Lecture Notes M 517
Introduction to Analysis

Donald J. Estep

Department of Mathematics
Colorado State University
Fort Collins, CO 80523

estep@math.colostate.edu
http://www.math.colostate.edu/∼estep

Contents

CHAPTER 1

Introduction

What is analysis? The definition depends on your point of view. To an applied mathematician, analysis means approximation and estimation. They try to describe the physical world in mathematical terms so as to understand how it works and make predictions on future behavior/events. But, the world is so complex, we have to resort to approximations both to describe it and to understand the mathematical descriptions themselves. To a pure mathematician, analysis is the study of the limiting behavior of an an infinite process.

Many mathematical objects, such as numbers, derivatives, and integrals, are defined as the limit of an infinite process. Dealing with such limits in a rigorous way is what disinguishes modern mathematics, beginning around the time of Newton and Leibniz, from classical mathematics.

These are really the same notions; approximation and estimation requires the idea of a limit and, likewise, a limit implies the ability to approximate.

The analysis we learn in this course makes up a basic bag of tools for a working mathematician. These tools are needed for any deeper study in applied mathematics, differential equations, topology, functional analysis,... . As with any kind of tool, you can only really learn them by using them, i.e. by doing problems.

1.1. Review of Real Numbers

This course is like an advanced calculus course, except that we work with the more general notion of a "metric space" rather than only the real numbers. Examples include $\mathbb{R}$ as well as spaces of functions. We first talk about the structure of the space itself - as we might begin a rigorous caclulus course by constructing real numbers and discussing sequences - and then we develop properties of functions on our space. It will probably be useful for you to compare our developments to what holds for reals and functions on reals.

Recall

Definition 1.1.1. A **sequence** of real numbers

$$(a_1, a_2, ...) = \{a_i\}_{i=1}^{\infty} = \{a_{bike}\}_{bike=1}^{\infty}$$

is a set of numbers listed in a specific order. Here, "∞" denotes a neverending list. The subscript "i" or "$bike$" is called the **index**.

Example 1.1.2.

$$\frac{1}{9} = .11111\bar{1}$$

So,

$$(.1, .11, .111, ..., \underbrace{.11...1_n}_{n\ 1's}, ...)$$

is a sequence of reals that approximates $\frac{1}{9}$ in the sense that

$$\left|\frac{1}{9} - .1...1_n\right| < 10^{-n}$$

for all n.

Sequences can have many forms.

Example 1.1.3.

$$(2, 4, 6, ...) = \{2n\}_{n=1}^{\infty}$$
$$(-1, 1, -1, 1, ...) = \{(-1)^n\}_{n=1}^{\infty}$$
$$(1, \frac{1}{2}, \frac{1}{3}, ...) = \{\frac{1}{n}\}_{n=1}^{\infty}$$
$$(1, 1, 1, ...) = \{1\}_{n=1}^{\infty}$$
$$(3, 4.182, \frac{3}{5}, -14.68, ...) = \{?\}_{n=1}^{\infty}$$

Some of these sequences have the special property that as the index increases, the corresponding numbers become closer and closer to one number.

Definition 1.1.4. If the numbers in the sequence $\{a_n\}$ approach a number a as n increases, then we call a the **limit** of the sequence and write

$$\lim_{n\to\infty} a_n = a \text{ and } a_n \xrightarrow[n\to\infty]{} a$$

We say the sequence **converges**.

Example 1.1.5.

$$\lim_{n\to\infty} .11...1_n = \frac{1}{9}$$
$$\lim_{n\to\infty} \frac{1}{n} = 0$$
$$\lim_{n\to\infty} 5 = 5$$

To make this definition more precise, we use mathematical language:

Definition 1.1.6. $\lim\limits_{n\to\infty} a_n = a$ means that for every $\epsilon > 0$, there is an N such that $|a_n - a| < \epsilon$ for $n \geq N$.

Exercise 1.1.7. Verify $\lim\limits_{n\to\infty} .111...1_n = \dfrac{1}{9}$. (Find a concrete relation between the ϵ and the N.)

Definition 1.1.8. If a sequence does not converge, we say it **diverges**.

There is a myriad of interesting behavior associated with divergence. Here are some examples of diverging sequences:

Example 1.1.9.

$$(1, 2, 3, 4, ...)$$
$$(1, -1, 1, -1, ...)$$
$$(\text{random numbers})$$

After making this definition, the usual thing to do is to prove some basic properties. Some of these are:

THEOREM 1.1.10. *Limits of Sequences of Real Numbers*

(1) *Limits are unique.*
(2) *Arithmetic Properties: Suppose* $a_n \to a$ *and* $b_n \to b$. *Then*
 (a) $a_n + b_n \to a + b$.
 (b) $a_n b_n \to ab$.

An important concept related to convergence is the notion of a Cauchy sequence.

DEFINITION 1.1.11. A sequence $\{a_n\}$ of real numbers is a **Cauchy Sequence**, or is **Cauchy**, if for every $\epsilon > 0$, there is an N such that

$$|a_i - a_j| < \epsilon \text{ for } i, j > N.$$

The interest in the Cauchy criterion (Def. 1.1.11) is that *it does not require the value of the limit*, i.e., what the sequence converges to. The standard criterion for convergence is practically useless. We are able to prove

THEOREM 1.1.12. *A sequence of real numbers converges if and only if it is Cauchy.*

Sequences and limits are inherent to the notion of the real number system. Recall that rational numbers, in general, have infinite decimal expansions. But, at least these decimal expansions are periodic, e.g., .123412341234... . Most real numbers (in a precise sense) have infinite, non-repeating decimal expansions. These are called the irrational numbers. All kinds of difficulties arise because of this fact. For example, how should we add the irrational numbers e and $\sqrt{2}$?

We add numbers with finite decimal expansions by starting at the rightmost digit. But, that is impossible with irrationals! What we do in practice is add finite truncations, or approximations, of e and $\sqrt{2}$, and count on this being an approximation of $\sqrt{2} + e$.

In fact, this is exactly the way we construct real numbers from the rationals, which we understand better. Just as with "+" above, the main point is to show the real numbers inherit the usual properties of rational numbers. The tricky part is that we can't write down reals until we have defined them! This is why the Cauchy criterion is so important.

All of these difficulties carry over to defining functions of real numbers (how should we compute $\sqrt{e}$?) and the central role of convergence is why continuity is such an important property.

In this course, we replace real numbers by a more abstract space, which includes $\mathbb{R}$ and $\mathbb{R}^n$ as well as other collections like functions. Then we develop the basic properties of the space and functions on the space.

1.2. Functions and Sets

DEFINITION 1.2.1. Let A , B be two sets and suppose to each element a in A there is associated one element b in B, which we write as $b = f(a)$. f is a **function** or **map** from A into B. We write $f : A \to B$.

Definition 1.2.2. Let A, B be sets and $f : A \to B$. If $C \subset A$ (read C is a subset of A), then we define

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

to be the **image** of C under f. We call A the **domain** of f and B the **range** of f.

Definition 1.2.3. If A, B are sets and $f(A) = B$, we say f maps A **onto** B.

Definition 1.2.4. Suppose A, B are sets and $f : A \to B$. If $C \subset B$, then

$$f^{-1}(C) = \{a \in A \mid f(a) \in C\}$$

is the **inverse image** of C under f. $f^{-1}(C)$ may be the empty set.

Definition 1.2.5. Let A, B be sets and $f : A \to B$. If $y \in B$, then $f^{-1}(y) = \{a \in A \mid f(a) = y\}$. If for each $y \in B$, $f^{-1}(y)$ consists of at most one element of A, then f is a **1-1 (one-to-one) map** of A into B. This is equivalent to saying that

$$f(a_1) \neq f(a_2) \text{ when } a_1, a_2 \in A, a_1 \neq a_2.$$

The concept of a function is fundamental in many ways. As an appplication:

Definition 1.2.6. A **sequence** in a set A is a function f from the natural numbers $\mathbb{N}$ into A. If $f(n) = a_n$ for $n \in N$, then we write the sequence as

$$\{a_n\}_{n=1}^{\infty} = (a_1, a_2, ...)$$

The values of f are called the **terms** or **elements** of the sequence.

> *Note:* $\mathbb{N}$ *may be replaced by* $\mathbb{N} \cup \{0\} = \mathbb{Z}_0$ *or even* $\mathbb{Z}$. *Also, the* a_n *need not be distinct.*

Infinite sets and sequences have some nice properties that make them relatively easy to use, as we will see. However, we do have to work with more complicated sets in analysis, and though it is somewhat unfamiliar, the index notation can be generalized to help out:

Definition 1.2.7. Let A and B be sets and suppose that for each $a \in A$ there is associated a subset of B called C_a. The set whose elements are the sets C_a for $a \in A$ is denoted

$$\{C_a \mid a \in A\} = \{C_a\}_{a \in A} = \{C_a\}$$

Example 1.2.8. Let $A = [-1, 1]$ and $C_a = \{x : \sin(x) = a\}$ for $a \in A$.. For example,

$$C_0 = \{..., -3\pi, -2\pi, -\pi, 0, \pi, 2\pi, ...\}$$

has a lot of points in it!

Exercise 1.2.9. What is $\{C_a\}_{a \in [-1,1]}$?

> Note: the relation $a \to C_a$ in Def. 1.2.7 is not a function of one number to one number. We call such this a **set valued function**.

1.3. Unions and Intersection

DEFINITION 1.3.1. Let A be a set and $\{C_a\}_{a\in A}$ a collection of subsets of A such that $C_a \subset A$. The **union** of the sets $\{C_a\}$ is the set S such that $s \in S$ if and only if $s \in C_a$ for at least one $a \in A$. We write

$$S = \bigcup_{a\in A} C_a.$$

If $A = \mathbb{N}$, then we write

$$S = \bigcup_{m=1}^{\infty} C_m = \bigcup_m C_m = \bigcup C_m.$$

If $A = \{1, 2, ..., n\}$, then we write

$$S = \bigcup_{m=1}^{n} C_m$$

The **intersection** of the sets $\{C_A\}$ is the set T such that $t \in T$ if and only if $t \in C_a$ for all $a \in A$. We write

$$T = \bigcap_{a\in A} C_a = \bigcap_{m=1}^{\infty} C_m = \bigcap_{m=1}^{n} C_m.$$

EXAMPLE 1.3.2.

$$\{-1, 2, 5, 8\} \bigcup \{3, 5, 8, 10\} = \{-1, 2, 3, 5, 8, 10\}$$

$$\{-1, 2, 5, 8\} \bigcap \{3, 5, 8, 10\} = \{5, 8\}$$

EXAMPLE 1.3.3. Let $A = (0, 1]$ and $C_a = (0, a)$ for $a \in A$. Then

(1) $C_a \subset C_b$ if and only if $0 < a \leq b < 1$
(2) $\bigcup_{a\in A} C_a = C_1$
(3) $\bigcap_{a\in A} C_a = \emptyset$ (the empty set)

EXERCISE 1.3.4. What is $\bigcup_{a\in A} C_a$ and $\bigcap_{a\in A} C_a$ in Example 1.2.8?

DEFINITION 1.3.5. If A and B are two sets and $A \bigcap B$ is not empty, then we say A and B **intersect**. Otherwise, they are **disjoint**.

There are many important properties of $\bigcup$ and $\bigcap$ that you can read on page 28 of Rudin's "Principles of Mathematical Analysis".

1.4. Cardinality

Cardinality refers to the "number of points" in a set. "Number of points" is in quotes because we want to talk about infinite sets as well. Moreover, we want to distinguish different kinds of infinite sets.

EXAMPLE 1.4.1. { 1, 2, 5,π, -8}, $\mathbb{Z}$, and $\mathbb{R}$ turn out to have different cardinalities.

We start the discussion by developing a mechanism for comparing the number of elements between two sets. We use functions:

DEFINITION 1.4.2. If there is a 1-1 map of a set A onto a set B, then we say that A and B **are in 1-1 correspondence**, **have the same cardinality or cardinal number**, or **are equivalent**, and we write A B.

This relation has the following (obvious) properties:

THEOREM 1.4.3. *Let A, B, C be sets. Then*

(1) $A \sim A$ (reflexive)
(2) $A \sim B \Leftrightarrow B \sim A$ (symmetric)
(3) *If $A \sim B$ and $B \sim C$, then $A \sim C$.* (transitive)

DEFINITION 1.4.4. A relation with the properties in Thm. 1.4.3 is called an **equivalence relation**.

We now write down the basic classification of cardinality:

DEFINITION 1.4.5. Let A be a set.

(1) A is **finite** if $A \sim \{1, 2, ..., n\}$ for some natural number n or if A is empty.
(2) A is **infinite** if A is not finite.
(3) A is **countable** if $A \sim \mathbb{N}$.
(4) A is **uncountable** if A is neither finite nor countable.
(5) A is **at most countable** if A is finite or countable

Note: some texts use countable to mean at most countable.

EXAMPLE 1.4.6. $\mathbb{Z}$ is countable.

$$\begin{array}{cccccccccc} \mathbb{Z} & 0, & 1, & \text{-1}, & 2, & \text{-2}, & 3, & \text{-3}, & ... \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \mathbb{N} & 1, & 2, & 3, & 4, & 5, & 6, & 7, & ... \end{array}$$

Here,

$$f(n) = \begin{cases} \frac{n}{2} & \text{for } n \text{ even} \\ -\frac{n-1}{2} & \text{for } n \text{ odd} \end{cases}.$$

takes $\mathbb{N}$ to $\mathbb{Z}$ and is 1-1 and onto.

Note: this shows that an infinite set can be equivalent to a proper subset of itself! This only happens on infinite sets.

EXAMPLE 1.4.7. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence with distinct terms, and observe $\{a_{2n}\}_{n=1}^{\infty}$ is a subsequence. Then the function $f : \{a_n\} \to \{a_{2n}\}$ given by $f(a_m) = a_{2m}$ is a bijection and we see $\{a_n\}$ is equivalent to $\{a_{2n}\}$ as sets.

LEMMA 1.4.8. Every countable set can be represented as a sequence.

PROOF. Let A be a countable set. This means $A \sim \mathbb{N}$. So, there exists $f : \mathbb{N} \to A$, a 1-1 and onto map. Let

$$\begin{aligned} a_1 &= f(1) \\ a_2 &= f(2) \\ &\vdots \\ a_n &= f(n) \\ &\vdots \end{aligned}$$

Since f is 1-1, we have A represented as $\{a_n\}_{n=1}^{\infty}$. $\square$

Note: The representation given in Lemma 1.4.8 is not unique!

THEOREM 1.4.9. *Every infinite subset of a countable set is countable.*

PROOF. Let A be countable and suppose $C \subset A$ is infinite. Let A be arranged

$$A = (a_1, a_2, ...)$$

where the $\{a_i\}$ are distinct. Let $n_1 \in \mathbb{N}$ be the smallest integer such that $a_{n_1} \in C$. Having chosen $n_1, n_2, ..., n_{k-1}$ with $k \geq 2$, let n_k be the smallest integer larger than n_{k-1} such that $a_{n_k} \in C$. We let $f : \mathbb{N} \to C$ be given by

$$f(k) = a_{n_k}.$$

Notice f is 1-1 and onto, and therefore C is countable. □

Thm. 1.4.9, along with the fact that every infinite set contains a countable subset, says that countable sets are the "smallest" kind of infinite set in terms of cardinality. The following result is very significant, though this won't be readily apparent until one studies measure theory.

THEOREM 1.4.10. *Suppose $\{E_n\}_{n=1}^{\infty}$ is a sequence of countable sets. Then*

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

is countable.

PROOF. We let E_n be denoted by $\{X_{n_k}\}_{k=1}^{\infty} = \{X_{n_1}, X_{n_2}, ...\}$. Since S is the union of all these elements, we can think of it as an infinite array:

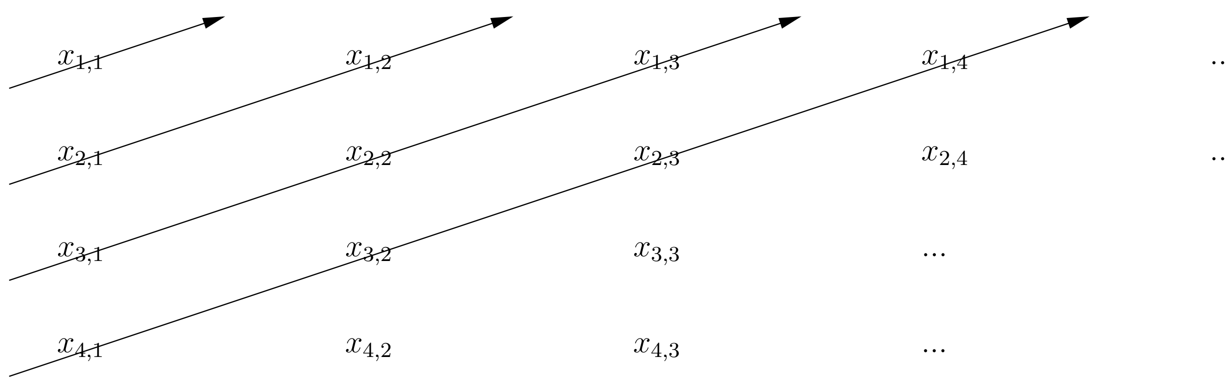


FIGURE 1.1

We want to number the elements in S using $\mathbb{N}$. consider the ordering shown in Figure 1.1, which is listing

$$\begin{array}{cccccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & ... \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & ... \end{array}$$

We can find a $1-1$ and onto map between the elements and $\mathbb{N}$ in this way. □

Note: there is an obvious corollary for at most countable unions of at most countable sets.

Next we show that the rational numbers, $\mathbb{Q}$, are countable. To do this, we show the following:

THEOREM 1.4.11. *Let A be a countable set. Let B_n be the set of n-tuples $(a_1, ..., a_n)$, where $a_i \in A$ for $1 \leq i \leq n$, and the $\{a_i\}$ in an n-tuple need not be distinct. Then B_n is countable.*

PROOF. By induction. Since $B_1 = A$, B_1 is countable. For the induction step, assume B_{n-1} is countable where $n \in \{2, 3, ...\}$. The elements of B_n correspond to $\{(b, a) \mid b \in B_{n-1}, a \in A\}$. In particular, for any fixed $b \in B_{n-1}$, the set $\{(b, a)\} \mid a \in A\} \equiv A$ and hence is countable. Thus, B_n is the countable union of countable sets. □

COROLLARY 1.4.12. $\mathbb{Q}$ is countable.

PROOF. Consider $A = \mathbb{Z}$ and $n = 2$ in Theorem 1.4.11. Then $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}\}$, corresponding to a subset of $\{(a, b) \mid a, b \in \mathbb{Z}\}$. We know the latter set is countable, so by Theorem 1.4.9, we have that $\mathbb{Q}$ is countable. □

Next, we will prove a fact about the cardinality of $\mathbb{R}$.

THEOREM 1.4.13. *$\mathbb{R}$ is not countable.*

PROOF. The proof has one tricky point, which is that a real number can have two decimal expansions. For example, $.2\overline{0} = .1\overline{9}$.

It suffices to show that the set of numbers $\{x \mid 0 < x < 1\} = (0, 1)$ is uncountable. If these numbers were countable, then there would be a sequence $\{s_n\}_{n=1}^{\infty}$ that gives them all. We show this is impossible by constructing a number in $(0, 1)$ but not in $\{s_n\}_{n=1}^{\infty}$.

We write each s_n as a decimal expansion:

$$s_n = 0.d_{n_1} d_{n_2} d_{n_3} ...$$

where $d_{n_i} \in \{0, 1, 2, ..., 9\}$ for each i and n. Now define $y = 0.e_1 e_2 e_3 ...$, where

$$e_m = \begin{cases} 1 & \text{if } d_{n_m} \neq 1 \\ 2 & \text{if } d_{n_m} = 1 \end{cases}.$$

Notice $y \in (0, 1)$, but $y \notin \{s_n\}_{n=1}^{\infty}$ (see Example 1.4.14). □

EXAMPLE 1.4.14. Let

$$s_1 = .\underline{1}23456789...$$
$$s_2 = .2\underline{4}6812461...$$
$$s_3 = .69\underline{1}284823...$$
$$s_4 = .444\underline{4}44444...$$
$$\vdots$$
$$y = .2121...$$

Now y is different than s_1 in the first digit, s_2 in the second, s_3 in the third, and so on. Note that situations like $s_n = .199999...$ and $y = .2000...$ cannot occur by the choice of $\{e_m\}$.

This type of argument is called a *Cantor diagonalization argument*. It is a powerful technique, which can be used to prove the following theorem:

THEOREM 1.4.15. *Let A be the set of all sequences whose elements are the digits 0 and 1. Then A is uncountable.*

This theorem also implies that the reals are uncountable by using binary expansions. It is also important in probability. We can denote an infinite sequence of coin tosses as a sequence of 0's and 1's, and we see that the number of possibilities is uncountable.

CHAPTER 2

Metric Spaces

The real numbers have properties that make it natural to discuss sequences. Indeed, the reals are defined in terms of sequences. We would like to create an abstract notion of a space in which it makes sense to talk about sequences and in which sequences play as an important of a role as they do in the real numbers. In a sense, we would like to abstract the "sequential nature" of the reals, as opposed to their other properties.

Our abstract space should certainly include the reals as an example. It should also include $\mathbb{R}^n$. So, we have to lose the order properties (this is, properties pertaining to the relation "¡") in our abstract space. But, $\mathbb{R}^n$ itself has special properties that will not be present in the abstract notion. Therefore, $\mathbb{R}^n$ will always be a special, important example of our abstract notion of a space.

2.1. Quick Reivew of $\mathbb{R}^n$

Recall that

$$\mathbb{R}^n = \{(x_1, x_2, ..., x_n) \mid x_i \in \mathbb{R}, 1 \leq i \leq n\},$$

where we define many properties of the n-tuple $(x_1, x_2, ..., x_n)$.

There are at least four important structures on $\mathbb{R}^n$:

(1) **Algebraic Structure**

$\mathbb{R}^n$ is a vector space over the scalars $\mathbb{R}$ with the usual definitions of addition, subtraction, and scalar multiplication carried out "component by component" on n-tuples. These definitions satisfy all the properties expected of a vector space.

(2) **Inner Product Spaces**

Consideration of how to "multiply" two vectors in $\mathbb{R}^n$ lead to the dot product, or inner product:

$$(x_1, ..., x_n) \cdot (y_1, ..., y_n) = x_1 y_1 + ... + x_n y_n$$

This defines another structure on $\mathbb{R}^n$, beginning with the fundamental notion of orthogonality. Two vectors are orthogonal if their inner product is zero. Beginning with this definition, we can develop all kinds of properties of $\mathbb{R}^n$.

(3) **Length Structure**

The inner product on $\mathbb{R}^n$ induces a natural length via

$$\parallel s \parallel = \sqrt{x \cdot x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

$\parallel \cdot \parallel$ is called a norm and it has some important properties. For $\alpha \in \mathbb{R}, x, y \in \mathbb{R}^n$,

(a) $\| x \| \geq 0$ and $\| x \| = 0 \Leftrightarrow x = 0$
(b) $\| \alpha x \| = |\alpha| \, \| x \|$
(c) (Triangle Inequality) $\| x + y \| \leq \| x \| + \| y \|$
(d) (Cauchy-Schwarz Inequality)$|x \cdot y| \leq \| x \| \, \| y \|$

(4) **Distance Structure** Finally, the norm induces a natural distance function. We define the distance between x and y in $\mathbb{R}^n$ as

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

This distance function satisfies
(a) $d(x, y) \geq 0$ and $d(x, y) = 0 \Leftrightarrow x = y$
(b) $d(x, y) = d(y, x)$
(c) $d(x, z) \leq d(x, y) + d(y, z)$
for $x, y, z \in \mathbb{R}^n$.

The distance function is certainly critical for talking about the convergence of sequences of points in $\mathbb{R}^n$: Convergence is tied to the notion of points becoming closer and distance between points becoming smaller.

To define our abstract space, we skip past the other properties of $\mathbb{R}^n$ and go right to distance.

2.2. Definition and Initial Examples of a Metric Space

DEFINITION 2.2.1. A set $\mathbf{X}$ whose elements are called **points** is a **metric space** if to any two points $x, y \in \mathbf{X}$ there is associated a real number $d(x, y)$, called the **distance between** x **and** y, such that for $x, y, z \in \mathbf{X}$,,

(1) $d(x, y) > 0$ if $x \neq y$ and
$d(x, y) = 0$ if $x = y$
(2) $d(x, y) = d(y, x)$
(3) $d(x, z) \leq d(x, y) + d(y, z)$.

d is called the **metric** on $\mathbf{X}$. Property (3) is called the **triangle inequality**.

EXAMPLE 2.2.2. $\mathbb{R}$ with $d(x, y) = |x - y|$

EXAMPLE 2.2.3. $\mathbb{R}^n$ with $d(x, y) = \| x - y \| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$

EXAMPLE 2.2.4. $\mathbb{R}^n$ with $d(x, y) = \| x - y \|_1 = \sum_{i=1}^n |x_i - y_i|$
($\| \, \|_1$ is called the 1-norm.)

EXERCISE 2.2.5. Verify $\| \, \|_1$ is a norm and that this defines a metric.

EXAMPLE 2.2.6. $\mathbb{R}^n$ with $d(x, y) = \| x - y \|_\infty = \max_{1 \leq i \leq n} |x_i - y_i|$
($\| \, \|_\infty$ is called the ∞ norm.)

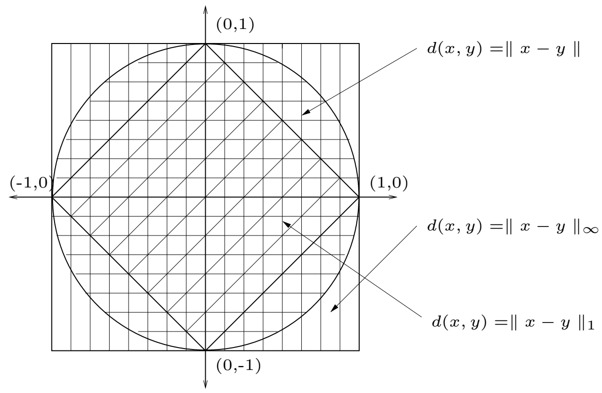
EXERCISE 2.2.7. Verify that $\| \, \|_\infty$ is a norm and that this induces a metric.

