Mathematical Methods of Engineering Analysis

Erhan Çinlar Robert J. Vanderbei

February 2, 2000

Contents

Sets and Functions

This introductory chapter is devoted to general notions regarding sets, functions, sequences, and series. The aim is to introduce and review the basic notation, terminology, conventions, and elementary facts.

1 Sets

A *set* is a collection of some objects. Given a set, the objects that form it are called its *elements*. Given a set A, we write $x \in A$ to mean that x is an element of A. To say that $x \in A$, we also use phrases like x is in A, x is a member of A, x belongs to A, and A includes x.

To specify a set, one can either write down all its elements inside curly brackets (if this is feasible), or indicate the properties that distinguish its elements. For example, $A = \{a, b, c\}$ is the set whose elements are a, b, and c, and $B = \{x : x > 2.7\}$ is the set of all numbers exceeding 2.7. The following are some special sets:

$\emptyset$: The *empty set*. It has no elements.

$\mathbb{N} = \{1, 2, 3, \ldots\}$: Set of *natural numbers*.

$\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$: Set of *integers*.

$\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$: Set of *positive integers*.

$\mathbb{Q} = \{\frac{m}{n} : m \in \mathbb{Z}, n \in \mathbb{N}\}$: Set of *rationals*.

$\mathbb{R} = (-\infty, \infty) = \{x : -\infty < x < +\infty\}$: Set of *reals*.

$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$: *Closed intervals*.

$(a, b) = \{x \in \mathbb{R} : a < x < b\}$: *Open intervals*.

$\mathbb{R}_+ = [0, \infty) = \{x \in \mathbb{R} : x \geq 0\}$: Set of *positive reals*.

Subsets

A set A is said to be a *subset* of a set B if every element of A is an element of B. We write $A \subset B$ or $B \supset A$ to indicate it and use expressions like A is contained in B, B contains A, to the same effect. The sets A and B are the same, and then we write $A = B$, if and only if $A \subset B$ and $A \supset B$. We write $A \neq B$ when A and B are not the same. The set A is called a *proper subset* of B if A is a subset of B and A and B are not the same.

The empty set is a subset of every set. This is a point of logic: let A be a set; the claim is that $\emptyset \subset A$, that is, that every element of $\emptyset$ is also an element of A, or equivalently, there is no element of $\emptyset$ that does not belong to A. But the last is obviously true simply because $\emptyset$ has no elements.

Set Operations

Let A and B be sets. Their *union*, denoted by $A \cup B$, is the set consisting of all elements that belong to either A or B (or both). Their *intersection*, denoted by $A \cap B$, is the set of all elements that belong to both A and B. The *complement* of A in B, denoted by $B \setminus A$, is the set of all elements of B that are not in A. Sometimes, when B is understood from context, $B \setminus A$ is also called the complement of A and is denoted by A^c. Regarding these operations, the following hold:

Commutative laws:

$$\begin{aligned} A \cup B &= B \cup A, \\ A \cap B &= B \cap A. \end{aligned}$$

Associative laws:

$$\begin{aligned} (A \cup B) \cup C &= A \cup (B \cup C), \\ (A \cap B) \cap C &= A \cap (B \cap C). \end{aligned}$$

Distributive laws:

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). \end{aligned}$$

The associative laws show that $A \cup B \cup C$ and $A \cap B \cap C$ have unambiguous meanings.

Definitions of unions and intersections can be extended to arbitrary collections of sets. Let I be a set. For each $i \in I$, let A_i be a set. The *union* of the sets A_i, $i \in I$, is the set A such that $x \in A$ if and only if $x \in A_i$ for some i in I. The following notations are used to denote the union and intersection respectively:

$$\bigcup_{i \in I} A_i, \qquad \bigcap_{i \in I} A_i.$$

When $I = \mathbb{N} = \{1, 2, 3, \ldots\}$, it is customary to write

$$\bigcup_{i=1}^{\infty} A_i, \qquad \bigcap_{i=1}^{\infty} A_i.$$

All of these notations follow the conventions for sums of numbers. For instance,

$$\bigcup_{i=1}^{n} A_i = A_1 \cup \cdots \cup A_n, \qquad \bigcap_{i=5}^{13} A_i = A_5 \cap A_6 \cap \cdots \cap A_{13}$$

stand, respectively, for the union over $I = \{1, \ldots, n\}$ and the intersection over $I = \{5, 6, \ldots, 13\}$.

Disjoint Sets

Two sets are said to be *disjoint* if their intersection is empty; that is, if they have no elements in common. A collection $\{A_i : i \in I\}$ of sets is said to be *disjointed* if A_i and A_j are disjoint for all i and j in I with $i \neq j$.

Products of Sets

Let A and B be sets. Their *product*, denoted by $A \times B$, is the set of all pairs (x, y) with x in A and y in B. It is also called the *rectangle* with sides A and B.

If $A_1, \ldots, A_n$ are sets, then their product $A_1 \times \cdots \times A_n$ is the set of all n-tuples $(x_1, \ldots, x_n)$ where $x_1 \in A_1, \ldots, x_n \in A_n$. This product is called, variously, a rectangle, or a box, or an n-dimensional box. If $A_1 = \cdots = A_n = A$, then $A_1 \times \cdots \times A_n$ is denoted by A^n. Thus, $\mathbb{R}^2$ is the plane, $\mathbb{R}^3$ is the three-dimensional space, $\mathbb{R}^2_+$ is the positive quadrant of the plane, etc.

Exercises:

1.1 Let E be a set. Show the following for subsets A, B, C, and A_i of E. Here, all complements are with respect to E; for instance, $A^c = E \setminus A$.

1. $(A^c)^c = A$
2. $B \setminus A = B \cap A^c$
3. $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C)$
4. $(A \cup B)^c = A^c \cap B^c$
5. $(A \cap B)^c = A^c \cup B^c$
6. $(\bigcup_{i \in I} A_i)^c = \bigcap_{i \in I} A_i^c$
7. $(\bigcap_{i \in I} A_i)^c = \bigcup_{i \in I} A_i^c$

1.2 Let a and b be real numbers with $a < b$. Find

$$\bigcup_{n=1}^{\infty} [a + \frac{1}{n}, b - \frac{1}{n}], \quad \bigcap_{n=1}^{\infty} [a - \frac{1}{n}, b + \frac{1}{n}]$$

1.3 Describe the following sets in words and pictures:

1. $A = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}$
2. $B = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$
3. $C = B \setminus A$
4. $D = C \times B$
5. $S = C \times C$

1.4 Let A_n be the set of points $(x, y) \in \mathbb{R}^2$ lying on the curve $y = 1/x^n$, $0 < x < \infty$. What is $\bigcap_{n \geq 1} A_n$?

2 Functions and Sequences

Let E and F be sets. With each element x of E, let there be associated a unique element $f(x)$ of F. Then f is called a *function* from E into F, and f is said to *map* E into F. We write $f : E \mapsto F$ to indicate it.

Let f be a function from E into F. For x in E, the point $f(x)$ in F is called the *image* of x or the value of f at x. Similarly, for $A \subset E$, the set

$$\{y \in F : y = f(x) \text{ for some } x \in A\}$$

is called the *image* of A. In particular, the image of E is called the *range* of f. Moving in the opposite direction, for $B \subset F$,

$$f^{-1}(B) = \{x \in E : f(x) \in B\} \tag{2.1}$$

is called the *inverse image* of B under f. Obviously, the inverse of F is E.

Terms like mapping, operator, transformation are synonyms for the term “function” with varying shades of meaning depending on the context and on the sets E and F. We shall become familiar with them in time. Sometimes, we write $x \mapsto f(x)$ to indicate the mapping f; for instance, the mapping $x \mapsto x^3 + 5$ from $\mathbb{R}$ into $\mathbb{R}$ is the function $f : \mathbb{R} \mapsto \mathbb{R}$ defined by $f(x) = x^3 + 5$.

Injections, Surjections, Bijections

Let f be a function from E into F. It is called an *injection*, or is said to be *injective*, or is said to be *one-to-one*, if distinct points have distinct images (that is, if $x \neq y$ implies $f(x) \neq f(y)$). It is called a *surjection*, or is said to be *surjective*, if its range is F, in which case f is said to be from E *onto* F. It is called a *bijection*, or is said to be *bijective*, if it is both injective and surjective.

These terms are relative to E and F. For examples, $x \mapsto e^x$ is an injection from $\mathbb{R}$ into $\mathbb{R}$, but is a bijection from $\mathbb{R}$ into $(0, \infty)$. The function $x \mapsto \sin x$ from $\mathbb{R}$ into $\mathbb{R}$ is neither injective nor surjective, but it is a surjection from $\mathbb{R}$ onto $[-1, 1]$.

Sequences

A *sequence* is a function from $\mathbb{N}$ into some set. If f is a sequence, it is customary to denote $f(n)$ by something like x_n and write (x_n) or $(x_1, x_2, \ldots)$ for the sequence (instead of f). Then, the x_n are called the *terms* of the sequence. For instance, $(1, 3, 4, 7, 11, \ldots)$ is a sequence whose first, second, etc. terms are $x_1 = 1$, $x_2 = 3$, ...

If A is a set and every term of the sequence (x_n) belongs to A, then (x_n) is said to be a sequence in A or a sequence of elements of A, and we write $(x_n) \subset A$ to indicate this.

A sequence (x_n) is said to be a *subsequence* of (y_n) if there exist integers $1 \leq k_1 < k_2 < k_3 < \cdots$ such that

$$x_n = y_{k_n}$$

for each n. For instance, the sequence $(1, 1/2, 1/4, 1/8, \ldots)$ is a subsequence of $(1, 1/2, 1/3, 1/4, 1/5, \ldots)$.

Exercises:

2.1 Let f be a mapping from E into F. Show that

1. $f^{-1}(\emptyset) = \emptyset$,
2. $f^{-1}(F) = E$,
3. $f^{-1}(B \setminus C) = f^{-1}(B) \setminus f^{-1}(C)$,
4. $f^{-1}(\bigcup_{i \in I} B_i) = \bigcup_{i \in I} f^{-1}(B_i)$,
5. $f^{-1}(\bigcap_{i \in I} B_i) = \bigcap_{i \in I} f^{-1}(B_i)$,

for all subsets B, C, B_i of F.

2.2 Show that $x \mapsto e^{-x}$ is a bijection from $\mathbb{R}_+$ onto $(0, 1]$. Show that $x \mapsto \log x$ is a bijection from $(0, \infty)$ onto $\mathbb{R}$. (Incidentally, $\log x$ is the logarithm of x to the base e, which is nowadays called the natural logarithm. We call it the logarithm. Let others call their logarithms "unnatural.")

2.3 Let f be defined by the arrows below:

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ 0 & -1 & 1 & -2 & 2 & -3 & 3 & \cdots \end{array}$$

This defines a bijection from $\mathbb{N}$ onto $\mathbb{Z}$. Using this, construct a bijection from $\mathbb{Z}$ onto $\mathbb{N}$.

2.4 Let $f : \mathbb{N} \times \mathbb{N} \mapsto \mathbb{N}$ be defined by the table below where $f(i, j)$ is the entry in the i^{th} row and the j^{th} column. Use this and the preceding exercise to construct a bijection from $\mathbb{Z} \times \mathbb{Z}$ onto $\mathbb{N}$.

$i \backslash j$	1	2	3	4	5	6	$\cdots$
1	1	3	6	10	15	21	
2	2	5	9	14	20		
3	4	8	13	19			
4	7	12	18				
5	11	17					
6	16						
$\vdots$							

2.5 *Functional Inverses.* Let f be a bijection from E onto F. Then, for each y in F there is a unique x in E such that $f(x) = y$. In other words, in the notation of (2.1), $f^{-1}(\{y\}) = \{x\}$ for each y in F and some unique x in E. In this case, we drop some brackets and write $f^{-1}(y) = x$. The resulting function f^{-1} is a bijection from F onto E; it is called the functional inverse of f. This particular usage should not be confused with the general notation of f^{-1}. (Note that (2.1) defines a function f^{-1} form $\mathcal{F}$ into $\mathcal{E}$, where $\mathcal{F}$ is the collection of all subsets of F and $\mathcal{E}$ is the collection of all subsets of E.)

3 Countability

Two sets A and B are said to have the same cardinality, and then we write $A \sim B$, if there exists a bijection from A onto B. Obviously, having the same cardinality is an equivalence relation; it is

1. reflexive: $A \sim A$,
2. symmetric: $A \sim B \Rightarrow B \sim A$,
3. transitive: $A \sim B$ and $B \sim C \Rightarrow A \sim C$.

A set is said to be *finite* if it is empty or has the same cardinality as $\{1, 2, \ldots, n\}$ for some n in $\mathbb{N}$; in the former case it has 0 elements, in the latter exactly n. It is said to be *countable* if it is finite or has the same cardinality as $\mathbb{N}$; in the latter case it is said to have a countable infinity of elements.

In particular, $\mathbb{N}$ is countable. So are $\mathbb{Z}$, $\mathbb{N} \times \mathbb{N}$ in view of exercises 2.3 and 2.4. Note that an infinite set can have the same cardinality as one of its proper subsets. For instance, $\mathbb{Z} \sim \mathbb{N}$, $\mathbb{R}_+ \sim (0, 1]$, $\mathbb{R} \sim \mathbb{R}_+ \sim (0, 1)$; see exercise 2.2 for the latter. Incidentally, $\mathbb{R}_+$, $\mathbb{R}$, etc. are uncountable, as we shall show shortly.

A set is countable if and only if it can be injected into $\mathbb{N}$, or equivalently, if and only if there is a surjection from $\mathbb{N}$ onto it. Thus, a set A is countable if and only if there is a sequence (x_n) whose range is A. The following lemma follows easily from these remarks.

3.1 LEMMA. *If A can be injected into B and B is countable, then A is countable. If A is countable and there is a surjection from A onto B, then B is countable.*

3.2 THEOREM. *The product of two countable sets is countable.*

PROOF. Let A and B be countable. If one of them is empty, then $A \times B$ is empty and there is nothing to prove. Suppose that neither is empty. Then, there exist injections $f : A \mapsto \mathbb{N}$ and $g : B \mapsto \mathbb{N}$. For each pair (x, y) in $A \times B$, let $h(x, y) = (f(x), g(y))$; then h is an injection from $A \times B$ into $\mathbb{N} \times \mathbb{N}$. Since $\mathbb{N} \times \mathbb{N}$ is countable (see Exercise (2.4)), this implies via the preceding lemma that $A \times B$ is countable □

3.3 COROLLARY. *The set of all rational numbers is countable.*

PROOF. Recall that the set $\mathbb{Q}$ of all rationals consists of ratios m/n with $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. Thus, $f(m, n) = m/n$ defines a surjection from $\mathbb{Z} \times \mathbb{N}$ onto $\mathbb{Q}$. Since $\mathbb{Z}$ and $\mathbb{N}$ are countable, so is $\mathbb{Z} \times \mathbb{N}$ by the preceding theorem. Hence, $\mathbb{Q}$ is countable by Lemma 3.1. □

3.4 THEOREM. *The union of a countable collection of countable sets is countable.*

PROOF. Let I be a countable set, and let A_i be a countable set for each i in I. The claim is that $A = \bigcup_{i \in I} A_i$ is countable. Now, there is a surjection $f_i : \mathbb{N} \mapsto A_i$ for each i, and there is a surjection $g : \mathbb{N} \mapsto I$; these follow from the countability of I and the A_i. Note that, then, $h(m, n) = f_{g(m)}(n)$ defines a surjection h from $\mathbb{N} \times \mathbb{N}$ onto A. Since $\mathbb{N} \times \mathbb{N}$ is countable, this implies via Lemma 3.1 that A is countable. □

The following theorem exhibits an uncountable set. As a corollary, we show that $\mathbb{R}$ is uncountable.

3.5 THEOREM. *Let E be the set of all sequences whose terms are the digits 0 and 1. Then, E is uncountable.*

PROOF. Let A be a countable subset of E. Let $x_1, x_2, \ldots$ be an enumeration of the elements of A, that is, A is the range of (x_n). Note that each x_n is a sequence of zeros and ones, say $x_n = (x_{n,1}, x_{n,2}, \ldots)$ where each term $x_{n,m}$ is either 0 or 1. We define a new sequence $y = (y_n)$ by letting $y_n = 1 - x_{n,n}$. The sequence y differs from every one of the sequences $x_1, x_2, \ldots$ in at least one position. Thus, y is not in A but is in E.

We have shown that if $A \subset E$ and is countable, then there is a $y \in E$ such that $y \notin A$. If E were countable, the preceding would hold for $A = E$, which would be absurd. Hence, E must be uncountable. $\square$

3.6 COROLLARY. *The set of all real numbers is uncountable.*

PROOF. It is enough to show that the interval $[0,1)$ is uncountable. For each $x \in [0,1)$, let $0.x_1x_2x_3\cdots$ be the binary expansion of x (in case x is dyadic, say $x = k/2^n$ for some k and n in $\mathbb{N}$, there are two such possible binary expansions, in which case we take the expansion with infinitely many zeros), and we identify the binary expansion with the sequence $(x_1, x_2, \ldots)$ in the set E of the preceding theorem. Thus, to each x in $[0,1)$ there corresponds a unique element $f(x)$ of E. In fact, f is a surjection onto the set $E \setminus D$ where D denotes the set of all sequences of zeros and ones that are eventually all ones. It is easy to show that D is countable and hence that $E \setminus D$ is uncountable. From this it follows that $[0,1)$ is uncountable. $\square$

Exercises:

3.1 *Dyadics.* A number is said to be dyadic if it has the form $k/2^n$ for some integers k and n in $\mathbb{Z}_+$. Show that the set of all dyadic numbers is countable. Of course, every dyadic number is rational.

3.2 Let D denote the set of all sequences of zeros and ones that are eventually all ones. Show that D is countable.

3.3 Suppose that A is uncountable and that B is countable. Show that $A \setminus B$ is uncountable.

3.4 Let x be a real number. For each $n \in \mathbb{Z}_+$, let x_n be the smallest dyadic number of the form $k/2^n$ that exceeds x. Show that $x_0 \geq x_1 \geq x_2 \geq \cdots$ and that $x_n > x$ for each n. Show that, for every $\epsilon > 0$, there is an n_ϵ such that $x_n - x < \epsilon$ for all $n \geq n_\epsilon$.

4 On the Real Line

The object is to review some facts and establish some terminology regarding the set $\mathbb{R}$ of all real numbers and the set $\bar{\mathbb{R}} = [-\infty, +\infty]$ of all extended real numbers. The *extended real number system* consists of $\mathbb{R}$ and two extra symbols, $-\infty$ and ∞. The relation $<$ is extended to $\bar{\mathbb{R}}$ by postulating that $-\infty < x < +\infty$ for every real number x. The arithmetic operations are extended to $\bar{\mathbb{R}}$ as follows: for each $x \in \mathbb{R}$,

$$\begin{aligned} x + \infty = x - (-\infty) &= \infty \\ x + (-\infty) = x - \infty &= -\infty \\ x \cdot \infty &= \begin{cases} \infty & \text{if } x > 0 \\ -\infty & \text{if } x < 0 \end{cases} \end{aligned}$$

$$\begin{aligned} x\cdot(-\infty) &= (-x)\cdot\infty \\ x/\infty = x/(-\infty) &= 0 \\ \infty+\infty &= \infty \\ (-\infty)+(-\infty) &= -\infty \\ \infty\cdot\infty = (-\infty)\cdot(-\infty) &= \infty \\ \infty\cdot(-\infty) &= -\infty. \end{aligned}$$

The operations $0\cdot(\pm\infty)$, $(-\infty)-(-\infty)$, $+\infty/+\infty$, and $-\infty/-\infty$ are undefined.

Positive and Negative

We call x in $\bar{\mathbb{R}}$ *positive* if $x \geq 0$ and *strictly positive* if $x > 0$. By symmetry, then, x is *negative* if $x \leq 0$ and strictly negative if $x < 0$. A function $f : E \mapsto \bar{\mathbb{R}}$ is said to be *positive* if $f(x) \geq 0$ for all x in E and *strictly positive* if $f(x) > 0$ for all x in E. Negative and strictly negative functions are defined similarly. This usage is in accord with modern tendencies, though at variance with common usage[1].

Increasing, Decreasing

A function $f : \bar{\mathbb{R}} \mapsto \bar{\mathbb{R}}$ is said to be *increasing* if $f(x) \leq f(y)$ whenever $x \leq y$. It is said to be *strictly increasing* if $f(x) < f(y)$ whenever $x < y$. Decreasing and strictly decreasing functions are defined similarly by reversing the inequalities.

These notions carry over to functions $f : E \mapsto \bar{\mathbb{R}}$ with $E \subset \bar{\mathbb{R}}$. In particular, since a sequence is a function on $\mathbb{N}$, these notions apply to sequences in $\bar{\mathbb{R}}$. Thus, for example, $(x_n) \subset \bar{\mathbb{R}}$ is increasing if $x_1 \leq x_2 \leq \cdots$ and is strictly decreasing if $x_1 > x_2 > \cdots$.

Bounds

Let $A \subset \bar{\mathbb{R}}$. A real number b is called an *upper bound* for A provided that $A \subset [-\infty, b]$, and then A is said to be *bounded above* by b. Lower bounds and being bounded below are defined similarly. The set A is said to be *bounded* if it is bounded above and below; that is, if $A \subset [a, b]$ for some real interval $[a, b]$.

These notions carry over to functions and sequences as follows. Given $f : E \mapsto \bar{\mathbb{R}}$, the function f is said to be bounded above, below, etc. according as its range is bounded above, below, etc. Thus, for instance, f is bounded if there exist real numbers $a \leq b$ such that $a \leq f(x) \leq b$ for all x in E.

Supremum and Infimum

If $A \subset \bar{\mathbb{R}}$ is bounded above, then it has a least upper bound, that is, an upper bound b such that no number less than b is an upper bound; we call that least upper bound the *supremum* of A. If A is not bounded above, we define the supremum to be $+\infty$. The

[1]Often used concepts should have the simpler names. Mindbending double negatives should be avoided, and as much as possible, the mathematical usage should not conflict with the ordinary language.

infimum of A is defined similarly; it is $-\infty$ if A has no lower bound and is the greatest lower bound otherwise. We let

$$\inf A, \quad \sup A$$

denote the infimum and supremum of A, respectively. For example,

$$\inf\{1, 1/2, 1/3, \ldots\} = 0, \quad \sup\{1, 1/2, 1/3, \ldots\} = 1,$$

$$\inf(a, b] = \inf[a, b] = a, \quad \sup(a, b) = \sup(a, b] = b.$$

In particular, $\inf \emptyset = +\infty$ and $\sup \emptyset = -\infty$. If A is finite, then $\inf A$ is the smallest element of A, and $\sup A$ is the largest. Even when A in infinite, it is possible that $\inf A$ is an element of A, in which case it is called the *minimum* of A. Similarly, if $\sup A$ is an element of A, then it is also called the *maximum* of A.

If $f : E \mapsto \bar{\mathbb{R}}$, it is customary to write

$$\inf_{x \in D} f(x) = \inf\{f(x) : x \in D\}$$

and call it the infimum (or maximum) of f over $D \subset E$, and similarly with the supremum. In the case of sequences $(x_n) \subset \bar{\mathbb{R}}$,

$$\inf x_n, \quad \sup x_n$$

denote, respectively, the infimum and supremum of the range of (x_n). Other such notations are generally self-explanatory; for example,

$$\inf_{n \geq k} x_n = \inf\{x_k, x_{k+1}, \ldots\}, \quad \sup_{k \geq 1} x_{nk} = \sup\{x_{n1}, x_{n2}, \ldots\}.$$

Limits

If (x_n) is an increasing sequence in $\bar{\mathbb{R}}$, then $\sup x_n$ is also called the *limit* of (x_n) and is denoted by $\lim x_n$. If it is a decreasing sequence, then $\inf x_n$ is called the limit of (x_n) and again denoted by $\lim x_n$.

Let $(x_n) \subset \bar{\mathbb{R}}$ be an arbitrary sequence. Then

4.1 $$\underline{x}_m = \inf_{n \geq m} x_n, \quad \bar{x}_m = \sup_{n \geq m} x_n, \quad m \in \mathbb{N},$$

define two sequences; $(\underline{x}_n)$ is increasing, and $(\bar{x}_n)$ is decreasing. Their limits are called the *limit inferior* and the *limit superior*, respectively, of the sequence (x_n):

4.2 $$\liminf x_n = \lim \underline{x}_n = \sup_m \inf_{n \geq m} x_n,$$

4.3 $$\limsup x_n = \lim \bar{x}_n = \inf_m \sup_{n \geq m} x_n,$$

Figure 1 is worthy of careful study. Note that, in general,

4.4 $$-\infty \leq \liminf x_n \leq \limsup x_n \leq +\infty.$$

If $\liminf x_n = \limsup x_n$, then the common value is called the *limit* of (x_n) and is denoted by $\lim x_n$. Otherwise, if limits inferior and superior are not equal, the sequence (x_n) does not have a limit.

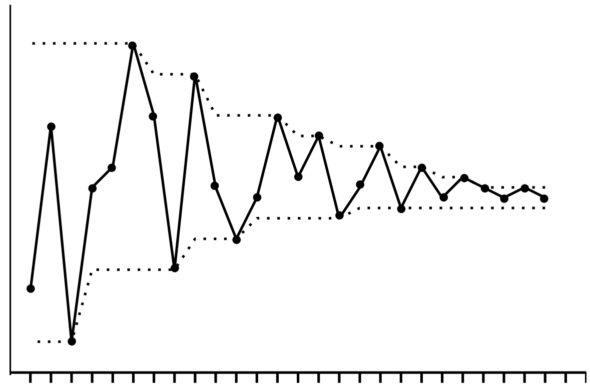

Figure 1: Lim Sup and Lim Inf. The pairs (n, x_n) are connected by the solid lines for clarity. The pairs $(n, \underline{x}_n)$ form the lower dotted line and $(n, \bar{x}_n)$ the upper dotted line.

Convergence of Sequences

A sequence (x_n) of real numbers is said to be *convergent* if $\lim x_n$ exists and is a real number.

An examination of Figure 1 shows that this is equivalent to the classical definition of convergence: (x_n) converges to x if for every $\epsilon > 0$, there is an n_ϵ such that $|x_n - x| < \epsilon$ for all $n \geq n_\epsilon$. The phrase "there is n_ϵ ... for all $n \geq n_\epsilon$" can be expressed in more geometric terms by phrases like "the number of terms outside $(x - \epsilon, x + \epsilon)$ is finite," or "all but finitely many terms are in $(x - \epsilon, x + \epsilon)$," or "$|x_n - x| < \epsilon$ for all n large enough."

The following is a summary of the relations between convergence and algebraic operations. The proof will be omitted.

4.5 THEOREM. *Let (x_n) and (y_n) be convergent sequences with limits x and y respectively. Then,*

1. $\lim cx_n = cx$,
2. $\lim(x_n + y_n) = x + y$,
3. $\lim x_n y_n = xy$,
4. $\lim x_n / y_n = x/y$ *provided that* $y_n, y \neq 0$.

In practice, we do not have the sequence laid out before us. Instead, some rule is given for generating the sequence and the object is to show that the resulting sequence will converge. For instance, a function may be specified somehow and a procedure described to find its maximum; starting from some point, the procedure will give the successive points $x_1, x_2, \ldots$ which are meant to form the sequence that converges to the point x where the maximum is achieved.

Often, to find the limit of (x_n), one starts with a search for sequences that bound (x_n) from above and below and whose limits can be computed easily: suppose that

$$y_n \le x_n \le z_n \quad \text{for all n,} \qquad \lim y_n = \lim z_n,$$

then $\lim x_n$ exists and is equal to the limit of the other two. The art involved is in finding such sequences (y_n) and (z_n).

4.6 EXAMPLE. This example illustrates the technique mentioned above. We want to show that $(n^{1/n})$ converges. Note that $n^{1/n} \ge 1$ always, and put $x_n = n^{1/n} - 1$, and consider the sequnce (x_n). Now, $(1 + x_n)^n = n$, and by the binomial theorem

$$\begin{aligned}(a+b)^n &= a^n + na^{n-1}b + \frac{n(n-1)}{2}a^{n-2}b^2 + \cdots + b^n \\ &\ge \frac{n(n-1)}{2}a^{n-2}b^2\end{aligned}$$

for $a, b \ge 0$ and $n \ge 2$. So,

$$n = (1 + x_n)^n \ge \frac{n(n-1)}{2}x_n^2,$$

or

$$0 \le x_n \le \sqrt{\frac{2}{n-1}}.$$

It follows that $\lim x_n = 0$, and hence

$$\lim n^{1/n} = 1.$$

Exercises:

4.1 Show that if $A \supset B$ then $\inf A \le \inf B \le \sup B \le \sup A$. Use this to show that, if $A_1 \supset A_2 \supset \cdots$, then

$$\inf A_1 \le \inf A_2 \le \cdots \le \inf A_n \le \cdots \le$$

$$\le \sup A_n \le \cdots \le \sup A_2 \le \sup A_1.$$

Use this to show that $(\underline{x}_n)$ is increasing, $(\bar{x}_n)$ is decreasing, and $\lim \underline{x}_n \le \lim \bar{x}_n$ (see (4.1) – (4.3) for definitions).

4.2 Show that $\sup(-x_n) = -\inf x_n$ for any sequence (x_n) in $\bar{\mathbb{R}}$. Conclude that $\lim\sup(-x_n) = -\lim\inf x_n$.

4.3 *Cauchy Criterion.* Sequence (x_n) is convergent if and only if for every $\epsilon > 0$ there is an n_ϵ such that $|x_m - x_n| \leq \epsilon$ for all $m \geq n \geq n_\epsilon$. Prove this by examining Figure 1 on the definition of the limit.

4.4 *Monotone Sequences.* If (x_n) is increasing, then $\lim x_n$ exists (but could be $+\infty$). Thus, such a sequence converges if and only if it is bounded above. Show this. State the version of this for decreasing sequences.

4.5 *Iterative Sequences.* Often, x_{n+1} is obtained from x_n via some rule, that is, $x_{n+1} = f(x_n)$ for some function f. If (x_n) is so obtained from some function f, it is said to be iterative. If (x_n) is such and f is continuous and $\lim x_n = x$ exists, then $x = f(x)$. This works well for identifying the limit especially when f is simple and $x = f(x)$ has only one solution. In general, with complicated functions f, the reverse is true: To find x satisfying $x = f(x)$, one starts at some point x_0, computes $x_1 = f(x_0)$, $x_2 = f(x_1)$, ..., and tries to show that $x = \lim x_n$ exists and satisfies $x = f(x)$.

4.6 *Domination.* A sequence (x_n) is said to be dominated by a sequence (y_n) if $x_n \leq y_n$ for each n. Show that, if so

1. $\inf x_n \leq \inf y_n$,
2. $\sup x_n \leq \sup y_n$,
3. $\liminf x_n \leq \liminf y_n$,
4. $\limsup x_n \leq \limsup y_n$.

In particular, if the limits exist, $\lim x_n \leq \lim y_n$.

Incidentally, $(\underline{x}_n)$ defined by (4.1) is the maximal increasing sequence dominated by (x_n), and $(\bar{x}_n)$ is the minimal decreasing sequence dominating (x_n).

4.7 *Comparisons.* Let (x_n) be a positive sequence. Then, (x_n) converges to 0 if and only if it is dominated by a sequence (y_n) with $\limsup y_n = 0$. Show this.

Favorite sequences (y_n) used in this role are given by $y_n = 1/n$, $y_n = r^n$ for some fixed number $r \in (0, 1)$, and $y_n = n^p r^n$ with $p \in (-\infty, +\infty)$ and $r \in (0, 1)$.

4.8 *Existence of Least Upper Bounds.* Let A be a nonempty subset of $\mathbb{R}$ and let $B = \{b : b \text{ is an upper bound of } A\}$. Assuming that B is nonempty, show that B has a minimum element.

