ of the rectangle into an integration of the boundary $\partial \mathbb{R}^k_+$. By the previous discussion we notice that they are off by exactly $(-1)^k$ since $\partial \mathbb{R}^k_+ = (-1)^k \mathbb{R}^{k-1}$. This completes the proof of the theorem. $\square$

Now, we discuss how can we recover all the results that we talked about before from Stokes' theorem.

- Fundamental Theorem of Calculus: Let $\omega = f(x)$, and we obtain from Stokes' theorem

  $$\int_{[a,b]} d\omega = \int_a^b f'(x) dx = \int_{\partial[a,b]} f(x) = f(b) - f(a).$$

  Notice that $\partial[a, b]$ has only two points a, b. Hence the integral degenerates into a sum. Since the normal at a of the line segment $[a, b]$ is opposite to the direction we are moving along the line segment from a to b, we assign a $-$ to $f(a)$. Since the normal at b of the line segment $[a, b]$ is the same as the direction we are moving along the line segment from a to b, we assign a $+$ to $f(b)$.

- Fundamental Theorem of Line integral: Let $\omega = f(\mathbf{x})$, a smooth function $f : \mathbb{R}^3 \to \mathbb{R}$. define our curve $C = \mathbf{x}(s) = (x_1(s), x_2(s), x_3(s))$. parametrized by arclength. Then

  $$\int_C d\omega = \int_C \sum_j \frac{\partial f}{\partial x_j} dx_j = \int_C \sum \frac{\partial f}{\partial x_j} x_j'(s) ds = \int_C \nabla f \cdot \mathbf{T} ds$$

where $\mathbf{T} = (x_1'(s), x_2'(s), x_3'(s))$ is the unit tangent vector along C, since the curve is parametrized by arclength (arclength parametrization gives unit speed curve). On the other hand we have

$$\int_{\partial C} f(\mathbf{x}) = f(\mathbf{x}(b)) - f(\mathbf{x}(a)).$$

The orientation will induce a minus sign in front of $f(\mathbf{x}(a))$, similar to the Fundamental Theorem of Calculus. Combine them together, we get the desired result, which is equivalent to what we have before.

- Green's Theorem: Just let $\omega = P\,dx + Q\,dy$. The result immediately follows.

- Divergence Theorem: We first introduce what dS means. The area element dS eats two vectors $\mathbf{v}, \mathbf{w}$, and splits out the the area of the parallelogram with $\mathbf{v}, \mathbf{w}$ being its two sides. Hence $dS(\mathbf{v}, \mathbf{w}) = |\mathbf{v} \times \mathbf{w}|$. If we multiply dS by the normal vector $\mathbf{n}$ to the plane containing $\mathbf{v}, \mathbf{w}$, then we get a vector $d\mathbf{S}(\mathbf{v}, \mathbf{w}) = \mathbf{n}dS(\mathbf{v}, \mathbf{w}) = \mathbf{v} \times \mathbf{w}$. Then this suggests the definition of area element to be

$$d\mathbf{S} = (dy \wedge dz, dz \wedge dx, dx \wedge dy).$$

Then

$$dy \wedge dz(\mathbf{v}, \mathbf{w}) = \begin{vmatrix} dy(\mathbf{v}) & dz(\mathbf{v}) \\ dy(\mathbf{w}) & dz(\mathbf{w}) \end{vmatrix} = \begin{vmatrix} v_y & v_z \\ w_y & w_z \end{vmatrix} = v_y w_z - v_z w_y.$$

Then one can check that $d\mathbf{S}(\mathbf{v}, \mathbf{w}) = \mathbf{v} \times \mathbf{w}$ by similar procedure. Let $\mathbf{F} = (F_x, F_y, F_z)$ be a vector field. Then we consider the differential 2 form

$$\omega = \mathbf{F} \cdot d\mathbf{S} = F_x dy \wedge dz + F_y dz \wedge dx + F_z dx \wedge dy.$$

Then

$$d\omega = \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\right) dx \wedge dy \wedge dz = (\nabla \cdot \mathbf{F}) dx \wedge dy \wedge dz.$$

Hence let M be a region in $\mathbb{R}^3$. Stokes' theorem says

$$\int_M d\omega = \int_M \nabla \cdot \mathbf{F} dV = \int_{\partial M} \omega = \int_{\partial M} \mathbf{F} \cdot d\mathbf{S}.$$