This is an interesting development: One space $\mathbf{X}$ might have several difference metrics!

DEFINITION 2.2.8. Let $(\mathbf{X}, d)$ be a metric space. The **ball of radius** r **centered at** $x \in \mathbf{X}$ is defined

$$B_r(x) = \{y \in \mathbf{X} \mid d(y, x) \leq r\}.$$

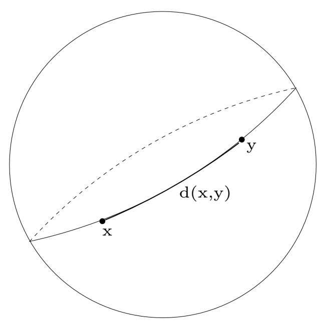
The **unit ball** is $B_1(o)$.



EXAMPLE 2.2.9. It is interesting to plot the unit "balls" in $\mathbb{R}^2$ with respect to the three metrics:

EXERCISE 2.2.10. Reproduce this picture.

EXAMPLE 2.2.11. Let S be a sphere in $\mathbb{R}^3$. We define a metric on its surface by setting $d(x, y)$ to be the shortest distance along the surface of S between x and y (which are on the surface). The shortest path along the surface is called a geodesic.



Showing this is a metric is hard!

DEFINITION 2.2.12. For any set $\mathbf{X}$, we define the **discrete metric** by

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

EXERCISE 2.2.13. Verify that the function given in Def. 2.2.12 is a metric.

DEFINITION 2.2.14. Consider the set of sequences $\{(x_1, x_2, x_3, ...) \mid x_i \in \mathbb{R}\}$. The **Hilbert Space** l_2 is the subset of such sequences for which the terms are "finitely square summable", i.e.,

$$l_2 = \{(x_1, x_2, x_3, ...) \mid x_i \in \mathbb{R} \text{ for } 1 \leq i \text{ and } \sum_{i=1}^{\infty} x_i^2 < \infty\}$$

Notice that the terms in $\sum_{i=1}^{\infty} x_i^2$ are positive. This implies, among other things, that

$$\lim_{i\to\infty} x_i = 0 \text{ and } \sum_{i=1}^{\infty} x_i^2 = \lim_{N\to\infty} \sum_{i=1}^{N} x_i^2$$

In view of the latter fact, we can think of l_2 as something like $(\mathbb{R}^n, \| \cdot \|)$ where "$n \to \infty$". (Recall the usual definitions of +, etc. for sequences.)

THEOREM 2.2.15. *With* $d(x, y) = \sqrt{\sum_{i=1}^{\infty}(x_i - y_i)^2}$ *for* $x, y \in l_2$, (l_2, d) *is a metric space.*

Note: Actually, $\| x \| = \sqrt{\sum_{i=1}^{\infty} x_i^2}$ defines a norm on l_2, and $d(x, y) = \| x - y \|$. However, we do not pursue that.

PROOF. The proof is not so hard, except that we have to be careful to treat the infinite sequences in a mathematically correct way. For example, to verify (1) of Defn. 2.2.1, we want to show that $d(x, y) \geq 0$ and $d(x, y) = 0 \Leftrightarrow x = y$ (two sequences are equal if and only if all the elements are equal) for $x, y \in l_2$.

However, we first have to show that $d(x, y)$ is defined and finite! For this we require the following basic inequality:

LEMMA 2.2.16. Let $a, b \in \mathbb{R}$. Then $2|a||b| \leq a^2 + b^2$.

Proof: $0 \leq (|a| - |b|)^2 = |a|^2 - 2|a||b| + |b|^2$.

Now, for $x, y \in l_2$,

$$\{\sum_{i=1}^{N}(x_i - y_i)^2\}_{n=1}^{\infty}$$

is a monotone, nondecreasing sequence. If we prove the terms are bounded above uniformly for all N, then it has a limit, i.e.,

$$\lim_{N\to\infty} \sum_{i-1}^{\infty}(x_i - y_i)^2$$

is defined and is finite. For any N, we have

$$\begin{aligned} \sum_{i=1}^{N}(x_i - y_i)^2 &\leq 2\sum_{i=1}^{N} x_i^2 + 2\sum_{i=1}^{N} y_i^2 \\ &\leq 2\sum_{i=1}^{\infty} x_i^2 + 2\sum_{i=1}^{\infty} y_i^2, \end{aligned}$$

where the sum on the right is finite by assumption.

We conclude that if $x, y \in l_2$, then $d(x, y)$ is well-defined and finite.

Clearly,

$$d(x, y) = \lim_{N\to\infty} \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \geq 0.$$

If $x = y$, then $d(x, y) = 0$. Vice versa, if $d(x, y) = 0$, then

$$\lim_{N\to\infty} \sum_{i=1}^{N} (x_i - y_i)^2 = 0.$$

But the terms of $\{\sum_{i=1}^{n} (x_i - y_i)^2\}_{N=1}^{\infty}$ are nonnegative and nondecreasing, hence

$$\sum_{i=1}^{N} (x_i - y_i)^2 = 0$$

for all N. This implies $x_i = y_i$ for all i.

For property (2), it is clear that $d(x, y) = d(y, x)$.

Finally, for property (3), suppose $x, y, z \in l_2$. By the properties of the standard Euclidean distance on $\mathbb{R}^n$,

$$\sqrt{\sum_{i=1}^{N} (x_i - z_i)^2} \leq \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} + \sqrt{\sum_{i=1}^{N} (y_i - z_i)^2}$$

for all N. Letting $N \to \infty$, we see

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

□

An interesting fact about l_2 is that it must be "infinite dimensional" (whatever that means!). Another infinite dimensional example:

Definition 2.2.17. We let $\mathcal{C}([a, b])$ denote the vector space of continuous functions on the closed interval $[a, b]$.

Why do I claim this is infinite dimensional? Recall we can write a continuous function on $[a, b]$ in terms of a unique Fourier series, i.e. as a linear combination of the orthonormal Fourier basis functions derived from sin and cos.

Definition 2.2.18. The **maximum norm** or **sup norm** on $\mathcal{C}([a, b])$ is defined as

$$\| f \|_\infty = \max_{a \leq x \leq b} |f(x)|.$$

(Actually, we should use

$$\sup_{a \leq x \leq b} |f(x)|,$$

but in the case of continuous functions, these are equivalent. We will discuss the difference between sup and max later.)

Theorem 2.2.19. *$\mathcal{C}([a, b])$ with $d(f, g) = d_\infty(f, g) = \| f - g \|_\infty$ is a metric space.*

Proof. First, we recall that a continuous function on a closed, bounded interval has a maximum value and a minimum value. Hence, if $f \in \mathcal{C}([a, b])$, then $\| f \|_\infty$ is well defined, and likewise if $f, g \in \mathcal{C}([a, b])$, then $|f(x) - g(x)|$ is continuous on $[a, b]$, and $d(f, g)$ is well-defined. Clearly, $d(f, g) \geq 0$ and $d(f, g) = 0 \Leftrightarrow |f(x) - g(x)| = 0$ for all x, or $f(x) = g(x)$. Also, $d(f, g) = d(g, f)$. Finally, if $f, g, h \in \mathcal{C}([a, b])$, then for any $a \leq x \leq b$,

$$|f(x) - h(x)| \leq |f(x) - g(x)| + |g(x) - h(x)|.$$

So,

$$\max_{a\le x\le b}|f(x)-g(x)| \le \max_{a\le x\le b}(|f(x)-g(x)|+|g(x)-h(x)|)$$
$$\le \max_{a\le x\le b}|f(x)-g(x)|+\max_{a\le x\le b}|g(x)-h(x)|,$$

which shows property (3). □

EXERCISE 2.2.20. Given $f,g\in\mathcal{C}([a,b])$, we have

$$\max_{a\le x\le b}(|f(x)|+|g(x)|)\le \max_{a\le x\le b}|f(x)|+\max_{a\le x\le b}|g(x)|.$$

Construct an example that gives strict inequality in this bound.

2.3. Components of a Metric Space

Let $(\mathbf{X},d)$ be a metric space and $A\subset\mathbf{X}$ a subset.

DEFINITION 2.3.1. A **neighborhood** of a point $x\in\mathbf{X}$ is the set $N_r(x)$ of all points $y\in\mathbf{X}$ with $d(x,y)<r$.

EXAMPLE 2.3.2. In $\mathbb{R}^n$, $N_r(x)$ is the open ball of radius r centered at x. In the discrete metric, $N_r(x)=x$ when $r\le 1$ and $N_r(x)=\mathbf{X}$ when $r>1$.

DEFINITION 2.3.3. A point x is a **limit point** of A if every neighborhood of x contains a point $y\ne x$ in A.

EXAMPLE 2.3.4. Suppose $a,b,c\in(\mathbb{R},|\cdot|)$ with $a<b<c$. Then any $x\in(a,b)$ is a limit point of (a,b). a and b are limit points of (a,b). c is not a limit point of (a,b).

DEFINITION 2.3.5. If $x\in A$ and x is not a limit point of A, then x is an **isolated point** of A.

EXAMPLE 2.3.6. In example 2.3.4, if $A=(a,b)\bigcup\{c\}$, then c is an isolated point of A.

DEFINITION 2.3.7. A is **closed** if every limit point of A belongs to A.

EXAMPLE 2.3.8. In $(\mathbb{R},|\cdot|)$, $[a,b]$ is closed while (a,b), $(a,b]$, and $[a,b)$ are not.

DEFINITION 2.3.9. A point $x\in A$ is an **interior point** if there is a neighborhood $N_r(x)\subset A$ (for some r, not all r!).

EXAMPLE 2.3.10. In $(\mathbb{R},|\cdot|)$, any $x\in(a,b)$ is an interior point of (a,b), but a and b are not.

DEFINITION 2.3.11. A is **open** if every point of A is an interior point.

EXAMPLE 2.3.12. In $(\mathbb{R},|\cdot|)$, (a,b) is open, but $[a,b]$, $[a,b)$, and $(a,b]$ are not.

DEFINITION 2.3.13. The **complement** A^c of A is the set of points in $\mathbf{X}$ but not in A. This is also written $A^c=\mathbf{X}\smallsetminus A$.

Before we define further components, we develop some basic connections between neighborhoods, open sets, and closed sets.

THEOREM 2.3.14. *Any neighborhood is open.*

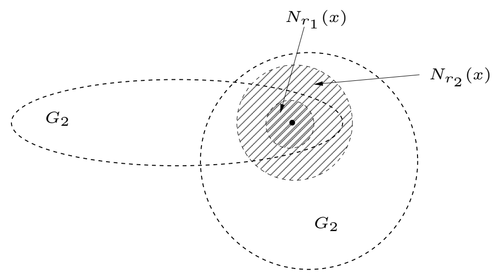


FIGURE 2.1

PROOF. Let $N_r(x)$ be a neighborhood of $x \in \mathbf{X}$ and choose $y \in N_r(x)$. Set $\rho = r - d(x,y) > 0$. For all z such that $d(y,z) < \rho$, we have

$$d(x,z) \leq d(x,y) + d(y,z) < r - \rho + \rho = r,$$

so $z \in N_r(x)$. In other words, $N_\rho(y) \subset N_r(x)$. So, y is an interior point of $N_r(x)$. □

THEOREM 2.3.15.

(1) *Let $\{A_\alpha\}$ be a finite or infinite collection of sets $A_\alpha \subset \mathbf{X}$. Then*

$$(\bigcup_\alpha A_\alpha)^c = \bigcap_\alpha A_\alpha^c.$$

(2) *A set $A \subset \mathbf{X}$ is open if and only if A^c is closed.*
(3) *If $\{G_\alpha\}$ is a collection of open sets, then $\bigcup_\alpha G_\alpha$ is open.*
(4) *If $\{F_\alpha\}$ is a collection of closed sets, then $\bigcap_\alpha F_\alpha$ is closed.*
(5) *If $\{G_1, ..., G_n\}$ is a finite collection of open sets, then $\bigcap_{i=1}^n G_i$ is open.*
(6) *If $\{F_1, ..., F_n\}$ is a finite collection of closed sets, then $\bigcup_{i=1}^n F_i$ is closed.*

PROOF.

(1) If $x \in (\bigcup_\alpha A_\alpha)^c$, then $x \notin \bigcup_\alpha A_\alpha$, i.e., $x \notin A_\alpha$ for any α. So, $x \in A_\alpha^c$ for all α. Hence, $(\bigcup_\alpha A_\alpha)^c \subset \bigcap_\alpha A_\alpha^c$. The reverse relation is very similar to prove.

   EXERCISE 2.3.16. Prove $\bigcap_\alpha A_\alpha^c \subset (\bigcup_\alpha A_\alpha)^c$.

(2) Suppose A^c is closed and $x \in A$. $x \notin A^c$ and so x is not a limit point of A^c. This implies there is a neighborhood N of x such that $A^c \bigcap N$ is empty. So, $N \subset A$. Hence, x is an interior point of A and A is open.

   If A is open and x is a limit point of A^c, then every neighborhood of x contains a point of A^c, so x is not an interior point of A. Since A is open, $x \in A^c$. So, A^c is closed.
(3) Set $G = \bigcup_\alpha G_\alpha$. If $x \in G$, then $x \in G_\alpha$ for some α. x is an interior point of G_α, which implies x is an interior point of G, and so G is open.
(4) Now, $(\bigcap_\alpha F_\alpha)^c = \bigcup_\alpha F_\alpha^c$, and F_α^c is open for all α. So, $(\bigcap_\alpha F_\alpha)^c$ is open and $\bigcap_\alpha F_\alpha$ is closed.
(5) Set $G = \bigcap_{i=1}^n G_i$. For $x \in G$, there are neighborhoods $N_{r_i}(x) \subset G_i, i = 1, ..., n$, for some radii $\{r_i\}_{i=1}^n$. Set $r = \min\{r_1, ..., r_n\}$, so $N_r(x) \subset G_i, 1 \leq i \leq n$. $N_r(x) \subset G$, so G is open. (See Figure 2.1.)
(6)

EXERCISE 2.3.17. Use complements in (5) to prove this.

□

Finiteness of the collection is essential in (5) and (6) of Thm. 2.3.15.

EXAMPLE 2.3.18. If $G_n = (-\frac{1}{n}, \frac{1}{n})$ in $(\mathbb{R}, |\cdot|)$, then $\bigcap_{n=1}^{\infty} G_n = \{0\}$ is not open.

EXERCISE 2.3.19. Construct an analogous example for (6).

DEFINITION 2.3.20. If $A \subset \mathbf{X}$, where $(\mathbf{X}, d)$ is a metric space, then let A' be the set of limit points of A. The **closure** of A is $\overline{A} = A \bigcup A'$.

EXAMPLE 2.3.21. In $(\mathbb{R}, |\cdot|)$, $\overline{(a, b)} = [a, b]$.

THEOREM 2.3.22. *Let $A \subset \mathbf{X}$, where $(\mathbf{X}, d)$ is a metric space. Then*

(1) *$\overline{A}$ is closed.*
(2) *$A = \overline{A} \Leftrightarrow A$ is closed.*
(3) *$\overline{A} \subset F$ for every closed set $F \subset \mathbf{X}$ such that $A \subset F$.*

(3) says that $\overline{A}$ is the "smallest" closed set in $\mathbf{X}$ that contains A.

PROOF.

(1) If $x \in \mathbf{X}$ and $x \notin \overline{A}$, then x is not in A and is not a limit point of A. Hence, x has a neighborhood in $\mathbf{X}$ that does not intersect A. So, $(\overline{A})^c$ is open.
(2) If $A = \overline{A}$, then (1) implies A is closed. If A is closed, then $A' \subset A$ and $A = A \bigcup A' = \overline{A}$.
(3) If F is closed and $A \subset F$, then since $F \supset F'$, $F \supset A'$, and so $F \supset \overline{A}$. This has an important consequence in the case of $(\mathbb{R}, |\cdot|)$.

□

THEOREM 2.3.23. *Let A be a nonempty set of real numbers that is bounded above. Let $y = \sup A$. Then, $y \in \overline{A}$ and $y \in A$ if A is closed.*

(We say that bounded closed sets of real numbers have maximum values.)

PROOF. If $y \in A$, then $y \in \overline{A}$. Assume $y \notin A$. For $r > 0$, there is an $x \in A$ such that $y - r < x < y$ (otherwise $y - r$ would be an upper bound for A). So, y is a limit point of A! Hence, $y \in \overline{A}$. □

In general, boundedness is often important.

DEFINITION 2.3.24. A set $A \subset \mathbf{X}$, where $(\mathbf{X}, d)$ is a metric space, is **bounded** if there is a point $y \in \mathbf{X}$ and a number M such that

$$d(x, y) < M$$

for all $x \in A$. In other words, $A \subset N_M(y)$.

EXAMPLE 2.3.25. $\mathbb{R}^n$ with d_2, d_1, or d_∞ is not bounded, but $B_r(x)$ in any of these metric spaces is bounded.

We next characterize limit points in terms of sequences. First, a preliminary result:

THEOREM 2.3.26. *Let $A \subset \mathbf{X}$, where $(\mathbf{X}, d)$ is a metric space. Then $x \in \mathbf{X}$ is a limit point of A if and only if every neighborhood of x contains infinitely many points of A. In particular, a finite set of points has no limit points.*

PROOF. The "if" direction is just the definition. Suppose x has a neighborhood N that contains only a finite number of points of A. Call the points $\{y_1, ..., y_n\}$ in $A \bigcap N$ that are distinct from x and set

$$r = \min_{1 \leq m \leq n} d(x, y_m) > 0.$$

Now $N_r(x)$ contains no point x of A such that $x \neq z$. Hence, x cannot be a limit point of A. □

Question: Why can we use "min" instead of "inf", and why is this important?

DEFINITION 2.3.27. A **sequence** of points in a metric space $(\mathbf{X}, d)$ is an ordered set of points $\{x_1, x_2, x_3, ...\}$ with $x_i \in \mathbf{X}$ for all i. If $x_i \in A \subset \mathbf{X}$ for all i, then we say the sequence is in A and write $\{x_n\} \in A$.

DEFINITION 2.3.28. A sequence $\{x_n\}_{n=1}^{\infty}$ in a metric space $(\mathbf{X}, d)$ converges to the limit $x \in \mathbf{X}$, written

$$\lim_{n \to \infty} x_n = x \text{ or } x_n \to x,$$

if for every $\epsilon > 0$ there exists an N such that $d(x_n, x) < \epsilon$ for $n \geq N$.

The characterizations we seek are

THEOREM 2.3.29. *Let $A \subset \mathbf{X}$, where $(\mathbf{X}, d)$ is a metric space. Then*

(1) *A is closed if and only if every convergent sequence in A has its limit in A.*
(2) *x is a limit point of A if and only if there is a sequence of* distinct *points $\{x_n\}$ in A such that $x_n \to x$.*

PROOF. (1) Suppose A is closed , $\{x_n\} \in A$, and $x_n \to x$. If $B = \{x_n\}$ is infinite, then every neighborhood of x contains an infinite number of points of B hence of A. So, x is a limit point of A and $x \in A$.

For the converse, let x be a limit point of A. choose $x_1 \in N_1(x) \bigcap A, x_1 \neq x$. Suppose $x_1, ..., x_n$ have been chosen to be distinct points in $\mathbf{X}$ such that

$$x_i \in N_{\frac{1}{i}}(x) \bigcap A$$

for $1 \leq i \leq n$. We choose $x_{n+1} \in N_{\frac{1}{n+1}}(x) \bigcap A$ such that $x_{n+1} \neq x_1, x_2, ..., x_n$. Then, $\{x_n\}$ is a sequence of distinct points in A such that $x_n \to x$. By assumption, $x \in A$, so A is closed.

(2) We actually proved the "only if" above in the proof of (1). If $\{x_n\}$ is a sequence of distinct points in A such that $x_n \to x$, then every neighborhood of x contains an infinite number of the x_n's, so x is a limit point of A.

□

There are a couple of definitions that will be useful later on:

DEFINITION 2.3.30. If $(\mathbf{X}, d)$ is a metric space, $A \subset \mathbf{X}$ is **dense** in $\mathbf{X}$ if every point in $\mathbf{X}$ is a limit point of A or is in A or both.

EXAMPLE 2.3.31. $\mathbb{Q}$ is dense in $\mathbb{R}$.

DEFINITION 2.3.32. A subset $A \subset \mathbf{X}$ of a metric space $(\mathbf{X}, d)$ is **perfect** if it is closed and every point in A is a limit point of A.

EXAMPLE 2.3.33. In $(\mathbb{R}, |\cdot|)$, let $a < b < c$. Then $[a, b]$ is perfect, but $[a, b] \bigcup \{c\}$ is closed and not perfect.

Finally, we address a subtle point about openness that is one reason that we do not use open/closed sets as the fundamental tool in our investigations.

EXAMPLE 2.3.34. Consider $(0, 1) \subset (\mathbb{R}, |\cdot|)$, which is open. $(0, 1) \times \{0\}$ is not open in $(\mathbb{R}^2, \|\cdot\|)$. Note that $(0, 1) \times \{1\} \subset \mathbb{R}^1 \subset \mathbb{R}^2$ and the metric induced by $|\cdot|$ in $\mathbb{R}^1$ is the same as restricting the metric induced by $\|\cdot\|$ in $\mathbb{R}^2$ to $\mathbb{R}^1$.

DEFINITION 2.3.35. Suppose $(\mathbf{X}, d)$ is a metric space and $\mathbf{Y} \subset \mathbf{X}$ is a subset. Then $(\mathbf{Y}, d)$ is also a metric space which we say is **induced** by the metric on $\mathbf{X}$.

Now, returning to Ex. 2.3.34:

DEFINITION 2.3.36. Suppose $(\mathbf{X}, d)$ is a metric space and $A \subset \mathbf{Y} \subset \mathbf{X}$. A is **open relative to Y** if to each $x \in A$ there is an r such that $y \in A$ when $d(x, y) < r$ and $y \in \mathbf{Y}$. In other words, $N_r(x) \bigcap \mathbf{Y} \subset A$.

EXAMPLE 2.3.37. In Ex. 2.3.34, $(0, 1)$ is open relative to $\mathbb{R}^1 \subset \mathbb{R}^2$, but is not open in $\mathbb{R}^2$.

In the situation in Def. 2.3.36, it is natural to ask for the relations between being open relative to $\mathbf{Y}$ and openness in $\mathbf{X}$.

THEOREM 2.3.38. *Let $(\mathbf{X}, d)$ be a metric space and $\mathbf{Y} \subset \mathbf{X}$. $A \subset \mathbf{Y}$ is open relative to $\mathbf{Y}$ if and only if $A = \mathbf{Y} \bigcap G$ for some open set $G \subset \mathbf{X}$.*

PROOF. If A is open relative to $\mathbf{Y}$, then to each $x \in A$ there is an $r_x > 0$ such that $d(x, y) < r_x$ and $y \in \mathbf{Y} \Rightarrow y \in A$.

Let $B_x = \{y \in \mathbf{X} \mid d(y, x) < r_x\}$ and set $G = \bigcup_{x \in A} B_x$. Since B_x is open for every x, G is open. Moreover, $x \in B_x$ for all $x \in A$, so $A \subset G \bigcap \mathbf{Y}$. Also, $B_x \bigcap \mathbf{Y} \subset A$ for all $x \in A$, so $G \bigcap \mathbf{Y} \subset A$.

Now, if G is open in $\mathbf{X}$ and $A = G \bigcap \mathbf{Y}$, then every $x \in A$ has a neighborhood $B_x \subset G$. Then $B_x \bigcap \mathbf{Y} \subset A$, so A is open relative to $\mathbf{Y}$. $\square$

CHAPTER 3

Compactness

One frequent goal of analysis is to generalize a property of some function that holds at each point, or small part of a set, so that it holds *uniformly* over the entire set at once. We might say that we are trying to draw a global conclusion from local information. It turns out that the properties of the underlying set are critical. This might seem a little abstract, but in fact you are very familiar with one example covered in calculus.

3.1. Compactness Arguments and Uniform Boundedness

DEFINITION 3.1.1. A function $f : \mathbb{R} \to \mathbb{R}$ is **bounded** or **locally bounded** at each point in a set $A \subset \mathbb{R}$ if for each $x \in A$ there are constants δ_x, M_x such that $|f(y)| \leq M_x$ for $x - \delta_x \leq y \leq x + \delta_x$.

In this definition, the subscript "x" on δ_x and M_x is usually left off. We include it to emphasize that each x may require different values of δ_x and M_x. Contrast this with

DEFINITION 3.1.2. A function $f : \mathbb{R} \to \mathbb{R}$ is **uniformly continuous** on a set $A \subset \mathbb{R}$ if there is a constant M such that $|f(y)| \leq M$ for all $y \in A$.

In this case, M may depend on the set A, but does not vary with each choice of point in A.

A basic problem that arises is:

> Given a function f that is locally bounded on a set A, can we conclude that f is uniformly bounded on A?

The answer depends on properties of A.

EXAMPLE 3.1.3. $f(x) = \frac{1}{x}$ is locally bounded on $(0, 1)$. For $x \in (0, 1)$, set $\delta_x = \frac{x}{2}$ and $M = \frac{2}{x}$. But, $\frac{1}{x}$ is not uniformly bounded on $(0, 1)$.

What goes wrong? $(0, 1)$ has the limit point 0 (where $\frac{1}{x}$ is undefined), but $0 \notin (0, 1)$ and we have no assumption of $\frac{1}{x}$ being bounded at 0. We could avoid this by assuming the set is closed.

EXAMPLE 3.1.4. $\frac{1}{x}$ is locally bounded on $[\epsilon, 1]$ for any $1 > \epsilon > 0$ and is also uniformly bounded on $[\epsilon, 1]$ for $1 > \epsilon > 0$. Choose $M = \frac{1}{\epsilon}$.

EXAMPLE 3.1.5. $f(x) = x$ is locally bounded on $[0, \infty)$ but is not uniformly bounded on $[0, \infty)$. For local boundedness, take $\delta_x = 1$ and $M_x = x +_1$.

What goes wrong? $[0, \infty)$ is closed and contains its limit points, but it is too "big", allowing x to grow without bound. We can avoid this by assuming the set itself is bounded.

EXAMPLE 3.1.6. $f(x) = x$ is uniformly bounded on $[0, N]$ for any $N > 0$. Choose $M = N$.