5 Series

Given a sequence $(x_n) \subset \mathbb{R}$, the sequence (s_n) defined by

5.1
$$s_n = \sum_{i=1}^{n} x_i$$

is called the sequence of partial sums of (x_n), and the symbolic expression

5.2
$$\sum x_n$$

is called the *series* associated with (x_n). The series is said to *converge* to s, and then we write

5.3
$$\sum_{1}^{\infty} x_n = s$$

if and only if the sequence (s_n) converges to s.

Sometimes, we write $x_1 + x_2 + \cdots$ for the series (5.2). Sometimes, for convenience of notation, we shall consider series of the form $\sum_0^\infty$ or $\sum_m^\infty$, depending on the index set. Here are a few examples:

$$\begin{aligned}
\sum_{n=0}^{\infty} x^n &= \frac{1}{1-x} \quad \text{for } x \in (-1,1), \\
\sum_{n=0}^{\infty} \frac{x^n}{n!} &= e^x \quad \text{for } x \in \mathbb{R}, \\
\sum_{n=1}^{\infty} \frac{1}{n^2} &= \frac{\pi^2}{6}, \\
\sum_{n=m}^{\infty} x^n &= \frac{x^m}{1-x} \quad \text{for } x \in (-1,1).
\end{aligned}$$

The following result is obtained by applying the Cauchy Criterion (Exercise 4.3) to the sequence of partial sums.

5.4 THEOREM. *The series $\sum x_n$ converges if and only if for every $\epsilon > 0$ there is an n_ϵ such that*

5.5
$$|\sum_{i=n}^{m} x_i| \le \epsilon$$

for all $m \ge n \ge n_\epsilon$.

In particular, taking $m = n$ in (5.5) we obtain $|x_n| \le \epsilon$. Thus we have obtained the following:

5.6 COROLLARY. *If $\sum x_n$ converges, then $\lim x_n = 0$.*

The converse is not true. For example, $\lim 1/n = 0$ but $\sum 1/n$ is divergent. In the case of series with positive terms, partial sums form an increasing sequence, and hence, the following holds (see Exercise 4.4):

5.7 PROPOSITION. *Suppose that the x_n are positive. Then $\sum x_n$ converges if and only if the sequence of partial sums is bounded.*

In many cases, we encounter series whose terms are positive and decreasing. The following theorem due to Cauchy is helpful in such cases, especially if the terms involve powers. Note the way a rather thin sequence determines the convergence or divergence of the whole series.

5.8 THEOREM. *Suppose that (x_n) is decreasing and positive. Then $\sum x_n$ converges if and only if the series*

$$x_1 + 2x_2 + 4x_4 + 8x_8 + \cdots$$

converges.

PROOF. Let $s_n = x_1 + \cdots + x_n$ as usual and put $t_k = x_1 + 2x_2 + \cdots + 2^k x_{2^k}$. Now, for $n \leq 2^k$, since $x_1 \geq x_2 \geq \cdots \geq 0$,

$$\begin{aligned} s_n &\leq x_1 + (x_2 + x_3) + (x_4 + \cdots x_7) + \cdots + (x_{2^k} + \cdots + x_{2^{k+1}-1}) \\ &\leq x_1 + 2x_2 + 4x_4 + \cdots + 2^k x_{2^k} \\ &= t_k, \end{aligned}$$

and for $n \geq 2^k$,

$$\begin{aligned} s_n &\geq x_1 + x_2 + (x_3 + x_4) + (x_5 + \cdots x_8) + \cdots + (x_{2^{k-1}+1} + \cdots + x_{2^k}) \\ &\geq \frac{1}{2}x_1 + x_2 + 2x_4 + \cdots + 2^{k-1}x_{2^k} \\ &= \frac{1}{2}t_k. \end{aligned}$$

Thus, the sequences (s_n) and (t_n) are either both bounded or both unbounded, which completes the proof via Proposition 5.7 □

5.9 EXAMPLE. $\sum 1/n^p$ converges if $p > 1$ and diverges if $p \leq 1$. For $p \leq 0$, the claim is trivial to see. For $p > 0$, the terms $x_n = 1/n^p$ form a decreasing positive sequnce, and thus, the preceding theorem applies. Now,

$$\sum_{k=0}^{\infty} 2^k x_{2^k} = \sum (2^{1-p})^k,$$

which converges if $2^{1-p} < 1$ and diverges otherwise. Since $2^{1-p} < 1$ if and only if $p > 1$, we are done.

5.10 EXAMPLE. The series

$$\sum_{2}^{\infty} \frac{1}{n(\log n)^p}$$

converges if $p \in (1, \infty)$ and diverges otherwise. Here we start the series with $n = 2$ since $\log 1 = 0$. Since the logarithm function is monotone increasing, Theorem 5.8 applies. Now, $x_n = 1/n(\log n)^p$ and so

$$\sum_{k=1}^{\infty} 2^k x_{2^k} = \sum_{1}^{\infty} 2^k \frac{1}{2^k (\log 2^k)^p} = \frac{1}{(\log 2)^p} \sum_{1}^{\infty} \frac{1}{k^p},$$

which converges if and only if $p > 1$ in view of the preceding example.

Ratio Test, Root Test

The ratio test ties the convergence of $\sum x_n$ to the behavior of the ratios x_{n+1}/x_n for large n; it is highly useful.

5.11 THEOREM.

1. *If* $\limsup |x_{n+1}/x_n| < 1$*, then* $\sum x_n$ *converges.*
2. *If* $\liminf |x_{n+1}/x_n| > 1$*, then* $\sum x_n$ *diverges.*

PROOF. (1) If $\limsup |x_{n+1}/x_n| < 1$, then there is a number $r \in [0, 1)$ and an integer n_0 such that $|x_{n+1}/x_n| \leq r$ for all $n \geq n_0$. Thus $|x_{n_0+k}| \leq |x_{n_0}| r^k$ for all $k \geq 0$, and therefore, for $m > n > n_0$,

$$|\sum_{i=n}^{m} x_i| \leq \sum_{i=n}^{\infty} |x_i| \leq |x_{n_0}| \sum_{i=n}^{\infty} r^{i-n_0} = |x_{n_0}| \frac{r^{n-n_0}}{1-r}.$$

Given $\epsilon > 0$ choose n_ϵ so that $|x_{n_0}| r^{n_\epsilon - n_0}/(1-r) < \epsilon$. Then Cauchy's criterion works with this n_ϵ and $\sum x_n$ converges.

(2) If $\liminf |x_{n+1}/x_n| > 1$ then there is an integer n_0 such that $|x_{n+1}| \geq |x_n|$ for all $n \geq n_0$. Hence, $|x_n| \geq |x_{n_0}|$ for all $n \geq n_0$ which shows that (x_n) does not converge to 0 as it must in order for $\sum x_n$ to converge (see Corollary 5.6). $\square$

The preceding test gives no information in cases where

$$\liminf |x_{n+1}/x_n| \leq 1 \leq \limsup |x_{n+1}/x_n|.$$

For instance, for the two series $\sum 1/n$ and $\sum 1/n^2$, both the $\liminf$ and the $\limsup$ are equal to 1, but the first series diverges whereas the second converges. Also, the series

5.12
$$\frac{1}{2}+\frac{1}{3}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{2^3}+\frac{1}{3^3}+\frac{1}{2^4}+\frac{1}{3^4}+\cdots$$

obviously converges to $3/2$; yet, the ratio test is miserably inconclusive:

$$\liminf \frac{x_{n+1}}{x_n} = \lim \left(\frac{2}{3}\right)^n = 0$$

$$\limsup \frac{x_{n+1}}{x_n} = \lim \left(\frac{3}{2}\right)^n = \infty.$$

The following test, called the *root test*, is a stronger test — if the ratio test works, so does the root test. But the root test works in some situations where the ratio test fails; for example, the root test works for the series (5.12).

5.13 THEOREM. *Let* $a = \limsup |x_n|^{1/n}$. *Then* $\sum x_n$ *converges if* $a < 1$, *and diverges if* $a > 1$.

PROOF. Suppose that $a < 1$. Then, there is a $b \in (a, 1)$ such that $|x_n|^{1/n} \le b$ for all $n \ge n_0$, where n_0 is some integer. Then, $|x_n| \le b^n$ for all $n \ge n_0$, and comparing $\sum x_n$ with the geometric series $\sum b^n$ shows that $\sum x_n$ converges.

Suppose that $a > 1$. Then, a subsequence of $|x_n|$ must converge to $a > 1$, which means that $|x_n| \ge 1$ for infinitely many n. So, (x_n) does not converge to zero, and hence, $\sum x_n$ cannot converge. □

Power Series

Given a sequence (c_n) of complex numbers, the series

5.14
$$\sum_0^\infty c_n z^n$$

is called a *power series*. The numbers $c_0, c_1, \ldots$ are called the coefficients of the power series; here z is a complex number.

In general, the series will converge or diverge, depending on the choice of z. As the following theorem shows, there is a number $r \in [0, \infty]$, called the radius of convergence, such that the series converges if $|z| < r$ and diverges if $|z| > r$. The behavior for $|z| = r$ is much more complicated and cannot be described easily.

5.15 THEOREM. *Let* $a = \limsup |c_n|^{1/n}$ *and* $r = 1/a$.

1. *If* $|z| < r$, *then* $\sum c_n z^n$ *converges.*

2. *If $|z| > r$, then $\sum c_n z^n$ diverges.*

PROOF. Put $x_n = c_n z^n$ and apply the root test with

$$\limsup |x_n|^{1/n} = |z| \limsup |c_n|^{1/n} = a|z| = \frac{|z|}{r}.$$

□

5.16 EXAMPLE.

1. $\sum z^n/n! = e^z$ and $r = \infty$.
2. $\sum z^n$ converges for $|z| < 1$ and diverges for $|z| \geq 1$; $r = 1$.
3. $\sum z^n/n^2$ converges for $|z| \leq 1$ and diverges for $|z| > 1$; $r = 1$.
4. $\sum z^n/n$ converges for $|z| < 1$ and diverges for $|z| > 1$; $r = 1$; for $z = 1$ the series diverges, but for $|z| = 1$ but $z \neq 1$ it converges.

Absolute Convergence

The series $\sum x_n$ is said to *converge absolutely* if $\sum |x_n|$ is convergent. If the x_n are all positive numbers, then absolute convergence is the same as convergence. Using Cauchy's criterion (see Theorem 5.4) on both sides of

$$|\sum_{i=n}^{m} x_i| \leq \sum_{i=n}^{m} |x_i|$$

shows that if $\sum x_n$ converges absolutely then it converges. But the converse is not true: for example,

$$\sum (-1)^n/n$$

converges but is not absolutely convergent.

The comparison tests above, as well as the root and ratio tests, are in fact tests for absolute convergence. If a series is not absolutely convergent, one has to study the sequence of partial sums to determine whether the series converges at all.

Rearrangements

Let $(k_1, k_2, \ldots)$ be a sequence in which every integer $n \geq 1$ appears once and only once, that is, $n \mapsto k_n$ is a bijection from $\mathbb{N}$ onto $\mathbb{N}$. If

$$y_n = x_{k_n}, \quad n \in \mathbb{N},$$

for such a sequence (k_n), then we say that (y_n) is a rearrangement of (x_n).

Let (y_n) be a rearrangement of (x_n). In general, the series $\sum y_n$ and $\sum x_n$ are quite different. However, if $\sum x_n$ is absolutely convergent, then so is $\sum y_n$ and it converges to the same number as does $\sum x_n$. The converse is also true: if every rearrangement of the series $\sum x_n$ converges, then the series $\sum x_n$ is absolutely convergent and all its rearrangements converge (to the same sum).

On the other hand, if $\sum x_n$ is not absolutely convergent, its various rearrangements may converge or diverge, and in the case of convergence, the sum generally depends on the rearrangement chosen. For instance,

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \cdots$$

is convergent, but not absolutely so. Its rearrangement

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \cdots$$

(with $+ + - + + - + + -$ pattern) is again convergent, but not to the same sum. In fact, the following theorem due to Riemann shows that one can create rearrangements that are as bizarre as one wants.

5.17 THEOREM. *Let $\sum x_n$ be convergent but not absolutely. Then, for any two numbers $a \leq b$ in $\bar{\mathbb{R}}$ there is a rearrangement $\sum y_n$ of $\sum x_n$ such that*

$$\liminf \sum_1^n y_i = a, \quad \limsup \sum_1^n y_i = b.$$

We omit the proof. Note that, in particular, taking $a = b$ we can find a rearrangement $\sum y_n$ with sum a, no matter what a is.

Exercises:

5.1 Determine the convergence or divergence of the following:

1. $\sum(\sqrt{n+1} - \sqrt{n})$
2. $\sum(\sqrt{n+1} - \sqrt{n})/n$
3. $\sum(\sin n)/(n\sqrt{n})$
4. $\sum(-1)^n n/(n^2+1)$.

In case of convergence, indicate whether it is absolute convergence.

5.2 Show that if $\sum x_n$ converges then so does $\sum \sqrt{x_n}/n$.

5.3 Show that if $\sum x_n$ converges and (y_n) is bounded and monotone (either increasing or decreasing), then $\sum x_n y_n$ converges.

5.4 Find the radius of convergence of each of the following power series:

1. $\sum n^2 z^n$,
2. $\sum 2^n z^n/n!$,
3. $\sum 2^n z^n/n^2$,
4. $\sum n^3 z^n/3^n$.

5.5 Suppose that $f(z) = \sum c_n z^n$. Express the sum of the even terms, $\sum c_{2n} z^{2n}$, and the sum of the odd terms, $\sum c_{2n+1} z^{2n+1}$, in terms of f.

5.6 Suppose that $f(z) = \sum c_n z^n$. Express $\sum c_{3n} z^{3n}$ in terms of f.

5.7 *Rearrangements.* Let $\sum x_n$ be a series that converges absolutely. Prove that every rearrangement of $\sum x_n$ converges, and that they all converge to the same sum.

5.8 *Riemann's Theorem.* Prove Riemann's theorem 5.17 by filling in the details in the following outline:

1. Let (x_n^+) denote the subsequence consisting of the positive elements of (x_n) and let (x_n^-) denote the subsequence of negative elements of (x_n). Both of these sequences must be infinite.
2. Both sequences (x_n^+) and (x_n^-) converge to zero.
3. Both series $\sum x_n^+$ and $\sum x_n^-$ diverge.
4. Suppose that $a, b \in \mathbb{R}$ and define a rearrangement as follows: start with the positive elements and choose elements from this set until the partial sum exceeds b. Then, choose elements from the set of negative elements until the partial sum is less than a. Then, choose elements from the set of positive elements until the partial sum exceeds b. Continue this proceedure of alternating between elements of the positive and negative sets indefinitely.
5. Prove that the procedure described above can be continued ad infinitum.
6. Prove that this rearrangement has the properties stated in Riemann's theorem.
7. Extend the above arguments to the case where $a, b = \pm\infty$.

5.9 *Poisson distribution.* Let $p_n = e^{-\lambda}\lambda^n/n!$ where λ is a positive real. Show that

1. $p_n > 0$,

2. $\sum_{n=0}^{\infty} p_n = 1$,
3. $\sum_{n=0}^{\infty} n p_n = \lambda$.

5.10 *Borel Summability.* Consider a series $\sum_{n=0}^{\infty} x_n$ with partial sums $s_n = \sum_{i=0}^{n} x_i$. We say that the series is *Borel summable* if

$$\lim_{\lambda \to \infty} \sum_{n=0}^{\infty} s_n p_n$$

converges, where p_n are the Poisson probabilities defined in Exercise 5.9. For what values of z is the geometric series $\sum_{n=0}^{\infty} z^n$ Borel summable?

Metric Spaces

Basic questions of analysis on the real line are tied to the notions of closeness and distances between points. The same issue of closeness comes up in more complicated settings, for instance, like when we try to approximate a function by a simpler function. Our aim is to introduce the idea of distance in general, so that we can talk of the distance between two functions with the same conceptual ease as when we talk of the distance between two points in the plane. After that, we discuss the main issues: convergence, continuity, approximations. All along, there will be examples of different spaces and different ways of measuring distances.

6 Euclidean Spaces

This section is to review the space $\mathbb{R}^n$ together with its Euclidean distance. Recall that each element of $\mathbb{R}^n$ is an n-tuple $x = (x_1, \ldots, x_n)$, where the x_i are real numbers. The elements of $\mathbb{R}^n$ are called *points* or *vectors*, and we are familiar with the operations like addition of vectors and multiplication by scalars.

Inner Product and Norm

For x and y in $\mathbb{R}^n$, their *inner product* $x \cdot y$ is the number

6.1 $$x \cdot y = \sum_1^n x_i y_i.$$

If we regard x and y as column vectors, then $x \cdot y = x^T y$. For x in $\mathbb{R}^n$, the *norm* of x is defined to be the positive number

6.2 $$\|x\| = \sqrt{x \cdot x} = \sqrt{\textstyle\sum_1^n x_i^2}.$$

The norm satisfies the following:

6.3 $$\|x\| \geq 0 \text{ for every } x \text{ in } \mathbb{R}^n,$$

6.4 $$\|x\| = 0 \text{ if and only if } x = 0,$$

6.5 $$\|x + y\| \leq \|x\| + \|y\| \text{ for all } x \text{ and } y \text{ in } \mathbb{R}^n.$$

Of these, 6.3 and 6.4 are obvious, and 6.5 is immediate from the following, which is called the *Schwartz inequality*.

6.6 PROPOSITION. $|x \cdot y| \leq \|x\|\|y\|$ *for all* x *and* y *in* $\mathbb{R}^n$.

PROOF. Consider the function

$$\begin{aligned} f(\lambda) &= \|\lambda y - x\|^2 \\ &= \lambda^2\|y\|^2 - 2\lambda(x \cdot y) + \|x\|^2. \end{aligned}$$

This function is clearly positive and quadratic and its minimum occurs at

$$\lambda = \frac{x \cdot y}{\|y\|^2}.$$

For this value of λ we have

$$0 \leq f(\frac{x \cdot y}{\|y\|^2}) = -\frac{(x \cdot y)^2}{\|y\|^2} + \|x\|^2$$

from which Schwartz's inequality follows immediately. □

Euclidean Distance

For x and y in $\mathbb{R}^n$, the *Euclidean distance* between x and y is defined to be the number $\|x - y\|$. It follows from the properties given above that, for all x, y, z in $\mathbb{R}^n$,

1. $\|x - y\| \geq 0$,
2. $\|x - y\| = \|y - x\|$,
3. $\|x - y\| = 0$ if and only if $x = y$,
4. $\|x - y\| + \|y - z\| \geq \|x - z\|$.

The last is called the *triangle inequality*: on $\mathbb{R}^2$, if the points x, y, z are the vertices of a triangle, this is simply the well-known fact that the sum of the lengths of two sides is greater than or equal to the length of the third side.

The set $\mathbb{R}^n$ together with the Euclidean distance is called n*-dimensional Euclidean space*. The Euclidean spaces are important examples of metric spaces.

Exercises:

6.1 Show that the mapping $(x, y) \mapsto x \cdot y$ from $\mathbb{R}^n \times \mathbb{R}^n$ into $\mathbb{R}$ is a linear transformation in x and is a linear transformation in y (and therefore is said to be bilinear). Conclude that

$$(x + y) \cdot (x + y) = x \cdot x + 2x \cdot y + y \cdot y.$$

Use this and the Schwartz inequality to prove (6.5).

6.2 Show that $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$. Interpret this in geometric terms, on $\mathbb{R}^2$, as a statement about parallelograms.

6.3 Points x and y are said to be *orthogonal* if $x \cdot y = 0$. Show that this is equivalent to saying that the lines connecting the origin to x and y are perpendicular. In general, letting α be the angle between the lines through x and y, we have $x \cdot y = \|x\|\|y\| \cos\alpha$.

7 Metric Spaces

Let E be a set. A *metric* on E is a function $d : E \times E \mapsto \mathbb{R}_+$ that satisfies the following for all x, y, z in E:

1. $d(x, y) = d(y, x)$,
2. $d(x, y) = 0$ if and only if $x = y$,
3. $d(x, y) + d(y, z) \geq d(x, z)$.

A *metric space* is a pair (E, d) where E is a set and d is a metric on E. In this context, we think of E as a space, call the elements of E points, and refer to $d(x, y)$ as the distance from x to y.

EXAMPLES.

7.1 *Euclidean spaces.* Consider $\mathbb{R}^n$ with the Euclidean distance $d(x, y) = \|x - y\|$ on it. It follows from (1)–(4) that d is a metric on $\mathbb{R}^n$. Thus, $(\mathbb{R}^n, d)$ is a metric space and is called n-dimensional Euclidean space.

7.2 *Manhattan metric.* On $\mathbb{R}^n$ define a metric d by

$$d(x, y) = \sum_1^n |x_i - y_i|.$$

This d is called the Manhattan metric, or l_1-metric, on $\mathbb{R}^n$, and $(\mathbb{R}^n, d)$ is a metric space again. Note that for $n > 1$ this metric is different from the Euclidean metric of the preceding example.

7.3 *Space $\mathcal{C}$.* Let $\mathcal{C}$ denote the set of all continuous functions from the interval $[0, 1]$ into $\mathbb{R}$. For x and y in $\mathcal{C}$, let

$$d(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

It is clear that $d(x, y)$ is a positive real number, that $d(x, y) = d(y, x)$, and that $d(x, y) = 0$ if and only if $x = y$. As for the triangle inequality, we note that

$$|x(t) - z(t)| \leq |x(t) - y(t)| + |y(t) - z(t)| \leq d(x, y) + d(y, z)$$

for every t in $[0, 1]$, from which we have $d(x, y) + d(y, z) \geq d(x, z)$. Thus, d is a metric on $\mathcal{C}$, and $(\mathcal{C}, d)$ is a metric space. This metric space is important in analysis.

Usage

In the literature, it is common practice to call E a metric space if (E, d) is a metric space for some metric d. If there is only one metric under consideration, this is harmless and saves time. For instance, the phrase "Euclidean space $\mathbb{R}^n$" refers to $(\mathbb{R}^n, d)$ where d is the Euclidean metric. For a while at least, we shall indicate the metric involved in each case in order to avoid all possible confusion.

Distances from Points to Sets and from Sets to Sets

Let (E, d) be a metric space. For x in E and $A \subset E$, let

7.4 $$d(x, A) = \inf\{d(x, y) : y \in A\};$$

this is called the distance from the point x to the set A. For $A \subset E$ and $B \subset E$, the distance from A to B is defined by

7.5 $$d(A, B) = \inf\{d(x, y) : x \in A, y \in B\}.$$

The *diameter* of a set $A \subset E$ is defined to be

7.6 $$\text{diam } A = \sup\{d(x, y) : x \in A, y \in A\}.$$

A set is said to be *bounded* if its diameter is finite.

Balls

Let (E, d) be a metric space. For x in E and r in $(0, \infty)$,

7.7 $$B(x, r) = \{y \in E : d(x, y) < r\}$$

is called the *open ball* with *center* x and *radius* r, and

7.8 $$\bar{B}(x, r) = \{y \in E : d(x, y) \leq r\}$$

is the corresponding *closed ball*.

For example, if $E = \mathbb{R}^3$ and d is the usual Euclidean metric, then $B(x, r)$ becomes the set of all points inside the sphere with center x and radius r, and $\bar{B}(x, r)$ is the set of all points inside or on that sphere.

Exercises and Complements:

7.1 *Discrete metric.* Let E be an arbitrary set. Define

$$d(x,y) = \left[\begin{array}{ll} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{array} \right.$$

Show that this d is a metric on E. It is called the discrete metric on E.

7.2 *Metrics on* $\mathbb{R}^n$. For each number $p \geq 1$,

$$d_p(x,y) = (\sum_1^n |x_i - y_i|^p)^{1/p}$$

defines a metric d_p on $\mathbb{R}^n$. Note that d_1 is the Manhatten metric, and d_2 is the Euclidean metric. Finally,

$$d_\infty(x,y) = \sup_{1 \leq i \leq n} |x_i - y_i|$$

is again a metric on $\mathbb{R}^n$. Show this.

7.3 *Equivalent Metrics.* Two metrics d and d' are equivalent if there exist strictly positive constants c_1 and c_2 such that for all x, y:

$$c_1 d'(x,y) \leq d(x,y) \leq c_2 d'(x,y).$$

Show that d_1, d_2, and d_∞ are all equivalent to each other.

7.4 *Weighted Metrics on* $\mathbb{R}^n$. The metrics introduced in the preceding exercise treat all components of $x-y$ equally. This is reasonable if $\mathbb{R}^n$ is thought of geometrically and the selection of a coordinate system is unimportant. On the other hand, if $x = (x_1, \ldots, x_n)$ stands for a shopping list that requires buying x_1 units of product one, and x_2 units of product two, and so on, then it would make much better sense to define the distance between two shopping lists x and y by

$$d(x,y) = \sum_1^n w_i |x_i - y_i|$$

where $x_1, \ldots, w_n$ are fixed, strictly positive numbers, with w_i being the value of one unit of product i. Show that this d is indeed a metric. More generally, paralleling the metrics introduced in the previous exercise,

$$d_p(x,y) = (\sum_i^n w_i |x_i - y_i|^p)^{1/p}, \quad x, y \in \mathbb{R}^n,$$

is a metric on $\mathbb{R}^n$ for each $p \geq 1$ and each fixed, strictly positive vector w (the latter means $w_1 > 0, \ldots, w_n > 0$).

7.5 *l^2-Spaces.* Instead of $\mathbb{R}^n$, now consider the space $\mathbb{R}^\infty$ of all infinite sequences in $\mathbb{R}$, that is, each x in $\mathbb{R}^\infty$ is a sequence $x = (x_1, x_2, \ldots)$ of real numbers. In analogy with the d_2 metrics introduced on $\mathbb{R}^n$ in Exercises 7.2 and 7.4, we define

$$d_2(x, y) = (\sum_1^\infty |x_i - y_i|^2)^{1/2}.$$

This d_2 satisfies all the conditions for a metric except that $d_2(x, y)$ can be ∞ for some x and y. To remedy the latter, we let E be the set of all x in $\mathbb{R}^\infty$ with

$$\sum_1^\infty x_i^2 < \infty.$$

Then, by an easy generalization of the Schwartz inequality, it follows that $d_2(x, y) < \infty$ for all x and y in E. Thus, (E, d_2) is a metric space. It is generally denoted by l^2.

7.6 *Metrics on $\mathcal{C}$.* Consider the set $\mathcal{C}$ of all continuous functions from $[0, 1]$ into $\mathbb{R}$. The interval $[0, 1]$ can be replaced by any bounded interval $[a, b]$, in which case one writes $\mathcal{C}([a, b])$. A number of metrics can be defined on $\mathcal{C}$ in analogy with those in Exercise 7.2. The analogy is provided by the following observation: every x in $\mathbb{R}^n$ can be thought of as a function x from $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$ into $\mathbb{R}$, namely, the function x with $x(t) = x_i$ for $t = i/n$. Thus, replacing the set $\{\frac{1}{n} \frac{2}{n}, \ldots, \frac{n}{n}\}$ with the interval $[0, 1]$ and replacing the summation by integration, we obtain

$$d_p(x, y) = (\int_0^1 |x(t) - y(t)|^p dt)^{1/p}$$

for all x and y in $\mathcal{C}$. Since any continuous function on $[0, 1]$ is bounded, the integral here is finite and it is easy to check the conditions for this d_p to be a metric, except perhaps for the triangle inequality. So, for each $p \geq 1$, this d_p is a metric on $\mathcal{C}$. Incidentally, the metric of Example 7.3 can be denoted by d_∞ in analogy with d_∞ in Exercise 7.2.

7.7 *Open Balls.* Let $E = \mathbb{R}^2$. Describe the open ball $B(x, r)$, for fixed x and r, under each of the following metrics:

1. d_2 of Exercise 7.2.
2. d_1 of Exercise 7.2.
3. d_∞ of Exercise 7.2.
4. d_2 of Exercise 7.4 with $w_1 = 1$ and $w_2 = 5$.

7.8 *Open Balls in $\mathcal{C}$.* For the metric space of Example 7.3, describe $B(x, r)$ for a fixed function x and fixed number $r > 0$. Draw pictures!

7.9 *Product Spaces.* Let (E_1, d_1) and (E_2, d_2) be arbitrary metric spaces. Let $E = E_1 \times E_2$ and define, for $x = (x_1, x_2)$ in E and $y = (y_1, y_2)$ in E,

$$d(x, y) = [d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2]^{1/2}.$$

Show that d is a metric on E. The metric space (E, d) is called the product of the metric spaces (E_1, d_1) and (E_2, d_2).

8 Open and Closed Sets

Let (E, d) be a metric space. All points mentioned below are points of E, all sets are subsets of E. Recall the definition 7.7 of the open ball $B(x, r)$ with center x and radius r.

8.1 DEFINITION. A set A is said to be *open* if for every x in A there is an $r > 0$ such that $B(x, r) \subset A$. A set is said to be *closed* if its complement is open.

For example, if $E = \mathbb{R}$ with the usual distance, the intervals (a, b), $(-\infty, b)$, (a, ∞) are open, the intervals $[a, b]$, $(-\infty, b]$, $[a, \infty)$ are closed, and the interval $(a, b]$ is neither open nor closed.

8.2 PROPOSITION. *Every open ball is open.*

PROOF. Fix x and r. To show that $B(x, r)$ is open, we need to show that for every y in $B(x, r)$ there is a $q > 0$ such that $B(y, q) \subset B(x, r)$. This is accomplished by picking $q = r - d(x, y)$. Since y is in $B(x, r)$, we have $d(x, y) < r$ and, hence, $q > 0$. And, every point of $B(y, q)$ is a point of $B(x, r)$, because $z \in B(y, q)$ means $d(z, y) < q$ which implies that

$$d(z, x) \leq d(z, y) + d(y, x) < q + d(y, x) = r.$$

$\square$

8.3 THEOREM. *The sets $\emptyset$ and E are open. The intersection of a finite number of open sets is open. The union of an arbitrary collection of open sets is open.*

PROOF. The first assertion is trivial from the definition.

We prove the second assertion for the intersection of two open sets. The general case follows from the repeated aplication of the case for two. Let A and B be open. Let $x \in A \cap B$. Since A is open and x is in A, there is $p > 0$ such that $B(x, p) \subset A$.

Similarly, there is a $q > 0$ such that $B(x, q) \subset B$. Let $r = p \wedge q$, the smaller of p and q. Then, $B(x, r) \subset B(x, p) \subset A$ and $B(x, r) \subset B(x, q) \subset B$. Hence, $B(x, r) \subset A \cap B$. So, $A \cap B$ is open.

For the last assertion, let $\{A_i : i \in I\}$ be an arbitrary collection of open sets. We want to show that $A = \cup_i A_i$ is open. Let x be in A. Then, $x \in A_i$ for some $i \in I$. Since A_i is open, there is an $r > 0$ such that $B(x, r) \subset A$. Since $A_i \subset A$, this shows that $B(x, r) \subset A$. So, A is open. □

The following characterization is immediate from the preceding theorem together with Proposition 8.2.

8.4 PROPOSITION. *A set is open if and only if it is the union of a collection of open balls.*

PROOF. If A is the union of a collection of open balls, then A must be open in view of 8.2 and 8.3. To show the converse, let A be open. Then, for every x in A, there is an open ball $A_x = B(x, r(x))$ contained in A. Obviously, the union of all these A_x is exactly A. □

Closed Sets

Recall that a subset of E is closed if and only if its complement is open. Thus, the following theorem is immediate from Theorem 8.3 above and the fact that the complement of a union is the intersection of complements and vice versa.

8.5 THEOREM. *The sets $\emptyset$ and E are closed. The union of finitely many closed sets is closed. The intersection of an arbitrary collection of closed sets is closed.*

Every closed ball is closed. This last observation can be proved along the lines of 8.2: if $y \in E \setminus \bar{B}(x, r)$ then $d(y, x) > r$, and picking $p = d(x, y) - r > 0$ we see that $B(y, p) \subset E \setminus \bar{B}(x, r)$, which proves that $E \setminus \bar{B}(x, r)$ is open. In particular, for each x in E, the singleton $\{x\}$ is closed. It follows from this and the preceding theorem that every finite set is closed.

