Chapter 1

Fourier Series

宋人有善為不龜手之藥者，世世以洴澼絖為事。客聞之，請買其方百金。聚族而謀曰： 我世世為洴澼絖，不過數金；今一朝而鬻技百金，請與之。 客得之，以説吳王。越有難，吳王使之將。冬與越人水戰，大敗越人，裂地而封之。 能不龜手，一也；或以封，或不免於洴澼絖，則所用之異也。 莊子 逍遙遊

In this chapter we study Fourier series. Basic definitions and examples are given in Section 1. In Section 2 we prove the fundamental Riemann-Lebesgue lemma and discuss the Fourier series from the mapping point of view. Pointwise and uniform convergence of the Fourier series of a function to the function itself under various regularity assumptions are studied in Section 3. In Section 1.5 we establish the L^2-convergence of the Fourier series without any additional regularity assumption. There are two applications. In Section 1.4 it is shown that every continuous function can be approximated by polynomials in a uniform manner. In Section 1.6 a proof of the classical isoperimetric problem for plane curves is presented. In the two appendices basic facts on series of functions and sets of measure zero are present.

1.1 Definition and Examples

The concept of series of functions and their pointwise and uniform convergence were discussed in Mathematical Analysis II. Power series and trigonometric series are the most important classes of series of functions. We learned power series in Mathematical Analysis II and now we discuss Fourier series. You are referred to Appendix I for basic definitions of series and series of functions.

First of all, a **trigonometric series** on $[-\pi, \pi]$ is a series of functions of the form

$$\sum_{n=0}^{\infty}(a_n \cos nx + b_n \sin nx), \quad a_n, b_n \in \mathbb{R}.$$

As $\cos 0x = 1$ and $\sin 0x = 0$, we always set $b_0 = 0$ and express the series as

$$a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx).$$

It is called a **cosine series** if all b_n vanish and **sine series** if all a_n vanish. Trigonometric series form an important class of series of functions. In Mathematical Analysis II, we studied the convergence of the series of functions. We recall

- Uniform convergence implies pointwise convergence of a series of functions,
- Absolute convergence implies pointwise convergence of a series of functions,
- Weierstrass M-Test for uniform and absolute convergence (see Appendix I).

For instance, using the fact that $|\cos nx|, |\sin nx| \leq 1$, Weierstrass M-Test tells us that a trigonometric series is uniformly and absolutely convergent when its coefficients satisfy $\sum_n |a_n|, \sum_n |b_n| < \infty$, and this is the case when $|a_n|, |b_n| \leq Cn^{-s}, \forall n \geq 1$, for some constant C and $s > 1$. Since the partial sums are continuous functions and uniform convergence preserves continuity, the infinite series

$$\phi(x) \equiv a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)$$

is a continuous function on $[-2\pi, 2\pi]$. ϕ is also of period 2π. For, by pointwise convergence, we have

$$\begin{aligned}
\phi(x + 2\pi) &= \lim_{n\to\infty} \sum_{k=0}^{n} \big(a_k \cos(kx + 2k\pi) + b_k \sin(kx + 2k\pi)\big) \\
&= \lim_{n\to\infty} \sum_{k=0}^{n} (a_k \cos kx + b_k \sin kx) \\
&= \phi(x),
\end{aligned}$$

hence it is 2π-periodic.

In the literature there are many interesting convergence results concerning trigonometric series. Nevertheless, we will not go into this direction. Here our attention is on a special class of trigonometric series called Fourier series. Each Fourier series is associated with an integrable, periodic function.

Given a 2π-periodic function which is Riemann integrable function f on $[-\pi, \pi]$, its

Fourier series or **Fourier expansion** is the trigonometric series given by

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(y)\cos ny\,dy, \quad n \geq 1 \\ b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(y)\sin ny\,dy, \quad n \geq 1 \qquad \text{and} \\ a_0 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)\,dy. \end{aligned} \tag{1.1}$$

Note that a_0 is the average of the function over the interval. From this definition we gather two basic information. First, the Fourier series of a function involves the integration of the function over an interval, hence any modification of the values of the function over a subinterval, not matter how small it is, may change the Fourier coefficients a_n and b_n. This is unlike power series which only depend on the local properties (derivatives of all order at a designated point). We may say Fourier series depend on the global information but power series only depend on local information. Second, recalling from the theory of Riemann integral, we know that two integrable functions which are equal almost everywhere have the same integral. (We will see the converse is also true, namely, two functions with the same Fourier series are equal almost everywhere.) In Appendix II we recall the concept of a measure zero set and some of its basic properties. Therefore, the Fourier series of two such functions are the same. In particular, the Fourier series of a function is completely determined with its value on the open interval $(-\pi, \pi)$, regardless its values at the endpoints.

The motivation of the Fourier series comes from the belief that for a "nice function" of period 2π, its Fourier series converges to the function itself. In other words, we have

$$f(x) = a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)\ , \quad \forall x \in \mathbb{R}. \tag{1.2}$$

When this holds, the coefficients a_n, b_n are given by (1.1). To see this, we multiply (1.2) by $\cos mx$ and then integrate over $[-\pi, \pi]$. Using the relations

$$\begin{aligned} \int_{-\pi}^{\pi} \cos nx \cos mx\,dx &= \begin{cases} \pi, & n = m \\ 0, & n \neq m \end{cases}, \\ \int_{-\pi}^{\pi} \cos nx \sin mx\,dx &= 0 \quad (n, m \geq 1), \qquad \text{and} \\ \int_{-\pi}^{\pi} \cos nx\,dx &= \begin{cases} 2\pi, & n = 0 \\ 0, & n \neq 0 \end{cases}, \end{aligned}$$

we *formally* arrive at the expression of a_n, $n \geq 0$, in (1.2). Similarly, by multiplying (1.2) by $\sin mx$ and then integrate over $[-\pi, \pi]$, one obtain the expression of b_n, $n \geq 1$, in (1.2) after using

$$\int_{-\pi}^{\pi} \sin nx \sin mx\,dx = \begin{cases} \pi, & n = m \\ 0, & n \neq m \end{cases}.$$

Of course, (1.2) arises from the hypothesis that every sufficiently nice function of period 2π is equal to its Fourier expansion. The study of under which "nice conditions" this could happen is one of the main objects in the theory of Fourier series.

We can associate a Fourier series for any integrable function on $[-2\pi, 2\pi]$. Indeed, it suffices to extend the function as a function of period 2π. The extension is straightforward; simply let $\tilde{f}(x) = f(x-(n+1)\pi)$ where n is the unique integer satisfying $n\pi < x \leq (n+2)\pi$. It is clear that $\tilde{f}$ is equal to f on $(-\pi, \pi]$. As the function is defined on $[-\pi, \pi]$, apparently an extension in strict sense is possible only if $f(-\pi) = f(\pi)$. Since the function value at one point does not change the Fourier series, from now on it will be understood that the extension of a function to a 2π-periodic function refers to the extension for the restriction of this function on $(-\pi, \pi]$. Note that for the 2π-periodic extension of a continuous function on $[-\pi, \pi]$ has a jump discontinuity at $\pm\pi$ when $f(\pi) \neq f(-\pi)$. It is is continuous on $\mathbb{R}$ if and only if $f(-\pi) = f(\pi)$. In the following we will not distinguish f with its extension $\tilde{f}$.

We will use

$$f(x) \sim a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

to denote the fact that the right hand side of this expression is the Fourier series of f.

Example 1.1 We consider the function $f_1(x) = x$. Its extension is a piecewise smooth function with jump discontinuities at $(2n+1)\pi, n \in \mathbb{Z}$. As f_1 is odd and $\cos nx$ is even,

$$\pi a_n = \int_{-\pi}^{\pi} x \cos nx \, dx = 0, \quad n \geq 0,$$

and

$$\begin{aligned} \pi b_n &= \int_{-\pi}^{\pi} x \sin nx \, dx \\ &= -x \, \frac{\cos nx}{n} \Big|_{-\pi}^{\pi} + \int_{-\pi}^{\pi} \frac{\cos nx}{n} \, dx \\ &= (-1)^{n+1} \frac{2\pi}{n}. \end{aligned}$$

Therefore,

$$f_1(x) \sim 2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin nx.$$

Since f_1 is an odd function, it is reasonable to see that no cosine functions are involved in its Fourier series. How about the convergence of this Fourier series? Although the terms decay like $O(1/n)$ as $n \to \infty$, its convergence is not clear at this moment. On the other hand, this Fourier series is equal to 0 at $x = \pm\pi$ but $f_1(\pm\pi) = \pi$. So, one thing is sure, namely, the Fourier series is not always equal to its function. It is worthwhile to observe

that the bad points $\pm\pi$ are precisely the discontinuity points of f_1.

Notation The big O and small $\circ$ notations are very convenient in analysis. We say a sequence $\{x_n\}$ satisfies $x_n = O(n^s)$ means that there exists a constant C independent of n such that $|x_n| \le Cn^s$ as $n \to \infty$, in other words, the growth (resp. decay $s \ge 0$) of $\{x_n\}$ is not faster (resp. slower $s < 0$) the s-th power of n. On the other hand, $x_n = \circ(n^s)$ means $|x_n|n^{-s} \to 0$ as $n \to \infty$.

Example 1.2 Next consider the function $f_2(x) = x^2$. Unlike the previous example, its 2π-periodic extension is continuous on $\mathbb{R}$. After performing integration by parts, the Fourier series of f_2 is seen to be

$$f_2(x) \equiv x^2 \sim \frac{\pi^2}{3} - 4\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} \cos nx.$$

As f_2 is an even function, this is a cosine series. The rate of decay of the Fourier series is like $O(1/n^2)$. Using Weierstrass M-test, this series converges uniformly to a continuous function. In fact, due to the following result, it converges uniformly to f_2. Note that f_2 is smooth on $(n\pi, (n+1)\pi), n \in \mathbb{Z}$.

Convergence Criterion. *The Fourier series of a continuous, 2π-periodic function which is C^1-piecewise on $[-\pi, \pi]$ converges to the function uniformly.*

A function is called C^1-**piecewise** on some interval $I = [a, b]$ if there exists a partition of I into subintervals $\{I_j\}_{j=1}^N$ and there are C^1-function f_j defined on I_j such that $f = f_j$ on each (a_j, a_{j+1}) where $I_j = [a_j, a_{j+1}]$. This convergence criterion is a special case of Theorem 1.7 in Section 3.

We list more examples of Fourier series of functions and leave them for you to verify.

(a) $f_3(x) \equiv |x| \sim \dfrac{\pi}{2} - \dfrac{4}{\pi}\displaystyle\sum_{n=1}^{\infty} \frac{1}{(2n-1)^2} \cos(2n-1)x,$

(b) $f_4(x) = \begin{cases} 1, & x \in [0, \pi] \\ -1, & x \in (-\pi, 0) \end{cases} \sim \dfrac{4}{\pi}\displaystyle\sum_{n=1}^{\infty} \frac{1}{2n-1} \sin(2n-1)x,$

(c) $f_5(x) = \begin{cases} x(\pi - x), & x \in [0, \pi) \\ x(\pi + x), & x \in (-\pi, 0) \end{cases} \sim \dfrac{8}{\pi}\displaystyle\sum_{n=1}^{\infty} \frac{1}{(2n-1)^3} \sin(2n-1)x.$

Let $\{c_n\}_{-\infty}^{\infty}$ be a bisequence of complex numbers. (A bisequence is a map from $\mathbb{Z}$ to $\mathbb{C}$.) A (complex) trigonometric series is the infinite series associated to the bisequence

$\{c_n e^{inx}\}_{-\infty}^{\infty}$ and is denoted by $\sum_{-\infty}^{\infty} c_n e^{inx}$. To be in line with the real case, it is said to be convergent at x if

$$\lim_{n\to\infty} \sum_{k=-n}^{n} c_n e^{nix}$$

exists. Now, a complex Fourier series can be associated to a complex-valued function. Let f be a 2π-periodic complex-valued function which is integrable on $[-\pi, \pi]$. Its Fourier series is given by the series

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx},$$

where the Fourier coefficients c_n are given by

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx, \quad n \in \mathbb{Z}.$$

Here for a complex function f, its integration over some $[a, b]$ is defined to be

$$\int_a^b f(x)dx = \int_a^b f_1(x)dx + i \int_a^b f_2(x)dx,$$

where f_1 and f_2 are respectively the real and imaginary parts of f. It is called integrable if both real and imaginary parts are integrable. The same as in the real case, formally the expression of c_n is obtained as in the real case by first multiplying the relation

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

with e^{imx} and then integrating over $[-\pi, \pi]$ with the help from the relation

$$\int_{-\pi}^{\pi} e^{imx} e^{-inx}\, dx = \begin{cases} 2\pi, & n = m \\ 0, & n \neq m \end{cases} .$$

When f is of real-valued, there are two Fourier series, that is, the real and the complex ones. To relate them it is enough to observe the Euler's formula $e^{i\theta} = \cos\theta + i \sin\theta$, so for $n \geq 1$

$$\begin{aligned} 2\pi c_n &= \int_{-\pi}^{\pi} f(x) e^{-nx} dx \\ &= \int_{-\pi}^{pi} f(x) \left(\cos nx + i \sin nx\right) dx \\ &= \int_{-\pi}^{pi} f(x) \cos nx dx - i \in_{-\pi}^{\pi} f(x) \sin nx dx \\ &= \pi(a_n - i b_n) \ . \end{aligned}$$

we see that

$$c_n = \frac{1}{2}(a_n - ib_n), \quad n \geq 1, \quad c_0 = a_0 \ .$$

By a similar computation, we have

$$c_n = \frac{1}{2}(a_{-n} + ib_{-n}), \quad n \leq -1 \ .$$

It follows that $c_{-n} = \overline{c_n}$ for all n. In fact, the converse is true, that is, a complex Fourier series is the Fourier series of a real-valued function if and only if $c_{-n} = \overline{c_n}$ holds for all n. Indeed, letting

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

be the Fourier series of f, it is straightforward to verify that

$$\overline{f(x)} \sim \sum_{n=-\infty}^{\infty} d_n e^{inx}, \quad d_n = \overline{c_{-n}} \ .$$

Hence when f is real-valued, $\overline{f} = f$ so $c_n = \overline{c_{-n}}$ holds. The complex form of Fourier series sometimes makes expressions and computations more elegant. We will use it whenever it makes things simpler.

We have been working on the Fourier series of 2π-periodic functions. For functions of $2T$-period, their Fourier series are not the same. They can be found by a scaling argument. Let f be $2T$-periodic. The function $g(x) = f(Tx/\pi)$ is a 2π-periodic function. Thus,

$$f\left(\frac{Tx}{\pi}\right) = g(x) \sim a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

where $a_0, a_n, b_n, n \geq 1$ are the Fourier coefficients of g. By a change of variables, we can express everything inside the coefficients in terms of f, $\cos n\pi x/T$ and $\sin n\pi x/T$. The result is

$$f(x) \sim a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi}{T}x + b_n \sin\frac{n\pi}{T}x\right),$$

where

$$\begin{aligned} a_n &= \frac{1}{T}\int_{-T}^{T} f(y)\cos\frac{n\pi}{T}y\,dy, \\ b_n &= \frac{1}{T}\int_{-T}^{T} f(y)\sin\frac{n\pi}{T}y\,dy, \quad n \geq 1, \qquad \text{and} \\ a_0 &= \frac{1}{2T}\int_{-T}^{T} f(y)\,dy. \end{aligned}$$

It reduces to (1.1) when T is equal to π.

1.2 Riemann-Lebesgue Lemma

From the examples of Fourier series of functions in the previous section we see that the coefficients decay to 0 eventually. We will show that this is generally true. This is the content of the following result.

Theorem 1.1 (**Riemann-Lebesgue Lemma**). *The Fourier coefficients of a 2π-periodic function integrable on $[-\pi, \pi]$ converge to 0 as $n \to \infty$.*

We point out this theorem still holds when $[-\pi, \pi]$ is replaced by any $[a, b]$. The proof is essentially the same.

We will use $R[-\pi, \pi]$ to denote the vector space of all integrable functions. To prepare for the proof we study how to approximate an integrable function by step functions. Let $a_0 = -\pi < a_1 < \cdots < a_N = \pi$ be a partition of $[-\pi, \pi]$. A **step function** s satisfies $s(x) = s_j$, $\forall x \in (a_j, a_{j+1}]$, $\forall j \geq 0$. The value of s at $-\pi$ is not important, but for definiteness let's set $s(-\pi) = s_0$. We can express a step function in a better form by introducing the characteristic function χ_E for a set $E \subset \mathbb{R}$:

$$\chi_E = \begin{cases} 1, & x \in E, \\ 0, & x \notin E. \end{cases}$$

Then,

$$s(x) = \sum_{j=0}^{N-1} s_j \chi_{I_j}, \quad I_j = (a_j, a_{j+1}], \quad j \geq 1, \ I_0 = [a_0, a_1].$$

Lemma 1.2. *For every step function s in $R[-\pi, \pi]$, there exists some constant C independent of n such that*

$$|a_n|, \ |b_n| \leq \frac{C}{n}, \quad \forall n \geq 1,$$

where a_n, b_n are the Fourier coefficients of s.

Proof. Let $s(x) = \sum_{j=0}^{N-1} s_j \chi_{I_j}$. We have

$$\begin{aligned} \pi a_n &= \int_{-\pi}^{\pi} \sum_{j=0}^{N-1} s_j \chi_{I_j} \cos nx \, dx \\ &= \sum_{j=0}^{N-1} s_j \int_{a_j}^{a_{j+1}} \cos nx \, dx \\ &= \frac{1}{n} \sum_{j=0}^{N-1} s_j (\sin n a_{j+1} - \sin n a_j). \end{aligned}$$

It follows that

$$|a_n| \le \frac{C}{n}, \quad \forall n \ge 1, \qquad C = \frac{2}{\pi}\sum_{j=0}^{N-1} |s_j|.$$

Clearly a similar estimate holds for b_n. □

Lemma 1.3. *Let $f \in R[-\pi, \pi]$. Given $\varepsilon > 0$, there exists a step function s such that $s \le f$ on $[-\pi, \pi]$ and*

$$\int_{-\pi}^{\pi} (f - s) < \varepsilon.$$

Proof. As f is integrable, it can be approximated from below by its Darboux lower sums. In other words, for $\varepsilon > 0$, we can find a partition $-\pi = a_0 < a_1 < \cdots < a_N = \pi$ such that

$$\left| \int_{-\pi}^{\pi} f - \sum_{j=0}^{N-1} m_j (a_{j+1} - a_j) \right| < \varepsilon,$$

where $m_j = \inf \{f(x) : x \in [a_j, a_{j+1}]\}$. It follows that

$$\left| \int_{-\pi}^{\pi} (f - s) \right| < \varepsilon$$

after setting

$$s(x) = \sum_{j=0}^{N-1} m_j \chi_{I_j}, \quad I_j = (a_j, a_{j+1}], \;\; j \ge 1, \; I_0 = [a_0, a_1].$$

□

Now we prove Theorem 1.1.

Proof. For $\varepsilon > 0$, we can find s as constructed in Lemma 1.3 such that $0 \le f - s$ and

$$\int_{-\pi}^{\pi} (f - s) < \frac{\varepsilon}{2}.$$

Let a'_n be the n-th Fourier coefficient of s. By Lemma 1.2,

$$|a'_n| < \frac{\varepsilon}{2},$$

for all $n \ge n_0 = [2C/\varepsilon] + 1$.

$$\begin{aligned} |\pi(a_n - a'_n)| &= \left| \int_{-\pi}^{\pi} (f - s) \cos nx \, dx \right| \\ &\le \int_{-\pi}^{\pi} |f - s| \\ &= \int_{-\pi}^{\pi} (f - s) \\ &< \frac{\varepsilon}{2}. \end{aligned}$$

It follows that for all $n \geq n_0$,

$$|a_n| \leq |a_n - a'_n| + |a'_n| < \frac{\varepsilon}{2\pi} + \frac{\varepsilon}{2} < \varepsilon.$$

The same argument applies to b_n too. □

It is useful to bring in a "mapping" point of view between functions and their Fourier series. Let $R_{2\pi}$ be the collection of all 2π-periodic complex-valued functions integrable on $[-\pi, \pi]$ and $\mathcal{C}$ consisting of all complex-valued bisequences $\{c_n\}$ satisfying $c_n \to 0$ as $n \to \pm\infty$. The Fourier series sets up a mapping Φ from $R_{2\pi}$ to $\mathcal{C}$ by sending f to $\{\hat{f}(n)\}$ where, to make things clear, we have let $\hat{f}(n) = c_n$, the n-th Fourier coefficient of f. When real-functions are considered, restricting to the subspace of $\mathcal{C}$ given by those satisfying $c_{-n} = \overline{c_n}$, Φ maps all real functions into this subspace. Perhaps the first question we ask is: Is Φ one-to-one? Clearly the answer is no, for two functions which differ on a set of measure zero have the same Fourier coefficients. However, we have the following result, to be proved in Section 5.

Uniqueness Theorem. *The Fourier series of two functions in $R_{2\pi}$ coincide if and only if they are equal almost everywhere.*

Thus Φ is essentially one-to-one. One can show that it is not onto. Despite of this, we may still study how various structures on $R_{2\pi}$ and $\mathcal{C}$ are associated under Φ. Observe that both $R_{2\pi}$ and $\mathcal{C}$ form vector spaces over $\mathbb{C}$. In fact, there are obvious and surprising ones. Some of them are listed below and more can be found in the exercise.

Property 1. Φ is a linear map. Observe that both $R_{2\pi}$ and $\mathcal{C}$ form vector spaces over $\mathbb{R}$ or $\mathbb{C}$. The linearity of Φ is clear from its definition.

Property 2. When $f \in R_{2\pi}$ is k-th differentiable and all derivatives up to k-th order belong to $R_{2\pi}$, $\hat{f^k}(n) = (in)^k \hat{f}(n)$ for all $n \in \mathbb{Z}$. See Proposition 1.4 below for a proof. This property shows that differentiation turns into the multiplication of a factor $(in)^k$ under Φ. This is amazing!

Property 3. Every translation in $\mathbb{R}$ induces a "translation operation" on functions defined on $\mathbb{R}$. More specifically, for $a \in \mathbb{R}$, set $f_a(x) = f(x + a)$, $x \in \mathbb{R}$. Clearly f_a belongs to $R_{2\pi}$. We have $\hat{f}_a(n) = e^{ina}\hat{f}(n)$. This property follows directly from the definition. It shows that a translation in $R_{2\pi}$ turns into the multiplication of a factor e^{ina} under Φ.

Proposition 1.4. *Let f be a 2π-periodic function which is differentiable on $[-\pi, \pi]$ with $f' \in R_{2\pi}$. If*

$$f'(x) \sim c_0 + \sum_{n=1}^{\infty} \big(c_n \cos nx + d_n \sin nx\big),$$

then $c_n = nb_n$, and $d_n = -na_n$. In complex notations, $\hat{f'}(n) = in\hat{f}(n)$.

Proof. We compute

$$\begin{aligned}\pi c_n &= \int_{-\pi}^{\pi} f'(y)\cos ny\,dy \\ &= f(y)\cos ny|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f(y)(-n\sin ny)\,dy \\ &= n\int_{-\pi}^{\pi} f(y)\sin ny\,dy \\ &= \pi n a_n.\end{aligned}$$

Similarly,

$$\begin{aligned}\pi d_n &= \int_{-\pi}^{\pi} f'(y)\sin ny\,dy \\ &= f(y)\sin ny|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f(y)n\cos ny\,dy \\ &= -n\int_{-\pi}^{\pi} f(y)\cos ny\,dy \\ &= -\pi n a_n.\end{aligned}$$

□

Property 2 links the regularity of the function to the rate of decay of its Fourier coefficients. This is an extremely important property. When f is a 2π-periodic function whose derivatives up to k-th order belong to $R_{2\pi}$, applying Riemann-Lebesgue lemma to $f^{(k)}$ we know that $\hat{f^{(k)}}(n) = \circ(1)$ as $n \to \infty$. By Property 2 it follows that $\hat{f}(n) = \circ(n^{-k})$, that is, the Fourier coefficients of f decay faster that n^{-k}. Since $\sum_{n=1}^{\infty} n^{-2} < \infty$, an application of Weierstrass M-test establishes the following result: *The Fourier series of f converges uniformly provided f, f' and f'' belong to $R_{2\pi}$.* Although the Fourier series converges uniformly, at this point we cannot conclude that the limit function is f. This fact will be proved in the next section.

1.3 Convergence of Fourier Series

In this section we study the convergence of the Fourier series of a function to itself. Recall that the series $a_0 + \sum_{n=1}^{\infty}(a_n\cos nx + b_n\sin nx)$, or $\sum_{n=-\infty}^{\infty} c_n e^{inx}$, where a_n, b_n, c_n are the Fourier coefficients of a function f converges to f at x means that the n-th partial sum

$$(S_n f)(x) = a_0 + \sum_{k=1}^{n}(a_k\cos kx + b_k\sin kx)$$

or

$$(S_n f)(x) = \sum_{k=-n}^{n} c_k e^{ikx}$$

converges to $f(x)$ as $n \to \infty$.

We start by expressing the partial sums in closed form. Indeed,

$$\begin{aligned}
(S_n f)(x) &= a_0 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f + \sum_{k=1}^{n} \frac{1}{\pi}\int_{-\pi}^{\pi} f(y)(\cos ky \cos kx + \sin ky \sin ky)\ dy \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi}\Big(\frac{1}{2} + \sum_{k=1}^{n}\cos k(y-x)\Big) f(y)\ dy \\
&= \frac{1}{\pi}\int_{x-\pi}^{x+\pi}\Big(\frac{1}{2} + \sum_{k=1}^{n}\cos kz\Big) f(x+z)\ dz \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi}\Big(\frac{1}{2} + \sum_{k=1}^{n}\cos kz\Big) f(x+z)\ dz\ ,
\end{aligned}$$

where in the last step we have used the fact that the integrals over any two periods are the same. Using the elementary formula

$$\cos\theta + \cos 2\theta + \cdots + \cos n\theta = \frac{\sin\left(n+\frac{1}{2}\right)\theta - \sin\frac{1}{2}\theta}{2\sin\frac{\theta}{2}}, \quad \theta \neq 0,$$

we obtain

$$(S_n f)(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{\sin\left(n+\frac{1}{2}\right)z}{2\sin\frac{1}{2}z} f(x+z)\, dz\ .$$

We express it in the form

$$(S_n f)(x) = \int_{-\pi}^{\pi} D_n(z) f(x+z)\, dy,$$

where D_n is the **Dirichlet kernel**

$$D_n(z) = \begin{cases} \dfrac{\sin\left(n+\frac{1}{2}\right)z}{2\pi\sin\frac{1}{2}z}, & z \neq 0 \\ \dfrac{2n+1}{2\pi}, & z = 0. \end{cases}$$

Using $\sin\theta/\theta \to 1$ as $\theta \to 0$, we see that D_n is continuous on $[-\pi, \pi]$.

Taking $f \equiv 1$, $S_n f = 1$ for all n. Hence

$$1 = \int_{-\pi}^{\pi} D_n(z)\, dz.$$

Using it we can write

$$(S_n f)(x) - f(x) = \int_{-\pi}^{\pi} D_n(z)(f(x+z) - f(x))\, dz. \tag{1.3}$$

In order to show $S_n f(x) \to f(x)$, it suffices to show the right hand side of (1.3) tends to 0 as $n \to \infty$.

Thus, the Dirichlet kernel plays a crucial role in the study of the convergence of Fourier series. We list some of its properties as follows.

Property I. $D_n(z)$ is an even, continuous, 2π-periodic function vanishing at $z = 2k\pi/(2n+1), -n \le k \le n$, on $[-\pi, \pi]$.

Property II. D_n attains its maximum value $(2n+1)/2$ at 0.

Property III.

$$\int_{-\pi}^{\pi} D_n(z) dz = 1$$

.

Property IV. For every $\delta > 0$,

$$\int_0^{\delta} |D_n(z)| dz \to \infty, \quad \text{as } n \to \infty.$$

Only the last property needs a proof. Indeed, for each n we can fix an N such that $\pi N \le (2n+1)\delta/2 \le (N+1)\pi$, so $N \to \infty$ as $n \to \infty$. We compute

$$\begin{aligned}
\int_0^{\delta} |D_n(z)| dz &= \int_0^{\delta} \frac{|\sin(n+\frac{1}{2})z|}{2\pi |\sin \frac{z}{2}|} dz \\
&\ge \frac{1}{\pi} \int_0^{(n+\frac{1}{2})\delta} \frac{|\sin t|}{t} dt \\
&\ge \frac{1}{\pi} \int_0^{N\pi} \frac{|\sin t|}{t} dt \\
&= \frac{1}{\pi} \sum_{k=1}^{N} \int_{(k-1)\pi}^{k\pi} \frac{|\sin t|}{t} dt \\
&= \frac{1}{\pi} \sum_{k=1}^{N} \int_0^{\pi} \frac{|\sin s|}{s + (k-1)\pi} ds \\
&\ge \frac{1}{\pi} \sum_{k=1}^{N} \int_0^{\pi} \frac{|\sin s|}{\pi k} ds \\
&= c_0 \sum_{k=1}^{N} \frac{1}{k}, \quad c_0 = \frac{1}{\pi^2} \int_0^{\pi} |\sin s| ds > 0, \\
&\to \infty,
\end{aligned}$$

as $N \to \infty$.

To elucidate the effect of the kernel, we fix a small $\delta > 0$ and split the integral into two parts:

$$\int_{-\pi}^{\pi} \chi_A(z) D_n(z)(f(x+z) - f(x))\, dz,$$

and

$$\int_{-\pi}^{\pi} \chi_B(z) D_n(z)(f(x+z) - f(x))\, dz,$$

where $A = (-\delta, \delta)$ and $B = [-\pi, \pi] \setminus A$. The second integral can be written as

$$\int_{-\pi}^{\pi} \frac{\chi_B(z)(f(x+z) - f(x))}{2\pi \sin \frac{z}{2}} \sin\big(n + \frac{1}{2}\big) z\, dz.$$

As $|\sin z/2|$ has a positive lower bound on B, the function

$$\frac{\chi_B(z)(f(x+z) - f(x))}{2\pi \sin \frac{z}{2}}$$

belongs to $R[-\pi, \pi]$ and the second integral tends to 0 as $n \to \infty$ in view of Riemann-Lebesgue lemma. The trouble lies on the first integral. It can be estimated by

$$\int_{-\delta}^{\delta} |D_n(z)||f(x+z) - f(x)| dz.$$

In view of Property IV, No matter how small δ is, this term may go to ∞ so it is not clear how to estimate this integral.

The difficulty can be resolved by imposing a further regularity assumption on the function. First a definition. For a function f defined on $[a, b]$ is called **Lipschitz continuous** at $x \in [a, b]$ if there exist L and δ such that

$$|f(y) - f(x)| \le L\,|y - x|\,, \quad \forall y \in [a, b],\ |y - x| \le \delta. \tag{1.4}$$

Here both L and δ depend on x. We point out that if $f \in C[a, b]$ is Lipschitz continuous at x, there exists some L' such that

$$|f(y) - f(x)| \le L'|y - x|, \quad \forall y \in [a, b].$$

In fact, this comes from (1.4) if $|y - x| \le \delta$. For y satisfying $|y - x| > \delta$, we have

$$|f(y) - f(x)| \le \frac{|f(y)| + f(x)|}{\delta} |y - x|,$$

hence we could take

$$L' = \max\left\{L, \frac{2M}{\delta}\right\}\ ,$$

where $M = \sup\{|f(y)| :\ y \in [a, b]\}$.