This suggests the following theorem:

THEOREM 3.1.7. *Suppose* $f : \mathbb{R} \to \mathbb{R}$ *is locally bounded on a set* $A \subset \mathbb{R}$. *If* A *is closed and bounded, then* f *is uniformly bounded on* A.

We present several proofs of this theorem, all of which are called "compactness arguments". Each argument rests on formulating a property of A ("compactness") that follows from the assumption of A being closed and bounded. These properties generalize to abstract metric spaces, and this will be our lead into compactness.

I want to emphasize that assuming a set of *real numbers* is closed and bounded is somewhat natural, but this may or may not be natural in an abstract metric space. Nevertheless, the properties that derive from being closed and bounded are natural.

The first property is:

DEFINITION 3.1.8. A set $A \subset \mathbb{R}$ has the **Bolzano-Weierstrass property** if every sequence of points in A has a subsequence that converges to a point in A.

EXAMPLE 3.1.9. Suppose $A \subset \mathbb{R}$ is finite. Then A has the Bolzano-Weierstrass property. This is because $\{a_n\} \in A$ implies there exists $x \in A$ such that $B := \{i \mid a_i = x\}$ is infinite. Let $B = \{b_1, b_2, ...\}$ with $i > j \Rightarrow b_i > b_j$, and we see $(a_{b_1}, a_{b_2}, ...)$ is a subsequence of $\{a_n\}$ that converges to x.

EXAMPLE 3.1.10. $\mathbb{N}$ does *not* have the Bolzano-Weierstrass property. Take the sequence $\{a_n\}_{n=1}^{\infty}$ where $a_n = n$.

EXERCISE 3.1.11. Does $\{\frac{1}{a} \mid a \in \mathbb{Z}, a \neq 0\}$ have the Bolzano-Weierstrass property? Justify your answer.

The first proof of Theorem ‘3.1.7 is based on the following theorem:

THEOREM 3.1.12. *A set* $A \subset \mathbb{R}$ *is closed and bounded if and only if it has the Bolzano-Weierstrass property.*

Now, we are not going to prove this property here. It will follow from our more general results later. What we will do is show how to use this to prove Theorem 3.1.7.

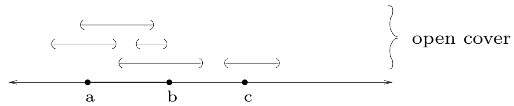
PROOF. Suppose f is not uniformly bounded. So, for any $n \in \mathbb{N}$, there exists $x \in A$ such that $|f(x)| > n$. Let $\{x_n\}$ be a sequence such that

$$|f(x_n)| > n$$

for all n. By Theorem 3.1.12, there must be a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ that converges to some $y \in A$. Choose $N > y$. There exists N_0 such that $m > N_0$ implies $|x_{n_m} - y| < N - y$. But if $n_l > \max\{N, N_0\}$, then we have $x_{n_l} < N$ and $x_{n_l} > N$. This is contradiction. □

The second property uses the notion of a covering:

DEFINITION 3.1.13. Let $A \subset \mathbb{R}$ and $\mathcal{O}$ be a set of open intervals in $\mathbb{R}$. If for every $x \in A$ there is at least one interval $I \in \mathcal{O}$ such that $x \in I$, then $\mathcal{O}$ is an **open cover** of A.



EXAMPLE 3.1.14. $A = (a, b) \cup \{c\}$

EXAMPLE 3.1.15. Since $\mathbb{Q}$ is countable, we can write $\mathbb{Q} = \{a_1, a_2, a_3, ...\}$. Let $\mathcal{O} = \{(a_n - \epsilon, a_n + \epsilon)\}$. Then $\mathcal{O}$ is an open cover of $\mathbb{Q}$ for any $\epsilon > 0$.

DEFINITION 3.1.16. A set $A \subset \mathbb{R}$ has the **Heine-Borel property** if *every* open cover of A can be reduced to an open cover with a finite number of sets, i.e., if $\mathcal{O}$ is an open cover of A, then there are intervals $I_1, ..., I_n \in \mathcal{O}$ such that $A \subset I_1 \bigcup ... \bigcup I_n$.

EXAMPLE 3.1.17. Every finite set has the Heine-Borel Property.

EXAMPLE 3.1.18. $\mathbb{N}$ does not have the Heine-Borel property. Take the cover $\{(n - \epsilon, n + \epsilon) \mid n \in \mathbb{N}\}$ for some small $\epsilon > 0$.

EXAMPLE 3.1.19. $A = \{\frac{1}{n} \mid n \in \mathbb{N}\}$ does not have the Heine-Borel property.

EXERCISE 3.1.20. Prove the statement made in Example 3.1.19.

Now we will observe one more tool that we need in order to explore the second proof of Theorem 3.1.7.

THEOREM 3.1.21. *A set $A \subset \mathbb{R}$ has the Heine-Borel property if and only if A is closed and bounded*

Again this follows from our more general result later. We just use this result right now.

PROOF. Of Theorem 3.1.7.

Since f is locally bounded on A, for each $x \in A$, there is an open interval $I_x \ni x$ and a number M_x such that

$$|f(y)| \leq M_x$$

for $y \in I_x \bigcap A$. Let $\mathcal{O} = \{I_x \mid x \in A\}$.

$\mathcal{O}$ is an open cover of A, so there are $I_{x_1}, I_{x_2}, ..., I_{x_3}$ such that

$$A \subset I_{x_1} \bigcup ... \bigcup I_{x_n}.$$

Let $M = \max\{M_{x_1}, ..., M_{x_n}\}$ *(Why is having a finite number important?)*. Choose $x \in A$, so $x \in I_{x_i}$ for some $1 \leq i \leq n$. Hence,

$$|f(x)| \leq M_{x_i} \leq M$$

□

The next property uses the idea of a descending sequence of sets.

DEFINITION 3.1.22. A sequence of sets $\{A_n\}_{n=1}^{\infty}$ of $\mathbb{R}$ is **descending** if

$$A_1 \supset A_2 \supset A_3 \supset ....$$

Cantor asked: Under what conditions do we have $\bigcap_{n=1}^{\infty} A_n \neq \emptyset$?

DEFINITION 3.1.23. A descending sequence of sets $\{A_n\}_{n=1}^{\infty}$ in $\mathbb{R}$ has the **Cantor Intersection property** if $\bigcap_{n=1}^{\infty} A_n \neq \emptyset$.

EXAMPLE 3.1.24. Let $A_n = (4 - \frac{1}{n}, 5 + \frac{1}{n})$ for $n \in \mathbb{N}$. Then $\bigcap_{n=1}^{\infty} A_n = [4, 5]$.

EXAMPLE 3.1.25. Let $A_n = [-\frac{1}{n}, \frac{1}{n}]$ for some $a \in \mathbb{R}$. Then $\bigcap_{n=1}^{\infty} A_n = \{0\}$.

EXAMPLE 3.1.26. Let $A_n = (0, \frac{1}{n}) \subset \mathbb{R}$. Then $\{A_n\}$ is descending, but $\bigcap_{n=1}^{\infty} A_n = \emptyset$. To see this, notice $0 \notin \bigcap_{n=1}^{\infty} A_n$. Also, if $x \neq 0$, then there exists $m \in \mathbb{N}$ such that $\frac{1}{m} > |x|$. $x \notin A_m$, so $x \notin \bigcap_{n=1}^{\infty} A_n$. So, $\{A_n\}_{n=1}^{\infty}$ does not have the Cantor Intersection property.

EXAMPLE 3.1.27. Let $A_n = [n, \infty) \subset \mathbb{R}$ for $n \in \mathbb{N}$. Then $\{A_n\}$ is descending but $\bigcap_{n=1}^{\infty} A_n = \emptyset$.

In these two examples, the sets in the sequence are not open and bounded respectively. Cantor proved

THEOREM 3.1.28. *Let $\{A_n\}$ be a descending sequence of nonempty, closed, and bounded subsets of $\mathbb{R}$. Then*

$$\bigcap_{n=1}^{\infty} A_n \neq \emptyset.$$

Again, we present a more general result later. For now, we use this property for our third proof of Theorem 3.1.7.

PROOF. Suppose f is not uniformly bounded on A. Since A is bounded, $A \subset [a, b]$ for some a, b. Divide $[a, b]$ into two sub-intervals of length $\frac{(b-a)}{2}$. f must be unbounded on at least one of these sub-intervals. Call this sub-interval $[a_1, b_1]$. Now repeat the division argument to get a new subinterval $[a_2, b_2]$ of $[a_1, b_1]$ of length $\frac{(b-a)}{2^2}$ such that f is unbounded on $[a_2, b_2] \bigcap A$. Inductively, we obtain a sequence $\{[a_n, b_n]\}$ with $(b_n - a_n) = \frac{b-a}{2^n}$, $[a_n, b_n] \subset [a_{n-1}, b_{n-1}]$ (so the sequence is descending), and f is unbounded on $[a_n, b_n] \bigcap A$.

By Theorem 3.1.28, there is a point $x \in \bigcap_{n=1}^{\infty} [a_n, b_n]$. x is a limit point of A (why?) and so $x \in A$. By local boundedness, there are δ_x and M_x such that $|f(y)| \leq M_x$ for $y \in (x - \delta_x, x + \delta_x) \bigcap A$. But, for n sufficiently large, $[a_n, b_n] \subset (x - \delta_x, x + \delta_x)$, which gives a contractiction. □

The last property does not really give a new compactness argument. It is a restatement of the Bolzano-Weierstrass property in terms of Cauchy sequences. (Recall Definition 1.1.11.)

We will base our fourth proof of Theorem 3.1.7 on the following theorem:

THEOREM 3.1.29. *A set $A \subset \mathbb{R}$ is closed and bounded if and only if*

(1) *every sequence in A contains a Cauchy subsequence and*
(2) *every sequence in A that is a Cauchy sequence converges to a limit in A.*

Both of these properties are needed, as we will see later. The proof is only a slight modification of the first proof based on Theorem 3.1.12.

PROOF. Of Theorem 3.1.7

If f is not bounded on A, there is a sequence $\{x_n\} \in A$ such that $|f(x_n)| > n$. There must be a Cauchy subsequence $\{x_{n_k}\}$ that converges to $y \in A$. There are δ_y and M_y such that $|f(z)| \leq M_y$ for $z \in (y - \delta_y, y + \delta_y) \bigcap A$. But, for all k large, $x_{n_k} \in (y - \delta_y, y + \delta_y)$ which yields a contradiction. □

3.2. Sequential Compactness

The "compactness arguments" presented in section 3.1 turn out to be powerful tools in abstract metric spaces. But, the necessary properties of a set needed to use these arguments do not follow from the assumption of being closed and bounded, as they do in $\mathbb{R}^n$ (we will see this). Closed and bounded sets in $\mathbb{R}^n$ have special properties because of the underlying properties of $\mathbb{R}^n$.

So, our strategy is to assume the equivalent conditions for being closed and bounded presented in Theorems 3.1.12, 3.1.21, and 3.1.28, i.e., the analogs of the Bolzano-Weierstrass, the Heine-Borel, and the Cantor intersection properties, in our general abstract metric spaces. After all, it is these conditions that are used in our compactness arguments.

Nominally, these three conditions define three different characteristic properties of the set in question. A major result we prove is that these are equivalent and define just one characteristic property called compactness.

The first type of compactness we define is the analog of the Bolzano-Weierstrass property. We define this and then explore some of the consequences that follow from having the property.

DEFINITION 3.2.1. A subset K of a metric space $(\mathbf{X}, d)$ is **sequentially compact** if every sequence of points in K has a subsequence that converges to a point in K. If $\mathbf{X}$ itself is sequetially compact, we call it a **sequentially compact space**.

It is common to use K for a compact set. Note: since (K, d) is also a metric space, we can think of a sequentially compact subset as a sequentially compact space.

EXAMPLE 3.2.2. A closed, bounded interval in $\mathbb{R}$ is sequentially compact. (See Theorem 3.1.12.)

EXAMPLE 3.2.3. The set of rational numbers in $[0, 1]$ is not a sequentially compact subset of $(\mathbb{R}, |\cdot|)$. Consider the equence obtained by taking a finite number of terms of the decimal expansion of $\frac{1}{\sqrt{2}}$:

$$(.7, .70, .707, .7071, .70710, .707106, ...),$$

which converges to an irrational number. Any subsequence also converges to $\frac{1}{\sqrt{2}}$.

The first consequence of this definition that we prove is

THEOREM 3.2.4. *Let $K \subset \mathbf{X}$ be a (sequentially) compact subset of a metric space $(\mathbf{X}, d)$. Then K is closed and bounded.*

PROOF. First we prove closed. Suppose x were a limit point of K. Then there exists a sequence of elements in K that converge to x. This sequence must have a subsequence that converges to an element in K. However, the sequence and the subsequence must have the same limit, so x must be in K.

Second, we prove bounded. Assume K is not bounded. Choose $x_0 \in K$. For each $n \in \mathbb{N}$, let $x_n \in K$ with $d(x_0, x_n) > n$. Now, the sequence $\{x_n\}$ has the property that every subsequence is unbounded, and hence cannot converge. So, K cannot be sequentially compact. □

> Note: being closed and bounded is not sufficient to guarantee sequential compactness.

EXAMPLE 3.2.5. For $x, y \in \mathbb{R}$, define

$$d(x,y) = \min\{|x-y|, 1\}.$$

$(\mathbb{R}, d)$ is a metric space and $\{x_n\}$ converges to x in $(\mathbb{R}, |\cdot|)$ if and only if it converges to x in $(\mathbb{R}, d)$.

EXERCISE 3.2.6. Show this last statement.

Now, $(\mathbb{R}, d)$ is closed and bounded. (Boundedness follows since $d(x, o) \leq 1$ for all $x \in \mathbb{R}$.) However, $(\mathbb{R}, d)$ is not sequentially compact. For example, $\{1, 2, 3, ...\}$ contains no convergent subsequence.

3.3. Separability

We saw that sequential compactness imples being closed and bounded, and the closed and bounded sets in $\mathbb{R}$ are special with respect to making compactness arguments. Another property of real numbers is that the rational numbers are dense. This turns out to be even more fundamentally important than is obvious. It also turns out that sequential compactness is related to this property.

DEFINITION 3.3.1. A metric space $(\mathbf{X}, d)$ is **separable** if it contains a countable, dense subset. A subset $A \subset \mathbf{X}$ of a metric space $(\mathbf{X}, d)$ is **separable** if it is separable considered as a metric space with the metric d.

EXAMPLE 3.3.2. $(\mathbb{R}, |\cdot|)$ and $(\mathbb{R}^n, \|\cdot\|)$ are separable. So are closed, bounded subsets of these metric spaces.

THEOREM 3.3.3. *Let $K \subset \mathbf{X}$ be a sequentially compact subset of a metric space $(\mathbf{X}, d)$. Then K is separable.*

We prove this by introducing a useful notion and proving another theorem.

DEFINITION 3.3.4. Let $A \subset \mathbf{X}$, wehre $(\mathbf{X}, d)$ is a metric space. If for $\epsilon > 0$ there are points $x_1, ..., x_n \in A$ such that

$$A \subset \bigcup_{i=1}^{n} N_\epsilon(x_i),$$

then $\{x_1, ..., x_n\}$ is an **ϵ-net** for A.

> Note: if $\{x_1, ..., x_n\}$ is an ϵ-net for A, then for any $x \in A$, $d(x, x_i) < \epsilon$ for at least one i.

EXAMPLE 3.3.5. Given any set of real numbers that is bounded and any $\epsilon > 0$, we can find an ϵ-net consisting of points with finite decimal expansions. This is very important for computation with real numbers on computers.

Note that the existence of an ϵ-net for any $\epsilon > 0$ is a stronger condition than boundedness. Not only is K contained in some large ball, which is obtained by choosing some ϵ-net $\{x_1, ..., x_n\}$ and taking a ball that contains $\bigcup_i N_\epsilon(x_i)$, but it is contained in the union of a finite number of neighborhoods of any small size.

EXAMPLE 3.3.6. Consider the closed unit "ball" in l_2 (see Definition 2.2.14.):

$$\overline{N_1(0)} = \{x \in l_2 \mid \sqrt{\sum_{n=1}^{\infty} x_n^2} \leq 1\}.$$

The sequence $(x_1, x_2, x_3, ...)$ with

$$\begin{aligned} x_1 &= (1, 0, 0, 0, ...) \\ x_2 &= (0, 1, 0, 0, ...) \\ x_3 &= (0, 0, 1, 0, ...) \\ &\vdots \end{aligned}$$

is clearly in $\overline{N_1(0)}$. However,

$$d(x_m, x_n) = \sqrt{2}$$

for any $m \neq n$, hence $N_{\frac{1}{2}}(x)$ can contain at most one of the $\{x_n\}$ for any $x \in l_2$. If $x_m \in N_{\frac{1}{2}(x)}$ for some m, i.e.,

$$d(x_m, x) < \frac{1}{2},$$

then for $m \neq n$, we have

$$\begin{aligned} \sqrt{2} = d(x_n, x_m) &\leq d(x_n, x) + d(x_m, x) \\ &\leq d(x_n, x) + \frac{1}{2}. \end{aligned}$$

This yields

$$d(x_n, x) \geq \sqrt{2} - \frac{1}{2} > \frac{1}{2}.$$

Thus, there can be no ϵ-net for $\overline{N_1(0)}$ in l_2 with $\epsilon = \frac{1}{2}$.

EXAMPLE 3.3.7. $(\mathbb{R}, |\cdot|)$ does not have an ϵ-net for any $\epsilon > 0$.

DEFINITION 3.3.8. Let $(\mathbf{X}, d)$ be a metric space. A set $A \subset \mathbf{X}$ is **totally bounded** if for every $\epsilon > 0$, there is an ϵ-net for A.

Theorem 3.3.3 follows from

THEOREM 3.3.9. *Let $K \subset \mathbf{X}$ be a sequentially compact subset of a metric space $(\mathbf{X}, d)$. Then K is totally bounded.*

PROOF. Of Theorem 3.3.9 Assume for some $\epsilon > 0$ there does not exist any finite set of points $\{x_1, ..., x_n\}$ with

$$K \subset \bigcup_{i=1}^{n} N_\epsilon(x_i).$$

We can choose a sequence of points $\{x_m\}_{m=1}^{\infty}$ in K such that $x_n \notin \bigcup_{m=1}^{n-1} N_\epsilon(x_m)$ for all n. This sequence can have no convergent subsequence. This is because if $\{x_{m_k}\}_{k=1}^{\infty} \to x$, then there exists l such that $k > l$ implies $d(x, x_{m_k}) < \frac{\epsilon}{2}$. But now $x_{l+1} \in N_\epsilon(x_l)$, contradicting our choice of x_{l+1}. This proves the contrapositive. □

EXERCISE 3.3.10. This proof is stated a little more easily if we use the notion of Cauchy sequences and the fact that a sequence that converges is Cauchy. Do this.

PROOF. Of Theorem 3.3.3. By Theorem 3.3.9, for $\epsilon = \frac{1}{m}, m \in \mathbb{N}$, there is an ϵ-net for K. Call it

$$\{x_{m,1}, x_{m,2}, ..., x_{m,n_m}\}.$$

So,

$$K \subset \bigcup_{i=1}^{n_m} N_{\frac{1}{m}}(x_{m,i}).$$

The set of points

$$A = \{x_{1,1}, ..., x_{1,n_1}, x_{2,1}, ..., x_{2,n_2}, x_{3,1}, ..., x_{3,n_3}, x_{4,1}, ...\}$$

is at most countable and dense in K.

A is at most countable because it is the countable union of finite sets. (The sets may intersect, so the end result could be finite.)

We need a little more justification to say A is dense in K. We have to show that every point in $K - A$ is a limit point of A. Assume $x \in K, x \notin A$. (Note, if K is finite, then $x \in A$ is forced.) We construct a sequence of points in A that converges to x.

For $m \in \mathbb{N}$, choose $x_m = x_{m,i}$ such that $1 \leq i \leq n_m$ and $x \in N_{\frac{1}{m}}(x_{m,i})$. Then we see that $\{x_i\}_{i=1}^{\infty} \in A$ and $\{x_i\}_{i=1}^{\infty} \to x$. □

Summing up, so far we have that sequential compactness implies closed, bounded, totally bounded, and separable.

There is one last fact about separability we will use. Separability is related to open covers.

DEFINITION 3.3.11. An **open cover** of a set A contained in a metric space $(\mathbf{X}, d)$ is a collection of open subsets $\{G_\alpha\}_{\alpha \in a}$ of $\mathbf{X}$ such that

$$A \subset \bigcup_{\alpha \in a} G_\alpha.$$

A **sub-cover** of $\{G_\alpha\}_{\alpha \in a}$ is a sub-collection $\{G_\alpha\}_{\alpha \in a' \subset a}$ that still covers A. A **countable** or **finite (sub)cover** has a countable or finite number of sets respectively.

THEOREM 3.3.12. *$Lindel\ddot{o}f'sTheorem$ Every open cover of a separable metric space has a countable or finite subcover.*

PROOF. Let $(\mathbf{X}, d)$ be a separable metric space and let $\{x_n\}_{n \in \mathbb{N}}$ be a dense set of points in $\mathbf{X}$. The set of neighborhoods

$$\{N_{\frac{1}{m}}(x_n)\}_{n,m \in \mathbb{N}}$$

is an at most countable collection of open sets. We number these sets by $\{N_1, N_2, N_3, ...\}$.

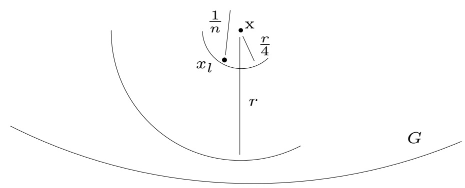


FIGURE 3.1

We first show that if $G \subset \mathbf{X}$ is an open set that contains $x \in \mathbf{X}$, then

$$x \in N_{n'} \subset G$$

for some n'. Since G is open, $N_r(x) \subset G$ for some $2 > r > 0$. We choose x_l from the sequence $\{x_m\}$ so that

$$|x_l - x| < \frac{r}{4}.$$

We choose $n \in \mathbb{N}$ with

$$\frac{4}{r} > n > \frac{2}{r} \text{ or } \frac{r}{4} < \frac{1}{n} < \frac{r}{2}.$$

(See figure 3.1). The triangle inequality implies

$$x \in N_{\frac{1}{n}}(x_l) \subset N_r(x).$$

So, we have $N_{n'} = N_{\frac{1}{n}}(x_l)$.

Let $\{G_\alpha\}_{\alpha \in a}$ be an open cover of $\mathbf{X}$. We use $\{N_i\}$ to extract a countable, or finite, subcover. We select a sequence $\{G_{\alpha_1}, G_{\alpha_2}, G_{\alpha_3}\}$ from $\{G_\alpha, ...\}$ by choosing, when possible, any G_{α_k} such that $N_k \subset G_{\alpha_k}$. We skip values of k when necessary to get an index set K. $\{G_{\alpha_k}\}_{\alpha \in K}$ is at most countable. (Note, there may be "gaps" in the indices $k \in K$. For example,

$$\alpha_1, \alpha_2, \alpha_{10}, \alpha_{11}, \alpha_{153}, ...$$

could be the indices. But, there is at most a countable number.

Any $x \in \mathbf{X}$ is contained in some $G_{\alpha'} \in \{G_\alpha\}$. But, then $x \in N_k \subset G_{\alpha'}$ for some k, and hence there is a $G_{\alpha_{k'}} \in \{G_{\alpha_k}\}_{k \in K}$ such that $x \in G_{\alpha_{k'}}$. Hence, $\{G_{\alpha_k}\}_{k \in K}$ covers $\mathbf{X}$. □

3.4. Notions of Compactness

We now generalize the other two conditions equivalent to being closed and bouned for intervals in $\mathbb{R}$ in Theorems 3.1.21 and 3.1.28, the Heine-Borel property and Cantor's intersection property.

DEFINITION 3.4.1. (Compare to Definition 3.1.16).

A subset $K \subset \mathbf{X}$ of a metric space $(\mathbf{X}, d)$ is **compact** if every open cover of K contains a finite subcover. If a metric space $\mathbf{X}$ is compact, then we call it a **compact metric space**.

DEFINITION 3.4.2. A collection $\{F_\alpha\}_{\alpha \in a}$ of closed subsets of a metric space $(\mathbf{X}, d)$ has the **finite intersection property** if every finite subcollection of the closed subsets has a nonempty intersection.

EXAMPLE 3.4.3. Define $I_\alpha = [0, \alpha]$ for $\alpha > 0$ and $I = [3, 4]$. Then $\{I_\alpha\}_{\alpha > 0}$ has the finite intersection property, but $\{\{I_\alpha\}_{\alpha > 0} \bigcup I\}$ does not.

We now show part of the Borel-Lebesgue Theorem. The rest will be proven in Chapter 4.

THEOREM 3.4.4. **Borel-Lebesgue Theorem, part I**
Let $(\mathbf{X}, d)$ be a metric space, and $K \subset \mathbf{X}$. The following are equivalent.

(1) *K is compact.*
(2) *Every collection of closed subsets of K with the finite intersection property has a nonempty intersection.*
(3) *K is sequentially compact.*

In light of this result, we only use the term "compact" and no longer refer to "sequential compactness". We may restate the results in Section 3.2 and 3.3:

THEOREM 3.4.5. *Let $K \subset \mathbf{X}$ be a compact subset of a metric space $(\mathbf{X}, d)$. Then,*

(1) *K is closed.*
(2) *K is bounded.*
(3) *K is separable.*
(4) *K is totally bounded.*

PROOF. Of Theorem 3.4.4
We show that (1) → (2) → (3) → (1).
(1) → (2)
We assume K is compact. Let $\{F_\alpha\}_{\alpha \in a}$ be a collection of closed subsets of K such that $\bigcap_{\alpha \in a} F_\alpha = \emptyset$. Consider the open sets $\{G_\alpha\}_{\alpha \in a}$ with $G_\alpha = F_\alpha^c$, the complement of F_α in $\mathbf{X}$. Then,

$$\bigcup_\alpha G_\alpha = \bigcup_\alpha F_\alpha^c = (\bigcap_\alpha F_\alpha)^c = \emptyset^c = \mathbf{X}.$$

So,

$$K \subset \bigcup_\alpha G_\alpha.$$

By compactness,

$$K \subset G_{\alpha_1} \bigcup ... \bigcup G_{\alpha_n}$$

for some $\alpha_1, ..., \alpha_n$. Moreover, since $F_{\alpha_i} \subset K$,

$$F_{\alpha_i} = K \setminus G_{\alpha_i}$$

so

$$\begin{aligned} F_{\alpha_1} \bigcap ... \bigcap F_{\alpha_n} &= K \setminus G_{\alpha_1} \bigcap ... \bigcap K \setminus G_{\alpha_n} \\ &= K \setminus (G_{\alpha_1} \bigcup ... \bigcup G_{\alpha_n}) = K \setminus K = \emptyset. \end{aligned}$$

Hence, $\{F_\alpha\}_{\alpha \in a}$ cannot have the finite intersection property.