Interior, Closure, and Boundary

Let A be a subset of E. The collection of all closed sets containing A is not empty (since E belongs to that collection.) The intersection $\bar{A}$ of that collection is a closed set by the last theorem. Clearly, $\bar{A}$ is the smallest closed set that contains A, that is, if $B \supset A$ and B is closed then $B \supset \bar{A}$. The set $\bar{A}$ is called the *closure* of A.

We define the *interior* of A similarly as the largest open set contained in A, and we denote it by A°. In other words, A° is the union of all open sets contained in A. Note that

8.6 $$A^\circ \subset A \subset \bar{A}.$$

We define the *boundary* of A to be the set $\partial A = \bar{A} \setminus A^\circ$.

For example, if A is the open ball $B(x,r)$ in the Euclidean space $E = \mathbb{R}^n$, the $A^\circ = A$, $\bar{A} = \bar{B}(x,r)$, and ∂A is the sphere of radius r centered at x. If $E = \mathbb{R}$ with the usual metric, and if $A = (a,b]$, then $\bar{A} = [a,b]$ and $A^\circ = (a,b)$ and $\partial A = \{a,b\}$. The following seems self evident.

8.7 PROPOSITION. *A set is closed if and only if it is equal to its closure. A set is open if and only if it is equal to its interior.*

Open Subsets of the Real Line

We take $E = \mathbb{R}$ with the usual distance. Then, every open ball is an open interval, and according to Proposition 8.4, every open set is the union of a collection of open balls. The following sharpens the picture by taking into account the special nature of the real line.

8.8 THEOREM. *A subset of $\mathbb{R}$ is open if and only if it is the union of a countable collection of disjoint open intervals.*

PROOF. The "if" part is immediate from Proposition 8.4 and the fact that every open ball is an interval in this case.

To prove the "only if" part, let A be an open subset of $\mathbb{R}$. Recall that the set $\mathbb{Q}$ of all rationals is countable. For each q in $\mathbb{Q} \cap A$, let

$$a_q = \sup\{y \le q : y \notin A\}, \quad b_q = \inf\{y \ge q : y \notin A\}.$$

Then,

$$B = \bigcup_{q \in \mathbb{Q} \cap A} (a_q, b_q)$$

is the union of a countable collection of open intervals. We show next that $A = B$ by showing that $A \subset B$ and $B \subset A$.

Let x be in A. Since A is open, there is a ball $B(x,r)$ contained in A. Take a rational number q in this ball. Clearly, $B(x,r) \subset (a_q, b_q)$. Thus, x is in B. Since this is true for every x in A, we have that $A \subset B$.

Fix $q \in \mathbb{Q} \cap A$. Clearly, $(a_q, b_q) \subset A$. Hence, $B \subset A$.

We have shown that $A = B$, and B has the desired form except that the intervals (a_q, b_q) are not necessarily disjoint. Note that if $r \in (a_q, b_q)$ then $(a_r, b_r) = (a_q, b_q)$ and $q \in (a_r, b_r)$. Let us write $q \approx r$ if and only if $(a_q, b_q) = (a_r, b_r)$. This defines an equivalence relation on the set $\mathbb{Q} \cap A$. Thus, by picking exactly one q from each

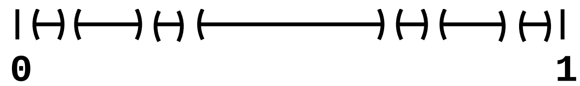

Figure 2: The set $D = \cup D_q$.

equivalence class, we can form a set $I \subset \mathbb{Q} \cap A$ such that $(a_q, b_q) \cap (a_r, b_r) = \emptyset$ for all distinct q and r in I, and

$$A = B = \bigcup_{q \in I} (a_q, b_q).$$

$\square$

8.9 EXAMPLE.

The Cantor Set. Start with the unit interval $\mathbb{B} = [0, 1]$. To each q in the set $I = \{1/2;\ 1/4,\ 3/4;\ 1/8,\ 3/8,\ 5/8,\ 7/8;\ 1/16,\ 3/16, \ldots, 15/16; \ldots\}$ we associate an open interval D_q in the following fashion: $D_{1/2}$ is the open interval $(1/3, 2/3)$ which is the middle third of $\mathbb{B}$. Deleting it from $\mathbb{B}$ leaves two closed intervals, $[0,\ 1/3]$ and $[1/3,\ 1]$. Let $D_{1/4}$ be the interval $(1/9,\ 2/9)$, which is the middle third of $[0,\ 1/3]$, and let $D_{3/4}$ be $(7/9,\ 8/9)$, which is the middle third of $[2/3,\ 1]$. Deleting those middle thirds, we are left with four closed intervals of length $1/9$ each. Let $D_{1/8}$, $D_{3/8}$, $D_{5/8}$, $D_{7/8}$ be the open intervals that make up the middle thirds of those closed intervals. Delete the middle thirds, and continue in this manner (see Figure 2). Then,

$$D = \bigcup_{q \in I} D_q$$

is the union of the countably many disjoint open intervals D_q, $q \in I$. It is an example of a non-trivial open set. Incidentally, note that the lengths of the D_q sum to

$$\frac{1}{3} + (\frac{1}{9} + \frac{1}{9}) + (\frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27}) + \cdots = 1.$$

Thus, the "length" of D is 1. But the set $C = \mathbb{B} \setminus D$ is not empty.

The set $C = \mathbb{B} \setminus D$ is called the *Cantor set*. It is obviously a closed set. The construction above shows that C is obtained by starting with $\mathbb{B}$ and deleting the middle third of every interval we can find. Thus, there is no open interval contained in C. That is, there are no open balls in C. Hence, the interior of C must be empty, and C is pure boundary:

$$C^\circ = \emptyset, \quad \bar{C} = C, \quad \partial C = C.$$

Also, since the length of D is equal to the length of $\mathbb{B}$, the length of $C = \mathbb{B} \setminus D$ must be 0. In summary, the Cantor set is very thin.

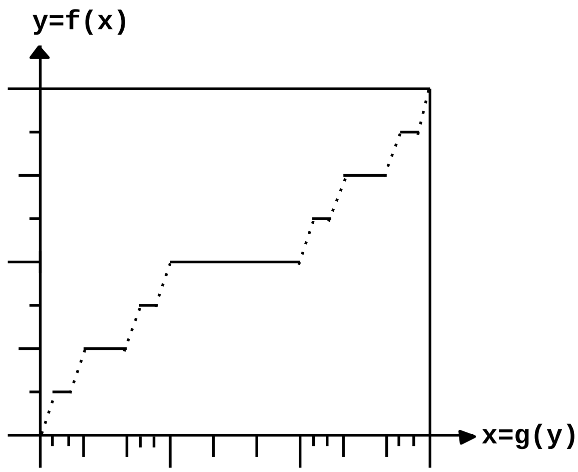

Figure 3: The cantor function.

Nevertheless, C has at least as many points as the interval $[0, 1]$. We prove this next by showing, via construction, that there exists an injection g from $[0, 1]$ into C.

To this end, we start by defining an increasing function f from D into $[0, 1]$ by letting

$$f(x) = q, \quad \text{if } x \in D_q.$$

Then, we define the function g on $[0, 1]$ by setting $g(1) = 1$ and

$$g(y) = \inf\{x \in D : f(x) > y\}, \quad 0 \leq y < 1.$$

We show first that $g(y) \in C$ for every y. This is obvious for $y = 1$. Let $y \in [0, 1)$; note that $g(y)$ is the infimum of the union of all intervals D_q with $q > y$; clearly, that infimum cannot belong to D; so $g(y)$ must belong to C (since it is obvious that $g(y) \in \mathbb{B}$). Finally, we show that $g : [0, 1] \mapsto C$ is an injection by showing that if $y < z$, then $g(y) < g(z)$. Fix $y < z$. Note that there is at least one q in I such that $y < q < z$, and the corresponding set D_q is contained in $\{x \in D : f(x) > y\}$ but not in $\{x \in D : f(x) > z\}$. It follows that the number $g(y)$ is to the left of the interval D_q whereas $g(z)$ is to the right. So, $g(y) < g(z)$ if $y < z$. Hence, $g : [0, 1] \mapsto C$ is an injection.

Exercises and Complements:

8.1 Let (E, d) be a metric space. Show that

$$\begin{aligned}
\bar{A} &= \{x \in E : d(x, A) = 0\} \\
A^\circ &= \{x \in E : d(x, A^c) > 0\} \\
\partial A &= \{x \in E : d(x, A) = 0 \text{ and } d(x, A^c) = 0\}.
\end{aligned}$$

8.2 Let (E, d) be a metric space. Fix $A \subset E$. Show that $A_\epsilon = \{x \in E : d(x, A) < \epsilon\}$ is an open set containing A for each $\epsilon > 0$. Show that $\bar{A} = \cap_{\epsilon>0} A_\epsilon$.

8.3 *Boundedness.* Let (E, d) be a metric space. Show that a subset A of E is bounded if and only if it is contained in some ball, that is, if and only if $A \subset B(x, r)$ for some x and r.

8.4 Take $E = \mathbb{R}$ and d the usual metric. Let $A \subset E$. Show that if A is closed and bounded above, then $\sup A$ belongs to A (that is, A has a maximum). Similarly, if A is closed and bounded below, then it has a minimum. Show that an open set A cannot have a minimum, that is, $\inf A$ cannot belong to A.

8.5 Let D be the open set of Example 8.9. Find its interior and boundary.

8.6 *Denseness.* A set D is said to be *dense* in E if $\bar{D} = E$. Let D be dense in E. Show that every x in E is at 0 distance from D. Thus, every open ball has at least one point of D. Show that the set $\mathbb{Q}$ of all rationals is dense in $\mathbb{R}$, the set of all pairs of rationals is dense in $\mathbb{R}^2$, etc.

8.7 *Separability.* The metric space E is said to be separable if there exists a countable set D that is dense in E. So, for example, the Euclidean spaces $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, ... are separable.

8.8 *Discrete metric spaces.* Let E be arbitrary and suppose that d is the discrete metric (see (7.1) for it) on E. Show that each subset A is both open and closed. For $r \leq 1$, every open ball $B(x, r)$ consists of exactly the point x. Note that $B(x, 1) = \{x\}$, $\bar{B}(x, 1) = E$ for every x (Moral: $\bar{B}(x, r)$ is not necessarily the closure of $B(x, r)$). If E is countable, then it is separable (trivially). If E is uncountable, it is not separable. Show this.

9 Convergence

Let (E, d) be a metric space. Our goal is to discuss the notion of convergence for a sequence of points in E. We do so by employing the concept of convergence in $\mathbb{R}$, for which we refer to Section 4 of Chapter .

9.1 DEFINITION. A sequence (x_n) in E is said to be *convergent* in E if there exists a point x in E such that $\lim d(x_n, x) = 0$. And, then, (x_n) is said to *converge* to x, the point x is called the *limit* of (x_n), and the notation $x = \lim x_n$ is used to indicate it.

REMARK: The preceding definition includes, implicit in it, the fact that a convergent sequence has exactly one limit. To see this, suppose that (x_n) converges to x and to y, that is, $\lim d(x_n, x) = 0$ and $\lim d(x_n, y) = 0$. Then,

$$0 \leq d(x, y) \leq d(x, x_n) + d(x_n, y)$$

by the triangle inequality, and the right side converges to zero. Thus, $d(x, y) = 0$, which means that $x = y$.

The following brings together a number of re-wordings of convergence. Each is a slight alteration of the others. No proof seems needed.

9.2 THEOREM. *The following statements are equivalent:*

1. (x_n) *converges to* x.
2. *For every* $\epsilon > 0$ *there is an* n_ϵ *such that* $d(x_n, x) < \epsilon$ *for all* $n \geq n_\epsilon$.
3. *The set* $\{n : d(x_n, x) \geq \epsilon\}$ *is finite for each* $\epsilon > 0$.
4. *For every* $\epsilon > 0$, *the ball* $B(x, \epsilon)$ *includes all but a finite number of the terms* x_n.

9.3 COROLLARY. *Every convergent sequence is bounded.*

PROOF. Let (x_n) be convergent and x its limit. In view of the equivalence of 1 and 4 in Theorem 9.2, $B(x, 1)$ includes all but a finite number of the terms x_n. Let r be the maximum of the distances from x to those terms x_n outside $B(x, 1)$, if there are any; otherwise, set $r = 1$. Clearly $r < \infty$ and $B(x, r)$ contains (x_n), which means that (x_n) is bounded. □

Subsequences

It follows from Theorem 9.2 that we may remove a finite number of terms, or rearrange the terms, without affecting the convergence. The following generalizes this.

9.4 PROPOSITION. *If a sequence converges to* x, *then every subsequence of it converges to the same* x.

PROOF. Let (x_n) be a sequence with limit x. Let (y_n) be a subsequence of it, that is, $y_n = x_{k_n}$ for some $k_1 < k_2 < \cdots$. Now, by Theorem 9.2, for every $\epsilon > 0$ the ball $B(x, \epsilon)$ includes all the terms x_n except for some finite number of them; therefore the same must be true for the terms y_n. So, by Theorem 9.2, the subsequence (y_n) converges to x. □

Convergence and Closed Sets

Think of a particle that moves in E by jumps: first it is at x_1, then at x_2, then at x_3, and so on. The following gives meaning to the term "closed set" if you think of sequences in this fashion.

9.5 THEOREM. *A set is closed if and only if it includes the limit of every sequence in it.*

PROOF. *"Only if" part.* Suppose that A is a closed set and that (x_n) is a sequence in A with limit x. We show that, then, x must belong to A. For, otherwise, if x were in A^c, there would exist an $\epsilon > 0$ such that $B(x, \epsilon) \subset A^c$ since A^c is open and $B(x, \epsilon)$ would include infinitely many terms since x is the limit, which would contradict the fact that all the x_n are in A.

"If" part. We show that if A is not closed then there is a sequence (x_n) in A that converges to some point x in A^c. Suppose that A is not closed. Then A^c is not open. Thus, there exists an x in A^c such that $B(x, r) \cap A$ has at least one point for each $r > 0$. Hence, for each n in $\mathbb{N}$, there is an x_n in A such that $d(x_n, x) < 1/n$. Obviously, (x_n) is in A and converges to x which is not in A. □

Exercises:

9.1 *Discrete metric spaces.* Suppose that d is the discrete metric on E. Show that (x_n) is convergent if and only if it is ultimately *stationary*, that is, if and only if it has the form $(x_1, x_2, \ldots, x_n, x, x, x, \ldots)$ for some n.

9.2 Let (E, d) be arbitrary. Show that if (x_n) converges to x and (y_n) converges to y, then $d(x_n, y_n)$ converges to $d(x, y)$. Hint: first show that, for arbitrary x, y, z in E,

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

Use this to write

$$\begin{aligned} |d(x_n, y_n) - d(x, y)| &\leq |d(x_n, y_n) - d(x_n, y)| \\ &\quad + |d(x_n, y) - d(x, y)| \\ &\leq d(y_n, y) + d(x_n, x), \end{aligned}$$

and take limits.

9.3 Show that if (x_n) converges to x, then $d(x_n, A)$ converges to $d(x, A)$ for each fixed subset A of E.

10 Completeness

Let (E, d) be a metric space. Recall that a sequence (x_n) in E is convergent if there is an x in E such that $\lim d(x_n, x) = 0$. This definition has two shortcomings. First, starting with (x_n), we rarely have a candidate x for the limit. Second, often we are not interested in computing the limit itself; it is generally sufficient to know that the limit exists and has such and such properties. This section is aimed at rectifying these shortcomings.

Cauchy Sequences

10.1 DEFINITION. A sequence (x_n) in E is said to be *Cauchy* if for every $\epsilon > 0$ there is an n_ϵ such that $d(x_m, x_n) < \epsilon$ for all $m > n \geq n_\epsilon$.

The following is nearly a re-statement of this definition in slightly more geometric terms.

10.2 LEMMA. *A sequence (x_n) is Cauchy if and only if for every $\epsilon > 0$ there is a ball of radius ϵ that contains all but finitely many of the terms x_n.*

PROOF. Suppose that (x_n) is Cauchy. Let $\epsilon > 0$. Then, there is n_ϵ such that $d(x_m, x_n) < \epsilon$ for all $m > n \geq n_\epsilon$. Thus, in particular, the ball $B(x_{n_\epsilon}, \epsilon)$ contains all the terms except possibly $x_1, \ldots, x_{n_\epsilon - 1}$. This proves the necessity of the condition.

Conversely, suppose that for every $\epsilon > 0$ there is a ball $B(x, \epsilon)$ with some x as its center such that all but a finite number of the terms are in the ball. Given $\epsilon > 0$, now pick x so that $B(x, \epsilon/2)$ contains all the x_n except perhaps finitely many, that is, there is n_ϵ such that $x_n \in B(x, \epsilon/2)$ for all $n \geq n_\epsilon$. Now, if $m > n \geq n_\epsilon$, then

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \epsilon/2 + \epsilon/2 = \epsilon.$$

Hence, (x_n) is Cauchy. This proves the sufficiency. □

10.3 THEOREM.

1. *Every convergent sequence is Cauchy.*
2. *Every Cauchy sequence is bounded.*
3. *Every subsequence of a Cauchy sequence is Cauchy.*

PROOF. The first claim is immediate from the preceding lemma and Theorem 9.2. The second claim is proved, via the preceding lemma, by following the proof of Corollary 9.3. The last claim is immediate from the preceding lemma. □

The following shows that if a sequence is Cauchy and you can find a subsequence of it that converges to some point x, then the original sequence converges to x.

10.4 PROPOSITION. *A Cauchy sequence that has a convergent subsequence is itself convergent.*

PROOF. Let (x_n) be Cauchy. Let x be the limit of a convergent subsequence of it. Pick $\epsilon > 0$. By Lemma 10.2, there is a ball $B(y, \epsilon)$ that contains all but a finite number of the x_n. That ball $B(y, \epsilon)$ must contain all but a finite number of the subsequence as well. Thus, x must be in $\bar{B}(y, \epsilon)$. Then, $B(x, 3\epsilon)$ contains $\bar{B}(y, \epsilon)$ and hence contains all but a finite number of the x_n. Thus, (x_n) is convergent and $x = \lim x_n$ in view of Theorem 9.2. □

Complete Metric Spaces

All the results above suggest that all Cauchy sequences should be convergent, which is in fact what we hope for. Unfortunately, this is not true in general. Here is an example.

Suppose that $E = \mathbb{Q}$, the set of all rationals, with the metric it inherits from the real line. Let $x = \sqrt{2}$, which is not a rational number, and let (x_n) be a sequnce in $\mathbb{Q}$ that converges to x in the sense of convergence in $\mathbb{R}$: for instance, pick x_n to be a rational number in the interval $(x, x + 1/n)$ for each n. Over the metric space $\mathbb{Q}$, the sequence (x_n) is Cauchy, but fails to be convergent in $\mathbb{Q}$ simply because x is not in $\mathbb{Q}$. The problem here is not with the Cauchy sequence, but with the space $\mathbb{Q}$. The space $\mathbb{Q}$ has holes in it!

The following introduces the extra notion we want.

10.5 DEFINITION. The metric space (E, d) is said to be *complete* if every Cauchy sequence in E converges to a point of E.

The following is immediate from Theorem 9.5.

10.6 PROPOSITION. *If (E, d) is complete and $D \subset E$ is closed, then (D, d) is a complete metric space.*

The following shows that familiar spaces are complete. Other examples are listed in exercises.

10.7 THEOREM. *Every Euclidean space is complete.*

PROOF. We start with the one-dimensional Euclidean space, namely $\mathbb{R}$. Let $(x_n) \subset \mathbb{R}$ be Cauchy. Then, for every $\epsilon > 0$ there is a ball of radius ϵ (namely an open interval of length 2ϵ) that contains all but finitely many of the x_n. Therefore, the numbers $x = \liminf x_n$ and $y = \limsup x_n$ must belong to that ball, which means that $0 \leq y - x < 2\epsilon$. Since this is true for every $\epsilon > 0$, we must have $x = y$, that is, (x_n) is convergent. This proves that $\mathbb{R}$ is complete.

Now, fix $k \geq 2$ and consider the Euclidean space $\mathbb{R}^k$. We write $x = (a, b, \ldots, c)$ for each x in $\mathbb{R}^k$ for simplicity of notation (in other words, the coordinates of x are $a, b, \ldots, c$).

Consider a Cauchy sequence of points $x_n = (a_n, b_n, \ldots, c_n)$ in $\mathbb{R}^k$. Given $\epsilon > 0$, then, for all m and n large enough, we have

$$d(x_m, x_n) = (|a_m - a_n|^2 + |b_m - b_n|^2 + \cdots + |c_m - c_n|^2)^{1/2} < \epsilon,$$

which shows that

$$|a_m - a_n| < \epsilon, \quad |b_m - b_n| < \epsilon, \quad \ldots, |c_m - c_n| < \epsilon.$$

In other words, the sequences (a_n), (b_n), ..., (c_n) in $\mathbb{R}$ are Cauchy. We have just shown that $\mathbb{R}$ is complete. So, these sequences must be convergent in $\mathbb{R}$, say, with limits $a, b, \ldots, c$ respectively. Now, let $x = (a, b, \ldots, c)$ and note that

$$d(x_n, x)^2 = |a_n - a|^2 + |b_n - b|^2 + \cdots + |c_n - c|^2$$

converges to 0. Hence, $\lim d(x_n, x) = 0$, and (x_n) is convergent. This completes the proof that $\mathbb{R}^k$ is complete. $\square$

Exercises and Complements:

10.1 Show that the following metric spaces are complete:

1. $E = \mathbb{R}^2$ with the Manhattan metric d.
2. E arbitrary, d is the discrete metric.

In fact, each $\mathbb{R}^k$ is a complete metric space with any one of the metrics d_p introduced in Exercises 7.2 and 7.4.

10.2 Show that the space l^2 introduced in Exercise 7.5 is complete. Incidently, so is the space $\mathcal{C}$ of Example 7.3 and Exercise 7.6.

10.3 Two Cauchy sequences (x_n) and (y_n) are said to be equivalent if their merger $(x_1, y_1, x_2, y_2, \ldots)$ is Cauchy. In this case, we write $(x_n) \equiv (y_n)$. Show that this defines an requivalence relation. That is,

1. $(x_n) \equiv (x_n)$
2. $(x_n) \equiv (y_n)$ implies that $(y_n) \equiv (x_n)$
3. $(x_n) \equiv (y_n)$, $(y_n) \equiv (z_n)$ implies that $(x_n) \equiv (z_n)$.

11 Compactness

Let (E, d) be a metric space. It will be convenient to refer to E as a metric space, without mentioning d. We shall use the picturesque phrase "the collection $\{A_i : i \in I\}$ covers B" to mean that $\cup_{i \in I} A_i \supset B$.

11.1 DEFINITION. A set $C \subset E$ is said to be *compact* if every collection of open sets that covers C has a finite sub-collection that covers C. The metric space (E, d) is said to be compact if E is so.

We shall show that, for many metric spaces, compact sets are precisely the sets that are bounded and closed. The following are aimed in that direction. The proofs are excessively detailed in order to facilitate understanding.

11.2 PROPOSITION. *Every compact set is bounded.*

PROOF. Let C be compact. For each x in C, let B_x be a ball of radius 1 centered at x. Obviously, then, the collection $\{B_x : x \in C\}$ of open sets covers C. Hence, there must be a finite sub-collection, say of sets $B_{x_1}, \ldots, B_{x_n}$, that covers C. Since the union of balls $B_{x_1}, \ldots, B_{x_n}$ must be bounded, this implies that C is bounded as well. □

11.3 PROPOSITION. *Every closed subset of a compact set is compact.*

PROOF. Let D be compact. Let $C \subset D$ be closed. Fix a collection of open sets that covers C. Adding the open set $E \setminus C$ to that collection, we obtain a collection of open sets that covers D. Since D is compact, the latter collection has a finite sub-collection that still covers D. Removing $E \setminus C$ from that sub-collection (if it were in), we obtain a finite sub-collection of the original collection that covers C. Thus, C must be compact. □

Compact Subspaces

Recall that every subset D of E can be regarded as a metric space by itself, with the metric it inherits from E. Whether D is open or not as a subset of E, it is open automatically when it is regarded as a metric space. The concept of compactness does not suffer from such foolishness.

11.4 PROPOSITION. *A set D is compact as a metric space if and only if it is compact as a subset of E.*

PROOF. A subset of D is an open ball in the space D if and only if it has the form $B \cap D$ for some open ball B of the space E. Since an open set is the union of all the open balls it contains, it follows that A is an open subset of the space D if and only if $A = B \cap D$ for some open subset B of the space E. Now, the definition of compactness does the rest. □

Cluster Points, Convergence, Completeness

This is to look into the connections between compactness and convergence.

11.5 DEFINITION. A point x in E is called a *cluster point* [2] of a subset A of E provided that every open ball centered at x contains infinitely many points of A.

11.6 THEOREM. *Every infinite subset of a compact set has at least one cluster point in that compact set.*

PROOF. We shall show that if C is compact, and $A \subset C$, and A has no cluster point in C, then A is finite. Let A and C be such. Since no x in C is a cluster point of A, for every x in C there is an open ball $B(x, r)$ that contains only finitely many points of A. Those open balls cover C obviously. Since C is compact, there must be a finte number of them that cover C and, therefore, A. Since each one of those finitely many balls has a finte number of points of A, the total number of points in A must be finite. □

The following is the way compactness helps in discussing convergence. In particular, together with Proposition 10.4, it shows that every Cauchy sequence in a compact set is convergent.

11.7 THEOREM. *Every sequence in a compact set has a subsequence that converges to some point of that set.*

PROOF. Let C be compact. Let $(x_n) \subset C$. If the set $A = \{x_1, x_2, \ldots\}$ is finite, then at least one point of A, say x, appears infinitely often in the sequence, and hence $(x, x, \ldots)$ is a subsequence, which obviously converges to $x \in A \subset C$. Now suppose that A is infinite. By the preceding theorem, then A has a cluster point x in C. Since each one of the balls $B(x, 1/n)$, $n = 1, 2, \ldots$, has infinitely many points in C, we may pick k_1 so that x_{k_1} is in $B(x, 1)$, pick $k_2 > k_1$ so that x_{k_2} is in $B(x, 1/2)$, pick $k_3 > k_2$ so that x_{k_3} is in $B(x, 1/3)$, and so on. Obviously, (x_{k_n}) converges to x. □

[2] Other terms in common use include limit point, adherence point, point of accumulation, etc.

11.8 COROLLARY. *Every compact set is closed.*

PROOF. Let C be compact. The preceding theorem implies that every convergent sequence in C converges to a point of C. Thus, C is closed by Theorem 9.5. $\square$

11.9 COROLLARY. *Every compact metric space is complete. Every Cauchy sequence in a compact metric space is convergent.*

PROOF. The second statement is immediate from Theorem 11.7 and Proposition 10.4. The first follows from the second by the definition of completeness. $\square$

Compactness in Euclidean Spaces

We have seen that, for an arbitrary metric space, every compact set is bounded and closed (Proposition 11.2 and Corollary 11.8). In the case of Euclidean spaces, the converse is true as well. This is called the *Heine-Borel Theorem*.

11.10 THEOREM. *A subset of a Euclidean space is compact if and only if it is bounded and closed.*

We start by listing an auxiliary result that is trivial at least for $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$. We omit its proof.

11.11 LEMMA. *Let B be a bounded subset of a Euclidean space E. Then, for every $\epsilon > 0$ there is a finite collection of closed balls of radius ϵ that covers B.*

Here is the proof of Theorem 11.10.

PROOF. As mentioned above, 11.2 and 11.8 prove the necessity part. We now prove the sufficiency of the condition.

Let E be a Euclidean space and let C be a closed and bounded subset of E. Suppose that C is not compact. Then, there is a collection $\{A_i : i \in I\}$ of open sets that covers C but is such that

11.12 no finite sub-collection $\{A_i : i \in I\}$ covers C.

(a) Let $\epsilon = 1/2$. By the preceding lemma, we can find a finite number m of closed balls $B_1, \ldots, B_m$ of radius ϵ that cover C. Then, $C = (C \cap B_1) \cup \cdots \cup (C \cap B_m)$. In view of (11.12), at least one of $C \cap B_1, \ldots, C \cap B_m$ cannot ever be covered by a finite sub-collection of the A_i; let that one be denoted by C_1. Now, C_1 is closed, its diameter is at most $2\epsilon = 1$ (since the B_k have diameter 1), and (11.12) is true for C_1.

(b) Applying the arguments of the preceding paragraph with $\epsilon = 1/4$ to the set C_1 we get a new set $C_2 \subset C_1$ that is closed, has diameter at most $1/2$, and (11.12) holds for C_2. Repeating this with $\epsilon = 1/6,\ 1/8,\ 1/10, \ldots$ we obtain further sets $C_3, C_4, C_5, \ldots$ with the same properties but with diameters at most $1/3,\ 1/4,\ 1/5, \ldots$. Clearly $C_1 \supset C_2 \supset C_3 \supset \cdots$.

(c) Since (11.12) holds for each C_n, it must be that no C_n is empty (covering an empty set takes no effort). Thus, we may pick x_1 from C_1, x_2 from C_2, and so on to obtain a sequence (x_n).

(d) This sequence is Cauchy: given $\epsilon > 0$ choose n so that $1/2n < \epsilon$, and then $x_n, x_{n+1}, \ldots$ are all in a ball of radius ϵ since all these terms are in C_n which has diameter less than $1/n$. Since E is Euclidean, it is complete (see Theorem 10.7), which means that every Cauchy sequence converges. Hence, the sequence (x_n) converges to some point x_0 in E. Since, for each n, $(x_m : m \geq n) \subset C_n$ and C_n is closed, the limit x_0 belongs to C_n by Theorem 9.5.

(e) Since the A_i cover C, there must exist an i in I such that x_0 is in A_i. Fix that i. Since A_i is open, there is an $\epsilon > 0$ such that

$$B(x_0, \epsilon) \subset A_i.$$

Now choose n large enough that $1/n < \epsilon/2$. Since, $x_0 \in C_n$ and diam $C_n \leq 1/n < \epsilon/2$, we see that

$$C_n \subset B(x_0, \epsilon).$$

In other words, A_i covers C_n. This contradicts the earlier assertion that (11.12) holds for all C_n. This completes the proof. □

Exercises:

11.1 *Supremums.* Let A be a non-empty subset of $\mathbb{R}$. Suppose that A is bounded above but has no greatest element. Show that, then, $\sup A$ is a cluster point of A.

11.2 Show that the union of a finite number of compact sets is again compact.

11.3 Give an example of an infinite subset of $\mathbb{R}$ that has no cluster points. Give an example of one with exaclty two cluster points. Identify the cluster points of the set

$$A = \{x \in \mathbb{R} : x = \frac{1}{m} + \frac{1}{n} \text{ for some } m, n \text{ in } \mathbb{N}\}.$$

11.4 *Sequences in* $\mathbb{R}$. By the Heine-Borel theorem, every closed interval $[a, b] \subset \mathbb{R}$ is compact. Thus, every bounded sequence in $\mathbb{R}$ has a convergent subsequence (cf. Theorem 11.7). Another consequence is the following useful result:

Let (x_n) be a bounded sequence in $\mathbb{R}$. Suppose that all convergent subsequences of it have the same limit x. Then, (x_n) converges to x.

Prove this by following the steps below.

(a) Show that $\underline{x} = \liminf x_n$ and $\bar{x} = \limsup x_n$ are cluster points of (x_n).

(b) Show that there is a subsequence of (x_n) that converges to $\underline{x}$. Similarly, then, there is a subsequence that converges to $\bar{x}$.