Theorem 1.5. *Let f be a 2π-periodic function integrable on $[-\pi,\pi]$. Suppose that f is Lipschitz continuous at x. Then $\{S_nf(x)\}$ converges to $f(x)$ as $n\to\infty$.*

Proof. Let Φ_δ be a cut-off function satisfying (a) $\Phi_\delta\in C(\mathbb{R})$, $\Phi_\delta\equiv 0$ outside $(-\delta,\delta)$, (b) $\Phi_\delta\geq 0$ and (c) $\Phi_\delta=1$ on $(-\delta/2,\delta/2)$. We write

$$\begin{aligned}(S_nf)(x)-f(x) &= \int_{-\pi}^{\pi} D_n(z)(f(x+z)-f(x))\,dz\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sin(n+\frac12)z}{\sin\frac{z}{2}}(f(x+z)-f(x))\,dz\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_\delta(z)\frac{\sin(n+\frac12)z}{\sin\frac{z}{2}}(f(x+z)-f(x))\,dz\\ &\quad+\frac{1}{\pi}\int_{-\pi}^{\pi}(1-\Phi_\delta(z))\frac{\sin(n+\frac12)z}{\sin\frac{z}{2}}(f(x+z)-f(x))\,dz\\ &\equiv I+II\ .\end{aligned}$$

By our assumption on f, there exists $\delta_0>0$ such that

$$|f(x+z)-f(x)|\leq L\,|z|\,,\quad \forall\,|z|<\delta_0.$$

Using $\sin\theta/\theta\to 1$ as $\theta\to 0$, there exists δ_1 such that $2|\sin z/2|\geq |z/2|$ for all $z,|z|<\delta_1$. For $z,|z|<\delta\equiv\min\{\delta_0,\delta_1\}$, we have $|f(x+z)-f(x)|/|\sin z/2|\leq 4L$ and

$$\begin{aligned}|I| &\leq \frac{1}{2\pi}\int_{-\delta}^{\delta}\Phi_\delta(z)\frac{\left|\sin(n+\frac12)z\right|}{\left|\sin\frac{z}{2}\right|}\,|f(x+z)-f(x)|\,dz\\ &\leq \frac{1}{2\pi}\int_{-\delta}^{\delta}4L\,dz\\ &= \frac{4\delta L}{\pi}.\end{aligned}\tag{1.5}$$

For $\varepsilon>0$, we fix δ so that

$$\frac{4\delta L}{\pi}<\frac{\varepsilon}{2}.\tag{1.6}$$

After fixing δ, we turn to the second integral

$$\begin{aligned}II &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(1-\Phi_\delta(z))(f(x+z)-f(x))}{\sin\frac{z}{2}}\sin\Big(n+\frac12\Big)z\,dz\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(1-\Phi_\delta(z))(f(x+z)-f(x))}{\sin\frac{z}{2}}\Big(\cos\frac{z}{2}\sin nz+\sin\frac{z}{2}\cos nz\Big)\,dz\\ &\equiv \int_{-\pi}^{\pi}F_1(x,z)\sin nz\,dz+\int_{-\pi}^{\pi}F_2(x,z)\cos nz\,dz.\end{aligned}$$

As $1-\Phi_\delta(z)=0$, for $z\in(-\delta/2,\delta/2)$, $|\sin z/2|$ has a positive lower bound on $(-\pi,-\delta/2)\cup(\delta/2,\pi)$, and so F_1 and F_2 are integrable on $[-\pi,\pi]$. By Riemann-Lebesgue lemma, for $\varepsilon>0$, there is some n_0 such that

$$\left|\int_{-\pi}^{\pi} F_1\sin nz\,dz\right|,\ \left|\int_{-\pi}^{\pi} F_2\cos nz\,dz\right|<\frac{\varepsilon}{4},\quad \forall n\ge n_0. \tag{1.7}$$

Putting (1.5), (1.6) and (1.7) together,

$$|S_nf(x)-f(x)|<\frac{\varepsilon}{2}+\frac{\varepsilon}{4}+\frac{\varepsilon}{4}=\varepsilon,\quad \forall n\ge n_0.$$

We have shown that $S_nf(x)$ tends to $f(x)$ when f is Lipschitz continuous at x. □

We leave some remarks concerning this proof. First, the cut-off function Φ_δ can be replaced by $\chi_{[-\delta,\delta]}$ without affecting the proof. Second, the regularity condition Lipschitz continuity is used to kill off the growth of the kernel at x. Third, this method used in this proof is a standard one. It will appear in many other places. For instance, a careful examination of it reveals a convergence result for functions with jump discontinuity after using the evenness of the Dirichlet kernel.

Theorem 1.6. *Let f be a 2π-periodic function integrable on $[-\pi,\pi]$. Suppose at some $x\in[-\pi,\pi]$, $\lim_{y\to x_+} f(y)$ and $\lim_{y\to x_-} f(y)$ exist and there are $\delta>0$ and constant L such that*

$$|f(y)-f(x^+)|\le L(y-x),\quad \forall y,\quad 0<y-x<\delta,$$

and

$$|f(y)-f(x^-)|\le L(x-y),\quad \forall y,\quad 0<x-y<\delta.$$

Then $\{S_nf(x)\}$ converges to $(f(x^+)+f(x^-))/2$ as $n\to\infty$.

Here $f(x^+)$ and $f(x^-)$ stand for $\lim_{y\to x^+} f(y)$ and $\lim_{y\to x^-} f(y)$ respectively. We leave the proof of this theorem as an exercise.

A function f defined on $[a,b]$ is called to satisfy a **Lipschitz condition** if there exists an L such that

$$|f(x)-f(y)|\le L|x-y|\ ,\quad \forall x,y\in[a,b]\ .$$

When f satisfies a Lipschitz condition, it is Lipschitz continuous everywhere. The Lipschitz condition is some kind of "uniformly Lipschitz" condition. Every continuously differentiable function on $[a,b]$ satisfies a Lipschitz condition. In fact, by the fundamental theorem of calculus, for $x,y\in[a,b]$,

$$\begin{aligned}|f(y)-f(x)| &= \Big|\int_x^y f'(t)dt\Big|\\ &\le M|y-x|,\end{aligned}$$

where $M=\sup\{|f'(t)|:\ t\in[a,b]\}$. Similarly, every piecewise C^1-function satisfies a Lipschitz condition.

Now, we have a theorem on the uniform convergence of the Fourier series of a function to the function itself.

Theorem 1.7. *Let f a 2π-periodic function satisfying a Lipschitz condition. Its Fourier series converges to f uniformly as $n \to \infty$.*

Proof. Observe that when f is Lipschitz continuous on $[-\pi, \pi]$, δ_0 and δ_1 can be chosen independent of x and (1.5), (1.6) hold uniformly in x. In fact, δ_0 only depends on L, the constant appearing in the Lipschitz condition. Thus the theorem follows if n_0 in (1.7) can be chosen uniformly in x. This is the content of the lemma below. We apply it by taking $f(x, y)$ to be $F_1(x, z)$ or $F_2(x, z)$.

□

Lemma 1.8. *Let $f(x, y)$ be periodic in y and $f \in C([-\pi, \pi] \times [-\pi, \pi])$. For any fixed x,*

$$c(n, x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x, y) e^{-iny}\, dy \to 0$$

uniformly in x as $n \to \infty$.

Proof. We need to show that for every $\varepsilon > 0$, there exists some n_0 independent of x such that

$$|c(n, x)| < \varepsilon, \quad \forall n \geq n_0.$$

Observe that

$$\begin{aligned} 2\pi c(n, x) &= \int_{-\pi}^{\pi} f(x, y) e^{-iny}\, dy \\ &= \int_{-\pi-\frac{\pi}{n}}^{\pi-\frac{\pi}{n}} f\left(x, z + \frac{\pi}{n}\right) e^{-in(z+\frac{\pi}{n})}\, dz \quad y = z + \frac{\pi}{n}, \\ &= -\int_{-\pi}^{\pi} f\left(x, z + \frac{\pi}{n}\right) e^{-inz}\, dz \quad (f \text{ is } 2\pi\text{-periodic}). \end{aligned}$$

We have

$$c(n, x) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(f(x, y) - f\left(x, y + \frac{\pi}{n}\right) \right) e^{-iny}\, dy.$$

As $f \in C([-\pi, \pi] \times [-\pi, \pi])$, it is uniformly continuous in $[-\pi, \pi] \times [-\pi, \pi]$. For $\varepsilon > 0$, there exists a δ such that

$$|f(x, y) - f(x', y')| < \varepsilon \quad \text{if } |x - x'|,\ |y - y'| < \delta.$$

We take n_0 so large that $\pi/n_0 < \delta$. Then, using $|e^{-iny}| = 1$,

$$\begin{aligned} |c(n, x)| &\leq \frac{1}{4\pi} \int_{-\pi}^{\pi} \left| f(x, y) - f\left(x, y + \frac{\pi}{n}\right) \right| dy \\ &\leq \frac{\varepsilon}{4\pi} \int_{-\pi}^{\pi} dy = \frac{\varepsilon}{2} \\ &< \varepsilon, \qquad \forall n \geq n_0. \end{aligned}$$

□

Example 1.3. We return to the functions discussed in Examples 1.1 and 1.2. Indeed, f_1 is smooth except at $n\pi$. According to Theorem 1.5, the series

$$2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin nx$$

converges to x for every $x \in (-\pi, \pi)$. On the other hand, we observed before that the series tend to 0 at $x = \pm\pi$. As $f_1(\pi_+) = -\pi$ and $f(\pi_-) = \pi$, we have $f_1(\pi_+) + f(\pi_-) = 0$, which is in consistency with Theorem 1.5. In the second example, f_2 is continuous, 2π-periodic. By Theorem 1.7, its Fourier series

$$\frac{\pi^2}{3} - 4\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2}\cos nx$$

converges to x^2 uniformly on $[-\pi, \pi]$.

So far we have been working on the Fourier series of 2π-periodic functions. It is clear that the same results apply to the Fourier series of $2T$-periodic functions for arbitrary positive T.

We have shown the convergence of the Fourier series under some additional regularity assumptions on the function. But the basic question remains, that is, is the Fourier series of a continuous, 2π-periodic function converges to itself? It turns out the answer is negative. A not-so-explicit example can be found in Stein-Shakarchi and an explicit but complicated one was given by Fejér (see Zygmund "Trigonometric Series"). You may google for more. In fact, using the uniform boundedness principle in functional analysis, one can even show that "most" continuous functions have divergent Fourier series. The situation is very much like in the case of the real number system where transcendental numbers are uncountable while algebraic numbers are countable despite the fact that it is difficult to establish a specific number is transcendental.

We present another convergence result where is concerned with pointwise convergence. It replaces regularity by monotonicity in the function under consideration. Theorem 1.9 and Proposition 1.10 are for optional reading.

Theorem 1.9. *Let f be a 2π-periodic function integrable on $[-\pi, \pi]$. Suppose that it is piecewise continuous and increasing near some point x. Its Fourier series converges to $\big(f(x^+) + f(x^-)\big)/2$ at x.*

Proof. In the following proof we will take $x = 0$ for simplicity. We first write, using the

evenness of D_n,

$$\begin{aligned}(S_nf)(0) &= \int_{-\pi}^{\pi} D_n(z)f(z)\,dz \\ &= \int_0^{\pi} (f(z)+f(-z))D_n(z)\,dz.\end{aligned}$$

So,

$$(S_nf)(0) - \frac{1}{2}(f(0^+)+f(0^-)) = \int_0^{\pi} (f(z)-f(0^+)+f(-z)-f(0^-))D_n(z)\,dz.$$

We will show that

$$\int_0^{\pi} (f(z)-f(0^+))D_n(z)\,dz \to 0 \tag{1.8}$$

and

$$\int_0^{\pi} (f(-z)-f(0^-))D_n(z)\,dz \to 0 \tag{1.9}$$

as $n\to\infty$. Indeed, for a small $h>0$, we consider

$$\begin{aligned}\int_0^h (f(z)-f(0^+))\frac{\sin\frac{2n+1}{2}z}{\sin\frac{z}{2}}\,dz &= \int_0^h (f(z)-f(0^+))\frac{\sin\frac{2n+1}{2}z}{\frac{z}{2}}\,dz \\ &\quad + \int_0^h (f(z)-f(0^+))\left(\frac{\sin\frac{2n+1}{2}z}{\sin\frac{z}{2}} - \frac{\sin\frac{2n+1}{2}z}{\frac{z}{2}}\right)dz.\end{aligned}$$

Using L'Hospital's rule,

$$\frac{1}{\sin\frac{z}{2}} - \frac{1}{\frac{z}{2}} = \frac{z-2\sin\frac{z}{2}}{z\sin\frac{z}{2}} \to 0 \quad \text{as } z\to 0.$$

Therefore, for $\varepsilon>0$, we can find h_1 such that

$$\begin{aligned}&\int_0^{h_1} \left|f(z)-f(0^+)\right| \left|\frac{1}{\sin\frac{z}{2}} - \frac{1}{\frac{z}{2}}\right| \left|\sin\frac{2n+1}{2}z\right| dz \\ &\le \int_0^{h_1} \left|f(z)-f(0^+)\right| \left|\frac{1}{\sin\frac{z}{2}} - \frac{1}{\frac{z}{2}}\right| dz < \frac{\varepsilon}{3},\end{aligned}$$

where h_1 is independent of n. Next, by the second mean-value theorem for integral (see below),

$$\int_0^h (f(z)-f(0^+))\frac{\sin\frac{2n+1}{2}z}{\frac{z}{2}}\,dz = (f(h)-f(0^+))\int_k^h \frac{\sin\frac{2n+1}{2}z}{\frac{z}{2}}\,dz$$

for some $k\in(0,h)$. As

$$\begin{aligned}\left|\int_k^h \frac{\sin\frac{2n+1}{2}z}{\frac{z}{2}}\,dz\right| &= \left|2\int_{\ell k}^{\ell h}\frac{\sin t}{t}\,dt\right|, \quad \ell = \frac{2n+1}{2} \\ &\le 2\left|\int_0^{\ell h}\frac{\sin t}{t}\,dt\right| + 2\left|\int_0^{\ell k}\frac{\sin t}{t}\,dt\right| \\ &\le 4\sup_T\left|\int_0^T \frac{\sin t}{t}\,dt\right| \equiv 4L,\end{aligned}$$

and we can find $h_2 \le h_1$ such that

$$4L\left|f(h) - f(0^+)\right| < \frac{\varepsilon}{3}, \quad \forall 0 < h \le h_2,$$

we have

$$\left|\int_0^{h_2} (f(z) - f(0^+)) \frac{\sin \frac{2n+1}{2} z}{\sin \frac{z}{2}}\, dz\right| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \frac{2\varepsilon}{3}.$$

Now, by Riemann-Lebesgue lemma, there exists some n_0 such that

$$\left|\int_{h_2}^{\pi} (f(z) - f(0^+)) \frac{\sin \frac{2n+1}{2} z}{\sin \frac{z}{2}}\, dz\right| < \frac{\varepsilon}{3}, \quad \forall n \ge n_0.$$

Putting things together,

$$\left|\int_0^{\pi} (f(z) - f(0^+)) \frac{\sin \frac{2n+1}{2} z}{\sin \frac{z}{2}}\, dz\right| < \frac{2\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon, \quad \forall n \ge n_0.$$

We have shown that (1.8) holds. To prove (1.9), it suffices to apply (1.8) to the function $g(z) = f(-z)$.

□

Proposition 1.10 (**Second Mean-Value Theorem**). *Let $f \in R[a,b]$ and g be monotone on $[a,b]$ and satisfy $g(a) = 0$. There exists some $c \in (a,b)$ such that*

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

Proof. Without loss of generality, we assume g is increasing. Let

$$a = x_0 < x_1 < \cdots < x_n = b$$

be a partition P on $[a,b]$.

$$\int_a^b fg = \sum_{j=1}^n g(x_j) \int_{x_{j-1}}^{x_j} f + \sum_{j=1}^n \int_{x_{j-1}}^{x_j} f(x)(g(x) - g(x_j))\, dy.$$

In case $\|P\| \to 0$, it is not hard to show that the second integral tends to zero, so

$$\int_a^b fg = \lim_{\|P\| \to 0} \sum_{j=1}^n g(x_j) \int_{x_{j-1}}^{x_j} f.$$

Letting $F(x) = \int_x^b f$ and using $F(x_n) = F(b) = 0$, we have

$$\begin{aligned}\sum_{j=1}^n g(x_j) \int_{x_{j-1}}^{x_j} f &= \sum_{j=1}^n g(x_j)(F(x_j) - F(x_{j-1})) \\ &= g(x_1)F(x_0) + \sum_{j=1}^{n-1} (g(x_{j+1}) - g(x_j))F(x_j).\end{aligned}$$

Let $m = \inf_{[a,b]} F$ and $M = \sup_{[a,b]} F$. As g is increasing,

$$mg(b) \leq g(x_1)F(x_0) + \sum_{j=1}^{n-1}(g(x_{j+1}) - g(x_j))F(x_j) \leq Mg(b).$$

Letting $\|P\| \to 0$, we conclude that

$$mg(b) \leq \int_a^b fg \leq Mg(b).$$

As $c \mapsto \int_c^b f$ is continuous and bounded between m and M, there is some c such that

$$\frac{1}{g(b)} \int_a^b fg = \int_c^b f.$$

□

1.4 Weierstrass Approximation Theorem

As an application of Theorem 1.7, we prove a theorem of Weierstrass concerning the approximation of continuous functions by polynomials. First we consider how to approximate a continuous function by continuous, piecewise linear functions. A continuous function defined on $[a, b]$ is **piecewise linear** if there exists a partition $a = a_0 < a_1 < \cdots < a_n = b$ such that f is linear on each subinterval $[a_j, a_{j+1}]$.

Proposition 1.11. *Let f be a continuous function on $[a, b]$. For every $\varepsilon > 0$, there exists a continuous, piecewise linear function g such that $\|f - g\|_\infty < \varepsilon$.*

Recall that $\|f - g\|_\infty = \sup\{|f(x) - g(x)| : \ x \in [a, b]\}$.

Proof. As f is uniformly continuous on $[a, b]$, for every $\varepsilon > 0$, there exists some δ such that $|f(x) - f(y)| < \varepsilon/2$ for $x, y \in [a, b], |x - y| < \delta$. We partition $[a, b]$ into subintervals $I_j = [a_j, a_{j+1}]$ whose length is less than δ and define g to be the piecewise linear function satisfying $g(a_j) = f(a_j)$ for all j. For $x \in [a_j, a_{j+1}]$, g is given by

$$g(x) = \frac{f(a_{j+1}) - f(a_j)}{a_{j+1} - a_j}(x - a_j) + f(a_j).$$

We have

$$\begin{aligned} |f(x) - g(x)| &= |f(x) - \frac{f(a_{j+1}) - f(a_j)}{a_{j+1} - a_j}(x - a_j) + f(a_j)| \\ &\leq |f(x) - f(a_j)| + |\frac{f(a_{j+1}) - f(a_j)}{a_{j+1} - a_j}(x - a_j)| \\ &\leq |f(x) - f(a_j)| + |f(a_{j+1}) - f(a_j)| \\ &< \varepsilon, \end{aligned}$$

the result follows. □

Next we study how to approximate a continuous function by trigonometric polynomials (or, equivalently, finite Fourier series).

Proposition 1.12. *Let f be a continuous function on $[0,\pi]$. For $\varepsilon > 0$, there exists a trigonometric polynomial h such that $\|f-h\|_\infty < \varepsilon$.*

Proof. First we extend f to $[-\pi,\pi]$ by setting $f(x) = f(-x)$ (using the same notation) to obtain a continuous function on $[-\pi,\pi]$ with $f(-\pi) = f(\pi)$. By the previous proposition, we can find a continuous, piecewise linear function g such that $\|f-g\|_\infty < \varepsilon/2$. Since $g(-\pi) = f(-\pi) = f(\pi) = g(\pi)$, g can be extended as a Lipschitz continuous, 2π-periodic function. By Theorem 1.7, there exists some N such that $\|g - S_N g\|_\infty < \varepsilon/2$. Therefore, $\|f - S_N g\|_\infty \le \|f-g\|_\infty + \|g - S_N g\|_\infty < \varepsilon/2 + \varepsilon/2 = \varepsilon$. The proposition follows after noting that every finite Fourier series is a trigonometric polynomial (see Exercise 1). □

Theorem 1.13 (Weierstrass Approximation Theorem). *Let $f \in C[a,b]$. Given $\varepsilon > 0$, there exists a polynomial p such that $\|f-p\|_\infty < \varepsilon$.*

Proof. Consider $[a,b] = [0,\pi]$ first. Extend f to $[-\pi,\pi]$ as before and, for $\varepsilon > 0$, fix a trigonometric polynomial h such that $\|f-h\|_\infty < \varepsilon/2$. This is possible due to the previous proposition. Now, we express h as a finite Fourier series $a_0 + \sum_{n=1}^{N}(a_n \cos nx + b_n \sin nx)$. Using the fact that

$$\cos\theta = \sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n}}{(2n)!}, \quad \text{and } \sin\theta = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}\theta^{2n-1}}{(2n-1)!},$$

where the convergence is uniform on $[-\pi,\pi]$, each $\cos nx$ and $\sin nx, n = 1,\cdots,N$, can be approximated by polynomials. Putting all these polynomials together we obtain a polynomial $p(x)$ satisfying $\|h-p\|_\infty < \varepsilon/2$. It follows that $\|f-p\|_\infty \le \|f-h\|_\infty + \|h-p\|_\infty < \varepsilon/2 + \varepsilon/2 = \varepsilon$.

When f is continuous on $[a,b]$, the function $\varphi(t) = f(\frac{b-a}{\pi}t + a)$ is continuous on $[0,\pi]$. From the last paragraph, we can find a polynomial $p(t)$ such that $\|\varphi - p\|_\infty < \varepsilon$ on $[0,\pi]$. But then the polynomial $q(x) = p(\frac{\pi}{b-a}(x-a))$ satisfies $\|f-q\|_\infty = \|\varphi - p\|_\infty < \varepsilon$ on $[a,b]$. □

1.5 Mean Convergence of Fourier Series

In Section 2 we studied the uniform convergence of Fourier series. Since the limit of a uniformly convergent series of continuous functions is again continuous, we do not expect results like Theorem 1.6 applies to functions with jumps. In this section we will measure the distance between functions by a norm weaker than the uniform norm. Under the new

L^2-distance, you will see that every integrable function is equal to its Fourier expansion almost everywhere.

Recall that there is an inner product defined on the n-dimensional Euclidean space called the Euclidean metric

$$\langle x, y\rangle_2 = \sum_{j=1}^{n} x_j y_j, \qquad x, y \in \mathbb{R}^n.$$

With this inner product, one can define the concept of orthogonality and angle between two vectors. Likewise, we can also introduce a similar product on the space of integrable functions. Specifically, for $f, g \in R[-\pi, \pi]$, the **L^2-product** is given by

$$\langle f, g\rangle_2 = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

The L^2-product behaves like the Euclidean metric on $\mathbb{R}^n$ except at one aspect, namely, the condition $\langle f, f\rangle_2 = 0$ does not imply $f \equiv 0$. This is easy to see. In fact, when f is equal to zero except at finitely many points, then $\langle f, f\rangle_2 = 0$. From Appendix II $\langle f, f\rangle_2 = 0$ if and only if f is equal to zero except on a set of measure zero. This minor difference with the Euclidean inner product will not affect our discussion much. Parallel to the Euclidean case, we define the L^2-norm of an integrable function f to be

$$\|f\|_2 = \sqrt{\langle f, f\rangle_2},$$

and the **L^2-distance** between two integrable functions f and g by $\|f - g\|_2$. (When f, g are complex-valued, one should define the L^2-product to be

$$\langle f, g\rangle_2 = \int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx\ ,$$

so that $\langle f, f\rangle_2 \geq 0$. We will be restricted to real functions in this section.) One can verify that the triangle inequality

$$\|f + g\|_2 \leq \|f\|_2 + \|g\|_2$$

holds. We can also talk about $f_n \to f$ in L^2-sense, i.e., $\|f_n - f\|_2 \to 0$, or equivalently,

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} |f - f_n|^2 = 0, \qquad \text{as } n \to \infty\ .$$

This is a convergence in an average sense. It is not hard to see that when $\{f_n\}$ tends to f uniformly, $\{f_n\}$ must tend to f in L^2-sense. A moment's reflection will show that the converse is not always true. Hence convergence in L^2-sense is weaker than uniform convergence. We will discuss various metrics and norms in Chapter 2.

Our aim in this section is to show that the Fourier series of every integrable function converges to the function in the L^2-sense.

Just like the canonical basis $\{e_1, \ldots, e_n\}$ in $\mathbb{R}^n$, the functions

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos nx, \frac{1}{\sqrt{\pi}} \sin nx \right\}_{n=1}^{\infty}$$

forms an "orthonormal basis" in $R[-\pi, \pi]$, see Section 1.1. In the following we denote by

$$E_n = \left\langle \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos nx, \frac{1}{\sqrt{\pi}} \sin nx \right\rangle_{j=1}^{n}$$

the $(2n+1)$-dimensional vector space spanned by the first $2n+1$ trigonometric functions.

We start with a general result. Let $\{\phi_n\}_{n=1}^{\infty}$ be an orthonormal set (or orthonormal family) in $R[-\pi, \pi]$, i.e.,

$$\int_{-\pi}^{\pi} \phi_n \phi_m = \delta_{nm}, \quad \forall n, m \geq 1.$$

Let

$$\mathcal{S}_n = \langle \phi_1, \ldots, \phi_n \rangle$$

be the n-dimensional subspace spanned by $\phi_1, \ldots, \phi_n$. For a general $f \in R[-\pi, \pi]$, we consider the minimization problem

$$\inf \{ \|f - g\|_2 : g \in \mathcal{S}_n \}. \tag{1.10}$$

From a geometric point of view, this infimum gives the L^2-distance from f to the finite dimensional subspace $\mathcal{S}_n$.

Proposition 1.14. *The unique minimizer of (1.10) is attained at the function* $g = \sum_{j=1}^{n} \alpha_j \phi_j$, *where* $\alpha_j = \langle f, \phi_j \rangle$.

Proof. To minimize $\|f - g\|_2$ is the same as to minimize $\|f - g\|_2^2$. Every g in $\mathcal{S}_n$ can be written as $g = \sum_{j=1}^{n} \beta_j \phi_j$, $\beta_j \in \mathbb{R}$. Let

$$\begin{aligned} \Phi(\beta_1, \ldots, \beta_n) &= \int_{-\pi}^{\pi} |f - g|^2 \\ &= \int_{-\pi}^{\pi} \Big(f - \sum_{j=1}^{n} \beta_j \phi_j \Big)^2 \\ &= \int_{-\pi}^{\pi} f^2 - 2 \sum_{j=1}^{n} \beta_j \alpha_j + \sum_{j=1}^{n} \beta_j^2. \end{aligned}$$

be a function from $\mathbb{R}^n$ to $\mathbb{R}$. We use elementary inequality $2ab \leq a^2 + b^2$ in a tricky way,

$$\begin{aligned} 2 \sum_{j=1}^{n} \beta_j \alpha_j &= 2 \sum_{j=1}^{n} \frac{\beta_j}{\sqrt{2}} \sqrt{2} \alpha_j \\ &\leq \sum_{j=1}^{n} \frac{\beta_j^2}{2} + 2 \sum_{j=1}^{n} \alpha_j^2. \end{aligned}$$

Therefore,

$$\begin{aligned}\Phi(\beta) &\geq \int_{-\pi}^{\pi} f^2 - \frac{1}{2}\sum_{j=1}^{n}\beta_j^2 - 2\sum_{j=1}^{n}\alpha_j^2 + \sum_{j=1}^{n}\beta_j^2 \\ &= \int_{-\pi}^{\pi} f^2 - 2\sum_{j=1}^{n}\alpha_j^2 + \frac{1}{2}|\beta|^2 \\ &\to \infty,\end{aligned}$$

as $|\beta| \to \infty$. It implies that Φ must attain a minimum at some finite point γ. At this point γ, $\nabla\Phi(\gamma) = (0, \ldots, 0)$. We compute

$$\frac{\partial \Phi}{\partial \beta_i} = -2\alpha_i + 2\beta_i.$$

Hence, $\beta = \alpha$. As there is only one critical point, it must be the minimum of Φ. □

Given an orthonormal set $\{\phi_n\}_{n=1}^{\infty}$, one may define the "Fourier series" of an L^2-function f with respect to the orthnormal set $\{\phi_n\}$ to be the series $\sum_{n=1}^{\infty}\langle f, \phi_n\rangle \phi_n$ and set $P_n f = \sum_{k=1}^{n}\langle f, \phi_k\rangle \phi_k$. This proposition asserts that the distance between f and $\mathcal{S}_n$ is realized at $\|f - S_n f\|_2$. The function $P_n f$ is sometimes called the orthogonal projection of f on $\mathcal{S}_n$. As a special case, taking $\{\phi_n\} = \left\{1/\sqrt{2\pi}, \cos nx/\sqrt{\pi}, \sin nx/\sqrt{\pi}\right\}$ and $\mathcal{S}_{2n+1} = E_n$, a direct computation shows that $P_{2n+1} f = S_n f$, where $S_n f$ is the n-th partial sum of the Fourier series of f. Thus we can rewrite Proposition 1.14 in this special case as

Corollary 1.15. *For $f \in R_{2\pi}$, for each $n \geq 1$,*

$$\|f - S_n f\|_2 \leq \|f - g\|_2$$

for all g of the form

$$g = c_0 + \sum_{k=1}^{n}(c_j \cos kx + d_j \sin kx), \quad c_0, c_k, d_k \in \mathbb{R}.$$

Here is the main result of this section.

Theorem 1.16. *For every $f \in R_{2\pi}$,*

$$\lim_{n\to\infty} \|S_n f - f\|_2 = 0.$$

Proof. Let $f \in R[-\pi, \pi]$. For $\varepsilon > 0$, we can find a 2π-periodic, Lipschitz continuous function g such that

$$\|f - g\|_2 < \frac{\varepsilon}{2}.$$

Indeed, g can be obtained by first approximating f by a step function and then modifying the step function at its jumps. By Theorem 1.7, we can fix an N so that in sup-norm

$$\|g - S_N g\|_\infty < \frac{\varepsilon}{2\sqrt{2\pi}}.$$

Thus

$$\|g - S_N g\|_2 = \sqrt{\int_{-\pi}^{\pi} (g - S_N g)^2} \leq \|g - S_N g\|_\infty \sqrt{2\pi} < \frac{\varepsilon}{2}.$$

It follows from Corollary 1.15 that

$$\begin{aligned} \|f - S_N f\|_2 &\leq \|f - S_N g\|_2 \\ &\leq \|f - g\|_2 + \|g - S_N g\|_2 \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

As $\mathcal{S}_N \subset \mathcal{S}_n$ for all $n \geq N$, by Corollary 1.15 again, we have

$$\|f - S_n f\|_2 \leq \|f - S_N f\|_2 < \varepsilon.$$

□

We have the following result concerning the uniqueness of the Fourier expansion.

Corollary 1.17. *(a) Suppose that f_1 and f_2 in $R_{2\pi}$ have the same Fourier series. Then f_1 and f_2 are equal almost everywhere.*

(b) Suppose that f_1 and f_2 in $\mathcal{C}_{2\pi}$ have the same Fourier series. Then $f_1 \equiv f_2$.

Proof. Let $f = f_2 - f_1$. The Fourier coefficients of f all vanish, hence $S_n f = 0$, for all n. By Theorem 1.16, $\|f\|_2 = 0$. From Appendix II we know that f^2, hence f, must vanish almost everywhere. In other words, f_2 is equal to f_1 almost everywhere. (a) holds. To prove (b), letting f be continuous and assuming $f(x)$ is not equal to zero at some x, by continuity it is non-zero for all points near x. Hence we may assume x belongs to $(-\pi, \pi)$ and $|f(y)| > 0$ for all $y \in (x - \delta, x + \delta)$ for some $\delta > 0$. But then $\|f\|_2$ would be greater or equal to the integral of $|f|$ over $(x - \delta, x + \delta)$, which is positive. This contradiction shows that $f \equiv 0$. □

Another interesting consequence of Theorem 1.16 is the Parseval's identity. In fact, this identity is equivalent to Theorem 1.16.

Corollary 1.18 (Parseval's Identity). *For every $f \in R_{2\pi}$,*

$$\|f\|_2^2 = 2\pi a_0^2 + \pi \sum_{n=1}^{\infty} \left(a_n^2 + b_n^2\right),$$

where a_n and b_n are the Fourier coefficients of f.

Proof. Making use of the relations such as $\langle f, \cos nx/\sqrt{\pi}\rangle_2 = \sqrt{\pi}a_n, n \geq 1$, we have $\langle f, S_n f\rangle_2 = \|S_n f\|_2^2 = 2\pi a_0^2 + \pi\sum_{k=1}^n (a_k^2 + b_k^2)$. By Theorem 1.15,

$$\begin{aligned} 0 = \lim_{n\to\infty} \|f - S_n f\|_2^2 &= \lim_{n\to\infty} \left(\|f\|_2^2 - 2\langle f, S_n f\rangle_2 + \|S_n f\|_2^2\right) \\ &= \lim_{n\to\infty} \left(\|f\|_2^2 - \|S_n f\|_2^2\right) \\ &= \|f\|_2^2 - \left[2\pi a_0^2 + \pi\sum_{n=1}^{\infty}\left(a_n^2 + b_n^2\right)\right]. \end{aligned}$$

□

The norm of f can be regarded as the length of the "vector" f. Parseval's Identity shows that the square of the length of f is equal to the sum of the square of the length of the orthogonal projection of f onto each one-dimensional subspace spanned by the sine and cosine functions. This is an infinite dimensional version of the ancient Pythagoras theorem. It is curious to see what really comes out when you plug in some specific functions. For instance, we take $f(x) = x$ and recall that its Fourier series is given by $\sum 2(-1)^{n+1}/n \sin nx$. Therefore, $a_n = 0, n \geq 0$ and $b_n = 2(-1)^{n+1}/n$ and Parseval's identity yields Euler's summation formula

$$\frac{\pi^2}{6} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots .$$

You could find more interesting identities by applying the same idea to other functions.

The following result will be used in the next section.

Corollary 1.19 (Wirtinger's Inequality). *For every $f \in R_{2\pi}$ satisfying $f' \in R_{2\pi}$,*

$$\int_{-\pi}^{\pi} (f(x) - \overline{f})^2 dx \leq \int_{-\pi}^{\pi} f'^2(x) dx \ ,$$

and equality holds if and only if $f(x) = a_0 + a_1 \cos x + b_1 \sin x$.

Here

$$\overline{f} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f$$

is the average or mean of f over $[-\pi, \pi]$.

Proof. Noting that

$$\overline{f} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\cos 0x dx = a_0,$$

$$f(x) - \overline{f} = \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

by Theorem 1.6. By Parseval's identity,

$$\int_{-\pi}^{\pi}(f(x)-\overline{f})^2dx=\pi\sum_{n=1}^{\infty}(a_n^2+b_n^2)\ ,$$

and

$$\int_{-\pi}^{\pi}(f(x)-\overline{f})'^2dx=\int_{-\pi}^{\pi}f'(x)^2dx=\pi\sum_{n=1}^{\infty}n^2(a_n^2+b_n^2).$$

Therefore, we have

$$\int_{-\pi}^{\pi}f'(x)^2dx-\int_{-\pi}^{\pi}(f(x)-\overline{f})^2dx=\pi\sum_{n=1}^{\infty}(n^2-1)(a_n^2+b_n^2)\ ,$$

and the result follows.