EXAMPLE 3.4.6. Consider $[0, 1] \subset (\mathbb{R}, |\cdot|)$. Define

$$F_n = [\frac{1}{n+3}, \frac{1}{n+1}], n \in \mathbb{N},$$

so $\{F_n\} = \{[\frac{1}{4}, \frac{1}{2}], [\frac{1}{5}, \frac{1}{3}], [\frac{1}{6}, \frac{1}{4}], [\frac{1}{7}, \frac{1}{5}], ...\}$ (see Figure 3.2).
Now, if $G_n = F_n^c$, then

$$G_n = (-\infty, \frac{1}{n+3}) \bigcup (\frac{1}{n+1}, \infty)$$

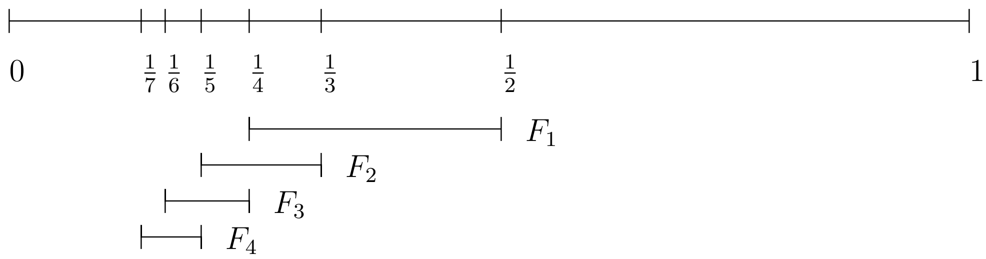


FIGURE 3.2

or

$$\begin{aligned} G_1 &= (-\infty, \frac{1}{4}) \bigcup (\frac{1}{2}, \infty) \\ G_2 &= (-\infty, \frac{1}{5}) \bigcup (\frac{1}{3}, \infty) \\ G_3 &= (-\infty, \frac{1}{6}) \bigcup (\frac{1}{4}, \infty) \\ G_4 &= (-\infty, \frac{1}{7}) \bigcup (\frac{1}{5}, \infty) \end{aligned}$$

Clearly, $\bigcap_{n=1}^{\infty} F_n = \emptyset$, since $F_4 \bigcap F_1 = \emptyset$.

Moreover, plotting $\{G_n\}$ shows that

$$[0,1] \subset G_1 \bigcup G_2 \bigcup G_3 \bigcup G_4.$$

$\{G_1, G_2, G_3, G_4\}$ is the finite subcover constructed in the proof (see Figure 3.3).

Back to the proof: <u>(2) → (3)</u>

Let $\{x_m\}$ be a sequence in K. Define

$$F_n = \overline{\{x_n, x_{n+1}, ...\}} = \overline{\{x_m\}_{m=n}^{\infty}}$$

This is a descending sequence of closed sets with the finite intersection property, since

$$x_{n'} \in F_{n_1} \bigcap ... \bigcap F_{n_k}$$

when $n' = \max\{n_1, ..., n_k\}$ for any $n_1, ..., n_k \in \mathbb{N}$.

Hence, there is an $x \in K$ with $x \in \bigcap_{n=1}^{\infty} F_n$. There are two possibilities. If $x = x_n$ for infinitely many n, the we extract the subsequence consisting of repeated valued of x_n, which obviously converges to $x = x_n$.

If $x = x_n$ for finitely many n, then we construct a subsequence that converges to x as follows: Choose n_1 large enough that $x \neq x_m$ for $m \geq n_1$. Given $n_k \in \mathbb{N}$, us the fact that

$$x \in \overline{\{x_{n_k+1}, x_{n_k+2}, x_{n_k+3}, ...\}}$$

is a limit point to choose $n_{k+1} > n_k$ with $0 < d(x, x_{n_k+1}) < \frac{1}{k+1}$. Clearly, $\{x_{n_k}\} \to x$.

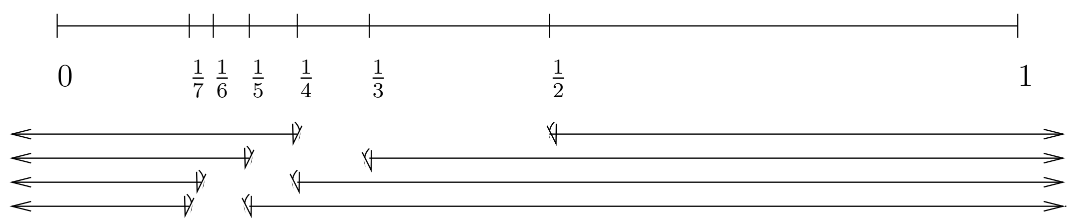


FIGURE 3.3

EXAMPLE 3.4.7. First, consider $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ...)$.

$$\begin{aligned} F_1 &= \{1, \tfrac{1}{2}, \tfrac{1}{3}, ...\} \bigcup \{0\} \\ F_2 &= \{\tfrac{1}{2}, \tfrac{1}{3}, ...\} \bigcup \{0\} \\ F_3 &= \{\tfrac{1}{3}, \tfrac{1}{4}, ...\} \bigcup \{0\} \\ &\vdots \\ \bigcap F_n &= \{0\} \end{aligned}$$

In the proof, $x = 0$ and the sequence itself converges to x.

Next, consider $\{1, \frac{1}{2}, 1, \frac{1}{3}, 1, \frac{1}{4}, 1, \frac{1}{5}, ...\}$.

$$\begin{aligned} F_1 &= \{1, \tfrac{1}{2}, 1, \tfrac{1}{3}, 1, \tfrac{1}{4}, 1, \tfrac{1}{5}, ...\} \bigcup \{0\} \\ F_2 &= \{\tfrac{1}{2}, 1, \tfrac{1}{3}, ...\} \bigcup \{0\} \\ F_3 &= \{1, \tfrac{1}{3}, 1, \tfrac{1}{4}, ...\} \bigcup \{0\} \\ &\vdots \\ \bigcap F_n &= \{1, 0\} \end{aligned}$$

If we choose $x = 1$, then we choose the subsequence of all 1's. If we choose $x = 0$, then we choose the subsequence $\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ...\}$.

Back to the proof: (3) $\rightarrow$ (1).

Assume K is sequentially compact and let $\{G_\alpha\}_{\alpha\in a}$ be an open cover of K. By Theorem 3.3.3, K is separable and by Theorem 3.3.12, $\{G_\alpha\}_{\alpha\in a}$ can be reduced to a finite or countble subcover.

Assume we have a countable cover $\{G_n\}_{n=1}^{\infty}$. We show this can be reduced to a finite subcover. Assume this is not true. For each $n \in \mathbb{N}$, there is an $x_n \in K$ but not in $\bigcup_{m=1}^{n} G_m$. Otherwise, $\{G_m\}_{m=1}^{n}$ covers K. The sequence $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ that converges to a limit $x \in K$. Since $K \subset \bigcup_{m=1}^{\infty} G_m, x \in G_M$ for some $M > 0$. Since G_M is open, $x_{n_k} \in G_M$ for all sufficiently large k. But, this contradicts the construction by which $x_{n_k} \notin G_M$ for $n_k > M$. □

3.5. Some Properties of Compactness

We now show some easy, but characteristic properties of compactness of subsets.

Theorem 3.5.1. *Let* $(\mathbf{X}, d)$ *be a metric space.*

(1) *If* $\{K_1, ..., K_n\}$ *are compact subsets of* $\mathbf{X}$*, then* $\bigcup_{m=1}^{n} K_m$ *is compact.*
(2) *If* $\{K_\alpha\}_{\alpha\in a}$ *is a collection of compact subsets of* $\mathbf{X}$*, then* $\bigcap_{\alpha\in a} K_\alpha$ *is compact.*

Exercise 3.5.2. Prove Theorem 3.5.1.

Recall the "flaw" concerning openness and subsets discussed in Example 2.3.34. If $(\mathbf{X}, d)$ is a metric space and $\mathbf{Y} \subset \mathbf{X}$, then $(\mathbf{Y}, d)$ is a metric space. A set $G \subset \mathbf{Y}$ may be open in $\mathbf{Y}$, but this does not mean it is open in $\mathbf{X}$. Contrast this to

Theorem 3.5.3. *Suppose* $(\mathbf{X}, d)$ *is a metric space and* $K \subset \mathbf{Y} \subset \mathbf{X}$. K *is a compact subset of* $\mathbf{X}$ *if and only if* K *is a compact subset of* $\mathbf{Y}$.

Proof. Suppose $K \subset \mathbf{X}$ is compact and let $\{A_\alpha\}_{\alpha\in a}$ be a collection of sets that are open relative to $\mathbf{Y}$ such that $K \subset \bigcup_{\alpha\in a} A_\alpha$. By Theorem 2.3.38, there are open sets $\{G_\alpha\}_{\alpha\in a}$ in $\mathbf{X}$ with

$$A_\alpha = \mathbf{Y} \bigcap G_\alpha$$

for $\alpha \in a$. Since K is compact in $\mathbf{X}$ and is covered by $\{A_\alpha\}_{\alpha\in a}$,

$$K \subset G_{\alpha_1} \bigcup ... \bigcup G_{\alpha_n}$$

for some $\alpha_1, ..., \alpha_n \in a$. But, this implies

$$K \subset A_{\alpha_1} \bigcup ... \bigcup A_{\alpha_n},$$

and K is compact in $\mathbf{Y}$.

The other direction is simply the reverse of this argument. □

Finally, recall that Example 3.2.5 shows that being closed and bounded is not sufficient to guarantee compactness. Interestingly,

Theorem 3.5.4. *Closed subsets of a compact set in a metric space are compact.*

Proof. Let $(\mathbf{X}, d)$ be a metric space, $K \subset \mathbf{X}$ compact, and $F \subset K$ closed. Let $\{G_\alpha\}_{\alpha\in a}$ be an open cover of F. If F^c is added to $\{G_\alpha\}_{\alpha\in a}$, then we obtain an open cover $\{\{G_\alpha\}_{\alpha\in a} \bigcup F^c\}$ of K (see Figure 3.4).

Since K is compact, there is a finite subcollection from $\{G_\alpha\}_{\alpha\in a} \bigcup \{F\}$ that covers K, and hence covers F. We can remove F^c and still retain a cover of F. Thus, a finite subcollection of $\{G_\alpha\}_{\alpha\in a}$ covers F, and F is compact. □

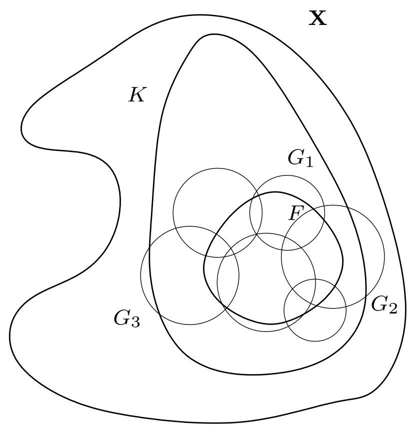


FIGURE 3.4

3.6. Compact Sets in $\mathbb{R}^n$

As a special case, we consider $\mathbb{R}^n$. First, we prove the generalization of any closed and bounded interval is compact.

DEFINITION 3.6.1. Let $a_m > b_m$ be numbers in $\mathbb{R}$ for $m = 1, 2, ..., n$. The set

$$\{x = (x_1, ..., x_n) \in \mathbb{R}^n \mid a_m \leq x_m \leq b_m, 1 \leq m \leq n\}$$

is an **n-cell** in $\mathbb{R}^n$.

THEOREM 3.6.2. *Every n-cell of* $\mathbb{R}^n$ *is compact. This implies in particular that* $[a, b] \subset \mathbb{R}, a < b$, *is compact*

PROOF. We first show a modified form of the finite intersection property for n-cells. We begin with intervals in $\mathbb{R}^1$:

Let $\{I_m\}_{m=1}^{\infty}$ be a descending sequence of closed intervals in $\mathbb{R}^1$, i.e., $I_1 \supset I_2 \supset ....$ This implies $\{I_m\}_{m=1}^{\infty}$ has the finite intersection property. We prove that $\bigcap_{m=1}^{\infty} I_m$ is nonempty.

Let $I_m = [a_m, b_m]$. The sequence $\{a_m\}$ is bounded above by b_1, hence

$$x = \sup_{m \in \mathbb{N}} a_m < \infty.$$

We show that $x \in \bigcap_{m=1}^{\infty} I_m$. For positive integers m, l,

$$a_l \leq a_{l+m} \leq b_{l+m} \leq b_m,$$

so $x \leq b_m$ for all m. Since $a_m \leq x$ for all m, $a_m \leq x \leq b_m$ for all m, and $x \in \bigcap_{m \in \mathbb{N}} I_m$.

Now suppose $\{I_m\}$ is a descending sequence of n-cells in $\mathbb{R}^n$. Let

$$I_m = \{x \mid a_{m,l} \leq x_l \leq b_{m,l}, 1 \leq l \leq n, m \in \mathbb{N}\}$$

Set

$$I_{m,l} = [a_{m,l}, b_{m,l}] \subset \mathbb{R}^1$$

where $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$. So, for $a \leq l \leq n, m \in \mathbb{N}$. In other words,

$$I_m = I_{m,1} \times ... \times I_{m,n}.$$

For each l, $\{I_{m,l}\}_{m\in\mathbb{N}}$ is a descending sequence of intervals in $\mathbb{R}^1$. Hence, there are real numbers x_l such that

$$a_{m,l} \le x_l \le b_{m,l}$$

for $1 \le l \le n, m \in \mathbb{N}$. We have $x = (x_1, x_2, ..., x_n) \in I_m$ for all m.

Now we prove that the n-cell

$$I = \{x \in \mathbb{R}^n \mid a_n \le x_m \le b_m, 1 \le m \le n\}$$

is compact. Set $\delta =\parallel b - a \parallel$, so

$$\parallel x - y \parallel \le \delta$$

for all $x, y \in I$. Suppose there is an open cover $\{G_\alpha\}_{\alpha\in a}$ of I that contains no finite subcover. Set

$$c_m = \frac{a_m + b_m}{2}$$

for $1 \le m \le n$. The intervals $\{[a_m, c_m], [c_m, b_m]\}$ determine 2^n n-cells $\{I_l\}$ whose union is I.

One of these n-cells, at least, cannot be covered by any finite collection from $\{G_\alpha\}_{\alpha\in a}$. Call this J_1. We next subdivide J_1 into 2^n n-cells in the same way. Again, one of the resulting n-cells cannot be covered by any finite collection from $\{G_\alpha\}_{\alpha\in a}$, and we call this J_2.

Inductively, we obtain a sequence of n-cells $\{J_m\}_{m=1}^{\infty}$ with the properties

(1) $J_1 \supset J_2 \supset ...$
(2) J_m is not covered by any finite collection from $\{G_\alpha\}$.
(3) If $x \in J_m$ and $y \in J_m$, then $\parallel x - y \parallel \le 2^{-m}\delta$.

By the discussion above, we know there is a point $x \in \bigcap_{m=1}^{\infty} J_M, x \in G_\alpha$ for some α. Since G_α is open, there is an $r > 0$ with $N_r(x) \subset G_\alpha$. If we choose m with $2^{-m}\delta < r$, then $J_m \subset G_\alpha$. But, this is a contradiction. □

From this, it is easy to prove

Theorem 3.6.3. *A set $K \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Proof. By Theorem 3.4.5, if K is compact, then it is closed and bounded. On the other hand, if K is bounded, then it is contained in some compact n-cell. Since K is a closed subset of a compact set, Theorem 3.5.4 shows it is compact. □

These last two theorems complete the proofs of Theorems 3.1.12(Bolzano-Weierstrass property), 3.1.21(Heine-Borel property), and 3.1.28(Cantor Intersection Property).

CHAPTER 4

Cauchy Sequences in Metric Spaces

We have defined the notions of sequences, subsequences, and convergence, and explored their relation to the essential property of compactness. Recall that a serious flaw in the definition of convergence, from the point of application, is that this definition requires the (possibly unknown) limit. Cauchy sequences are a way to get around this.

4.1. A Few Facts and Boundedness

First, we present a few facts about sequences and discuss the property of boundedness.

THEOREM 4.1.1. *The limit of a convergent sequence in a metric space is unique.*

PROOF. Suppose that $\{x_n\}$ is a sequence in a metric space $(\mathbf{X}, d)$ that converges to x and y. Choose $\epsilon > 0$. There are N and M such that $d(x_n, x) < \epsilon$ for $n \geq N$ and $d(x_n, y) \leq \epsilon$ for $n \geq M$. Hence, for $n \geq \max\{N, M\}, d(x, y) \leq d(x, x_n) + d(x_n, y) \leq 2\epsilon$. Since $d(x, y) \leq 2\epsilon$ for any $\epsilon > 0, d(x, y) = 0$ and so $x = y$. □

THEOREM 4.1.2. *Let $\{x_n\}$ be a sequence in a metric space $(\mathbf{X}, d)$. $\{x_n\}$ converges to $x \in \mathbf{X}$ if and only if every neighborhood of x contains all but finitely many terms of $\{x_n\}$*

PROOF. Suppose $\{x_n\} \to x$ and let $N_\epsilon(x)$ be a neighborhood of x. By convergence, there is an N such that $d(x_n, x) < \epsilon$ for $n \geq N$. Hence, $x_n \in N_\epsilon(x)$ for $n \geq N$.

Now assume every neighborhood of x contains all but a finite number of $\{x_n\}$. For $\epsilon > 0$, consider $N_\epsilon(x)$. By assumption, there is an $N > 0$ such that $x_n \in N_\epsilon(x)$, i.e., $d(x_n, x) < \epsilon$ for $n \geq N$. □

Recalling the notions of subsequences and subsequential limits, we have the following useful fact.

THEOREM 4.1.3. *The subsequential limits of a sequence in a metric space form a closed subset of the space.*

PROOF. Let $\{x_n\}$ be a sequence in a metric space $(\mathbf{X}, d)$. Let A be the set of all subsequential limits of $\{x_n\}$. Let x be a limit point of A. We want to show that $x \in A$, which means showing that x is a subsequential limit of $\{x_n\}$.

Choose n_1 so that $x_{n_1} \neq x$. If no such n_1 exists, then A has one point, and we are done. Let $\delta = d(x, x_1)$. Suppose $n_1, ..., n_{m-1}$ are chosen. Since x is a limit point of A, there is a $y \in A$ with $d(x, y) < 2^{-m}\delta$. Since $y \in A$, there is an $n_m > n_{m-1}$ with $d(y, x_{n_m}) < 2^{-m}\delta$.

We have

$$\begin{aligned} d(x, x_{n_m}) &\leq d(x,y) + d(y, x_{n_m}) \\ &\leq 2^{-m}\delta + 2^{-m}\delta = 2^{1-m}\delta \end{aligned}$$

We conclude that $\{x_{n_m}\} \to x$. □

Finally, we discuss the connection between boundedness (Defn. 2.3.24) and convergence of sequences.

Definition 4.1.4. A sequence $\{x_n\}$ in a metric space $(\mathbf{X}, d)$ is **bounded** if its range forms a bounded set in $\mathbf{X}$. Otherwise, it is **unbounded**. Equivalently, $\{x_n\}$ is bounded if and only if there exists $A \subset \mathbf{X}$,a bounded set, such that $x_n \in A$ for all n.

Example 4.1.5. In $(\mathbb{R}, |\cdot|)$,
(1) $\{\frac{1}{n}\}_{n=1}^{\infty}$ converges and is bounded.
(2) $\{n^2\}_{n=1}^{\infty}$ diverges and is unbounded.
(3) $\{1 + (-1)^n\}_{n=1}^{\infty}$ diverges and is bounded.

We have the relation

Theorem 4.1.6. *A sequence in a metric space that converges is bounded.*

Proof. Suppose $\{x_n\}$ is a sequence in a metric space $(\mathbf{X}, d)$ that converges to x. There is an integer N such that $d(x_n, x) < 1$ for $n \geq N$. Set

$$r = \max\{1, d(x, x_1), \ldots, d(x, x_{n-1})\}.$$

Then, $d(x_n, x) \leq r$ for all n. □

4.2. Cauchy Sequences

The practical trouble with the standard definition of convergence is that it involves the (usually) unknown limit.

Example 4.2.1. The sequence

$$\{ \quad \sum_{m=1}^{n} \frac{4}{n} \sqrt{e^{\frac{4m}{n}} - \sin(\frac{2}{1 + \frac{4m}{n}})} \quad \}_{n=1}^{\infty}$$

converges, because it converges to

$$\int_{1}^{4} \sqrt{e^{x} - \sin(\frac{2}{1+x}})dx,$$

which we can prove exists by standard Calculus results. However, we do not know the value of this integral and *cannot* verify this by the definition of convergence.

The notion of a Cauchy sequence gives a way around this difficulty.

The idea is that if a sequence converges to a limit, i.e., the terms in the sequence approach a limit as the index increases, then the terms also approach each other as the index increases.

Definition 4.2.2. A sequence $\{x_n\}$ in a metric space $(\mathbf{X}, d)$ is a **Cauchy sequence** if for every $\epsilon > 0$ there is an $N > 0$ such that $d(x_n, x_m) < \epsilon$ for $n, m \geq N$.

EXAMPLE 4.2.3. $\{\frac{1}{n}\}$ is a Cauchy sequence in $(\mathbb{R}, |\cdot|)$. This is because $d(\frac{1}{n}, \frac{1}{m}) = |\frac{1}{n} - \frac{1}{m}| \leq \frac{2}{\min\{n,m\}}$.

So, given $\epsilon > 0$, if we choose $N > \frac{2}{\epsilon}$, then for $n, m \geq N$, $|\frac{1}{n} - \frac{1}{m}| < \epsilon$.

EXAMPLE 4.2.4. $\{\frac{\sin(nx)}{n}\}$ is a Cauchy sequence in $\mathcal{C}([0, \pi])$ since

$$|\frac{\sin(nx)}{n} - \frac{\sin(mx)}{m}| \leq \frac{2}{\min\{n, m\}}$$

for $0 \leq x \leq \pi$. Hence, given $\epsilon > 0$, if $N > \frac{1}{\epsilon}$, then

$$d(\frac{\sin(nx)}{n}, \frac{\sin(mx)}{m}) < \epsilon$$

for $n, m \geq N$.

Fitting our intuition,

THEOREM 4.2.5. *Any sequence in a metric space that converges is a Cauchy sequence.*

PROOF. Assume $\{x_m\}$ is a sequence in a metric space $(\mathbf{X}, d)$ and $x_n \to x$. Choose $\epsilon > 0$. There is an N such that $d(x_n, x) < \epsilon$ for $n \geq N$. Hence,

$$d(x_n, x_m) \leq d(x_n, x) + d(x_m, x) \leq 2\epsilon$$

for $n, m \geq \mathbb{N}$, and $\{x_n\}$ is Cauchy. □

Moreover,

THEOREM 4.2.6. *If a subsequence of a Cauchy sequence in a metric space converges to a limit, then the Cauchy sequence itself converges to the same limit.*

PROOF. Let $\{x_n\}$ be a Cauchy sequence in a metric space $(\mathbf{X}, d)$. Suppose the subsequence $\{x_{n_k}\}$ converges to x. Given $\epsilon > 0$, choose N so that $d(x_n, x_m) < \frac{\epsilon}{2}$ for $n, m \geq \mathbb{N}$. Choose K such that $k \geq K$ implies $n_k \geq N$ and $d(d_{n_k}, x) < \frac{\epsilon}{2}$. Then for all $n \geq N$ and $k \geq K$, $d(x_n, x) < d(x_n, x_{n_k}) + d(x_{n_k}, x) < \epsilon$. □

Finally, we observe one more characteristic of Cauchy sequences:

THEOREM 4.2.7. *Cauchy sequences are bounded.*

PROOF. Suppose $\{x_n\}$ is a Cauchy sequence in the metric space $(\mathbf{X}, d)$. Then there exists $N \in \mathbb{N}$ such that $n, m \geq N$ implies $d(x_n, x_m) < 1$. We now have

$$d(x_N, x_k) \leq \max\{d(x_N, x_1), d(x_N, x_2), ..., d(x_N, x_{N-1}), 1\}$$

for all $k \in \mathbb{N}$. □

4.3. Completeness

Unfortunately, the converse to Theorem 4.2.5 just does *not* hold. Not every Cauchy sequence in a metric space must converge to a point in the space.

EXAMPLE 4.3.1. Consider $(0, 1) \subset (\mathbb{R}, |\cdot|)$. $\{\frac{1}{n}\}$ is a Cauchy sequence in $(0, 1)$, but does not converge to a limit in $(0, 1)$.

EXAMPLE 4.3.2. Consider $\mathbb{Q} \subset (\mathbb{R}, |\cdot|)$. $\{(1 + \frac{1}{n})^n\}$ is a Cauchy sequence in $\mathbb{Q}$ because we know that $(1 + \frac{1}{n})^n \to e$ in $(\mathbb{R}, |\cdot|)$, but its limit $e \not\in \mathbb{Q}$.

EXAMPLE 4.3.3. Consider the space of polynomials on $[a,b]$: $\mathcal{P}([a,b]) \subset \mathcal{C}([a,b])$. The sequence $\{1+x+\frac{x^2}{2!}+...+\frac{x^n}{n!}\}$ converges uniformly to e^x on $[a,b]$, i.e.,

$$\max_{a\le x\le b}|1+x+...+\frac{x^n}{n!}-e^x| \xrightarrow[n\to\infty]{} 0$$

by Taylor's theorem. Hence, $\{1+x+...+\frac{x^n}{n!}\}$ is a Cauchy sequence in $\mathcal{P}([a,b])$, but its limit is $e^x \notin \mathcal{P}([a,b])$.