(c) By the hypothesis that all convergent subsequences have the same limit, we conclude that $\underline{x} = \bar{x}$, which means that $\lim x_n$ exists (and is in $\mathbb{R}$ since (x_n) is bounded).

Functions on Metric Spaces

Elementary analysis is mostly about functions from $\mathbb{R}$ into $\mathbb{R}$, or functions from $\mathbb{R}^n$ into $\mathbb{R}$, or, somewhat more generally, functions from $\mathbb{R}^n$ into $\mathbb{R}^m$. Our aim is to consider functions from one metric space to another. Replacing Euclidean spaces by metric spaces introduces no new difficulties and is immensely useful for dealing with various problems concerning differential and integral equations.

For mappings from a metric space to another we employ either notations like T, S, U or notations like f, g, h. Generally, the transformation notation is cleaner: we write Tx for the image of x under T and $T^{-1}B$ for the inverse image of B, which become $f(x)$ and $f^{-1}(B)$ in the standard function notation.

12 Continuous Mappings

Throughout this section, E, E', etc. will be metric spaces with corresponding metrics d, d', etc. Given a mapping T from E into E', we write Tx for the image of the point x of E and $T^{-1}B$ for the inverse image of the subset B of E'. On a first reading, the reader may wish to take $E' = \mathbb{R}$ and $d'(x, y) = |x - y|$ as usual.

12.1 DEFINITION. A mapping $T : E \mapsto E'$ is said to be *continuous at the point* x of E provided that for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$y \in E, \quad d(x, y) < \delta \quad \Rightarrow \quad d'(Tx, Ty) < \epsilon.$$

The mapping T is said to be *continuous* if it is continuous at every x of E.

REMARKS: (a) In the definition, δ is allowed to depend on ϵ and x.

(b) When $E = E' = \mathbb{R}$ with the usual metric, the preceding is the classical definition of continuity.

(c) The condition for T to be continuous at x can be rephrased in more geometric terms as follows: for every $\epsilon > 0$ there is a $\delta > 0$ such that T maps the open ball $B(x, \delta)$ of E into the open ball $B'(Tx, \epsilon)$ of E'. Here,

$$B(x, \delta) = \{y \in E : d(x, y) < \delta\}, \quad B'(Tx, \epsilon) = \{y \in E' : d'(Tx, y) < \epsilon\}.$$

Continuity and Open Sets

12.2 THEOREM. *A mapping $T : E \mapsto E'$ is continuous if and only if $T^{-1}B$ is an open subset of E for every open subset B of E'.*

PROOF. Suppose that T is continuous. Let $B \subset E'$ be open. We want to show that, then, $A = T^{-1}B$ is open, that is, for every x in A there is $\delta > 0$ such that $B(x, \delta) \subset A$. To this end, fix x in A, note that $y = Tx$ is in B, and therefore, there is $\epsilon > 0$ such that $B'(y, \epsilon) \subset B$ (since B is open). By the continuity of T, for that ϵ, there is a $\delta > 0$ such that T maps $B(x, \delta)$ into $B'(y, \epsilon)$. Since $B'(y, \epsilon) \subset B$, we have $B(x, \delta) \subset A$ as needed.

Suppose that $T^{-1}B$ is open in E for every open subset B of E'. Let x in E be arbitrary. We want to show that, then, T is continuous at x. To this end, fix $\epsilon > 0$. Since $B'(Tx, \epsilon)$ is open, its inverse image is open, that is $A = T^{-1}B'(Tx, \epsilon)$ is an open subset of E. Note that x is in A; therefore, there is a $\delta > 0$ such that $B(x, \delta) \subset A$, and then T maps $B(x, \delta)$ into $B'(Tx, \epsilon)$. So, T is continuous at x. □

Continuity and Convergence

If (x_n) is a sequence in E, we write $x_n \xrightarrow{d} x$ to mean that (x_n) converges to x in E in the metric d, that is, $d(x_n, x) \to 0$. Similarly, we write $y_n \xrightarrow{d'} y$ to mean that the sequence (y_n) in E' converges to y in the metric d'. The following is probably the most useful characterization of continuity.

12.3 THEOREM. *A mapping $T : E \mapsto E'$ is continuous at the point x of E if and only if*

$$(x_n) \subset E, \quad x_n \xrightarrow{d} x \quad \Rightarrow \quad Tx_n \xrightarrow{d'} Tx.$$

PROOF. Suppose that T is continuous at x. Let $(x_n) \subset E$ be such that $x_n \xrightarrow{d} x$. We want to show that, then, $Tx_n \xrightarrow{d'} Tx$, which is equivalent to showing that for every $\epsilon > 0$ the ball $B'(Tx, \epsilon)$ contains all but finitely many of the points Tx_n. To this end, fix $\epsilon > 0$. By the continuity of T at x, there is $\delta > 0$ such that T maps $B(x, \delta)$ into $B'(Tx, \epsilon)$. Since $x_n \in B(x, \delta)$ for all but finitely many n, it follows that $Tx_n \in B'(Tx, \epsilon)$ for all but finitely many n, which is as desired.

Suppose that T is not continuous at x. Then, there is $\epsilon > 0$ such that for every $\delta > 0$ there is y in E such that $d(x, y) < \delta$ and $d'(Tx, Ty) \geq \epsilon$. Thus, for that ϵ, taking $\delta = 1,\ 1/2,\ 1/3,\ \ldots$ we can pick $y = x_1, x_2, x_3, \ldots$ such that $d(x_n, x) < 1/n$ and $d'(Tx_n, Tx) \geq \epsilon$. Hence, there is a sequence $(x_n) \subset E$ such that $x_n \xrightarrow{d} x$ but (Tx_n) does not converge to Tx. □

Compositions

The following result is recalled best by the phrase "a continuous function of a continuous function is continuous".

12.4 THEOREM. *If $T : E \mapsto E'$ is continuous at $x \in E$ and $S : E' \mapsto E''$ is continuous at $Tx \in E'$, then $S \circ T : E \mapsto E''$ is continuous at $x \in E$. If T is continuous and S is continuous, then $S \circ T$ is continuous.*

PROOF. The second assertion is immediate from the first. To show the first, let $(x_n) \subset E$ be such that $x_n \xrightarrow{d} x$. If T is continuous at x, the $Tx_n \xrightarrow{d'} Tx$ by the last theorem; and if S is continuous at Tx, this in turn implies that $S(Tx_n) \xrightarrow{d''} S(Tx)$ by the last theorem again, which means that $S \circ T$ is continuous at x. □

EXAMPLES.

12.5 *Constants.* Let $T : E \mapsto E'$ be defined by $Tx = b$ where b in E' is fixed. This T is continuous.

12.6 *Identity.* Let $T : E \mapsto E$ be defined by $Tx = x$. This T is continuous, as is easy to see from Theorem 12.2 or 12.3.

12.7 *Restrictions.* Let $T : E \mapsto E'$ be continuous. For $D \subset E$, the restriction of T to D is the mapping $S : D \mapsto E'$ defined by putting $Sx = Tx$ for each $x \in D$. Obviously, the continuity of T implies that of S.

12.8 *Discontinuity.* Let $f : \mathbb{R} \mapsto \mathbb{R}$ be defined by setting $f(x) = 1$ if x is rational and $f(x) = 0$ if x is irrational. This function is discontinuous at every $x \in \mathbb{R}$. To see it, fix x in $\mathbb{R}$. For every $\delta > 0$, the ball $B(x, \delta)$ has infinitely many rationals and infinitely many irrationals. Thus, it is impossible to satisfy the condition for continuity at x (for any $\epsilon < 1$).

12.9 *Lipschitz continuity.* A mapping $T : E \mapsto E'$ is said to satisfy a Lipschitz condition if there exists a constant $K \in (0, \infty)$ such that

$$d'(Tx, Ty) \leq Kd(x, y)$$

for all x, y in E. Every such mapping is continuous: given $\epsilon > 0$, choose $\delta = \epsilon/K$ no matter what x is.

12.10 *Coordinate mappings.* Let $E = \mathbb{R}^n$, the n-dimensional Euclidean space, fix i in $\{1, \ldots, n\}$, and define $P_i : \mathbb{R}^n \mapsto \mathbb{R}$ by $P_i x = x_i$, the i^{th} coordinate of x. Then, P_i satisfies the Lipschitz condition above with $K = 1$ and, thus, is continuous.

Real-Valued Functions

Functions f from a metric space E into $\mathbb{R}$ can be combined through arithmetic operations to obtain new functions. For instance, $f + g$ is the function whose value at x is $f(x) + g(x)$. In defining f/g, however, one must exercise some caution at points x where $g(x) = 0$. It is best to limit the definition of f/g to the set $\{x \in E : g(x) \neq 0\}$. The following is immediate from Theorem 12.3.

12.11 PROPOSITION. *If $f : E \mapsto \mathbb{R}$ and $g : E \mapsto \mathbb{R}$ are continuous, then so are $f + g$, $f - g$, $f \cdot g$, f/g except that, in the last case, f/g should be treated as a function on $\{x : g(x) \neq 0\}$.*

$\mathbb{R}^n$-Valued Functions

These are functions from a metric space E into the Euclidean space $\mathbb{R}^n$ (with the Euclidean distance). The following reduces the notion of continuity for such mappings to the case of real-valued functions. We use the projection mappings P_i introduced in Example 12.10: $P_i x$ is the i^{th} coordinate of the vector x in $\mathbb{R}^n$.

12.12 PROPOSITION. *A mapping $T : E \mapsto \mathbb{R}^n$ is continuous if and only if the mappings $P_1 \circ T, \ldots, P_n \circ T$ from E into $\mathbb{R}$ are continuous.*

PROOF. Let T be continuous. Then, $P_i \circ T$ is continuous for each i because a continuous function of a continuous function is continuous.

Suppose that $P_1 \circ T, \ldots, P_n \circ T$ are continuous. To show that, then, T is continuous, we start by observing that

12.13 $$\|u - v\| = \sqrt{\sum_1^n |P_i u - P_i v|^2}, \quad u, v \in \mathbb{R}^n.$$

Now, fix $x \in E$ and $\epsilon > 0$. Using the definition of continuity for $P_i \circ T$ at x with $\epsilon_i = \epsilon/\sqrt{n}$, we find $\delta_i > 0$ such that

$$d(x, y) < \delta_i \Rightarrow |P_i T x - P_i T y| < \epsilon/\sqrt{n}.$$

Let $\delta = \min\{\delta_1, \dots, \delta_n\}$. Then $\delta > 0$ and

$$\begin{aligned} d(x,y) < \delta \quad &\Rightarrow \quad |P_iTx - P_iTy| < \epsilon/\sqrt{n} \text{ for each } i \\ &\Rightarrow \quad \|Tx - Ty\| < \epsilon \end{aligned}$$

in view of 12.13 used with $u = Tx$ and $v = Ty$. □

Exercises:

12.1 *Continuity of metrics.* Recall the definition of the product space $E \times E$ from Exercise 7.9 in Chapter with $(E_1, d_1) = (E_2, d_2) = (E, d)$. Show that $d : E \times E \mapsto \mathbb{R}_+$ is continuous.

12.2 *Continuity of pairs.* Let $f : E \mapsto E'$ and $g : E \mapsto E'$ be continuous. Define $h : E \mapsto E' \times E'$ by $h(x) = (f(x), g(x))$. Show that h is continuous.

12.3 *Closed sets.* If $T : E \mapsto E'$ is continuous, then $T^{-1}B$ is a closed subset of E for every closed subset B of E'. Show. For $f : E \mapsto \mathbb{R}$ continuous, show that the sets $\{x \in E : f(x) \le b\}$, $\{x \in E : f(x) = b\}$, $\{x \in E : f(x) \ge b\}$ are closed in E.

12.4 *Indicators.* For $A \subset E$ let 1_A be the indicator of A, that is, $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ if $x \notin A$. Show that 1_A is continuous at all points $x \in E$ except for $x \in \partial A$.

12.5 *Left and Right Continuity.* Let $f : \mathbb{R} \mapsto E'$. Order properties of the real line enable us to refine the notion of continuity as follows. The function f is said to be *right-continuous* at $x \in \mathbb{R}$ provided that $f(x_n) \xrightarrow{d'} f(x)$ for every decreasing sequence $(x_n) \subset \mathbb{R}$ with limit x. Similarly, f is said to be *left-continuous* at x if $f(x_n) \xrightarrow{d'} f(x)$ for every increasing sequence (x_n) with limit x.

Show that f is continuous at x if and only if it is both right-continuous and left-continuous at x.

12.6 *Functional inverses.* Let $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be a continuous and strictly increasing bijection. Let $f^{-1}(y)$ be that point x for which $f(x) = y$. Show that the function f^{-1} is continuous and strictly increasing.

12.7 *Legendre Transforms.* A function $f : \mathbb{R} \mapsto \mathbb{R}$ is called *convex* if

$$f(px + qy) \le pf(x) + qf(y)$$

for all $x, y \in \mathbb{R}$ and all $p, q \in (0,1)$ satisfying $p + q = 1$. The *Legendre transform* of a convex function f is the function $g : \mathbb{R} \mapsto \mathbb{R}$ defined by

$$g(y) = \max_x (xy - f(x)).$$

Show that g is convex and that

$$f(x) = \max_y (xy - g(y)).$$

State any extra "smoothness" assumptions you might need.

12.8 *Sections.* Let $f : E_1 \times E_2 \mapsto \mathbb{R}$ be continuous. Show that, for each y in E_2, the mapping $x \mapsto f(x, y)$ from E_1 into $\mathbb{R}$ is continuous. Similarly, $y \mapsto f(x, y)$ is continuous for each x. Unfortunately, the converse does not hold: it is possible to have $x \mapsto f(x, y)$ continous for each y and $y \mapsto f(x, y)$ continuous for each x even though f is not continuous. Give an example of such a function.

13 Compactness and Uniform Continuity

As before, E, E', etc. are metric spaces with metrics d, d', etc. This section is on the effect of compactness on continuity.

13.1 THEOREM. *Let $T : E \mapsto E'$ be continuous. If E is compact, then the range of T is a compact subset of E'.*

PROOF. Let $D \subset E'$ be the range of T. Assuming that E is compact, we need to show that D is compact. Let $\{B_i : i \in I\}$ be a collection of open subsets of E' that covers D. Then, the continuity of T implies via Theorem 12.2 that the sets $A_i = T^{-1}B_i$, $i \in I$, are open. Moreover, $\{A_i : i \in I\}$ covers E: if x is in E then Tx is in D, and hence, Tx is in B_i for some i, which implies that x is in the corresponding A_i. Now the compactness of E implies that there exists a finte set $J \subset I$ such that $\{A_i : i \in J\}$ covers E. Thus, if $x \in E$, then $x \in A_i$ for some i in J and therefore $Tx \in B_i$ for some i in J. That is, $\{B_i : i \in J\}$ covers D. So, D must be compact. □

Recall that every compact set is closed and bounded. Thus, if $f : E \mapsto \mathbb{R}$ is continuous and E is compact, then the range of f is bounded and closed, which implies that f attains a maximum and a minimum, that is, there are x_0 and x_1, such that $f(x_0) \le f(x) \le f(x_1)$ for all $x \in E$ (see Exercise 11.1 in Chapter to the effect that if $D \subset \mathbb{R}$ is closed and bounded then $\inf A$ and $\sup A$ belong to D). We have thus shown the following:

13.2 COROLLARY. *Let E be compact and $f : E \mapsto \mathbb{R}$ continuous. Then, f is bounded and attains a maximum and a minimum.*

The conclusion fails if E is not compact. For instance, $f(x) = x$ on $E = (0, 1)$ is bounded but has neither a maximum nor a minimum. Also, $f(x) = 1/x$ on $E = (0, 1)$ is not bounded and has neither a maximum nor a minimum.

Uniform Continuity

Recall the definition of continuity: $T : E \mapsto E'$ is continuous provided that for every x in E and every $\epsilon > 0$ there is a $\delta > 0$ (depending on x and ϵ) such that $d(x, y) < \delta$ implies $d'(Tx, Ty) < \epsilon$ for all y in E. The importance of the following is to remove the dependence of δ on x.

13.3 DEFINITION. A mapping $T : E \mapsto E'$ is said to be *uniformly continuous* provided that for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$x, y \in E, \quad d(x, y) < \delta \quad \Rightarrow \quad d'(Tx, Ty) < \epsilon.$$

Obviously, every uniformly continuous function is continuous. The converse is false. For example, the function $f : (0, 1) \mapsto \mathbb{R}$ defined by $f(x) = 1/x$ is continuous but not uniformly so. The failure here is not due to the unboundedness of f. For instance, the function $f : (0, 1) \mapsto [-1, 1]$ defined by $f(x) = \sin 1/x$ is continuous but not uniformly so. The mappings of Examples 12.5, 12.6, 12.9, and 12.10 are uniformly continuous. In fact, they are all special cases of 12.9 on Lipschitz continuity. Being Lipschitz almost encapsulates the notion of uniform continuity

13.4 PROPOSITION. *Let $T : E \mapsto E'$ be Lipschitz continuous. Then T is uniformly continuous.*

PROOF. Fix $\epsilon > 0$ and choose $\delta = \epsilon/K$. This δ works and is independent of x. □

(Exercise 13.6 provides an "almost converse" to this result). The following shows the important role of compactness on uniform continuity.

13.5 THEOREM. *Let $T : E \mapsto E'$ be continuous. If E is compact, then T is uniformly continuous.*

PROOF. Fix $\epsilon > 0$. We search for $\delta > 0$ that will fulfill the condition for uniform continuity. Since T is continuous, for each x in E there is $\delta(x) > 0$ such that

13.6 $$d(x, y) < \delta(x) \Rightarrow d'(Tx, Ty) < \epsilon/2.$$

The collection of open balls $B(x, \delta(x)/2)$, $x \in E$, covers E. Since E is compact, there must exist a finite number of them, say those corresponding to $x_1, \ldots, x_n$, that cover E. Define

$$\delta = \frac{1}{2} \min\{\delta(x_1), \ldots, \delta(x_n)\}.$$

Then, $\delta > 0$ and it remains to show that this δ works. Let x, y in E be arbitrary and suppose that $d(x, y) < \delta$. By the way the $x_1, \ldots, x_n$ are chosen, there is an i such that x is in $B(x_i, \delta(x_i)/2)$, that is,

$$d(x, x_i) < \frac{1}{2}\delta(x_i).$$

Moreover, for the same i,

$$d(y, x_i) \leq d(y, x) + d(x, x_i) \leq \delta + \frac{1}{2}\delta(x_i) \leq \delta(x_i).$$

Thus, $d(x, x_i) < \delta(x_i)$ and $d(y, x_i) < \delta(x_i)$, which by 13.6 imply that

$$d'(Tx, Ty) < \epsilon/2, \text{ and } d'(Ty, Tx_i) < \epsilon/2.$$

Thus, $d'(Tx, Ty) < \epsilon$ by the triangle inequality. □

Exercises:

13.1 *Metrics.* Show that, for fixed x_0 in E, the function $x \mapsto d(x, x_0)$ from E into $\mathbb{R}_+$ is uniformly continuous.

13.2 *Compositions.* Let $T : E \mapsto E'$ and $S : E' \mapsto E''$ be uniformly continuous. Show that, then, $S \circ T : E \mapsto E''$ is uniformly continuous.

13.3 *Homeomorphisms.* Recall that for a bijection $f : E \mapsto E'$ we define the functional inverse f^{-1} by setting $f^{-1}(y) = x$ if and only if $f(x) = y$. A *homeomorphism* from E onto E' is a bijection that is continuous and whose functional inverse is also continuous. Incidentally, two spaces E and E' are said to be *homeomorphic* if there exists a homeomorphism from one to the other. Compactness helps in checking for homeomorphisms. Show that if $f : E \mapsto E'$ is a continuous bijection and E is compact, then f is a homeomorphism.

13.4 *Extensions.* Let D be dense in E (see Exercise 8.6 in Chapter for the definition). Note that this means that every point of $E \setminus D$ is a cluster point of D. Suppose that $f : D \mapsto \mathbb{R}$ is uniformly continuous. Show that, then, there exists a unique continuous function $\bar{f} : E \mapsto \mathbb{R}$ such that $\bar{f}(x) = f(x)$ for all x in D. Then, $\bar{f}$ is called the *continuous extension* of f onto E.

13.5 *Cantor function.* Let $E = [0, 1]$, and C be the Cantor set, and $D = E \setminus C$; see Example 8.9 in Chapter . Note that D is dense in E, since C has no open intervals contained in it.

Show that the function f constructed in 8.9 of Chapter is a uniformly continuous function from D into $[0, 1]$. By the preceding exercise, then,

f has a continuous extension $\bar{f}$ onto $E = [0, 1]$. In fact, $\bar{f}$ is uniformly continuous (why?).

The function $\bar{f}$ is called the *Cantor function*. It is increasing and continuous. Its derivative exists at every x in D and is equal to 0. So, although $\bar{f}$ increases from 0 to 1 in a continuous fashion, all its increase is on the set C, and C has "length" 0.

13.6 *Lipschitz Continuity.* A mapping $T : \mathbb{R}^n \mapsto \mathbb{R}$ is uniformly continuous if and only if for every $\epsilon > 0$ there exists K_ϵ such that

$$|Tx - Ty| \leq K_\epsilon \cdot \|x - y\| + \epsilon$$

for all x and y in $\mathbb{R}^n$. Prove this.

Hints: (a) The "if" part is easy. Choose

$$\delta = \frac{\epsilon/2}{K_{\epsilon/2}}.$$

(b) For the "only if" part: fix $\epsilon > 0$ and x and y; choose a chain of points $x = x_0, x_1, x_2, \ldots, x_m = y$ with distances $\|x_i - x_{i+1}\| < \delta$; ask, how many such points do we need, and note that

$$|Tx - Ty| \leq \sum_1^m |Tx_i - Tx_{i+1}| \leq n\epsilon;$$

figure out m needed and then what K_ϵ should be.

14 Sequences of Functions

Let E and E' be metric spaces with respective metrics d and d'. Let (T_n) be a sequence of mappings from E into E'.

14.1 DEFINITION. The sequence (T_n) is said to *converge pointwise* to a mapping $T : E \mapsto E'$ provided that the sequence $(T_n x)$ converges to Tx in E' for each point x in E.

In other words, for each x in E, we must have

14.2
$$\lim_n d'(T_n x, Tx) = 0,$$

that is, for every $\epsilon > 0$ there must be an $n_{\epsilon,x}$ such that $d'(T_n x, Tx) < \epsilon$ for all $n \geq n_{\epsilon,x}$. If $n_{\epsilon,x}$ can be chosen to be free of x, we obtain the following stronger concept of convergence:

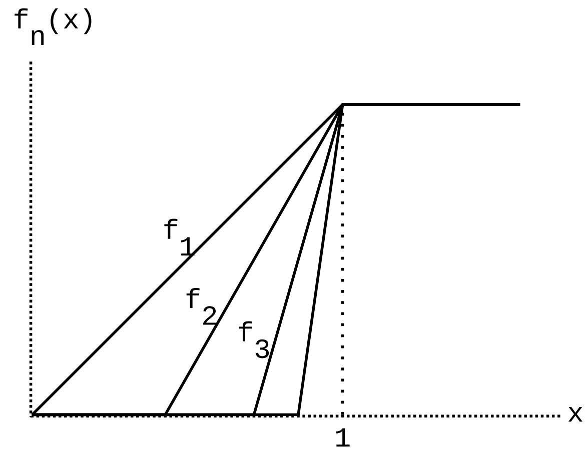

Figure 4: Here (f_n) converges to f, where $f(x) = 0$ for $x < 1$ and $f(x) = 1$ for $x \geq 1$. Convergence is pointwise but not uniform.

14.3 DEFINITION. The sequence (T_n) is said to *converge uniformly* to a mapping T provided that

$$\lim_n \sup_{x \in E} d'(T_n x, Tx) = 0.$$

Obviously, uniform convergence of (T_n) implies pointwise convergence (and the limit T is the same). That the converse is generally false can be seen from Figures 4 and 5 below: here the functions $f_n : [0, \infty) \mapsto [0, 1]$ converge pointwise, but not uniformly.

Cauchy Criterion

As with sequences of points, it is important to have a criterion for the uniform convergence of (T_n) expressed in terms of the T_n themselves. The following Cauchy criterion does this:

14.4 THEOREM. *Suppose that E' is complete. Then, (T_n) is uniformly convergent if and only if for every $\epsilon > 0$ there is an n_ϵ with*

14.5 $$\sup_x d'(T_n x, T_m x) < \epsilon \quad \textit{for all } m > n \geq n_\epsilon.$$

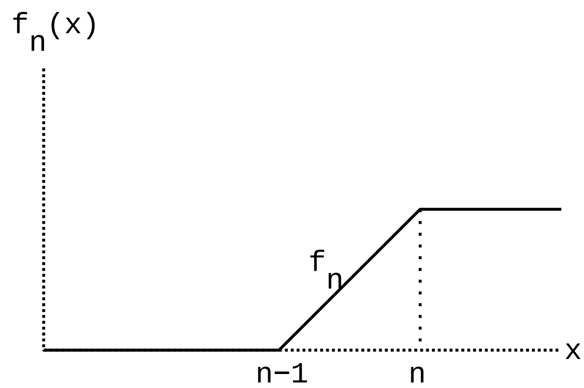

Figure 5: These f_n converge to $f = 0$ pointwise, but not uniformly.

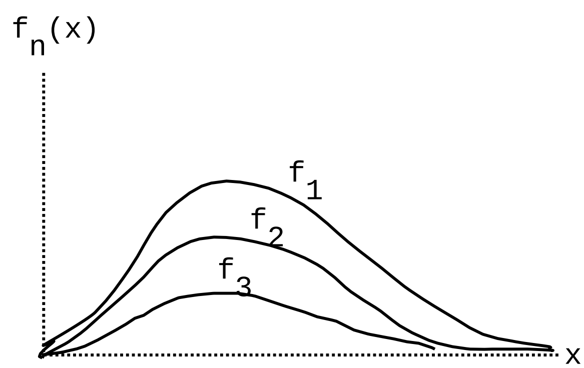

Figure 6: These f_n converge to 0 uniformly (and hence pointwise).

PROOF. Suppose that (T_n) converges uniformly, say, to T. Then, for every $\epsilon > 0$, there is an n_ϵ such that $d'(T_nx, Tx) < \epsilon/2$ for all $n \geq n_\epsilon$. Thus, for $m, n \geq n_\epsilon$,

$$d'(T_nx, T_mx) \leq d'(T_nx, Tx) + d'(Tx, T_mx) < \epsilon/2 + \epsilon/2 = \epsilon$$

for all x. So, (T_n) is Cauchy (for every $\epsilon > 0$ there is n_ϵ such that 14.5 holds).

Let (T_n) be Cauchy. Then, in particular, for each x in E the sequence (T_nx) in E' is Cauchy. Since E' is complete, this implies that (T_nx) converges to some point of E', call it Tx. This defines a mapping $T : E \mapsto E'$. We want to show that (T_n) converges to T uniformly. Since (T_n) is Cauchy, for every $\epsilon > 0$ there is an n_ϵ such that

$$d'(T_nx, T_mx) < \epsilon \quad \text{for all } m, n \geq n_\epsilon$$

for all x. Now, let $m \mapsto \infty$; then, (T_mx) converges to Tx and the continuity of $y \mapsto d'(T_nx, y)$ implies that $d'(T_nx, T_mx) \mapsto d'(T_nx, Tx)$. Thus, as we needed to show, for $\epsilon > 0$ there is an n_ϵ with

$$d'(T_nx, Tx) < \epsilon \quad \text{for all } n \geq n_\epsilon \text{ and all } x \in E.$$

□

Continuity of Limit Functions

As can be seen from Figure 4, the pointwise limit of a sequence of continuous functions is not necessarily continuous. In fact, the primary use of uniform convergence is to ensure the continuity of the limit function.

14.6 THEOREM. *Suppose that each T_n is continuous and (T_n) converges to T uniformly. Then, T is continuous.*

PROOF. Fix x in E. Note that for all n and y

$$d'(Tx, Ty) \leq d'(Tx, T_nx) + d'(T_nx, T_ny) + d'(T_ny, Ty).$$

Given $\epsilon > 0$, there is an n_ϵ such that the first and third terms on the right side are less than $\epsilon/3$ each for $n = n_\epsilon$; This comes from the uniform convergence of (T_n) of T. Moreover, the continuity of T_{n_ϵ} at the point x implies the existence of $\delta = \delta_{\epsilon,x}$ such that the second term on the right with $n = n_\epsilon$ is less than $\epsilon/3$ for all $y \in B(x, \delta)$. Hence, for every $\epsilon > 0$ there is a $\delta = \delta_{\epsilon,x}$ such that $d(x, y) < \delta$ implies that $d'(Tx, Ty) < \epsilon$ for all y; that is, T is continuous at x. □

Exercises:

14.1 Let $0 \leq a < b < 1$. Let $f_n : [a, b] \mapsto \mathbb{R}_+$ be defined by $f_n(x) = x^n$. Show that (f_n) converges uniformly to $f = 0$.

14.2 Let $T_n : [0,1] \mapsto [0,1]$ be defined by $T_n x = x^n(1-x)$. Show that (T_n) is uniformly convergent.

14.3 Let $f : \mathbb{R} \mapsto \mathbb{R}$ be uniformly continuous. Define $f_n(x) = f(x + 1/n)$. Show that (f_n) converges uniformly to f.

14.4 Let (f_n) be defined as a sequence of functions from $\mathbb{R}_+$ into $\mathbb{R}_+$ by $f_1(x) = \sqrt{x}$, $f_2(x) = \sqrt{x + \sqrt{x}}$, $f_3(x) = \sqrt{x + \sqrt{x + \sqrt{x}}}, \ldots$ Show that (f_n) is convergent and find the limit function.

15 Spaces of Continuous Functions

Throughout this section (E,d) will be a compact metric space, and all functions are from E into $\mathbb{R}$. On a first reading, the reader should take $E = [a,b]$, a closed interval. Our aim is to illustrate the uses of the foregoing concepts in the analysis of the function space $\mathcal{C}(E,\mathbb{R})$ of all continuous functions from E into $\mathbb{R}$. For brevity, we write $\mathcal{C}$ for $\mathcal{C}(E,\mathbb{R})$.

The set $\mathcal{C}$ is a vector space: if f and g are in $\mathcal{C}$ then so is $af + bg$ for each a in $\mathbb{R}$ and b in $\mathbb{R}$. Moreover, various arithmetic operations are well-defined on $\mathcal{C}$: $f + g$, $f - g$, $f \cdot g$, and f/g all belong to $\mathcal{C}$ if f and g are in $\mathcal{C}$, except that in the case of f/g one must worry about $g(x) = 0$.

Although each point of $\mathcal{C}$ is a function, in many respects $\mathcal{C}$ is like a Euclidean space. We may, for instance, define a norm of $\mathcal{C}$ as follows. Let $f \in \mathcal{C}$. Being a continuous function on a compact metric space, f is bounded and attains its maximum and minimum. It follows that

15.1 $$\|f\| = \max_{x \in E} |f(x)|$$

is a well-defined positive real-number; it is called the *norm* of f. It is indeed a norm:

15.2 $$\|f\| \geq 0; \quad \|f\| = 0 \text{ if and only if } f = 0;$$

15.3 $$\|cf\| = |c| \cdot \|f\|;$$

15.4 $$\|f + g\| \leq \|f\| + \|g\|.$$

As with Euclidean spaces, we may use the norm above to define a metric on $\mathcal{C}$. We define the distance between f and g to be

15.5 $$d(f,g) = \|f - g\|.$$

Convergence in $\mathcal{C}$

The following shows that the convergence in the metric space $\mathcal{C}$ is equivalent to the uniform convergence of functions on E.