□

This inequality is also known as Poincaré's Inequality.

1.6 The Isoperimetric Problem

The classical isoperimetric problem known to the ancient Greeks asserts that only the circle maximizes the enclosed area among all simple, closed curves of the same perimeter. In this section we will present a proof of this inequality by Fourier series. To formulate this geometric problem in analytic terms, we need to recall some facts from advanced calculus.

Indeed, a parametric C^1-curve is a map γ from some interval $[a,b]$ to $\mathbb{R}^2$ such that x and y belong to $C^1[a,b]$ where $\gamma(t)=(x(t),y(t))$ and $x'^2(t)+y'^2(t)>0$ for all $t\in[a,b]$. In the following a curve is always refereed to a parametric C^1-curve. For such a curve, its length is defined to be

$$L[\gamma]=\int_a^b\sqrt{x'(t)^2+y'(t)^2}dt,\quad \gamma=(x,y).$$

A curve is closed if $\gamma(a)=\gamma(b)$ and simple if $\gamma(t)\neq\gamma(s),\ \forall t\neq s$ in $[a,b)$. The length of a closed curve is called the perimeter of the curve.

When a closed, simple curve is given, the area it encloses is also fixed. Hence one should be able to express this enclosed area by a formula involving γ only. Indeed, this can be accomplished by the Green's theorem. Recalling that the Green's theorem states that for every pair of C^1-functions P and Q defined on the curve γ and the region enclosed by the curve, we have

$$\int_\gamma Pdx+Qdy=\iint_D\Big(\frac{\partial Q}{\partial x}(x,y)-\frac{\partial P}{\partial y}(x,y)\Big)dxdy\ ,$$

where the left hand side is the line integral along γ and D is the domain enclosed by γ (see Fritzpatrick, p.543). Taking $P \equiv 0$ and $Q = x$, we obtain

$$\iint_D \Big(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\Big) dxdy = \iint_D 1 = \text{ area of } D,$$

so

$$A[\gamma] = \iint_D 1 dxdy = \int_\gamma x dy = \int_a^b x(t)y'(t)dt \ .$$

The classical isoperimetric problem is: Among all simple, closed curves with a fixed perimeter, find the one whose enclosed area is the largest. We will see that the circle is the only solution to this problem.

To proceed further, let us recall the concept of reparametrization. Indeed, a curve γ_1 on $[a_1, b_1]$ is called a reparametrization of the curve γ on $[a, b]$ if there exists a C^1-map ξ from $[a_1, b_1]$ to $[a, b]$ with non-vanishing derivative so that $\gamma_1(t) = \gamma(\xi(t)),\ \forall t \in [a_1, b_1]$. It is known that the length remains invariant under reparametrizations.

Another useful concept is the parametrization by arc-length. A curve $\gamma = (x, y)$ on $[a, b]$ is called in arc-length parametrization if $x'^2(t) + y'^2(t) = 1,\ \forall t \in [a, b]$. We know that every curve can be reparametrized in arc-length parametrization. Let $\gamma(t) = (x(t), y(t)),\ t \in [a, b]$, be a parametrization of a curve. We define a function φ by setting

$$\varphi(\tau) = \int_a^\tau \big(x'^2(t) + y'^2(t)\big)^{1/2} dt,$$

it is readily checked that φ is a C^1-map from $[a, b]$ to $[0, L]$ with positive derivative, and $\gamma_1(s) = \gamma(\xi(s)),\ \xi = \varphi^{-1}$, is an arc-length reparametrization of γ on $[0, L]$ where L is the length of γ.

We now apply the Wirtinger's Inequality to give a proof of the classical isoperimetric problem.

Let $\gamma : [a, b] \to \mathbb{R}^2$ be a closed, simple C^1-curve bounding a region D. Without loss of generality we may assume that it is parametrized by arc-length. Assuming the perimeter of γ is equal to 2π, we want to find the region that encloses the maximal area. The perimeter is given by

$$L[\gamma] = \int_0^{2\pi} \sqrt{x'^2(s) + y'^2(s)} ds = 2\pi \ ,$$

and the area is given by

$$A[\gamma] = \int_0^{2\pi} x(s)y'(s)ds \ .$$

Extending γ_1 and γ_2 as 2π-periodic functions, we compute

$$\begin{aligned}
2A[\gamma] &= \int_{-\pi}^{\pi} 2x(s)y'(s)ds \\
&= \int_{-\pi}^{\pi} 2(x(s)-\overline{x})y'(s)ds \\
&\leq \int_{-\pi}^{\pi} (x(s)-\overline{x})^2 ds + \int_{-\pi}^{\pi} y'^2(s)ds \qquad (\text{by } 2ab \leq a^2+b^2) \\
&\leq \int_{-\pi}^{\pi} x'^2(s)ds + \int_{-\pi}^{\pi} y'^2(s)ds \qquad (\text{by Wirtinger's Inequality}) \\
&= \int_{-\pi}^{\pi} \big(x'^2(s)+y'^2(s)\big)ds \\
&= 2\pi, \qquad (\text{use } x'^2(s)+y'^2(s)=1)
\end{aligned}$$

whence $A[\gamma] \leq \pi$. We have shown that the enclosed area of a simple, closed C^1-curve with perimeter 2π cannot exceed π. As π is the area of the unit circle, the unit circle solves the isoperimetric problem.

Now the uniqueness case. We need to examine the equality signs in our derivation. We observe that the second equality holds if and only if $a_n = b_n = 0$ for all $n \geq 2$ in the Fourier series of $x(s)$. So, $x(s) = a_0 + a_1 \cos s + b_1 \sin s$, or

$$x(s) = a_0 + r\cos(s - x_0),$$

where

$$r = \sqrt{a_1^2 + b_1^2},\ \cos x_0 = \frac{a_1}{r}\ .$$

(Note that $(a_1, b_1) \neq (0,0)$. For if $a_1 = b_1 = 0$, $x(s)$ is constant and $x'^2 + y'^2 = 1$ implies $y'^2(s) = \pm s + b$, and y can never be periodic.) Now we determine y. From the above calculation, when the first equality holds ($2ab = a^2 + b^2$ means $a - b = 0$),

$$x - \overline{x} - y' = 0.$$

So $y'(s) = x(s) - \overline{x} = r\cos(s - x_0)$, which gives

$$y(s) = r\sin(s - x_0) + c_0\ , \quad c_0 \text{ constant.}$$

It follows that γ describes a circle of radius r centered at (a_0, c_0). Using the fact that the perimeter is 2π, we conclude that $r = 1$, so the maximum must be a unit circle.

Summarizing, we have the following solution to the classical isoperimetric problem.

Theorem 1.20. *Among all closed, simple C^1-curves of the same perimeter, only the circle encloses the largest area.*

The same proof also produces a dual statement, namely, among all regions which enclose the same area, only the circle has the shortest perimeter.

Appendix I Series of Functions

This appendix serves to refresh your memory after the long, free summer.

A sequence is a mapping φ from $\mathbb{N}$ to $\mathbb{R}$. For $\varphi(n) = a_n$, we usually denote the sequence by $\{a_n\}$ rather than φ. This is a convention. We say the sequence is convergent if there exists a real number a satisfying, for every $\varepsilon > 0$, there exists some n_0 such that $|a_n - a| < \varepsilon$ for all $n, n \geq n_0$. When this happens, we write $a = \lim_{n\to\infty} a_n$.

An (infinite) series is always associated with a sequence. Given a sequence $\{x_n\}$, set $s_n = \sum_{k=1}^{n} x_k$ and form another sequence $\{s_n\}$. This sequence is the infinite series associated to $\{x_n\}$ and is usually denoted by $\sum_{k=1}^{\infty} x_k$. The sequence $\{s_n\}$ is also called the sequence of n-th partial sums of the infinite series. By definition, the infinite series is convergent if $\{s_n\}$ is convergent. When this happens, we denote the limit of $\{s_n\}$ by $\sum_{k=1}^{\infty} x_k$, in other words, we have

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_k = \sum_{k=1}^{\infty} x_k.$$

So the notation $\sum_{k=1}^{\infty} x_k$ has two meanings, first, it is the notation for an infinite series and, second, the limit of its partial sums (whenever it exists).

When the target $\mathbb{R}$ is replaced by $\mathbb{C}$, we obtain a sequence or a series of complex numbers, and the above definitions apply to them after replacing the absolute value by the complex absolute value or modulus.

Let $\{f_n\}$ be a sequence of real- or complex-valued functions defined on some non-empty E on $\mathbb{R}$. It is called convergent pointwisely to some function f defined on the same E if for every $x \in E$, $\{f_n(x)\}$ converges to $f(x)$ as $n \to \infty$. Keep in mind that $\{f_n(x)\}$ is sequence of real or complex numbers, so its convergence has a valid meaning. A more important concept is the uniform convergence. The sequence $\{f_n\}$ is uniformly convergent to f if, for every $\varepsilon > 0$, there exists some n_0 such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq n_0$. Equivalently, uniform convergence holds if, for every $\varepsilon > 0$, there exists some n_1 such that $\|f_n - f\|_\infty < \varepsilon$ for all $n \geq n_1$. Here $\|f\|_\infty$ denotes the sup-norm of f on E.

An (infinite) series of functions is the infinite series given by $\sum_{k=1}^{\infty} f_k(x)$ where f_k are defined on E. Its convergence and uniform convergence can be defined via its partial sums $s_n(x) = \sum_{k=1}^{n} f_k(x)$ as in the case of sequences of numbers.

Among several criteria for uniform convergence, the following test is very useful.

Weierstrass M-Test. *Let $\{f_k\}$ be a sequence of functions defined on some $E \subset \mathbb{R}$. Suppose that there exists a sequence of non-negative numbers, $\{x_k\}$, such that*

1. *$|f_k(x)| \leq x_k$ for all $k \geq 1$, and*
2. *$\sum_{k=1}^{\infty} x_k$ is convergent.*

Then $\sum_{k=1}^{\infty} f_k$ converges uniformly and absolutely on E.

Appendix II Sets of Measure Zero

Let E be a subset of $\mathbb{R}$. It is called of measure zero, or sometimes called a null set, if for every $\varepsilon > 0$, there exists a (finite or infinite) sequence of intervals $\{I_k\}$ satisfying (1) $E \subset \cup_{k=1}^{\infty} I_k$ and (2) $\sum_{k=1}^{\infty} |I_k| < \varepsilon$. (When the intervals are finite, the upper limit of the summation should be changed accordingly.) Here I_k could be an open, closed or any other interval and its length $|I_k|$ is defined to the $b - a$ where $a \leq b$ are the endpoints of I_k.

The empty set is a set of measure zero from this definition. Every finite set is also null. For, let $E = \{x_1, \cdots, x_N\}$ be the set. For $\varepsilon > 0$, the intervals $I_k = (x_1 - \varepsilon/(4N), x_k + \varepsilon/(4N))$ clearly satisfy (1) and (2) in the definition.

Next we claim that every countable set is also of measure zero. Let $E = \{x_1, x_2, \cdots\}$ be a countable set. We choose

$$I_k = \left(x_k - \frac{\varepsilon}{2^{k+2}}, x_k + \frac{\varepsilon}{2^{k+2}}\right) .$$

Clearly, $E \subset \cup_{k=1}^{\infty} I_k$. On the other hand,

$$\begin{aligned} \sum_{k=1}^{\infty} |I_k| &= \sum_{k=1}^{\infty} \frac{\varepsilon}{2^{k+1}} \\ &= \frac{\varepsilon}{2} \\ &< \varepsilon . \end{aligned}$$

We conclude that every countable set is a null set.

There are uncountable sets of measure zero. For instance, the Cantor set which plays an important role in analysis, is of measure zero. Here we will not go into this.

The same trick in the above proof can be applied to the following situation.

Proposition A.1. *The union of countably many null sets is a null set.*

Proof. Let $E_k, k \geq 1$, be sets of measure zero. For $\varepsilon > 0$, there are intervals satisfying $\{I_j^k\}, E_k \subset \cup_j I_j^k$, and $\sum_j |I_j^k| < \varepsilon/2^k$. It follows that $E \equiv \cup_k E_k \subset \cup_{j,k} I_j^k = \cup_k \cup_j I_j^k$ and

$$\sum_k \sum_j |I_j^k| < \sum_k \frac{\varepsilon}{2^k} = \varepsilon.$$

□

The concept of a null set comes up naturally in the theory of Riemann integration. A theorem of Lebsegue asserts that a bounded function is Riemann integrable if and only if its discontinuity set is null. (This result can be found in an appendix of Bartle-Sherbert and also in my 2060 notes. It will be proved again in Real Analysis. Presently you may simply take it for granted.) Let us prove the following result.

Proposition A.2. *Let f be a non-negative integrable function on $[a, b]$. Then $\int_a^b f = 0$ if and only if f is equal to 0 except on a null set. Consequently, two integrable functions f, g satisfying*

$$\int_a^b |f - g| = 0,$$

if and only if f is equal to g except on a null set.

Proof. We set, for each $k \geq 1$, $A_k = \{x \in [a, b] : \ f(x) > 1/k\}$. It is clear that

$$\{x : \ f(x) > 0\} = \bigcup_{k=1}^{\infty} A_k \ .$$

By Proposition A.1., it suffices to show that each A_k is null. Thus let us consider A_{k_0} for a fixed k_0. Recall from the definition of Riemann integral, for every $\varepsilon > 0$, there exists a partition $a = x_1 < x_2 < \cdots < x_n = b$ such that

$$0 \leq \sum_{k=1}^{n-1} f(z_k)|I_k| = \left| \sum_{k=1}^{n-1} f(z_k)|I_k| - \int_a^b f \right| < \frac{\varepsilon}{k_0} \ ,$$

where $I_k = [x_k, x_{k+1}]$ and z_k is an arbitrary tag in $[x_j, x_{j+1}]$. Let $\{k_1, \cdots, k_m\}$ be the index set for which I_{k_j} contains a point z_{k_j} from A_{k_0}. Choosing the tag point to be z_{k_j}, we have $f(z_{k_j}) = 1/k_0$. Therefore,

$$\frac{1}{k_0} \sum_{k_j} |I_{k_j}| = \sum_{k_j} f(z_{k_j})|I_{k_j}| \leq \sum_{k=1}^{n-1} f(z_k)|I_k| < \frac{\varepsilon}{k_0},$$

so

$$\sum_{k_j} |I_{k_j}| < \varepsilon.$$

We have shown that A_{k_0} is of measure zero.

Conversely, let D be the set consisting of all discontinuity points of f. We claim that $A \subset D$. Assuming this, the aforementioned Lebsegue's theorem tells us that D is of measure zero, so is A. To prove the claim, let us assume that there is a some $x_0 \in A$ at which f is continuous. We have $f(x_0) > 0$. Without loss of generality we may also assume x_0 is not one of the endpoints of the interval. By the definition of continuity, there exists some small $\delta > 0$ such that $|f(x) - f(x_0)| < f(x_0)/2$ for $x \in (x_0 - \delta, x_0 + \delta)$. Consequently, $f(x) \geq f(x_0) - f(x_0)/2 = f(x_0)/2$ for x in this interval. We have

$$\begin{aligned} 0 &= \int_a^b f \\ &= \int_a^{x_0-\delta} f + \int_{x_0-\delta}^{x_0+\delta} f + \int_{x_0+\delta}^b f \\ &\geq \int_{x_0-\delta}^{x_0+\delta} f \\ &\geq \frac{f(x_0)}{2} \times 2\delta > 0, \end{aligned}$$

contradiction holds. □

A property holds **almost everywhere** if it holds except on a null set. For instance, this proposition asserts that the integral of a non-negative function is equal to zero if and only if it vanishes almost everywhere.

Comments on Chapter 1. Historically, the relation (1.2) comes from a study on the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

where $u(x,t)$ denote the displacement of a string at the position-time (x,t). Around 1750, D'Alembert and Euler found that a general solution of this equation is given by

$$f(x - ct) + g(x + ct)$$

where f and g are two arbitrary twice differentiable functions. However, D. Bernoulli found that the solution could be represented by a trigonometric series. These two different ways of representing the solutions led to a dispute among the mathematicians at that time, and it was not settled until Fourier gave many convincing examples of representing functions by trigonometric series in 1822. His motivation came from heat conduction. After that, trigonometric series have been studied extensively and people call it Fourier series in honor of the contribution of Fourier. Nowadays, the study of Fourier series has matured into a branch of mathematics called harmonic analysis. It has equal importance in theoretical and applied mathematics, as well as other branches of natural sciences and engineering.

The book by R.T. Seely, "An Introduction to Fourier Series and Integrals", W.A. Benjamin, New York, 1966, is good for further reading.

In some books the Fourier series of a function is written in the form

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

instead of

$$a_0 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

so that the formula for a_0 is the same as the other a_n's (see (1.1)). However, our notation has the advantage that a_0 has a simple meaning, i.e., it is the average of the function over a period.

Concerning the convergence of a Fourier series to its function, we point out that an example of a continuous function whose Fourier series diverges at some point can be found in Stein-Sharachi. More examples are available by googling. The classical book by A. Zygmund, "Trigonometric Series" (1959) reprinted in 1993, contains most results before 1960. After 1960, one could not miss to mention Carleson's sensational work in 1966. His result in particular implies that the Fourier series of every function in $R_{2\pi}$ converges to the function itself almost everywhere.

There are several standard proofs of the Weierstrass approximation theorem, among them Rudin's proof in "Principles" by expanding an integral kernel and Bernstein's proof based on binomial expansion are both worth reading. Recently the original proof of Weierstrass by the heat kernel is available on the web. It is nice to take a look too. In Chapter 3 we will reproduce Rudin's proof and then discuss Stone-Weierstrass theorem, a far reaching generalization of Weierstrass approximation theorem.

The elegant proof of the Isoperimetric Inequality by Fourier series presented here is due to Hurwitz (1859-1919). You may google under "proofs of the Isoperimetric Inequality" to find several different proofs in the same spirit. The Isoperimetric Inequality has a higher dimensional version which asserts that the ball has the largest volume among all domains having the same surface area. However, the proof is much more complicated.

The aim of this chapter is to give an introduction to Fourier series. It will serve the purpose if your interest is aroused and now you consider to take our course on Fourier analysis in the future. (Watch out! This course will not be offered in 2016/17.) Not expecting a thorough study, I name Stein-Shakarchi as the only reference.

Chapter 2

Metric Spaces

狙公賦芧曰：朝三而莫四。衆狙皆怒。曰：然則朝四而莫三。衆狙皆悦。
莊子 齊物論

A metric space is a mathematical object in which the distance between two points is meaningful. Metric spaces constitute an important class of topological spaces. We introduce metric spaces and give some examples in Section 1. In Section 2 open and closed sets are introduced and we discuss how to use them to describe the convergence of sequences and the continuity of functions. Relevant notions such as the boundary points, closure and interior of a set are discussed. Compact sets are introduced in Section 3 where the equivalence between the Bolzano-Weierstrass formulation and the finite cover property is established. In Sections 4 and 5 we turn to complete metric spaces and the Contraction Mapping Principle. Two applications of the Contraction Mapping Principle are subsequently given, first a proof of the Inverse Function Theorem in Section 6 and, second, a proof of the fundamental existence and uniqueness theorem for the initial value problem of differential equations in Section 7.

2.1 Definitions and Examples

Throughout this chapter X always denotes a non-empty set. We would like to define a concept of distance which assigns a positive number to every two points in X, that is, the distance between them. In analysis the name metric is used instead of distance. (But "d" not "m" is used in notation. I have no idea why it is so.) A **metric** on X is a function from $X \times X$ to $[0, \infty)$ which satisfies the following three conditions: $\forall x, y, z \in X$,

M1. $d(x, y) \geq 0$ and equality holds if and only if $x = y$,

M2. $d(x, y) = d(y, x)$, and

M3. $d(x, y) \leq d(x, z) + d(z, y)$.

The last condition, the triangle inequality, is a key property of a metric. M2 and M3 together imply another form of triangle inequality,

$$|d(x,y) - d(x,z)| \le d(y,z).$$

The pair (X,d) is called a **metric space**. Let $x \in X$ and $r > 0$, the **metric ball** (of radius r centered at x) or simply the ball $B_r(x)$ is the set $\{y \in X : d(y,x) < r\}$.

Here are some examples of metric spaces.

Example 2.1. Let $\mathbb{R}$ be the set of all real numbers. For $x, y \in \mathbb{R}$, we define $d(x,y) = |x-y|$ where $|x|$ denotes the absolute value of x. It is easily seen that $d(\cdot,\cdot)$ satisfies M1-M3 above and so it defines a metric. In particular, M3 reduces to the usual triangle inequality. Thus $(\mathbb{R}, d)$ is a metric space. From now on whenever we talk about $\mathbb{R}$, it is understood that it is a metric space endowed with this metric.

Example 2.2. More generally, let $\mathbb{R}^n$ be the n-dimensional real vector space consisting of all n-tuples $x = (x_1, \dots, x_n)$, $x_j \in \mathbb{R}$, $j = 1, \dots, n$. For $x, y \in \mathbb{R}^n$, introduce the **Euclidean metric**

$$\begin{aligned} d_2(x,y) &= \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} \\ &= \big(\sum_{j=1}^{n} (x_j - y_j)^2\big)^{\frac{1}{2}}. \end{aligned}$$

It reduces to Example 1 when $n = 1$. Apparently, M3 and M2 are fulfilled. To verify the triangle inequality, letting $u = x - z$ and $v = x - y$, M3 becomes

$$\big(\sum_{1}^{n} (u_j + v_j)^2\big)^{1/2} \le \big(\sum_{1}^{n} u_j^2\big)^{1/2} + \big(\sum_{1}^{n} v_j^2\big)^{1/2}.$$

Taking square, we see that it follows from Cauchy-Schwarz inequality

$$\Big|\sum_{1}^{n} u_j v_j\Big| \le \big(\sum_{1}^{n} u_j^2\big)^{1/2} \big(\sum_{1}^{n} v_j^2\big)^{1/2}.$$

In case you do not recall its proof, look up a book. We need to use mathematics you learned in all previous years. Take this as a chance to refresh them.

Example 2.3. It is possible to have more than one metrics on a set. Again consider $\mathbb{R}^n$. Instead of the Euclidean metric, we define

$$d_1(x,y) = \sum_{j=1}^{n} |x_j - y_j|$$

and

$$d_\infty(x,y) = \max\left\{|x_1 - y_1|, \ldots, |x_n - y_n|\right\}.$$

It is not hard to verify that d_1 and d_∞ are also metrics on $\mathbb{R}^n$. We denote the metric balls in the Euclidean, d_1 and d_∞ metrics by $B_r(x)$, $B_r^1(x)$ and $B_r^\infty(x)$ respectively. $B_r(x)$ is the standard ball of radius r centered at x and $B_r^\infty(x)$ is the cube of length r centered at x. I let you draw and tell me what $B_r^1(x)$ looks like.

Example 2.4. Let $C[a,b]$ be the real vector space of all continuous, real-valued functions on $[a,b]$. For $f, g \in C[a,b]$, define

$$d_\infty(f,g) = \|f-g\|_\infty \equiv \max\left\{|f(x) - g(x)| : x \in [a,b]\right\}.$$

It is easily checked that d_∞ is a metric on $C[a,b]$. The metric ball $B_r(f)$ in the uniform metric consists of all continuous functions sitting inside the "tube"

$$\left\{(x,y) : |y - f(x)| < r, x \in [a,b]\right\}.$$

Another metric defined on $C[a,b]$ is given by

$$d_1(f,g) = \int_a^b |f-g|.$$

It is straightforward to verify M1-M3 are satisfied. In Section 1.5 we encountered the L^2-distance. Indeed,

$$d_2(f,g) = \sqrt{\int_a^b |f-g|^2},$$

really defines a metric on $C[a,b]$. The verification of M3 to similar to what we did in Example 2.2, but Cauchy-Schwarz inequality is now in integral form

$$\int_a^b |fg| \leq \sqrt{\int_a^b f^2}\sqrt{\int_a^b g^2}.$$

In passing we point out same notations such as d_1 and d_2 have been used to denote different metrics. They arise in quite different context though. It should not cause confusion.

Example 2.5. Let $R[a,b]$ be the vector space of all Riemann integrable functions on $[a,b]$ and consider $d_1(f,g)$ as defined in the previous example. One can show that M2 and M3 are satisfied, but not M1. In fact,

$$\int_a^b |f-g| = 0$$

does not imply that f is equal to g. It tells us they differ on a set of measure zero. This happens, for instance, they are equal except at finitely many points. To construct a metric

space out of d_1, we introduce a relation on $R[a,b]$ by setting $f \sim g$ if and only if f and g differ on a set of measure zero. It is routine to verify that $\sim$ is an equivalence relation. Let $\widetilde{R}[a,b]$ be the equivalence classes of $R[a,b]$ under this relation. We define a metric on $\widetilde{R}[a,b]$ by, $\forall \overline{f}, \overline{g} \in \widetilde{R}[a,b]$,

$$\widetilde{d_1}(\overline{f}, \overline{g}) = d_1(f,g), \quad f \in \overline{f},\ g \in \overline{g}.$$

Then $(\widetilde{R}[a,b], \widetilde{d_1})$ forms a metric space. I let you verify that $\widetilde{d_1}$ is well-defined, that is, it is independent of the choices of f and g, and is a metric on $(\widetilde{R}[a,b], \widetilde{d_1})$. A similar consideration applies to the L^2-distance to get a metric $\widetilde{d_2}$.

A **norm** $\|\cdot\|$ is a function on a real vector space X to $[0,\infty)$ satisfying the following three conditions, for all $x, y \in X$ and $\alpha \in \mathbb{R}$,

N1. $\|x\| \geq 0$ and "=" 0 if and only if $x = 0$

N2. $\|\alpha x\| = |\alpha|\ \|x\|$, and

N3. $\|x + y\| \leq \|x\| + \|y\|$.

The pair $(X, \|\cdot\|)$ is called a **normed space**. There is always a metric associated to a norm. Indeed, letting

$$d(x,y) = \|x - y\|,$$

it is readily checked that d defines a metric on X. This metric is called the **metric induced** by the norm. In all the five examples above the metrics are induced respectively from norms. I leave it to you to write down the corresponding norms. Normed spaces will be studied in MATH4010 Functional Analysis. In the following we give two examples of metrics defined on a set without the structure of a vector space. Hence they cannot be metrics induced by norms.

Example 2.6. Let X be a non-empty set. For $x, y \in X$, define

$$d(x,y) = \begin{cases} 1, & x \neq y, \\ 0, & x = y. \end{cases}$$

The metric d is called the **discrete metric** on X. The metric ball $B_r(x)$ consists of x itself for all $r \in (0,1)$.

Example 2.7. Let H be the collection of all strings of words in n digits. For two strings of words in H, $a = a_1 \cdots a_n$, $b = b_1 \cdots b_n$, $a_j, b_j \in \{0,1,2,\ldots,9\}$. Define

$$d_H(a,b) = \text{the number of digits at which } a_j \text{ is not equal to } b_j.$$

By using a simple induction argument one can show that (H, d_H) forms a metric space. Indeed, the case $n = 1$ is straightforward. Let us assume it holds for n-strings and

show it for $(n+1)$-strings. Let $a = a_1 \cdots a_n a_{n+1}, b = b_1 \cdots b_n b_{n+1}, c = c_1 \cdots c_n c_{n+1}, a' = a_1 \cdots a_n, b' = b_1 \cdots b_n$, and $c' = c_1 \cdots c_n$. Consider the case that $a_{n+1} = b_{n+1} = c_{n+1}$. We have $d_H(a,b) = d_H(a',b') \leq d_H(a',c') + d_H(c',b') = d_H(a,c) + d_H(c,b)$ by induction hypothesis. When a_{n+1} is not equal to one of b_{n+1}, c_{n+1}, say, $a_{n+1} \neq b_{n+1}$, we have $d_H(a,b) = d_H(a',b')+1 \leq d_H(a',c')+d_H(c',b')+1$. If $b_{n+1} = c_{n+1}$, $d_H(a,c) = d_H(a',c')+1$. Therefore, $d_H(a',c') + d_H(c',b') + 1 \leq d_H(a,c) + d_H(c,b)$. If $b_{n+1} \neq c_{n+1}$, $d_H(c',b') = d_H(c,b) - 1$ and the same inequality holds. Finally, if $a_{n+1} \neq c_{n+1}$ and $a_{n+1} = b_{n+1}$, $d_H(a,b) = d_H(a',b') \leq d_H(a',c') + d_H(c',b') \leq d_H(a,c) - 1 + d_H(c,b) - 1 \leq d_H(a,c) + d_H(c,b)$. The metric d_H is called the **Hamming distance**. It measures the error in a string during transmission.

Let Y be a non-empty subset of (X,d). Then $(Y, d|_{Y\times Y})$ is again a metric space. It is called a **metric subspace** of (X,d). The notation $d|_{Y\times Y}$ is usually written as d for simplicity. Every non-empty subset of a metric space forms a metric space under the restriction of the metric. In the following we usually call a metric subspace a subspace for simplicity. Note that a metric subspace of a normed space needs not be a normed space. It is so only if the subset is also a vector subspace.

Recall that convergence of sequences of real numbers and uniform convergence of sequences of functions are main themes in Mathematical Analysis I and II and sequences of vectors were considered in Advanced Calculus I and II. With a metric d on a set X, it makes sense to talk about limits of sequences in a metric space. Indeed, a sequence in (X,d) is a map φ from $\mathbb{N}$ to (X,d) and usually we write it in the form $\{x_n\}$ where $\varphi(n) = x_n$. We call $\{x_n\}$ **converges to** x if $\lim_{n\to\infty} d(x_n, x) = 0$, that's, for every $\varepsilon > 0$, there exists n_0 such that $d(x_n, x) < \varepsilon$, for all $n \geq n_0$. When this happens, we write or $\lim_{n\to\infty} x_n = x$ or $x_n \to x$ in X.

Convergence of sequences in $(\mathbb{R}^n, d_2)$ reduces to the old definition we encountered before. From now on, we implicitly refer to the Euclidean metric whenever convergence of sequences in $\mathbb{R}^n$ is considered. For sequences of functions in $(C[a,b], d_\infty)$, it is simply the uniform convergence of sequences of functions in $C[a,b]$.

As there could be more than one metrics defined on the same set, it is natural to make a comparison among these metrics. Let d and ρ be two metrics defined on X. We call ρ is **stronger** than d, or d is **weaker** than ρ, if there exists a positive constant C such that $d(x,y) \leq C\rho(x,y)$ for all $x, y \in X$. They are **equivalent** if d is stronger and weaker than ρ simultaneously, in other words,

$$d(x,y) \leq C_1\rho(x,y) \leq C_2 d(x,y), \quad \forall x, y \in X,$$

for some positive C_1 and C_2. When ρ is stronger than d, a sequence converging in ρ is also convergent in d. When d and ρ are equivalent, a sequence is convergent in d if and only if it is so in ρ.

Take d_1, d_2 and d_∞ on $\mathbb{R}^n$ as an example. It is elementary to show that for all $x, y \in \mathbb{R}^n$,

$$d_2(x,y) \le n^{1/2} d_\infty(x,y) \le n^{1/2} d_2(x,y),$$

and

$$d_1(x,y) \le n d_\infty(x,y) \le n d_1(x,y),$$

hence d_1, d_2 and d_∞ are all equivalent. The convergence of a sequence in one metric implies its convergence in other two metrics.

It is a basic result in functional analysis that every two induced metrics in a finite dimensional normed space are equivalent. We will divide the proof of this fact in several parts in the exercise. Consequently, examples of inequivalent metrics can only be found when the underlying space is of infinite dimensional.

Let us display two inequivalent metrics on $C[a,b]$. For this purpose it suffices to consider d_1 and d_∞. On one hand, clearly we have

$$d_1(f,g) \le (b-a) d_\infty(f,g), \quad \forall f, g \in C[a,b],$$

so d_∞ is stronger than d_1. But the converse is not true. Consider the sequence given by (taking $[a,b] = [0,1]$ for simplicity)

$$f_n(x) = \begin{cases} -n^3 x + n, & x \in [0, 1/n^2], \\ 0, & x \in (1/n^2, 1]. \end{cases}$$

We have $d_1(f_n, 0) \to 0$ but $d_\infty(f_n, 0) \to \infty$ as $n \to \infty$. Were $d_\infty(f_n,0) \le C d_1(f_n,0)$ true for some positive constant C, $d_1(f_n,0)$ must tend to ∞ as well. Now it tends to 0, so d_1 cannot be stronger than d_2 and these two metrics are not equivalent.

Now we define continuity in a metric space. Recalling that for a real-valued function defined on some set E in $\mathbb{R}$, there are two equivalent ways to define the continuity of the function at a point. We could use either the behavior of sequences or the ε-δ formulation. Specifically, the function f is continuous at $x \in E$ if for every sequence $\{x_n\} \subset E$ satisfying $\lim_{n\to\infty} x_n = x$, $\lim_{n\to\infty} f(x_n) = f(x)$. Equivalently, for every $\varepsilon > 0$, there exists some $\delta > 0$ such that $|f(y) - f(x)| < \varepsilon$ whenever $y \in E, |y - x| < \delta$. Both definition can be formulated on a metric space. Let (X,d) and (Y,ρ) be two metric spaces and $f : (X,d) \to (Y,\rho)$. Let $x \in X$. We call f is **continuous at** x if $f(x_n) \to f(x)$ in (Y,ρ) whenever $x_n \to x$ in (X,d). It is **continuous** on a set $E \subset X$ if it is continuous at every point of E.