In these examples, the Cauchy sequence "acts" like it converges, but its limit is not defined in the space we are working in.

DEFINITION 4.3.4. A metric space is **complete** if every Cauchy sequence converges to an element in the space.

Completeness is a property that has to be established, and this may or may not be easy to do! As a first example, we prove

THEOREM 4.3.5. $\mathbb{R}^n$ *is complete.*

PROOF. Let $\{x_n\}$ be a Cauchy sequence in $\mathbb{R}^n$. By Theorem 4.2.7, $\{x_n\}$ is contained in a compact n-cell. This means that is has a convergent subsequence and by Theorem 4.2.6 converges itself. □

As a second example, we present

EXAMPLE 4.3.6. Recall that a real valued function f is bounded on an interval $[a,b]$ if there is a constant M such that

$$|f(x)| \le M \text{ for } a \le x \le b.$$

Also, a continuous function is bounded on a finite closed interval but a function that is bounded does not have to be continuous.

We define

$$\mathcal{M}([a,b]) = \{f : [a,b] \to \mathbb{R} \mid f \text{ is bounded}\}$$

and introduce the metric

$$d(f,g) = \sup_{[a,b]}|f(x)-g(x)|$$

for $f,g \in \mathcal{M}([a,b])$.

EXERCISE 4.3.7. Show $(\mathcal{M}([a,b]),d)$ is a metric space.

We show that $\mathcal{M}([a,b])$ is complete. This is a three step process:

(1) Find a natural candidate for a limit for a Cauchy sequence.
(2) Verify that the limit is in the metric space.
(3) Prove the Cauchy sequence converges to this limit.

Let $\{f_k\}$ be a Cauchy sequence in $\mathcal{M}([a,b])$. For a fixed x in $[a,b]$, consider the sequence of numbers $\{f_k(x)\}$. This is a Cauchy sequence in $(\mathbb{R},|\cdot|)$, since

$$|f_n(x)-f_m(x)| \le \sup_{a\le y\le b}|f_n(y)-f_m(y)| = d(f_n,f_m).$$

Since $\mathbb{R}$ is complete, $\{f_n(x)\}$ converges to a real number. Define the function $f:[a,b]\to\mathbb{R}$ by

$$f(x) = \lim_{n\to\infty} f_n(x) \qquad a \le x \le b.$$

This is our candidate for the limit.

We next show $f \in \mathcal{M}([a,b])$, i.e., f is bounded. Choose N so that $d(f_m, f_n) \leq 1$ for $n, m \geq N$. In particular,

$$\sup_{a \leq x \leq b} |f_N(x) - f_m(x)| \leq 1 \qquad \text{for } m \geq N.$$

We let $m \to \infty$ in this inequality and obtain

$$\sup_{a \leq x \leq b} |f_N(x) - f(x)| \leq 1.$$

Since $f_N \in \mathcal{M}([a,b])$, there is an M such that $\sup_{a \leq x \leq b} |f_N(x)| \leq M$ and so

$$\sup_{a \leq x \leq b} |f(x)| \leq M + 1.$$

Finally, we show $\{f_n\} \to f$ in $(\mathcal{M}([a,b]), d)$. Note that we have *pointwise* convergence by construction, but we do not automatically have *uniform* convergence, which is the convergence notion in $(\mathcal{M}([a,b]), d)$, automatically.

Let $\epsilon > 0$. There is an N such that

$$d(f_m, f_N) < \epsilon \qquad \text{for } m \geq N.$$

This means

$$|f_m(x) - f_N(x)| < \epsilon$$

for $a \leq x \leq b$ and $m \geq N$. Taking the limit as $m \to \infty$ yields

$$|f(x) - f_N(x)| < \epsilon, \qquad a \leq x \leq b.$$

For $n \geq N$,

$$\begin{aligned} |f_n(x) - f(x)| &\leq |f_n(x) - f_N(x)| + |f_N(x) - f(x)| \\ &< 2\epsilon, \end{aligned}$$

for $a \leq x \leq b$. Hence, $d(f_n, f) < 2\epsilon$ for $n \geq N$.

We can characterize completeness in terms of a generalization of the Cantor Intersection Property (recall Definition 3.1.23).

Definition 4.3.8. Let A be a nonempty subset of a metric space (χ, d). The **diameter** of A is defined by

$$diam(A) = \sup_{x,y \in A} d(x,y).$$

Notice that $diam(A)$ can be infinite and the diameter of a set consisting of a single point is zero.

We have

Theorem 4.3.9. *A metric space is complete if and only if the intersection of every descending sequence of nonempty, closed sets whose diameters approach zero consists of a single point.*

Proof. Let $\{F_n\}$ be a sequence of closed sets in a metric space (χ, d) with $F_1 \supset F_2 \supset F_3 \supset ...$ and $diam(F_n) \to 0$, where $F_n \neq 0$ for all n.

Choose $x_n \in F_n$ for $n \geq 1$. Note that since the sets are descending,

$$\sup_{x,y,\in F_n} d(x,y) \geq \sup_{x,y \in F_{n+1}} d(x,y).$$

Given $\epsilon > 0$, choose N such that $diam(F_n) < \epsilon$ for $n \geq N$. This implies that $d(x_n, x_m) < \epsilon$ for $n, m \geq N$. Hence $\{x_n\}$ is a Cauchy sequence and there is a point

x such that $x_n \to x$. We claim $\{x\} = \bigcap_{n=1}^{\infty} F_n$. First note that if $x, y \in \bigcap_{n=1}^{\infty} F_n$, then $x, y \in F_n$ for all n, and since $d(x,y) \leq diam(F_n)$ for all n, $d(x,y) = 0$, i.e., $x = y$. Hence, $\bigcap_{n=1}^{\infty} F_n$ can consist of *at most* one point.

Now we claim $x \in F_n$ for all n. If not, since F_n^c is open for all n, there is an n and an $\epsilon > 0$ such that $N_\epsilon(x) \subset F_n^c$. But, then $d(x,y) \geq \epsilon$ for $y \in F_n$. But, this means $d(x, x_m) \geq \epsilon > 0$ for $m \geq N$, contradicting $x_n \to x$.

Now let $\{x_n\}$ be a Cauchy sequence in (χ, d). Choose $N_1 > 0$ such that $n, m \geq N_1$ implies $d(x_n, x_m) < \frac{1}{2}$. Set x_{N_1} as the first term in a subsequence. Given $N_1, N_2, ..., N_{k-1}$, choose $N_k > N_{k-1}$ such that $m, n \geq N_k$ implies $d(x_n, x_m) < \frac{1}{2^k}$. Let x_{N_k} be the k^{th} term in the subsequence. Now

$$d(x_{N_k}, x_{N_{k+1}}) < \frac{1}{2^k}$$

since $N_{k+1} > N_k$.

Define the sequence of closed "balls"

$$\begin{aligned} B_1 &= \overline{N}_1(x_{N_1}) \\ B_2 &= \overline{N}_{\frac{1}{2}}(x_{N_2}) \\ B_3 &= \overline{N}_{\frac{1}{4}}(x_{N_3}) \\ &\vdots \\ B_k &= \overline{N}_{\frac{1}{2^{k-1}}}(x_{N_k}) \\ &\vdots \end{aligned}$$

These closed sets are nonempty, since $x_{N_k} \in B_k$. Moreover, they are descending. If $y \in B_{k+1}$, then

$$\begin{aligned} d(y, x_{N_k}) &\leq d(y, x_{N_{k+1}}) + d(x_{n_{k+1}}, x_{N_k}) \\ &\leq \tfrac{1}{2^k} + \tfrac{1}{2^k} = \tfrac{1}{2^{k-1}}, \end{aligned}$$

so $B_{k+1} \subset B_k$. By assumption, there is a unique point $x \in \bigcap_{k=1}^{\infty} B_k$. Since $d(x, x_{N_k}) \leq \frac{1}{2^{k-1}}, x_{N_k} \to x$. Since $\{x_n\}$ is Cauchy, $x_n \to x$. □

We can characterize complete subsets of a complete metric space very nicely:

THEOREM 4.3.10. *Let (χ, d) be a complete metric space. A subspace $\mathbf{Y} \subset \chi$ is complete if and only if $\mathbf{Y}$ is closed.*

PROOF. Suppose $\mathbf{Y}$ is closed and $\{x_n\}$ is a Cauchy sequence in $\mathbf{Y}$. Since χ is complete, $x_k \to x \in \chi$. But, $\mathbf{Y}$ is a closed, so $x \in \mathbf{Y}$. This means $\mathbf{Y}$ is complete.

If $\mathbf{Y}$ is complete, then let x be a limit point of $\mathbf{Y}$. There is a sequence $\{x_n\}$ in $\mathbf{Y}$ with $x_n \to x$. This sequence converges in χ, so it is a Cauchy sequence in χ, and therefore in $\mathbf{Y}$. This means $x_n \to \tilde{x} \in \mathbf{Y}$ and $\mathbf{Y}$ is closed. □

4.4. Cauchy Sequences and Compactness

In this section, we develop the Cauchy sequence analog of sequential compactness, and complete the last part of Theorem 3.4.4:

THEOREM 4.4.1. **Borel-Lebesgue Theorem, part II** *Let* $(\mathbf{X}, d)$ *be a metric space and* $K \subset \mathbf{X}$*.* K *is compact if and only if it is complete and every sequence in* K *has a Cauchy subsequence.*

PROOF. Suppose K is compact and let $\{x_n\}$ be a Cauchy sequence in K. $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$ with limit in K. By Theorem 4.2.6, $\{x_n\}$ converges to the same limit. So, K is complete.

For the other direction, assume $\{x_n\}$ is a sequence in K. We know that $\{x_n\}$ has a Cauchy subsequence $\{x_{n_k}\}$ that converges to $x \in K$, since K is complete. This shows K is (sequentially) compact. □

We conclude by relating the concepts of the existence of a Cauchy subsequence of an arbitrary sequence and total boundedness.

THEOREM 4.4.2. *A subset of a metric space is totally bounded if and only if every sequence in the subset has a Cauchy subsequence.*

PROOF. Assume $(\mathbf{X}, d)$ is a metric space and $A \subset \mathbf{X}$ is totally bounded. Let $\{x_n\}$ be a sequence in A.

For $\epsilon_1 = 1$, choose $y_{1,1}, ..., y_{1,N_1} \in \mathbf{X}$ such that $A \subset \bigcup_{k=1}^{N_1} N_{\epsilon_1}(y_{1,k})$. At least one of the neighborhoods $N_{\epsilon_1}(y_{1,k})$ contains an infinite number of the $\{x_n\}$. Suppose this is $N_{\epsilon_1}(y_{1,k_1})$. Choose x_{n_1} to be one of the infinite number of $\{x_n\}$ in $N_{\epsilon_1}(y_{1,k_1})$.

For $\epsilon_2 = \frac{1}{2}$, choose $y_{2,1}, ..., y_{2,N_2}$ such that $A \subset \bigcup_{k=1}^{N_2} N_{\epsilon_2}(y_{2,k})$. For some k, $N_{\epsilon_2}(y_{2,k}) \bigcap N_{\epsilon_1}(y_{1,k_1})$ contains an infinite number of the $\{x_n\}$. Say this is $N_{\epsilon_2}(y_{2,k_2}) \bigcap N_{\epsilon_1}(y_{1,k_1})$ and choose x_{n_2} to be one of the points $\{x_n\}$ there.

Inductively, for $\epsilon_m = \frac{1}{m}$, we choose $y_{m,1}, ..., y_{m,N_m}$ so that $A \subset \bigcup_{k=1}^{N_m} N_{\epsilon_m}(y_{m,k})$. For some $1 \leq k_m \leq N_m, N_{\epsilon_m}(y_{m,k_m}) \bigcap N_{\epsilon_{m-1}}(y_{m-1,k_{m-1}}) \bigcap ... \bigcap N_{\epsilon_1}(y_{1,k_1})$ contains an infinite number of $\{x_n\}$. Choose x_{n_m} to be one of these points. By construction, $\{x_{n_m}\}$ is Cauchy since for $\epsilon > 0$,

$$d(x_{n_m}, x_{n_k}) < \epsilon_{\max(m,k)} = \min\{\frac{1}{m}, \frac{1}{k}\} < \epsilon$$

for $k, m > \frac{1}{\epsilon}$.

Next, assume every sequence has a Cauchy subsequence. For $\epsilon > 0$, choose $x_1 \in K$. Then choose $x_2 \in K$ such that $x_2 \notin N_\epsilon(x_1)$. Choose $x_3 \in K$ so $x_3 \notin N_\epsilon(x_2) \bigcup N_\epsilon(x_1)$. Proceed inductively. This process *must* stop after a finite number of steps since otherwise we would obtain a sequence that can have no Cauchy subsequence. Thus, $K \subset \bigcup_{m=1}^{N} N_\epsilon(x_m)$ for some N. □

We can restate

THEOREM 4.4.3. **Borel-Lebesguq Theorem II** *A subset of a metric space is compact if and only if it is complete and totally bounded.*

CHAPTER 5

Sequences in $\mathbb{R}^n$

As an application of sequences, we consider $\mathbb{R}^n$. Because of the vector space structure on $\mathbb{R}^n$ and the order structure on $\mathbb{R}$, we can say more about sequences and convergence.

5.1. Arithmetic Properties and Convergence

The first issue is to develop the arithmetic properties of convergence:

THEOREM 5.1.1. *Let $\{a_n\}$ and $\{b_n\}$ be convergent sequences in $(\mathbb{R}, |\cdot|)$ with $a_n \to a$ and $b_n \to b$. Then,*

(1) $\lim\limits_{n\to\infty} (a_n + b_n) = a + b$.
(2) $\lim\limits_{n\to\infty} ca_n = ca$ *and* $\lim\limits_{n\to\infty} (c + a_n) = c + a$ *for any* $c \in \mathbb{R}$.
(3) $\lim\limits_{n\to\infty} a_n b_n = ab$.
(4) $\lim_{n\to\infty} \frac{1}{a_n} = \frac{1}{a}$, *provided* $a \neq 0$ *and* $a_n \neq 0$ *for all* n.

PROOF. (1) Let $\epsilon > 0$ be given. There exists $N_1 \in \mathbb{N}$ such that $n \geq N_1$ implies $|a_n - a| < \frac{\epsilon}{2}$. Also, there exists $N_2 \in \mathbb{N}$ such that $n \geq N_2$ implies $|b_n - b| < \frac{\epsilon}{2}$. So, $n \geq \max(N_1, N_2)$ implies

$$|a_n + b_n - (a + b)| \leq |a_n - a| + |b_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

(2) Let $c \in \mathbb{R}$ and $\epsilon > 0$. Notice the result is obvious for $c = 0$, so assume $c \neq 0$. There exists N such that $n \geq N$ implies $|a_n - a| < \min(\epsilon, \frac{\epsilon}{|c|})$. Then $n \geq N$ implies

$$|c + a_n - (c + a)| = |a_n - a| < \epsilon$$

and

$$|ca_n - ca| = |c||a_n - a| < |c|\frac{\epsilon}{|c|} = \epsilon.$$

(3) Let $\epsilon > 0$. We have two cases. First, assume $ab = 0$. Then a or b is zero. Without loss of generality assume $a = 0$. We know by Theorem 4.2.7 that there exists $B > 0$ such that $|b_n| < B$ for all n. There exists $N \in \mathbb{N}$ such that $n \geq N$ implies $|a_n - a| = |a_n| < \frac{\epsilon}{B}$. But now $n \geq N$ implies

$$|a_n b_n - ab| = |a_n b_n| \leq |a_n||b_n| \leq \frac{\epsilon}{B} B = \epsilon.$$

Second, assume $ab \neq 0$. Again by Theorem 4.2.7, there exist $B > \sqrt{3\epsilon}$ such that $|a_n| < B$ and $|b_n| < B$ for all n. Also, there exist N_1 and N_2

such that $n \geq N_1$ implies $|a_n - a| < \min(|a|, \frac{\epsilon}{3B})$ and $n \geq N_2$ implies $|b_n - b| < \min(|b|, \frac{\epsilon}{3B})$. So, $n \geq \max(N_1, N_2)$ implies

$$\begin{aligned} |a_n b_n - ab| &= ||a_n b_n| - |ab|| \\ &< |(|a| + \frac{\epsilon}{3B})(|b| + \frac{\epsilon}{3B}) - |ab|| \\ &< ||ab| - \frac{\epsilon}{3} - \frac{\epsilon}{3} - \frac{\epsilon}{3} - |ab|| \\ &= \epsilon. \end{aligned}$$

(4) Let $\epsilon > 0$. There exists $N \in \mathbb{N}$ such that $n \geq N$ implies $|a - a_n| < \min(\frac{|a|^2\epsilon}{2}, \frac{|a|}{2})$. Then $n \geq N$ also implies

$$|\frac{1}{a_n} - \frac{1}{a}| = \frac{|a - a_n|}{aa_n} < \frac{|a - a_n|}{\frac{|a|^2}{2}} < \epsilon.$$

□

THEOREM 5.1.2. *Suppose* $\{x_m\}$ *is a sequence in* $(\mathbb{R}^n, \|\cdot\|)$. *Then,* $x_m \to x$ *if and only if each component of* x_m *converges to the corresponding component of* x.

PROOF. Let $x_m = (x_{m,1}, ..., x_{m,n})$ and $x = (y_1, ..., y_n)$. The claim is

$$(x_m \to x) \Leftrightarrow (x_{m,k} \to y_k), \qquad 1 \leq k \leq n.$$

If $x_m \to x$, then the inequality

$$|x_{m,k} - y_k| \leq \| x_m - x \|, \qquad 1 \leq k \leq n$$

shows $x_{m,k} \to y_k$ for $1 \leq k \leq n$.

If $x_{m,k} \to y_k, 1 \leq k \leq n$, then given $\epsilon > 0$ there is an N_k such that

$$|x_{m,k} - y_k| < \epsilon \text{ for } m \geq N_k$$

for each $1 \leq k \leq n$. For $m \geq \max\{N_1, ..., N_n\}$,

$$\| x_m - x \| = (\sum_{k=1}^{n} (x_{m,k} - y_k)^2)^{\frac{1}{2}} \leq \sqrt{n}\epsilon.$$

□

THEOREM 5.1.3. *Suppose* $\{x_m\}$ *and* $\{y_m\}$ *are sequences in* $(\mathbb{R}^n, \|\cdot\|)$ *and* $\{a_m\}$ *is a sequence in* $(\mathbb{R}, |\cdot|)$ *such that* $x_m \to x, y_m \to y, a_m \to a$. *Then*

(1) $\lim_{m\to\infty} (x_m + y_m) = x + y$.
(2) $\lim_{m\to\infty} x_m \cdot y_m = x \cdot y$.
(3) $\lim_{m\to\infty} a_m x_m = ax$.

PROOF. (1) For each $1 \leq k \leq n$, the kth components of $\{x_m\}$ and $\{y_m\}$ converge to the kth components of x and y, respectively. By Theorem 5.1.1, then, the kth component of $\{x_m + y_m\}$ converges to $x + y$ for all $1 \leq k \leq n$. Now by Theorem 5.1.2, we have the desired result.

(2) Let $x_m = (x_{m,1}, x_{m,2}, ..., x_{m,n}), y_m = (y_{m,1}, y_{m,2}, ..., y_{m,n}), x = (c_1, c_2, ..., c_n),$ and $y = (d_1, d_2, ..., d_n)$ for all $m \in \mathbb{N}$. Then

$$\begin{aligned}\lim_{m\to\infty} x_m \cdot y_m &= \lim_{m\to\infty} \sum_{k=1}^{n} x_{m,k} y_{m,k} \\ &= \sum_{k=1}^{n} \lim_{m\to\infty} x_{m,k} y_{m,k} \\ &= \sum_{m=1}^{n} c_m d_m \qquad \text{by Theorem 5.1.1} \\ &= x \cdot y.\end{aligned}$$

(3) Using the notation above, notice $a_m x_m = (a_m x_{m,1}, ..., a_m x_{m,n})$. By Theorem 5.1.1, each of these components converges to ac_k, so by Theorem 5.1.2, we have the desired result.

$\square$

5.2. Sequences in $\mathbb{R}$ and Order.

Order can be used to say a lot about convergence. For example,

DEFINITION 5.2.1. A sequence $\{x_n\}$ in $(\mathbb{R}, |\cdot|)$ is

(1) **monotonically increasing** if $x_n \leq x_{n+1}$ for $n = 1, 2, 3, \ldots.$
(2) **monotonically decreasing** if $x_n \geq x_{n+1}$ for $n = 1, 2, 3, \ldots.$

If it is either, we say it is **monotonic**.

THEOREM 5.2.2. *Suppose* $\{x_n\}$ *in* $(\mathbb{R}, |\cdot|)$ *is monotonic. Then* $\{x_n\}$ *converges if and only if it is bounded.*

PROOF. We know convergence implies boundedness. Assume $x_n \leq x_{n+1}$ for all n. Let $A = \{x_n \mid n \in \mathbb{N}\}$ (the range of $\{x_n\}$, i.e., the set of elements in the sequence), and since $\{x_n\}$ is bounded, let $x = \sup A < \infty$. We claim $x_n \to x$.

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

$$x - \epsilon < x_N \leq x,$$

otherwise $x - \epsilon$ is an upper bound for A smaller than x. But

$$x - \epsilon < x_n \leq x \qquad \text{for all } n \geq N.$$

Now, if $\{x_n\}$ is monotonically decreasing, then $\{-x_n\}$ is monotonically increasing and converges, and by Theorem 5.1.1, part (2), $\{x_n\}$ converges. □

We would like to characterize the "size" of the range of a sequence. To do this, we also have to consider the subsequential limits and the possibility that the sequence is unbounded. For this, we consider two special instances of divergence.

DEFINITION 5.2.3. If $\{x_n\}$ is a sequence in $(\mathbb{R}, |\cdot|)$ such that for any M there is an N with $x_n \geq M$ for $n \geq N$, we write $x_n \to \infty$ and say $\{x_n\}$ **diverges to infinity**.

If $\{x_n\}$ is a sequence in $(\mathbb{R}, |\cdot|)$ such that for any M there is an N with $x_n \leq M$ for $n \geq N$, we write $x_n \to -\infty$ and say that $\{x_n\}$ **diverges to** $-\infty$.

We can now define the "bounds" on the range of a sequence in the limit of large index.

DEFINITION 5.2.4. Let $\{x_n\}$ be sequence in $\mathbb{R}$ and let A consist of all the subsequential limits of $\{x_n\}$ plus ∞ or $-\infty$ if there is a subsequence that diverges to ∞ or $-\infty$ respectively. Let

$$x^* = \sup A, \quad x_* = \inf A.$$

x^* and x_* are called the **upper** and **lower limits** of $\{x_n\}$ and we write

$$x^* = \limsup_{n\to\infty} x_n$$

$$x_* = \liminf_{n\to\infty} x_n.$$

(Note: Here, we use ∞ as the supremum of a set of numbers unbounded above and $-\infty$ as the infimum of a set of numbers unbounded below.)

EXAMPLE 5.2.5. $\{x_n\} = (1, -1, 1^2, 1, -1, 2^2, 1, -1, 3^2, 1, -1, 4^2, \ldots\}$
Here, $A = \{1, -1, \infty\}, \limsup_{n\to\infty} x_n = \infty$, and $\liminf_{n\to\infty} x_n = -1$.

EXAMPLE 5.2.6. Let $\{x_n\} = \mathbb{Q}$. Then $\limsup_{n\to\infty} x_n = \infty$ and $\liminf_{n\to\infty} x_n = -\infty$.

EXAMPLE 5.2.7. $\{x_n\} = \{(-1)^n(1+\frac{1}{n})\} = \{-2, \frac{3}{2}, \frac{-4}{3}, \frac{5}{4}, \frac{-6}{5}, \frac{7}{6}, ...\}$. Here, $\limsup_{n\to\infty} x_n = 1$ and $\liminf_{n\to\infty} x_n = -1$.

We can immediately connect these notions to convergence:

THEOREM 5.2.8. *A sequence $\{x_n\}$ in $(\mathbb{R}, |\cdot|)$ converges to x if and only if $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = x$.*

EXERCISE 5.2.9. Prove Theorem 5.2.8.

There are other ways to define lim sup and lim inf and these give alternatives for computing their values.

THEOREM 5.2.10. *Let $\{x_n\}$ be a sequence in $(\mathbb{R}, |\cdot|)$. Then*

(1) $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_n, x_{n+1}, x_{n+2}, ...\}$.
(2) $\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_n, x_{n+1}, x_{n+2}, ...\}$.
(3) *If $x = \limsup_{n\to\infty} x_n$, then for every $\epsilon > 0$ there is an N such that $x_n \leq x + \epsilon$ for $n \geq N$.*
(4) *If $x = \liminf_{n\to\infty} x_n$, then for every $\epsilon > 0$ there is an N such that $x_n \geq x - \epsilon$ for $n \geq N$.*

EXERCISE 5.2.11. Prove Theorem 5.2.10.

The following result is interesting and important.

THEOREM 5.2.12. *Let $\{x_n\}$ be a sequence in $(\mathbb{R}, |\cdot|), x^* = \limsup_{n\to\infty} x_n, x_* = \liminf_{n\to\infty} x_n$, and $A =$ the set of subsequential limits of $\{x_n\}$. Then x^* and x_* are in A.*

EXERCISE 5.2.13. Prove Theorem 5.2.12.

We can use these ideas to relate two different sequences.

THEOREM 5.2.14. *Let $\{x_n\}$ and $\{y_n\}$ be sequences in $(\mathbb{R}, |\cdot|)$.*

(1) *If $x_n \leq y_n$ for all $n \geq N$ for some fixed N, then*

$$\liminf_{n\to\infty} x_n \leq \liminf_{n\to\infty} y_n$$
$$\limsup_{n\to\infty} x_n \leq \limsup_{n\to\infty} y_n.$$

(2) *If $0 \leq x_n \leq y_n$ for $n \geq N$ where N is some fixed N and $y_n \to 0$, then $x_n \to 0$.*

EXERCISE 5.2.15. Prove Theorem 5.2.14.

5.3. Series

Series are just a special kind of sequence.