15.6 THEOREM. *A sequence* (f_n) *in* $\mathcal{C}$ *is convergent if and only if the sequence of functions* $f_n : E \mapsto \mathbb{R}$ *is uniformly convergent.*

PROOF. The definition of convergence for a sequence of points in a metric space and the definition of uniform convergence for a sequence of functions $f_n : E \mapsto \mathbb{R}$ are such that the claim is simply that

$$\lim_n d(f_n, f) = 0 \quad \Leftrightarrow \quad \lim_n \sup_{x\in E} |f_n(x) - f(x)| = 0.$$

But this is obvious in view of 15.5 and 15.1. □

Conceptually, then, the somewhat complex concept of uniform convergence of a sequence of functions is equivalent to the simpler concept of convergence of a sequence in a metric space.

Lipschitz Continuous Functions

A function $f \in \mathcal{C}$ is said to be *Lipschitz continuous* if there exists a constant K such that

15.7 $$|f(x) - f(y)| \leq K \cdot d(x, y) \quad \text{for all } x, y \in E.$$

Let B_K be the set of all f in $\mathcal{C}$ satisfying 15.7. Then, clearly, the set of all Lipschitz continuous functions is exactly the union of the B_K's.

If $E = [a, b]$, f is differentiable, and the derivative f' is bounded (that is, there exists a K such that $|f'(x)| \leq K$ for all $x \in [a, b]$), then f is Lipschitz continuous. Consider a fixed K and let A_K denote the set of all differentiable functions f whose derivatives f' are continuous and bounded by K. The set A_K is not closed, which can be seen from Figure 7 where $(f_n) \subset A_K$, (f_n) converges to f in $\mathcal{C}$, but f is not in A_K. In fact, the closure of A_K is precisely B_K. We leave this without proof. Instead, we show the following partial result with general E.

15.8 PROPOSITION. B_K *is a closed subset of* $\mathcal{C}$.

PROOF. We use the characterization Theorem 9.5 from Chapter . Let $(f_n) \subset B_K$ converge to the point f in $\mathcal{C}$. We need to show that f is in B_K. Now, for arbitrary x and y in E,

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &\leq \|f - f_n\| + Kd(x, y) + \|f_n - f\| \end{aligned}$$

for all n. Since $\|f_n - f\| \mapsto 0$, this shows that f satisfies 15.7. □

As mentioned above, the set of all Lipschitz continuous functions coincides exactly with $\cup_K B_K$. Even though each B_K is closed, the union is not. This fact can be seen from the sequence of functions shown in Figure 8. In fact, its closure is precisely $\mathcal{C}$, that is, every f in $\mathcal{C}$ is the limit of a sequence of Lipschitz continuous functions.

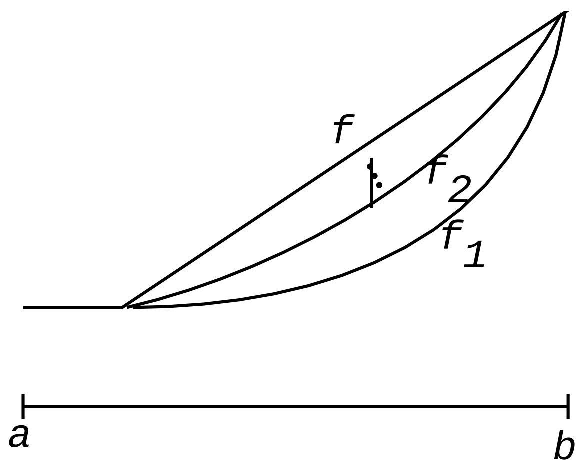

Figure 7: A sequence of differentiable functions whose derivatives are bounded but whose limit is not differentiable.

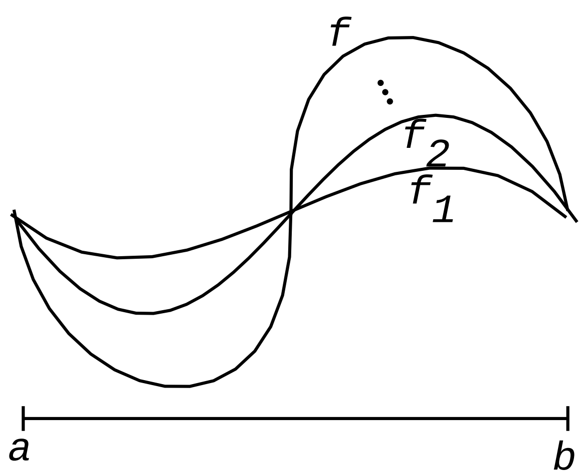

Figure 8: A sequence of Lipschitz continuous functions converging to a continuous function that is not Lipschitz.

Completeness

The space $\mathcal{C}$ is not bounded. Therefore it cannot be compact. But, at least, it is complete.

15.9 THEOREM. *The space $\mathcal{C}$ is complete.*

PROOF. Let $(f_n) \subset \mathcal{C}$ be Cauchy, that is, for every $\epsilon > 0$ there is an n_ϵ such that $\|f_n - f_m\| \leq \epsilon$ for all $m > n \geq n_\epsilon$. This is equivalent to the condition 14.5 (here $E' = \mathbb{R}$ which is complete). Thus, by Theorem 14.4, (f_n) is uniformly convergent as a sequence of functions on E. But, by Theorem 15.6, uniform convergence is equivalent to convergence in $\mathcal{C}$. So, (f_n) is convergent in $\mathcal{C}$. □

Functionals

Since $\mathcal{C}$ is a metric space, we may speak of functions defined on $\mathcal{C}$ as we speak of functions defined on E. For linguistic clarity, a function from $\mathcal{C}$ into $\mathbb{R}$ is called a *functional*. Here are some examples of functionals: for $f \in \mathcal{C}$,

15.10 $$M(f) = \max_{x \in E} f(x)$$

15.11 $$P_x(f) = f(x), \quad x \in E \text{ fixed}$$

15.12 $$F(f) = \phi(f(x_1), \ldots, f(x_k)),$$

where $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ is fixed and $x_1, \ldots, x_k$ are fixed in E.

Here are some further examples of functionals, in the particular case where $E = [a, b]$:

15.13 $$L(f) = \int_a^b f(x)dx,$$

15.14 $$L_\phi(f) = \int_a^b \phi(x)f(x)dx,$$

where $\phi \in \mathcal{C}$ is some fixed function.

The functional M is uniformly continuous; in fact, it is Lipschitz continuous with Lipschitz constant $K = 1$:

$$\begin{aligned} |M(f) - M(g)| &= |\max_x f(x) - \max_x g(x)| \\ &\leq \max_x |f(x) - g(x)| \\ &= \|f - g\| \\ &= d(f, g). \end{aligned}$$

Even easier is the Lipschitz continuity of the coordinate mapping P_x:

$$|P_x(f) - P_x(g)| = |f(x) - g(x)| \leq \|f - g\|.$$

Assuming that the function $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ is continuous, the function F is continuous: if $\|f_n - f\| \mapsto 0$, then the sequence of points $(f_n(x_1), \ldots, f_n(x_k)) \in \mathbb{R}^k$ converges to the point $(f(x_1), \ldots, f(x_k)) \in \mathbb{R}^k$ as $n \mapsto \infty$, and the continuity of ϕ implies that $F(f_n) \mapsto F(f)$.

The functional L is a linear transformation from $\mathcal{C}$ into $\mathbb{R}$. It is uniformly continuous; in fact, it is Lipschitz continuous with Lipschitz constant $K = b - a$. So is L_ϕ with Lipschitz constant $K = \int_a^b |\phi(x)| dx$.

Exercises:

15.1 If f and g are two continuous functions on a compact metric space, show that

$$|\max_x f(x) - \max_x g(x)| \leq \max_x |f(x) - g(x)|.$$

Differential and Integral Equations

The aim of this chapter is to discuss several applications of metric space ideas to some classical problems of engineering analysis.

We shall start with one theorem, the fixed point theorem for contractions on a metric space, and show how various problems can be beaten to submission with it.

16 Contraction Mappings

The aim of this section is to prepare the stage for some applications to differential and integral equations encountered frequently in engineering. Throughout, E is a metric space with some metric d.

We shall use the term "transformation on E" to mean a mapping from E into E. If T is a transformation on E, then the image Tx of x is a point in E, and the image of Tx is $T(Tx)$, for which we will write T^2x. In other words, we are writing T^2 for $T \circ T$. Similarly, we define further iterates by

$$T^{n+1}x = T(T^n x), \quad x \in E, n \geq 0,$$

with $T^0x = x$ for all x. So, T^0 is the identity, T^1 is T, etc.

Given a point x in E, if we write $x_0 = x$, $x_1 = Tx$, $x_2 = T^2x$, $x_3 = T^3x$, $\ldots$, we obtain a sequence (x_n) in E; this sequence is called the *orbit* starting at x. One should think of $x_n = T^n x$ as the position at time n of a particle that starts at x and moves successively to Tx, T^2x, $\ldots$.

16.1 DEFINITION. A transformation T on E is said to be a *contraction* if it is Lipschitz continuous with some Lipschitz constant $\alpha < 1$.

In other words, T is a contraction of E if there exists a constant $\alpha \in [0, 1)$ such that

16.2 $$d(Tx, Ty) \leq \alpha d(x, y) \quad \text{for all } x, y \in E.$$

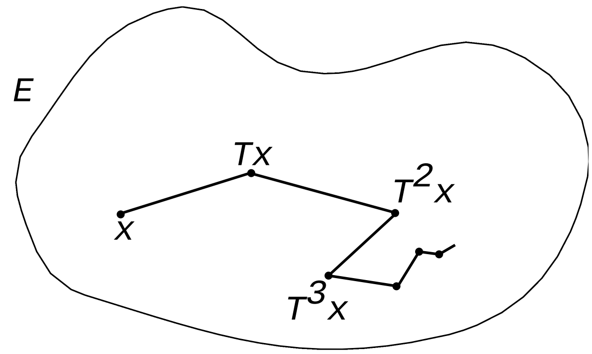

Figure 9: The orbit of x under the map T.

Fixed Point Theorem

A point x is said to be a *fixed point* of a transformation T if $Tx = x$. Figure 10 shows a transformation T on $E = [0, 1]$; there, x^* is the unique fixed point of T, and the orbit $(T^n x_0)$ of x_0 converges to the fixed point x^*.

The following theorem shows that every contraction of a complete metric space has a unique fixed point. Its proof shows how to obtain the fixed point by a method of successive approximations.

16.3 THEOREM. *Suppose that E is complete. Let T be a contraction on E. Then, T has a unique fixed point and for each point x_0 in E, the orbit $(T^n x_0)$ converges to that fixed point.*

PROOF. Fix x_0 in E and let $(x_0, x_1, x_2, \ldots)$ be its orbit. We show first that this sequence is Cauchy. Indeed, suppose that $m < n$. Then $x_m = T^m x_0$ and $x_n = T^n x_0 = T^m T^{n-m} x_0 = T^m x_{n-m}$. Hence, since $d(T^m x, T^m y) \le \alpha^m d(x, y)$ in view of 16.2, we have

$$\begin{aligned} d(x_m, x_n) &\le \alpha^m d(x_0, x_{n-m}) \\ &\le \alpha^m [d(x_0, x_1) + d(x_1, x_2) + \cdots + d(x_{n-m-1}, x_{n-m})]. \end{aligned}$$

Now note that $d(x_i, x_{i+1}) = d(T^i x_0, T^i x_1) \le \alpha^i d(x_0, x_1)$. Thus,

$$\begin{aligned} d(x_m, x_n) &\le \alpha^m d(x_0, x_1)[1 + \alpha + \alpha^2 + \cdots + \alpha^{n-m-1}] \\ &= \alpha^m d(x_0, x_1) \frac{1 - \alpha^{n-m}}{1 - \alpha} \\ &\le \alpha^m \frac{d(x_0, x_1)}{1 - \alpha}. \end{aligned}$$

Since $\alpha < 1$, the right side goes to 0 as $m \mapsto \infty$. Hence, the sequence (x_n) is Cauchy.

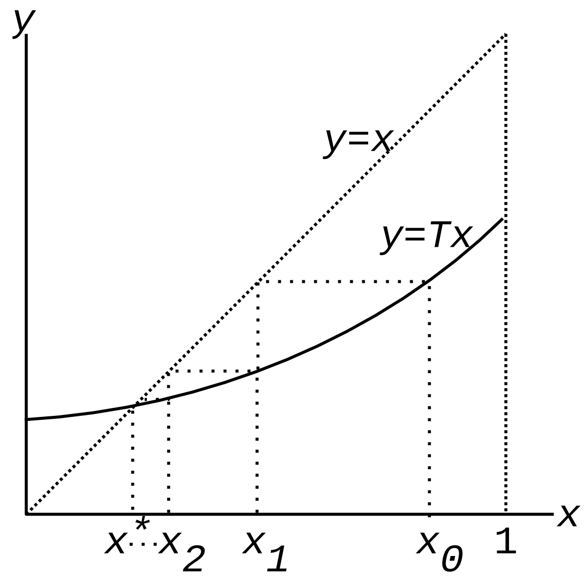

Figure 10: A contraction on $[0, 1]$.

Since E is complete, the sequence (x_n) must converge to some point x in E. Then, by the continuity of T,

$$Tx = T(\lim x_n) = \lim Tx_n = \lim x_{n+1} = x,$$

that is, x is a fixed point. To complete the proof, we now show that the fixed point is unique. To this end, let y be another fixed point. Then,

$$Tx = x \quad \text{and} \quad Ty = y,$$

and hence, by the contraction condition, $d(x, y) = d(Tx, Ty) \leq \alpha d(x, y)$. Since $\alpha < 1$, this is possible only if $d(x, y) = 0$, that is, $x = y$. □

The preceding theorem can be used to prove existence and uniqueness of solutions to a wide variety of equations. Besides showing that $Tx = x$ has a solution, the proof gives a practical method for arriving at it. Indeed, start from an arbitrary point x_0 and successively compute $x_1 = Tx$, $x_2 = Tx_1$, $x_3 = Tx_2$, $\ldots$. The x_n get close to x (geometrically fast):

$$d(x_{n+1}, x) = d(Tx_n, Tx) \leq \alpha d(x_n, x),$$

which shows that

16.4 $$d(x_n, x) \leq \alpha^n d(x_0, x).$$

Exercises:

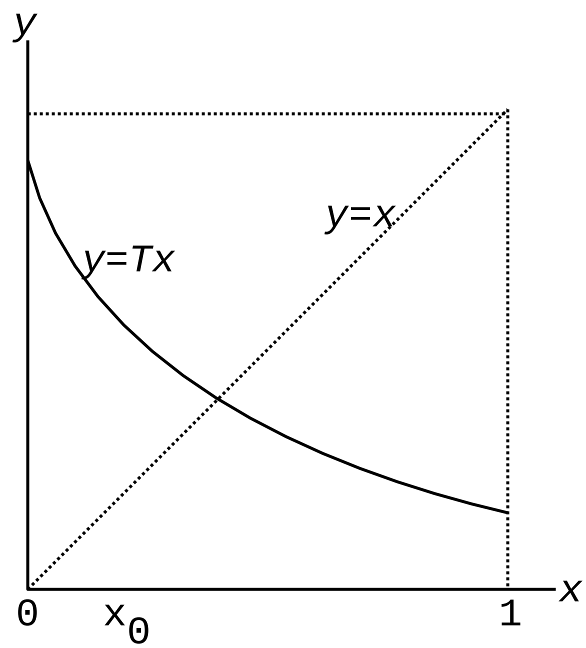

Figure 11: Exercise 16.1.

16.1 For the transformation $T : [0,1] \mapsto [0,1]$ shown in Figure 11 find the orbit of the point x_0 indicated.

16.2 For the transformation $T : [0,1] \mapsto [0,1]$ given by $Tx = 0.3 + 0.2x + 0.5x^3$, Figure 12 shows that there are exactly two fixed points. Find them. Show that, for arbitrary $x_0 \neq 1$, the orbit of x_0 converges to the smaller fixed point x^*.

16.3 *Branching processes.* In a chain reaction, each particle gives rise to a random number of new particles. Each of these new particles act independently and produces random numbers of newer particles. And this continues indefinitely. Let p_k be the probability that a particle produces k particles; here $p_0, p_1, p_2, \ldots$ are positive numbers with $\sum p_k = 1$. Starting with one particle, we now consider the probability that the chain reaction fizzles out, that is, the population of particles becomes extinct. Let x_n be the probability that the n^{th} generation is extinct already. Note that the $(n+1)^{\text{th}}$ generation consists of particles that are n^{th} generation offspring of the individuals of the first generation. In order for the population to be extinct at or before the $(n+1)^{\text{th}}$ generation, populations initiated by the

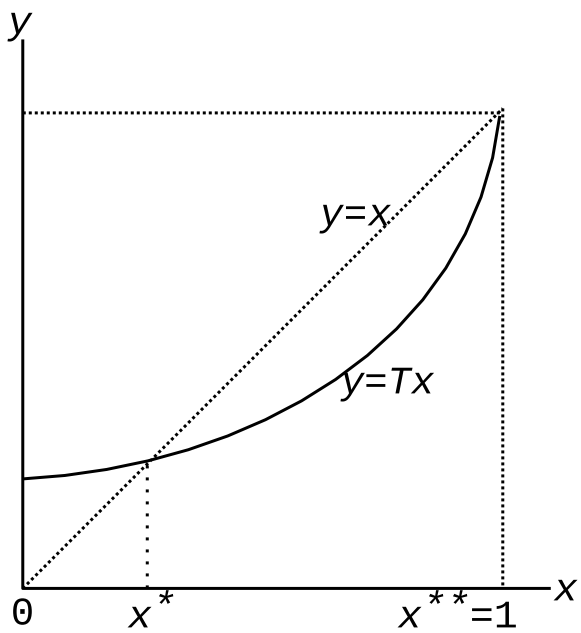

Figure 12: Exercise 16.2.

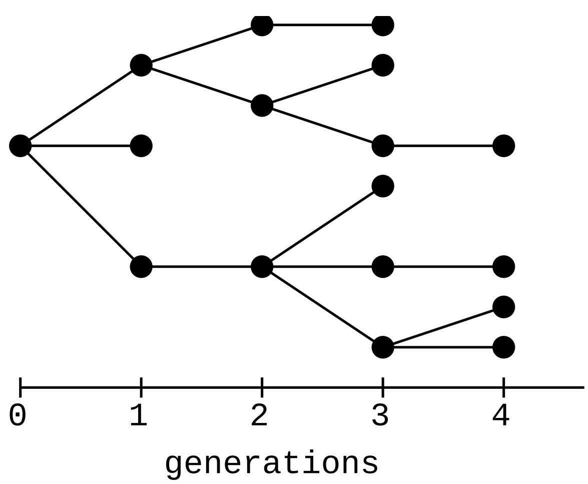

Figure 13: Exercise 16.3.

particles of the first generation must all become extinct. Thus,

$$x_{n+1} = \sum_{k=0}^{\infty} p_k (x_n)^k.$$

In other words, $x_{n+1} = Tx_n$ where $T : [0,1] \mapsto [0,1]$ is defined by

$$Tx = \sum_{k=0}^{\infty} p_k x^k, \quad x \in [0,1].$$

Now, the probability x^* of eventual extinction for the population is the limit of x_n, and thus satisfies

$$x^* = Tx^*.$$

(a) Show that $x_1 = p_0$. Show that the sequence (x_n) increases to the extinction probability x^*.

(b) Assume that $p_0 > 0$. If $p_0 + p_1 = 1$ (so that $p_2 = p_3 = \cdots = 0$) show that $x^* = 1$.

(c) Show that the mapping $x \mapsto Tx$ is increasing and convex.

(d) Let $a = \sum_{k=1}^{\infty} p_k k$, that is, a is the expected number of particles produced by one particle. Show that if $a \leq 1$, then $x = Tx$ has only one solution and the fixed point is $x^* = 1$.

(e) Suppose that $a > 1$. Then, show that $x = Tx$ has exactly two solutions. One solution is 1, the other is the extinction probability x^*. Show this by examining the graph of T and using (a).

16.4 Let $T : [0,1] \mapsto [0,1]$ be defined by

$$Tx = 4x(1-x).$$

Show that T has exactly two fixed points. Compute them. Give an example of an orbit that converges to the fixed point $x^* = 0$. Note the highly chaotic nature of the orbits.

16.5 Let $T : [0,1] \mapsto [0,1]$ be defined by $Tx = 2x \pmod 1$, that is, $Tx = 2x$ if $2x < 1$ and $Tx = 2x - 1$ if $2x \geq 1$. The only fixed point is $x^* = 0$.

Incidentally, if $x = 0.\omega_1\omega_2\omega_3\cdots$ is the binary representation of x then $Tx = 0.\omega_2\omega_3\omega_4\cdots$ and $T^2x = 0.\omega_3\omega_4\omega_5\cdots$, etc. Note the highly chaotic nature of the orbits by plotting $(T^n x)$.

16.6 Let $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a linear transformation, say $Tx = Ax$ where A is some $n \times n$ matrix. Give a condition on A that guarantees T to be a contraction (with the Euclidean metric on $\mathbb{R}^n$).

16.7 Let $Tx = Ax + b$ where A is $n \times n$ matrix and b is a fixed vector in $\mathbb{R}^n$. Consider $E = \mathbb{R}^n$ with the weighted Manhattan metric $d(x,y) = \sum_{i=1}^n w_i \cdot |x_i - y_i|$ where the weights $w_1, \ldots, w_n$ are strictly positive. Show that, to assume that T is a contraction of this metric space E, it is sufficient to have

$$\sum_{i=1}^n w_i |a_{ij}| < w_j, \quad j = 1, \ldots, n.$$

17 Systems of Linear Equations

In this section we discuss the use of the fixed point theorem in solving systems of linear equations. As a by-product, we get a chance to discuss the importance of choosing the right metric for a particular application.

Let $E = \mathbb{R}^n$; we do not specify the metric just yet. Fix $b \in \mathbb{R}^n$ and consider the system of linear equations

17.1
$$x_i = \sum_{j=1}^n a_{ij} x_j + b_i, \quad i = 1, \ldots, n,$$

where the a_{ij} are real numbers. Writing A for the $n \times n$ matrix of elements a_{ij}, the system 17.1 is equivalent to

17.2
$$x = Ax + b.$$

In other words, the problem is to find the fixed point of the transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ defined by

17.3
$$Tx = Ax + b.$$

If T is a contraction, then we can use Theorem 16.3 and obtain the unique solution of $Tx = x$ by the method of successive approximations.

The conditions under which T is a contraction depend on the choice of metric on $E = \mathbb{R}^n$. We discuss three cases.

Maximum Norm

Suppose that d is the metric associated with the maximum norm:

$$d(x,y) = \max_{1 \le i \le n} |x_i - y_i|.$$

Then, since $Tx - Ty = Ax - Ay = A(x-y)$,

$$d(Tx, Ty) \quad = \quad \max_i |\sum_{j=1}^n a_{ij}(x_j - y_j)|$$

$$\begin{aligned}
&\leq \max_i \sum_j |a_{ij}| \cdot |x_j - y_j| \\
&\leq \max_i \sum_j |a_{ij}| \max_k |x_k - y_k| \\
&= (\max_i \sum_j |a_{ij}|) d(x, y).
\end{aligned}$$

Thus, the contraction condition 16.2 is satisfied if

17.4 $$\alpha = \max_i \sum_j |a_{ij}| < 1.$$

Manhattan Metric

Suppose that d is the Manhattan metric:

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|.$$

Then,

$$\begin{aligned}
d(Tx, Ty) &= \sum_i |\sum_j a_{ij}(x_j - y_j)| \\
&\leq \sum_i \sum_j |a_{ij}| \cdot |x_j - y_j| \\
&\leq (\max_j \sum_i |a_{ij}|) d(x, y),
\end{aligned}$$

and the contraction condition is satisfied if

17.5 $$\alpha = \max_j \sum_i |a_{ij}| < 1.$$

Euclidean Metric

Suppose that d is the ordinary Euclidean distance. Then,

$$\begin{aligned}
d(Tx, Ty)^2 &= \sum_i \left(\sum_j a_{ij}(x_j - y_j) \right)^2 \\
&\leq \sum_i \left(\sum_j a_{ij}^2 \right) \left(\sum_j (x_j - y_j)^2 \right) \\
&= (\sum_i \sum_j a_{ij}^2) d(x, y)^2,
\end{aligned}$$

where we used Schwartz's inequality at the second step. Thus, the contraction condition 16.2 is satisfied if

17.6 $$\alpha = \sum_i \sum_j a_{ij}^2 < 1.$$

Conclusion

Under each of the metrics discussed, $\mathbb{R}^n$ is a complete metric space. Hence, if at least one of the conditions 17.4–17.6 holds, Theorem 16.3 applies to show that there exists a unique solution to 17.1. The sequence of successive approximations $x^{(0)}, x^{(1)}, \dots$ (whose limit is the fixed point x) has the following form:

17.7 $$x^{(k+1)} = Ax^{(k)} + b, \quad k = 0, 1, \dots,$$

and we can choose any point $x^{(0)} \in \mathbb{R}^n$ as the initial point.

Each of the conditions 17.4–17.6 is sufficient for applying this method. None is necessary; it is easy to give examples of A where one condition holds but not the others.

18 Integral Equations

The most interesting applications of fixed point theorems arise when the underlying metric space is a function space. Here we discuss the existence and uniquencess of solutions to Fredholm and Volterra equations.

Fredholm Equation

A *Fredholm equation* (of the second kind) is an integral equation of the form

18.1 $$f(x) = \phi(x) + \lambda \int_a^b K(x,y) f(y) dy.$$

Here, the functions $K : [a,b] \times [a,b] \mapsto \mathbb{R}$ and $\phi : [a,b] \mapsto \mathbb{R}$ are given, $\lambda \in \mathbb{R}$ is an arbitrary parameter, and $f : [a,b] \mapsto \mathbb{R}$ is the unknown function. The function K is called the *kernel* of the equation. The equation is said to be *homogeneous* if $\phi = 0$ and *non-homogeneous* otherwise.

The Fredholm equation is the continuous version of the system of linear equations 17.1. To see this, suppose that the interval is discretized and is replaced by $n+1$ equidistant points $a = x_0 < x_1 < \cdots < x_n = b$. Then, writing $y_i = f(x_i)$ and $b_i = \phi(x_i)$ and $a_{ij} = \lambda K(x_i, x_j)/n$, we see that 18.1 becomes

$$y_i = b_i + \sum_j a_{ij} y_j.$$

Whether this discretization is appropriate is a different matter.

Let $\mathcal{C}$ be the collection of all continuous functions f from $[a, b]$ into $\mathbb{R}$, and let the metric on $\mathcal{C}$ be defined through the maximum norm:

18.2 $$d(f, g) = \|f - g\| = \sup_{a \leq x \leq b} |f(x) - g(x)|.$$

With this metric, $\mathcal{C}$ is a complete metric space (see Theorem 15.9 in Chapter).

Suppose that K is continuous on the square $[a, b] \times [a, b]$ and that ϕ is continuous on $[a, b]$. Then, the function Tf defined by

18.3 $$Tf(x) = \phi(x) + \lambda \int_a^b K(x, y) f(y) dy$$

is continuous on $[a, b]$ for each continuous function f on $[a, b]$. In other words, the mapping $f \mapsto Tf$ is a transformation on $\mathcal{C}$. Now, the Fredholm equation 18.1 becomes

18.4 $$f = Tf,$$

and thus, solving 18.1 is equivalent to finding the fixed points of the transformation T on $\mathcal{C}$.

To this end, in order to apply the fixed point theorem 16.3, all we need to show is that T is a contraction (recall that $\mathcal{C}$ is complete). The following shows that T is indeed so if the parameter λ is small enough.

18.5 THEOREM. *Suppose that ϕ and K are continuous. Then there exists $\lambda_0 > 0$ such that the equation 18.1 has a unique solution f for each λ in $(-\lambda_0, \lambda_0)$. Moreover, the solution f is continuous.*

PROOF. Since K is continuous on the square $[a, b] \times [a, b]$, it is bounded there (continuous functions on compact spaces are bounded). So, there is a constant $c > 0$ such that $|K(x, y)| \leq c$ for all x, y. Thus,

$$\begin{aligned} \|Tf - Tg\| &= \max_x |\lambda \int_a^b K(x, y)(f(y) - g(y))| \\ &\leq |\lambda| \cdot c \cdot (b - a) \max_y |f(y) - g(y)| \\ &= |\lambda| \cdot c \cdot (b - a) \cdot \|f - g\|. \end{aligned}$$

Choose $\lambda_0 = 1/c \cdot (b - a)$. Then, for each $\lambda \in (-\lambda_0, \lambda_0)$, the preceding shows that T is a contraction on $\mathcal{C}$. By Theorem 16.3, consequently, there is a unique fixed point f in $\mathcal{C}$ of the transformation T. □

18.6 EXAMPLE. Suppose that $K(x, y) = xy$ on $[0, 1] \times [0, 1]$. Let $\phi \in \mathcal{C}$ be arbitrary and consider the Fredholm equation

18.7 $$f(x) = \phi(x) + \lambda \int_0^1 xy f(y) dy.$$

The proof of 18.5 shows that, for $|\lambda| < 1$, there is a unique solution f. And the solution is the limit of the sequence

$$f_0 = \phi, \quad f_1 = Tf_0, \quad f_2 = Tf_1, \quad f_3 = Tf_2, \dots$$

where, in general,

$$Tf(x) = \phi(x) + \lambda x \int_0^1 yf(y)dy.$$

Now, we start computing. Defining $a = \int_0^1 y\phi(y)dy$, we have

$$\begin{array}{lclcl}
f_0(x) & = & \phi(x) & & \\
f_1(x) & = & Tf_0(x) & = & \phi(x) + \lambda x \int_0^1 y\phi(y)dy \\
 & & & = & \phi(x) + a\lambda x \\
f_2(x) & = & Tf_1(x) & = & \phi(x) + \lambda x \int_0^1 y(\phi(y) + a\lambda y)dy \\
 & & & = & \phi(x) + a\lambda x + a\frac{\lambda^2}{3}x \\
f_3(x) & = & Tf_2(x) & = & \phi(x) + \lambda x \int_0^1 y(\phi(y) + a\lambda y + a\frac{\lambda^2}{3}y)dy \\
 & & & = & \phi(x) + a\lambda x + a\frac{\lambda^2}{3}x + a\frac{\lambda^3}{9}x \\
 & & \vdots & & \\
f_n(x) & = & Tf_{n-1}(x) & = & \phi(x) + a\lambda x \left(1 + \frac{\lambda}{3} + \left(\frac{\lambda}{3}\right)^2 + \cdots + \left(\frac{\lambda}{3}\right)^{n-1}\right).
\end{array}$$

In fact, it becomes clear from this that a fixed point f exists for all $\lambda \in (-3, 3)$ and the solution to 18.7 is

18.8
$$f(x) = \lim_n f_n(x) = \frac{3a\lambda}{3-\lambda}x + \phi(x)$$

with $a = \int_0^1 \phi(y)dy$.

Going back to 18.7, the special form of the kernel K suggests a quicker method. Indeed, let

$$c = \int_0^1 yf(y)dy.$$

Then, using 18.7 in the form

$$f(x) = \phi(x) + \lambda xc,$$

we get

$$c = \int_0^1 xf(x)dx = \int_0^1 x\phi(x)dx + \int_0^1 x\lambda xcdx = a + \frac{\lambda}{3}c.$$

Solving this for c, we see that

$$f(x) = \phi(x) + \lambda xc = \phi(x) + \frac{3a\lambda}{3-\lambda}x$$

as before provided that $\lambda \neq 3$. Note that this is the solution for arbitrary $\lambda \neq 3$. But the method of successive approximations works for $|\lambda| < 3$ only.