Proposition 2.1. *Let f be a mapping from (X,d) to (Y,ρ) and $x_0 \in X$. Then f is continuous at x_0 if and only if for every $\varepsilon > 0$, there exists some $\delta > 0$ such that $\rho(f(x), f(x_0)) < \varepsilon$ for all $x, d(x, x_0) < \delta$.*

Proof. $\Leftarrow$) Let ε be given and δ is chosen accordingly. For any $\{x_n\} \to x_0$, given $\delta > 0$, there exists some n_0 such that $d(x_n, x_0) < \delta$ $\forall n \ge n_0$. It follows that $\rho(f(x_n), f(x_0)) < \varepsilon$ for all $n \ge n_0$, so f is continuous at x_0.

$\Rightarrow$) Suppose that the implication is not valid. There exist some $\varepsilon_0 > 0$ and $\{x_k\} \in X$ satisfying $d(f(x_k), f(x_0)) \geq \varepsilon_0$ and $d(x_k, x_0) < 1/k$. However, the second condition tells us that $\{x_k\} \to x_0$, so by the continuity at x_0 one should have $d(f(x_k), f(x_0)) \to 0$, contradiction holds. □

We will shortly use open/closed sets to describe continuity in a metric space.

As usual, continuity of functions is closed under compositions of functions.

Proposition 2.2. *Let $f : (X, d) \to (Y, \rho)$ and $g : (Y, \rho) \to (Z, m)$ be given.*

(a) If f is continuous at x and g is continuous at $f(x)$, then $g \circ f : (X, d) \to (Z, m)$ is continuous at x.

(b) If f is continuous in X and g is continuous in Y, then $g \circ f$ is continuous in X.

Proof. It suffices to prove (a). Let $x_n \to x$. Then $f(x_n) \to f(x)$ as f is continuous at x. Then $(g \circ f)(x_n) = g(f(x_n)) \to g(f(x)) = (g \circ f)(x)$ as g is continuous at $f(x)$. □

2.2 Open and Closed Sets

The existence of a metric on a set enables us to talk about convergence of a sequence and continuity of a map. It turns out that, in order to define continuity, it requires a structure less stringent then a metric structure. It suffices the set is endowed with a topological structure. In a word, a metric induces a topological structure on the set but not every topological structure comes from a metric. In a topological space, continuity can no longer be defined via the convergence of sequences. Instead one uses the notion of open and closed sets in the space. As a warm up for topology we discuss how to use the language of open/closed sets to describe the convergence of sequences and the continuity of functions in this section.

Let (X, d) be a metric space. A set $G \subset X$ is called an **open set** if for each $x \in G$, there exists some ρ such that $B_\rho(x) \subset G$. The number ρ may vary depending on x. We also define the empty set ϕ to be an open set.

Proposition 2.3. *Let (X, d) be a metric space. We have*

(a) X and ϕ are open sets.

(b) If $\bigcup_{\alpha \in \mathcal{A}} G_\alpha$ is an open set provided that all G_α, $\alpha \in \mathcal{A}$, are open where $\mathcal{A}$ is an arbitrary index set.

(c) If $G_1, \ldots, G_N$ are open sets, then $\bigcap_{j=1}^{N} G_j$ is an open set.

Note the union in (b) of this proposition is over an arbitrary collection of sets while the intersection in (c) is a finite one.

Proof. (a) Obvious.

(b) Let $x \in \bigcup_{\alpha\in\mathcal{A}} G_\alpha$. There exists some α_1 such that $x \in G_{\alpha_1}$. As G_{α_1} is open, there is some $B_\rho(x) \subset G_{\alpha_1}$. But then $B_\rho(x) \subset \bigcup_{\alpha\in\mathcal{A}} G_\alpha$, so $\bigcup_{\alpha\in\mathcal{A}} G_\alpha$ is open.

(c) Let $x \in \bigcap_{j=1}^N G_j$. For each j, there exists $B_{\rho_j}(x) \subset G_j$. Let $\rho = \min\{\rho_1, \ldots, \rho_N\}$. Then $B_\rho(x) \subset \bigcap_{j=1}^N G_j$, so $\bigcap_{j=1}^N G_j$ is open. □

The complement of an open set is called a **closed set**. Taking the complement of Proposition 2.2, we have

Proposition 2.4. *Let (X, d) be a metric space. We have*

(a) X and ϕ are closed sets.

(b) If F_α, $\alpha \in \mathcal{A}$, are closed sets, then $\bigcap_{\alpha\in\mathcal{A}} F_\alpha$ is a closed set.

(c) If $F_1, \ldots, F_N$ are closed sets, then $\bigcup_{j=1}^N F_j$ is a closed set.

Note that X and ϕ are both open and closed.

Example 2.8. Every ball in a metric space is an open set. Let $B_r(x)$ be a ball and $y \in B_r(x)$. We claim that $B_\rho(y) \subset B_r(x)$ where $\rho = r - d(y, x) > 0$. For, if $z \in B_\rho(y)$,

$$\begin{aligned} d(z, x) &\leq d(z, y) + d(y, x) \\ &< \rho + d(y, x) \\ &= r \ , \end{aligned}$$

by the triangle inequality, so $z \in B_r(x)$ and $B_\rho(y) \subset B_r(x)$ holds. Next, the set $E = \{y \in X : \ d(y, x) > r\}$ for fixed x and $r \geq 0$ is an open set. For, let $y \in E$, $d(y, x) > r$. We claim $B_\rho(y) \subset E$, $\rho = d(y, x) - r > 0$,. For, letting $z \in B_\rho(y)$,

$$\begin{aligned} d(z, x) &\geq d(y, x) - d(y, z) \\ &> d(y, x) - \rho \\ &= r, \end{aligned}$$

shows that $B_\rho(y) \subset E$, hence E is open. Finally, consider $F = \{x \in X : \ d(x, z) = r > 0\}$ where z and r are fixed. Observing that F is the complement of the two open sets $B_r(z)$ and $\{x \in X : \ d(x, z) > r\}$, we conclude that F is a closed set.

Example 2.9. In the real line every open interval (a, b), $-\infty \leq a \leq b \leq \infty$, is an open set. Other intervals such as $[a, b)$, $[a, b]$, $(a, b]$, $a, b \in \mathbb{R}$, are not open. It can be shown that every open set G in $\mathbb{R}$ can be written as a disjoint union of open intervals. Letting $(a_n, b_n) = (-1/n, 1/n)$,

$$\bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$$

is not open. It shows that Proposition 2.2(c) does not hold when the intersection is over infinite many sets. On the other hand, $[a, b]$ is a closed set since $\{a\} = \mathbb{R}\backslash(-\infty, a) \cup (a, \infty)$, a single point is always a closed set. Finally, some sets we encounter often are neither open nor closed. Take the set of all rational numbers as example, as every open interval containing a rational number also contains an irrational number, we see that $\mathbb{Q}$ is not open. The same reasoning shows that the set of all irrational numbers is not open, hence $\mathbb{Q}$ is also not a closed set.

Example 2.10. When we studied multiple integrals in Advanced Calculus II, we encountered many domains or regions as the domain of integration. These domains are open sets in $\mathbb{R}^n$. For instance, consider the set $G = \{(x_1, x_2) \in \mathbb{R}^2 : \ x_1^2/4 + x_2^2/9 < 1\}$ which is the set of all points lying inside an ellipse. We claim that it is an open set. Each point $(y_1, y_2) \in G$ satisfies the inequality

$$\frac{y_1^2}{4} + \frac{y_2^2}{9} < 1.$$

Since the function $(x_1, x_2) \mapsto x_1^2/4 + x_2^2/9 - 1$ is continuous, there exists some $\varepsilon > 0$ such that

$$\frac{z_1^2}{4} + \frac{z_2^2}{9} < 1,$$

for all (z_1, z_2), $d((z_1, z_2), (y_1, y_2)) < \varepsilon$. In other words, the ball $B_\varepsilon((y_1, y_2))$ is contained in G, so G is open. Similarly, one can show that the outside of the ellipse, denoted by H, is open (using the fact $x_1^2/4 + x_2^2/9 > 1$ in H) and the set composed of all points lying on the ellipse, denoted by S, is closed. Finally, the set $G \cup S$ is closed as its complement H is open. In general, most domains in $\mathbb{R}^2$ in advanced calculus consist of points bounded by one or several continuous curves. They are all open sets like G. All points lying outside of the boundary curves form an open set and those lying on the curves form a closed set. The points sitting inside and on the curves form an closed set. The situation extends to higher dimensional domains whose boundary are given by finitely many pieces of continuous hypersurfaces.

Example 2.11. Consider the set $E = \{f \in C[a, b] : \ f(x) > 0, \ \forall x \in [a, b]\}$ in $C[a, b]$. We claim that it is open. For $f \in E$, it is positive everywhere on the closed, bounded interval $[a, b]$, hence it attains its minimum at some x_0. It follows that $f(x) \geq m \equiv f(x_0) > 0$.

Letting $r = m/2$, for $g \in B_r(f),\ d_\infty(g, f) < r = m/2$ implies

$$\begin{aligned} g(x) &\geq f(x) - |g(x) - f(x)| \\ &> m - \frac{m}{2} \\ &= \frac{m}{2} > 0, \end{aligned}$$

for all $x \in [a, b]$, hence $g \in E$ which implies $B_r(f) \subset E$, E is open. Likewise, sets like $\{f :\ f(x) > \alpha,\ \forall x\}, \{f :\ f(x) < \alpha,\ \forall x\}$ where α is a fixed number. On the other hand, by taking complements of these open sets, we see that the sets $\{f :\ f(x) \geq \alpha,\ \forall x\}, \{f : f(x) \leq \alpha,\ \forall x\}$ are closed.

Example 2.12. Let us consider a general situation. Let A be a non-empty set in a metric space (X, d) and $x \in X$. The distance from x to A is defined to be

$$\text{dist } (x, A) = \inf\{d(y, x) :\ y \in A\}.$$

Then the set $\{x \in X :\ \text{dist } (x, A) > 0\}$ is open and $\{x \in X :\ \text{dist } (x, A) = 0\}$ is closed. To see this, we first note that

$$|d(x, A) - d(y, A)| \leq d(x, y), \quad \forall x, y \in X\}.$$

That is, the distance function satisfies a Lipschitz condition (see Chapter 1 for the definition). The proof runs as follows, for $\varepsilon > 0$, there exists some $z \in A$ such that $d(y, A) + \varepsilon > d(y, z)$. Therefore,

$$\begin{aligned} d(x, A) &\leq d(x, z) \\ &\leq d(x, y) + d(y, z) \\ &\leq d(x, y) + d(y, A) + \varepsilon. \end{aligned}$$

Letting $\varepsilon \to 0$, we obtain $d(x, A) - d(y, A) \leq d(x, y)$ and the desired result follows after noting the roles of x and y are exchangeable. Now, let $d(x, A) = \rho > 0$, we have $d(y, A) \geq \rho - d(x, y) > 0$ for all $y \in B_{\rho/2}(x)$, hence $\{x \in X :\ \text{dist } (x, A) > 0\}$ is open and its complement $\{x \in X :\ \text{dist } (x, A) = 0\}$ is closed.

Example 2.13. Consider the extreme case where the space X is endowed with the discrete metric. We claim that every set is open and closed. Clearly, it suffices to show that every singleton set $\{x\}$ is open. But, this is obvious because the ball $B_{1/2}(x) = \{x\}$ belongs to $\{x\}$. It is also true that $B_r(x) = X$ once $r > 1$.

We now use open sets to describe the convergence of sequences.

Proposition 2.5. *Let (X, d) be a metric space. A sequence $\{x_n\}$ converges to x if and only if for each open G containing x, there exists n_0 such that $x_n \in G$ for all $n \geq n_0$.*

Proof. Let G be an open set containing x. According to the definition of an open set, we can find $B_\varepsilon(x) \subset G$. It follows that there exists n_0 such that $d(x_n, x) < \varepsilon$ for all $n \geq n_0$, i.e., $x_n \in B_\varepsilon(x) \subset G$ for all $n \geq n_0$. Conversely, taking $G = B_\varepsilon(x)$, we see that $x_n \to x$. □

From this proposition we deduce the following result which explains better the terminology of a closed set.

Proposition 2.6. *The set A is a closed set in (X, d) if and only if whenever $\{x_n\} \subset A$ and $x_n \to x$ as $n \to \infty$ implies that x belongs to A.*

Proof. $\Rightarrow$). Assume on the contrary that x does not belong to A. As $X \setminus A$ is an open set, by Proposition 2.4 we can find a ball $B_\varepsilon(x) \subset X \setminus A$. However, as $x_n \to x$, there exists some n_0 such that $x_n \in B_\varepsilon(x)$ for all $n \geq n_0$, contradicting the fact that $x_n \in A$.

$\Leftarrow$). If $X \setminus A$ is not open, say, we could find a point $x \in X \setminus A$ such that $B_{1/n}(x) \bigcap A \neq \phi$ for all n. Pick $x_n \in B_{1/n}(x) \bigcap A$ to form a sequence $\{x_n\}$. Clearly $\{x_n\}$ converges to x. By assumption, $x \in A$, contradiction holds. Hence $X \setminus A$ must be open.

□

Now we use open sets to describe the continuity of functions.

Proposition 2.7. *Let $f : (X, d) \to (Y, \rho)$.*

(a) f is continuous at x if and only if for every open set G containing $f(x)$, $f^{-1}(G)$ contains $B_\varepsilon(x)$ for some $\varepsilon > 0$.

(b) f is continuous in X if and only if for every open G in Y, $f^{-1}(G)$ is an open set in X.

These statements are still valid when "open" is replaced by "closed".

Proof. We consider (a) and (b) comes from (a) easily.

$\Rightarrow$). Suppose there exists some open G such that $f^{-1}(G)$ does not contain $B_{1/n}(x)$ for all $n \geq 1$. Pick $x_n \in B_{1/n}(x)$, $x_n \notin f^{-1}(G)$. Then $x_n \to x$ but $f(x_n)$ does not converge to x, contradicting the continuity of f.

$\Leftarrow$). Let $\{x_n\} \to x$ in X. Given any open set G containing $f(x)$, we can find $B_r(x) \subset f^{-1}(G)$. Thus, there exists n_0 such that $x_n \in B_r(x)$ for all $n \geq n_0$. It follows that $f(x_n) \in G$ for all $n \geq n_0$. By Proposition 2.4, f is continuous at x.

□

Example 2.14. Consider Example 2.12 again. We showed that the function $f(x) = \text{dist}\,(x, A)$ is a continuous function in (X, d). By this proposition, we immediately deduce that the set $f^{-1}(0, \infty) = \{x \in X : \text{ dist}\,(x, A) > 0\}$ is open and $f^{-1}(\{0\}) = \{x \in X : \text{dist}\,(x, A) = 0\}$ is closed.

Let Y be a subspace of (X, d). We describe the open sets in Y. First of all, the metric ball in (Y, d) is given by $B'_r(x) = \{y \in Y : d(y, x) < r\}$ which is equal to $B_r(x) \cap Y$. For an open set E in Y, for each $x \in E$ there exists some $B'_{\rho_x}(x)$, such that $B'_{\rho_x}(x) \subset E$. Therefore,

$$E = \bigcup B'_{\rho_x}(x) = \bigcup (B_{\rho_x}(x) \cap Y) = \left(\bigcup B_{\rho_x}(x)\right) \cap Y.$$

We conclude

Proposition 2.8. *Let Y be a subspace of (X, d). A set E in Y is open in Y if and only if there exists an open set G in (X, d) satisfying $E = G \cap Y$. It is closed if and only if there exists a closed set F in (X, d) satisfying $E = F \cap Y$.*

Example 2.15. Let $[0, 1]$ be the subspace of $\mathbb{R}$ under the Euclidean metric. The set $[0, 1/2)$ is not open in $\mathbb{R}$ as every open interval of the form $(a, b), a < 0 < b$, is not contained in $[0, 1/2)$, so 0 is not an interior point of $[0, 1)$. However, it is relatively open in $[0, 1)$ because when regarded as a subset of $[0, 1)$, the set $[0, a), 1/2 > a > 0$, is an open set (relative in $[0, 1)$) contained in $[0, 1/2)$. For, by the proposition above, $[0, a) = (-1, a) \cap [0, 1)$ is relatively open.

We describe some further useful notions associated to sets in a metric space.

Let E be a set in (X, d). A point x is called a **boundary point** of E if $G \cap E$ and $G \setminus E$ are non-empty for every open set G containing x. Of course, it suffices to take G of the form $B_\varepsilon(x)$ for all sufficiently small ε or $\varepsilon = 1/n,\ n \geq 1$. We denote the boundary of E by ∂E. The **closure** of E, denoted by $\overline{E}$, is defined to be $E \cup \partial E$. Clearly $\partial E = \partial(X \setminus E)$. The boundary of the ball $B_r(x)$ in $\mathbb{R}^n$ is the sphere $S_r(x) = \{y \in \mathbb{R}^n : \ d_2(y, x) = r\}$. Hence, the closed ball $\overline{B_r(x)}$ is given by $B_r(x) \bigcup S_r(x)$, which is precisely the closure of $B_r(x)$.

Example 2.16. Let $E = [0, 1) \times [0, 1)$. It is easy to see that $\partial E = [0, 1] \times \{0, 1\} \cup \{0, 1\} \times [0, 1]$. Thus some points in ∂E belong to E and some do not. The closure of E, $\overline{E}$, is equal to $[0, 1] \times [0, 1]$.

It can be seen from definition that the boundary of the empty set is the empty set and the boundary of a set is always a closed set. For, let $\{x_n\}$ be a sequence in ∂E converging to some x. For any ball $B_r(x)$, we can find some x_n in it, so the ball $B_\rho(x_n),\ \rho =$

$r - d(x_n, x) > 0$, is contained in $B_r(x)$. As $x_n \in \partial E$, $B_\rho(x_n)$ has non-empty intersection with E and $X \setminus E$, so does $B_r(x)$ and $x \in \partial E$ too. The following proposition characterizes the closure of a set as the smallest closed set containing this set.

Proposition 2.9. *Let E be a set in (X, d). We have*

$$\overline{E} = \bigcap \{C : \ C \text{ is a closed set containing } E\}.$$

Proof. We first claim that $\overline{E}$ is a closed set. We will do this by showing $X \setminus \overline{E}$ is open. Indeed, for x lying outside $\overline{E}$, x does not belong to E and there exists an open ball $B_\rho(x)$ disjoint from E. Thus, $\overline{E}$ is disjoint from $B_{\rho/2}(x)$ and so $X \setminus \overline{E}$ is open. We conclude that $\overline{E}$ is closed. Next we claim that $\overline{E}$ is contained in any closed set C containing E. It suffices to show that $\partial E \subset C$. Indeed, if $x \in \partial E$, every ball $B_{1/n}(x)$ would have non-empty intersection with E. By picking a point x_n from $B_{1/n}(x) \cap E$, we obtain a sequence $\{x_n\}$ in E converging to x as $n \to \infty$. As C is closed, x belongs to C by Proposition 2.5. □

A point x is called an **interior point** of E if there exists an open set G containing x such that $G \subset E$. It can be shown that all interior points of E form an open set call the **interior** of E, denoted by E^o. It is not hard to see that $E^0 = E \setminus \partial E$. The interior of a set is related to its closure by the following relation: $E^o = X \setminus \big(\overline{X \setminus E}\big)$. Using this relation, one can show that the interior of a set is the largest open set sitting inside E. More precisely, $G \subset E^0$ whenever G is an open set in E.

Example 2.17. Consider the set of all rational numbers E in $[0, 1]$. It has no interior point since there are irrational numbers in every open interval containing a rational number, so I^o is the empty set. On the other hand, since every open interval contains some rational numbers, the closure of E, $\overline{E}$, is $[0, 1]$. It shows the interior and closure of a set could be very different.

Example 2.18. In Example 2.10 we consider domains in $\mathbb{R}^2$ bounded by several pieces of continuous curves. Let D be such a domain and the curves bounding it be S. It is routine to verify that $\partial D = S$, that is, the set of all boundary points of D is precisely the S and the closure of D, $\overline{D}$, is $D \cup S$. The interior of $\overline{D}$ is D.

Example 2.19. Let $S = \{f \in C[0, 1]) : \ 1 < f(x) \leq 5, x \in [0, 1]\}$. It is easy to see that $\overline{S} = \{f \in C[0, 1] : \ 1 \leq f(x) \leq 5, \ x \in [0, 1]\}$ and $S^o = \{f \in C[0, 1] : \ 1 < f(x) < 5, x \in [0, 1]\}$.

2.3 Compactness

Recall that Bolzano-Weierstrass Theorem asserts that every sequence in a closed bounded interval has a convergent subsequence in this interval. The result still holds for all closed,

bounded sets in $\mathbb{R}^n$. In general, a set $E \subset (X, d)$ is **compact** if every sequence has a convergent subsequence with limit in E. This property is also called **sequentially compact** to stress that the behavior of sequences is involved in the definition. The space (X, d) is called a **compact space** is X is a compact set itself. According to this definition, every interval of the form $[a, b]$ is compact in $\mathbb{R}$ and sets like $[a_1, b_1] \times [a_2, b_2] \times \cdots [a_n, b_n]$ and $\overline{B_r(x)}$ are compact in $\mathbb{R}^n$ under the Euclidean metric. In a general metric space, the notion of a bounded set makes perfect sense. Indeed, a set A is called a **bounded set** if there exists some ball $B_r(x)$ for some $x \in X$ and $r > 0$ such that $A \subset B_r(x)$. Now we investigate the relation between a compact set and a closed bounded set. First of all, we have

Proposition 2.10. *Every compact set in a metric space is closed and bounded.*

Proof. Let K be a compact set. To show that it is closed, let $\{x_n\} \subset K$ and $x_n \to x$. We need to show that $x \in K$. As K is compact, there exists a subsequence $\{x_{n_j}\} \subset K$ converging to some z in K. By the uniqueness of limit, we have $x = z \in K$, so $x \in K$ and K is closed. On the other hand, if K is unbounded, that is, for any fixed point x_0, K is not contained in the balls $B_n(x_0)$ for all n. Picking $x_n \in K \setminus B_n(x_0)$, we obtain a sequence $\{x_n\}$ satisfying $d(x_n, x_0) \to \infty$ as $n \to \infty$. By the compactness of K, there is a subsequence $\{x_{n_j}\}$ converging to some z in K. By the triangle inequality,

$$\begin{aligned} \infty > d(z, x_0) &= \lim_{j\to\infty} d(x_{n_j}, z) + d(z, x_0) \\ &\geq \lim_{j\to\infty} d(x_{n_j}, x_0) \to \infty, \end{aligned}$$

as $j \to \infty$, contradiction holds. Hence K must be bounded. □

As a consequence of Bolzano-Weierstrass Theorem every sequence in a bounded and closed set in $\mathbb{R}^n$ contains a convergent subsequence. Thus a set in $\mathbb{R}^n$ is compact if and only if it is closed and bounded. Proposition 2.10 tells that every compact set is in general closed and bounded, but the converse is not always true. To describe an example we need to go beyond $\mathbb{R}^n$ where we can be free of the binding of Bolzano-Weierstrass Theorem. Consider the set $\mathcal{S} = \{f \in C[0, 1] : 0 \leq f(x) \leq 1\}$. Clearly it is closed and bounded in $C[0, 1]$. We claim that it is not compact. For, consider the sequence $\{f_n\}$ in $(C[0, 1], d_\infty)$ given by

$$f_n(x) = \begin{cases} nx, & x \in [0, \frac{1}{n}] \\ 1, & x \in [\frac{1}{n}, 1]. \end{cases}$$

$\{f_n(x)\}$ converges pointwisely to the function $f(x) = 1$, $x \in (0, 1]$ and $f(0) = 0$ which is discontinuous at $x = 0$, that is, f does not belong to $C[0, 1]$. If $\{f_n\}$ has a convergent subsequences, then it must converge uniformly to f. But this is impossible because the uniform limit of a sequence of continuous functions must be continuous. Hence $\mathcal{S}$ cannot be compact. In fact, a remarkable theorem in functional analysis asserts that the closed unit ball in a normed space is compact if and only if the normed space if and only if the normed space is of finite dimension.

Since convergence of sequences can be completely described in terms of open/closed sets, it is natural to attempt to describe the compactness of a set in terms of these new notions. The answer to this challenging question is a little strange at first sight. Let us recall the following classical result:

Heine-Borel Theorem. *Let $\{I_j\}_{j=1}^{\infty}$ be a family of open intervals satisfying*

$$[a,b] \subset \bigcup_{j=1}^{\infty} I_j \ .$$

There is always a finite subfamily $\{I_{j_1}, \cdots, I_{j_K}\}$ such that

$$[a,b] \subset \bigcup_{k=1}^{K} I_{j_k} \ .$$

This property is not true for open (a,b). Indeed, the intervals $\{(a+1/j, b-1/j)\}$ satisfy $(a,b) \subset \bigcup_j (a+1/j, b-1/j)$ but there is no finite subcover. It is a good exercise to show that Heine-Borel Theorem is equivalent to Bolzano-Weierstrass Theorem. (When I was an undergraduate in this department in 1974, we were asked to show this equivalence, together with the so-called Nested Interval Theorem, in a MATH2050-like course.) This equivalence motivates how to describe compactness in terms of the language of open/closed sets.

We introduce some terminologies. First of all, an **open cover** of a subset E in a metric space (X,d) is a collection of open sets $\{G_\alpha\}, \alpha \in \mathcal{A}$, satisfying $E \subset \bigcup_{\alpha\in\mathcal{A}} G_\alpha$. A set $E \subset X$ satisfies the **finite cover property** if whenever $\{G_\alpha\}, \alpha \in \mathcal{A}$, is an open cover of E, there exist a subcollection consisting of finitely many $G_{\alpha_1}, \ldots, G_{\alpha_N}$ such that $E \subset \bigcup_{j=1}^{N} G_{\alpha_j}$. ("Every open cover has a finite subcover.") A set E satisfies the **finite intersection property** if whenever $\{F_\alpha\}, \alpha \in \mathcal{A}$, are relatively closed sets in E satisfying $\bigcap_{j=1}^{N} F_{\alpha_j} \neq \phi$ for any finite subcollection F_{α_j}, $\bigcap_{\alpha\in\mathcal{A}} F_\alpha \neq \phi$. Here a set $F \subset E$ is relatively closed means F is closed in the subspace E. We know that it implies $F = A \cap E$ for some closed set A. Therefore, when E is closed, a relatively closed subset is also closed.

Proposition 2.11. *A closed set has the finite cover property if and only if it has the finite intersection property.*

Proof. Let E be a non-empty closed set in (X,d).

$\Rightarrow$) Suppose $\{F_\alpha\}$, F_α closed sets contained in E, satisfies $\bigcap_{j=1}^{N} F_{\alpha_j} \neq \phi$ for any finite subcollection but $\bigcap_{\alpha\in\mathcal{A}} F_\alpha = \phi$. As E is closed, each F_α is closed in X, and

$$E = E \setminus \bigcap_{\alpha\in\mathcal{A}} F_\alpha = \bigcup_{\alpha\in\mathcal{A}} (E \cap F'_\alpha) \subset \bigcup_{\alpha\in\mathcal{A}} F'_\alpha.$$

By the finite covering property we can find $\alpha_1, \ldots, \alpha_N$ such that $E \subset \bigcup_{j=1}^{N} F'_{\alpha_j}$, but then $\phi = E \setminus E \supset E \setminus \bigcup_1^N F'_{\alpha_j} = \bigcap_{j=1}^{N} F_{\alpha_j}$, contradiction holds.

$\Leftarrow$) If $E \subset \bigcup G_{\alpha\in\mathcal{A}}$ but $E \subsetneq \bigcup_{j=1}^{N} G_{\alpha_j}$ for any finite subcollection of $\mathcal{A}$, then

$$\phi \neq E \setminus \bigcup_{j=1}^{N} G_{\alpha_j} = \bigcap_{j=1}^{N} \left(E \setminus G_{\alpha_j}\right)$$

which implies $\bigcap_{\alpha\in\mathcal{A}} \left(E \setminus G_\alpha\right) \neq \phi$ by the finite intersection property. Note that each $E \setminus G_{\alpha_j}$ is closed. Using $E \bigcap \left(\bigcup_{\alpha\in\mathcal{A}} G_\alpha\right)' = \bigcap_{\alpha\in\mathcal{A}} \left(E \setminus G_\alpha\right)$, we have $E \subsetneq \bigcup_{\alpha\in\mathcal{A}} G_\alpha$, contradicting our assumption.

□

Proposition 2.12. *Let E be compact in a metric space. For each $\alpha > 0$, there exist finitely many balls $B_\alpha(x_1), \ldots, B_\alpha(x_N)$ such that $E \subset \bigcup_{j=1}^{N} B_\alpha(x_j)$ where $x_j, 1 \leq j \leq N$, are in E.*

Proof. Pick $B_\alpha(x_1)$ for some $x_1 \in E$. Suppose $E \setminus B_\alpha(x_1) \neq \phi$. We can find $x_2 \notin B_\alpha(x_1)$ so that $d(x_2, x_1) \geq \alpha$. Suppose $E \setminus \left(B_\alpha(x_1) \bigcup B_\alpha(x_2)\right)$ is non-empty. We can find $x_3 \notin B_\alpha(x_1) \bigcup B_\alpha(x_2)$ so that $d(x_j, x_3) \geq \alpha$, $j = 1, 2$. Keeping this procedure, we obtain a sequence $\{x_n\}$ in E such that

$$E \setminus \bigcup_{j=1}^{n} B_\alpha(x_j) \neq \phi \quad \text{and} \quad d(x_j, x_n) \geq \alpha,\ j = 1, 2, \ldots, n-1.$$

By the compactness of E, there exists $\left\{x_{n_j}\right\}$ and $x \in E$ such that $x_{n_j} \to x$ as $j \to \infty$. But then $d(x_{n_j}, x_{n_k}) < d(x_{n_j}, x) + d(x_{n_k}, x) \to 0$, contradicting $d(x_j, x_n) \geq \alpha$ for all $j < n$. Hence one must have $E \setminus \bigcup_{j=1}^{N} B_\alpha(x_j) = \phi$ for some finite N.

□

Sometimes the following terminology is convenient. A set E is called **totally bounded** if for each $\varepsilon > 0$, there exist $x_1, \cdots, x_n \in X$ such that $E \subset \cup_{k=1}^{n} B_\varepsilon(x_k)$. Proposition 2.12 simply states that every compact set is totally bounded. We will use this property of a compact set again in the next chapter.

Theorem 2.13. *Let E be a closed set in (X, d). The followings are equivalent:*

(a) E is compact;

(b) E satisfies the finite cover property; and

(c) E satisfies the finite intersection property.

Proof. (a)$\Rightarrow$ (b). Let $\{G_\alpha\}$ be an open cover of E without finite subcover and we will draw a contradiction. By Proposition 2.11, for each $k \geq 1$, there are finitely many balls of radius $1/k$ covering E. We can find a set $B_{1/k} \cap E$ (suppress the irrelevant center)

which cannot be covered by finitely many members in $\{G_\alpha\}$. Pick $x_k \in B_{1/k} \cap E$ to form a sequence. By the compactness of E, we can extract a subsequence $\{x_{k_j}\}$ such that $x_{k_j} \to x$ for some $x \in E$. Since $\{G_\alpha\}$ covers E, there must be some G_β that contains x. As G_β is open and the radius of B_{1/k_j} tends to 0, we deduce that, for all sufficiently large k_j, $B_{1/k_j} \cap E$ is contained in G_β. In other words, G_β forms a single subcover of $B_{1/k} \cap E$, contradicting our choice of $B_{1/k_j} \cap E$. Hence (b) must be valid.

(b)$\Leftrightarrow$ (c). See Proposition 2.11.

(c)$\Rightarrow$ (a). Let $\{x_n\}$ be a sequence in E. Without loss of generality we may assume that it contains infinitely many distinct points, otherwise the conclusion is obvious. The balls $B_1(x_n)$ form an open cover of the set $\overline{\{x_n\}}$, hence it has a finite subcover. (Note that a closed subset of a set satisfying the finite cover property again satisfies the finite cover property.) Since there are infinitely many distinct points in this sequence, we can choose one of the these balls, denoted by B_1, which contains infinitely many points in E. Next we cover $\overline{\{x_n\}}$ by balls $B_{1/2}(x_n)$. Among a finite subcover, choose one of the balls $B_{1/2}$ which contains infinitely many distinct points from B_1. Keeping doing this, we obtain a sequence of balls $\{B_{1/k}\}, k \geq 1$, such that each $B_1 \cap \cdots \cap B_{1/k}$ contains infinitely many distinct points in E. We pick a subsequence $\{x_{n_k}\}$ from $B_1 \cap \cdots \cap B_{1/k} \cap E$. By the finite intersection property, $\bigcap_k \left(\overline{B}_{1/k} \cap E\right)$ is nonempty and in fact consists of one point $z \in E$. It is clear that $d(x_{n_k}, z) \leq 2/k \to 0$ as $k \to \infty$. We have succeeded in producing a convergent subsequence in the closed set E. Hence E is compact. □

In the proof of the following result, we illustrate how to prove the same statement by using the subsequence approach and the finite cover approach.