DEFINITION 5.3.1. Let $\{x_n\}$ be a sequence in $(\mathbb{R}| \cdot |)$. The **partial sums** associated to $\{x_n\}$ are defined

$$S_n = \sum_{m=1}^{n} x_m.$$

The sequence of partial sums may or may not converge. If it does, we define

DEFINITION 5.3.2. Let $\{x_n\}$ be a sequence in $(\mathbb{R}, |\cdot|)$ and $\{S_n\}$ the sequence of partial sums. If $\{S_n\}$ converges we define the **series** associated to $\{x_n\}$ as

$$\sum_{n=1}^{\infty} x_n = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{m=1}^{n} x_m$$

and say the series **converges**. Otherwise, we say the series **diverges**.

The Cauchy Criterion for convergence becomes

THEOREM 5.3.3. *A series* $\sum_{n=1}^{\infty} x_n$ *converges if and only if for every* $\epsilon > 0$*, there is an* N *such that*

$$|\sum_{k=n}^{m} a_k| < \epsilon \qquad \text{for } m > n \geq N.$$

EXERCISE 5.3.4. Prove Theorem 5.3.3.

This implies

THEOREM 5.3.5. *If a series* $\sum_{n=1}^{\infty} x_n$ *converges, then* $\lim_{n\to\infty} x_n = 0$.

EXERCISE 5.3.6. Prove Theorem 5.3.5

The converse does not hold.

EXAMPLE 5.3.7. $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

In general, if the summands in a series vary in sign, then convergence becomes a delicate and difficult issue. When the summands have the same sign, things are simpler. For example,

THEOREM 5.3.8. *A series of nonnegative terms converges if and only if the partial sums form a bounded sequence.*

EXERCISE 5.3.9. Prove Theorem 5.3.8

CHAPTER 6

Continuous Funtions on Metric Spaces

We now study functions from one metric space to another. Recall Definition 1.2.1.

EXAMPLE 6.0.10. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & (x_1, x_2) \neq (0,0) \\ 0 & (x_1, x_2) = (0,0). \end{cases}$$

EXAMPLE 6.0.11. Define $T : \mathcal{C}([a,b]) \to \mathbb{R}$ by

$$T(f) = \int_a^b f(s)ds.$$

EXAMPLE 6.0.12. Define $S : \mathcal{C}([a,b]) \to \mathcal{C}([a,b])$ by

$$(S(f))(t) = \int_a^t f(s)ds, \quad a \leq t \leq b.$$

EXAMPLE 6.0.13. Let $\mathcal{C}^1([a,b])$ consist of those functions in $\mathcal{C}([a,b])$ with a continuous first derivative. Define $D : \mathcal{C}^1([a,b]) \to \mathcal{C}([a,b])$ by

$$D(f)(x) = f'(x), \quad a \leq x \leq b.$$

EXAMPLE 6.0.14. Suppose $f \in \mathcal{C}([0,\pi])$. Recall that we define its Fourier sine series as

$$\sum_{k=1}^{\infty} a_k \sin(kx), \quad 0 \leq x \leq \pi$$

with

$$a_k = \frac{2}{\pi} \int_0^{\pi} f(s) \sin(ks) ds.$$

We consider the function F that takes a function f to its sequence of Fourier coefficients $\{a_1, a_2, a_3, ...\}$. Bessel's inequality states

$$\sum_{k=1}^{\infty} a_k^2 \leq \frac{2}{\pi} \int_0^{\pi} f^2(s) ds < \infty.$$

Hences, $F : \mathcal{C}([a,b]) \to l^2$.

This gives you some idea of the tremendous range of applications of functions on metric spaces.

6.1. Limit of a Function

Given our interest in sequences, it is no surprise that we consider functions that have special behavior when applied to sequences. Namely, if a sequence converges, then we want the image under a function in question to converge. We begin by defining the notion of a limit of a function of a sequence.

DEFINITION 6.1.1. Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, f a function with $f : A \to \mathbf{Y}$, and x a limit point of A.

We write

$$\lim_{s \to x} f(s) = y$$
$$f(s) \to y \text{ as } s \to x$$
$$\lim_{n \to \infty} f(x_n) = y$$

for every sequence $\{x_n\}$ in A such that $x_n \neq x$ for all n and $\lim_{n\to\infty} x_n = x$. We say that f **has a limit at** x.

Note: The convergence $\lim_{n\to\infty} x_n = x$ is in the metric of $\mathbf{X}$, i.e., $\lim_{n\to\infty} d_x(x_n, x) = 0$, while the convergence $\lim_{n\to\infty} f(x_n) = y$ is in the metric of $\mathbf{Y}$, i.e., $\lim_{n\to\infty} d_y(f(x_n), y) = 0$.

EXAMPLE 6.1.2. Note that in the function in Figure 6.1, $\lim_{s\to x} f(s) = y \neq f(x)$. This is one reason for having the condition $x_n \neq x$ on the sequence.

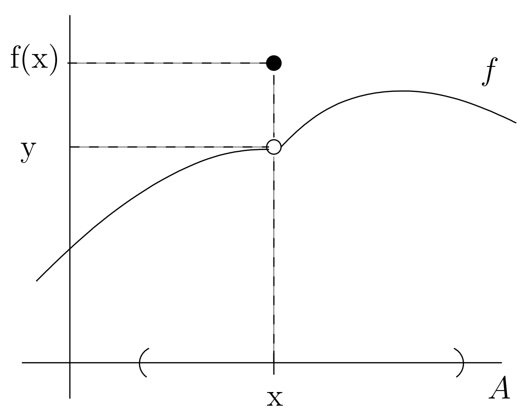


FIGURE 6.1

EXAMPLE 6.1.3. Consider T in Ex. 6.0.11. $T(f)$ has a limit at $f \in \mathcal{C}([a, b])$ if there is a number I with

$$\int_0^b f_n(s)ds \to I,$$

for all sequences $\{f_n\}$ in $\mathcal{C}([a, b])$ with $f_n \to f$ in $\mathcal{C}([a, b])$, i.e., uniform convergence.

Is this true? More later...

It follows immediately that

THEOREM 6.1.4. *Let $f : A \subset \mathbf{X} \to \mathbf{Y}$, where $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ are metric spaces. If f has a limit at a point x, then this limit is unique.*

EXAMPLE 6.1.5. Consider Ex. 6.0.10, and let $\{(x_{1,n}, x_{2,n})\}$ be a sequence in $\mathbb{R}^2$. If $(x_{1,n}, x_{2,n}) \to (0,0)$, with $x_{1,n} \neq 0$ for all n and $x_{2,n} = x_{1,n}$ for all n, then

$$f(x_{1,n}, x_{2,n}) = \frac{x_{1,n}^2}{2x_{1,n}^2} = \frac{1}{2} \qquad \text{for all } n.$$

If $\{(x_{1,n}, x_{2,n})\} = \{(0, x_{2,n})\}$, where $x_{2,n} \neq 0$ for all n, then

$$f(x_{1,n}, x_{2,n}) = 0 \qquad \text{for all } n.$$

f cannot have a limit at $(0,0)$.

This example points to the difficulties that can happen in a general metric space. The condition "for all sequences $\{x_n\}$ with $x_n \to x, x_n \neq x$ all n" is very strong.

We can characterize the notion of a limit at a point using sets.

THEOREM 6.1.6. *Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, f a function with $f : A \to \mathbf{Y}$, and x a limit point of A. f has limit y at x if and only if for every $\epsilon > 0$ there is a $\delta > 0$ such that $d_y(f(s), y) < \epsilon$ for all $s \in A$ with $0 < d_x(x, s) < \delta$.*

PROOF. Suppose the second property is false. There is an $\epsilon > 0$ such that for every $\delta > 0$ there is an $s \in A$ for which $0 < d_x(x, s) < \delta$ but $d_y(f(s), y) \geq \epsilon$. Take $\delta_n = \frac{1}{n}$ for $n \in \mathbf{N}$. We get a sequence $\{x_n\}$ with $x_n \to x$ but $\{f(x_n)\}$ cannot converge to y.

On the other hand, suppose the second property holds and $\{x_n\}$ is a sequence in A with $x_n \neq x$ for all n and $x_n \to x$. Given $\epsilon > 0$ there is a $\delta > 0$ such that

$$d_y(f(s), y) < \epsilon \qquad \text{for } s \in A, 0 < d_x(s, x) < \delta.$$

Given δ, there is an N such that

$$0 < d_x(x, x_n) < \delta, \quad n \geq N.$$

Hence,

$$d_y(f(x_n), y) < \epsilon, \quad n \geq N.$$

□

6.2. Continuous Functions

We now consider functions that "preserve" convergence of a sequence.

DEFINITION 6.2.1. Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, f a function with $f : A \to \mathbf{Y}$, and $x \in A$. f is **continuous at** x if for every sequence $\{x_n\}$ in A that converges to x, the sequence $\{f(x_n)\}$ converges to $f(x)$ in $\mathbf{Y}$. If f is continuous at every point in A, then f is **continuous on** A.

> Note: f must be defined at x in order to be continuous at x.
> Note: A function is continous at any isolated point at which it is defined.

EXAMPLE 6.2.2. The function in Ex. 6.0.10 is not continuous.

EXAMPLE 6.2.3. Regarding Ex. 6.1.3, ideally we would like to have

$$\int_a^b f_n(s)ds \to \int_a^b f(s)ds$$

for any sequence $\{f_n\}$ in $\mathcal{C}([a,b])$ that converges to $f \in \mathcal{C}([a,b])$. This would make the integral $\int_a^b \cdot ds$ continuous on $\mathcal{C}([a,b])$. More later...

There are other characterizations of continuity:

THEOREM 6.2.4. *Let* $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ *be metric spaces,* $A \subset \mathbf{X}$*,* $f : A \to \mathbf{Y}$*, and* $x \in A$*.*

(1) *f is continuous at x if and only if for every $\epsilon > 0$, there is a $\delta > 0$ such that*

$$d_y(f(s), f(x)) < \epsilon$$

for all $s \in A$ with $d_x(s,x) < \delta$.

(2) *f is continuous at x if and only if for every neighborhood V of $f(x)$ there is a neighborhood U of x such that $f(U \cap A) \subset V$.*

EXERCISE 6.2.5. Prove Theorem 6.2.4.

As Rudin remarks, none of these ideas depend on $\mathbf{X} \setminus A$, so we can restrict the discussion to the case $A = \mathbf{X}$. There is a useful characterization of continuity on a set:

THEOREM 6.2.6. *Let* $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ *be metric spaces. A map* $f : \mathbf{X} \to \mathbf{Y}$ *is continuous on* $\mathbf{X}$ *if and only if* $f^{-1}(G)$ *is open in* $\mathbf{X}$ *for every open set* $G \subset \mathbf{Y}$*.*

PROOF. Let f be continous on $\mathbf{X}$ and $G \subset \mathbf{Y}$ open. Suppose $x \in \mathbf{X}$ and $f(x) \in G$. There is an $\epsilon > 0$ such that $y \in G$ if $d_y(f(x), y) < \epsilon$. Since f is continuous at x, there is a $\delta > 0$ such that

$$d_y(f(s), f(x)) < \epsilon \quad \text{ if } d_x(s,x) < \delta.$$

Hence, $s \in f^{-1}(G)$ as soon as $d_x(s,x) < \delta$, and x is an interior point of $f^{-1}(G)$, and $f^{-1}(G)$ must be open.

Conversely, suppose $f^{-1}(G)$ is open in $\mathbf{X}$ for every open $G \subset \mathbf{Y}$. Choose $x \in \mathbf{X}$ and $\epsilon > 0$. Let G be the set of $y \in \mathbf{Y}$ such that $d_y(y, f(x)) < \epsilon$. G is open, so $f^{-1}(G)$ is open. There is a $\delta > 0$ such that $s \in f^{-1}(G)$ as soon as $d_x(s,x) < \delta$. But if $s \in f^{-1}(G)$, then $f(s) \in G$, so $d_y(f(s), f(x)) < \epsilon$. □

During the rest of this chapter, we discuss properties of continuous functions in various settings. As a first result,

THEOREM 6.2.7. *Let* $(\mathbf{W}, d_w), (\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ *be metric spaces,* $A \subset \mathbf{W}$*,* $f : A \to \mathbf{X}$*,* $g : f(A) \to \mathbf{Y}$ *and define* $h : A \to \mathbf{Y}$ *by* $h(w) = g(f(w))$ *for* $w \in A$*. If* f *is continuous at* $w \in A$ *and* g *is continuous at* $f(w)$*, then* h *is continuous at* w*.*

PROOF. Let $\epsilon > 0$ be given. There is a $\eta > 0$ such that

$$d_y(g(x), g(f(w))) < \epsilon$$

for $x \in f(A)$ with $d_x(x, f(w)) < \eta$. But, there is a $\delta > 0$ such that

$$d_x(f(\bar{w}, f(w)) < \eta$$

if $\bar{w} \in A$ and $d_w(\bar{w}, w) < \delta$. So, $d_y(h(\bar{w}, h(w)) < \epsilon$ for $\bar{w}, w \in A, d_w(\bar{w}, w) < \delta$. □

DEFINITION 6.2.8. The function h in Theorem 6.2.6 is called the **composition** of g with f.

6.3. Uniform Continuity

Recall that we say a function is continuous on a set A if it is continuous at each point of A. It is important to realize that this is actually a rather weak property because the "degree" of continuity is allowed to vary with the point. Specifically, if $f : A \to \mathbf{Y}$, where $A \subset \mathbf{X}$ and $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ are metric spaces, and f is continuous on A, then given $a \in A$ and $\epsilon > 0$, there is a $\delta = \delta_{x,\epsilon}$ that *depends on both x and ϵ* such that $d_y(f(y), f(x)) < \epsilon$ for all $y \in A$ with $d_x(y, x) < \delta_{x,\epsilon}$.

EXAMPLE 6.3.1. Consider $f(x) = \frac{1}{x}$ on $(0, 1)$ and fix $x \in (0, 1)$. We have

$$|f(x) - f(y)| = |\frac{1}{x} - \frac{1}{y}| = |\frac{y - x}{xy}| = \frac{|y - x|}{yx}$$

for $y \in (0, 1)$. If we restrict y so $y > \frac{x}{2}$ (which allows y to be close to x), we get

$$|f(y) - f(x)| \leq \frac{2}{x^2}|x - y|.$$

Given $\epsilon > 0$, if we assume that

$$|x - y| < \frac{x^2\epsilon}{2} = \delta_{x,\epsilon},$$

then $|f(y) - f(x)| < \epsilon$ for $y \in (0, 1), y > \frac{x}{2}, |x - y| < \delta$. To guarantee that f changes no more than ϵ by this argument, we have to restrict y to a neighborhood of x whose size decreases as x becomes closer to zero.

EXERCISE 6.3.2. Convince yourself that if we choose $y = x - \frac{x^2}{2}\epsilon$, then $|f(x) - f(y)|$ is on the order of ϵ.

DEFINITION 6.3.3. Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, and $f : A \to \mathbf{Y}$. f is **uniformly continuous** on A if for every $\epsilon > 0$ there is a $\delta > 0$ such that $d_y(f(x), f(y)) < \epsilon$ for all $x, y \in A$ with $d_x(x, y) < \delta$.

EXAMPLE 6.3.4. $\frac{1}{x}$ is not uniformly continuous on $(0, 1)$ (see the problem?!). So, continuity does not imply uniform continuity.

EXAMPLE 6.3.5. Consider x^2 on $(0, 1)$ with the usual metrics. Choosing $x, y \in (0, 1)$, we have

$$|f(x) - f(y)| = |x^2 - y^2| = |x + y||x - y| < 2|x - y|.$$

Given $\epsilon > 0$, if we choose any $x, y \in (0, 1)$ with $|x - y| < \delta = \frac{\epsilon}{2}$, then $|f(x) - f(y)| < \epsilon$.

EXAMPLE 6.3.6. Consider x^2 on $\mathbb{R}$. Given any $\delta > 0$, suppose we choose x and y so $y = x - \frac{\delta}{2}$. Then,

$$|x^2 - y^2| = |x + y||x - y| = |2x - \frac{\delta}{2}| \cdot \frac{\delta}{2}.$$

Given any $\epsilon > 0$, it is possible to choose x large enough that

$$|2x - \frac{\delta}{2}| \cdot \frac{\delta}{2} > \epsilon,$$

namely, $x > \frac{\epsilon}{\delta} + \frac{\delta}{4}$. So, x^2 is not uniformly continuous on $\mathbb{R}$.

There is a special class of functions that are uniformly continuous.

DEFINITION 6.3.7. Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, and $f : A \to \mathbf{Y}$. f is **Lipschitz continuous** on A if there is a constant L such that

$$d_y(f(x), f(y)) \leq L d_x(x, y)$$

for all $x, y \in A$.

A Lipschitz continuous function is automatically uniformly continuous. Lipschitz continuity is not often discussed in standard analysis texts, but it is very important in practice and is one of the most commonly encountered forms of continuity in differential equations, numerical analysis, and engineering.

It turns out that often continuity is too weak an assumption to guarantee needed properties and we must use uniform continuity. As one example,

EXAMPLE 6.3.8. The function $f(x) = \frac{x}{1-x}$ is continuous on $[0, 1)$ with the usual metrics, but is not uniformly continuous on $[0, 1)$.

EXERCISE 6.3.9. Show this.

The sequence $\{x_n\}$, with $x_n = 1 - \frac{1}{n}$, is a Cauchy sequence in $[0, 1)$.

EXERCISE 6.3.10. Show this.

However, $f(x_n) = n - 1$, so $\{f(x_n)\}$ is not a Cauchy sequence in $f([0, 1])$.

In other words, the image of a Cauchy sequence under a continuous function need not be a Cauchy sequence! However, we can prove

THEOREM 6.3.11. *Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}, f : A \to \mathbf{Y}$ uniformly continuous on A, and $\{x_n\}$ a Cauchy sequence in A. Then, $\{f(x_n)\}$ is a Cauchy sequence in $\mathbf{Y}$.*

PROOF. Given $\epsilon > 0$, there is a $\delta > 0$ such that $d_y(f(x_n), f(x_m)) < \epsilon$ for all x_n, x_m with $d_x(x_n, x_m) < \delta$. Since $\{x_n\}$ is Cauchy, given $\delta > 0$ there is an N such that $d_x(x_n, x_m) < \delta$ for $n, m > N$. □

6.4. Continuity and Compactness

If we consider the examples in Section 6.3, we see that the properties of the underlying set A are relevant to whether a function is merely continuous or uniformly continuous. It turns out that continuity interacts very well with compactness.

As a first result,

THEOREM 6.4.1. *Suppose $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ are metric spaces, $A \subset \mathbf{X}$, and $f : A \to \mathbf{Y}$ is continuous on A. If A is compact, then $f(A)$ is compact.*

PROOF. Let $\{G_\alpha\}$ be an open cover of $f(A)$. Since f is continuous, each set $f^{-1}(G_\alpha)$ is open. Moreover, $\{f^{-1}(G_\alpha)\}$ covers A. Hence, there are $\alpha_1, ..., \alpha_n$ such that $A \subset f^{-1}(G_{\alpha_1} \cup ... \cup f^{-1}(G_{\alpha_n})$. Since $f(f^{-1}(A)) \subset A$ for all $A \subset \mathbf{Y}$,

$$f(A) \subset G_{\alpha_1} \cup ... \cup G_{\alpha_n}.$$

□

In fact, it turns out that continuity and uniform continuity are the same on compact sets.

THEOREM 6.4.2. *Let* $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ *be metric spaces,* $A \subset \mathbf{X}$*, and* $f : A \to \mathbf{Y}$ *continuous on* A*. If* A *is compact, then* f *is uniformly continuous on* A*.*

PROOF. This is a compactness argument.

Let $\epsilon > 0$ be given. To each $x \in A$, there is a number δ_x such that

$$d_y(f(x), f(y)) < \epsilon \text{ for } y \in A, d_x(x, y) < \delta_x.$$

Let G_x be the set of $y \in \mathbf{X}$ such that $d_x(x, y) < \frac{1}{2}\delta_x$. G_x is open and since $x \in G_x, \{G_x\}_{x \in A}$ is an open cover of A. There are points $x_1, ..., x_n$ in A such that $A \subset G_{x_1} \cup ... \cup G_{x_n}$. Set $\delta = \frac{1}{2}\min\{\delta_{x_1}, ..., \delta_{x_n}\}$. It is crucial that we can use min rather than inf here!

Suppose $x, y \in A$ and $d_x(x, y) < \delta$. Now, $x \in G_{x_m}$ for some $1 \leq m \leq n$, so

$$d_x(x, x_m) < \frac{1}{2}\delta_{x_m}.$$

We have

$$\begin{aligned} d_x(y, x_m) &\leq d_x(y, x) = d_x(x, x_m) < \delta + \frac{1}{2}\delta_{x_m} \\ &< \delta_{x_m} \end{aligned}$$

Thus,

$$\begin{aligned} d_y(f(x), f(y)) &\leq d_y(f(x), f(x_m)) + d_y(f(x_m), f(y)) \\ &< 2\epsilon. \end{aligned}$$

□

This last theorem turns out to have tremendous consequences.

As a last result, we note that it is possible to say something about the continuity of the inverse map in the setting of a compact space.

THEOREM 6.4.3. *Suppose* f *is a continuous and 1-1 map of a* compact *metric space* $\mathbf{X}$ *onto a metric space* $\mathbf{Y}$*. The inverse map on* $\mathbf{Y}$ *is a continuous map from* $\mathbf{Y}$ *to* $\mathbf{X}$*.*

PROOF. We use Theorem 6.2.6 on f^{-1}, so we have to show that $f(G)$ is open in $\mathbf{Y}$ for every open set G in $\mathbf{X}$. Choose an open $G \subset \mathbf{X}$. Since G^c is a closed subset of a compact space, it is compact. Hence, $f(G^c)$ is compact in $\mathbf{Y}$ and therefore closed. Since f is 1-1 and onto, $f(G) = f(G^c)^c$, and so it is open. □

6.5. $\mathbb{R}^n$-valued Continuous Functions

We now consider the special case of functions from a metric space $(\mathbf{X}, d)$ into $\mathbb{R}$ or $\mathbb{R}^n$ with the usual metric.

First note that if f and g are such functions taking a metric space $(\mathbf{X}, d)$ into $\mathbb{R}$, then we can define $f + g, f - g, fg, \frac{f}{g}$ in a natural way. If $f, g : (\mathbf{X}, d) \to \mathbb{R}^n$, then we can define $f + g, f - g$, and $f \cdot g$ in a natural way as well. We can then write down the arithmetic properties of limits of such functions.

THEOREM 6.5.1. *Suppose* $(\mathbf{X}, d)$ *is a metric space,* $A \subset \mathbf{X}, f, g : A \to \mathbb{R}$*, with the usual metric on* $\mathbb{R}$*, and* x *is a limit point of* A*. Then*

(1) $\lim_{y \to x}(f + g)(y) = \lim_{y \to x} f(y) + \lim_{y \to x} g(y)$.
(2) $\lim_{y \to x}(fg)(y) = (\lim_{y \to x} f(y))(\lim_{y \to x} g(y))$.
(3) $\lim_{y \to x}(\frac{f}{g})(y) = \frac{\lim_{y \to x} f(y)}{\lim_{y \to x} g(y)}$, *provided* $\lim_{y \to x} g(y) \neq 0$.

PROOF. This follows from the properties of limits of sequences of real numbers. □

In the same way, we prove

THEOREM 6.5.2. *Let $(\mathbf{X}, d)$ be a metric space and $A \subset \mathbf{X}$.*

(1) *Suppose $f_1, ..., f_n : A \to \mathbb{R}$ with the usual metric and let $f : A \to \mathbb{R}^n$ be defined by $f(x) = (f_1(x), ..., f_n(x))$, where we take the usual metric on $\mathbb{R}^n$. Then f is continuous at $x \in A$, or on A, if and only if each component f_m is continuous at $x \in A$, or on A.*
(2) *If f and g are continuous maps from A into $(\mathbb{R}^n, \| \cdot \|)$, then $f + g, f \cdot g$ are continuous.*

EXAMPLE 6.5.3. If $x \in \mathbb{R}^n$ is written $x = (x_1, ..., x_n)$, then the **coordinate functions**

$$\phi_m(x) = x_m, \quad 1 \le m \le n,$$

are continuous since

$$|\phi_m(x) - \phi_m(y)| \le \| x - y \|, \quad x, y \in \mathbb{R}^n.$$

EXAMPLE 6.5.4. Repeated applications of Theorem 6.5.2 shows that **polynomials** $p : \mathbb{R}^n \to \mathbb{R}$,

$$p(x) = \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} ... \sum_{m_n=0}^{M_n} C_{m_1 m_2 ... m_n} x_1^{m_1} ... x_n^{m_n}$$

are continuous. Furthermore, all **rational functions** $\frac{p(x)}{q(x)}$, where p, q are polynomials, are continuous at all points where $q \neq 0$.

The latter result also requires Theorem 6.2.7.

EXAMPLE 6.5.5. Let $(\mathbf{X}, d)$ be a metric space and $z \in \mathbf{X}$. We can define $f : \mathbf{X} \to \mathbb{R}$ using d as

$$f(x) = d(x, z), \quad x \in \mathbf{X}.$$

The triangle inequality implies

$$\begin{aligned} d(x, z) &\le d(x, y) + d(y, z) \\ d(y, z) &\le d(x, y) + d(x, z), \quad x, y \in \mathbf{X} \end{aligned}$$

or

$$|d(x, z) - d(y, z)| \le d(x, y), \quad x, y \in \mathbf{X}$$

and therefore f is continuous on $\mathbf{X}$.

We now consider what happens for $\mathbb{R}^n$-valued continuous functions on compact sets.

DEFINITION 6.5.6. Let $(\mathbf{X}, d)$ be a metric space and $f : A \to \mathbb{R}^n$, with the usual metric on $\mathbb{R}^n$. f is **bounded on** A if there is a constant M such that

$$\| f(x) \| \le M \quad \text{for all } x \in A.$$

If f is bounded on $\mathbf{X}$, we say it is **bounded**.

THEOREM 6.5.7. *Let $(\mathbf{X}, d)$ be a metric space, $A \subset \mathbf{X}$. If $f : A \to (\mathbb{R}^n, \| \cdot \|)$ is continuous and A is compact, then f is bounded and $f(A)$ is closed.*

PROOF. See Theorem 6.4.1 and Theorem 3.6.3. □

When f is real-valued, we can say even more.