Studying the iterative method in the preceding example, we can get a theoretical understanding of the nature of solutions. To this end, we re-do the computations of $f_0 = \phi$, $f_1 = Tf_0$, $f_2 = Tf_1$, ... once more, now with an arbitrary kernel K, and omitting the limits of integration we get

$$
\begin{aligned}
f_0(x) &= \phi(x) \\
f_1(x) &= Tf_0(x) = \phi(x) + \lambda \int K(x,y)\phi(y)dy \\
f_2(x) &= Tf_1(x) = \phi(x) + \lambda \int K(x,y) f_1(y)dy \\
&= \phi(x) + \lambda \int K(x,y)[\phi(y) + \lambda \int K(y,z)\phi(z)dz]dy \\
&= \phi(x) + \lambda \int K(x,y)\phi(y)dy + \lambda^2 \int K_2(x,z)\phi(z)dz
\end{aligned}
$$

where

$$K_2(x,z) = \int K(x,y)K(y,z)dy.$$

Continuing,

$$
\begin{aligned}
f_3(x) &= Tf_2(x) \\
&= \phi(x) + \lambda \int K(x,y)[\phi(y) + \lambda \int K(y,z)\phi(z)dz \\
&\quad + \lambda^2 \int K_2(y,z)\phi(z)dz] \\
&= \phi(x) + \lambda \int K(x,z)\phi(z)dz \\
&\quad + \lambda^2 \int K_2(x,z)\phi(z)dz + \lambda^3 \int K_3(x,z)\phi(z)dz
\end{aligned}
$$

where

$$K_3(x,z) = \int K(x,y)K_2(y,z)dz.$$

The pattern is now clear. We have

18.9 $$f_n(x) = \phi(x) + \sum_{i=1}^{n} \lambda^i \int_a^b K_i(x,y)\phi(y)dy$$

with $K_1 = K$, and $K_2, K_3, \ldots$ defined recursively via

18.10 $$K_{i+1}(x,y) = \int_a^b K(x,z)K_i(z,y)dy.$$

Theorem 18.5 shows that when $|\lambda| < \lambda_0$, the sequence f_n converges to the fixed point f, where

18.11 $$f(x) = \phi(x) + \sum_{i=1}^{\infty} \lambda^i \int_a^b K_i(x,y)\phi(y)dy.$$

Since this is true for arbitrary ϕ, we can change the order of summation and integration. Thus, with

18.12 $$R_\lambda(x,y) = \sum_{i=1}^{\infty} \lambda^i K_i(x,y),$$

we have

18.13 $$f(x) = \phi(x) + \int_a^b R_\lambda(x,y)\phi(y)dy.$$

Although 18.10, 18.12, 18.13 together give an "explicit" solution to the Fredholm equation, this explicitness is only theoretical. For, computing R_λ is of the same order of difficulty as solving 18.1 (in fact, even harder).

On the other hand, if the kernel K is simple enough, analytic solutions might be possible. The following illustrates the computations for such a special case.

18.14 EXAMPLE. Suppose that

$$K(x,y) = \sum_{j=1}^{n} p_j(x)q_j(y) \quad x, y \in [a,b]$$

for some continuous functions $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ on $[a,b]$. For ϕ continuous on $[a,b]$, consider the Fredholm equations 18.1. Now, if $f \in \mathcal{C}$ satisfies 18.1, then

18.15 $$f(x) = \phi(x) + \lambda \sum_{j=1}^{n} z_j p_j(x)$$

where

18.16 $$z_j = \int_a^b q_j(y)f(y)dy, \quad j = 1, \ldots, n.$$

In view of 18.15, then

$$\begin{aligned} z_i &= \int_a^b q_i(x)f(x)dx \\ &= \int_a^b q_i(x)\phi(x)dx + \lambda \sum_{j=1}^{n} \left(\int_a^b q_i(x)p_j(x)dx \right) z_j. \end{aligned}$$

Thus, letting

18.17 $$c_i = \int_a^b q_i(x)\phi(x)dx, \quad a_{ij} = \int_a^b q_i(x)p_j(x)dx,$$

we obtain

18.18 $$z_i = c_i + \lambda \sum_{j=1}^{n} a_{ij} z_j, \quad i = 1, 2, \ldots, n.$$

Note that the c_i and a_{ij} are known. If we can solve 18.18 for the z_i's, then 18.15 gives the solution f.

In vector-matrix notation, 18.18 becomes

$$z = c + \lambda A z,$$

whose solution is easy to discern. We can solve it for z (for arbitrary c) as long as $I - \lambda A$ is invertible, that is, as long as $1/\lambda$ is not an eigenvalue for A. Thus, we have a solution z for arbitrary b provided that $\lambda \in (-1/\lambda_0, 1/\lambda_0)$ where λ_0 is the modulus of the largest eigenvalue of A.

Volterra Equation

Let K be a continuous function on $[a, b] \times [a, b]$ and let ϕ be a continuous function on $[a, b]$. Consider the equation

18.19 $$f(x) = \phi(x) + \lambda \int_a^x K(x, y) f(y) dy, \quad x \in [a, b].$$

It is called the *Volterra equation*. It differs from the Fredholm equation only slightly, and in form only. If we define

$$\hat{K}(x, y) = \begin{cases} K(x, y) & \text{if } y \le x, \\ 0 & \text{if } y > x, \end{cases}$$

then 18.19 becomes the Fredholm equation 18.1 with kernel $\hat{K}$. However, it is easier to attack 18.19 directly.

18.20 THEOREM. *For each $\lambda \in \mathbb{R}$, the Volterra equation 18.19 has a unique solution f that is continuous on $[a, b]$.*

PROOF. Let $\mathcal{C} = \mathcal{C}([a, b], \mathbb{R})$, the set of all continuous functions from $[a, b]$ into $\mathbb{R}$, with the usual uniform metric $\|f - g\|$. Let c be the maximum of $|K(x, y)|$ over all $x, y \in [a, b]$; this number is finite since K is continuous. Define the transformation $T : f \mapsto Tf$ on $\mathcal{C}$ by

$$Tf(x) = \phi(x) + \lambda \int_a^x K(x, y) f(y) dy.$$

Now, for f and g in $\mathcal{C}$,

$$\begin{aligned} |Tf(x) - Tg(x)| &= |\lambda \int_a^x K(x, y)[f(y) - g(y)] dy| \\ &\le |\lambda| c (x - a) \|f - g\|, x \in [a, b]. \end{aligned}$$

We use this, next, to bound $T^2f - T^2g = T(Tf - Tg)$:

$$\begin{aligned}|T^2f(x) - T^2g(x)| &= |\lambda \int_a^x K(x,y)[Tf(y) - Tg(y)]dy| \\ &\leq |\lambda| \int_a^x |K(x,y)||\lambda|c(y-a)\|f-g\|dy \\ &\leq |\lambda|^2c^2 \int_a^x (y-a)dy\|f-g\| \\ &\leq \frac{|\lambda|^2c^2(x-a)^2}{2}\|f-g\|.\end{aligned}$$

Iterating in this manner, we see that

$$|T^kf(x) - T^kg(x)| \leq \frac{|\lambda|^kc^k(x-a)^k}{k!}\|f-g\|$$

for all $x \in [a,b]$. Hence,

$$\|T^kf - T^kg\| \leq \frac{[|\lambda|c(b-a)]^k}{k!}\|f-g\|.$$

Recalling that $r^n/n!$ tends to 0 as $n \mapsto \infty$ for any $r \in \mathbb{R}$, we conclude that there exists k such that T^k is a contraction: simply take k large enough to have $[|\lambda|c(b-a)]^k/k! < 1$. Finally, the existence and uniqueness of $f \in \mathcal{C}$ satisfying $f = Tf$ follows from the next theorem. Obviously, if $f = Tf$, then f solves 18.19. $\square$

Generalization of the Fixed Point Theorem

18.21 THEOREM. *Let E be a complete metric space and let T be a continuous transformation on E. If T^k is a contraction for some $k \geq 1$, then T has a unique fixed point.*

PROOF. Fix k such that $U = T^k$ is a contraction. By Theorem 16.3, then, U has a unique fixed point x, and $\lim_n U^nx_0 = x$ for every point x_0 in E. Now, by the continuity of T,

$$\begin{aligned}Tx &= \lim_n TU^nx_0 \\ &= \lim_n TT^{kn}x_0 \\ &= \lim T^{kn}Tx_0 \\ &= \lim_n U^nTx_0 \\ &= x,\end{aligned}$$

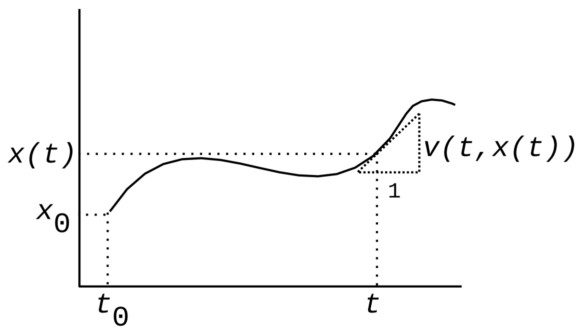

Figure 14: A moving particle.

that is, x is a fixed point of T. To show that it is the only fixed point of T we note that every fixed point of T is a fixed point of $T^k = U$, whereas U has only one fixed point, namely x. $\square$

Exercises:

18.1 Solve the Fredholm equation 18.1 for arbitrary ϕ, on $[a, b] = [0, 2\pi]$, with the kernel

$$K(x, y) = \sin(x + y).$$

18.2 Do the same with $[a, b] = [0, 1]$ and $K(x, y) = (x - y)^2$.

18.3 Let p be a continuous function of $[0, b]$. Show that

$$f(x) = \phi(x) + \int_0^x p(y)f(x - y)dy, \quad x \in [0, b],$$

has a unique solution f for each continuous function ϕ.

19 Differential Equations

We continue with applications of the fixed point theorem by discussing Picard's method of successive approximations for solving systems of differential equations.

We start with the simplest case where the differential equation describes the position of a particle moving on $\mathbb{R}$. The picture of the motion is given in Figure 14. The motion is described by the initial data t_0 and x_0 and by a continuous function $v : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ as follows. The particle starts from x_0 at time t_0; its velocity at time t is $v(t, x)$ if its position then is x. Thus, letting $x(t)$ denote the position of the particle at time t, we have

19.1 $$x(t) = x_0 + \int_{t_0}^t v(s, x(s))ds, \quad t \geq t_0.$$

The points t_0 and x_0 and the velocity function v are given. We are interested in the existence and uniqueness of the function x.

In the classical formulation of this problem, it is usual to express 19.1 as a differential equation:

19.2 $$\frac{dx}{dt} = v(t,x), \quad x(t_0) = x_0.$$

The following is *Picard's Theorem*:

19.3 THEOREM. *Let v be defined and continuous on $[t_0, \infty) \times [a,b]$, and x_0 be in (a,b), and suppose that v satisfies a Lipschitz condition in its spatial argument:*

19.4 $$|v(t,x) - v(t,y)| \le K|x-y|, \quad x,y \in [a,b].$$

Then, there is a $t_1 > t_0$ such that 19.1 has a unique solution $\{x(t) : t_0 \le t \le t_1\}$.

PROOF. By the continuity of v, we have

19.5 $$|v(t,x)| \le c, \quad t_0 \le t \le t_1', \quad a \le x \le b$$

for some constant c. Choose $\delta > 0$ so that

19.6 $$K\delta < 1 \quad \text{and} \quad a \le x_0 - c\delta < x_0 < x_0 + c\delta \le b.$$

Let $t_1 = \min\{t_1', t_0 + \delta\}$. Let $\mathcal{C}^*$ be the space of all continuous functions $x : [t_0, t_1] \mapsto [x_0 - c\delta, x_0 + c\delta]$ with the usual supremum metric; that is, $\|x-y\| = \sup_{t_0 \le t \le t_1} |x(t) - y(t)|$.

The set $\mathcal{C}^*$ is a closed subset of the space $\mathcal{C}([t_0, t_1], \mathbb{R})$. Since the latter is complete, $\mathcal{C}^*$ is complete.

Consider the transformation T defined by

19.7 $$Tx(t) = x_0 + \int_{t_0}^{t} v(s, x(s))ds, \quad t \in [t_0, t_1].$$

For $x \in \mathcal{C}^*$, we have from 19.5 that

$$|Tx(t) - x_0| \le \int_{t_0}^{t} |v(s,x(s))|ds \le c(t - t_0) \le c\delta,$$

which shows that $Tx \in \mathcal{C}^*$. Moreover, for $x, y \in \mathcal{C}^*$,

$$\begin{aligned} |Tx(t) - Ty(t)| &\le \int_{t_0}^{t} |v(s,x(s)) - v(s,y(s))|ds \\ &\le \int_{t_0}^{t} K|x(s) - y(s)|ds \\ &\le K\delta\|x - y\| \end{aligned}$$

in view of 19.4. Thus, $\|Tx - Ty\| \leq K\delta\|x - y\|$ and $K\delta < 1$ by the way δ was chosen. So, T is a contraction on $\mathcal{C}^*$. Since $\mathcal{C}^*$ is complete, Theorem 16.3 applies to show that T has a unique fixed point x. But, $x = Tx$ means that x solves 19.1. This completes the proof. $\square$

The preceding can be easily generalized to the case of systems of differential equations

19.8 $$\frac{dx_i}{dt} = v_i(t, x_1, \ldots, x_n), \quad i = 1, 2, \ldots, n.$$

Before listing it, we mention that the term "domain" means "an open and connected subset of a Euclidean space", and we note that 19.1 can be interpreted for $t < t_0$ by the convention that integrals from t_0 to t are the negatives of integrals from t to t_0. The following is the analog of Theorem 19.3 for motions in $\mathbb{R}^n$.

19.9 THEOREM. *Let v be a continuous function from some domain*

$$D \subset \mathbb{R} \times \mathbb{R}^n$$

into $\mathbb{R}^n$. Suppose that $(t_0, x_0) \in D$ and that $v(t, x) = (v_1(t, x), \ldots, v_n(t, x))$ satisfies the following Lipschitz condition for some K:

19.10 $$\max_{1 \leq i \leq n} |v_i(t, x) - v_i(t, y)| \leq K \max_{1 \leq j \leq n} |x_j - y_j|.$$

Then, there is an interval $[t_0 - \delta, t_0 + \delta]$ in which the system 19.8 has a unique solution $\{x(t) : t_0 - \delta \leq t \leq t_0 + \delta\}$ satisfying $x(t_0) = x_0$.

REMARK: In integral notation, we may write 19.8 as

$$x_i(t) = x_{0i} + \int_{t_0}^{t} v_i(s, x_1(s), \ldots, x_n(s))ds, \quad i = 1, \ldots, n.$$

The claim of the preceding theorem is that this has a unique solution $\{x(t) : t_0 - \delta \leq t \leq t_0 + \delta\}$. In vector notation, we may re-write this as

$$x(t) = x_0 + \int_{t_0}^{t} v(s, x(s))ds, \quad |t - t_0| \leq \delta,$$

which is exactly the same as 19.1 except that here $x : [t_0 - \delta, t_0 + \delta] \mapsto \mathbb{R}^n$ and $v : D \mapsto \mathbb{R}^n$.

Let the metric on $\mathbb{R}^n$ be

$$d(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|.$$

Then, the Lipschitz condition 19.10 can be written as

19.11 $$d(v(t, x), v(t, y)) \leq K d(x, y).$$

It should be clear by now that the proof of Theorem 19.3 will go through for Theorem 19.9 as well, with some notational changes. We give the proof for the sake of completeness.

PROOF. By the continuity of $v_1, \ldots, v_n$, we have

$$|v_i(t,x)| \leq c \quad i = 1, \ldots, n$$

for some $c > 0$, for all (t,x) in some domain $D' \subset D$ containing (t_0, x_0). Choose $\delta > 0$ so that

$$K\delta < 1$$

and

$$(t,x) \in D' \text{ if } t \in [t_0 - \delta, t_0 + \delta] \text{ and } d(x, x_0) \leq c\delta,$$

where the metric d is the usual maximum norm on $\mathbb{R}^n$.

Let $\mathcal{C}^*$ be the space of continuous functions $x : [t_0 - \delta, t_0 + \delta] \mapsto \bar{B}(x_0, c\delta)$, and let the metric on $\mathcal{C}^*$ be defined by

$$\|x - y\| = \max_t d(x(t), y(t)).$$

It is clear that $\mathcal{C}^*$ is complete. Define, for $x \in \mathcal{C}^*$,

$$Tx(t) = x_0 + \int_{t_0}^{t} v(s, x(s))ds, \quad t_0 - \delta \leq t \leq t_0 + \delta.$$

We proceed to show that T is a contraction on $\mathcal{C}^*$, which will complete the proof via Theorem 16.3.

First, we show that $Tx \in \mathcal{C}^*$ for $x \in \mathcal{C}^*$. For such x, it is clear that Tx is a continuous function, and

$$d(Tx(t), x_0) = \max_i |\int_{t_0}^{t} v_i(s, x(s))ds| \leq c\delta$$

for t in $[t_0 - \delta, t_0 + \delta]$ in view of the boundedness of v_i by c. Thus, $Tx \in \mathcal{C}^*$ if $x \in \mathcal{C}^*$. Moreover, for $x, y \in \mathcal{C}^*$,

$$\begin{aligned}
\|Tx - Ty\| &= \max_t d(Tx(t), Ty(t)) \\
&= \max_t \max_i |\int_{t_0}^{t} [v_i(s, x(s)) - v_i(s, y(s))]ds| \\
&\leq \max_t \int_{t_0}^{t} d(v(s, x(s)) - v(s, y(s)))ds \\
&\leq \max_t \int_{t_0}^{t} K d(x(s), y(s))ds \\
&\leq K\delta \|x - y\|,
\end{aligned}$$

which follows from the Lipschitz condition 19.11 on v. Since $K\delta < 1$, this shows that T is a contraction on $\mathcal{C}^*$. □

The preceding theorem ensures the existence and uniqueness of a solution x to the system 19.8 of differential equations. Successive approximations to x can be obtained as follows. Define

$$\begin{aligned} x^{(0)}(t) &= x_0, \quad t \in [t_0 - \delta, t_0 + \delta] \\ x^{(n+1)}(t) &= Tx^{(n)}(t) \\ &= x_0 + \int_{t_0}^{t} v(s, x^{(n)}(s))ds, \quad t \in [t_0 - \delta, t_0 + \delta]. \end{aligned}$$

Then, the sequence $x^{(n)}$ of functions converges to the solution x.

Exercises:

19.1 Solve the system

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^{n} a_{ij}x_j(t) + b_i(t), \quad i = 1, 2, \ldots, n$$

for smooth b and initial condition $x(0) = x_0$. How does the method of successive approximations work?

Convex Analysis

The aim of this chapter is to discuss basic concepts in convex analysis.

20 Convex Sets and Convex Functions

20.1 DEFINITION. A set $C \subset \mathbb{R}^n$ is called a *convex set* if

$$tx + (1-t)y \in C$$

for all $x, y \in C$ and $0 < t < 1$.

20.2 DEFINITION. An $\mathbb{R} \cup \{\infty\}$–valued function f defined on $\mathbb{R}^n$ is called a *convex function* if

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

for all $x, y \in \mathbb{R}^n$ and $0 < t < 1$.

An example of a convex set and function are shown in Figure 15. An example of a nonconvex set and function are shown in Figure 16. There are two important sets that one associates with functions defined on $\mathbb{R} \cup \{\infty\}$.

20.3 DEFINITION. The *epigraph* of an $\mathbb{R} \cup \{\infty\}$–valued function f, denote epi (f), is defined by

$$\text{epi } (f) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq r\}.$$

20.4 DEFINITION. Given a convex function f, The *effective domain* of an $\mathbb{R} \cup \{\infty\}$–valued function f, denote dom (f), is defined by

$$\text{dom } (f) = \{x \in \mathbb{R}^n : f(x) < \infty\}.$$

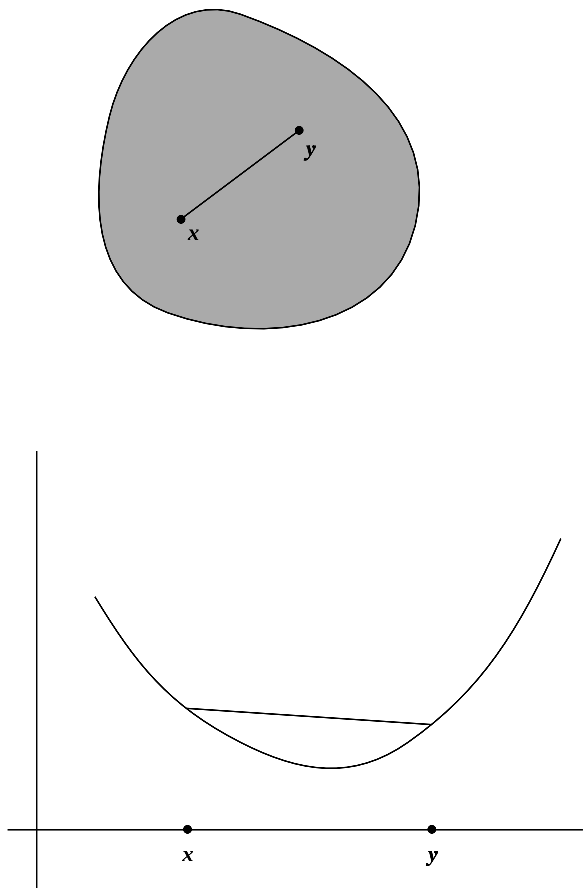

Figure 15: (a) A convex set. (b) A convex function.

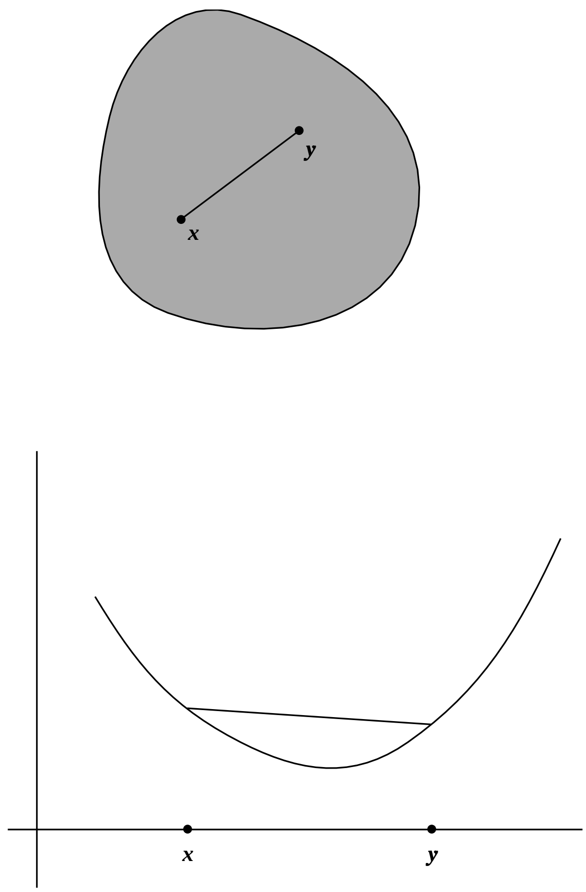

Figure 16: (a) A nonconvex set. (b) A nonconvex function.

The notions of set convexity and function convexity are closely related:

20.5 THEOREM. *A function is convex if and only if its epigraph is convex.*

PROOF. First suppose that f is convex. Fix (x, r) and (y, s) in epi (f) and fix $0 < t < 1$. Then

$$\begin{aligned} f(tx + (1-t)y) &\leq tf(x) + (1-t)f(y) \\ &\leq tr + (1-t)s. \end{aligned}$$

Therefore, $(tx + (1-t)y, tr + (1-t)s) \in$ epi (f). That is, epi (f) is convex.

Now, suppose that epi (f) is convex. Fix x, y in $\mathbb{R}^n$ and fix $0 < t < 1$. Then,

$$t(x, f(x)) + (1-t)(y, f(y)) \in \text{epi } (f).$$

Therefore, $tf(x) + (1-t)f(y) \geq f(tx + (1-t)y)$. That is, f is convex. $\square$

21 Projection

Given a point x in $\mathbb{R}^n$ and a convex set C, the following theorem establishes the existence and uniqueness of a point in C closest to the point x. Such a point is called the *projection* of x on C.

21.1 THEOREM. *Let C be a nonempty closed convex set in $\mathbb{R}^n$ and let x be a point in $\mathbb{R}^n$. Then, there exists a unique solution to*

$$\min_{z \in C} \|z - x\|^2.$$

PROOF. We start by proving existence. Fix $z_0 \in C$. Put $r = \|z_0 - x\|$ and let $B(r, x) = \{z : \|z - x \leq r\}$ denote the closed ball of radius r centered at x. Clearly,

$$\min_{z \in C} \|z - x\|^2 = \min_{z \in C \cap B(r,x)} \|z - x\|^2.$$

Put $f(z) = \|z - x\|^2$. As we saw in Theorem **??**, a continuous function on a nonempty compact set (in this case $C \cap B(r, x)$) attains its infimum. Therefore there exists an $x^* \in C$ such that

$$\|x^* - x\| \leq \|z - x\|$$

for all $z \in C$.

Now, consider the question of uniqueness. Suppose that x^* is not unique. That is, suppose that there exists an x^{**} in C that is distinct from x^* and for which

$$\|x^* - x\| = \|x^{**} - x\|.$$

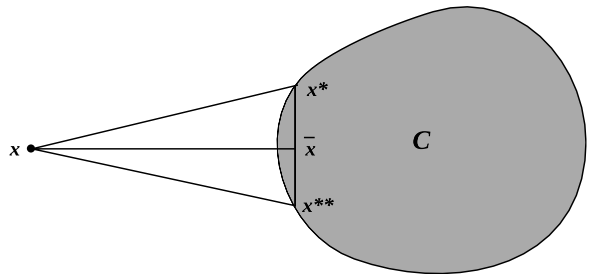

Figure 17: Clearly $x - \bar{x}$ is orthogonal to $x^* - x^{**}$ if x^* and x^{**} are equidistant from x.

Put $\bar{x} = (x^* + x^{**})/2$. By convexity of C, $\bar{x}$ belongs to C. Furthermore, $x - \bar{x}$ is orthogonal to $x^* - x^{**}$ (see Figure 17):

$$\begin{aligned}
(x-\bar{x})^T(x^*-x^{**}) &= \left(\frac{x-x^*}{2} + \frac{x-x^{**}}{2}\right)^T (x^*-x^{**}) \\
&= \frac{1}{2}\left((x-x^*)+(x-x^{**})\right)^T \left((x^*-x)+(x-x^{**})\right) \\
&= \frac{1}{2}\left(\|x-x^{**}\|^2 - \|x-x^*\|^2\right) \\
&= 0.
\end{aligned}$$

Now compare the distance to x^* with the distance to $\bar{x}$:

$$\begin{aligned}
\|x-x^*\|^2 &= (x-x^*)^T(x-x^*) \\
&= (x-\bar{x}+\bar{x}-x^*)^T(x-\bar{x}+\bar{x}-x^*) \\
&= \|x-\bar{x}\|^2 + 2(x-\bar{x})^T(\bar{x}-x^*) + \|\bar{x}-x^*\|^2 \\
&= \|x-\bar{x}\|^2 + \|\bar{x}-x^*\|^2 \\
&> \|x-\bar{x}\|^2.
\end{aligned}$$

The strict inequality contradicts the minimality of x^*. Therefore, the minimum must have been unique to start with. □

The next theorem gives a useful characterization of the projection of x on C.

21.2 THEOREM. *A point $\bar{x}$ is the projection of x on C if and only if $\bar{x}$ belongs to C and*

$$(z-\bar{x})^T(x-\bar{x}) \le 0$$

for all z in C.

Note that the above inequality can be interpreted geometrically as the statement that the vector from $\bar{x}$ to x makes an obtuse angle with the vector from $\bar{x}$ to any other

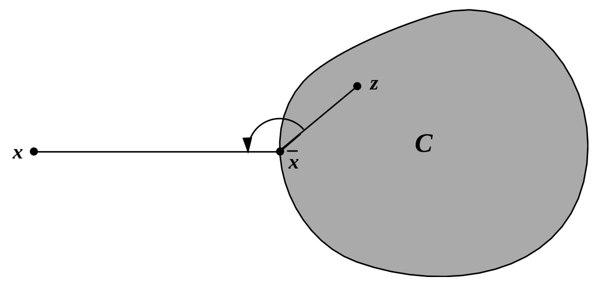

Figure 18: The angle between the vector from $\bar{x}$ to x and the vector from $\bar{x}$ to z is clearly obtuse if z is in C and C is convex.

point in C (see Figure 18).

PROOF. First suppose that $(z-\bar{x})^T(x-\bar{x}) \leq 0$ for all $z \in C$ and that $\bar{x}$ belongs to C. Fix z in C and compute as follows:

$$\begin{aligned} \|z-x\|^2 &= \|(z-\bar{x})+(\bar{x}-x)\|^2 \\ &= \|z-\bar{x}\|^2 + \|\bar{x}-x\|^2 + 2(z-\bar{x})^T(\bar{x}-x). \end{aligned}$$

Since all terms on the right are nonnegative, it follows that

$$\|z-x\|^2 \geq \|z-\bar{x}\|^2.$$

Since z was arbitrary, we see that the inequality holds for all z in C. Therefore, $\bar{x}$ is the projection of x on C.

Now, suppose that $\bar{x}$ is in C and that there exists a $z \in C$ for which

21.3 $$(z-\bar{x})^T(x-\bar{x}) > 0.$$

While z might be further from x than $\bar{x}$, we shall show that some points on the line segment connecting $\bar{x}$ to z are closer than $\bar{x}$ (see Figure 19). To this end, put

$$z(t) = tz + (1-t)\bar{x}$$

and

$$f(t) = \|z(t)-x\|^2.$$

It is easy to check that $f'(0) = 2(z-\bar{x})^T(\bar{x}-x)$, which is strictly negative. Therefore, there exists a $0 < \bar{t} < 1$ such that $f(\bar{t}) < f(0)$. But $z(\bar{t}) \in C$ and so $\bar{x}$ cannot be the projection of x on C. This contradiction implies that the strict inequality (21.3) must be wrong. □

When the set C is a linear subspace of $\mathbb{R}^n$, an explicit formula can be given for the projection onto C:

21.4 THEOREM. *Suppose that* $C = \{z : z = A^T y$ *for some* $y \in \mathbb{R}^m\}$ *where* A *is an* $m \times n$ *matrix of rank* m. *Then the following are equivalent:*

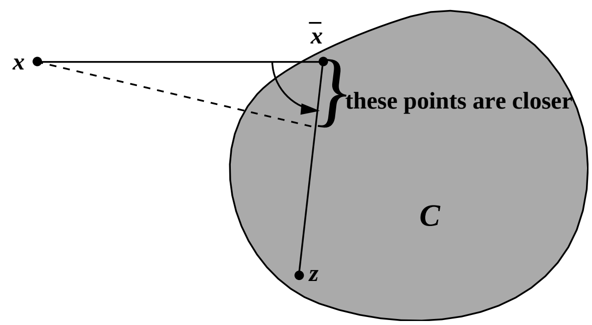

Figure 19: Clearly some points on the line segment connecting $\bar{x}$ to z lie closer to x than $\bar{x}$ when the angle is acute as shown here.

1. *$\bar{x}$ is the projection of x on C.*
2. *$\bar{x} = A^T(AA^T)^{-1}Ax$.*
3. *$\bar{x} \in C$ and $x^Tz = \bar{x}^Tz$ for all $z \in C$.*

Note: The set C is the span of the set of n-vectors given by the rows of A. The rank assumption simply means that these vectors are linearly independent. It is easy to check that A has rank m if and only if AA^T is nonsingular.

PROOF. *(1) implies (2):* By definition, $\bar{x}$ solves $\min_{y\in\mathbb{R}^n} f(y)$ where $f(y) = \|x - A^Ty\|^2 = x^Tx - 2(Ax)^Ty + y^TAA^Ty$. Let $\bar{y}$ denote a point at which the gradient of f vanishes:

$$\nabla f(\bar{y}) = -2Ax + 2AA^T\bar{y} = 0.$$

Since AA^T is nonsingular, $\bar{y}$ is uniquely given by

$$\bar{y} = (AA^T)^{-1}Ax.$$

Hence, $\bar{x} = A^T\bar{y} = A^T(AA^T)^{-1}Ax$.