Proposition 2.14. *Let K be a compact set and G be an open set, $K \subset G$, in the metric space (X, d). Then*

$$dist\ (K, \partial G) > 0,$$

where dist $(A, B) = \inf\{d(x, y) :\ x \in A,\ y \in B\}$.

Proof. First proof: Suppose on the contrary that dist $(K, \partial G) = 0$. By the definition of the distance between two sets, there are $\{x_n\} \subset K$ and $\{y_n\} \subset \partial G$ such that $d(x_n, y_n) \to 0$. By the compactness of K, there exists a subsequence $\{x_{n_j}\}$ and $x^* \in K$ such that $x_{n_j} \to x^*$. From $d(x^*, y_{n_j}) \leq d(x_{n_j}, y_{n_j}) + d(x^*, x_{n_j}) \to 0$ we see that $x^* \in \partial G$ (the boundary of a set is always closed). But then $G \cap \partial G$ is non-empty, which is impossible as G is open. So dist $(K, \partial G) > 0$.

Second proof: For $x \in K$, it is clear that dist $(x, \partial G) > 0$. We can find a small number $\rho_x > 0$ such that dist $(y, \partial G) > 0$ for all $y \in B_{\rho_x}(x)$. The collection of all balls $B_{\rho_x}(x),\ x \in K$, forms an open cover of K. Since K is compact, there exist $x_1, \cdots, x_N$, such that $B_{x_{\rho_j}}(x_j),\ j = 1, \cdots, N$, form a finite subcover of K. Taking $\rho = \min\{\rho_{x_1}, \cdots, \rho_{x_N}\}$, we conclude dist $(K, G) > \rho > 0$. □

We finally note

Proposition 2.15. *Let E be a compact set in (X, d) and $F : (X, d) \to (Y, \rho)$ be continuous. Then $f(E)$ is a compact set in (Y, ρ).*

Proof. Let $\{y_n\}$ be a sequence in $f(E)$ and let $\{x_n\}$ be in E satisfying $f(x_n) = y_n$ for all n. By the compactness of E, there exist some $\{x_{n_j}\}$ and x in E such that $x_{n_j} \to x$ as $j \to \infty$. By the continuity of f, we have $y_{n_j} = f(x_{n_j}) \to f(x)$ in $f(E)$. Hence $f(E)$ is compact. □

Can you prove this property by using the finite cover property of compact sets?

There are several fundamental theorems which hold for continuous functions defined on a closed, bounded set in the Euclidean space. Notably they include

- A continuous function on a closed, bounded set is uniformly continuous; and
- A continuous function on a closed, bounded set attains its minimum and maximum in the set.

Although they may no longer hold on arbitrary closed, bounded sets in a general metric space, they continue to hold when the sets are strengthened to compact ones. The proofs are very much like in the finite dimensional case. I leave them as exercises.

2.4 Completeness

In $\mathbb{R}^n$ a basic property is that every Cauchy sequence converges. This property is called the completeness of the Euclidean space. The notion of a Cauchy sequence is well-defined in a metric space. Indeed, a sequence $\{x_n\}$ in (X, d) is a **Cauchy sequence** if for every $\varepsilon > 0$, there exists some n_0 such that $d(x_n, x_m) < \varepsilon$, for all $n, m \geq n_0$. A metric space (X, d) is **complete** if every Cauchy sequence converges. A subset E is **complete** if $(E, d\big|_{E\times E})$ is complete.

Example 2.20. The interval $[a, b]$ is a complete space. For, if $\{x_n\}$ is a Cauchy sequence in $[a, b]$, it is also a Cauchy sequence in $\mathbb{R}$. By the completeness of the real line, $\{x_n\}$ converges to some x. Since $[a, b]$ is closed, x must belong to $[a, b]$, so $[a, b]$ is complete. In contrast, the set $[a, b),\ b \in \mathbb{R}$, is not complete. For, simply observe that the sequence $\{b - 1/k\}, k \geq k_0$ for some large k_0, is a Cauchy sequence in $[a, b)$ and yet it does not have a limit in $[a, b)$ (the limit is b, which lies outside $[a, b)$).

Example 2.21. In Mathematical Analysis II we learned that every Cauchy sequence in $C[a, b]$ with respect to the sup-norm implies that it converges uniformly, so the limit is again continuous and $C[a, b]$ is a complete space. The subset $E = \{f : f(x) \geq 0,\ \forall x\}$

is also complete. Let $\{f_n\}$ be a Cauchy sequence in E, it is also a Cauchy sequence in $C[a,b]$ and hence there exists some $f \in C[a,b]$ such that $\{f_n\}$ converges to f uniformly. As uniform convergence implies pointwise convergence, $f(x) = \lim_{n\to\infty} f_n(x) \geq 0$, so f belongs to E and E is complete. Next, let $P[a,b]$ be the collection of all restriction of polynomials on $[a,b]$. It forms a subspace of $C[a,b]$. Taking the sequence $h_n(x)$ given by

$$h_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!},$$

$\{h_n\}$ is a Cauchy sequence in $P[a,b]$ which converges to e^x. As e^x is not a polynomial, $P[a,b]$ is not a complete subset of $C[a,b]$.

Proposition 2.16. *Let (X,d) be a metric space.*

(a) Every closed set in X is complete provided X is complete.

(b) Every complete set in X is closed.

(c) Every compact set in X is complete.

Proof. (a) Let (X,d) be complete and E a closed subset of X. Every Cauchy sequence $\{x_n\}$ in E is also a Cauchy sequence in X. By the completeness of X, there is some x in X to which $\{x_n\}$ converges. However, as E is closed, x also belongs to E. So every Cauchy sequence in E has a limit in E.

(b) Let $E \subset X$ be complete and $\{x_n\}$ a sequence converging to some x in X. Since every convergent sequence is a Cauchy sequence, $\{x_n\}$ must converge to some z in E. By the uniqueness of limit, we must have $x = z \in C$, so C is closed.

(c) Let $\{x_n\}$ be a Cauchy sequence in the compact set K. By compactness, there is a subsequence $\{x_{n+j}\}$ converging to some x in X. As every compact set is also closed, x belongs to K. For $\varepsilon > 0$, there exists some n_0 such that $|x_n - x_m| < \varepsilon/2$, for all $n, m \geq n_0$ and $|x_{n_j} - x| \leq \varepsilon/2$, for $n_j \geq n_0$, it follows that

$$|x_n - x| \leq |x_n - x_{n_j}| + |x_{n_j} - x| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

for all $n \geq n_0$, that is, $\{x_n\}$ converges to x in K. □

To obtain a typical non-complete set, we consider the interval $[0,1]$ in $\mathbb{R}$ which is complete and, in fact, compact. Take away one point z from it to form $E = [a,b] \setminus \{z\}$. E is not complete, since every sequence in E converging to z is a Cauchy sequence which does not converge in E. In general, you may think of sets with "holes" being non-complete ones. Now, given a non-complete metric space, can we make it into a complete metric space by filling out all the holes? The answer turns out to affirmative. We can always enlarge

a non-complete metric space into a complete one by putting in some ideal points. The process of achieving this goal was long invented by Cantor (1845–1918) in his construction of the real numbers from rational numbers. We start with some formalities.

A metric space (X, d) is called **embedded** in (Y, ρ) if there is a mapping $\Phi : X \to Y$ such that $d(x, y) = \rho(\Phi(x), \Phi(y))$. The mapping Φ is sometimes called a **metric preserving** map. Note that it must be 1-1 and continuous. We call the metric space (Y, ρ) a **completion** of (X, d) if (X, d) is embedded in (Y, ρ) and $\overline{\Phi(X)} = Y$. The latter condition is a minimality condition; (X, d) is enlarged merely to accommodate those ideal points to make the space complete.

Theorem 2.17. *Every metric space has a completion.*

Before the proof we briefly describe the idea. When (X, d) is not complete, we need to invent ideal points and add them to X to make it complete. The idea goes back to Cantor's construction of the real numbers from rational numbers. Suppose now we have only rational numbers and we want to add irrationals. First we identify $\mathbb{Q}$ with a proper subset in a larger set as follows. Let $\mathcal{C}$ be the collection of all Cauchy sequences of rational numbers. Every point in $\mathcal{C}$ is of the form $(x_1, x_2, \cdots)$ where $\{x_n\}, x_n \in \mathbb{Q}$, forms a Cauchy sequence. A rational number x is identified with the constant sequence $(x, x, x, \dots)$ or any Cauchy sequence which converges to x. For instance, 1 is identified with $(1, 1, 1, \dots)$, $(0.9, 0.99, 0.999, \dots)$ or $(1.01, 1.001, 1.0001, \dots)$. Clearly, there are Cauchy sequences which cannot be identified with rational numbers. For instance, there is no rational number corresponding to $(3, 3.1, 3.14, 3.141, 3.1415, \dots)$, as we know, its correspondent should be the irrational number π. Similar situation holds for the sequence $(1, 1.4, 1.41, 1.414, \cdots)$ which should correspond to $\sqrt{2}$. Since the correspondence is not injective, we make it into one by introducing an equivalence relation on $\mathcal{C}$ Indeed, $\{x_n\}$ and $\{y_n\}$ are said to be equivalent if $|x_n - y_n| \to 0$ as $n \to \infty$. The equivalence relation $\sim$ forms the quotient $\mathcal{C}/\sim$ which is denoted by $\widetilde{\mathcal{C}}$. Then $x \mapsto \widetilde{x}$ sends $\mathbb{Q}$ injectively into $\widetilde{\mathcal{C}}$. It can be shown that $\widetilde{\mathcal{C}}$ carries the structure of the real numbers. In particular, those points not in the image of $\mathbb{Q}$ are exactly all irrational numbers. Now, for a metric space the situation is similar. We let $\widetilde{\mathcal{C}}$ be the quotient space of all Cauchy sequence in X under the relation $\{x_n\} \sim \{y_n\}$ if and only if $d(x_n, y_n) \to 0$. Define $\widetilde{d}(\widetilde{x}, \widetilde{y}) = \lim_{n\to\infty} d(x_n, y_n)$, for $x \in \widetilde{x}$, $y \in \widetilde{y}$. We have the embedding $(X, d) \to (\widetilde{X}, \widetilde{d})$, and we can further show that it is a completion of (X, d).

The following proof is for optional reading. In the exercise we will present a simpler but less instructive proof.

Proof of Theorem 2.16. Let $\mathcal{C}$ be the collection of all Cauchy sequences in (M, d). We introduce a relation $\sim$ on $\mathcal{C}$ by $x \sim y$ if and only if $d(x_n, y_n) \to 0$ as $n \to \infty$. It is routine to verify that $\sim$ is an equivalence relation on $\mathcal{C}$. Let $\widetilde{X} = \mathcal{C}/\sim$ and define a map:

$\widetilde{X} \times \widetilde{X} \mapsto [0, \infty)$ by

$$\widetilde{d}(\widetilde{x}, \widetilde{y}) = \lim_{n\to\infty} d(x_n, y_n)$$

where $x = (x_1, x_2, x_3, \cdots)$ and $y = (y_1, y_2, y_3, \cdots)$ are respective representatives of $\widetilde{x}$ and $\widetilde{y}$. We note that the limit in the definition always exists: For

$$d(x_n, y_n) \leq d(x_n, x_m) + d(x_m, y_m) + d(y_m, y_n)$$

and, after switching m and n,

$$|d(x_n, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_m, y_n).$$

As x and y are Cauchy sequences, $d(x_n, x_m)$ and $d(y_m, y_n) \to 0$ as $n, m \to \infty$, and so $\{d(x_n, y_n)\}$ is a Cauchy sequence of real numbers.

Step 1. (well-definedness of $\widetilde{d}$) To show that $\widetilde{d}(\widetilde{x}, \widetilde{y})$ is independent of their representatives, let $x \sim x'$ and $y \sim y'$. We have

$$d(x_n, y_n) \leq d(x_n, x'_n) + d(x'_n, y'_n) + d(y'_n, y_n).$$

After switching x and x', and y and y',

$$|d(x_n, y_n) - d(x'_n, y'_n)| \leq d(x_n, x'_n) + d(y_n, y'_n).$$

As $x \sim x'$ and $y \sim y'$, the right hand side of this inequality tends to 0 as $n \to \infty$. Hence $\lim_{n\to\infty} d(x_n, y_n) = \lim_{n\to\infty} d(x'_n, y'_n)$.

Step 2. ($\widetilde{d}$ is a metric). Let $\{x_n\}, \{y_n\}$ and $\{z_n\}$ represent $\widetilde{x}, \widetilde{y}$ and $\widetilde{z}$ respectively. We have

$$\begin{aligned}
\widetilde{d}(\widetilde{x}, \widetilde{z}) &= \lim_{n\to\infty} \big(d(x_n, z_n) \\
&\leq \lim_{n\to\infty} \big(d(x_n, y_n) + d(y_n, z_n)\big) \\
&= \lim_{n\to\infty} d(x_n, y_n) + \lim_{n\to\infty} d(y_n, z_n) \\
&= \widetilde{d}(\widetilde{x}, \widetilde{y}) + \widetilde{d}(\widetilde{y}, \widetilde{z})
\end{aligned}$$

Step 3. We claim that there is a metric preserving map $\Phi : X \mapsto \widetilde{X}$ satisfying $\overline{\Phi(X)} = \widetilde{X}$.

Given any x in X, the "constant sequence" $(x, x, x, \cdots)$ is clearly a Cauchy sequence. Let $\widetilde{x}$ be its equivalence class in $\mathcal{C}$. Then $\Phi x = \widetilde{x}$ defines a map from X to $\widetilde{X}$. Clearly

$$\widetilde{d}(\Phi(x), \Phi(y)) = \lim_{n\to\infty} d(x_n, y_n) = d(x, y)$$

since $x_n = x$ and $y_n = y$ for all n, so Φ is metric preserving and it is injective in particular.

To show that the closure of $\Phi(X)$ is $\widetilde{X}$, we observe that any $\widetilde{x}$ in $\widetilde{X}$ is represented by a Cauchy sequence $x = (x_1, x_2, x_3, \cdots)$. Consider the constant sequence $x^n = (x_n, x_n, x_n, \cdots)$ in $\Phi(X)$. We have

$$\widetilde{d}(\widetilde{x}, \widetilde{x_n}) = \lim_{m\to\infty} d(x_m, x_n).$$

Given $\varepsilon > 0$, there exists an n_0 such that $d(x_m, x_n) < \varepsilon/2$ for all $m, n \geq n_0$. Hence $\widetilde{d}(\widetilde{x}, \widetilde{x_n}) = \lim_{m\to\infty} d(x_m, x_n) < \varepsilon$ for $n \geq n_0$. That is $\widetilde{x^n} \to \widetilde{x}$ as $n \to \infty$, so the closure of $\Phi(M)$ is precisely M.

Step 4. We claim that $(\widetilde{X}, \widetilde{d})$ is a complete metric space. Let $\{\widetilde{x^n}\}$ be a Cauchy sequence in $\widetilde{X}$. As $\overline{\Phi(X)}$ is equal to $\widetilde{M}$, for each n we can find a $\widetilde{y}$ in $\Phi(X)$ such that

$$\widetilde{d}(\widetilde{x^n}, \widetilde{y^n}) < \frac{1}{n}\ .$$

So $\{\widetilde{y^n}\}$ is also a Cauchy sequence in $\widetilde{d}$. Let y_n be the point in X so that $y^n = (y_n, y_n, y_n, \cdots)$ represents $\widetilde{y^n}$. Since Φ is metric preserving, and $\{\widetilde{y^n}\}$ is a Cauchy sequence in $\widetilde{d}$, $\{y_n\}$ is a Cauchy sequence in X. Let $(y_1, y_2, y_3, \cdots) \in \widetilde{y}$ in $\widetilde{X}$. We claim that $\widetilde{y} = \lim_{n\to\infty} \widetilde{x^n}$ in $\widetilde{X}$. For, we have

$$\begin{aligned} \widetilde{d}(\widetilde{x^n}, \widetilde{y}) &\leq \widetilde{d}(\widetilde{x^n}, \widetilde{y^n}) + \widetilde{d}(\widetilde{y^n}, \widetilde{y}) \\ &\leq \frac{1}{n} + \lim_{m\to\infty} d(y_n, y_m) \to 0 \end{aligned}$$

as $n \to \infty$. We have shown that $\widetilde{d}$ is a complete metric on $\widetilde{X}$. □

Completion of a metric space is unique once we have clarified the meaning of uniqueness. Indeed, call two metric spaces (X, d) and (X', d') **isometric** if there exists a bijective embedding from (X, d) onto (X', d'). Since a metric preserving map is always one-to-one, the inverse of of this mapping exists and is a metric preserving mapping from (X', d') to (X, d). So two spaces are isometric provided there is a metric preserving map from one onto the other. Two metric spaces will be regarded as the same if they are isometric, since then they cannot be distinguish after identifying a point in X with its image in X' under the metric preserving mapping. With this understanding, the completion of a metric space is unique in the following sense: If (Y, ρ) and (Y', ρ') are two completions of (X, d), then (Y, ρ) and (Y', ρ') are isometric. We will not go into the proof of this fact, but instead leave it to the interested reader. In any case, now it makes sense to use "the completion" of X to replace "a completion" of X.

2.5 The Contraction Mapping Principle

Solving an equation $f(x) = 0$, where f is a function from $\mathbb{R}^n$ to itself frequently comes up in application. This problem can be turned into a problem for fixed points. Literally,

a fixed point of a mapping is a point which becomes unchanged under this mapping. By introducing the function $g(x) = f(x) + x$, solving the equation $f(x) = 0$ is equivalent to finding a fixed point for g. This general observation underlines the importance of finding fixed points. In this section we prove the Contraction Mapping Principle, one of the oldest fixed point theorems and perhaps the most well-known one. As we will see, it has a wide range of applications.

A map $T : (X, d) \to (X, d)$ is called a **contraction** if there is a constant $\gamma \in (0, 1)$ such that $d(Tx, Ty) \leq \gamma d(x, y)$, $\forall x, y \in X$. A point x is called a **fixed point** of T if $Tx = x$. Usually we write Tx instead of $T(x)$.

Theorem 2.18 (**Contraction Mapping Principle**). *Every contraction in a complete metric space admit a unique fixed point.*

This theorem is also called Banach's Fixed Point Theorem.

Proof. Let T be a contraction in the complete metric space (X, d). Pick an arbitrary $x_0 \in X$ and define a sequence $\{x_n\}$ by setting $x_n = Tx_{n-1} = T^n x_0$, $\forall n \geq 1$. We claim that $\{x_n\}$ forms a Cauchy sequence in X. First of all, by iteration we have

$$\begin{aligned} d(T^n x_0, T^{n-1} x_0) &\leq \gamma d(T^{n-1} x_0, T^{n-2} x_0) \\ &\cdot \\ &\cdot \\ &\leq \gamma^{n-1} d(Tx_0, x_0). \end{aligned} \tag{2.1}$$

Next, for $n \geq N$ where N is to be specified in a moment,

$$\begin{aligned} d(x_n, x_N) &= d(T^n x_0, T^N x_0) \\ &\leq \gamma d(T^{n-1} x_0, T^{N-1} x_0) \\ &\leq \gamma^N d(T^{n-N} x_0, x_0). \end{aligned}$$

By the triangle inequality and (2.1),

$$\begin{aligned} d(x_n, x_N) &\leq \gamma^N \sum_{j=1}^{n-N} d(T^{n-N-j+1} x_0, T^{n-N-j} x_0) \\ &\leq \gamma^N \sum_{j=1}^{n-N} \gamma^{n-N-j} d(Tx_0, x_0) \\ &< \frac{d(Tx_0, x_0)}{1-\gamma} \gamma^N. \end{aligned} \tag{2.2}$$

For $\varepsilon > 0$, choose N so large that $d(Tx_0, x_0)\gamma^N/(1-\gamma) < \varepsilon/2$. Then for $n, m \geq N$,

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_N) + d(x_N, x_m) \\ &< \frac{2d(Tx_0, x_0)}{1-\gamma} \gamma^N \\ &< \varepsilon, \end{aligned}$$

thus $\{x_n\}$ forms a Cauchy sequence. As X is complete, $x = \lim_{n\to\infty} x_n$ exists. By the continuity of T, $\lim_{n\to\infty} Tx_n = Tx$. But on the other hand, $\lim_{n\to\infty} Tx_n = \lim_{n\to\infty} x_{n+1} = x$. We conclude that $Tx = x$.

Suppose there is another fixed point $y \in X$. From

$$\begin{aligned} d(x,y) &= d(Tx,Ty) \\ &\leq \gamma d(x,y), \end{aligned}$$

and $\gamma \in (0,1)$, we conclude that $d(x,y) = 0$, i.e., $x = y$. □

Incidentally, we point out that this proof is a constructive one. It tells you how to find the fixed point starting from an arbitrary point. In fact, letting $n \to \infty$ in (2.2) and then replacing N by n, we obtain an error estimate between the fixed point and the approximating sequence $\{x_n\}$:

$$d(x,x_n) \leq \frac{d(Tx_0,x_0)}{1-\gamma}\gamma^n, \quad n \geq 1.$$

Example 2.22. Let us take X to be $\mathbb{R}$. Then T is nothing but a real-valued function on $\mathbb{R}$. Denoting the identity map $x \mapsto x$ by I. A point on the graph of T is given by (x,Tx) and a point on the graph of I is (x,x). So every intersection point of both graphs $(x,Tx) = (x,x)$ is a fixed point of T. From this point of view we can see functions may or may not have fixed points. For instance, the function $Tx = x + e^x$ does not have any fixed point. By drawing graphs one is convinced that there are functions with graphs lying below the diagonal line and yet whose slope is always less than one but tends to 1 at infinity (see exercise for a concrete one). It shows the necessity of $\gamma \in (0,1)$. On the other hand, functions like $Sx = x(x-1)(x+2)$ whose graph intersects the diagonal line three times, so it has three fixed points. The insight of Banach's Fixed Point Theorem is to single out a class of functions which admits one and only one fixed point. The contractive condition can be expressed as

$$\left|\frac{Tx-Ty}{x-y}\right| < \gamma, \quad \forall x,y.$$

It means that the slope of T is always bounded by $\gamma \in (0,1)$. Let (x,Tx) be a point of the graph of T and consider the cone emitting from this point bounded by two lines of slopes $\pm\gamma$. When T is a contraction, it is clear that its graph lies within this cone. A moment's reflection tells us that it must hit the diagonal line exactly once.

Example 2.23. Let $f : [0,1] \to [0,1]$ be a continuously differentiable function satisfying $|f'(x)| < 1$ on $[0,1]$. We claim that f admits a fixed point. For, by the mean value theorem, for $x,y \in [0,1]$ there exists some $z \in (0,1)$ such that $f(y) - f(x) = f'(z)(y-x)$. Therefore,

$$\begin{aligned} |f(y) - f(x)| &= |f'(z)||y-x| \\ &\leq \gamma|y-x|, \end{aligned}$$

where $\gamma = \sup_{t\in[0,1]} |f'(t)| < 1$ (Why?). We see that f is a contraction. By the Contraction Mapping Principle, it has a fixed point. In fact, by using the mean-value theorem one can show that *every continuous function* from $[0, 1]$ to itself admits at least one fixed point. This is a general fact. According to Brouwer's Fixed Point Theorem, every continuous maps from a compact convex set in $\mathbb{R}^n$ to itself admits one fixed point. This theorem surely includes the present case. However, when the set has "non-trivial topology", fixed points may not exist. For instance, take X to be $A = \{(x, y) : \ 1 \le x^2 + y^2 \le 4\}$ and T to be a rotation. It is clear that T has no fixed point in A. This is due to the topology of A, namely, it has a hole.

2.6 The Inverse Function Theorem

The Inverse Function Theorem and Implicit Function Theorem play a fundamental role in analysis and geometry. They illustrate the principle of linearization which is ubiquitous in mathematics. We learned these theorems in advanced calculus but the proofs were not emphasized. Now we fill out the gap. Adapting the notations in advanced calculus, a point $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ is sometimes called a vector and we use $|x|$ instead of $\|x\|_2$ to denote its Euclidean norm in this section.

All is about linearization. Recall that a real-valued function on an open interval I is differentiable at some $x_0 \in I$ if there exists some $a \in \mathbb{R}$ such that

$$\lim_{x\to x_0} \left| \frac{f(x) - f(x_0) - a(x - x_0)}{x - x_0} \right| = 0.$$

In fact, the value a is equal to $f'(x_0)$, the derivative of f at x_0. We can rewrite the limit above using the little o notation:

$$f(x_0 + z) - f(x_0) = f'(x_0)z + \circ(z), \quad \text{as } z \to 0.$$

Here $\circ(z)$ denotes a quantity satisfying $\lim_{z\to 0} \circ(z)/|z| = 0$. The same situation carries over to a real-valued function f in some open set in $\mathbb{R}^n$. A function f is called differentiable at p_0 in this open set if there exists a vector $a = (a_1, \cdots, a_n)$ such that

$$f(p_0 + z) - f(p_0) = \sum_{j=1}^{n} a_j z_j + \circ(z) \quad \text{as } z \to 0.$$

Again one can show that the vector a is uniquely given by the gradient vector of f at p_0

$$\nabla f(p_0) = \left(\frac{\partial f}{\partial x_1}(p_0), \cdots, \frac{\partial f}{\partial x_n}(p_0) \right).$$

More generally, a map F from an open set in $\mathbb{R}^n$ to $\mathbb{R}^m$ is called differentiable at a point p_0 in this open set if each component of $F = (f^1, \cdots, f^m)$ is differentiable. We can write

the differentiability condition collectively in the following form

$$F(p_0 + z) - F(p_0) = DF(p_0)z + o(z), \tag{2.3}$$

where $DF(p_0)$ is the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by

$$(DF(p_0)z)_i = \sum_{j=1}^{n} a_{ij}(p_0)x_j, \quad i = 1, \cdots, m,$$

where $\big(a_{ij}\big) = \big(\partial f^i/\partial x_j\big)$ is the Jabocian matrix of f. (2.3) shows near p_0, that is, when z is small, the function F is well-approximated by the linear map $DF(p_0)$ up to the constant $F(p_0)$ as long as $DF(p_0)$ is nonsingular. It suggests that the local information of a map at a differentiable point could be retrieved from its a linear map, which is much easier to analyse. This principle, called linearization, is widely used in analysis. The Inverse Function Theorem is a typical result of linearization. It asserts that a map is locally invertible if its linearization is invertible. Therefore, local bijectivity of the map is ensured by the invertibility of its linearization. When $DF(p_0)$ is not invertible, the first term on the right hand side of (2.3) may degenerate in some or even all direction so that $DF(p_0)z$ cannot control the error term $\circ(z)$. In this case the local behavior of F may be different from its linearization.

Theorem 2.19 (**Inverse Function Theorem**). *Let $F : U \to \mathbb{R}^n$ be a C^1-map where U is open in $\mathbb{R}^n$ and $p_0 \in U$. Suppose that $DF(p_0)$ is invertible. There exist open sets V and W containing p_0 and $F(p_0)$ respectively such that the restriction of F on V is a bijection onto W with a C^1-inverse. Moreover, the inverse is C^k when F is $C^k, 1 \le k \le \infty$, in U.*

Example 2.24. The Inverse Function Theorem asserts a local invertibility. Even if the linearization is non-singular everywhere, we cannot assert global invertibility. Let us consider the switching between the cartesian and polar coordinates in the plane:

$$x = r\cos\theta, \quad y = r\sin\theta\ .$$

The function $F : (0, \infty) \times (-\infty, \infty) \to \mathbb{R}^2$ given by $F(r, \theta) = (x, y)$ is a continuously differentiable function whose Jacobian matrix is non-singular except $(0, 0)$. However, it is clear that F is not bijective, for instance, all points $(r, \theta + 2n\pi), n \in \mathbb{Z}$, have the same image under F.

Example 2.25. An exceptional case is dimension one where a global result is available. Indeed, in Mathematical Analysis II we learned that if f is continuously differentiable on (a, b) with non-vanishing f', it is either strictly increasing or decreasing so that its global inverse exists and is again continuously differentiable.

Example 2.26. Consider the map $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by $F(x, y) = (x^2, y)$. Its Jacobian matrix is singular at $(0, 0)$. In fact, for any point $(a, b), a > 0$, $F(\pm\sqrt{a}, b) = (a, b)$. We cannot find any open set, no matter how small is, at $(0, 0)$ so that F is injective. On the other hand, the map $H(x, y) = (x^3, y)$ is bijective with inverse given by $J(x, y) = (x^{1/3}, y)$.

However, as the non-degeneracy condition does not hold at $(0,0)$ so it is not differentiable there. In these cases the Jacobian matrix is singular, so the nondegeneracy condition does not hold. We will see that in order the inverse map to be differentiable, the nondegeneracy condition must hold.

A map from some open set in R^n to $\mathbb{R}^m$ is C^k, $1 \leq k \leq \infty$ if all its components belong to C^k. It is called a C^∞-map or a smooth map if its components are C^∞.

The condition that $DF(p_0)$ is invertible, or equivalently the non-vanishing of the determinant of the Jacobian matrix, is called the nondegeneracy condition. Without this condition, the map may or may not be local invertible, see the examples below. Nevertheless, it is necessary for the differentiability of the local inverse. At this point, let us recall the general chain rule.

Let $G: \mathbb{R}^n \to \mathbb{R}^m$ and $F: \mathbb{R}^m \to \mathbb{R}^l$ be C^1 and their composition $H = F \circ G: \mathbb{R}^n \to \mathbb{R}^l$ is also C^1. We compute the first partial derivatives of H in terms of the partial derivatives of F and G. Letting $G = (g_1, \cdots, g_m)$, $F = (f_1, \cdots, f_l)$ and $H = (h_1, \cdots, h_l)$. From

$$h_k(x_1, \cdots, x_n) = f_k(g_1(x), \cdots, g_m(x)), \quad k = 1, \cdots, l,$$

we have

$$\frac{\partial h_k}{\partial y_i} = \sum_{i=1}^{n} \frac{\partial f_k}{\partial x_i} \frac{\partial g_i}{\partial x_j} \ .$$

Writing it in matrix form we have

$$DF(G(x))DG(x) = DH(x).$$

For, when the inverse is differentiable, we may apply this chain rule to differentiate the relation $F^{-1}(F(x)) = x$ to obtain

$$DF^{-1}(q_0)\ DF(p_0) = I\ , \quad q_0 = F(p_0),$$

where I is the identity map. We conclude that

$$DF^{-1}(q_0) = \big(DF(p_0)\big)^{-1},$$

in other words, the matrix of the derivative of the inverse map is precisely the inverse matrix of the derivative of the map. So when the inverse map is C^1, $DF(p_0)$ must be invertible.

Lemma 2.1. *Let L be a linear map from $\mathbb{R}^n$ to itself given by*

$$(Lz)_i = \sum_{j=1}^{n} a_{ij} z_j, \quad i = 1, \cdots n.$$

Then

$$|Lz| \leq \|L\|\ |z|, \quad \forall z \in \mathbb{R}^n,$$

where $\|L\| = \sqrt{\sum_{i,j} a_{ij}^2}$.

Proof. By Cauchy-Schwarz inequality,

$$\begin{aligned} |Lz|^2 &= \sum_i (Lz)_i^2 \\ &= \sum_i \big(\sum_j a_{ij} z_j\big)^2 \\ &\le \sum_i \big(\sum_j a_{ij}^2\big)\big(\sum_j z_j^2\big) \\ &= \|L\|^2 \, |z|^2 \, . \end{aligned}$$

□

Now we prove Theorem 2.19. We may take $p_0 = F(p_0) = 0$, for otherwise we could look at the new function $\overline{F}(x) = F(x+p_0) - F(p_0)$ instead of $F(x)$, after noting $D\overline{F}(0) = DF(p_0)$. First we would like to show that there is a unique solution for the equation $F(x) = y$ for y near 0. We will use the Contraction Mapping Principle to achieve our goal. After a further restriction on the size of U, we may assume that F is C^1 with $DF(x)$ invertible at all $x \in U$. For a fixed y, define the map in U by

$$T(x) = L^{-1}\left(Lx - F(x) + y\right)$$

where $L = DF(0)$. It is clear that any fixed point of T is a solution to $F(x) = y$. By the lemma,

$$\begin{aligned} |T(x)| &\leqslant \|L^{-1}\| \, |F(x) - Lx - y| \\ &\leqslant \|L^{-1}| \left(|F(x) - Lx| + |y|\right) \\ &\le \|L^{-1}\| \left(\left|\int_0^1 (DF(tx) - DF(0))dt \; x\right| + |y|\right), \end{aligned}$$

where we have used the formula

$$F(x) - DF(0)x = \int_0^1 \frac{d}{dt} F(tx)dt - DF(0) = \int_0^1 \big(DF(tx) - DF(0)\big)dt \; x,$$

after using the chain rule to get

$$\frac{d}{dt} F(tx) = DF(tx) \cdot x.$$

By the continuity of DF at 0, we can find a small ρ_0 such that

$$\|L^{-1}\| \|DF(x) - DF(0)\| \le \frac{1}{2}, \quad \forall x, \quad |x| \le \rho_0. \tag{2.4}$$

Then for for each y in $B_R(0)$, where R is chosen to satisfy $\|L^{-1}\|R \le \rho_0/2$, we have

$$\begin{aligned} |T(x)| &\le \|L^{-1}\| \left(\int_0^1 \|(DF(tx) - DF(0))dt\| |x| + |y|\right) \\ &\le \frac{1}{2}|x| + \|L^{-1}\||y| \\ &\le \frac{1}{2}\rho_0 + \frac{1}{2}\rho_0 = \rho_0, \end{aligned}$$

for all $x \in B_{\rho_0}(0)$. We conclude that T maps $\overline{B_{\rho_0}(0)}$ to itself. Moreover, for x_1, x_2 in $B_{\rho_0}(0)$, we have

$$\begin{aligned} |T(x_2) - T(x_1)| &= \left|L^{-1}\left(F(x_2) - Lx_2 - y\right) - L^{-1}\left(F(x_1) - Lx_1 - y\right)\right| \\ &\leqslant \|L^{-1}\|\ |F(x_2) - F(x_1) - DF(0)(x_2 - x_1)| \\ &\leqslant \|L^{-1}\|\ \left|\int_0^1 DF\left(x_1 + t(x_2 - x_1)\right)(x_2 - x_1)dt - DF(0)(x_2 - x_1)\right|, \end{aligned}$$

where we have used

$$\begin{aligned} F(x_2) - F(x_1) &= \int_0^1 \frac{d}{dt} F(x_1 + t(x_2 - x_1))dt \\ &= \int_0^1 DF(x_1 + t(x_2 - x_1))(x_2 - x_1)dt. \end{aligned}$$

Consequently,

$$|T(x_2) - T(x_1)| \leq \frac{1}{2}|x_2 - x_1|.$$

We have shown that $T : \overline{B_{\rho_0}(0)} \to \overline{B_{\rho_0}(0)}$ is a contraction. By the Contraction Mapping Principle, there is a unique fixed point for T, in other words, for each y in the ball $B_R(0)$ there is a unique point x in $\overline{B_{\rho_0}(0)}$ solving $F(x) = y$. Defining $G:\ B_R(0) \to \overline{B_{\rho_0}(0)} \subset X$ by setting $G(y) = x$, G is inverse to F.