THEOREM 6.5.8. *Let $(\mathbf{X}, d)$ be a metric space, $A \subset \mathbf{X}$ compact, and $f : A \to \mathbb{R}$ continuous. Set $M = \sup\limits_{x \in A} f(x)$ and $m = \inf_{x \in A} f(x)$. There are points $y, z \in A$ with $f(y) = M$ and $f(z) = m$.*

PROOF. $f(A)$ is closed and bounded. The result follows from Theorem 2.3.23. $\square$

There is also the famous

THEOREM 6.5.9. **Intermediate Value Theorem**

Suppose f is a continuous, real valued function defined on $[a, b]$ (where we take the usual metric) and $f(a) \neq f(b)$. Let c be any number between $f(a)$ and $f(b)$. Then $f(x) = c$ for some x between a and b.

PROOF. We treat the case $f(a) < f(b)$ and assume $f(a) < c < f(b)$. Define

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

Note that $a \in A$ and b is an upper bound for A, hence $x = \sup A$ is defined and $a \leq x \leq b$. We claim $f(x) = c$. There is a sequence $\{x_n\}$ in A with $x_n \leq$, for all n and $x_n \to x$.

Since $f(x_n) < c$ for all n and f is continuous at x,

$$f(x) = \lim_{n \to \infty} f(x_n) \leq c.$$

In particular, $x \neq b$! Choose any sequence $\{y_n\}$ with $x < y_n \leq b$ for all n and $y_n \to x$. Now $f(y_n) > c$ for all n, hence

$$f(x) = \lim_{n \to \infty} f(y_n) \geq c.$$

$\square$

CHAPTER 7

Sequences of Functions and $\mathcal{C}([a,b])$

As an application of all of the theory we have developed, we want to explore the properties of the metric space

$$\mathcal{C}([a,b]) = \{f \mid f \text{ is continuous on } [a,b]\}$$

with metric $d(f,g) = \sup_{[a,b]} |f(x) - g(x)|$. That is, we consider whether or not $\mathcal{C}([a,b])$ is

- closed
- complete
- separable
- compact

and if it does not have these properties, what kind of subsets do.

In all cases, we are investigating the properties of sequences of points in $\mathcal{C}([a,b])$, that is, sequences of continuous functions. We will be particularly interested in sequences that converge.

7.1. Convergent Sequences of Functions

Following Ex. 4.3.6, we define

Definition 7.1.1. Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, and $\{f_n\}$ a sequence of functions with $f_n : A \to \mathbf{Y}$ for all n. Suppose the sequence of *points* $\{f_n(x)\}$ in $\mathbf{Y}$ converges for all $x \in A$. We define a fucntion $f : A \to \mathbf{Y}$ by

$$f(x) = \lim_{n\to\infty} f_n(x), \quad x \in A.$$

We say that $\{f_n\}$ **converges pointwise to** f **on** A and f is the **pointwise limit** of $\{f_n\}$ on A.

Example 7.1.2. Consider

$$f_n(x) = \frac{1}{n}\sin(nx) \qquad \text{on } [0,\pi].$$

(See figure 7.1.)

We have

$$|f_n(x) - 0| = |\frac{1}{n}\sin(nx)| \le \frac{1}{n} \qquad \text{for all } x \in [0,\pi],$$

so $\{f_n\} \to 0$.

Example 7.1.3. Define on $[0,1]$ (See figure 7.2.) ,

$$f_n(x) = \begin{cases} n^2 x, & 0 \le x \le \frac{1}{n} \\ n^2(\frac{2}{n} - x), & \frac{1}{n} \le x \le \frac{2}{n} \\ 0, & \frac{2}{n} \le x \le 1 \end{cases}$$

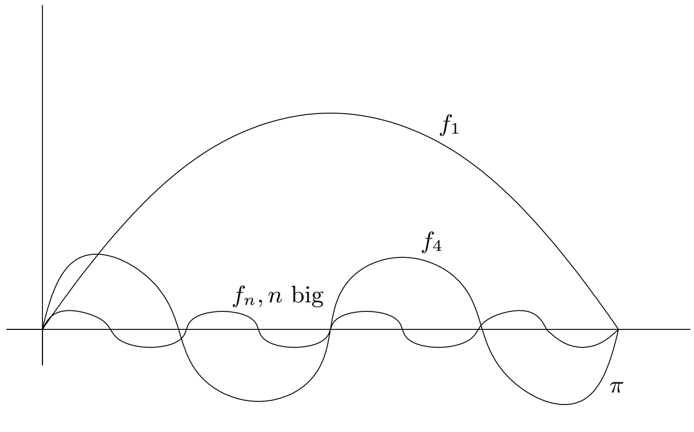


FIGURE 7.1

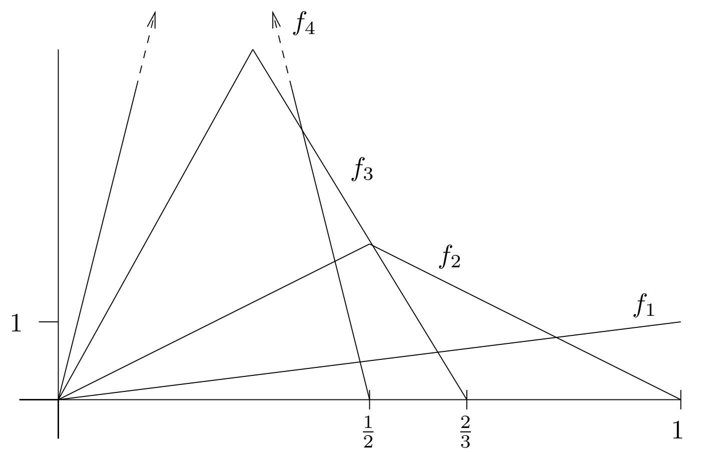


FIGURE 7.2

Given $x > 0, \frac{2}{n} \le x$ for all n sufficiently large, hence $f_n(x) \to 0$ pointwise on $[0,1]$.

This example shows that pointwise convergence can allow some pretty bad behavior!

In most situations, we want to know whether or not some particular properties of a sequence of functions is inherited by the limit.

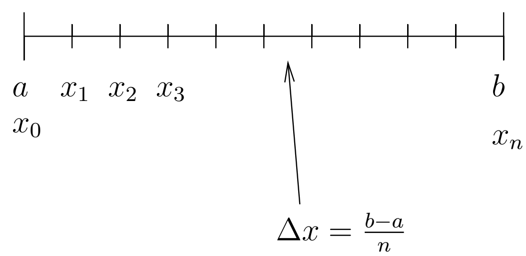
EXAMPLE 7.1.4. When talking about $\mathcal{C}([a,b])$, we want to know if the limit of a sequence of continuous functions is continuous.

EXAMPLE 7.1.5. In the numerical solution of differential equations, we want to know if the limit of a sequence of approximate solutions converge to the true solution. Consider the simplest equation

$$\begin{cases} y'(x) = f(x), & a \le x \le b \\ y(0) = y_0 \end{cases} \tag{7.1}$$

with solution $y(x) = y_0 + \displaystyle\int_a^x f(s)ds$.

We can define approximate solutions $\{y_n(x)\}$ using the rectangle rule to approximate the integral. Given n, define $\Delta x = \frac{b-a}{n}$ and $x_m = a + \Delta x \cdot m, 0 \le m \le n$;



If $x \in [a,b]$, then $x \in [x_{M-1}, x_M]$ for some $0 < M \le n$. Define

$$Y_n(x) = y_0 = \sum_{m=1}^{M-2} f(x_m)\Delta x + f(x_{M-1}) \cdot (x - x_{M-1}), \quad \text{for } a \le x \le b,$$

which you can understand from the plot:

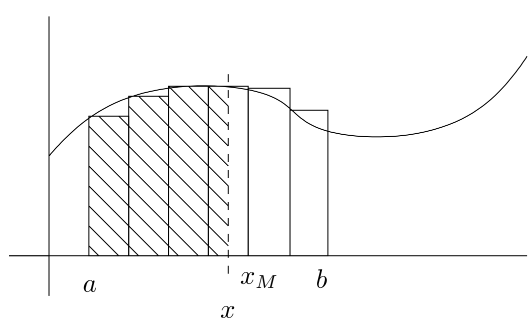


FIGURE 7.3

$Y_n(x)$ is the area of the rectangles shown above. $Y_n(x)$ is continuous on $[a,b]$ for all n if f is continuous, and we want to know if $\{Y_n(x)\} \to \{y(x)\}$, the solution of 7.1, which is also continuous if f is continuous. In most cases, we do *not* know y.

It is important to realize that often the inheritance of an analytic property by the limit of a sequence of functions is equivalent to whether or not it is justified to switch the order of two limiting processes.

EXAMPLE 7.1.6. Continuing Ex. 7.1.4, if the sequence $\{f_n\}$ in $\mathcal{C}([a,b])$ converges to f, we can rephrase the issue of inheriting continuity like this: Choose $a \le x \le b$. Then for $n \ge 1$,

$$\lim_{m\to\infty} f_n(x_m) = f_n(x)$$

for all sequences $\{x_m\}$ in $[a,b]$ with $x_m \to x$. Since $f_n \to f$ pointwise,

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty}\lim_{m\to\infty} f_n(x_m)$$

for all sequences $\{x_m\}$ in $[a,b]$ with $x_m \to x$. For f to be continuous at x, we must have

$$\lim_{m\to\infty} f(x_m) = f(x)$$

for all such sequences. In other words, we require

$$\lim_{m\to\infty}\lim_{n\to\infty} f_n(x_m) = \lim_{n\to\infty}\lim_{m\to\infty} f_n(x_m). \tag{7.2}$$

EXAMPLE 7.1.7. Continuing Ex. 7.1.5, it is easy to see that given $x \in [a,b]$, for all sufficiently fine meshes, i.e., sufficiently large n,

$$\lim_{h\to 0} \frac{Y_n(x+h) - Y_n(x)}{h} \approx f(x_{m-1})$$

and as $n \to \infty, x_{m-1} \to x$, so

$$\lim_{n\to\infty}\lim_{h\to 0} \frac{Y_n(x+h) - Y_n(x)}{h} = f(x).$$

This says $Y_n(x)$ is an approximate solution.

On the other hand, if $Y_n(x) \to y(x)$ and we want to show that y solves 7.1, then we want for $x \in [a,b]$,

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

or

$$\lim_{h\to 0}\lim_{n\to\infty} \frac{Y_n(x+h) - Y_n(x)}{h} = \lim_{n\to\infty}\lim_{n\to 0} \frac{Y_n(x+h) - Y_n(x)}{h}$$

General Principle: Whenever there is more than one limiting process in some situation, it is important to determine if the order of the limit matters.

EXAMPLE 7.1.8. Consider $\{\frac{m}{n+m}\}_{m,n=1}^{m,n=\infty}$.

$$\lim_{m\to\infty}\lim_{n\to\infty} \frac{m}{n+m} = \lim_{m\to\infty} 0 = 0$$
$$\lim_{n\to\infty}\lim_{m\to\infty} \frac{m}{n+m} = \lim_{n\to\infty} 1 = 1.$$

In fact, the limit of a sequence of continuous functions that converge pointwise is *not* necessarily continuous.

EXAMPLE 7.1.9. Consider $\{x^n\}_{n=0}^{\infty}$ on $[0,1]$. The sequence is in $\mathcal{C}([0,1])$. It converges pointwise on $[0,1]$ to

$$\chi_1(x) = \lim_{n\to\infty} x^n = \begin{cases} 0, & 0 \le x \le 1, \\ 1, & x = 1 \end{cases}$$

which is not continuous (see figure 7.4).

If $x < 1$, then given $\epsilon > 0$, choose $N > \frac{\epsilon}{-\log(x)}$ (recall $\log(x) < 0$), so $x^n < \epsilon$ for $n \ge N$. However, $x^n = 1$ for $x = 1$ and all n.

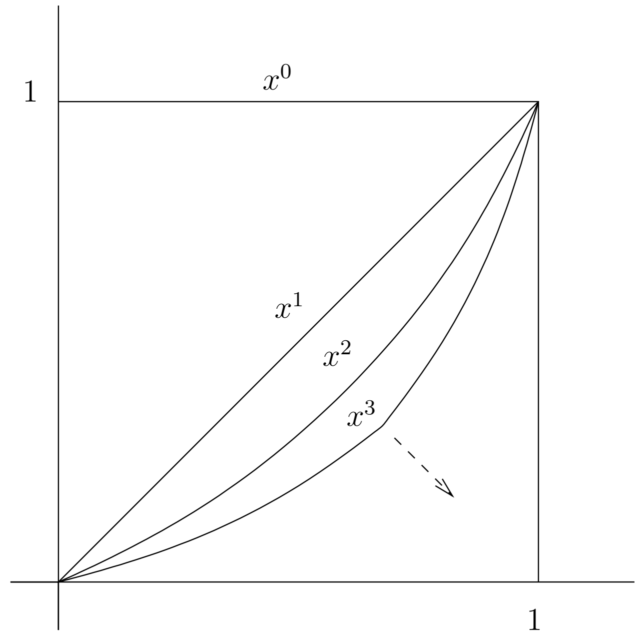


Figure 7.4

7.2. Uniform Convergence: $\mathcal{C}([a,b])$ is Closed, and Complete

Luckily, convergence in the sup metric of $\mathcal{C}([a,b])$ is stronger than pointwise convergence.

Example 7.2.1. The sequence $\{x^n\}$ in $\mathcal{C}([a,b])$ does not converge in $\mathcal{C}([a,b])$, although it does converge pointwise (see Example 7.1.9). For any $n \geq 1$,

$$\begin{aligned}\sup_{0\leq x\leq 1} |x^n - \chi_1(x)| &= \sup_{0\leq x\leq 1} |x^n - \begin{cases} 0, & 1 \leq x \leq 1 \\ 1, & x = 1 \end{cases} | \\ &= \sup_{0\leq x\leq 1} |x|^n = 1\end{aligned}$$

i.e.,

$$d(x^n, x_1) = 1 \text{ for all } n.$$

Convergence in $\mathcal{C}([a,b])$ is an example of uniform convergence.

Definition 7.2.2. Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, and $\{f_n\}$ a sequence of functions $f_n : A \to \mathbf{Y}$ for all n. $\{f_n\}$ **converges uniformly to** f **on** A if for every $\epsilon > 0$ there is an N such that

$$d_y(f_n(x), f(x)) < \epsilon \text{ for } x \in A \text{ and } n \geq N.$$

Compare this to Definition 7.1.1.

Example 7.2.3. The functions in Example 7.1.2 converge uniformly to 0 since $|\frac{1}{n}\sin(mx) - 0| \leq \frac{1}{n}$ for all $0 \leq x \leq \pi$.

EXAMPLE 7.2.4. The functions in Example 7.1.3 do *not* converge uniformly on $[0,1]$ since

$$\sup_{0\le x\le 1} |f_n(x) - 0| = n.$$

EXAMPLE 7.2.5. The sequence $\{x^n\}$ does not converge uniformly to $\chi_1(x)$ on $[0,1]$, but does on $[0,\frac{1}{2}]$.

Uniform convergence goes well with continuity.

THEOREM 7.2.6. *Let $(\mathbf{X}, d_x), (\mathbf{Y}, d_y)$ be metric spaces and $A \subset \mathbf{X}$. Suppose $\{f_n\}$ is a sequence of functions with $f_n : A \to \mathbf{Y}$ continuous on A for all n and $f_n \to (f : A \to \mathbf{Y})$ uniformly on A. Then f is continous of A.*

PROOF. Choose $x \in A$ and $\epsilon > 0$. We want to show we can make $d_y(f(y), f(x))$ smaller than ϵ by making $d_x(x, y)$ small. Uniform convergence means that we can make $d_y(f(x), f_n(x))$ and $d_y(f(y), f_n(y))$ small, so for $y \in A$, we write

$$d_y(f(x), f(y)) \le d_y(f(x), f_n(x)) + d_y(f_n(x), f_n(y)) + d_y(f_n(y), f(y)).$$

By uniform convergence, there is an N such that $d_y(f(x), f_n(x)) < \epsilon$ and $d_y(f(y), f_n(y)) < \epsilon$ for $n \ge N$, independent of x and y. Since f_n is continuous on A, there is a $\delta > 0$ such that for any *fixed* $n \ge N$,

$$d(f_n(x), f_n(y)) < \epsilon,$$

for all $y \in A, d_x(x, y) < \delta$. Hence, using that value of n, we conclude

$$d(f(x), f(y)) < 3\epsilon \text{ for all } y \in A, d_x(x, y) < \delta.$$

□

The functions in Example 7.1.3 show the converse does not hold: a sequence of continuous functions can converge to a continuous function without the convergence being uniform.

We now discuss the related topic of completeness. We state the Cauchy criterion for uniform convergence.

THEOREM 7.2.7. *Let $(\mathbf{X}, d_x)$ and $(\mathbf{Y}, d_y)$ be metric spaces, $\mathbf{Y}$ complete, $A \subset \mathbf{X}$, and $\{f_n\}$ a sequence with $f_n : A \to \mathbf{Y}$ for all n. $\{f_n\}$ converges uniformly on A if and only if for every $\epsilon > 0$ there is an N such that*

$$d_y(f_n(x), f_m(x)) < \epsilon \quad \text{for } n, m \ge N \text{ and } x \in A.$$

PROOF. Suppose $\{f_n\}$ converges uniformly on A to f. Given $\epsilon > 0$, there is an N such that

$$d_y(f_n(x), f(x)) < \epsilon, \quad x \in A, n \ge N.$$

So,

$$\begin{aligned} d_y(f_n(x), f_m(x)) &\le d_y(f_n(x), f(x)) + d_y(f(x), f_m(x)) \\ &< 2\epsilon \end{aligned}$$

for $x \in A, n, m \ge N$.

Conversely, suppose the Cauchy condition holds. The sequence of points $\{f_n(x)\}$ is a Cauchy sequence in $\mathbf{Y}$ for each $x \in A$, and therefore has a limit in $\mathbf{Y}$ that we call $f(x)$. This defines $f : A \to \mathbf{Y}$. $\{f_n\}$ converges pointwise to f on A and we have to show the convergence is uniform.

Let $\epsilon > 0$ and choose N so

$$d_y(f_n(x), f_m(x)) < \frac{\epsilon}{2} \quad \text{for } n, m \geq N, x \in A.$$

Fix n and let $m \to \infty$. Since $f_m \to f$ as $m \to \infty$ and $d_y(f(x), \cdot)$ is continuous, $d_y(f_n(x), f(x)) \leq \frac{\epsilon}{2} < \epsilon$ for $n \geq N, x \in A$. □

Now, we observe that convergence in $\mathcal{C}([a,b])$ is uniform convergence and $\mathbb{R}$ is complete. Theorems 7.2.6 and 7.2.7 imply

Theorem 7.2.8. *$\mathcal{C}([a,b])$ is closed and complete.*

7.3. $\mathcal{C}([a,b])$ is Separable

We know $\mathbb{R}$ is separable and, in particular, $\mathbb{Q}$ is dense in $\mathbb{R}$. This is extremely important from a practical point of view because it means we can approximate irrational numbers using rational numbers. This is what makes large scale scientific computing possible, for example.

We prove that $\mathcal{C}([a,b])$ is separable by first showing that continuous functions can be approximated arbitrarily well by polynomials.

Theorem 7.3.1. **Weierstrass Approximation Theorem** *Assume that f is continuous on $[a,b]$. Given $\epsilon > 0$, there is a polynomial p_n of sufficiently high degree n such that*

$$d(f, p_n) = \sup_{a \leq x \leq b} |f(x) - p_n(x)| < \epsilon.$$

Another way to state this result is that there is a sequence of polynomials $\{p_n\}$ (of course in $\mathcal{C}([a,b])$) that converges to f in $\mathcal{C}([a,b])$, that is, uniformly.

This theorem is profoundly important. It is the reason, for example, that the use of polynomials is so widespread in numerical analysis, i.e., approximation of functions, integrals, solutions of differential equations, and so on.

> Note: unlike Taylor's polynomials, this result does not require increasing smoothness of f to increase the accuracy of the polynomial approximations.

To prove that $\mathcal{C}([a,b])$ is separable, we first note that if

$$p(x) = \sum_{m=0}^{n} a_m x^m \text{ and } \tilde{p}(x) = \sum_{m=0}^{n} \tilde{a}_m x^m$$

are two polynomials on $[a,b]$, then

$$\begin{aligned} d(p, \tilde{p}) &= \sup_{a \leq x \leq b} |p(x) - \tilde{p}(x)| \\ &\leq c \cdot \sum_{m=0}^{n} |a_m - \tilde{a}_m| \\ &\leq (n+1) \cdot c \cdot \max_{0 \leq m \leq n} |a_m - \tilde{a}_m| \end{aligned}$$

where c is a constant that depends on a, b, and n. ($c = \max\limits_{0 \leq m \leq n} (\max(|a|, |b|)^m)$)

In particular, given a polynomial of degree n with real coefficients $\{a_m\}_{m=1}^{n}$ and $\epsilon > 0$, there is a polynomial $\tilde{p}$ of degree n with *rational* coefficients $\{\tilde{a}_m\}_{m=0}^{n}$ such that

$$d(p_n, \tilde{p}_n) \leq c(n+1) \max_{0 \leq m \leq n} |a_m - \tilde{a}_m| < \epsilon$$

since the rationals are dense in $\mathbb{R}$. (Note, however, that as the degree increases, the coefficients generally must be approximated to increasing accuracy.) This means that given a continuous function f on $[a,b]$, and $\epsilon > 0$, we can find a polynomial with rational coefficients p_n such that $d(p_n, f) < \epsilon$. We first use Theorem 7.3.1 to find a polynomial, with possibly real coefficients, that approximates f to within $\frac{\epsilon}{2}$ and then construct a polynomial with rational coefficients that approximates the first polynomial to within $\frac{\epsilon}{2}$.

Since the set of polynomials with rational coefficients is countable, this proves

THEOREM 7.3.2. *$\mathcal{C}([a,b])$ is separable.*

We first note that it suffices to prove Theorem 7.3.1 on $[0,1]$. We can map $[0,1]$ into $[a,b]$ by $y = (b-a)x + a$ and vice-versa by $x = \frac{a-y}{a-b}$. If g is continuous on $[a,b]$, then $f(x) = g((b-a)x+a)$ is continuous on $[0,1]$. If p_n approximates f to within ϵ on $[0,1]$, then $\tilde{p}_n(y) = p_n(\frac{a-y}{a-b})$ is a polynomial that approximates $g(y)$ to within ϵ on $[a,b]$.

We give a constructive proof that uses probability.

DEFINITION 7.3.3. Recall for $n \geq m \geq 0$, the binomial coefficient n **choose** m,

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

EXAMPLE 7.3.4.

$$\binom{4}{2} = \frac{4!}{2!2!} = 6 \quad \binom{3}{0} = \frac{3!}{3!0!} = 1$$

$\binom{n}{m}$ is the number of distince subsets with m objects that can be chosen from a set of n objects. This is very important in probability.

EXAMPLE 7.3.5. We compute the probability $\mathcal{P}$ of getting an ace of diamonds in a poker hand of 5 cards chosen at random from a deck of 52 cards using

$$\mathcal{P}(event) = \frac{\text{number of outcomes in the event}}{\text{total number of possible outcomes}}$$

when all outcomes are equally likely.

The total number of 5 card poker hands is $\binom{52}{5}$. Obtaining a "good" hand amounts to choosing any 4 cards from the remaining 51 cards after getting an ace of diamonds. So, there are $\binom{51}{4}$ good hands.

$$\mathcal{P} = \frac{\binom{51}{4}}{\frac{52}{5}} = \frac{5}{52}.$$

It is straightforward to show that

$$\binom{n}{m} = \binom{n}{n-m}, \quad \binom{n}{1} = \binom{n}{n-1} = n, \quad \binom{n}{n} = \binom{n}{0} = 1.$$

There is also an important result called the binomial exansion.

THEOREM 7.3.6. *For $n \in N$ and $a, b \in \mathbb{R}$,*

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^m b^{n-m}.$$

Using this, we can derive other formulas we need. Writing

$$(x+b)^n = \sum_{m=0}^{n} \binom{n}{m} x^m b^{n-m} \tag{7.3}$$

and differentiating with respect to x gives

$$n(x+b)^{n-1} = \sum_{m=0}^{n} m \binom{n}{m} x^{m-1} b^{n-m}.$$

Setting $x = a$ and multiplying by $\frac{a}{n}$ gives

$$a(a+b)^{n-1} = \sum_{m=0}^{n} \frac{m}{n} \binom{n}{m} a^m b^{n-m}. \tag{7.4}$$

Differentiating 7.3 twice and manipulating gives

$$(1-\frac{1}{n})a^2(a+b)^{n-2} = \sum_{m=0}^{n} (\frac{m^2}{n^2} - \frac{m}{n^2}) \binom{n}{m} a^m b^{n-m}. \tag{7.5}$$

The approximating polynomials used to show Theorem 7.3.1 are constructed using the so-called **binomial polynomials**. We set $b = 1-x$ in 7.3 to get

$$1 = (x+(1-x))^n = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m}$$

Definition 7.3.7. The $n+1$ binomial polynomials of degree n are defined

$$p_{n,m}(x) = \binom{n}{m} x^m (1-x)^{n-m}, \quad m = 0, 1, ..., n.$$

Example 7.3.8.

$$p_{2,0} = \binom{2}{0} x^0 (1-x)^2 = (1-x)^2$$
$$p_{2,1} = \binom{2}{1} x^1 (1-x)^1 = 2x(1-x)$$
$$p_{2,2} = x^2$$

We observe that if $0 \le x \le 1$ is the probability of an event E, then $p_{n,m}$ is the probability that E occurs exactly m times in n independent trials.

Example 7.3.9. Suppose we toss a coin with the probability X that heads H occurse and $1-X$ that tails T occurs. The coin is unfair if $X \neq \frac{1}{2}$. Consider a sequence of tosses

$$\underbrace{HTTHTHHHHTHTHHTTHTH}_{m \text{ heads in } n \text{ tosses.}}$$

The probability of any sequence occuring is

$$X^m(1-X)^m$$

if it has m heads. There are $\binom{n}{m}$ sequences with m heads in n tosses, so this shows the claim about $p_{n,m}(x)$.

The binomial polynomials have several useful properties following from 7.4 and 7.5:

$$\sum_{m=0}^{n} p_{n,m}(x) = 1 \tag{7.6a}$$

$$\sum_{m=0}^{n} m p_{n,m}(x) = nx \tag{7.6b}$$

$$\sum_{m=0}^{n} m^2 p_{n,m}(x) = (n^2 - n)x^2 + nx \tag{7.6c}$$

We next use these polynomials to prove the Law of Large Numbers.