- Stokes' Theorem in 3D: Let $\mathbf{F} = (F_x, F_y, F_z)$ be a vector field. let M be a surface, its boundary $\partial M = C = (x(t), y(t), z(t)))$ will be a curve. Then define the curve element

$$d\mathbf{r} = (x'(t), y'(t), z'(t))dt = (dx, dy, dz).$$

Now, consider the 1 form

$$\omega = F_x dx + F_y dy + F_z dz.$$

Then we have

$$\begin{aligned}
d\omega &= \frac{\partial F_x}{\partial y} dy \wedge dx + \frac{\partial F_x}{\partial z} dz \wedge dx + \frac{\partial F_y}{\partial x} dx \wedge dy \\
&\quad + \frac{\partial F_y}{\partial z} dz \wedge dy + \frac{\partial F_z}{\partial x} dx \wedge dz + \frac{\partial F_z}{\partial y} dy \wedge dz \\
&= \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) dy \wedge dz + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) dx \wedge dy + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) dz \wedge dx \\
&= (\nabla \times \mathbf{F}) \cdot d\mathbf{S}.
\end{aligned}$$

Then by generalized Stokes' theorem we have

$$\int_M d\omega = \int_M \nabla \times \mathbf{F} \cdot d\mathbf{S} = \int_{\partial M} \omega = \int_C \mathbf{F} \cdot d\mathbf{r}.$$

This completes our grand unifield tour of integration in $\mathbb{R}^n$. In the remaining part of this notes, we will study a highly developed theory of abstract integration, the Lebesgue theory.

Chapter 11

Lebesgue Theory

The only teaching that a professor can give, in my opinion, is that of thinking in front of his students.

-Henri Lebesgue

We will give a brief Introduction to Lebesgue's theory of integration. Many results are only briefly mentioned, whose proof will only be sketched, or sometimes omitted.

We first introduce some notation. Let A, B be two sets, then

$$A - B = \{x \in A \mid x \notin B\}.$$

This does not mean that $B \subseteq A$. However, if this is the case, we usually use $A \setminus B$. The emptyset is denoted by 0 and A and B are said to be disjoint if $A \cap B = 0$.

Definition 11.0.1. A family of sets R is called a ring if $A \in R$ and $B \in R$ implies $A \cup B \in R$ and $A - B \in R$. Notice that this implies $A \cup B = A - (A - B) \in R$. We call R a σ-ring if it is a ring and whenever $A_n \in R$ then

$$\bigcup_{n=1}^{\infty} A_i \in R.$$

Notice that this implies

$$\bigcap_{n=1}^{\infty} A_n = A_1 - \bigcup_{n=1}^{\infty} (A_1 - A_n) \in R.$$

Definition 11.0.2. A set function $\phi : R \to \mathbb{R} \cup \{\pm\infty\}$ sends every $A \in R$ to a number $\phi(A)$. We say ϕ is additive if $A \cap B = 0$ implies

$$\phi(A \cup B) = \phi(A) + \phi(B).$$

And we say ϕ is countably additive if $A_i \cap A_j = 0$ for all $i \neq j$ implies

$$\phi\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \phi\left(A_n\right).$$

Notice that if $\phi(A_j) = \infty$ for some j then the RHS will be meaningless. Therefore, we shall always assume the range of ϕ does not contain $\pm\infty$. Notice that the LHS is independent of the order of A_n. Hence RHS converges absolutely if it converges at all.

Proposition 11.0.3. Let ϕ be additive. Then the following are easily proved:

1. $\phi(0) = 0$.
2. Let $A_i \cap A_j = 0$ for $i \neq j$. Then $\phi(A_1 \cup \cdots \cup A_n) = \phi(A_1) + \cdots + \phi(A_n)$.
3. $\phi(A_1 \cup A_2) + \phi(A_1 \cap A_2) = \phi(A_1) + \phi(A_2)$.
4. If $\phi(A) \geq 9$ for all A and $A_1 \subseteq A_2$, then $\phi(A_1) \leq \phi(A_2)$.
5. If $B \subseteq A$ and $|\phi(B)| < \infty$, then $\phi(A - B) = \phi(A) - \phi(B)$.

Proof. For (3), notice that

$$\begin{aligned} &\phi\left((A_1 - A_2) \cup (A_1 \cap A_2) \cup (A_2 - A_1)\right) + \phi(A_1 \cap A_2) \\ &= \phi(A_1 - A_2) + \phi(A_1 \cap A_2) + \phi(A_2 - A_1) + \phi(A_1 \cap A_2) \\ &= \phi(A_1) + \phi(A_2). \end{aligned}$$

□

Theorem 11.0.4. *Suppose ϕ is countably additive on a ring R and $A_n \in R$ with $A_1 \subseteq A_2 \subseteq A_3 \cdots$, $A \in R$, and*

$$A = \bigcup_{n=1}^{\infty} A_n.$$

Then as $n \to \infty$ we have $\phi(A_n) \to \phi(A)$.