Next, we claim that G is continuous. In fact, for $G(y_i) = x_i,\ i = 1, 2$, (not to be mixed up with the x_i above),

$$\begin{aligned} |G(y_2) - G(y_1)| &= |x_2 - x_1| \\ &= |T(x_2) - T(x_1)| \\ &\leq \|L^{-1}\|\left(|F(x_2) - F(x_1) - L(x_2 - x_1)| + |y_2 - y_1|\right) \\ &\leq \|L^{-1}\|\left(\left|\int_0^1 \big(DF((1-t)x_1 + tx_2) - DF(0)\big)dt(x_2 - x_1)\right| + |y_2 - y_1|\right) \\ &\leq \frac{1}{2}|x_2 - x_1| + \|L^{-1}\||y_2 - y_1| \\ &= \frac{1}{2}|G(y_2) - G(y_1)| + \|L^{-1}\||y_2 - y_1|, \end{aligned}$$

where (4.2) has been used. We deduce

$$|G(y_2) - G(y_1)| \leqslant 2\|L^{-1}\||y_2 - y_1|\ , \tag{2.5}$$

that's, G is continuous on $B_R(0)$.

Finally, let's show that G is a C^1-map in $B_R(0)$. In fact, for $y_1, y_1 + y$ in $B_R(0)$, using

$$\begin{aligned} y &= F(G(y_1 + y)) - F(G(y_1)) \\ &= \int_0^1 DF(G(y_1) + t(G(y_1 + y) - G(y_1))dt\ (G(y_1 + y) - G(y_1)), \end{aligned}$$

we have

$$G(y_1+y)-G(y_1)=DF^{-1}(G(y_1))y+R,$$

where R is given by

$$DF^{-1}(G(y_1))\int_0^1\Big(DF(G(y_1))-DF(G(y_1)+t(G(y_1+y)-G(y_1))\Big)(G(y_1+y)-G(y_1))dt.$$

As G is continuous and F is C^1, we have

$$G(y_1+y)-G(y_1)-DF^{-1}(G(y_1))y=\circ(1)(G(y_1+y)-G(y_1))$$

for small y. Using (2.5), we see that

$$G(y_1+y)-G(y_1)-DF^{-1}(G(y_1))y=\circ(\|y\|)\ ,$$

as $\|y\|\to 0$. We conclude that G is differentiable with derivative equal to $DF^{-1}(G(y_1))$.

After we have proved the differentiability of G, from the formula $DF(G(y))DG(y)=I$ where I is the identity matrix we see that

$$DF^{-1}(y)=(DF(F^{-1}(y)))^{-1},\quad \forall y\in B_R(0).$$

From linear algebra we know that $DF^{-1}(y)$ can be expressed as a rational function of the entries of the matrix of $DF(F^{-1}(y)$. Consequently, F^{-1} is C^k in y if F is C^k in x for $1\le k\le\infty$.

The proof of the Inverse Function Theorem is completed by taking $W=B_R(0)$ and $V=F^{-1}(W)$.

Remark 2.1. It is worthwhile to keep tracking and see how ρ_0 and R are determined. Indeed, let

$$M_{DF}(\rho)=\sup_{x\in B_\rho(0)}\|DF(x)-DF(0)\|$$

be the modules of continuity of DF at 0. We have $M_{DF}(\rho)\downarrow 0$ as $\rho\to 0$. From this proof we see that ρ_0 and R can be chosen as

$$M_{DF}(\rho_0)\le\frac{1}{2\|L^{-1}\|},\ \text{ and } R\le\frac{\rho_0}{2\|L^{-1}\|}\ .$$

Example 2.27. Consider the system of equations

$$\begin{cases} x-y^2=a,\\ x^2+y+y^3=b.\end{cases}$$

We know that $x = y = 0$ is a solution when $(a, b) = (a, b)$. Can we find the range of (a, b) so that this system is solvable? Well, let $F(x, y) = (x - y^2, x^2 + y + y^3)$. We have $F(0, 0) = (0, 0)$ and DF is given by the matrix

$$\begin{pmatrix} 1 & -2y \\ 2x & 1 + 3y^2 \end{pmatrix},$$

which is nonsingular at $(0, 0)$. In fact the inverse matrix of $DF((0, 0))$ is given by the identity matrix, hence $\|L^{-1}\| = 1$ in this case. According to Remark 4.1 a good ρ_0 could be found by solving $M_{DF}(\rho_0) = 1/2$. We have $\|DF((x, y)) - DF((0, 0))\| = 4y^2 + 4x^2 + 9y^2$, which, in terms of the polar coordinates, is equal to $4r^2 + 9\sin^4\theta$. Hence the maximal value is given by $4r^2 + 9r^4$, and so ρ_0 could be chosen to be any point satisfying $4\rho_0^2 + 9\rho_0^4 \leq 1/2$. A simple choice is $\rho_0 = \sqrt{1/26}$. Then R is given by $\sqrt{26}/52$. We conclude that whenever a, b satisfy $a^2 + b^2 \leq 1/104$, this system is uniquely solvable in the ball $B_{\rho_0}((0, 0))$.

Example 2.28. Determine all points where the function $F(x, y) = (xy^2 - \sin \pi x, y^2 - 25x^2 + 1)$ has a local inverse and find the partial derivatives of the inverse. Well, the Jacobian matrix of F is given by

$$\begin{pmatrix} y^2 - \pi \cos \pi x & 2xy \\ -50x & 2y \end{pmatrix}.$$

Hence, F admits a local inverse at points (x, y) satisfying

$$2y(y^2 - \pi \cos \pi x) + 100x^2 y \neq 0 \ .$$

Derivatives of the inverse function, denoted by $G = (g_1, g_2)$, can be obtained by implicit differentiation of the relation

$$(u, v) = F(G(u, v)) = (g_1 g_2^2 - \sin \pi g_1, g_2^2 - 25g_1^2 + 1),$$

where g_1, g_2 are functions of (u, v). We have

$$\frac{\partial g_1}{\partial u} g_2^2 + 2g_1 g_2 \frac{\partial g_2}{\partial u} - \pi \cos \pi g_1 \ \frac{\partial g_1}{\partial u} = 1,$$

$$2g_2 \frac{\partial g_2}{\partial u} - 50 g_1 \frac{\partial g_1}{\partial u} = 0,$$

$$\frac{\partial g_1}{\partial v} g_2^2 + 2g_1 g_2 \frac{\partial g_2}{\partial v} - \pi \cos \pi g_1 \ \frac{\partial g_1}{\partial v} = 0,$$

$$2g_2 \frac{\partial g_2}{\partial v} - 50 g_1 \frac{\partial g_1}{\partial v} = 1.$$

The first and the second equations form a linear system for $\partial g_i / \partial u, i = 1, 2$, and the third and the fourth equations form a linear system for $\partial g_i / \partial v, \ i = 1, 2$. By solving it (the

solvability is ensured by the invertibility of the Jacobian matrix) we obtain the partial derivatives of the inverse function G. Nevertheless, it is too tedious to carry it out here. An alternative way is to find the inverse matrix of the Jacobian DF. In principle we could obtain all partial derivatives of G by implicit differentiation and solving linear systems.

Inverse Function Theorem may be rephrased in the following form.

A C^k-map F between open sets V and W is a "C^k-diffeomorphism" if F^{-1} exists and is also C^k. Let $f_1, f_2, \cdots, f_n$ be C^k-functions defined in some open set in $\mathbb{R}^n$ whose Jacobian matrix of the map $F = (f_1, \cdots, f_n)$ is non-singular at some point p_0 in this open set. By Theorem 4.1 F is a C^k-diffeomorphism between some open sets V and W containing p_0 and $F(p_0)$ respectively. To every function Φ defined in W, there corresponds a function defined in V given by $\Psi(x) = \Phi(F(x))$, and the converse situation holds. Thus every C^k-diffeomorphism gives rise to a "local change of coordinates".

Next we deduce Implicit Function Theorem from Inverse Function Theorem.

Theorem 2.20 (**Implicit Function Theorem**). *Consider C^1-map $F : U \to \mathbb{R}^m$ where U is an open set in $\mathbb{R}^n \times \mathbb{R}^m$. Suppose that $(p_0, q_0) \in U$ satisfies $F(p_0, q_0) = 0$ and $D_y F(p_0, q_0)$ is invertible in $\mathbb{R}^m$. There exist an open set $V_1 \times V_2$ in U containing (p_0, q_0) and a C^1-map $\varphi : V_1 \to V_2$, $\varphi(p_0) = q_0$, such that*

$$F(x, \varphi(x)) = 0 \ , \qquad \forall x \in V_1 \ .$$

The map φ belongs to C^k when F is $C^k, 1 \le k \le \infty$, in U. Moreover, if ψ is another C^1-map in some open set containing p_0 to V_2 satisfying $F(x, \psi(x)) = 0$ and $\psi(p_0) = q_0$, then ψ coincides with φ in their common set of definition.

The notation $D_y F(p_0, q_0)$ stands for the linear map associated to the Jocabian matrix $(\partial F_i / \partial y_j(p_0, q_0))_{i,j=1,\cdots,m}$ where p_0 is fixed.

Proof. Consider $\Phi : U \to \mathbb{R}^n \times R^m$ given by

$$\Phi(x, y) = (x, F(x, y)).$$

It is evident that $D\Phi(x, y)$ is invertible in $\mathbb{R}^n \times \mathbb{R}^m$ when $D_y F(x, y)$ is invertible in $\mathbb{R}^m$. By the Inverse Function Theorem, there exists a C^1-inverse $\Psi = (\Psi_1, \Psi_2)$ from some open W in $\mathbb{R}^n \times \mathbb{R}^m$ containing $(p_0, 0)$ to an open subset of U. By restricting W further we may assume $\Psi(W)$ is of the form $V_1 \times V_2$. For every $(x, z) \in W$, we have

$$\Phi(\Psi_1(x, z), \Psi_2(x, z)) = (x, z),$$

which, in view of the definition of Φ, yields

$$\Psi_1(x, z) = x, \text{ and } F((\Psi_1(x, z), \Psi_2(x, z)) = z.$$

In other words, $F(x, \Psi_2(x, z)) = z$ holds. In particular, taking $z = 0$ gives

$$F(x, \Psi_2(x, 0)) = 0, \quad \forall x \in V_1 \ ,$$

so the function $\varphi(x) \equiv \Psi_2(x, 0)$ satisfies our requirement.

By restricting V_1 and V_2 further if necessary, we may assume the matrix

$$\int_0^1 D_y F(x, y_1 + t(y_2 - y_1)dt$$

is nonsingular for $(x, y_1), (x, y_2) \in V_1 \times V_2$. Now, suppose ψ is a C^1-map defined near x_0 satisfying $\psi(p_0) = q_0$ and $F(x, \psi(x)) = 0$. We have

$$\begin{aligned} 0 &= F(x, \psi(x)) - F(x, \varphi(x)) \\ &= \int_0^1 D_y F(x, \varphi(x) + t(\psi(x) - \varphi(x))dt(\psi(x) - \varphi(x)), \end{aligned}$$

for all x in the common open set they are defined. This identity forces that ψ coincides with φ in this open set. The proof of the implicit function is completed, once we observe that the regularity of φ follows from Inverse Function Theorem. □

Example 2.29. Let $F : \mathbb{R}^5 \to \mathbb{R}^2$ be given by $F(x, y, z, u, v) = (xy^2 + xzu + yv^2 - 3, u^3yz + 2xv - u^2v^2 - 2)$. We have $F(1, 1, 1, 1, 1) = (0, 0)$. Show that there are functions $f(x, y, z), g(x, y, z)$ satisfying $f(1, 1, 1) = g(1, 1, 1) = 1$ and $F(x, y, z, f(x, y, z), g(x, y, z)) = (0, 0)$ for (x, y, z) near $(1, 1, 1)$. We compute the "partial" Jacobian matrix of F in (u, v):

$$\begin{pmatrix} xz & 2yv \\ 3u^2yz - 2uv^2 & 2x - 2u^2v \end{pmatrix}.$$

Its determinant at $(1, 1, 1, 1, 1)$ is equal to -2, so we can apply Implicit Function Theorem to get the desired result. The partial derivatives of f and g can be obtained by implicit differentiations. For instance, to find $\partial f/\partial y$ and $\partial g/\partial y$ we differentiate the relation

$$(xy^2 + xzf + yg^2 - 3, f^3yz + 2xg - f^2g^2 - 2) = (0, 0)$$

to get

$$2xy + xz\frac{\partial f}{\partial y} + g^2 + 2yg\frac{\partial g}{\partial y} = 0,$$

and

$$f^3z + 3f^2yz\frac{\partial f}{\partial y} + 2x\frac{\partial g}{\partial y} - 2fg^2\frac{\partial f}{\partial y} - 2f^2g\frac{\partial g}{\partial y} = 0.$$

By solving this linear system we can express $\partial f/\partial y$ and $\partial g/\partial y$ in terms of x, y, z, f and g. Similarly we can do it for the other partial derivatives.

It is interesting to note that the Inverse Function Theorem can be deduced from Implicit Function Theorem. Thus they are equivalent. To see this, keeping the notations used in Theorem 2.19. Define a map $\widetilde{F} : U \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\widetilde{F}(x, y) = F(x) - y.$$

Then $\widetilde{F}(p_0, q_0) = 0, q_0 = F(p_0)$, and $D\widetilde{F}(p_0, q_0)$ is invertible. By Theorem 2.20, there exists a C^1-function φ from near q_0 satisfying $\varphi(q_0) = p_0$ and $\widetilde{F}(\varphi(y), y) = F(\varphi(y)) - y = 0$, hence φ is the local inverse of F.

2.7 Picard-Lindelöf Theorem for Differential Equations

In this section we discuss the fundamental existence and uniqueness theorem for differential equations. I assume that you learned the skills of solving ordinary differential equations already so we will focus on the theoretical aspects.

Most differential equations cannot be solved explicitly, in other words, they cannot be expressed as the composition of elementary functions. Nevertheless, there are two exceptional classes which come up very often. Let us review them before going into the theory. The first one is linear equation.

$$\frac{dx}{dt} = a(t)x + b(t),$$

where a and b are continuous functions defined on some interval I. The general solution of this linear equation is given by the formula

$$x(t) = e^{A(t)} \left(x_0 + \int_{t_0}^{t} e^{-A(s)} b(s) ds \right), \quad A(t) = \int_{t_0}^{t} a(s) ds,$$

where $t_0 \in I, x_0 \in \mathbb{R}$, are arbitrary. The second class is the so-called separable equation

$$\frac{dx}{dt} = \frac{f(t)}{g(x)},$$

where f and $g \neq 0$ are continuous functions on intervals I and J respectively. Then the solution can be obtained by an integration

$$\int_{x_0}^{x} g(z) dz = \int_{t_0}^{t} f(s) ds, \quad t_0 \in I,\ x_0 \in J.$$

The resulting relation, written as $G(x) = F(t)$, can be converted into $x = G^{-1}F(t)$, a solution to the equation as immediately verified by the chain rule. These two classes of

equations are sufficient for our purpose. More interesting explicitly solvable equations can be found in texts on ODE's.

Well, let us consider the general situation. Numerous problems in natural sciences and engineering led to the initial value problem of differential equations. Let f be a function defined in the rectangle $R = [t_0 - a, t_0 + a] \times [x_0 - b, x_0 + b]$ where $(t_0, x_0) \in \mathbb{R}^2$ and $a, b > 0$. We consider the initial value problem or Cauchy Problem

$$\begin{cases} \dfrac{dx}{dt} = f(t, x), \\ x(t_0) = x_0. \end{cases} \tag{2.6}$$

(In some books the independent variable t is replaced by x and the dependent variable x is replaced by y. We prefer to use t instead of x as the independent variable in many cases is the time.) To solve the Cauchy Problem it means to find a function $x(t)$ defined in a perhaps smaller rectangle, that is, $x : [t_0 - a', t_0 + a'] \to [x_0 - b, x_0 + b]$, which is differentiable and satisfies $x(t_0) = x_0$ and $x'(t) = f(t, x(t)),\ \forall t \in [t_0 - a', t_0 + a']$, for some $0 < a' \leq a$. In general, no matter how nice f is, we do not expect there is always a solution on the entire $[t_0 - a, t_0 + a]$. Let us look at the following example.

Example 2.30. Consider the Cauchy Problem

$$\begin{cases} \dfrac{dx}{dt} = 1 + x^2, \\ x(0) = 0. \end{cases}$$

The function $f(t, x) = 1 + x^2$ is smooth on $[-a, a] \times [-b, b]$ for every $a, b > 0$. However, the solution, as one can verify immediately, is given by $x(t) = \tan t$ which is only defined on $(-\pi/2, \pi/2)$. It shows that even when f is very nice, a' could be strictly less than a.

The Picard-Lindelöf theorem, sometimes referred to as the fundamental theorem of existence and uniqueness of differential equations, gives a clean condition on f ensuring the unique solvability of the Cauchy Problem (2.6). This condition imposes a further regularity condition on f reminding what we did in the convergence of Fourier series. Specifically, a function f defined in R satisfies the **Lipschitz condition** (uniform in t) if there exists some $L > 0$ such that $\forall (t, x_i) \in R \equiv [t_0 - a, t_0 + a] \times [x_0 - b, x_0 + b],\ i = 1, 2,$

$$|f(t, x_1) - f(t, x_2)| \leq L\,|x_1 - x_2|\,.$$

Note that in particular means for each fixed t, f is Lipschitz continuous in x. The constant L is called a **Lipschitz constant.** Obviously if L is a Lipschitz constant for f, any number greater than L is also a Lipschitz constant. Not all continuous functions satisfy the Lipschitz condition. An example is given by the function $f(t, x) = tx^{1/2}$ is continuous. I let you verify that it does not satisfy the Lipschitz condition on any rectangle containing the origin.

In application, most functions satisfying the Lipschitz condition arise in the following manner. A C^1-function $f(t,x)$ in a closed rectangle automatically satisfies the Lipschitz condition. For, by the mean-value theorem, for some z lying on the segment between x_1 and x_2,

$$f(t,x_2) - f(t,x_1) = \frac{\partial f}{\partial x}(t,z)(x_2 - x_1).$$

Letting

$$L = \max\left\{ \left| \frac{\partial f}{\partial x}(t,x) \right| : \ (t,x) \in R \right\},$$

(L is a finite number because $\partial f/\partial y$ is continuous on R and hence bounded), we have

$$|f(t,x_2) - f(t,x_1)| \le L|x_2 - x_1|, \quad \forall (t,x_i) \in R, \ i = 1,2.$$

Theorem 2.21 (Picard-Lindelöf Theorem). *Consider (2.6) where $f \in C(R)$ satisfies the Lipschitz condition on $R = [t_0 - a, t_0 + a] \times [x_0 - b, x_0 + b]$. There exist $a' \in (0,a)$ and $x \in C^1[t_0 - a', t_0 + a']$, $x_0 - b \le x(t) \le x_0 + b$ for all $t \in [t_0 - a', t_0 + a']$, solving (2.6). Furthermore, x is the unique solution in $[t_0 - a', t_0 + a']$.*

From the proof one will see that $a' \in (0,a)$ can be taken to be any number satisfying

$$0 < a' < \min\left\{ \frac{b}{M}, \frac{1}{L} \right\},$$

where $M = \sup\{|f(t,x)| : \ (t,x) \in R\}$.

To prove Pircard-Lindelöf Theorem, we first convert (2.6) into a single integral equation.

Proposition 2.22. *Setting as in Theorem 2.21, every solution x of (2.6) from $[t_0 - a', t_0 + a']$ to $[x_0 - b, x_0 + b]$ satisfies the equation*

$$x(t) = x_0 + \int_{t_0}^{t} f(t, x(t))\, dt. \tag{2.7}$$

Proof. When x satisfies $x'(t) = f(t,x(t))$ and $x(t_0) = x_0$, (2.7) is a direct consequence of the Fundamental Theorem of Calculus (first form). Conversely, when $x(t)$ is continuous on $[t_0 - a', t_0 + a']$, $f(t,x(t))$ is also continuous on the same interval. By the Fundamental Theorem of Calculus (second form), the left hand side of (2.7) is continuously differentiable on $[t_0 - a', t_0 + a']$ and solves (2.6). □

Note that in this proposition we do not need the Lipschitz condition; only the continuity of f is needed.

Proof of Picard-Lindelöf Theorem. Instead of solving (2.6) directly, we look for a solution of (2.7). We will work on the metric space $X = \{\varphi \in C[t_0 - a', t_0 + a'] : \varphi(t) \in [x_0 - b, x_0 + b], \varphi(t_0) = x_0\}$ with the uniform metric. It is easily verified that it is a closed subset in the complete metric space $C[t_0 - a', t_0 + a']$ and hence complete. Recall that every closed subset of a complete metric space is complete. The number a' will be specified below.

We are going to define a contraction on X. Indeed, for $x \in X$, define T by

$$(Tx)(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds.$$

First of all, for every $x \in X$, it is clear that $f(t, x(t))$ is well-defined and $Tx \in C[t_0 - a', t_0 + a']$. To show that it is in X, we need to verify $x_0 - b \leq (Tx)(t) \leq x_0 + b$ for all $t \in [t_0 - a', t_0 + a']$. We claim that this holds if we choose a' satisfying $a' \leq b/M$, $M = \sup\{|f(t, x)| : (t, x) \in R\}$. For,

$$\begin{aligned} |(Tx)(t) - x_0| &= \left| \int_{t_0}^{t} f(t, x(t))\, dt \right| \\ &\leq M\, |t - t_0| \\ &\leq M a' \\ &\leq b. \end{aligned}$$

Next, we claim T is a contraction on X when a' is further restricted to $a' \leq \gamma/L$, where $\gamma \in (0, 1)$ and L is the Lipschitz constant for f. For,

$$\begin{aligned} |(Tx_2 - Tx_1)(t)| &= \left| \int_{t_0}^{t} f(t, x_2(t)) - f(t, x_1(t))\, dt \right| \\ &\leq \int_{t_0}^{t} \left| f(t, x_2(t)) - f(t, x_1(t)) \right| dt \\ &\leq L \int_{t_0}^{t} |x_2(t) - x_1(t)|\, dt \\ &\leq L \sup_{t \in I} |x_2(t) - x_1(t)|\, |t - t_0| \\ &\leq L a' \sup_{t \in I} |x_2(t) - x_1(t)| \\ &\leq \gamma \sup_{t \in I} |x_2(t) - x_1(t)|, \end{aligned}$$

where $I = [t_0 - a', t_0 + a']$. It follows that

$$d_\infty(Tx_2, Tx_1) \leq \gamma d_\infty(x_2, x_1)$$

where d_∞ is the uniform metric $d_\infty(f, g) \equiv \|f - g\|_\infty$ for $f, g \in C[t_0 - a', t_0 + a']$. Now we can apply the Contraction Mapping Principle to conclude that $Tx = x$ for some x, and x solves (2.6). We have shown that (2.6) admits a solution in $[t_0 - a', t_0 + a']$ where a' can be chosen to be any number less than $\min\{b/M, 1/L\}$. □

Next we discuss how to obtain a unique, maximal solution under a local Lipschitz condition. We start with a unique extension result. Proposition 2.23 to Theorem 2.28 are for optional reading.

Proposition 2.23. *Let x_1 and x_2 be solutions to (2.6) on open intervals I_1 and I_2 containing t_0 respectively. Under the Lipschitz condition on f, x_1 and x_2 coincide on $I_1 \cap I_2$. Therefore, the function y, which is equal to x_1 on I_1 and x_2 on I_2, is a solution to (2.6) on $I_1 \cup I_2$.*

Proof. Let $J = I_1 \cap I_2 \equiv (\alpha, \beta)$ and set $s = \sup\{t : \ x_1 \equiv x_2 \text{ on } [t_0, t]\}$. We claim that $s = \beta$. For, if $s < \beta$, by continuity $x_1(s) = x_2(s)$. For $t \in (\alpha, \beta)$, we have

$$\begin{aligned} |x_1(t) - x_2(t)| &= \left| \int_s^t f(\tau, x_1(\tau)) - f(\tau, x_2(\tau))\, d\tau \right| \\ &\le L \int_s^t |x_1(\tau) - x_2(\tau)| \ d\tau. \end{aligned}$$

Let $t \in J = [s - 1/(2L), s + 1/(2L)]$. By further enlarging L if necessary, we may assume that $J \subset (\alpha, \beta)$. Let t_1 satisfy $|x_1(t_1) - x_2(t_1)| = \max_{x \in J} |x_1(t) - x_2(t)|$. Then

$$\begin{aligned} \max_J |x_1(t) - x_2(t)| &= |x_1(t_1) - x_2(t_1)| \\ &\le L \max_J |x_1(t) - x_2(t)|\, |t_1 - s| \\ &\le \frac{1}{2} \max_J |x_1(t) - x_2(t)|\,, \end{aligned}$$

which forces $x_1 \equiv x_2$ on $[s - 1/(2L), s + 1/(2L)]$. It means that x_1 and x_2 coincide on $[t_0, z + 1/2L]$, contradicting the definition of z. Hence x_1 and x_2 must coincide on $[t_0, \beta)$. A similar argument shows that they coincide on $(\alpha, t_0]$. □

A consequence of Theorem 2.21 and Proposition 2.23 is the existence of a maximal solution. To describe it is convenient to enter a new definition. A function f defined in an open set in $\mathbb{R}^2$ is called to satisfy the **local Lipschitz condition** if it satisfies the Lipschitz condition on every compact subset of G. Here the Lipschtiz constants depend on the compact subset. It is common that they becomes larger and larger as the compact subsets swallow the open set G. If you feel the definition a bit complicated, we could put it the following way. We observe that every open set in $\mathbb{R}^n$ can be written as the countable union of compact subsets. Indeed, the subsets

$$K_n = \{(t, x) \in G : \ \text{dist}((t, x), \partial G) \ge 1/n\} \cap \overline{B_n(0)}, \ n \ge 1,$$

are compact and $G = \bigcup_{n=1}^{\infty} K_n$. Clearly every compact subset is contained in some K_n for sufficiently large n. With this understanding, the Lipschitz condition on f may be recast as, there exist $L_n, n \ge 1$, such that

$$\big|f(t, x_2) - f(t, x_1)\big| \le L_n |x_1 - x_1|, \quad \forall (t, x_1), \ (t, x_2) \in K_n.$$

In the exercise you are asked to show that a C^1-function defined in an open set satisfies the local Lipschitz condition.

Theorem 2.24. *Consider (2.6) where f is a continuous function on the open set G satisfying the local Lipschitz condition. Then there exists a solution x^* to (2.6) defined on some (α, β) satisfying*

(a) Whenever x is a solution of (2.6) on some interval I, $I \subset (\alpha, \beta)$ and $x = x^*$ on I .

(b) If β is finite, the solution escapes from every compact subset of G eventually. Similar results holds at α.

"The solution escapes from every compact subset of G eventually" means, for each compact $K \subset G$, there exists a small $\delta > 0$ such that $(t, x^*(x)) \in G \setminus K$ for $t \in [\beta - \delta, \beta)$. When G is $\mathbb{R}^2$, it means that x^* either tends to ∞ or $-\infty$ as t approaches β when β is finite. When β is infinite, the solution could tend to positive or negative infinity or oscillate up and down infinitely as x goes to ∞. In contrast, when β is finite, the solution could either goes to ∞ or $-\infty$ approaching β. The case of oscillation is excluded.

In view of this theorem, it is legal to call this *maximal solution the* solution of (2.6) and the interval (α, β) the maximal interval of existence.

Proof. Let $\mathcal{I}$ be the collection of all closed, bounded intervals I containing t_0 over which a solution of (2.6) exists and let I^* be the union of the intervals in $\mathcal{I}$. Clearly I^* is again an interval, denote its left and right endpoints by α and β respectively. By Proposition 2.20 there is a solution x^* of (2.6) defined on (α, β). When β is finite, let us show that the solution escapes from every compact subset eventually. Let K be a compact subset of G and suppose on the contrary that there exists $\{t_k\} \subset (\alpha, \beta)$, $t_k \to \beta$, but $(t_k, x^*(t_k)) \in K$ for all k. By compactness, we may assume $x^*(t_k)$ converges to some z in K (after passing to a subsequence if necessary). Since $\operatorname{dist}((\beta, z), \partial G) > 0$, we can find a rectangle $[\beta - r, \beta + r] \times [z - \rho, z + \rho]$ inside G. Then, as this is a compact subset, f satisfies the Lipschitz condition on this rectangle. By Theorem 2.21, we could use $(t_k, x^*(x_k))$ as the initial data to solve (2.6). The range of this solution would be some interval $[t_k - r', t_k + r']$ where r' is independent of k. Since t_k approaches β, for large k $\beta \in [t_k - r', t_k + r']$. But then by Proposition 2.23, the solution x^* can be extended beyond β, contraction holds. We conclude that the solution must escape from any compact subset eventually. A similar argument applies to the left endpoint α.

□

Example 2.31. Consider

$$f(t, x) = \frac{t}{1 - x}, \qquad (t, x) \in G \equiv (-\infty, \infty) \times (-\infty, 1)$$

and $t_0 = x_0 = 0$ in (2.6). Since $f \in C^1(G)$, the setting of Theorem 2.21 is satisfied. (Why?) This equation is separable and the solution is readily found to be

$$x(t) = 1 - \sqrt{1 - t^2}.$$

So the maximal interval of existence is given by $(-1, 1)$. As $t \to \pm 1$, $(t, x(t))$ hits the horizontal line $x = 1$ as asserted by Theorem 2.24.

There are two comparison principles which enable us to study the global existence or finite time blow-up of the maximal solution of the first order equation. In the first principle we compare solutions of the same equation with different data.

Proposition 2.25. *Consider (2.6) where f satisfies the Lipschitz condition in every compact subset of some open set G and $(t_0, a_1), (t_0, a_2) \in G$. Let $x_i, i = 1, 2$, be the solutions of (2.6) starting from a_i. If $a_1 < a_2$, then $x_1(t) < x_2(t)$ for $t > t_0$ as long as both solutions exist.*

Proof. By continuity, as $a_1 < a_2$, there exists some interval containing t_0 such that $x_1 < x_2$. Suppose that there exists some $t_1 > t_0$ such that $x_1(t_1) = x_2(t_1)$ and $x(t) < x_2(t),\ t \in [t_0, t_1)$. We can find a small, closed rectangle R in G containing (t_1, x_1) such that f satisfies the Lipschitz condition in R. Applying Proposition 2.23 to the equation by taking t_1 as the initial data, we conclude that x_1 and x_2 coincide on some open interval containing t_1, contradiction holds. Hence x_2 is always greater than x_1 as long as both of them exist. □

A special case is where there exists some γ such that $f(\gamma) = 0$ for all t (here f is independent of t). Then the constant $x(t) \equiv \gamma$ is a solution with initial data $x(t_0) = \gamma$. Such constant solutions are called steady states of the Cauchy Problem.

Corollary 2.26. *Let γ be a steady state of f which satisfies the Lipschitz condition locally in G and x is a solution of (2.6) satisfying $x(t_0) < \gamma$ (resp. $x(t_0) > \gamma$). Then $x(t) < \gamma$ (resp. $x(t) > \gamma$) as long as x exists.*

Example 2.32. Consider (2.6) where $f(x) = \alpha x(M - x),\ \alpha, M > 0$. This is a logistic model for the growth of population. The solution $x(t)$ gives the population of a species at time t. Obviously there are two steady states, namely, 0 and M. Therefore, any solution starting from $x(0) \in (0, M)$ is bounded between 0 and M. By Theorem 2.24 it exists for all time. In fact, as f is positive on $(0, M)$, the solution keeps increasing and it is easy to argue that it tends to M as $t \to \infty$. We call M an attracting steady state and 0 a repelling state. Incidentally, we point out that this equation is separable and the solution is given explicitly by

$$x(t) = \frac{Me^{M\alpha(t-t_0)+C}}{1 + e^{M\alpha(t-t_0)+C}}.$$

The main point here is that the behavior of the maximal solution can be studied without using the explicit expression.