(2) implies (3): Suppose that $\bar{x} = A^T(AA^T)^{-1}Ax$. Then, $\bar{x} = A^T\bar{y}$, where $\bar{y} = (AA^T)^{-1}Ax$. Hence, $\bar{x}$ belongs to C. Suppose that z also belongs to C. That is, $z = A^Ty$ for some $y \in \mathbb{R}^m$. Then,

$$z^T\bar{x} = y^TAA^T(AA^T)^{-1}Ax = y^TAx = z^Tx.$$

(3) implies (1): Suppose that $\bar{x} \in C$ and $x^Tz = \bar{x}^Tz$ for all $z \in C$. Picking $z = \bar{x}$, we see that $x^T\bar{x} = \bar{x}^T\bar{x}$. That is,

$$\bar{x}^T(x - \bar{x}) = 0.$$

Yet, for any z in C we have

$$z^T(x - \bar{x}) = 0.$$

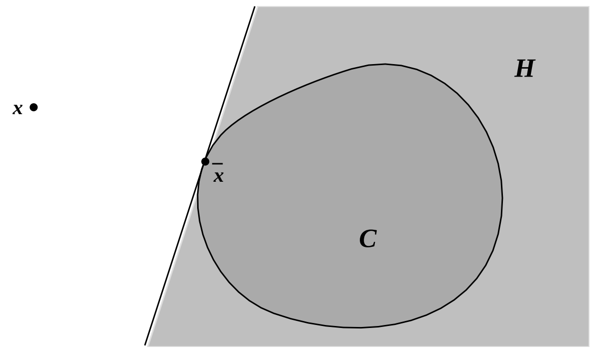

Figure 20: The separating hyperplane theorem.

Combining these two equations, we see that

$$(z-\bar{x})^T(x-\bar{x}) = 0.$$

Therefore, Theorem 21.2 implies that $\bar{x}$ is the projection of x on C. □

22 Supporting Hyperplane Theorem

22.1 DEFINITION. A *halfspace* H is a set of the form $\{z : a^T z \leq b\}$, where $a \neq 0$. The *boundary* ∂H is the hyperplane $\{z : a^T z = b\}$.

The projection theorems of the previous section provide the key tool to proving the important *supporting hyperplane theorem:*

22.2 THEOREM. *Suppose that C is a nonempty closed convex set in $\mathbb{R}^n$ and that x is a point not in C. Then there exists a halfspace H such that $C \subset H$, $C \cap \partial H \neq \emptyset$, and $x \notin H$.*

PROOF. Let $\bar{x}$ denote the projection of x on C. Let $a = x - \bar{x}$. Since $x \notin C$ and $\bar{x} \in C$, we see that $a \neq 0$. Put $H = \{z : a^T z \leq a^T \bar{x}\}$. By Theorem 21.2, C is a subset of H. Since $a^T x - a^T \bar{x} = \|a\|^2 > 0$, it follows that $x \notin H$. Since $\bar{x} \in C$ and $\bar{x} \in \partial H$, we get that $C \cap \partial H \neq \emptyset$. □

Measure and Integration

This chapter is devoted to integration on abstract spaces. As special cases, it covers the Riemann integral, line and surface integrals, and Stieltjes integrals.

23 Motivation

The integral introduced in elementary calculus courses is called the Riemann integral. Let's briefly review the definition of the integral from a to b of a real-valued function f. Let $\mathcal{P}$ denote a partition of the interval $[a,b]$:

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Associated with this partition, is an upper estimate of the integral

$$U(f,\mathcal{P}) = \sum_{i=1}^{n} \sup_{x_{i-1} \le x \le x_i} f(x)(x_i - x_{i-1})$$

and a lower estimate

$$L(f,\mathcal{P}) = \sum_{i=1}^{n} \inf_{x_{i-1} \le x \le x_i} f(x)(x_i - x_{i-1}).$$

Clearly,

$$L(f,\mathcal{P}) \le U(f,\mathcal{P}).$$

The function f is said to be ***Riemann integrable*** over the interval $[a,b]$ if

$$\sup_{\mathcal{P}} L(f,\mathcal{P}) = \inf_{\mathcal{P}} U(f,\mathcal{P}).$$

The basic result regarding Riemann integration is that if f is continuous, then the Riemann integral exists.

There are at least three problems with the Riemann integral. The first problem is that highly discontinuous functions aren't integrable. For example, consider the function f that is one at every irrational point and is zero at every rational. Then, for every partition $\mathcal{P}$,

$$U(f,\mathcal{P}) = b - a$$

and

$$L(f, \mathcal{P}) = 0.$$

The second problem is that one would like to be able to integrate functions whose domain is more general than simply the reals. Of course, Riemann integrals are extended to functions defined on $\mathbb{R}^n$, but even that is not as general as one would prefer.

The third problem is that one would often like to interchange a limit with an integral. Although it is not apparent from the definition given above, it turns out that justifying such an interchange for Riemann integrals is difficult.

To circumvent these difficulties, the idea is to partition the range instead of the domain (after all, the range is always the reals). Suppose first that f is a positive function defined on an arbitrary set E and partition $[0, n)$ using dyadic intervals $[(k-1)/2^n, k/2^n)$. Let

$$B_{k,n} = \{x \in E : f(x) \in [(k-1)/2^n, k/2^n)\}$$

denote the set of points in the domain that map into $[(k-1)/2^n, k/2^n)$. The following sum is a lower estimate of the area under f:

$$\sum_{k=1}^{n2^n} \frac{k}{2^n} \mu(B_{k,n}),$$

where $\mu(B_{k,n})$ denotes the length or, more generally, the measure of $B_{k,n}$. As n increases, this sum increases. Therefore, it has a limit (possibly infinite) which is called the *Lebesgue integral* of f over E:

$$\int_E f(x)\mu(dx) = \lim_n \sum_{k=1}^{n2^n} \frac{k}{2^n} \mu(B_{k,n}).$$

Note that μ is a function from subsets of E into $\mathbb{R}_+$. To capture the notion of being a "measure" of the subsets, μ should possess the following properties:

1. if $A_1, A_2, \ldots$, are disjoint subsets of E, then $\mu(\cup_n A_n) = \sum_n \mu(A_n)$;

2. $\mu(\emptyset) = 0$.

A function on subsets of E with these two properties is called a *measure* on E.

At this point the picture seems pretty clear. All that remains is to construct the measure μ in the cases of interest (such as the usual notion of length on $\mathbb{R}$). However, the following theorem due to Ulam shows that there aren't many measures that can be constructed this way.

23.1 THEOREM. *If μ is a finite measure defined on all subsets of $[0, 1]$, then there exists a countable collection of points $x_1, x_2, \ldots$ in $[0, 1]$ such that $\mu(\{x_1, x_2, \ldots\}^c) = 0$.*

Hence, there does not exist a measure defined on all subsets of $[0, 1]$ for which $\mu([a, b]) = b - a$. That is, there does not exist a measure which corresponds to our idea of length. The problem is that we have asked for too much. It is not necessary (and evidently not possible) to define our measures on all subsets of E. The collections of sets on which we will define our measures will be called algebras. This is the subject of the next section.

24 Algebras

Let E be a set (generally this set will be uncountably infinite although we by no means require this). We wish to assign "measures" to the sizes of various subsets of E. It would be nice to assign a measure to arbitrary subsets, but as we shall see this is impossible to do in such a way that certain natural additivity properties hold. Hence, we must restrict our attention only to certain subsets of E. We will call such subsets *measurable*. If a set A is measurable, it stands to reason that its complement should also be measurable (and its measure should be the total measure of E minus the measure of A). Given a finite disjoint collection of measurable sets, it makes sense that their union should be measurable since the measure of the union should be the sum of the measures of each set. Using the fact that complements of measurable sets are measurable, it is easy to see that finite non-disjoint unions of measurable set should also be measurable since they can be pieced together from disjoint measurable sets. Finally, it is reasonable to assume that countable unions of measurable sets should also be measurable, since the sums involved in the appropriate definition involves only positive numbers and so must either converge to a finite number or to infinity. A collection of measurable sets will be called a *σ-algebra* on E. To summarize the foregoing, a σ-algebra is a non-empty collection $\mathcal{E}$ of subsets of E with the following two properties:

$$A \in \mathcal{E} \Rightarrow E \setminus A \in \mathcal{E},$$

$$A_1, A_2, \ldots \in \mathcal{E} \Rightarrow \cup_1^\infty A_n \in \mathcal{E}.$$

In other words, a σ-algebra is a collection of subsets of E that is closed under the operations of complementation and countable unions. It follows that a σ-algebra is closed under finite unions, finite intersections, and countable intersections as well. In particular, the sets $\emptyset$ and E belong to every σ-algebra on E.

The simplest σ-algebra on E is $\mathcal{E} = \{\emptyset, E\}$; it is called the *trivial σ-algebra*. The largest is the collection of all subsets; it is called the *discrete σ-algebra*.

The intersection of an arbitrary family (countable or uncountable) of σ-algebra on E is again a σ-algebra. If $\mathcal{C}$ is a collection of subsets of E, the intersection of all σ-algebras containing $\mathcal{C}$ is the smallest σ-algebra that contains $\mathcal{C}$; it is called the σ-algebra *generated* by $\mathcal{C}$ and is denoted by $\sigma(\mathcal{C})$.

If E is a metric space, then the σ-algebra generated by the collection of all open subsets is called the *Borel σ-algebra* on E; it is denoted by $\mathcal{B}(E)$, and its elements are called Borel sets. Thus, every open set, every closed set, every set obtained from open and closed sets through various set operations are all Borel sets.

Monotone Class Theorem

This is a very useful theorem which simplifies the task of showing that a given colleciton is a σ-algebra. Throughout this subsection, E is an arbitrary set.

A collection $\mathcal{C}$ of subsets of E is called a *π-system* if it is closed under finite intersections, that is, if

24.1 $$A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}.$$

A collection $\mathcal{D}$ of subsets of E is called a *d-system* on E if

24.2
(i) $E \in \mathcal{D}$,
(ii) $A, B \in \mathcal{D}$ and $B \subset A \Rightarrow A \setminus B \in \mathcal{D}$,
(iii) $(A_n) \subset \mathcal{D}$ and $A_n \nearrow A \Rightarrow A \in \mathcal{D}$.

On the last line, we wrote $(A_n) \subset \mathcal{D}$ to mean that (A_n) is a sequence of elements of $\mathcal{D}$, and we wrote $A_n \nearrow A$ to mean that $A_1 \subset A_2 \subset \cdots$ and $\cup_n A_n = A$.

24.3 PROPOSITION. *Let $\mathcal{E}$ be a collection of subsets of E. Then, $\mathcal{E}$ is a σ-algebra on E if and only if $\mathcal{E}$ is both a π-systme and a d-system on E.*

PROOF. If $\mathcal{E}$ is σ-algebra then it is obviously a π-system and a d-system. To show the converse, suppose that $\mathcal{E}$ is both a π-system and a d-system. Now, 24.2i and 24.2ii show that $\mathcal{E}$ is closed under complements. Since $A \cup B = (A^c \cap B^c)^c$, this implies that $\mathcal{E}$ is closed under unions (if $A, B \in \mathcal{E}$ then $A^c, B^c \in \mathcal{E}$, and thus $A^c \cap B^c \in \mathcal{E}$ since $\mathcal{E}$ is a π-system, and hence $(A^c \cap B^c)^c \in \mathcal{E}$). This implies that $\mathcal{E}$ is closed under countable unios as well: if $A_1, A_2, \ldots \in \mathcal{E}$, put

$$B_1 = A_1, \quad B_2 = A_2, \quad B_3 = A_3, \ldots.$$

Each B_n belongs to $\mathcal{E}$ by what we have just shown. Obviously, $B_1 \subset B_2 \subset \cdots$ and $\cup_n B_n = cup_n A_n$. Thus, using property 24.2iii of athe d-system $\mathcal{E}$, we see that $\cup_n A_n \in \mathcal{E}$. □

The following lemma is needed in the proof of the main theorem. Its proof is obtained by checking the conditions of 24.2 one by one; we leave it as an exercise.

24.4 LEMMA. *Let $\mathcal{D}$ be a d-system on E. Fix $D \in \mathcal{D}$ and let*

$$|\hat{d} = \{A \in \mathcal{D} : A \cap D \in \mathcal{D}\}.$$

Then, $\cap\mathcal{D}$ is again a d-system.

The following the main result of this section. It is called Dynkin's monotome class theorem.

24.5 THEOREM. *If a d-system contains a π-system, then it contains also the σ-algebra generated by that π-system.*

PROOF. Let $\mathcal{C}$ be a π-system. Let $\mathcal{D}$ be the smallest d-system on E that contains $\mathcal{C}$. We need to show that $\mathcal{D} \supset \sigma(\mathcal{C})$. To that end, since $\sigma(\mathcal{C})$ is the smallest σ-algebra containing $\mathcal{C}$, it is sufficient to show that $\mathcal{D}$ is a σ-algebraḞor this, it is in turn sufficient to show that $\mathcal{D}$ is a π-system (and then Proposition 24.3 implies that the d-system $\mathcal{D}$ is a σ-algebra).

Fix $B \in \mathcal{C}$ and let $\mathcal{D}_1 = \{A \in \mathcal{D} : A \cap B \in \mathcal{D}\}$. Since $B \in \mathcal{C} \subset \mathcal{D}$, Lemma 24.4 shows that $\mathcal{D}_1$ is a d-system. Moreover, $\mathcal{D}_1 \supset \mathcal{C}$ since $A \cap B \in \mathcal{C} \subset \mathcal{D}$ for every $A \in \mathcal{C}$ by the fact that $\mathcal{C}$ is a π-system. So $\mathcal{D}_1$ must contain the smallest d-system containing $\mathcal{C}$, that is, $\mathcal{D}_1 \supset \mathcal{D}$. In other words, $A \cap B \in D$ for every $A \in \mathcal{D}$ and $B \in \mathcal{C}$.

Next, fix $A \in \mathcal{D}$ and let $\mathcal{D}_2 = \{B \in \mathcal{D} : A \cap B \in \mathcal{D}\}$. We have just shown that $\mathcal{D}_2 \supset \mathcal{C}$. Moreover, by Lemma 24.4 again, $\mathcal{D}_2$ is a d-system. THus, $\mathcal{D}_2 \supset \mathcal{D}$. In other words, $A \cap B \in \mathcal{D}$ for every $A \in \mathcal{D}$ and $\mathcal{B} \in \mathcal{D}$, that is, $\mathcal{D}$ is a π-system. This completes the proof. $\square$

Exercises:

24.1 *Partitions.* A partition of E is a countable disjointed collection of subsets whose union is E. It is called a finite partition if it has only finitely many elements.

1. Let $\{A, B, C\}$ be a partition of E. Describe the σ-algebra generated by this partition.
2. Let $\mathcal{C}$ be a partition of E. Let $\mathcal{E}$ be the collection of all countable unions of elements of $\mathcal{C}$. Show that $\mathcal{E}$ is a σ-algebra. Show that, in fact, $\mathcal{E} = \sigma(\mathcal{C})$.

Generally, if $\mathcal{C}$ is not a partition, the elements of $\sigma(\mathcal{C})$ cannot be obtained through such explicit constructions.

24.2 Let $\mathcal{B}$ and $\mathcal{C}$ be two collections of subsets of E. If $\mathcal{B} \subset \mathcal{C}$, then $\sigma(\mathcal{B}) \subset \sigma(\mathcal{C})$. If $\mathcal{B} \subset \sigma(\mathcal{C}) \subset \sigma(\mathcal{B})$, then $\sigma(\mathcal{B}) = \sigma(\mathcal{C})$. Show these.

24.3 *Borel σ-algebra on* $\mathbb{R}$. Show that $\mathcal{B}(\mathbb{R})$ is generated by the collection of all open intervals. Hint: recall that every open subset of $\mathbb{R}$ is a countable union of open intervals.

24.4 *Continuation.* Show that every interval of $\mathbb{R}$ is a Borel set. In particular, $(-\infty, x)$, $(-\infty, x]$, $(x, y]$, $[x, y]$ are all Borel sets. Every singleton $\{x\}$ is a Borel set.

24.5 Show that $\mathcal{B}(\mathbb{R})$ is also generated by any one of the following:

1. the collection of all intervals of the form (x, ∞),
2. the collection of all intervals of the form $(x, y]$,
3. the collection of all intervals of the form $[x, y]$,
4. the collection of all intervals of the form $(-\infty, x]$,
5. the collection of all intervals of the form (x, ∞) with x rational.

25 Measurable Spaces and Functions

A *measurable space* is a pair $(E, \mathcal{E})$ where E is a set and $\mathcal{E}$ is a σ-algebra on E. Then, the elements of $\mathcal{E}$ are called *measurable sets*. When E is a metric space and $\mathcal{E} = \mathcal{B}(E)$, the Borel σ-algebra on E, the measurable sets are also called *Borel sets*.

Let $(E, \mathcal{E})$ and $F, \mathcal{F})$ be measurable spaces and let f be a mapping from E into F. Then, f is said to be *measurable* relative to $\mathcal{E}$ and $\mathcal{F}$ if $f^{-1}(B) \in \mathcal{E}$ for every $B \in \mathcal{F}$ (these are the functions we wish to be able to integrate). If E and F are metric spaces and $\mathcal{E} = \mathcal{B}(E)$ and $\mathcal{F} = \mathcal{B}(F)$ and $f : E \mapsto F$ is measurable relative to $\mathcal{E}$ and $\mathcal{F}$, tthen f is also called a *Borel function*.

Measurable Functions

The following proposition reduces the checks for measurability:

25.1 PROPOSITION. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. In order for $f : E \mapsto F$ to be measurable relative to $\mathcal{E}$ and $\mathcal{F}$, it is necessary and sufficient that $f^{-1}(B) \in E$ for every $B \in \mathcal{F}_0$ for some collection $\mathcal{F}_0$ that generates $\mathcal{F}$.*

PROOF. Necessity part is trivial. To prove the sufficiency, let $\mathcal{F}_0 \subset \mathcal{F}$ be such that $\sigma(\mathcal{F}_0) = \mathcal{F}$ and suppose that $f^{-1}(B) \in \mathcal{E}$ for every $B \in \mathcal{F}_0$. We need to show that, then,

$$\mathcal{F}_1 = \{B \in \mathcal{F} : f^{-1}(B) \in \mathcal{E}\}$$

is equal to $\mathcal{F}$. For this, it is sufficient to show that $\mathcal{F}_1$ is a σ-algebra, since $\mathcal{F}_1 \supset \mathcal{F}_0$ by hypothesis and $\mathcal{F}$ is the smallest σ-algebra containing $\mathcal{F}_0$. But checking that $\mathcal{F}_1$ is a σ-algebra is easy in view of the relations given in Exercise 2.1. $\square$

Borel Functions

Let E and F be metric spaces and let $\mathcal{E}$ and $\mathcal{F}$ be their respective Borel σ-algebras. Let $f : E \mapsto F$. Since $\mathcal{F}$ is generated by the open subsets of F, in order for f to be a Borel function, it is necessary and sufficient that $f^{-1}(B) \in \mathcal{E}$ for every open subset B of F; this is an immediate corollary of the preceding proposition. In particular, if f is continuous, then $f^{-1}(B)$ is open in E for every open $B \subset F$. Thus, every continuous function $f : E \mapsto F$ is Borel measurable. The converse is generally false.

Compositions of Functions

Let $(E, \mathcal{E})$, $(F, \mathcal{F})$, and $(G, \mathcal{G})$ be measurable spaces. Let $f : E \mapsto F$ and $g : F \mapsto G$. Then, their composition $g \circ f : x \mapsto g(f(x))$ is a mapping from E into G. The following proposition will be recalled by the phrase "measurable functions of measurable functions are measurable".

25.2 PROPOSITION. *If f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and if g is measurable relative to $\mathcal{F}$ and $\mathcal{G}$, then $g \circ f$ is measurable relative to $\mathcal{E}$ and $\mathcal{G}$.*

PROOF. Recall that $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$ for every $C \subset G$. If $C \in \mathcal{G}$ and g is measurable, then $B = g^{-1}(C)$ is in $\mathcal{F}$. Therefore, if f is measurable, $f^{-1}(B) = f^{-1}(g^{-1}(C))$ is in $\mathcal{E}$ for every $C \in \mathcal{G}$. $\square$

Numerical Functions

By a *numerical function* on E, we mean a mapping from E into $\bar{\mathbb{R}}$ or some subset thereof. Such a function is said to be *positive* if all its values are in $\bar{\mathbb{R}}_+$ and is said to be real-valued if all its values are in $\mathbb{R}$. If $(E, \mathcal{E})$ is a measurable space and f is a numerical function on E, then f is said to be *$\mathcal{E}$-measurable* if it is measurable with respect to $\mathcal{E}$ and $\mathcal{B}(\bar{\mathbb{R}})$.

Let $(E, \mathcal{E})$ be a measurable space and let f be a numerical function on E. Using Proposition 25.1 with $F = \bar{\mathbb{R}}$ and $\mathcal{F} = \mathcal{B}(\bar{\mathbb{R}})$ and recalling Exercise 24.5, we see that the following holds.

25.3 PROPOSITION. *The numerical function f is $\mathcal{E}$-measurable if and only if any one of the following is true:*

1. $\{x : f(x) \leq r\} \in \mathcal{E}$ *for every* $r \in \mathbb{R}$,
2. $\{x : f(x) > r\} \in \mathcal{E}$ *for every* $r \in \mathbb{R}$,
3. $\{x : f(x) < r\} \in \mathcal{E}$ *for every* $r \in \mathbb{R}$, *etc.*

25.4 COROLLARY. *Suppose that* $f : E \mapsto F$ *where* F *is a countable subset of* $\bar{\mathbb{R}}$*. Then,* f *is* $\mathcal{E}$*-measurable if and only if* $\{x : f(x) = a\} \in \mathcal{E}$ *for every* $a \in F$*.*

PROOF. Necessity is trivial since each singleton $\{a\}$ is a Borel set. For the sufficiency, fix $r \in \mathbb{R}$ and note that $\{x : f(x) \leq r\}$ is the union of $\{x : f(x) = a\}$ over all $a \leq r$, $a \in F$, and therefore belongs to $\mathcal{E}$ since it is a countable union of the sets $\{x : f(x) = a\} \in \mathcal{E}$. Thus, f is $\mathcal{E}$-measurable by the preceeding proposition. □

Positive and Negative Parts of a Function

Let $(E, \mathcal{E})$ be a measurable space. Let f be a numerical function on E. Then,[3]

$$f^+ = f \vee 0, \quad f^- = -(f \wedge 0)$$

are called the positive part of f and negative part of f, respectively. Note that both f^+ and f^- are positive functions and

$$f = f^+ - f^-.$$

25.5 PROPOSITION. *The function* f *is* $\mathcal{E}$*-measurable if and only if both* f^+ *and* f^- *are* $\mathcal{E}$*-measurable.*

The proof is left as an exercise. The decomposition $f = f^+ - f^-$ enables us to state most results for positive functions only, since it is easy to obtain the corresponding result for arbitrary f.

Indicators and Simple Functions

Let $A \subset E$. Its *indicator*, denoted by 1_A, is defined by

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Obviously, 1_A is $\mathcal{E}$-measurable if and only if $A \in \mathcal{E}$.

A function f on E is said to be *simple* if it has the form

25.6 $$f = \sum_1^n a_i 1_{A_i}$$

[3]For $a, b \in \bar{\mathbb{R}}$ we write $a \vee b$ for the maximum of a and b, and $a \wedge b$ for the minimum. The notation extends to functions: $f \vee g$ is the function whose value at x is $f(x) \vee g(x)$; similarly for $f \wedge g$.

for some integer n, real numbers $a_1, \ldots, a_n$, and measurable sets $A_1, \ldots, A_n$. It is clear that, then, there exist an integer $m \geq 1$, distinct real numbers $b_1, \ldots, b_m$, and a measurable partition $\{B_1, \ldots, B_m\}$ of E such that $f = \sum_1^m b_i 1_{B_i}$, this latter representation is called the *canonical form* of the simple function f.

Every simple function of E is $\mathcal{E}$-measurable; this is immediate from Corollary 25.4 applied to the canonical form of f. Conversely, if f is $\mathcal{E}$-measurable, takes only finitely many values, and all those values are real, then f is simple.

In particular, every constant is a simple function. Moreover, if f and g are simple, then so are

$$f + g, \quad f - g, \quad fg, \quad f/g, \quad f \vee g, \quad f \wedge g,$$

except that, in the case of f/g one must make sure that g is never 0.

Approximations by Simple Functions

We start by constructing a sequence of simple functions that approximate the identity function d from $\bar{\mathbb{R}}_+$ into $\bar{\mathbb{R}}_+$. For each $n \in \mathbb{N}$, let

25.7 $$d_n(x) = \begin{cases} k/2^n & \text{if } \frac{k}{2^n} \leq x < \frac{k+1}{2^n}, \quad k \in \{0, 1, \ldots, n2^n - 1\}, \\ n & \text{if } x \geq n. \end{cases}$$

The figure below is for d_2. The following lemma should be self-evident.

25.8 LEMMA. *Each d_n is a simple Borel function on $\bar{\mathbb{R}}_+$. Each d_n is right-continuous and increasing. The sequence (d_n) is increasing pointwise to the function $d : x \mapsto x$.*

The following theorem characterizes all $\mathcal{E}$-measurable positive functions, and via Proposition 25.5, all $\mathcal{E}$-measurable functions.

25.9 THEOREM. *A positive function on E is $\mathcal{E}$-measruable if and only if it is the limit of an increasing sequence of simple positive functions.*

PROOF. *Necessity.* Let $f : E \mapsto \bar{\mathbb{R}}_+$ be $\mathcal{E}$-measurable. Let the d_n be defined by 25.7. Since each d_n is a measurable function from $\bar{\mathbb{R}}_+$ into $\bar{\mathbb{R}}_+$, and since measurable functions of measurable functions are measurable, the function $f_n = d_n \circ f$ is $\mathcal{E}$-measurable for each n. Since d_n is simple, so is f_n. Finally, $\lim f_n(x) = \lim d_n(f(x)) = f(x)$ since $\lim d_n(y) = y$ for all $\bar{\mathbb{R}}_+$. Thus, f is the limit of the sequence (f_n) of simple positive functions and $f_1 \leq f_2 \leq \cdots$ since $d_1 \leq d_2 \leq \cdots$.

Sufficiency. Let $f_1 \leq f_2 \leq \cdots$ be simple and positive and let $f = \lim f_n$. Now, for each $x \in E$ and $r \in \mathbb{R}$, we have $f(x) \leq r$ if and only if $f_n(x) \leq r$ for all n; thus,

$$\{x \in E : f(x) \leq r\} = \cap_{n=1}^{\infty} \{x \in E : f_n(x) \leq r\}$$

for each $r \in \mathbb{R}$. Since the f_n are simple (and therefore measurable), each factor on the right side belongs to $\mathcal{E}$ and, therefore, so does the intersection. Hence, f is $\mathcal{E}$-measurable by Proposition 25.3. $\square$

Limits of Sequences of Functions

Let $(E, \mathcal{E})$ be a measurable space and let (f_n) be a sequence of numerical functions on E.

25.10 THEOREM. *Suppose that each f_n is $\mathcal{E}$-measurable. The, each one of*

$$\inf f_n, \quad \sup f_n, \quad \liminf f_n, \quad \limsup f_n$$

is again $\mathcal{E}$-measurable. Moreover, if $\lim f_n$ exists, then it is $\mathcal{E}$-measurable.

PROOF. For $x \in E$ and $r \in \mathbb{R}$, we have $\inf f_n(x) \geq r$ if and only if $f_n(x) \geq r$ for all n. Thus, for each $r \in \mathbb{R}$,

$$\{x \in E : \inf f_n(x) \geq r\} = \cap_n \{x \in E : f_n(x) \geq r\}.$$

Now, $\{x : f_n(x) \geq r\} \in \mathcal{E}$ for each n by the measurability of f_n, and therefore the intersection on the right side belongs to $\mathcal{E}$ since $\mathcal{E}$ is closed under countable intersections. Thus, $\inf f_n$ is $\mathcal{E}$-measurable by Proposition 25.3.

The proof that $\sup f_n$ is $\mathcal{E}$-measurable follows via similar reasoning upon noting that

$$\{x \in E : \sup f_n(x) \leq r\} = \cap_n \{x \in E : f_n(x) \leq r\}.$$

It follows from these that

$$\liminf f_n = \sup_m \inf_{n \geq m} f_n, \quad \limsup f_n = \inf_m \sup_{n \geq m} f_n$$

are both $\mathcal{E}$-measurable. Finally, $\lim f_n$ exists if and only if $\liminf f_n = \limsup f_n$, and then $\lim f_n$ is the common limit; so, it must be $\mathcal{E}$-measurable. □

Monotone Classes of Functions

Often we are interested in showing that a certain property holds for all measurable functions. The following are useful in such quests.

Let $\mathcal{M}$ be a collection of positive functions on E. Then, $\mathcal{M}$ is called a *monotone class of functions* provided that

25.11
(i) $1 \in \mathcal{M}$,
(ii) $f, g \in \mathcal{M}$, and $a, b \in \mathbb{R}_+ \Rightarrow af + bg \in \mathcal{M}$,
(iii) $(f_n) \subset \mathcal{M}$, and $f_n \nearrow f \Rightarrow f \in \mathcal{M}$.

The following is called the monotone class theorem for functions.

25.12 THEOREM. *Let $\mathcal{M}$ be a monotone class of functions on E. Suppose that $1_A \in \mathcal{M}$ for every $A \in \mathcal{C}$ for some π-system $\mathcal{C}$ that generates the σ-algebra $\mathcal{E}$. Then, $\mathcal{M}$*

includes all positive $\mathcal{E}$-measurable functions and all bounded $\mathcal{E}$-measurable functions.

PROOF. We start by showing that $1_A \in \mathcal{M}$ for every $A \in \mathcal{E}$. To this end, let

$$\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{M}\}.$$

Using the properties 25.11 of $\mathcal{M}$, it is easy to check that $\mathcal{D}$ is a d-system. Moreover, $\mathcal{D} \supset \mathcal{C}$ by hypothesis. Thus, by Dynkin's monotone class theorem, $\mathcal{D} \supset \sigma(\mathcal{C}) = \mathcal{E}$. In other words, $1_A \in \mathcal{M}$ for every $A \in \mathcal{E}$.

Consequently, in view of property 25.11(ii), $\mathcal{M}$ includes all simple $\mathcal{E}$-measurable functions.

Let f be a positive $\mathcal{E}$-measurable function. By Theorem 25.9, there exists a sequence of positive simple functions $f_n \nearrow f$. Since each f_n in in $\mathcal{M}$ by the preceeding step, 25.11(iii) implies that f is in $\mathcal{M}$.

$\square$

Notation

We shall write $f \in \mathcal{E}$ to mean that f is an $\mathcal{E}$-measurable function. Thus, $\mathcal{E}$ stands both for a σ-algebra and for the collection of all numerical functions measurable with respect to it. Furthermore, we shall use the notation

$$\mathcal{F}_+ = \{f \in \mathcal{F} : f \geq 0\}$$

for any collection of $\mathcal{F}$ of numerical functions. Thus, in particular, $\mathcal{E}_+$ is the collection of all positive $\mathcal{E}$-measurable functions.

Exercises:

25.1 *Trace spaces.* Let $(E, \mathcal{E})$ be a measurable space and let $D \subset E$ be fixed. Show that

$$\mathcal{D} = \{A \cap D : A \in \mathcal{E}\}$$

is a σ-algebra on $\mathcal{D}$. Then, $\mathcal{D}$ is called the trace of $\mathcal{E}$ on D, and $(D, \mathcal{D})$ is called the trace of $(E, \mathcal{E})$ on D.