Proof. Let $B_1 = A_1$ and

$$B_n = A_n - A_{n-1} \quad (n = 2, 3, \ldots).$$

Then B_n are pairwise disjoint, and $A_n = \bigcup_{k=1}^{\infty} B_k$. Hence we have

$$\phi(A_n) = \sum_{i=1}^{n} \phi(B_i).$$

Taking the limit we have

$$\phi(A) = \sum_{i=1}^{\infty} \phi(B_i).$$

□

Now that we have some basic ideas on sets, rings, and set functions on rings, our next goal is to construct the so called Lebesgue measure.

Definition 11.0.5. Let $\mathbb{R}^p$ denote the p-dimensional Euclidean space. By an interval in $\mathbb{R}^p$ we mean the set of points $\mathbf{x} = (x_1, \cdots, x_p)$ such that

$$a_i \leq (<) x_i (<) \leq b_i \qquad i = 1, ..., p.$$

Notice that we do not rule out the possibility that the the interval is open or clopen. The emptyset, in particular, are includes among the intervals.

If A is the union of a finite number of intervals, A is said to be an elementary set. If I is an interval, we define

$$m(I) = \prod_{i=1}^{p} (b_i - a_i).$$

If $A = I_1 \cup \cdots \cup I_n$ and if these intervals are pairwise disjoint, then we set

$$m(A) = m(I_1) + \cdots + m(I_n).$$

Let $\mathscr{E}$ denote the family of elementary subsets of $\mathbb{R}^p$.

Notice that $m(I)$ gives the p-dimensional volumes of the rectangle I respectively. If A is a union of rectangles, then $m(A)$ also gives the volume of A. Then, for any arbitrary set E in $\mathbb{R}^p$, we can approximate it with rectangles, and we can define $m(E)$ to be the volumes of these rectangle approximation. Then as our approximation gets more accurate by taking the limit, this will give the volume of E. We make this argument formal:

Definition 11.0.6. A nonnegative additive set function $\phi : \mathscr{E} \to \mathbb{R}$ is said to be regular if to every $A \in \mathscr{E}$ and to every $\epsilon > 0$ there exists sets $F, G \in \mathscr{E}$ such that F is closed, G is open, $F \subseteq A \subseteq G$ and

$$\phi(G) - \epsilon \leq \phi(A) \leq \phi(F) + \epsilon.$$

Definition 11.0.7. Let μ be additive, regular, nonnegative, and finite on $\mathscr{E}$. Consider countable coverings of any set $E \subseteq \mathbb{R}^p$ by open elementary sets A_n (we can always do this, see HW7 Q1):

$$E \subset \bigcup_{n=1}^{\infty} A_n.$$

Define

$$\mu^*(E) = \inf \sum_{n=1}^{\infty} \mu(A_n)$$

where inf is taken over all countavle coverings of E by open elementary sets. We call $\mu^*(E)$ the outer measure of E corresponding to μ.

It is clear that $\mu^*(E) \geq 0$ for all E. And if $E_1 \subseteq E_2$ then $\mu^*(E_1) \leq \mu^*(E_2)$. This seems to be a reasonable extension of the volume measure on the general subset of $\mathbb{R}^p$. Now we investigate some properties of the outer measure, which is quite intuitive:

Theorem 11.0.8. *Let μ, μ^* be defined above.*

1. *For every $A \in \mathscr{E}$ we have $\mu^*(A) = \mu(A)$.*
2. *If $E = \bigcup_{n=1}^{\infty} E_n$, then $\mu^*(E) \leq \sum_{n=1}^{\infty} \mu^*(E_n)$.*

Proof. For (a), sinc μ is regular, A is contained in an open elementary set G such that $\mu(G) \leq \mu(A) + \epsilon$. Since G is an open covering of A by elementary set we have $\mu^*(A) \leq \mu(G)$. Since ϵ is arbitrary, this gives

$$\mu^*(A) \leq \mu(A).$$

The definition of μ^* shows that there is a sequence A_n of open elementary sets whose union contains A such that

$$\sum_{n=1}^{\infty} \mu\left(A_n\right) \leq \mu^*(A) + \epsilon.$$

Again by regularity, there us a closed elementary set F contained in A such that $\mu(F) \geq \mu(A) - \epsilon$. Since F is compact, we have

$$F \subset A_1 \cup \cdots \cup A_N$$

for some N. Hence

$$\mu(A) \leq \mu(F) + \epsilon \leq \mu\left(A_1 \cup \cdots \cup A_N\right) + \epsilon \leq \sum_{1}^{N} \mu\left(A_n\right) + \epsilon \leq \mu^*(A) + 2\epsilon.$$

This proves the claim.