Suppose we have an event E with probability X of occurring. How might we determine X experimentally? If we conduct many trials N, we might expect to see the event occurred *roughly* NX times *most of the time.* This is not clearly true, however. If we toss a fair coint 100,000 times, we expect to see around 50,000 heads. We could get all tails, however, with probability $(\frac{1}{2})^{100,000}$. On the other hand, we can show the probability of getting exactly half heads in n tosses goes like

$$\mathcal{P}\ \frac{1}{\sqrt{\pi n}} \quad (n\text{large})$$

hence also tends to zero.

The Law of Large Numbers encapsulates this.

THEOREM 7.3.10. **Law of Large Numbers** *Assume event E occurs with probability X and let m denote the number of times E occurs in n trials. Let $\epsilon > 0$ and $\delta > 0$ be given. The probability that $\frac{m}{n}$ differs from X by less than δ is greater than $1-\epsilon$, i.e.,*

$$\mathcal{P}(|\frac{m}{n} - x| < \delta) > 1 - \epsilon,$$

for all n sufficiently large.

> *Note: This does not say that E occurs exactly Xn times, nor that E* must *occur roughly Xn times.*

PROOF. In terms of binomial polynomials, we want to show that given $\epsilon, \delta > 0$,

$$\sum_{\substack{0 \le m \le n \\ |\frac{m}{n} - X| < \delta}} p_{n,m}(X) > 1 - \epsilon \quad \text{for all } n \text{ large.} \tag{7.7}$$

Since lower bounds are difficult in general, we consider the complementary sum giving the probability of what we don't want:

$$\sum_{\substack{0 \le m \le n \\ |\frac{m}{n} - X| \ge \delta}} p_{n,m}(X) = 1 - \sum_{\substack{0 \le m \le n \\ |\frac{m}{n} - X| < \delta}} p_{n,m}(x),$$

that we estimate as

$$\sum_{\substack{0\le m\le n\\ |\frac{m}{n}-X|\ge\delta}} p_{n,m} \le \frac{1}{\delta^2}\sum_{\substack{0\le m\le n\\ |\frac{m}{n}-X|\ge\delta}} (\frac{m}{n}-X)^2 p_{n,m}(X)$$
$$\le \frac{1}{n^2\delta^2}\sum_{m=0}^{n}(m-nX)^2 p_{n,m}(X)$$
$$\le \frac{1}{n^2\delta^2}(\sum_{m=0}^{n} m^2 p_{n,m}(X) - 2nX\sum_{m=0}^{n} m p_{n,m}(X) + n^2X^2\sum_{m=0}^{n} p_{n,m}(X)).$$

Using 7.6a - 7.6c, the sums on the right simplify to $nX(1-X)$. Since $X(1-X)\le\frac{1}{4}$ for $0\le X\le 1$,

$$\sum_{\substack{0\le m\le n\\ |\frac{m}{n}-X|>\delta}} p_{n,m}(x) \le \frac{1}{4n\delta^2} \tag{7.8}$$

and

$$\sum_{\substack{0\le m\le n\\ |\frac{m}{n}-X|<\delta}} p_{n,m}(x) \ge 1-\frac{1}{4n\delta^2}.$$

For given $\epsilon,\delta>0$, we can insure $(4n\delta^2)^{-1}<\epsilon$ by choosing $n>\frac{1}{4\delta^2\epsilon}$. □

Proof. Of Theorem 7.3.1. We first define the approximating polynomial, named after the person who made this proof.

Definition 7.3.11. We partition $[0,1]$ by a uniform mesh with $n+1$ nodes, $x_m=\frac{m}{n}, m=0,1,...,n$. The **Bernstein polynomial** of order n for f on $[0,1]$ is

$$B_n(f,x)=B_n(x)=\sum_{m=0}^{n} f(x_m)p_{n,m}(x).$$

Note that $\deg(B_n)\le n$.

The reason that B_n approximates f is intuitive

$$B_n(x)=\sum_{x_m\approx x} f(x_m)p_{n,m}(x)+\sum_{\substack{|x_m-x|\\ \text{large}}} f(x_m)p_{n,m}(x).$$

The first sum converges to f as n increases becasue we can find $\frac{m}{n}$ arbitrarily close to x while the second sum goes to zero by the Law of Large Numbers.

Example 7.3.12. Consider x^2 on $[0,1]$ with $n\ge 2$,

$$B_n(x)=\sum_{m=0}^{n}(\frac{m}{n})^2 p_{n,m}(x).$$

Using the identities 7.6a - 7.6c,

$$B_n(x)=(1-\frac{1}{n})x^2+\frac{1}{n}x.$$

Note that $B_n(x^2,x)\ne x^2$ and the error is

$$|x^2-B_n(x)|=\frac{1}{n}x(1-x)$$

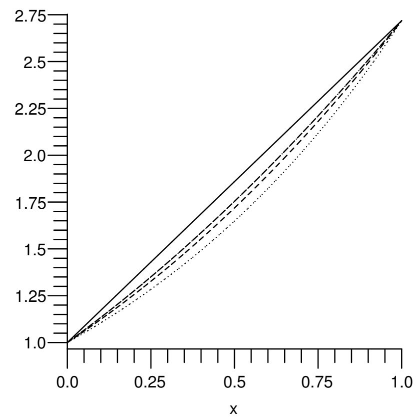


FIGURE 7.5. $B_1(x) > B_2(x) > B_3(x) > e^x$ on $(0,1)$

which tends to zero like $\frac{1}{n}$ on $(0,1]$. This contrasts with interpolating polynomials and Taylor polynomials, which both have the property that if f is a polynomial then $p_n = f$ for $n \geq \deg(f)$.

EXAMPLE 7.3.13. For e^x on $(0,1]$ (see Figure 7.5),

$$\begin{aligned}
B_1(x) &= (1-x) + ex \\
B_2(x) &= (1-x)^2 + 2e^{\frac{1}{2}}x(1-x) + ex^2 \\
B_3(x) &= (1-x)^3 + 3e^{\frac{1}{3}}x(1-x)^2 + 3e^{\frac{2}{3}}x^2(1-x) + ex^3 \\
&\vdots
\end{aligned}$$

We prove that given $\epsilon > 0$, there is an n such that

$$\sup_{0 \leq x \leq 1} |f(x) - B_n(x)| < \epsilon.$$

Using 7.6a, we write

$$\begin{aligned}
f(x) - B_n(x) &= \sum_{m=0}^{n} f(x)p_{n,m}(x) - \sum_{m=0}^{n} f(x_m)p_{n,m}(x) \\
&= \sum_{m=0}^{n} (f(x) - f(x_m))p_{n,m}(x).
\end{aligned}$$

We expect that we can make $f(x) - f(x_m)$ small when x is close to x_m by continuity. For $\delta > 0$, we write

(7.9)

$$f(x) - B_n(x) = \sum_{\substack{0 \leq m \leq n \\ |x - x_m| < \delta}} (f(x) - f(x_m))p_{n,m}(x) + \sum_{\substack{0 \leq m \leq n \\ |x - s_m| \geq \delta}} (f(x) - f(x_m))p_{n,m}(x).$$

Theorem 6.4.2 implies f is uniformly continuous on $[0,1]$. Given $\epsilon > 0$, there is a $\delta > 0$ such that

$$|f(x) - f(x_m)| < \frac{\epsilon}{2}$$

for all x, x_m in $[0,1]$ with $|x - x_m| \leq \delta$. Given δ, by the way, we can find x_m such that $|x - x_m| \leq \delta$ for all sufficiently large n, since the rationals are dense in $[0,1]$.

Thus,

$$|\sum_{\substack{0 \leq m \leq n \\ |x - x_m| < \delta}} (f(x) - f(x_m)) p_{n,m}(x)| \leq \sum_{\substack{0 \leq m \leq n \\ |x - x_m| < \delta}} |f(x) - f(x_m)| p_{n,m}(x)$$

$$\leq \frac{\epsilon}{2} \sum_{0 \leq m \leq n} p_{n,m}(x) = \frac{\epsilon}{2}.$$

Now the second sum on the right in 7.9 is bounded after we realize that Theorem 6.5.8 implies $|f|$ is bounded on $[0,1]$ by some constant M. Hence, 7.8 implies

$$|\sum_{\substack{0 \leq m \leq n \\ |x - x_m| \geq \delta}} (f(x) - f(x_m)) p_{n,m}(x)| \leq 2M \sum_{\substack{0 \leq m \leq n \\ |x - x_m| \geq \delta}} p_{n,m}(x)$$

$$\leq \frac{M}{2n\delta^2}.$$

Given δ from the first estimate, we can force $\frac{M}{2n\delta^2} < \frac{\epsilon}{2}$ by taking n sufficiently large. □

7.4. Compact Sets in $\mathcal{C}([a,b])$

Example 7.2.1 shows that $\mathcal{C}([a,b])$ is not compact.

EXAMPLE 7.4.1. The sequence $\{x^n \mid n \in \mathbb{N}\}$ does not have a convergent subsequence in $\mathcal{C}([a,b])$, i.e., there is no subsequence that converges uniformly.

This raises the issue of describing the compact subsets of $\mathcal{C}([a,b])$.

By Theorem 3.2.4, if $K \subset \mathcal{C}([a,b])$ is compact, then K is closed and bounded. Closed is rather obvious: K is closed if for any sequence of functions $\{f_n\}$ in K that conveges to f in the metric of $\mathcal{C}([a,b])$ (uniformly), we have $f \in K$.

EXAMPLE 7.4.2. Let

$$F = \{f \in \mathcal{C}([a,b]) \mid \sup_{a \leq x \leq b} |f(x)| < 1\}.$$

F is *not* closed since, for example,

$$\{1 - \frac{1}{n} \mid n = 1, 2, 3, ...\}$$

is a sequence of functions in F that converges uniformly to $f(x) \equiv 1$, which is not in F.

EXAMPLE 7.4.3. Let

$$F = \{f \in \mathcal{C}([a,b]) \mid \sup_{a \leq x \leq b} |f(x)| \leq 1\}.$$

We show F is closed. Choose a sequence $\{f_n\}$ in F that converges to f in $\mathcal{C}([a,b])$. We show $f \in F$, this is, $\sup_{a \leq x \leq b} |f(x)| \leq 1\}$.

There is an $\epsilon > 0$ and an $s \in [a,b]$ such that $|f(x)| > 1+\epsilon$. Because f is continuous, there is a $\delta > 0$ such that $|f(y)| > 1+\frac{\epsilon}{2}$ for $y \in (x-\delta, x+\delta) \cap [a,b]$. But, for $y \in (x-\delta, x+\delta) \cap [a,b]$ and *all* n,

$$|f(y) - f_n(y)| \geq ||f(y)| - |f_n(y)|| \geq |1 + \frac{\epsilon}{2} - 1| \geq \frac{\epsilon}{2},$$

which contradicts $\sup_{a \leq y \leq b} |f_n(y) - f(y)| \to 0$ as $n \to \infty$.

$K \subset \mathcal{C}([a,b])$ is bounded means there is a function $g \in \mathcal{C}([a,b])$ and an M such that

$$d(f,g) = \sup_{a \leq x \leq b} |f(x) - g(x)| \leq M$$

for all $f \in K$. Since such a g is itself bounded on $[a,b]$, we see there is an M such that

$$\sup_{a \leq x \leq b} |f(x)| \leq M \quad \text{for all } f \in K.$$

This motivates

Definition 7.4.4. Let $(\mathbf{X}, d)$ be a metric space, $A \subset \mathbf{X}$, and F a set of functions from A into $\mathbb{R}^n$ with the usual metric. F is **uniformly bounded** on A if there is an M such that

$$\sup_{x \in A} \| f(x) \| \leq M \quad \text{for all } f \in F.$$

We have shown that if $F \subset \mathcal{C}([a,b])$ is bounded, then F is uniformly bounded. The converse is obviously true.

Theorem 7.4.5. *Let $F \subset \mathcal{C}([a,b])$ be a set of continuous functions on $[a,b]$. Then, F is uniformly bounded on $[a,b]$ if and only if F is a bounded subset of $\mathcal{C}([a,b])$.*

Example 7.4.6. The set $F = \{x^n \mid n = 1,2,3...\}$ is bounded on $[0,1]$, but not on $[0,2]$. The qualification of boundedness does depend on $[a,b]$. On the other hand, we do *not* expect that being merely closed and bounded guarantees compactness. In fact, $\{x^n \mid n = 1,2,3,...\}$ is closed and bounded on $[0,1]$, but is not compact. We need something more.

Example 7.4.7. Consider a sequence $\{g_n\}$ in $\mathcal{C}$ that converges in $\mathcal{C}([a,b])$ to $g \in \mathcal{C}([a,b])$. The set

$$K = \{g_n \mid n = 1,2,3,...\} \cup \{g\}$$

is compact. If $\{f_n\}$ is a sequence in K, then if $\{f_n\}$ contains a function in K repeated infinitely often, then it has a subsequence that converges to an element of K. Otherwise, infinitely many of the functions $\{g_n\}$ are contained in $\{f_n\}$, and $\{f_n\}$ contains a subsequence that converges to g.

Since g is uniformly continuous on $[a,b]$, given $\epsilon > 0$, there is a δ_0 such that

$$|g(x) - g(y)| < \frac{\epsilon}{3} \quad \text{for } x, y \in [a,b], |x-y| < \delta_0.$$

Since $g_n \to g$ uniformly, there is an N such that

$$\sup_{a \leq x \leq b} |g(x) - g(y)| < \frac{\epsilon}{3} \quad \text{for } n \geq N.$$

Hence, for $n \geq N$, and $x, y \in [a,b]$ with $|x-y| < \delta_0$,

$$|g_n(x) - g_n(y)| \leq |g_n(x) - g(x)| + |g(x) - g(y)| + |g(y) - g_n(y)|$$
$$\frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

Now the functions $g_1, ..., g_{N-1}$ are also uniformly continuous. Hence, there are $\delta_1, ..., \delta_{N-1}$ such that $|g_m(x) - g_m(y)| < \epsilon$ for $x, y \in [a,b], |x-y| < \delta_m$, for $m = 1, 2, ..., N-1$.

Setting $\delta = \min\{\delta_0, \delta_1, ..., \delta_{N-1}\}$, we see that for $\epsilon > 0$ there is a $\delta > 0$ such that *for all* n,

$$|g_n(x) - g_n(y)| < \epsilon \quad \text{for } x, y \in [a,b], |x-y| < \delta.$$

The functions in K are uniformly continuous with the same ϵ and δ, sort of "uniformly uniformly continuous".

This motivates

DEFINITION 7.4.8. Let $(\mathbf{X}, d_x)$ and $(\mathbf{Y}, d_y)$ be metric spaces, $A \subset \mathbf{X}$, and F a set of functions from A into $\mathbf{Y}$. F is **equicontinuous on** A if for every $\epsilon > 0$ there is a $\delta > 0$ such that *for all* $f \in F$,

$$d_y(f(x, f(y)) < \epsilon \quad \text{for all } x, y \in A \text{ with } d_x(x,y) < \delta.$$

EXAMPLE 7.4.9. The functions in Example 7.4.7 are equicontinuous.

EXAMPLE 7.4.10. For fixed $L > 0$, let F be the set of functions

$$F = \{f : [a,b] \to \mathbb{R} \mid |f(x) - f(y)| \leq L|x-y| \text{ for all } x, y \in [a,b]\},$$

that is, F is the set of Lipschitz continuous functions on $[a,b]$ with constant L. We also say that F is uniformly Lipschitz continuous with constant L. Then, F is equicontinuous.

EXERCISE 7.4.11. Show F as above is equicontinuous.

EXAMPLE 7.4.12. Consider

$$F = \{x^n \mid n = 1, 2, 3, ...\}$$

on $0, 1]$. Take $\epsilon = \frac{1}{2}$ and choose $0 < \delta < 1$. Set $x = 1 - \frac{\delta}{2}$. Since $\lim_{n\to\infty} x^n = 0$, there is an N such that $|1 - x^n| > \frac{1}{2} = \epsilon$ for $n \geq N$.

Hence, $|1-x| < \delta$ while $|1 - x^n| > \frac{1}{2} = \epsilon$ for $n \geq N$. Since δ is arbitrary, F cannot be equicontinuous. The condition for equicontinuity fails at 1.

We now prove the important characterization of compactness in $\mathcal{C}([a,b])$.

THEOREM 7.4.13. **Arzela-Ascoli Theorem** *Let K be a closed subset of $\mathcal{C}([a,b])$. Then K is compact if and only if K is uniformly bounded and equicontinuous.*

PROOF. First note that since K is closed and $\mathcal{C}([a,b])$ is complete (Theorem 7.2.8), K is complete. By Theorem 4.4.3, K is compact if and only if it is totally bounded. We will prove that K is uniformly bounded and equicontinuous if and only if it is totally bounded.

Suppose first K is totally bounded. This means in particular that it is bounded and therefore uniformly bounded by Theorem 7.4.5. To show K is equicontinuous, for $\epsilon > 0$, let $\{f_1, ..., f_n\}$ be a set of functions in K such that

$$K \subset \bigcup_{m=1}^{n} N_\epsilon(f_m).$$

This means that for any $f \in K$, there is an m, $1 \le m \le n$, such that

$$\sup_{a \le x \le b} |f(x) - f_m(x)| < \epsilon.$$

Each of the f_m is uniformly continuous on $[a, b]$ and since there is a finite number of $\{f_1, ..., f_n\}$, there is a $\delta > 0$ such that

$$|f_m(x) - f_m(y)| < \epsilon \quad \text{for } 1 \le m \le n, x, y \in [a, b], |x - y| < \delta.$$

Choosing $f \in K$, we choose f_m as above and write

$$|f(x) - f(y)| \le |f(x) - f_m(x)| + |f_m(x) - f_m(y)| + |f_m(y) - f(y)|$$

and with δ chosen as above,

$$|f(x) - f(y)| < 3\epsilon,$$

for all $x, y \in [a, b]$ with $|x - y| < \delta$.

Since f was chosen arbitrarily, K is equicontinuous.

Now, we show that if K is uniformly bounded and equicontinuous, then it is totally bounded. So, given any $\epsilon > 0$, we construct a finite set of functions F such that

$$K \subset \bigcup_{f \in F} N_\epsilon(f),$$

which means that given $g \in K$ there is an $f \in F$ such that

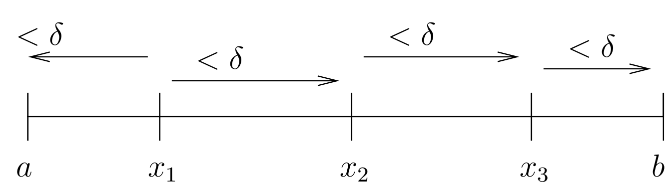
$$\sup_{a \le x \le b} |f(x) - g(x)| < \epsilon.$$

Note that F does not have to be contained in K, just in $\mathcal{C}([a, b])$.

Given $\epsilon > 0$, by equicontinuity, there is a $\delta > 0$ such that

$$|g(x) - g(y)| < \frac{\epsilon}{5} \quad \text{for all } g \in K, x, y \in [a, b] \text{ with } |x - y| < \delta.$$

Using this δ, we creat a mesh on $[a, b]$ with nodes $\{x_1, ..., x_n\}$ such that every point in $[a, b]$ is within δ of one of these points.



Note: we are using the fact that $[a, b]$ is totally bounded!

Now choose $M \in \mathbb{N}$ such that $|g(x)| \leq M$ for all $a \leq x \leq b$ and all $g \in K$. Choose $m \in \mathbb{N}$ with $\frac{1}{m} < \frac{\epsilon}{5}$ and partition $[-M, M]$ into $2Mm$ congruent intervals with nodes

$$-M = y_0 < y_1 < ... < y_{2Mm} = M.$$

(See figure 7.6.)

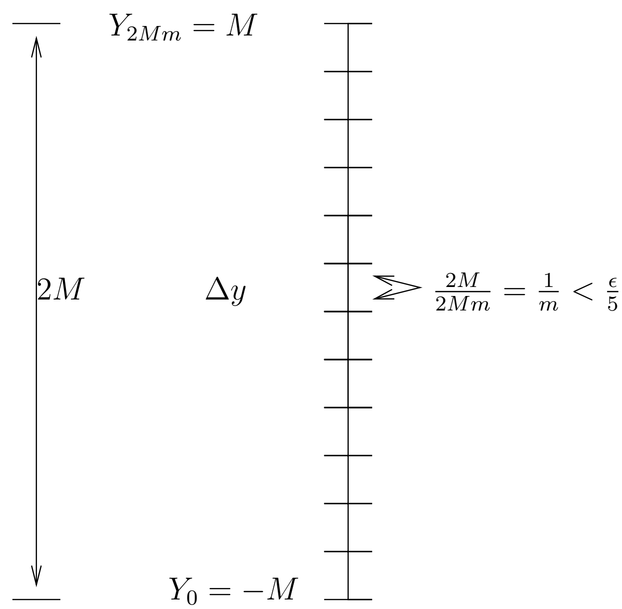
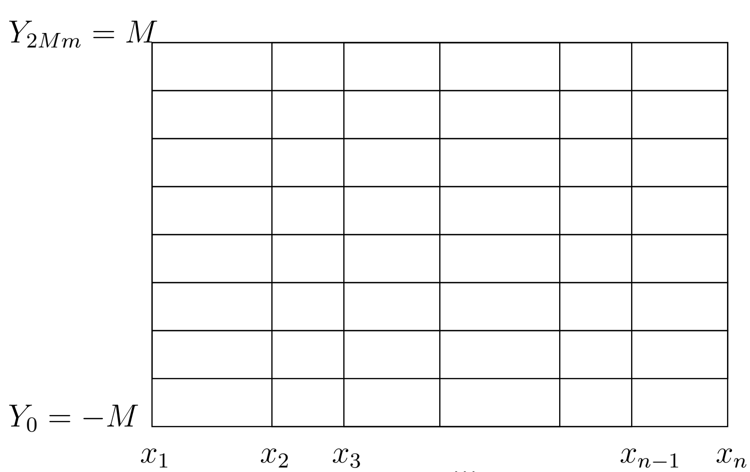


FIGURE 7.6

We now have a grid of points $\{(x_i, y_j) \mid 1 \leq i \leq n, 0 \leq j \leq 2Mm\}$ for the rectangle $[a,b] \times [-M, M]$.



Let F be the set of continuous functions on $[a,b]$ that are piecewise linear whose "corner points" occur at points on the grid (see Figure 7.7).

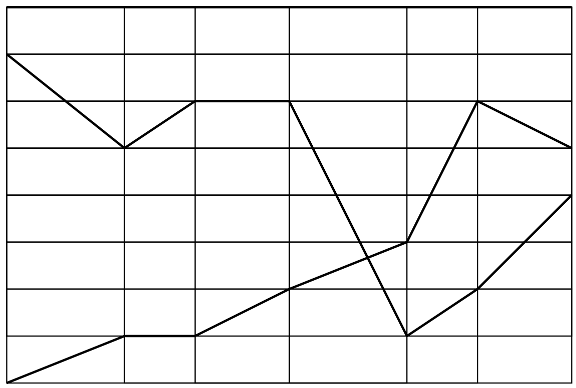


FIGURE 7.7. 3 examples

F has a finite number $((2Mm)^n)$ of elements. Choose $g \in K$. There is at least one $f \in F$ such that

$$|g(x_j) - f(x_j)| < \frac{\epsilon}{5} \quad j = 1, 2, ..., n.$$

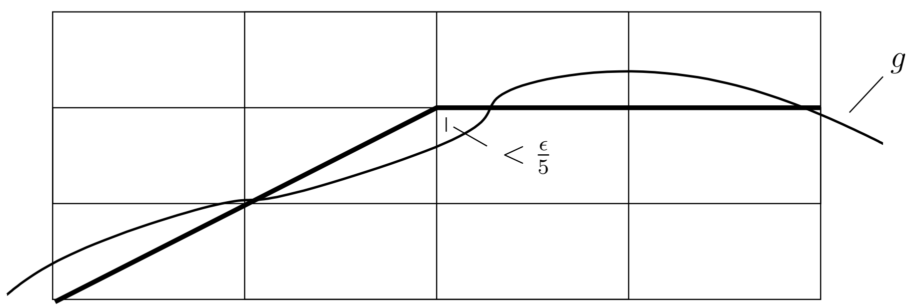


FIGURE 7.8

We want to show $g(x)$ is close to $f(x)$ for all $a \le x \le b$ (see Figure 7.8). Choose $a \le x \le b$. Now $x_j \le x \le x_{j+1}$ for some $1 \le j \le n$. By the equicontinuity of K and choice of δ, we know $|g(y) - g(x_j)| < \frac{\epsilon}{5}$ for $x_j \le y \le x_{j+1}$. It follows that

$$\begin{aligned} |f(x_{j+1}) - f(x_j)| &\le |f(x_{j+1}) - g(x_{j+1})| + |g(x_{j+1}) - g(x_j)| + |g(x_j) - f(x_j)| \\ &< \frac{\epsilon}{5} + \frac{\epsilon}{5} + \frac{\epsilon}{5} = \frac{3\epsilon}{5}. \end{aligned}$$

Since f is linear on $[x_j, x_{j+1}]$,

$$|f(y) - g(x_j)| < \frac{3\epsilon}{5}, \quad x_j \le y \le x_{j+1}.$$

So,

$$\begin{aligned}|g(x) - f(x)| &\leq |g(x) - g(x_j)| + |g(x_j) - f(x_j)| + |f(x_j) - f(x)| \\ &< \epsilon.\end{aligned}$$

F is an ϵ-net for K. □

EXAMPLE 7.4.14. We don't have time for details, but a classic application of the Arzela-Ascoli Theorem is to show that the forward Euler approximation

$$\begin{aligned}&Y_0 = y_0 \\ &Y_n = Y_{n-1} + \Delta t f(Y_{n-1}), \quad n = 1, 2, ..., N,\end{aligned}$$

where $\Delta t = \frac{T}{N}$, for the initial value problem

$$\begin{cases} y' = f(y), & 0 \leq t \leq T \\ y(0) = y_0 \end{cases}$$

converges to y for $0 \leq t \leq T$ if f is continuous.