We can also compare solutions of different equations.

Proposition 2.27. *Let $f_i, i = 1, 2$, be continuous in G and $f_1(t, x) \le f_2(t, x)$ for all $(x, t) \in G$. Further suppose that $f_2(\cdot, x)$ is increasing in x. Let $x_i, i = 1, 2$, be respectively the solution of (2.6) corresponding to f_i with $x_1(t_0) < x_2(t_0)$. Then $x_1(t) < x_2(t),\ t > t_0$, as long both solutions exist.*

Proof. By continuity $x_2 > x_1$ for t close to t_0. Suppose that there is some time they coincide. Let the first such time be t_1. We have $x_1(t_1) = x_2(t_1)$ and $x_1(t) < x_2(t)$, $t \in [t_0, t_1)$. Therefore,

$$\begin{aligned} 0 &= x_2(t_1) - x_1(t_1) \\ &= x_2(t_0) - x_1(t_0) + \int_{t_0}^{t_1} (f_2(s, x_2(s)) - f_1(s, x_1(s)))\, ds \\ &\geq x_2(t_0) - x_1(t_0) + \int_{t_0}^{t_1} (f_2(s, x_1(s)) - f_1(s, x_1(s)))\, ds \\ &\geq x_2(t_0) - x_1(t_0) > 0, \end{aligned}$$

contradiction holds. □

Example 2.33. Let f satisfy the local Lipschitz condition and the "sublinear growth condition"

$$|f(t, x)| \leq C(1 + |x|), \quad (x, t) \in \mathbb{R}^2.$$

We are going to show that the Cauchy Problem for f always admit a solution in $(-\infty, \infty)$. The idea is to dominate f by some linear function. Indeed, it suffices to consider the function $g(t, x) = C(1 + x)$. The solution satisfying $x(t_0) = \gamma$ is given explicitly by

$$y(t) = (1 + \gamma)e^{C(t-t_0)} - 1, \quad t \in (-\infty, \infty).$$

By taking $\gamma > x(t_0)$, we see that $x(t) < y(t)$ for all $t \in [t_0, \infty)$. It follows that x cannot blow up to ∞ at any finite time greater than t_0. Next consider the solution z of (2.6) for the function $h(t, x) = -C(1 + x)$ satisfying $z(t_0) < x(t_0)$. Then $x(t) > z(t)$ for t greater than t_0. It shows that x cannot blow up to $-\infty$ in any finite time beyond t_0. A similar argument shows that x cannot blow up in any finite time less than t_0.

We point out that the existence part of Picard-Lindelöf Theorem still holds without the Lipschitz condition. We will prove this in the next chapter. However, the solution may not be unique.

Example 2.34. Consider the Cauchy Problem $x' = |x|^\alpha$, $\alpha \in (0, 1)$, $x(0) = 0$. The function $f(x) = |x|^\alpha$ is Hölder continuous but not Lipschitz continuous. While $x_1 \equiv 0$ is a solution,

$$x_2 = (1 - \alpha)^{\frac{1}{1-\alpha}} |t|^{\frac{1}{1-\alpha}}$$

is also a solution. In fact, there are infinitely many solutions! Can you write them down?

Theorem 2.21, Propositions 2.22 –2.24 are valid for systems of differential equations. Without making things too clumsy, we put all results in a single theorem. First of all, the Cauchy Problem for systems of differential equations is

$$\begin{cases} \dfrac{dx_j}{dt} = f_j(t, x_1, x_2, \cdots, x_N), \\ x_j(t_0) = x_{j0}, \end{cases}$$

where $j = 1, 2, \cdots, N$. By setting $x = (x_1, x_2, \cdots, x_N)$ and $f = (f_1, f_2, \cdots, f_N)$, we can express it as in (2.3) but now both x and f are vectors.

Essentially following the same arguments as the case of a single equation, we have

Theorem 2.28 (**Picard-Lindelöf Theorem for Systems**). *Consider (2.6) where f satisfies the Lipschitz condition in every compact subset of the open set $G \subset \mathbb{R} \times \mathbb{R}^n$. There exists a solution x^* of (2.6) on (α, β) such that*

1. *If x is a solution of (2.6) on some interval I, then $I \subset (\alpha, \beta)$ and x is equal to x^* on I;*
2. *If $\beta < \infty$, then x^* escapes from any compact subset of G as t approaches β. Similar situation holds at α.*

Note that now the Lipschitz condition on f should be interpreted as

$$d_2(f(t, x_1), f(t, x_2)) \leq L d_2(x_1, x_2), \quad \forall t \in [t_0 - a, t_0 + a].$$

Finally, we remind you that there is a standard way to convert the Cauchy Problem for higher order differential equation ($m \geq 2$)

$$\begin{cases} x^{(m)} = f(t, x, x', \cdots, x^{(m-1)}), \\ x(t_0) = x_0, \ x'(t_0) = x_1, \cdots, x^{(m-1)}(t) = x_{m-1}, \end{cases}$$

into a system of first order differential equations. As a result, we also have a corresponding Picard-Lindelöf theorem for higher order differential equations as well as the existence of a maximal solution. I will let you formulate these results.

Comments on Chapter 2. A topology on a set X is a collection of sets τ consisting the empty set and X itself which is closed under arbitrary union and finite intersection. Each set in τ is called an open set. The pair (X, τ) is called a topological space. From Proposition 2.2 we see that the collection of all open sets in a metric space (X, d) forms a topology on X. This is the topological space induced by the metric. Metric spaces constitute a large class of topological spaces, but not every topological space comes from a metric. However, from the discussions in Section 2 we know that continuity can be defined solely in terms of open sets. It follows that continuity can be defined for topological spaces, and this is crucial for many further developments. In the past, metric spaces were covered in Introduction to Topology. Feeling that the notion of a metric space should be

learned by every math major, we move it here.

Wiki gives a nice summary of metric spaces under "metric space". Bolzano-Weierstrass Theorem states that every sequence in a closed, bounded set in $\mathbb{R}^n$ has a convergent subsequence. Heine-Borel Theorem says every open cover of a closed, bounded set in $\mathbb{R}^n$ admits a finite subcover. Bolzano-Weierstrass Theorem has motivated our definition of a compact set in a metric space, and yet Theorem 2.13 shows that one can also use the description in Heine-Borel Theorem to define compactness.

Many theorems in finite dimensional space are extended to infinite dimensional normed spaces when the underlying closed, bounded set is replaced by a compact set. Thus it is extremely important to study compact sets in a metric space. We will study compact sets in $C[a, b]$ in Chapter 3. A theorem of Arzela-Ascoli provides a complete characterization of compact sets in the space $C[a, b]$.

There are two popular constructions of the real number system, Dedekind cuts and Cantor's Cauchy sequences. Although the number system is fundamental in mathematics, we did not pay much attention to its rigorous construction. It is too dry and lengthy to be included in Mathematical Analysis I. Indeed, there are two sophisticate steps in the construction of real numbers from nothing, namely, the construction of the natural numbers by Peano's axioms and the construction of real numbers from rational numbers. Other steps are much easier. Cantor's construction of the irrationals from the rationals is very much like the proof of Theorem 2.15. You may google under the key words "Peano's axioms, Cantor's construction of the real numbers, Dedekind cuts" for more.

Contraction Mapping Principle, or Banach Fixed Point Theorem, was found by the Polish mathematician S. Banach (1892-1945) in his 1922 doctoral thesis. He is the founder of functional analysis and operator theory. According to P. Lax, "During the Second World War, Banach was one of a group of people whose bodies were used by the Nazi occupiers of Poland to breed lice, in an attempt to extract an anti-typhoid serum. He died shortly after the conclusion of the war." The interested reader should look up his biography at Wiki.

An equally famous fixed point theorem is Brouwer's Fixed Point Theorem. It states that every continuous map from a closed ball in $\mathbb{R}^n$ to itself admits at least one fixed point. Here it is not the map but the geometry, or more precisely, the topology of the ball matters. You will learn it in a course on topology.

Inverse and Implicit Function Theorems, which reduce complicated structure to simpler ones via linearization, are the most frequently used tool in the study of the local behavior of maps. We learned these theorems and some of its applications in Advanced

Calculus I already. In view of this, we basically provide detailed proofs here but leave out many standard applications. You may look up Fitzpatrick, "Advance Calculus", to refresh your memory. By the way, the proof in this book does not use Contraction Mapping Principle. I do know a third proof besides these two.

Picard-Lindelöf Theorem or the fundamental existence and uniqueness theorem of differential equations was mentioned in Ordinary Differential Equations and now its proof is discussed in details. Of course, the contributors also include Cauchy and Lipschitz. Further results without the Lipschitz condition can be found in Chapter 3. A classic text on ordinary differential equations is "Theory of Ordinary Differential Equations" by E.A. Coddington and N. Levinson. V.I. Arnold's "Ordinary Differential Equations" is also a popular text.

Although metric space is a standard topic, I found it difficult to fix upon a single reference book. Rudin's Principles covers some metric spaces, but his attention is mainly on the Euclidean space. Moreover, for a devoted student, this book should have been studied in a previous summer. Finally, I decide to list Dieudonne's old book "Foundation of Modern Analysis" as the only reference. This is the book from which I learned the subject, but it seems a bit out-dated and not easy to follow. Basic things on metric spaces have not changed at all in these years (despite delicate analysis on the convergence in metric spaces has become a hot research topic lately). Another good reference which is more comprehensible but contains less content is G.F. Simmons "Introduction to Topology and Modern Analysis". The chapters on metric and topological spaces are highly readable.

Chapter 3

The Space of Continuous Functions

仲尼適楚，出於林中，見痀僂者承蜩，猶掇之也。仲尼曰：子巧乎。 有道邪。曰：我有道也。五六月累丸二而不墜，則失者錙銖；累三而不墜，則失者十一； 累五而不墜，猶掇之也。吾處身也，若厥株拘；吾執臂也，若槁木之枝；雖天地之大，萬物之多，而唯蜩翼之知。吾不反不側，不以萬物易蜩之翼，何為而不得。 孔子顧謂弟子曰：用志不分，乃凝於神，其痀僂丈人之謂乎。 莊子 達生

In this chapter we study the space of continuous functions as a prototype of infinite dimensional normed spaces. In Section 1 we review these spaces. In Section 2 the notion of separability is introduced. A proof of Weierstrass approximation theorem different from the one given in Chapter 1 is present in Section 3, following by the general Stone-Weierstrass theorem. The latter is applied to establish the separability of the space of continuous functions when the underlying space is compact. Ascoli-Arezela theorem, which characterizes compact sets in the space of continuous functions, is established in Section 4. Finally in Section 5 we study complete metric spaces. Baire category theorem is proved and, as an application, it is shown that continuous, nowhere differentiable functions form a set of second category in the space of continuous functions.

3.1 Spaces of Continuous Functions

We studied continuous functions on an interval in MATH2050/60 and in a domain bounded by curves/surfaces in $\mathbb{R}^2$ or $\mathbb{R}^3$ in MATH2010/20. After the introduction of metric spaces, it is natural to consider the space of continuous functions defined on a metric space.

Let $C(X)$ denote the vector space of all continuous functions defined on X where (X, d) is a metric space. In the previous chapter we showed that there are many continuous functions in X. In general, in a metric space such as the real line, a continuous function may not be bounded. In order to turn continuous functions into a normed space, we need

to restrict to bounded functions. For this purpose let

$$C_b(X) = \{f : \ f \in C(X), \ |f(x)| \leq M, \ \forall x \in X \text{ for some } M\}.$$

It is readily checked that $C_b(X)$ is a normed space under the sup-norm. From now on, $C_b(X)$ is always regarded as a metric space under the metric induced by the sup-norm. In other words,

$$d_\infty(f, g) = \|f - g\|_\infty, \quad \forall f, g \in C_b(X).$$

Some basic properties of $C_b(X)$ are listed below.

First, $C_b(X)$ *is a Banach space.* Although the proof has no difference from its special case $C[a, b]$, we reproduce the proof here. Indeed, let $\{f_n\}$ be a Cauchy sequence in $C_b(X)$. For $\varepsilon > 0$, there exists some n_0 such that $\|f_n - f_m\|_\infty < \varepsilon/4$ for all $n \geq n_0$. In particular, it means for each x, $\{f_n(x)\}$ is a Cauchy sequence in $\mathbb{R}$. By the completeness of $\mathbb{R}$, the limit $\lim_{n\to\infty} f_n(x)$ exists and we define $f(x) \equiv \lim_{n\to\infty} f_n(x)$. Assuming that $f \in C_b(X)$, by taking $m \to \infty$ in the inequality above, we immediately obtain $\|f_n - f\|_\infty \leq \varepsilon/4 < \varepsilon$, hence $f_n \to f$ in $C_b(X)$. To show that $f \in C_b(X)$, we let $m \to \infty$ in $|f_n(x) - f_m(x)| < \varepsilon/4$ to get $|f_n(x) - f(x)| \leq \varepsilon/4$ for all x and $n \geq n_0$. Taking $n = n_0$ we get $|f(x)| \leq |f(x) - f_{n_0}(x)| + |f_{n_0}(x)| \leq \varepsilon/4 + \|f_{n_0}\|_\infty$, hence f is bounded. On the other hand, as f_{n_0} is continuous, for each x we can find a δ such that $|f_{n_0}(y) - f_{n_0}(x)| < \varepsilon/4$ whenever $d(y, x) < \delta$. It follows that for all y, $d(y, x) < \delta$,

$$|f(y) - f(x)| \leq |f(y) - f_{n_0}(y)| + |f_{n_0}(y) - f_{n_0}(x)| + |f_{n_0}(x) - f(x)| \leq \frac{3\varepsilon}{4} < \varepsilon.$$

From this proof we see that the completeness of $C_b(X)$ is inherited from the completeness of $\mathbb{R}$, so the underlying space X does not play any role in this aspect.

Second, $C_b(X) = C(X)$ *when X is a compact metric space.* Again this was done before and we reproduce a proof. We need to show every continuous function on a compact set is bounded. Assume on the contrary that for some continuous f, there are points $\{x_k\}$ such that $|f(x_k)| \to \infty$. By compactness, there is a subsequence $\{x_{k_j}\}$ and $z \in X$ such that $\lim_{j\to\infty} x_{k_j} = z$. But, by continuity we would have $\lim_{j\to\infty} |f(x_{k_j})| = |f(z)| < \infty$, contradiction holds. It is a good exercise to give another proof based on the finite cover property.

Third, $C_b(X)$ *is usually an infinite dimensional Banach space.* In particular this is true when X is $\mathbb{R}^n$ or a subset with non-empty interior. It is easy to construct an infinite set of linearly independent bounded, continuous functions in such X. For instance, when X is bounded and has non-empty interior, the restriction of all monomials on X are linearly independent and hence forms an infinite dimensional subspace in $C_b(X)$. On the other hand, $C_b(X)$ could be of finite dimensional in some extreme cases. For instance, take $X = \{x_1, \cdots, x_n\}$ be a finite set equipped with the discrete metric. Every function defined in X is continuous. Since a function is completely determined by its values, the correspondence $f \mapsto (f(x_1), \cdots, f(x_n))$ sets up a bijective linear map between $C_b(X)$ and $\mathbb{R}^n$. Therefore, the dimension of $C_b(X)$ is equal to n.

Finally, although $C(X)$ may contain unbounded functions, it is still possible to introduce a metric on $C(X)$ instead of a norm in some cases. Especially, we describe the metric when X is $\mathbb{R}^n$. Indeed, for $f \in C(\mathbb{R}^n)$, let $\|f\|_n = \sup_{x \in B_n(0)} |f(x)|$ and

$$d(f,g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\|f-g\|_n}{1+\|f-g\|_n}.$$

One can verify that d forms a complete metric on $C(\mathbb{R}^n)$. Whether a vector space admits a norm (normable) or a metric (metrizable) is a topic in functional analysis.

Other useful spaces of bounded, continuous or differentiable functions can be found in the exercise.

3.2 Separabilty

We start with a general metric space (X,d). A set E in X is **dense** if for every $x \in X$ and $\varepsilon > 0$, there exists some $y \in E$ such that $d(y,x) < \varepsilon$. Equivalently, E is dense if every metric ball contains some point in E. It is easy to see that E is dense if and only if $\overline{E} = X$. According to this definition, the space X is dense in X trivially. In the discrete metric, every single point is a metric ball and X is the only dense set. In other cases, there could be many dense sets. For instance, consider $\mathbb{R}$ the following three sets are dense: The set of all rational numbers, the set of all irrational numbers and the set formed by removing finitely many points from $\mathbb{R}$. Similar situations hold in $\mathbb{R}^n$. Next consider $C(R)$ where R is a closed, bounded (compact) rectangle in $\mathbb{R}^n$. Weierstrass approximation theorem asserts that the collection of all polynomials forms a dense set in $C(R)$. In Chapter 1 we showed that all finite trigonometric series are dense in $\mathcal{C}_{2\pi}$. In an exercise we showed that all finite double trigonometric series

$$\sum_{m,n=-N}^{N} a_{mn} e^{i(mx+ny)}$$

are dense in the space of continuous functions which are 2π-periodic in x and y. Generalization to higher dimensions is immediate.

The notion of a dense set is useful in the study of the structure of metric spaces. A metric space X is called a **separable space** if it admits a countable dense subset. Equivalently, X is separable if there is a countable subset E satisfying $\overline{E} = X$. A set is **separable** if it is separable as a metric subspace. When a metric space is separable, every element can be approximated by elements from a countable set. Hence its structure is easier to study than the non-separable ones. Here are two basic properties of separable spaces.

Proposition 3.1. *Every subset of a separable space is separable.*

Proof. Let Y be a subset of the separable space (X, d) and $D = \{x_j\}$ a countable dense subset of X. For each n, pick a point z_j^n from $Y \cap B_{1/n}(x_j)$ if it is non-empty to form the countable set $E = \{z_j^n\}$. We claim that E is dense in Y. For, let $y \in Y$ and each $\varepsilon > 0$, there is some $x_j \in D$ such that $d(y, x_j) < \varepsilon/2$. Therefore, for $n > 2/\varepsilon$, $B_{1/n}(x_j) \cap Y$ is nonempty and we can find some $y_j^n \in B_{1/n}(x_j) \cap Y$. It follows that $d(y, y_j^n) \leq d(y, x_j) + d(x_j, y_j^n) < \varepsilon/2 + 1/n < \varepsilon$.

□

Proposition 3.2. *Every compact metric space is separable.*

Proof. Every compact space is totally bounded. By Proposition 2.11, for each n, there exist finitely many points $x_1, \cdots, x_N$ such that the balls $B_{1/n}(x_j)$, $j = 1, \cdots N$, form an open cover of the space. It is clear that the countable set consisting of all centers of these balls when n runs from 1 to infinity forms a dense set of the space.

□

Now we give some examples of separable spaces.

Example 3.1. Consider the Euclidean space $\mathbb{R}^n$. The set of all rational numbers $\mathbb{Q}$ forms a countable dense subset of $\mathbb{R}$, so $\mathbb{R}$ is a separable space. Similarly, $\mathbb{R}^n$ is separable for all $n \geq 1$ because it contains the dense subset $\mathbb{Q}^n$. According to Proposition 3.1, all sets in the Euclidean space are separable.

Example 3.2. $C[a, b]$ is a separable space. Without loss of generality we take $[a, b] = [0, 1]$. Denote by $\mathcal{P}$ the restriction of all polynomials to $[0, 1]$. Let

$$\mathcal{S} = \{p \in \mathcal{P} : \text{ The coefficients of } p \text{ are rational numbers}\}.$$

It is clear that $\mathcal{S}$ is a countable set. Given any polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n$, $a_j \in \mathbb{R}, j = 1, \cdots, n$. For every $\varepsilon > 0$, we can choose some $b_j \in \mathbb{Q}$ such that $|a_j - b_j| < \varepsilon/(n+1)$ for all j. It follows that for $q(x) = \sum_j b_j x^j \in S$, we have

$$\begin{aligned} |p(x) - q(x)| &\leq \sum_j |a_0 - b_0| + |a_1 - b_1||x| + \cdots + |a_n - b_n||x|^n \\ &< (n+1)\frac{\varepsilon}{2(n+1)} \\ &= \frac{\varepsilon}{2} \end{aligned}$$

for all x. We conclude that $\|p - q\|_\infty \leq \varepsilon/2$ Now, for any $f \in C[0, 1]$ and $\varepsilon > 0$, we apply Weierstrass approximation theorem to obtain a polynomial p such that $\|f - p\|_\infty < \varepsilon/2$ and then find some $q \in \mathcal{S}$ such that $\|p - q\|_\infty \leq \varepsilon/2$. It follows that

$$\|f - q\|_\infty \leq \|f - p\|_\infty + \|p - q\|_\infty < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

that is, $\mathcal{S}$ is dense in $C[0,1]$. A straightforward generalization shows that $C(R)$ is separable when R is a closed, bounded (that is, a compact) rectangle in $\mathbb{R}^n$.

A general result in this direction is the following theorem. It is based on the Stone-Weierstrass theorem in Section 3. The proof is for optional reading.

Theorem 3.3. *The space $C(X)$ is separable when X is a compact metric space.*

Proof. When X consists of a singleton, $C(X)$ is equal to $\mathbb{R}$ and so separable. We will always assume X has more than one points below. By Proposition 2.11, we can find a sequence of balls $\{B_j\}$ whose centers $\{z_j\}$ form a dense set in X. Define $f_j(x) = d(x, z_j)$ and let $\mathcal{M} \subset C(X)$ consist of functions which are finite product of f_j's. Then let $\mathcal{A}$ consist of functions of the form

$$f = \sum_{k=1}^{N} a_k h_k, \qquad h_k \in \mathcal{M}, \quad a_j \in \mathbb{Q}.$$

It is readily checked that $\mathcal{A}$ forms a subalgebra of $C(X)$. To verify separating points property let x_1 and x_2 be two distinct points in X. The function $f(x) = d(x, x_1)$ satisfies $f(x_1) = 0$ and $f(x_2) \neq 0$. By density, we can find some z_k close to x_1 so that the function $f_k(x) = d(x, z_k)$ separates x_1 and x_2. On the other hand, given any point x_0 we can fix another distinct point y_0 so that the function $d(x, y_0)$ is nonvanishing at x_0. By density again, there is some z_j close to y_0 such that $f_j(x_0) \neq 0$. By Stone-Weierstrass theorem, $\mathcal{A}$ is dense in $C(X)$. The theorem will be proved if we can show that $\mathcal{A}$ is countable. To see this, let $\mathcal{A}_n$ be the subset of $\mathcal{A}$ which only involves finitely many functions $f_1, \cdots, f_n$. We have $\mathcal{A} = \cup_n \mathcal{A}_n$, so it suffices to show each $\mathcal{A}_n$ is countable. Each function in $\mathcal{A}_n$ is composed of finitely many terms of the form $f_{n_1}^{a_1} \cdots f_{n_k}^{a_k}$, $n_j \in \{1, \cdots, n\}$. Let $\mathcal{A}_n^m \subset \mathcal{A}_n$ consist of all those functions whose "degree" is less than or equal to m. It is clear that $\mathcal{A}_n^m$ is countable, so is $\mathcal{A}_n = \cup_m \mathcal{A}_n^m$. □

To conclude this section, we note the existence of non-separable spaces. Here is one.

Example 3.3. Consider the space of all bounded functions on $[a, b]$ under the supnorm. It forms a Banach space $B[a, b]$. We claim that it is not separable. For, let $f_z \in B[a, b]$ be given by $f_z(x) = 0$ for all $x \neq z$ and $f_z(z) = 1$. All f_z's form an uncountable set. Obviously the metric balls $B_{1/2}(f_z)$ are pairwise disjoint. If S is a dense subset of $B[a, b]$, $S \cap B_{1/2}(f_z)$ must be non-empty for each z. We pick $w_z \in S \cap B_{1/2}(f_z)$ to form an uncountable subset $\{w_z\}$ of S. We concldue that S must be uncountable, so there is no countable dense set of $B[a, b]$.

3.3 The Stone-Weierstrass Theorem

This section is for optional reading.

So far we have shown that trigonometric functions and polynomials are dense in the space of periodic, continuous functions and the space of continuous functions respectively. In this section we will establish a far-reaching generalization of these results in the space of continuous functions defined in a compact metric space. In such a space both trigonometric functions and polynomials are not available, so we need to seek a reasonable formulation. The answer relies on an extra algebraic structure we have exploited explicitly.

Observe that the space $C_b(X)$ carries an extra property, namely, it is an algebra under pointwise product. A subspace $\mathcal{A}$ is called a subalgebra of $C_b(X)$ if it is closed under this product. It is readily checked that all polynomials on $[a, b]$ form a subalgebra of $C[a, b]$, so does the subalgebra consisting of all polynomials with rational coefficients. Similarly, the vector space consisting of all trigonometric polynomials and its subspace consisting of all trigonometric polynomials with rational coefficients are algebras in in the space of 2π-periodic continuous functions in $\mathcal{C}_{2\pi}$. Note that $\mathcal{C}_{2\pi}$ can be identified with $C(S^1)$ where $S^1 = \{(\cos t, \sin t) \in \mathbb{R}^2 : \ t \in [0, 2\pi]\}$ is the unit circle.

Before proceeding further, recall that aside from the constant ones, there are many continuous functions in a metric space. For instance, given any two distinct points x_1 and x_2 in X, it is possible to find some $f \in C(X)$ such that $f(x_1) \neq f(x_2)$. We simply take $f(x) = d(x, x_1)$ and $f(x_1) = 0 < f(x_2)$. It is even possible to find one in $C_b(X)$, e.g., $g(x) = f(x)/(1 + f(x))$ serves this purpose. We now consider what conditions a subalgebra must possess in order that it becomes dense in $C_b(X)$. A subalgebra is called to satisfy the **separating points property** if for any two points x_1 and x_2 in X, there exists some $f \in \mathcal{A}$ satisfying $f(x_1) \neq f(x_2)$. From the discussion above, it is clear that $\mathcal{A}$ must satisfy the separating points property if its closure is $C_b(X)$. Thus the separating points property is a necessary condition for a subalgebra to be dense. On the other hand, the polynomials of the form $\sum_{j=0}^{n} a_j x^{2j}$ form an algebra which does not have the separating point property, for it is clear that $p(-x) = p(x)$ for such p. Another condition is that, whenever $x \in X$, there must be some $g \in \mathcal{A}$ such that $g(x) \neq 0$. We will call this the **non-vanishing property.** The non-vanishing property fails to hold for the algebra consisting of all polynomials of the form $\sum_{j=1}^{n} a_j x^j$, for $p(0) = 0$ for all these p. The non-vanishing property is also a necessary condition for an algebra to be dense in $C(X)$. For, if for some particular $z \in X$, $f(z) = 0$ holds for all $f \in \mathcal{A}$, it is impossible to approximate the constant function 1 in the supnorm by functions in $\mathcal{A}$. Surprisingly, it turns out these two conditions are also sufficient when the underlying space is compact, and this is the content of the following theorem. This is for an optional reading.

Theorem 3.4 (**Stone-Weierstrass Theorem**). *Let $\mathcal{A}$ be a subalgebra of $C(X)$ where X is a compact metric space. Then $\mathcal{A}$ is dense in $C(X)$ if and only if it has the separating*

points and non-vanishing properties.

Recall that $C_b(X) = C(X)$ when X is compact. For the proof of this theorem two lemmas are needed.

Lemma 3.5. *Let $\mathcal{A}$ be a subalgebra of $C_b(X)$. For every pair $f, g \in \mathcal{A}$, $|f|$, $f \vee g$ and $f \wedge g$ belong to the closure of $\mathcal{A}$.*

Proof. Observing the relations

$$f \vee g \equiv \max\{f, g\} = \frac{1}{2}(f + g) + \frac{1}{2}|f + g|,$$

and

$$f \wedge g \equiv \min\{f, g\} = \frac{1}{2}(f + g) - \frac{1}{2}|f + g|,$$

it suffices to show that $|f|$ belongs to the closure of $\mathcal{A}$. Indeed, given $\varepsilon > 0$, since $t \mapsto |t|$ is continuous on $[-M, M]$, $M = \sup\{|f(x)| : x \in X\}$, by Weierstrass approximation theorem, there exists a polynomial p such that $\big||t| - p(t)\big| < \varepsilon$ for all $t \in [-M, M]$. It follows that $\||f| - p(f)\|_\infty \leq \varepsilon$. As $p(f) \in \mathcal{A}$, we conclude that $\mathcal{A}$ is dense in $C(X)$. □

Lemma 3.6. *Let $\mathcal{A}$ be a subalgebra in $C(X)$ which separates points and non-vanishing at all points. For $x_1, x_2 \in X$ and $\alpha, \beta \in \mathbb{R}$, there exists $\varphi \in \mathcal{A}$ such that $\varphi(x_1) = \alpha$ and $\varphi(x_2) = \beta$.*

Proof. Since $\mathcal{A}$ separates points, we can find some $\psi \in \mathcal{A}$ such that $\psi(x_1) \neq \psi(x_2)$. We claim that one can further choose ψ such that $\psi(x_1), \psi(x_2)$ are both non-zero. For, if, for instance, $\psi(x_1) = 0$, fix some $\xi \in \mathcal{A}$ satisfying $\xi(x_1) \neq 0$. This is possible due to the non-vanishing property. Consider a function $\psi_1 \in \mathcal{A}$ of the form $\psi + t\xi$. We would like to find $t \in \mathbb{R}$ such that (a) $\psi_1(x_1) \neq \psi_1(x_2)$, (b) $\psi_1(x_1) \neq 0$, and (c) $\psi_1(x_2) \neq 0$. There are two cases; when $\xi(x_2) \neq 0$, it suffices to choose t such that $t \neq 0, -\psi(x_2)/\xi(x_2)$ (if $\xi(x_2) \neq 0$). When $\xi(x_2) = 0$, we choose t such that $t \neq \psi(x_2)/\xi(x_1)$. Replacing ψ by ψ_1, we obtain our desired function which satisfies (a)–(c).

Now, we can find a and b such that the combination $\varphi = a\psi + b\psi^2 \in \mathcal{A}$ satisfies the requirement in the lemma. Indeed, what we need are the conditions $a\psi(x_1) + b\psi^2(x_1) = \alpha$ and $a\psi(x_2) + b\psi^2(x_2) = \beta$. As the determinant of this linear system (viewing a and b as the unknowns) is equal to $\psi(x_1)\psi(x_2)(\psi(x_1) - \psi(x_2))$ which is not equal to 0, a and b can always be found. □

Proof of Theorem 3.5. It remains to establish the necessary part of the theorem. Let $f \in C(X)$ be given. For each pair of x, y, there exists a function $\varphi_{x,y} \in \mathcal{A}$ satisfying $\varphi_{x,y}(x) = f(x)$ and $\varphi_{x,y}(y) = f(y)$. This is due to the previous lemma when x and y are

distinct. When x is equal to y, such function still exists. Now, for each $\varepsilon > 0$, there exists an open set $U_{x,y}$ containing x and y such that

$$|f(t) - \varphi_{x,y}(t)| < \varepsilon, \quad \forall t \in U_{x,y}.$$

For fixed y, the sets $\{U_{x,y} : x \in X\}$ form an open cover of X. By the compactness of X, it admits a finite subcover $\{U_{x_j,y}\}_{j=1}^N$. The function $\varphi_y = \varphi_{x_1,y} \vee \cdots \vee \varphi_{x_N,y}$ belongs to $\overline{\mathcal{A}}$ according to Lemma 3.3. Furthermore, $\varphi_y > f - \varepsilon$ in X. For, let $x \in X$, there is some $U_{x_j,y}$ containing x. Therefore, $\varphi_y(x) \geq \varphi_{x_j,y}(x) > f(x) - \varepsilon$. Next, $G_y \equiv \cap_{j=1}^N U_{x_j,y}$ is an open set containing y and all these open sets together form an open cover of X when y runs over X. Note that $\varphi_y < f + \varepsilon$ on X since $\varphi_{x_j,y} < f + \varepsilon$ in G_y for all $j = 1, \cdots, N$. By compactness, we can extract $y_1, \cdots, y_M$ such that $\{G_{y_k}\}_{k=1}^M$, cover X. Define $\varphi = \varphi_{y_1} \wedge \cdots \wedge \varphi_{y_M}$. By Lemma 3.3 it belongs to $\overline{\mathcal{A}}$ and $\varphi > f - \varepsilon$ in X. On the other hand, each x belongs to some G_{y_k}, so $\varphi(x) \leq \varphi_{y_k}(x) < f(x) + \varepsilon$ holds. We conclude that $\|f - \varphi\|_\infty < \varepsilon, \varphi \in \overline{\mathcal{A}}$. □

3.4 Compactness and Arzela-Ascoli Theorem

We pointed out before that not every closed, bounded set in a metric space is compact. In Section 2.3 a bounded sequence without any convergent subsequence is explicitly displayed to show that a closed, bounded set in $C[a, b]$ needs not be compact. In view of numerous theoretic and practical applications, it is strongly desirable to give a characterization of compact sets in $C[a, b]$. The answer is given by the fundamental Arzela-Ascoli theorem. This theorem gives a necessary and sufficient condition when a closed and bounded set in $C[a, b]$ is compact. In order to have wider applications, we will work on a more general space $C(K)$, where K is a closed, bounded subset of $\mathbb{R}^n$, instead of $C[a, b]$. Recall that $C(K)$ is a complete, separable space under the sup-norm.