For (b), suppose $E = \bigcup E_n$ and assume $\mu^*(E_n) < \infty$ for all n. Given $\epsilon > 0$, there are covering $A_{nk}, k = 1, 2, 3, ...$ of E_n by open elementary sets such that

$$\sum_{k=1}^{\infty} \mu\left(A_{nk}\right) \leq \mu^*\left(E_n\right) + 2^{-n}\epsilon.$$

Then

$$\mu^*(E) \leq \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mu\left(A_{nk}\right) \leq \sum_{n=1}^{\infty} \mu^*\left(E_n\right) + \epsilon.$$

And the claim follows. □

Definition 11.0.9. For $A, B \in \mathbb{R}^p$, we define the symmetric difference

$$S(A, B) = (A - B) \cup (B - A).$$

And let

$$d(A, B) = \mu^*(S(A, B)).$$

We write $A_n \to A$ if

$$\lim_{n\to\infty} d\left(A, A_n\right) = 0.$$

Then we further make the following definition:

- If there is a sequence A_n of elementary set such that $A_n \to A$, we say A is finitely μ-measurable and write $A \in \mathfrak{M}_F(\mu)$.

- If A is the union of a countable collection of finitely μ-measurable sets, we say A is μ-measurable and write $A \in \mathfrak{M}(\mu)$.

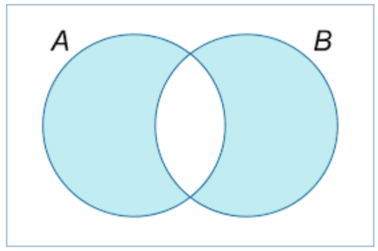


Figure 11.1: The symmetric difference of two sets A, B are shaded blue. If μ^* represents the volume of these two sets, then we can see that as the symmetric difference goes to zero, A, B are becoming "identical".

In fact, the d defined above is close to a metric. It can be shown that (pg. 306 in the textbook) that d satisfies the following properties:

- $d(A, B) = d(B, A)$;
- $d(A, A) = 0$ (But note that $d(A, B) = 0$ does not imply $A = B$);
- $d(A, B) \leq d(A, C) + d(C, B)$. Namely, d satisfies the triangle inequality.

This allows us to make one more observation of finitely μ-measurable sets: Their values on μ^* is finite. Suppose that $A_n \to A$. Then there must exist some N such that $d(A_N, A) < 1$. And suppose $\mu^*(A_N) = \mu(A_N) < M$ since elementary sets have finite volume. Then,

$$\begin{aligned}\mu^*(A) &= \mu^*(S(A, A_N) \cup (A \cap A_N)) \\ &\leq \mu^*(S(A, A_N)) \cup \mu^*(A \cap A_N) \\ &\leq d(A, A_N) + \mu(A_N) < M + 1 < \infty.\end{aligned}$$

Finally, it can be shown that

$$|\mu^*(A) - \mu^*(B)| \leq d(A, B).$$

Next, we claim that every regular set function on $\mathscr{E}$ can be extended to a countably additive set function on a σ-ring which contains $\mathscr{E}$. The proof is a bit technical, and for the sake of not distracting the readers from the main thread, we refer the proof of the following theorem to the book (pg. 307):

Theorem 11.0.10. *$\mathfrak{M}(\mu)$ is a σ-ring, and μ^* is countably additive on $\mathfrak{M}(\mu)$.*

Now we replace $\mu^*(A)$ by $\mu(A)$ if $A \in \mathfrak{M}(\mu)$. Thus μ, originally only defined on $\mathfrak{E}$, is extended to a countably additive set function on the σ-ring $M(\mu)$. This extended set function is called a measure. The special case $\mu = m$ is called the Lebesgue measure on $\mathbb{R}^p$.

11.1 Measurable Functions and Simple Functions

In this section we establish some preparatory work to define the Lebesgue intergal.

Definition 11.1.1. Let X be a set. X is said to be a measure space if there exists a σ-ring $\mathfrak{M}$ of subsets of X (called measurable sets) and a nonnegative countably additive set function μ (called a measure) defined on $\mathfrak{M}$.

Definition 11.1.2. Let X be a measurable space and $f : X \to \mathbb{R} \cup \{\pm\infty\}$. Then the function f is said to be measurable if the set

$$f^{-1}((a, +\infty)) = \{x |\ f(x) > a\}$$

is measurable for every real a.

We note that countable intersection of measurable sets is measurable, and complement of a measurable set is measurable.

Theorem 11.1.3. *The following are equivalent:*

1. $\{x |\ f(x) > a\}$ *is measurable for every real* a*;*
2. $\{x |\ f(x) \geq a\}$ *is measurable for every real* a*;*
3. $\{x |\ f(x) < a\}$ *is measurable for every real* a*;*
4. $\{x |\ f(x) \leq a\}$ *is measurable for every real* a*.*

Proof. The relations

$$\begin{aligned}
\{x \mid f(x) \geq a\} &= \textstyle\bigcap_{n=1}^{\infty} \left\{x \mid f(x) > a - \frac{1}{n}\right\} \\
\{x \mid f(x) < a\} &= X - \{x \mid f(x) \geq a\} \\
\{x \mid f(x) \leq a\} &= \textstyle\bigcap_{n=1}^{\infty} \left\{x \mid f(x) < a + \frac{1}{n}\right\} \\
\{x \mid f(x) > a\} &= X - \{x \mid f(x) \leq a\}
\end{aligned}$$

shows successively that 1 implies 2, 2 implies 3, 3 implies 4, and 4 implies 1. $\square$

Theorem 11.1.4. *If f is measurable, then $|f|$ is measurable.*

Proof. $\{x || f(x) |< a\} = \{x \mid f(x) < a\} \cap \{x \mid f(x) > -a\}$. $\square$

Theorem 11.1.5. *Let f_n be a sequence of measurable functions, for $x \in X$, put $g(x) = \sup f_n(x)$, and $h(x) = \limsup_{n\to\infty} f_n(x)$. Then both g and h are measurable. The same is true for* inf *and* lim inf.

Proof. We have $\{x \mid g(x) > a\} = \bigcup_{n=1}^{\infty} \{x \mid f_n(x) > a\}$, which proves the first claim. And

$$h(x) = \inf g_m(x) \qquad \text{where} \quad g_m(x) = \sup f_n(x)\ (n \geq m).$$

This proves the second claim. $\square$

Corollary 11.1.6. *If f, g are measurable, then* $\max(f, g)$ *and* $\min(f, g)$ *are measurable. If*

$$f^+ = \max(f, 0), \quad f^- = -\min(f, 0),$$

it follows then that f^+ and f^- are measurable.

Finally, the limit of a convergent sequence of measurable functions is measurable.

Theorem 11.1.7. *Let f, g be measurable real-valued functions on X. Let F be real and continous function on $\mathbb{R}^2$ and put*

$$h(x) = F(f(x), g(x))$$

Then h is measurable. In particular, $f + g$ and fg are measurable.

Proof. see textbook pg. 312. □

To summarize, *all ordinary operations of analysis, including limit operations, when applied to measurable functions, lead to measurable functions. In other words, all functions that are ordinarily met with are measurable.*

Definition 11.1.8. Let s be a real-valued function on X. If the range of s is finite, then we say that s is a simple function.

Let $E \subseteq X$, and put

$$\chi_E(x) = \begin{cases} 1 & (x \in E) \\ 0 & (x \notin E). \end{cases}$$

χ_E is called the characteristic function on E. Then it is clear that every simple function is a finite linear combination of characteristic functions:

$$s = \sum_{i=1}^{n} c_i \chi_{E_i}.$$

Now come the key theorem, which says that every function can be approximated by simple functions.

Theorem 11.1.9. *Let $f : X \to \mathbb{R}$. There exists a sequence s_n of simple functions such that $s_n(x) \to f(x)$ as $n \to \infty$, for every $x \in X$. If f is measurable, s_n may be chosen to be a sequence of measurable functions. If $f \geq 0$, s_n can be chosen to be a monotonically increasing sequence.*

Proof. If $f \geq 0$, define

$$E_{ni} = \left\{ x \mid \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n} \right\}, \quad F_n = \{x \mid f(x) \geq n\}$$

for $n = 1, 2, 3, ...$ and $i = 1, 2, ..., n2^n$. Put

$$s_n = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{E_{ni}} + n\chi_{F_n}.$$

In the general case, let $f = f^+ - f^-$. and apply the preceding construction to f^+ and to f^-. The sequence s_n converges uniformly to f is f is bounded, since $\|f - f_n\|_\infty < 2^{-n}$. □

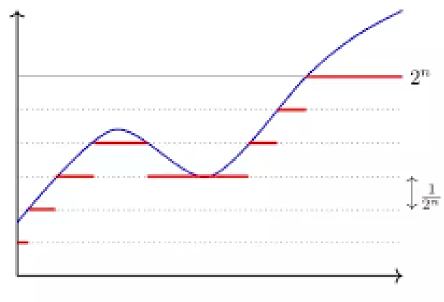


Figure 11.2: Simple function approximation illustration.

11.2 Lebesgue Integration

Definition 11.2.1. Suppose $\mathfrak{M}$ is the σ-ring of measurable sets, and μ is the measure, and suppose

$$s(x) = \sum_{i=1}^{n} c_i \chi_{E_i}(x) \quad (x \in X, c_i > 0)$$

is measurable with $E \in \mathfrak{M}$. We define

$$I_E(s) = \sum_{i=1}^{n} c_i \mu(E \cap E_i).$$

If f is measurable and **nonnegative**, then define

$$\int_E f d\mu = \sup I_E(s),$$

where the sup is taken over all measurable simple function s such that $0 \le s \le f$. The LHS is called the Lebesgue integral of f, with respect to the measure μ, over the set E. It should be noted that the integral may diverge.