The crux for compactness for continuous functions lies on the notion of equicontinuity. Let X be a subset of $\mathbb{R}^n$. A subset $\mathcal{F}$ of $C(X)$ is **equicontinuous** if for every $\varepsilon > 0$, there exists some δ such that

$$|f(x) - f(y)| < \varepsilon, \quad \text{for all } f \in \mathcal{F}, \text{ and } |x - y| < \delta, \ x, y \in X.$$

Recall that a function is uniformly continuous in X if for each $\varepsilon > 0$, there exists some δ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta, \ x, y \in X$. So, equicontinuity means that δ can further be chosen independent of the functions in $\mathcal{F}$.

There are various ways to show that a family of functions is equicontinuous. Recall that a function f defined in a subset X of $\mathbb{R}^n$ is called Hölder continuous if there exists some $\alpha \in (0, 1)$ such that

$$|f(x) - f(y)| \leq L|x - y|^\alpha, \quad \text{for all } x, y \in X, \tag{3.1}$$

for some constant L. The number α is called the Hölder exponent. The function is called Lipschitz continuous if (3.1) holds for α equals to 1. A family of functions $\mathcal{F}$ in $C(X)$ is said to satisfy a uniform Hölder or Lipschitz condition if all members in $\mathcal{F}$ are Hölder continuous with the same α and L or Lipschitz continuous and (3.1) holds for the same constant L. Clearly, such $\mathcal{F}$ is equicontinuous. The following situation is commonly encountered in the study of differential equations. The philosophy is that equicontinuity can be obtained if there is a good, uniform control on the derivatives of functions in $\mathcal{F}$.

Proposition 3.7. *Let $\mathcal{F}$ be a subset of $C(X)$ where X is a convex set in $\mathbb{R}^n$. Suppose that each function in $\mathcal{F}$ is differentiable and there is a uniform bound on the partial derivatives of these functions in $\mathcal{F}$. Then $\mathcal{F}$ is equicontinuous.*

Proof. For, x and y in X, $(1-t)x+ty$, $t \in [0,1]$, belongs to X by convexity. Let $\psi(t) \equiv f((1-t)x+ty)$. From the mean-value theorem

$$\psi(1)-\psi(0) = \psi'(t^*)(1-0),\ t^* \in [0,1],$$

for some mean value $t^* \in (0,1)$ and the chain rule

$$\psi'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}((1-t)x+ty)(y_j-x_j),$$

we have

$$f(y)-f(x) = \psi(1)-\psi(0) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}((1-t^*)x+t^*y)(y_j-x_j).$$

Therefore,

$$|f(y)-f(x)| \le \sqrt{n}M|y-x|,$$

where $M = \sup\{|\partial f/\partial x_j(x)| :\ x \in X,\ j=1,\dots,n,\ f \in \mathcal{F}\}$ after using Cauchy-Schwarz inequality. So $\mathcal{F}$ satisfies a uniform Lipschitz condition with Lipschitz constant $n^{1/2}M$. □

Example 3.4. Let $A = \{y :\ y \text{ solves } y' = \sin(xy),\ x \in [-1,1]\}$. From the discussion in the last section of Chapter 2, given any $y_0 \in \mathbb{R}$, there is a unique solution y solving the equation and $y(0) = y_0$, so A contains many elements. We e have the uniform estimate on its derivative, namely,

$$|y'(x)| = |\sin xy| \le 1,$$

hence it follows from the above discussion that A is equicontinuous.

Example 3.5. Let $B = \{f \in C[0,1] : |f(x)| \le 1, x \in [0,1]\}$. Clearly B is closed and bounded. However, we do not have any uniform control on the oscillation of the functions in this set, so it should not be equicontinuous. In fact, consider the sequence $\{\sin nx\}, n \ge 1$, in B. We claim that it is not equicontinuous. In fact, suppose for $\varepsilon = 1/2$, there exists some δ such that $|\sin nx - \sin ny| < 1/2$, whenever $|x-y| < \delta$ for all n. Pick a large n such that $n\delta > \pi$. Taking $x = 0$ and $y = \pi/2n$, $|x-y| < \delta$ but $|\sin nx - \sin ny| = |\sin \pi/2| = 1 > 1/2$, contradiction holds. Hence B is not equicontinuous.

—

Example 3.6. Let $C = \{f \in C[-1,1] : \ f(x) = ax^2 + be^x, 0 \le f(0) \le 1, 0 \le f(1) \le 10,\}$. Although there is no obvious control on the derivative, this set basically consists of functions with two parameters a and b only. If we consider the map $f \mapsto (a,b)$, this map sets up a one-to-one correspondence with C and a subset of $\mathbb{R}^2$. Since every closed, bounded subset in $\mathbb{R}^2$ is compact, it gives hope that C is equicontinuous. First, we show that C is bounded. From $f(0) \in [0,1]$ and $f(1) \in [0,10]$ we have $0 \le b \le 1$ and $-e \le a \le 10$. Therefore, $|f(x)| \le 10 + e$ for all $f \in C$, so C is bounded. Now, we claim that it is equicontinuous. In fact, since the functions $x \mapsto x^2$ and $x \mapsto e^x$ are uniformly continuous on $[-1.1]$, for $\varepsilon > 0$, there exists some δ such that $|x^2 - y^2|, |e^x - e^y| < \varepsilon/2(10+e)$ whenever $|x - y| < \delta$. It follows that

$$|f(x) - f(y)| \le |a||x^2 - y^2| + |b||e^x - e^y| \le \varepsilon.$$

More examples of equicontinuous families can be found in the exercise.

Theorem 3.8 (**Arzela-Ascoli**). *Let $\mathcal{F}$ be a closed set in $C(K)$ where K is a closed and bounded set in $\mathbb{R}^n$. Then $\mathcal{F}$ is compact if and only if it is bounded and equicontinuous.*

A set $\mathcal{E} \subset C(K)$ is bounded means it is contained in a ball, or, more specifically, there exists $M > 0$ such that

$$|f(x)| \le M, \quad \text{for all } f \in \mathcal{E} \ \text{ and } x \in K.$$

We need the following lemma from elementary analysis. It will be used in many occasions.

Lemma 3.9. *Let $\{z_j, \ j \ge 1\}$ be a sequence in $\mathbb{R}^n$ and $\{f_n\}$ be a sequence of functions defined on $\{z_j, \ j \ge 1\}$. Suppose that for each j, there exists an M_j such that $|f_n(z_j)| \le M_j$ for all $n \ge 1$. There is a subsequence of $\{f_n\}$, $\{g_n\}$, such that $\{g_n(z_j)\}$ is convergent for each j.*

Proof. Since $\{f_n(z_1)\}$ is a bounded sequence, we can extract a subsequence $\{f_n^1\}$ such that $\{f_n^1(z_1)\}$ is convergent. Next, as $\{f_n^1(z_2)\}$ is bounded, it has a subsequence $\{f_n^2\}$ such that $\{f_n^2(z_2)\}$ is convergent. Keep doing in this way, we obtain sequences $\{f_n^j\}$ satisfying (i) $\{f_n^{j+1}\}$ is a subsequence of $\{f_n^j\}$ and (ii) $\{f_n^j(z_1)\}, \{f_n^j(z_2)\}, \cdots, \{f_n^j(z_j)\}$ are convergent. Then the diagonal sequence $\{g_n\}, \ g_n = f_n^n, \ $ for all $n \ge 1$, is a subsequence of $\{f_n\}$ which converges at every z_j. □

The subsequence selected in this way is sometimes called Cantor's diagonal sequence.

Proof of Arzela-Ascoli Theorem. Assuming boundedness and equicontinuity of $\mathcal{F}$, we would like to show that $\mathcal{F}$ is compact.

Since K is compact in $\mathbb{R}^n$, it is totally bounded. By Proposition 2.11, for each $j \geq 1$, we can cover K by finitely many balls $B_{1/j}(x_1^j), \cdots, B_{1/j}(x_N^j)$ where the number N depends on j. All $\{x_k^j\}$, $j \geq 1, 1 \leq k \leq N$, form a countable set. For any sequence $\{f_n\}$ in $\mathcal{F}$, by Lemma 3.11, we can pick a subsequence denoted by $\{g_n\}$ such that $\{g_n(x_k^j)\}$ is convergent for all x_k^j's. We claim that $\{g_n\}$ is a Cauchy sequence in $C(K)$. For, due to the equicontinuity of $\mathcal{F}$, for every $\varepsilon > 0$, there exists a δ such that $|g_n(x) - g_n(y)| < \varepsilon$, whenever $|x - y| < \delta$. Pick j_0, $1/j_0 < \delta$. Then for $x \in K$, there exists $x_k^{j_0}$ such that $|x - x_k^{j_0}| < 1/j_0 < \delta$,

$$\begin{aligned} |g_n(x) - g_m(x)| &\leq |g_n(x) - g_n(x_k^{j_0})| + |g_n(x_k^{j_0}) - g_m(x_k^{j_0})| + |g_m(x_k^{j_0}) - g_m(x)| \\ &< \varepsilon + |g_n(x_k^{j_0}) - g_m(x_k^{j_0})| + \varepsilon. \end{aligned}$$

As $\{g_n(x_k^{j_0})\}$ converges, there exists n_0 such that

$$|g_n(x_k^{j_0}) - g_m(x_k^{j_0})| < \varepsilon, \quad \text{for all } n, m \geq n_0. \tag{3.2}$$

Here n_0 depends on $x_k^{j_0}$. As there are finitely many $x_k^{j_0}$'s, we can choose some N_0 such that (3.2) holds for all $x_k^{j_0}$ and $n, m \geq N_0$. It follows that

$$|g_n(x) - g_m(x)| < 3\varepsilon, \quad \text{for all } n, m \geq N_0,$$

i.e., $\{g_n\}$ is a Cauchy sequence in $C(K)$. By the completeness of $C(K)$ and the closedness of $\mathcal{F}$, $\{g_n\}$ converges to some function in $\mathcal{F}$.

Conversely, let $\mathcal{F}$ be compact. By Proposition 2.11, for each $\varepsilon > 0$, there exist $f_1, \cdots, f_N \in \mathcal{F}$ such that $\mathcal{F} \subset \bigcup_{j=1}^N B_\varepsilon(f_j)$ where N depends on ε. So for any $f \in \mathcal{F}$, there exists f_j such that

$$|f(x) - f_j(x)| < \varepsilon, \quad \text{for all } x \in K.$$

As each f_j is continuous, there exists δ_j such that $|f_j(x) - f_j(y)| < \varepsilon$ whenever $|x - y| < \delta_j$. Letting $\delta = \min\{\delta_1, \cdots, \delta_N\}$, then

$$|f(x) - f(y)| \leq |f(x) - f_j(x)| + |f_j(x) - f_j(y)| + |f_j(y) - f(y)| < 3\varepsilon,$$

for $|x - y| < \delta$, so S is equicontinuous. As $\mathcal{F}$ can be covered by finitely many balls of radius 1, it is also bounded. We have completed the proof of Arzela-Ascoli theorem.

□

The following special case of Arzela-Ascoli theorem, sometimes called Ascoli's theorem, is the result we usually apply. When it comes to applications, the sufficient condition is more relevant than the necessary condition.

Theorem 3.10 (**Ascoli's Theorem**). *A sequence in $C(K)$ where K is a closed, bounded set in $\mathbb{R}^n$ has a convergent subsequence if it is uniformly bounded and equicontinuous.*

Proof. Let $\mathcal{F}$ be the closure of the sequence $\{f_n\}$. We would like to show that $\mathcal{F}$ is bounded and equicontinuous. First of all, by the uniform boundedness assumption, there is some M such that

$$|f_n(x)| \leq M, \quad \forall x \in K, \ n \geq 1.$$

As every function in $\mathcal{F}$ is either one of these f_n or the limit of its subsequence, it also satisfies this estimate, so $\mathcal{F}$ is bounded in $C(K)$. On the other hand, by equicontinuity, for every $\varepsilon > 0$, there exists some δ such that

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{2}, \quad \forall x, y \in K, \ |x - y| < \delta.$$

As every $f \in \mathcal{F}$ is the limit of a subsequence of $\{f_n\}$, f satisfies

$$|f(x) - f(y)| \leq \frac{\varepsilon}{2} < \varepsilon, \quad \forall x, y \in K, \ |x - y| < \delta,$$

so $\mathcal{F}$ is also equicontinuous. Now the conclusion follows from the Arzela-Ascoli theorem.

□

We present an application of Arzela-Ascoli theorem to ordinary differential equations. Consider the Cauchy problem (2.3) again,

$$\begin{cases} \dfrac{dy}{dx} = f(x, y), \\ y(x_0) = y_0. \end{cases} \tag{2.3}$$

where f is a continuous function defined in the rectangle $R = [x_0 - a, x_0 + a] \times [y_0 - b, y_0 + b]$. In Chapter 2 we proved that this Cauchy problem has a *unique* solution when f satisfies the Lipschitz condition. Now we show that the existence part of Picard-Lindelöf theoorem is still valid without the Lipschitiz condition.

Theorem 3.11 (**Cauchy-Peano Theorem**). *Consider (2.3) where f is continuous on $R = [x_0 - a, x_0 + a] \times [y_0 - b, y_0 + b]$. There exist $a' \in (0, a)$ and a C^1-function $y(x) : [x_0 - a', x_0 + a'] \to [y_0 - b, y_0 + b]$, solving (2.3).*

From the proof we will see that a' can be taken to be $0 < a' < \min\{a, b/M\}$ where $M = \sup\{|f(x, y)| : \ (x, y) \in R\}$. The theorem is also valid for systems.

Proof. Recalling in the proof of Picard-Lindelöf theorem we showed that under the Lipschitz condition the unique solution exists on the interval $[x_0 - a', x_0 + a']$ where $0 < a' <$

$\min\{a, b/M, 1/L^*\}$ where L^* is the Lipschitz constant. Let us first argue that the maximal solution in fact exists in the interval $[x_0 - a', x_0 + a']$ where $0 < a' < \min\{a, b/M\}$. In other words, the Lipschitz condition does not play any role in the range of existence. Although this was done in the exercise, we include it here for the sake of completeness.

Take $x_0 = y_0 = 0$ to simplify notations. The functions $w(x) = Mx$ and $z(x) = -Mx$ satisfy $y' = \pm M$, $y(0) = 0$, respectively. By comparing them with y, our maximal solution to (2.3), we have $z(x) \leq y(x) \leq w(x)$ as long as y exists. In case y exists on $[0, \alpha)$ for some $\alpha < \min\{a, b/M\}$, $(x, y(x))$ would be confined in the triangle bounded by $y = Mx, y = -Mx$, and $x = \alpha$. As this triangle is compactly contained in the interior of R, the Lipschitz constant ensures that the solution could be extended beyond α. Thus the solution exists up to $\min\{a, b/M\}$. Similarly, one can show that the solution exists in $(-\min\{a, b/M\}, 0]$.

With this improvement at our disposal, we prove the theorem. First, of all, by Weierstrass approximation theorem, there exists a sequence of polynomials $\{f_n\}$ approaching f in $C([-a, a] \times [-b, b])$ uniformly. In particular, it means that $M_n \to M$, where $M_n = \max\{|f_n(x, y)| : \ (x, y) \in [-a, a] \times [-b, b]$. As each f_n satisfies the Lipschitz condition (why?), there is a unique solution y_n defined on $I_n = (-a_n, a_n), a_n = \min\{a, b/M_n\}$ for the initial value problem

$$\frac{dy}{dx} = f_n(x, y), \quad y(0) = 0.$$

From $|dy_n/dx| \leq M_n$ and $\lim_{n\to\infty} M_n = M$, we know from Proposition 3.8 that $\{y_n\}$ forms an equicontinuous family. Clearly, it is also bounded. By Ascoli's theorem, it contains a subsequence $\{y_{n_j}\}$ converging uniformly to a continuous function $y \in I$ on every subinterval $[\alpha, \beta]$ of I and $y(0) = 0$ holds. It remains to check that y solves the differential equation for f.

Indeed, each y_n satisfies the integral equation

$$y_n(x) = \int_0^x f(t, y_n(t))dt, \quad x \in I_n.$$

As $\{y_{n_j}\} \to y$ uniformly, $\{f(x, y_{n_j}(x))\}$ also tends to $f(x, y(x))$ uniformly. By passing to limit in the formula above, we conclude that

$$y(x) = \int_0^x f(t, y(t))dt, \quad x \in I$$

holds. By the fundamental theorem of calculus, y is differentiable and a solution to (2.3). □

3.5 Completeness and Baire Category Theorem

In this section we discuss Baire category theorem, a basic property of complete metric spaces. It is concerned with the decomposition of a metric space into a countable union

of subsets. The motivation is somehow a bit strange at first glance. For instance, we can decompose the plane $\mathbb{R}^2$ as the union of strips $\mathbb{R}^2 = \bigcup_{k\in\mathbb{Z}} S_k$ where $S_k = (k, k+1]\times\mathbb{R}$. In this decomposition each S_k is not so sharply different from $\mathbb{R}^2$. Aside from the boundary, the interior of each S_k is just like the interior of $\mathbb{R}^2$. On the other hand, one can make the more extreme decomposition: $\mathbb{R}^2 = \bigcup_{\alpha\in\mathbb{R}} l_\alpha$ where $l_\alpha = \{\alpha\}\times\mathbb{R}$. Each l_α is a vertical straight line and is very different from $\mathbb{R}^2$. It is simpler in the sense that it is one-dimensional and has no area. The sacrifice is now we need an uncountable union. The question is: Can we represent $\mathbb{R}^2$ as a countable union of these sets (or sets with lower dimension)? It turns out that the answer is no. The obstruction comes from the completeness of the ambient space.

We need one definition. Let (X, d) be a metric space. A subset E of X is called **nowhere dense** if its closure does not contain any metric ball. Equivalently, E is nowhere dense if $X \setminus \overline{E}$ is dense (and open) in X. Note that a set is nowhere dense if and only if its closure is nowhere dense. Also every subset of a nowhere dense set is nowhere dense.

Theorem 3.12 (**Baire Category Theorem**). *Let $\{E_j\}_1^\infty$ be a sequence of nowhere dense subsets of (X, d) where (X, d) is complete. Then $\bigcup_{j=1}^\infty \overline{E_j}$ has empty interior.*

A set with empty interior means that it does not contain any ball. It is so if and only if its complement is a dense set.

Proof. Replacing E_j by its closure if necessary, we may assume all E_j's are closed sets. Let B_0 be any ball. The theorem will be established if we can show that $B_0 \bigcap (X \setminus \bigcup_j E_j) \neq \phi$. As E_1 is nowhere dense, there exists some point $x \in B_0$ lying outside E_1. Since E_1 is closed , we can find a closed ball $\overline{B}_1 \subset B_0$ centering at x such that $\overline{B}_1 \cap E_1 = \phi$ and its diameter $d_1 \leq d_0/2$, where d_0 is the diameter of B_0. Next, as E_2 is nowhere dense and closed, by the same reason there is a closed ball $\overline{B}_2 \subset B_1$ such that $\overline{B}_2 \cap E_2 = \phi$ and $d_2 \leq d_1/2$. Repeating this process, we obtain a sequence of closed balls $\overline{B}_j$ satisfying (a) $\overline{B}_{j+1} \subset B_j$, (b) $d_j \leq d_0/2^j$, and (c) $\overline{B}_j$ is disjoint from $E_1, \cdots, E_j$. Pick x_j from $\overline{B}_j$ to form a sequence $\{x_j\}$. As the diameters of the balls tend to zero, $\{x_j\}$ is a Cauchy sequence. By the completeness of X, $\{x_j\}$ converges to some x^*. Clearly x^* belongs to all $\overline{B}_j$. If x^* belongs to $\bigcup_j E_j$, x^* belongs to some E_{j_1}, but then $x^* \in \overline{B}_{j_1} \bigcap E_{j_1}$ which means that $\overline{B}_{j_1}$ is not disjoint from E_{j_1}, contradiction holds. We conclude that $B_0 \bigcap (X \setminus \bigcup_j E_j) \neq \phi$. □

Some remarks are in order.

First, taking complement in the statement of the theorem, it asserts that the intersection of countably many open, dense sets is again a dense set. Be careful it may not be open. For example, let $\{q_j\}$ be the set of all rational numbers in $\mathbb{R}$ and $D_k = \mathbb{R} \setminus \{q_j\}_{j=1}^k$. Each D_k is an open, dense set. However, $\bigcap_k D_k = \mathbb{R}\setminus\mathbb{Q}$ is the set of all irrational numbers. Although it is dense, it is not open any more.

Second, that the set $\bigcup_j \overline{E}_j$ has no interior in particular implies $X \setminus \bigcup_j \overline{E}_j$ is nonempty, that is, it is impossible to decompose a complete metric space into a countable union of

nowhere dense subsets.

Third, the above remark may be formulated as, if X is complete and $X = \bigcup_j A_j$ where A_j are closed, then one of the A_j's must contain a ball.

When we describe the size of a set in a metric space, we could use the notion of a dense set or a nowhere dense set. However, sometimes some description is too rough. For instance, consider $\mathbb{Q}, \mathbb{I}$ and Y, the set obtained by removing finitely many points from $\mathbb{R}$. All of them are dense in $\mathbb{R}$. However, everyone should agree that they are are very different. $\mathbb{Q}$ is countable, $\mathbb{I}$ is uncountable and Y is open. From a measure-theoretic point of view, $\mathbb{Q}$ is a set of measure zero and yet $\mathbb{I}$ has infinite measure. Y should be "more dense" than $\mathbb{I}$, and $\mathbb{I}$ "more dense" than $\mathbb{Q}$. Thus simply calling them dense sets is not precise enough. Baire category theorem enables us to make a more precise description of the size of a set in a complete metric space. A set in a metric space is called **of first category** if it can be expressed as a countable union of nonwhere dense sets. Note that by definition any subset of a set of first category is again of first category. A set is **of second category** if its complement is of first category. According to Baire category theorem, a set of second category is a dense set when the underlying space is complete.

Proposition 3.13. *If a set in a complete metric space is of first category, it cannot be of second category, and vice versa.*

Proof. Let E be of first category and let $E = \cup_{k=1}^{\infty} E_k$ where E_k are nowhere dense sets. If it is also of second category, its complement is of first category. Thus, $X \setminus E = \cup_{k=1}^{\infty} F_k$ where F_k are nowhere dense. It follows that $X = E \cup (X \setminus E) = \cup_k (E_k \cup F_k)$ so the entire space is a countable union on of nowhere dense sets, contradicting the completeness of the space and the Baire category theorem. □

Applying to $\mathbb{R}$, we see that $\mathbb{Q}$ is of first category and $\mathbb{I}$ is of second category although they both are dense sets. Indeed, $\mathbb{Q} = \cup_k \{x_k\}$ where k runs through all rational numbers and $\mathbb{I} = \mathbb{R} \setminus \mathbb{Q}$ is of second category.

Baire category theorem has many applications. We end this section by giving two standard ones. It is concerned with the existence of continuous, but nowhere differentiable functions. We knew that Weierstrass is the first person who constructed such a function in 1896. His function is explicitly given in the form of an infinite series

$$W(x) = \sum_{n=1}^{\infty} \frac{\cos 3^n x}{2^n}.$$

Here we use an implicit argument to show there are far more such functions than continuously differentiable functions.

We begin with a lemma.

Lemma 3.14. *Let $f \in C[a,b]$ be differentiable at x. Then it is Lipschitz continuous at x.*

Proof. By differentiability, for $\varepsilon = 1$, there exists some δ_0 such that

$$\left| \frac{f(y) - f(x)}{y - x} - f'(x) \right| < 1, \quad \forall y \neq x, |y - x| < \delta_0.$$

We have

$$|f(y) - f(x)| \leq L|y - x|, \quad \forall y, |y - x| < \delta_0,$$

where $L = |f'(x)| + 1$. For y lying outside $(x - \delta_0, x + \delta_0)$, $|y - x| \geq \delta_0$. Hence

$$\begin{aligned} |f(y) - f(x)| &\leq |f(x)| + |f(y)| \\ &\leq \frac{2M}{\delta_0}|y - x|, \quad \forall y \in [a,b] \setminus (x - \delta_0, x + \delta_0), \end{aligned}$$

where $M = \sup\{|f(x)| : \ x \in [a,b]\}$. It follows that f is Lipschitz continuous at x with an Lipschitz constant not exceeding $\max\{L, 2M/\delta_0\}$.

□

Proposition 3.15. *The set of all continuous, nowhere differentiable functions forms a set of second category in $C[a,b]$ and hence dense in $C[a,b]$.*

Proof. For each $L > 0$, define

$$S_L = \big\{ f \in C[a,b] : \ f \text{ is Lipschitz continuous at some } x \text{ with the Lipschitz constant } \leq L \big\}.$$

We claim that S_L is a closed set. For, let $\{f_n\}$ be a sequence in S_L which is Lipschitz continuous at x_n and converges uniformly to f. By passing to a subsequence if necessary, we may assume $\{x_n\}$ to some x^* in $[a,b]$. We have, by letting $n \to \infty$,

$$\begin{aligned} |f(y) - f(x^*)| &\leq |f(y) - f_n(y)| + |f_n(y) - f(x^*)| \\ &\leq |f(y) - f_n(y)| + |f_n(y) - f_n(x_n)| + |f_n(x_n) - f_n(x^*)| + |f_n(x^*) - f(x^*)| \\ &\leq |f(y) - f_n(y)| + L|y - x_n| + L|x_n - x^*| + |f_n(x^*) - f(x^*)| \\ &\to L|y - x^*| \end{aligned}$$

Next we show that each S_L is nowhere dense. Let $f \in S_L$. By Weierstrass approximation theorem, for every $\varepsilon > 0$, we can find some polynomial p such that $\|f - p\|_\infty < \varepsilon/2$. Since every polynomial is Lipschitz continuous, let the Lipschitz constant of p be L_1. Consider the function $g(x) = p(x) + (\varepsilon/2)\varphi(x)$ where φ is the jig-saw function of period r satisfying $0 \leq \varphi \leq 1$ and $\varphi(0) = 1$. The slope of this function is either $1/r$ or $-1/r$. Both will become large when r is chosen to be small. Clearly, we have $\|f - g\|_\infty < \varepsilon$. On the other hand,

$$\begin{aligned} |g(y) - g(x)| &\geq \frac{\varepsilon}{2}|\varphi(y) - \varphi(x)| - |p(y) - p(x)| \\ &\geq \frac{\varepsilon}{2}|\varphi(y) - \varphi(x)| - L_1|y - x| \ . \end{aligned}$$

For each x sitting in $[a,b]$, we can always find some y nearby so that the slope of φ over the line segment between x and y is greater than $1/r$ or less than $-1/r$. Therefore, if we choose r so that

$$\frac{\varepsilon}{2}\frac{1}{r} - L_1 > L,$$

we have $|g(y) - g(x)| > L|y - x|$, that is, g does not belong to S_L.

Denoting by S the set of functions in $C[a,b]$ which are differentiable at at least one point, by Lemma 3.15 it must belong to S_L for some L. Therefore, $S \subset \cup_{k=1}^{\infty} S_k$ is of first category. □

Though elegant, a drawback of this proof is that one cannot assert which particular function is nowhere differentiable. On the other hand, the example of Weierstrass is a concrete one.

Our second application is concerned with the basis of a vector space. Recall that a basis of a vector space is a set of linearly independent vectors such that every vector can be expressed as a linear combination of vectors from the basis. The construction of a basis in a finite dimensional vector space was done in MATH2040. However, in an infinite dimensional vector space the construction of a basis is difficult. Nevertheless, using Zorn's lemma, a variant of the axiom of choice, one shows that a basis always exists. Some authors call a basis for an infinite dimensional basis a Hamel basis. The difficulty in writing down a Hamel basis is explained in the following result.

Proposition 3.16. *Any basis of an infinite dimensional Banach space consists of uncountably many vectors.*

Proof. Let V be an infinite dimensional Banach space and $\mathcal{B} = \{w_j\}$ be a countable basis. We aim for a contradiction. Indeed, let W_n be the subspace spanned by $\{w_1, \cdots, w_n\}$. We have the decomposition

$$V = \bigcup_n W_n.$$

If one can show that each W_n is closed and nowhere dense, since V is complete, Baire category theorem tells us this decomposition is impossible. To see that W_n is nowhere dense, pick a unit vector v_0 outside W_n. For $w \in W_n$ and $\varepsilon > 0$, $w + \varepsilon v_0 \in B_\varepsilon(w) \cap (V \setminus W_n)$, so W_n is nowhere dense. Next, letting v_j be a sequence in W_n and $v_j \to v_0$, we would like to show that $v \in W_n$. Indeed, every vector $v \in W_n$ can be uniquely expressed as $\sum_{j=1}^n a_j w_j$. The map $v \mapsto a \equiv (a_1, \cdots, a_n)$ sets up a linear bijection between W_n and $\mathbb{R}^n$ and $|||a||| \equiv \|v\|$ defines a norm on $\mathbb{R}^n$. Since any two norms in $\mathbb{R}^n$ are equivalent, a convergent (resp. Cauchy) sequence in one norm is the same in the other norm. Since now $\{v_j\}$ is convergent in V, it is a Cauchy sequence in V. The corresponding sequence $\{a^j\}$, $a^j = (a_1^j, \cdots, a_n^j)$, is a Cauchy sequence in $\mathbb{R}^n$ with respect to $|||\cdot|||$ and hence in $\|\cdot\|_2$, the Euclidean norm. Using the completeness of $\mathbb{R}^n$ with respect to the Euclidean

norm, $\{a^j\}$ converges to some $a^* = (a_1^*, \cdots, a_n^*)$. But then $\{v_j\}$ converges to $v^* = \sum_j a_j^* w_j$ in W_n. By the uniqueness of limit, we conclude that $v_0 = v^* \in W_n$, so W_n is closed. □

Comments on Chapter 3. Three properties, namely, separability, compactness, and completeness of the space of continuous functions are studied in this chapter.

Separability is established by various approximation theorems. For the space $C[a, b]$, Weierstrass approximation theorem is applied. Weierstrass (1885) proved his approximation theorem based on the heat kernel, that is, replacing the kernel Q_n in our proof in Section 1 by the heat kernel. The argument is a bit more complicated but essentially the same. It is taken from Rudin, Principles of Mathematical Analysis. A proof by Fourier series is already presented in Chapter 1. Another standard proof is due to Bernstein, which is constructive and had initiated a branch of analysis called approximation theory. The Stone-Weierstrass theorem is due to M.H. Stone (1937, 1948). We use it to establish the separability of the space $C(X)$ where X is a compact metric space. You can find more approximation theorem by web-surfing.

Arzela-Ascoli theorem plays the role in the space of continuous functions the same as Bolzano-Weierstrass theorem does in the Euclidean space. A bounded sequence of real numbers always admits a convergent subsequence. Although this is no longer true for bounded sequences of continuous functions on $[a, b]$, it does hold when the sequence is also equicontinuous. Ascoli's theorem (Proposition 3.11) is widely applied in the theory of partial differential equations, the calculus of variations, complex analysis and differential geometry. Here is a taste of how it works for a minimization problem. Consider

$$\inf \left\{ J[u] : \ u(0) = 0, u(1) = 5, \ u \in C^1[0, 1] \right\},$$

where

$$J[u] = \int_0^1 \left(u'^2(x) - \cos u(x) \right) dx.$$

First of all, we observe that $J[u] \geq -1$. This is clear, for the cosine function is always bounded by ± 1. After knowing that this problem is bounded from -1, we see that $\inf J[u]$ must be a finite number, say, γ. Next we pick a minimizing sequence $\{u_n\}$, that is, every u_n is in $C^1[0, 1]$ and satisfies $u(0) = 0, u(1) = 5$, such that $J[u_n] \to \gamma$ as $n \to \infty$. By

Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\big|u_n(x) - u_n(y)\big| &\le \int_x^y \big|u_n'(x)\big|dx \\
&\le \sqrt{\int_x^y 1^2 dx}\sqrt{\int_x^y u_n'^2(x)dx} \\
&\le \sqrt{\int_x^y 1^2 dx}\sqrt{\int_0^1 u_n'^2(x)dx} \\
&\le \sqrt{J[u_n]+1}\sqrt{|y-x|} \\
&\le \sqrt{\gamma+2}\ |y-x|^{1/2}
\end{aligned}$$

for all large n. From this estimate we immediately see that $\{u_n\}$ is equicontinuous and bounded (because $u_n(0) = 0$). By Ascoli's theorem, it has a subsequence $\{u_{n_j}\}$ converging to some $u \in C[0,1]$. Apparently, $u(0) = 0, u(1) = 5$. Using knowledge from functional analysis, one can further argue that $u \in C^1[0,1]$ and is the minimum of this problem.

Arzela showed the necessity of equicontinuity and boundedness for compactness while Ascoli established the compactness under equicontinuity and boundedness. Google under Arzela-Ascoli theorem for details.

There are some fundamental results that require completeness. The contraction mapping principle is one and Baire category theorem is another. The latter was first introduced by Baire in his 1899 doctoral thesis. It has wide, and very often amazing applications in all branches of analysis. Some nice applications are available on the web. Google under applications of Baire category theorem for more.

Weierstrass' example is discussed in some detailed in Hewitt-Stromberg, "Abstract Analysis". An simpler example can be found in Rudin's Principles.

Being unable to locate a single reference containing these three topics, I decide not to name any reference but let you search through the internet.