It can be easily seen that

$$\int_E s d\mu = I_E(s).$$

Now, let f be measurable, we consider two integrals defined above:

$$\int_E f^+ d\mu, \quad \int_E f^- d\mu.$$

And then we define

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu.$$

If both integrals on the RHS are finoite, then the LHS is finite, and we say f is integral on E in the Lebesgue sense, with respect to μ, and we write $f \in \mathscr{L}(\mu)$ on E. If $\mu = m$, the usual notation is $f \in \mathscr{L}$ on E.

Note that if the RHS is infinite, then $\int_E f d\mu$ is defined, but is not integrable in the Lebesgue sense. We stipulate that f is integrable only if its integral over E is finite.

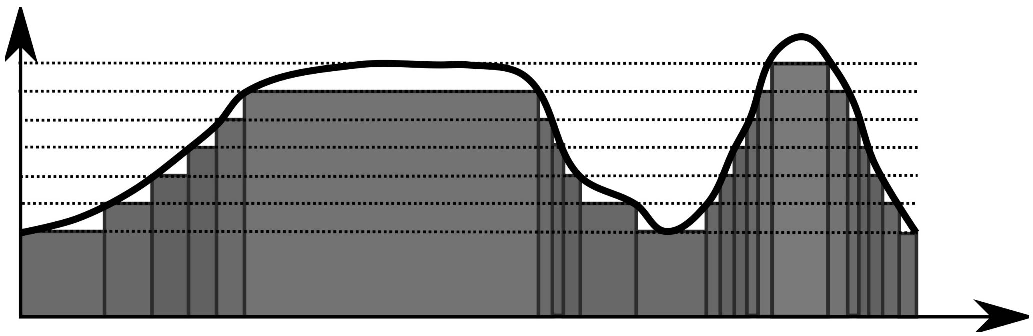


Figure 11.3: Lebesgue integration illustration

Notice the difference and yet the similarities between Riemann and Lebesgue integral. In both cases, we try to approximate the area under curves by rectangles. In the Riemann case, we divide the domain into small intervals $\Delta x = (x_i, x_{i+1})$, and assign each interval the value $f(x_i)$ of the function on the left endpoint, and then we sum up the area of these rectangles: $\sum f(x_i)\Delta x$ and then take the sup. In Lebesgue case, we are actually dividing the range of the function into small pieces $\Delta c = (c_i, c_{i+1})$, and we look for the volume of the preimage $\mu(f^{-1}(\Delta c))$ of this small piece. And then we assign this volume the value c_i of the function at the lower end point. Then we get the sum

$$\sum_{i=1}^{n} c_i \mu(f^{-1}(\Delta c)).$$

i.e. we approximate f by simple functions, calculate its area. Then we take the limit and get finer approximation.

Proposition 11.2.2. Properties of Lebesgue integral:

(a) If f is measurable and bounded on E and if $\mu(E) < \infty$, then $f \in \mathscr{L}(\mu)$ on E.

(b) If $a \leq f(x) \leq b$ for $x \in E$, and $\mu(E) < \infty$, then

$$a\mu(E) \leq \int_E f d\mu \leq b\mu(E).$$

(c) If $f, g \in \mathscr{L}(\mu)$ on E, and $f(x) \leq g(x)$ for $x \in E$, then

$$\int_E f d\mu \leq \int_E g d\mu.$$

(d) If $f \in \mathscr{L}(\mu)$ on E, then $cf \in \mathscr{L}(\mu)$ on E for every finite constant c, and

$$\int_E cf d\mu = c \int_E f d\mu.$$

(e) If $\mu(E) = 0$ and f is measurable, then $\int_E f d\mu = 0$.

(f) If $f \in \mathscr{L}(\mu)$ on E, $A \in \mathfrak{M}$, and $A \subseteq E$, then $f \in \mathscr{L}(\mu)$ on A.

Theorem 11.2.3. *Suppose f is measurable and nonnegative on X. For $A \in \mathfrak{M}$, we use the shorthand notation*

$$\phi(A) = \int_A f d\mu.$$

Then ϕ is countably additive on $\mathfrak{M}$. The same conclusion holds if $f \in \mathscr{L}(\mu)$ on X.

Proof. To prove the first claim, we have to show that

$$\phi(A) = \sum_{n=1}^{\infty} \phi\left(A_n\right)$$

if $A_n \in \mathfrak{M}$, with $A_i \cap A_j = 0$ for $i \neq j$ and $A = \bigcup A_n$.

If f is a characteristic function, then since μ is countably additive by assumption and

$$\int_A \chi_E d\mu = \mu(A \cap E),$$

we see that the claim holds. If f is simple, then it is a finite linear combination of characteristic functions, hence the claim also holds.

In general, for every measurable simple function s such that $0 \leq s \leq f$,

$$\int_A s d\mu = \sum_{n=1}^{\infty} \int_{A_n} s d\mu \leq \sum_{n=1}^{\infty} \phi\left(A_n\right) = \sum_{n=1}^{\infty} \int_{A_n} f d\mu.$$

Theredore, taking the sup of all such simple functions gives

$$\phi(A) \leq \sum_{n=1}^{x} \phi\left(A_n\right).$$

We assume the RHS is finite, otherwise the claim trivially holds. Given $\epsilon > 0$, we can choose a measurable function s such that $0 \leq s \leq f$ and

$$\int_{A_1} s d\mu \geq \int_{A_1} f d\mu - \epsilon, \quad \int_{A_2} s d\mu \geq \int_{A_2} f d\mu - \epsilon.$$

Hence

$$\phi\left(A_1 \cup A_2\right) \geq \int_{A_1 \cup A_2} s d\mu = \int_{A_1} s d\mu + \int_{A_2} s d\mu \geq \phi\left(A_1\right) + \phi\left(A_2\right) - 2\epsilon.$$

Since ϵ is arbitrary, we have

$$\phi(A_1 \cup A_2) \geq \phi(A_1) + \phi(A_2).$$

Then by induction,

$$\phi\left(A_1 \cup \cdots \cup A_n\right) \geq \phi\left(A_1\right) + \cdots + \phi\left(A_n\right)$$

for every n. Since $A \supset A_1 \cup \cdots \cup A_n$, we have

$$\phi(A) \geq \sum_{n=1}^{n} \phi(A_n)$$

for every n. Taking the limit we see

$$\phi(A) \geq \sum_{n=1}^{\infty} \phi(A_n).$$

This proves the first claim. The last claim follows immediately if we write $f = f^+ - f^-$, and apply the first claim to f^+ and f^-. □

Corollary 11.2.4. *If $A, B \in \mathfrak{M}$, $B \subseteq A$, and $\mu(A - B) = 0$, then*

$$\int_A f d\mu = \int_B f d\mu.$$

Proof. This follows from $A = B \cup (A - B)$, the previous theorem, and the previous proposition (e). This shows that the set of measure zero are negligible in integration. □

Example 11.2.5. The Dirichlet function $\mathscr{D} : [0, 1] \to \mathbb{R}$ defined by

$$\mathscr{D}(x) = \begin{cases} 1 & x \in [0, 1] \cap \mathbb{Q} \\ 0 & x \in [0, 1] - \mathbb{Q}. \end{cases}$$

Notice that rationals have measure zero on [0, 1]. Since a point has measure zero, $\mathbb{Q}$ is countable on [0, 1], and μ is countably additive, we have

$$\mu(\mathbb{Q}) = \mu\left(\bigcup_{n=1}^{\infty} \{q_n\}\right) = \sum_{n=1}^{\infty} \mu(\{q_n\}) = \sum_{n=1}^{\infty} 0 = 0.$$

Therefore,

$$\int_{[0,1]} \mathscr{D} d\mu = 0.$$

If a property P holds for evert $x \in E - A$, and if $\mu(A) = 0$, we says that P holds almost everywhere on E.

Theorem 11.2.6. *If $f \in \mathscr{L}(\mu)$ on E, then $|f| \in \mathscr{L}(\mu)$ on E, and*

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu.$$

Proof. Write $E = A \cup B$, where $f(x) \geq 0$ on A and $f(x) < 0$ on B. Then

$$\int_E |f| d\mu = \int_A |f| d\mu + \int_B |f| d\mu = \int_A f^+ d\mu + \int_B f^- d\mu < +\infty.$$

Hence $|f| \in \mathscr{L}(\mu)$. Since $f \leq |f|$ and $-f \leq |f|$, we have

$$\int_E f d\mu \leq \int_E |f| d\mu, \quad -\int_E f d\mu \leq \int_E |f| d\mu.$$

And the claim follows. □

The rest of this section is dedicated to show that Lebesgue integral behaves well under limit operations, which shows its advantage compared with the usual Riemann integral.

Theorem 11.2.7 (Lebesgue's Monotone Convergence Theorem)**.** *Suppose $E \in \mathfrak{M}$. Let f_n be a sequence of measurable functions such that*

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots \quad (x \in E)$$

and let f denote the limit of f_n as $n \to \infty$. Then

$$\int_E f_n d\mu \to \int_E f d\mu \quad (n \to \infty).$$

Proof. Clearly as $n \to \infty$ we have

$$\int_E f_n d\mu \to \alpha$$

for some α, and since $\int f_n \leq \int f$ we have

$$\alpha \leq \int_E f d\mu.$$

Now we want to show the reverse inequality. Choose $0 < c < 1$ amd let s be a simple measurable function such that $0 \leq s \leq f$ and put

$$E_n = \{x \mid f_n(x) \geq cs(x)\} \quad (n = 1, 2, 3, \ldots).$$

Notice that this set is nonempty, since $cs \leq f$ and $f_n \to f$, there exists N such that $f_N(x) \geq cs$ for some x. Clearly

$$E_1 \subset E_2 \subset E_3 \subset \cdots \qquad E = \bigcup_{n=1}^{\infty} E_n.$$

For every n we have

$$\int_E f_n d\mu \geq \int_{E_n} f_n d\mu \geq c \int_{E_n} s d\mu.$$

Let $n \to \infty$. Since integral is a countably additive set function and we have $\int_{E_n} s d\mu \to \int_E s d\mu$. Thus

$$\alpha \geq c \int_E s d\mu.$$

As $c \to 1$ we have

$$\alpha \geq \int_E s d\mu.$$

Thus we have

$$\alpha \geq \int_E f d\mu$$

since integral is defined to be a sup. $\square$

Theorem 11.2.8. *Suppose $f = f_1 + f_2$ with $f_1, f_2 \in \mathscr{L}(\mu)$ on E. Then $f \in \mathscr{L}(\mu)$ on E and*

$$\int_E f d\mu = \int_E f_1 d\mu + \int_E f_2 d\mu.$$

Proof. Omitted, see pg. 319-320. □

Theorem 11.2.9. *Let $E \in \mathfrak{M}$ and f_n is a sequence of nonnegative measurable functions and*

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad (x \in E).$$

Then

$$\int_E f d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu.$$

Proof. The partial sum is a monotonically increasing sequence. □

Theorem 11.2.10 (Fatou's Theorem)**.** *Suppose $E \in \mathfrak{M}$. If f_n is a sequence of nonnegative measurable functions and*

$$f(x) = \liminf_{n\to\infty} f_n(x) \quad (x \in E),$$

then

$$\int_E f d\mu \le \liminf_{n\to\infty} \int_E f_n d\mu.$$

Proof. Put

$$g_n(x) = \inf f_i(x) \quad (i \ge n).$$

Then g_n is measurable on E and

$$\begin{aligned} 0 \le g_1(x) \le g_2(x) \le \cdots \\ g_n(x) \le f_n(x), \\ g_n(x) \to f(x) \quad (n \to \infty). \end{aligned}$$

Hence

$$\int_E g_n d\mu \to \int_E f d\mu.$$

And we have

$$\begin{aligned} \int_E g_n d\mu &\le \int_E f_n d\mu \\ \Rightarrow \liminf_{n\to\infty} \int g_n d\mu &\le \liminf_{n\to\infty} \int f_n d\mu \\ \Rightarrow \lim_{n\to\infty} \int g_n d\mu &\le \liminf_{n\to\infty} \int f_n d\mu. \end{aligned}$$

This proves the claim. □

Theorem 11.2.11 (Lebesgue's Dominated Convergence Theorem)**.** *Suppose $E \in \mathfrak{M}$. If f_n is a sequence of nonnegative measurable functions such that $f_n \to f$. If there exists a function $g \in \mathscr{L}(\mu)$ on E such that*

$$|f_n(x)| \leq g(x) \quad (n = 1, 2, 3, \ldots, x \in E),$$

then

$$\lim_{n\to\infty} \int_E f_n d\mu = \int_E f d\mu.$$

Proof. Since $f_n + g \geq 0$, Fatou's theorem implies

$$\int_E (f+g) d\mu \leq \liminf_{n\to\infty} \int_E (f_n + g)\, d\mu \Rightarrow \int_E f d\mu \leq \liminf_{n\to\infty} \int_E f_n d\mu.$$

Also $g - f_n \geq 0$, hence

$$\int_E (g-f) d\mu \leq \liminf_{n\to\infty} \int_E (g - f_n)\, d\mu \Rightarrow -\int_E f d\mu \leq \liminf_{n\to\infty} \left[-\int_E f_n d\mu\right],$$

which is the same as

$$\int_E f d\mu \geq \limsup_{n\to\infty} \int_E f_n d\mu.$$

The existence of $\lim \int f_n d\mu$ implies

$$\limsup_{n\to\infty} \int_E f_n d\mu = \liminf_{n\to\infty} \int_E f_n d\mu.$$

This completes the proof. □