25.2 *σ-algebra generated by a function.* Let E be a set and let $(F, \mathcal{F})$ be a measurable space, Let f be a mapping from E into F and set

$$f^{-1}(\mathcal{F}) = \{f^{-1}(B) : B \in \mathcal{F}\}.$$

Use Exercise 2.1 to show that $f^{-1}(\mathcal{F})$ is a σ-algebra on E; it is called the σ-algebra on E generated by f. It is the smallest σ-algebra on E such that f is measurable relative to it and $\mathcal{F}$.

25.3 *Product spaces.* Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A rectangle $A \times B$ is said to be measurable if $A \in \mathcal{E}$ and $B \in \mathcal{F}$. Show that the collection of all measurable rectangles form a π-system. The σ-algebra on $E \times F$ generated by that π-system is denoted by $\mathcal{E} \otimes \mathcal{F}$ and is called the product σ-algebra. Further, $(E \times F, \mathcal{E} \otimes \mathcal{F})$ is called the product of $(E, \mathcal{E})$ and $(F, \mathcal{F})$, and is denoted by $(E, \mathcal{E}) \times (F, \mathcal{F})$ also. If $(E, \mathcal{E}) = (F, \mathcal{F})$, then it is usual to write E^2 for $E \times F$ and $\mathcal{E}^2 = \mathcal{E} \otimes \mathcal{F}$. In particular, $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) = (\mathbb{R}, \mathcal{B}(\mathbb{R})) \times (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and by an obvious extension, $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) = (\mathbb{R}, \mathcal{B}(\mathbb{R})) \times \cdots \times (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, n times.

25.4 *Continuation.* Let $(E, \mathcal{E})$, $(F, \mathcal{F})$, $(G, \mathcal{G})$ be measurable spaces. Let $f : E \mapsto F$ be measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and let $g : E \mapsto G$ be measurable relative to $\mathcal{E}$ and $\mathcal{G}$. Then,

$$h(x) = (f(x), g(x)), \quad x \in E,$$

defines a mapping from E into $F \times G$. Show that h is measurable relative to $\mathcal{E}$ and $\mathcal{F} \otimes \mathcal{G}$.

In particular, a function $f : E \mapsto \mathbb{R}^n$ is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R}^n)$ if and only if its coordinates are measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R})$; recall that the coordinates of f are the functions $f_1, \ldots, f_n$ such that $f(x) = (f_1(x), \ldots, f_n(x))$, $x \in E$.

25.5 *Discrete spaces.* A measurable space $(E, \mathcal{E})$ is said to be *discrete* if E is countable and $\mathcal{E}$ is the σ-algebra of all subsets of E. Then, show that every numerical function of E is $\mathcal{E}$-measurable.

25.6 Suppose that $\mathcal{E}$ is generated by a countable partition of E. Show that, then, a numerical function on E is $\mathcal{E}$-measurable if and only if it is constant over each member of that partition.

25.7 *Approximation by simple functions.* Show that a numerical function of E is $\mathcal{E}$-measruable if and only if it is the limit of a sequence of simple functions.

25.8 *Arithmetic operations.* Let f and g be $\mathcal{E}$-measurable. Show that, then, each one of

$$f + g, \quad f - g, \quad f \cdot g, \quad f/g, \quad f \vee g, \quad f \wedge g$$

is $\mathcal{E}$-measurable provided that it be well-defined.

25.9 *Functions on* $\mathbb{R}$. Let $f : \mathbb{R} \mapsto \mathbb{R}_+$ be increasing. Show that it is a Borel function.

25.10 *Step functions.* A function $f : \mathbb{R} \mapsto \mathbb{R}$ is called a step function if it has the form

$$f = \sum_1^\infty a_i 1_{A_i}$$

where each A_i is an interval. Show that every such f is a Borel function.

25.11 *Right-continuous functions.* Show that every right-continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ is Borel measurable. SImilarly, every left-continuous function is Borel. Hint for right-continuous f: define $d_n(x) = (k+1)/2^n$ if $k/2^n \le x < (k+1)/2^n$ for some $k = 0, 1, 2, \ldots$ for $n = 1, 2, \ldots$. Show that d_n is Borel. Let $f_n(x) = f(d_n(x))$. Show that each f_n is a step function, and show that $f_n \to f$ as $n \to \infty$.

26 Measures

Let $E, \mathcal{E})$ be a measurable space. A *measure* on $(E, \mathcal{E})$ is a mapping $\mu : \mathcal{E} \mapsto \bar{\mathbb{R}}_+$ such that

1. $\mu(\emptyset) = 0$,
2. $\mu(\cup_n A_n) = \sum_n \mu(A_n)$ for every disjointed sequence $(A_n) \subset \mathcal{E}$.

The latter condition is called *countable additivity*.

A *measure space* is a triplet $(E, \mathcal{E}, \mu)$ where E is a set, $\mathcal{E}$ is a σ-algebra on E, and μ is a measure on $(E, \mathcal{E})$.

26.1 PROPOSITION. *Let μ be a measure on $(E, \mathcal{E})$. Then, the following hold for all measurable sets A, B, and A_n, $n \ge 1$:*

Finite additivity: $A \cap B = \emptyset$ *implies that* $\mu(A \cup B) = \mu(A) + \mu(B)$.

Monotonicity: $A \subset B$ *implies that* $\mu(A) \le \mu(B)$.

Sequential continuity: $A_n \nearrow A$ *implies that* $\mu(A_n) \nearrow \mu(A)$.

Boole's inequality: $\mu(\cup_n A_n) \le \sum_n \mu(A_n)$.

PROOF. Finite additivity is a particular instance of the countable additivity of μ: take $A_1 = A$, $A_2 = B$, $A_3 = A_4 = \cdots = \emptyset$. Monotonicity follows from it and the positivity of μ: if $A \subset B$,

$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A)$$

since $\mu(B \setminus A) \ge 0$. Sequential continuity follows from (and is equivalent to) countable additivity: suppose that $A_n \nearrow A$; then,

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus A_2, \cdots$$

are disjoint, their union is A, and the union of the first n is A_n; hence, the sequence of numbers $\mu(A_n)$ increases by the monotonicity of μ, and

$$\lim \mu(A_n) = \lim \mu(\cup_1^n B_i) = \lim_n \sum_1^n \mu(B_i) = \sum_1^\infty \mu(B_i) = \mu(\cup_1^\infty B_i) = \mu(A).$$

Finally, Boole's inequality follows from the observation that

$$\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \leq \mu(A) + \mu(B).$$

□

Arithmetic of Measures

Let $(E, \mathcal{E})$ be a measurable space. If μ is a measure on it and if $c \geq 0$ is a constant, then $c\mu$ is again a measure. If μ and ν are measures on $(E, \mathcal{E})$, so is $\mu + \nu$. If $\mu_1, \mu_2, \ldots$ are measures, then so is $\mu = \sum \mu_m$: it is obvious that $\mu(\emptyset) = 0$, and if $A_1, A_2, \ldots$ are disjoint then

$$\begin{aligned} \mu(\cup_n A_n) &= \sum_m \mu_m(\cup_n A_n) \\ &= \sum_m \sum_n \mu_m(A_n) \\ &= \sum_n \sum_m \mu_m(A_n) \\ &= \sum_n \mu(A_n), \end{aligned}$$

where the crucial step (where the order of summation is changed) is justified by the elementary fact that

$$\sum_m \sum_n a_{mn} = \sum_n \sum_m a_{mn}$$

if $a_{mn} \geq 0$ for all m, n.

Finite, σ-finite, Σ-finite measures

Let μ be a measure on $(E, \mathcal{E})$. It is said to be *finite* if $\mu(E) < \infty$. It is called a *probability measure* if $\mu(E) = 1$. It is said to be σ-finite if there exists a measurable partition (E_n) of E such that $\mu(E_n) < \infty$ for each n. It is said to be Σ-finite if there exist finite measures $\mu_1, \mu_2, \ldots$ such that $\mu = \sum \mu_n$. Note that every finite measure is trivially σ-finite, every σ-finite measure is Σ-finite. The converses are false (see Exercise 26.4).

Specification of Measures

It is generally difficult to specify $\mu(A)$ for each A, simply because there are too many A in a σ-algebra. The following proposition is helpful in reducing the task to specifying $\mu(A)$ for those A belonging to a π-system that generates the given σ-algebra.

26.2 PROPOSITION. *Let μ and ν be measures on $(E, \mathcal{E})$. Suppose that $\mu(E) = \nu(E) < \infty$, and that μ and ν agree on a π-system generating $\mathcal{E}$. Then, $\mu = \nu$.*

PROOF. Let $\mathcal{C}$ be a π-system with $\sigma(\mathcal{C}) = \mathcal{E}$. Suppose that $\mu(A) = \nu(A)$ for every $A \in \mathcal{C}$. We need to show that, then, $\mu(A) = \nu(A)$ for every $A \in \mathcal{E}$. This amounts to showing that

$$\mathcal{D} = \{A \in \mathcal{E} : \mu(A) = \nu(A)\}$$

contains $\mathcal{E}$. Now, $\mathcal{D} \supset \mathcal{C}$ by hypothesis, and it is straightforward to check that $\mathcal{D}$ is a d-system. Thus, by Dynkin's monotone class theorem, $\mathcal{D} \supset \sigma(\mathcal{C}) = \mathcal{E}$. □

26.3 COROLLARY. *LEt μ and ν be probability measures on $\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, $\mu = \nu$ if and only if, for every $x \in \mathbb{R}$,*

$$\mu((-\infty, x]) = \nu((-\infty, x]).$$

PROOF. The collection $\mathcal{C}$ of all intervals of the form $(-\infty, x]$ is a π-system generating $\mathcal{B}(\mathbb{R})$. THus, the preceding proposition applies to prove sufficiency. Necessity is trivial. □

The following proposition extends 26.2 to σ-finite measures.

26.4 PROPOSITION. *Let μ and ν be σ-finite measures on $(E, \mathcal{E})$. Suppose that they agree on a π-system $\mathcal{C}$ generating $\mathcal{E}$. Suppose further that there is a partition (E_n) of E such that $E_n \in \mathcal{C}$ and $\mu(E_n) = \nu(E_n) < \infty$ for every n. Then, $\mu = \nu$.*

PROOF. For each n, define the measures μ_n and ν_n on $(E, \mathcal{E})$ by

$$\mu_n(A) = \mu(A \cap E_n), \quad \nu_n(A) = \nu(A \cap E_n), \quad A \in \mathcal{E}.$$

SInce $E_n \in \mathcal{C}$, and since $A \cap E_n \in \mathcal{C}$ for every $A \in \mathcal{C}$, we have

$$\mu_n(A) = \mu(A \cap E_n) = \nu(A \cap E_n) = \nu_n(A) \text{ for } A \in \mathcal{C}.$$

And, by hypothesis, $\mu_n(E) = \mu(E) = \nu(E) = \nu_n(E) < \infty$. Thus, the last proposition applies to show that $\mu_n = \nu_n$ for each n. This completes the proof since $\mu = \sum \mu_n$ and $\nu = \sum \nu_n$. □

Image of Measure

Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces. Let μ be a measure on $(E,\mathcal{E})$ and let $f : E \mapsto F$ be measurable relative to $\mathcal{E}$ and $\mathcal{F}$. Then,

26.5
$$\mu \circ f^{-1}(B) = \mu(f^{-1}(B)), \quad B \in \mathcal{F},$$

is well-defined since $f^{-1}(B) \in \mathcal{E}$ for each $B \in \mathcal{F}$. It is easy to check that $\nu = \mu \circ f^{-1}$ is a measure on $(F,\mathcal{F})$. It is called the *image of* μ *under* f.

Almost Everywhere

Often we face situations where a certain statement is true for every $x \in E_0$ and E_0 is almost the same as E in the sense that $E_0 \in \mathcal{E}$ and $\mu(E \setminus E_0) = 0$. In that case, we say that the statement is true for *almost every* x in E or that the statement is true almost everywhere.

Incidentally, a set $N \subset E$ is said to be neglibible if there is an $A \in \mathcal{E}$ such that $N \subset A$ and $\mu(A) = 0$. So, a statement holds almost everywhere if and only if it fails only over a neglibible set.

EXAMPLES.

26.6 *Dirac measure.* Let $(E,\mathcal{E})$ be a measurable space. Fix $x \in E$. Define

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

for each $A \in \mathcal{E}$. Then, δ_x is a measure on $(E,\mathcal{E})$. It is called the *Dirac measure* sitting at x.

26.7 *Counting measures.* Let $(E,\mathcal{E})$ be a measurable space and let D be a countable subset of E. Define a measure ν on $(E,\mathcal{E})$ by

$$\nu = \sum_{x \in D} \delta_x.$$

Note that $\nu(A)$ is the number of points in $A \cap D$. Such measures are called counting measures.

26.8 *Discrete measure spaces.* Let E be countable and $\mathcal{E}$ be the collection of all subsets of E. Specifying a measure on $(E,\mathcal{E})$ is equivalent to assigning a number $m(x)$ in $\bar{\mathbb{R}}_+$ to each point x in E and then letting

$$\mu(A) = \sum_{x \in A} m(x), \quad A \in \mathcal{E}.$$

Then, m is called the mass function corresponding to μ. In particular, if $E = \{1, 2, \ldots, n\}$, every measure μ on $(E, \mathcal{E})$ can be regarded a a vector in $\mathbb{R}^n$.

26.9 *Purely atomic measures.* Let $(E, \mathcal{E})$ be a measurable space, let D be a countable subset of E, and let $m(x)$ be a positive number for each $x \in D$. Define

$$\mu(A) = \sum_{x \in D} m(x)\delta_x(A), \quad A \in \mathcal{E}.$$

Then, μ is a measure on $(E, \mathcal{E})$. It puts the mass $m(x)$ at the point x, and there are only countable many points x like that. Such μ are said to be purely atomic, the points x with $\mu(\{x\}) > 0$ are called the atoms of μ.

26.10 *Lebesgue measures.* A measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called the *Lebesgue measure* on $\mathbb{R}$ if $\mu(A)$ is the length of A for every interval A. The collection $\mathcal{C}$ of all intervals form a π-system that generates $\mathcal{B}(\mathbb{R})$ and thus, by Proposition 26.4, there can be at most one such measure. The whole point of all measure theory is the following theorem which, unfortunately, we don't prove.

26.11 THEOREM. *There exists a measure on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *which assigns to each interval* A *its length.*

It is impossible to display $\mu(A)$ explicity for each Borel set A, but countable additivity and various properties list in Proposition 26.1 enable us to figure $\mu(A)$ out for most reasonable sets A. For instance, $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}$, $\mu(A) = 0$ for every countable set $A \subset \mathbb{R}$, $\mu(A) = 0$ for the cantor set A, and so on. Of course, there are many sets with strictly positive measure.

Similarly, Lebesgue measure on $\mathbb{R}^2$ is the "area" measure, Lebesgue measure on $\mathbb{R}^3$ is the "volume" measure, and so on. All Lebesgue measures on $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, etc. are σ-finite.

More generally, given an interval $E \subset \mathbb{R}$, it makes sense to talk of Lebesgue measure on $(E, \mathcal{B}(\mathcal{E}))$; this is the restriction of Lebesgue measure on $\mathbb{R}$ to the trace space $(E, \mathcal{B}(E))$. Similarly, one can talk of Lebesgue measure on a domain in $\mathbb{R}^2$ or on a domain in $\mathbb{R}^n$. In all cases we shall use λ_n to denote the Lebesgue measure on a domain in $\mathbb{R}^n$.

Exercises:

26.1 Show that $\mathcal{D}$ in the proof of 26.2 is a d-system.

26.2 *Restrictions.* Let $(E, \mathcal{E}, \mu)$ be a measure space. Let $D \in \mathcal{E}$ and let $\mathcal{D} = \{A \in \mathcal{E} : A \subset D\}$. Then, $(D, \mathcal{D})$ is the trace of $(E, \mathcal{E})$ on D. Define $\nu(A) = \mu(A)$ for $A \in \mathcal{D}$. Then, ν is a measure on $(D, \mathcal{D})$; it is called the *restriction* of μ to D.

26.3 *Uniform distribution.* Let $D \subset \mathbb{R}$ be an interval of finite length. Let $\mu(B) = \lambda_1(B)/\lambda_1(D)$ for Borel subsets B of D. Show that μ is a probability measure on $(D, \mathcal{D})$ where $\mathcal{D} = \mathcal{B}(D)$. It is called the *uniform distribution* on D.

26.4 Σ*-finiteness.* Let $E = \{a, b\}$ with the discrete σ-algebra, and define $\mu(\{a\}) = 0$, $\mu(\{b\}) = +\infty$. Show that this defines a Σ-finite measure μ that is not σ-finite.

26.5 *Atoms, atomic measures, diffuse measures.* Let $(E, \mathcal{E})$ be such the $\{x\} \in \mathcal{E}$ for every $x \in E$. A point x is said to be an *atom* for the measure μ if $\mu(\{x\}) > 0$. If μ has no atoms, then it is said to be *diffuse*. If μ puts no mass outside the set of its atoms, then it is *purely atomic*. In general, μ will have some atomic part and some diffuse part. This is to show this decomposition.

1. Let μ be finite. Show that it has at most countably many atoms. Hint: let D be the set of atoms, note that $D = \cup_n D_n$ where $D_n = \{x : \mu(\{x\}) \in [1/n,\ 1/(n-1)),\ n = 1, 2, \ldots$. Use the finiteness of μ to conclude that each D_n is a finite set, and therefore, that D must be countable.
2. Let μ be Σ-finite. Show that it has at most countably many atoms.
3. Let D be the set of atoms of a Σ-finite measure μ. Define

$$\nu(A) = \mu(A \cap D), \quad \lambda(A) = \mu(A \cap D^c), \quad A \in \mathcal{E}.$$

Then, ν is purely atomic, λ is diffuse, and

$$\mu = \nu + \lambda.$$

27 Integration

Let $(E, \mathcal{E})$ be a measurable space. Recall that $\mathcal{E}$ stands also for the collection of all $\mathcal{E}$-measurable functions and that $\mathcal{E}_+$ is the sub-collection consisting of positive $\mathcal{E}$-measurable functions. Given a measure μ on $(E, \mathcal{E})$, our aim is to define the "integral of f with respect to μ" for all reasonable functions f in $\mathcal{E}$. We shall denote it by any of the following:

$$\mu f = \int_E \mu(dx) f(x) = \int_E f d\mu.$$

When E is an interval of $\mathbb{R}$ and f is continuous and μ is the Lebesgue measure, the integral will coincide with the usual Riemann integral of f on E. When $E = \{1, \ldots, n\}$ and $\mathcal{E}$ is the discrete σ-algebra, every measure μ is specified by a row vector $(\mu_1, \ldots, \mu_n)$ with μ_i denoting $\mu(\{i\})$, and every function $f \in \mathcal{E}$ corresponds to a column vector $(f_1, \ldots, f_n)$ with $f_i = f(i)$; in this case the integral μf will coincide with the product of the row vector $(\mu_1, \ldots, \mu_n)$ with the column vector with entires $f_1, \ldots, f_n$. As this last case illustrates, it is best to think of the integral as a product. After we define it, we shall show that it has the properties of products.

Definition of the Integral

We define the integral μf in three steps: first for simple positive f, then for $f \in \mathcal{E}_+$, finally for reasonable $f \in \mathcal{E}$.

Step 1. Let f be a nonnegative simple function. If its cannonical form is $f = \sum_1^n a_i 1_{A_i}$, then we define

27.1
$$\mu f = \sum_1^n a_i \mu(A_i).$$

Step 2. Let $f \in \mathcal{E}_+$. Let (d_n) be defined by 25.7 and recall from the proof of Theorem 25.9 that $\lim d_n \circ f = f$. Now, for each n, the function $d_n \circ f$ is simple and positive, and the integral $\mu(d_n \circ f)$ is defined by the preceding step. We shall show in the remarks below that the numbers $\mu(d_n \circ f)$ form an increasing sequence, and hence, $\lim \mu(d_n \circ f)$ exists (it may be $+\infty$). Since $f = \lim d_n \circ f$, we define

27.2
$$\mu f = \lim \mu(d_n \circ f).$$

Step 3. Let $f \in \mathcal{E}$ be arbitrary. Then, f^+ and f^- belong to $\mathcal{E}_+$, and their integrals are defined by the preceding step. Noting that $f = f^+ - f^-$, we define

27.3
$$\mu f = \mu f^+ - \mu f^-$$

provided that at least one term on the right is finite. Otherwise, if $\mu f^+ = \mu f^- = +\infty$, then μf does not exist.

REMARKS: (a) Formula 27.1 holds for nonnegative simple functions even when $\sum_1^n a_i 1_{A_i}$ is not the canonical representation for f:

$$f = \sum_1^n a_i 1_{A_i} = \sum_1^m b_j 1_{B_j} \quad \Rightarrow \quad \mu f = \sum_1^n a_i \mu(A_i) = \sum_1^m b_j \mu(B_j).$$

This is easy to check using the finite additivity of μ.

(b) If f and g are nonnegative simple functions and $a, b \in \mathbb{R}_+$, then $af + bg$ is again a nonnegative simple function, and

$$\mu(af + bg) = a\ \mu f + b\ \mu g.$$

This can be checked using the preceding remark.

(c) If f is a nonnegative simple function, then 27.1 shows that $\mu f \geq 0$ (it can be $+\infty$).

(d) If f and g are nonnegative simple functions and $f \leq g$, then the preceding two remarks applied to f and $g - f$ show that $\mu f \leq \mu g$.

(e) In Step 2 of the definition, we have $d_1 \circ f \leq d_2 \circ f \leq \cdots$ and the preceding remark shows that $\mu(d_1 \circ f) \leq \mu(d_2 \circ f) \leq \cdots$ as claimed.

Integral over a Set

Let f be a measurable function and A a measurable set. Then, $f1_A \in \mathcal{E}$. The *integral of f over A* is defined to be the integral of $f1_A$; it exists if and only if $\mu(f1_A)$ exists. The following notations are used for it:

27.4
$$\mu(f1_A) = \int_A \mu(dx)f(x) = \int_A f d\mu.$$

Integrability

A function $f \in \mathcal{E}$ is said to be *integrable* if μf exists and is a finite number. Thus, $f \in \mathcal{E}$ is integrable if and only if $\mu f^+ < \infty$ and $\mu f^- < \infty$, or equivalently, if and only if $\mu|f| < \infty$ (note that $|f| = f^+ + f^-$).

Elementary Properties

Here are some familiar properties of the integrals. A few others are put into the exercises.

27.5 PROPOSITION.

(a) Positivity. *If $f \in \mathcal{E}_+$, then $\mu f \geq 0$.*

(b) For $f \in \mathcal{E}_+$, $\mu f = 0$ if and only if $f = 0$ almost everywhere.

(c) Monotonicity. *If $f, g \in \mathcal{E}_+$ and $f \leq g$, then $\mu f \leq \mu g$. If $f, g \in \mathcal{E}$ and f, g are integrable, and $f \leq g$, then $\mu f \leq \mu g$.*

(d) Finite additivity over sets. *Let $f \in \mathcal{E}_+$. If $\{A_1, \ldots, A_m\}$ is a measurable partition of $A \in \mathcal{E}$, then*

27.6
$$\int_A f d\mu = \sum_{i=1}^{m} \int_{A_i} f d\mu.$$

PROOF. (a) If $f \geq 0$, then the definition of μf yields $\mu f \geq 0$.

(c) If $0 \leq f \leq g$, then $d_n \circ f \leq d_n \circ g$ and so

$$\mu(d_n \circ f) \leq \mu(d_n \circ g)$$

by the monotonicity of integration for simple functions. Now, the left-hand side converges to μf and the right-hand side converges to μg. Hence $\mu f \leq \mu g$. The general case is similar.

(b) Linearity for simple functions and monotonicity imply the following chain of inequalities:

$$0 \leq \frac{1}{n}\mu(\{x : f(x) \geq \frac{1}{n}\}) = \frac{1}{n}\mu(1_{f \geq \frac{1}{n}}) = \mu(\frac{1}{n}1_{f \geq \frac{1}{n}}) \leq \mu(f1_{f \geq \frac{1}{n}}) \leq \mu f = 0.$$

Since the two ends of this chain of inequalities are equal, it follows that all the inequalities are in fact equalities. Hence,

$$\mu(\{x : f(x) \geq 1/n\}) = 0 \quad \forall n$$

and so

$$\{x : f(x) > 0\} = \cup_n \{x : f(x) \geq 1/n\}.$$

Taking the measure of both sides, we get

$$0 \leq \mu(\{x : f(x) > 0\}) \leq \sum_n \mu(\{x : f(x) \geq 1/n\}) = 0.$$

Again, equating this anchored chain of inequalities, we see that $f = 0$ a.e.

(d) Fix $f \in \mathcal{E}_+$. Let $A_1, \ldots, A_m \in \mathcal{E}$ be disjoint with union A. If f is simple, 27.6 is immediate from Remark b applied to the simple functions $f1_{A_1}, \ldots, f1_{A_m}$ whose sum is $f1_A$. Applying this to simple functions $d_n \circ f$, we see that

$$\sum_1^m \mu(1_{A_i} d_n \circ f) = \mu(1_A d_n \circ f).$$

Note that $1_B(x) d_n \circ f(x) = d_n(1_B(x) f(x))$ for each x by the way the function d_n is defined. Putting this observation into the preceding expression and letting $n \to \infty$ we obtain

$$\begin{aligned}
\sum_1^m \mu(f1_{A_i}) &= \sum_1^m \lim_n \mu(d_n \circ (f1_{A_i})) \\
&= \lim_n \sum_1^m \mu(d_n \circ (f1_{A_i})) \\
&= \lim_n \mu(d_n \circ (f1_A)) \\
&= \lim_n \mu(f1_A),
\end{aligned}$$

where the interchange of the limit and the sum is justified by the finiteness of m. □

Monotone Convergence Theorem

This is the key result in the theory of integration. It allows interchanging the order of taking limits and integrals under reasonable conditions.

27.7 THEOREM. *Let $(f_n) \subset \mathcal{E}_+$ be increasing. Then,*

$$\mu(\lim f_n) = \lim \mu f_n.$$

PROOF. Let $f = \lim f_n$; it is well-defined since $f_1 \le f_2 \le \cdots$ and is positive and $\mathcal{E}$-measurable. So, μf is well-defined. By the monotonicity of integration, $\mu f_1 \le \mu f_2 \le \cdots \le \mu f$. Therefore $\lim \mu f_n$ exists and

$$\lim_n \mu f_n \le \mu f.$$

It remains to show that $\lim_n \mu f_n \ge \mu f$. This is accomplished in steps.

Step 1. If $b \in \mathbb{R}_+$, $B \in \mathcal{E}$, and $f(x) > b$ for $x \in B$, then $\lim_n \mu(f_n 1_B) \ge b\mu(B)$.

First, note that $\{f_1 > b\} \subset \{f_2 > b\} \subset \cdots$ and that

$$\cup_n \{f_n > b\} = \{x : f_n(x) > b \text{ for some } n\} = \{f > b\}.$$

Put $B_n = \{f_n > b\} \cap B$. Then, $B_n \nearrow$ and $\cup_n B_n = \{f > b\} \cap B = B$. Thus,

27.8 $$\lim_n \mu(B_n) = \mu(B)$$

by the sequential continuity of μ under increasing limits. Now, note that

$$f_n 1_B \ge f_n 1_{B_n} \ge b 1_{B_n},$$

and so the monotonicity of integration yields that

$$\mu(f_n 1_B) \ge \mu(b 1_{B_n}) = b\mu(B_n).$$

Taking limits on both sides and using 27.8, we get

27.9 $$\lim \mu(f_n 1_B) \ge b\mu(B).$$

Step 2. The same inequality holds even if $f(x) \ge b$ for $x \in B$.

For $b = 0$, this is trivial. For $b > 0$, apply Step 1 with $b-\epsilon$ to see that $\lim_n \mu(f_n 1_B) \ge (b-\epsilon)\mu(B)$. Since ϵ is arbitrary, we can let it go to zero to obtain the desired inequality.

Step 3. If g is a simple function and $g \le f$, then $\lim_n \mu f_n \ge \mu g$.

Let $\sum_1^m b_i 1_{B_i}$ denote the canonical representation for g. Then, our assumptions imply that $f(x) \ge g(x) = b_i$ for $x \in B_i$. Hence, we may apply the result of Step 2 to conclude that

$$\lim_n \mu(f_n 1_{B_i}) \ge b_i \mu(B_i) \quad i = 1, \ldots, m.$$

Hence, by Proposition 27.5d applied to the function f_n, we see that

$$\begin{aligned} \lim_n \mu f_n &= \lim_n \sum_1^m \mu(f_n 1_{B_i}) \\ &= \sum_1^m \lim \mu(f_n 1_{B_i}) \\ &\ge \sum_1^m b_i \mu(B_i) = \mu g. \end{aligned}$$

Step 4. $\lim_n \mu f_n \geq \mu f$.

Put $g = d_m \circ f$. Step 3 applied with this g yields $\lim_n \mu f_n \geq \mu(d_m \circ f)$. Letting $m \to \infty$ we get the desired result. $\square$

A particular consequence of the monotone convergence theorem is that, in definition 27.2, the special sequence $(d_n \circ f)$ can be replaced by any sequence $(f_n) \subset \mathcal{E}_+$ increasing to f.

Linearity of Integration

27.10 PROPOSITION. *If $f, g \in \mathcal{E}_+$ and $a, b \in \mathbb{R}_+$, then*

$$\mu(af + bg) = a\mu f + b\mu g.$$

The same holds for arbitrary $f, g \in \mathcal{E}$ and $a, b \in \mathbb{R}$ provided that both sides are well-defined. It holds, in particular, if f and g are integrable.

PROOF. If f, g are simple, the result is established by direct checking as was remarked in b. For $f, g \in \mathcal{E}_+$, and $a, b \in \mathbb{R}_+$, choose (f_n) and (g_n) to be sequences of simple positive functions increasing to f and g, respectively. Then,

$$\mu(af_n + bg_n) = a\mu f_n + b\mu g_n,$$

and $af_n + bg_n \nearrow af + bg$, $f_n \nearrow f$, $g_n \nearrow f$. Taking limits on both sides and using the monotone convergence theorem completes the proof. If $f, g \in \mathcal{E}$ are arbitrary, write $f = f^+ - f^-$ and $g = f^+ - g^-$ and go through the same steps. $\square$

Fatou's Lemma

This gives a useful inequlaity for arbitrary sequences of positive measurable functions.

27.11 LEMMA. *Let $(f_n) \subset \mathcal{E}_+$. Then,* $\mu(\liminf f_n) \leq \liminf \mu f_n$.

PROOF. Define $g_m = \inf_{n \geq m} f_n$. Then, $\liminf f_n$ is the limit of the increasing sequence $(g_m) \subset \mathcal{E}_+$, and thus

$$\mu(\liminf f_n) = \mu(\lim g_m) = \lim \mu g_m$$

by the monotone convergence theorem. On the other hand, $g_m \leq f_n$ for all $n \geq m$, which yields $\mu g_m \leq \mu f_n$ for all $n \geq m$, which in turn means that $\mu g_m \leq \inf_{n \geq m} \mu f_n$. Hence, as needed,

$$\lim \mu g_m \leq \lim_m \inf_{n \geq m} \mu f_n = \liminf \mu f_n.$$

□

27.12 COROLLARY.
(a) Let $(f_n) \subset \mathcal{E}$. *If* $f_n \geq g$ *for all* n *for some integrable function* g, *then*

$$\mu(\liminf f_n) \leq \liminf \mu f_n.$$

(b) Let $(f_n) \subset \mathcal{E}$. *If* $f_n \leq g$ *for all* n *for some integrable function* g, *then*

$$\mu(\limsup f_n) \geq \limsup \mu f_n.$$

PROOF. Let g be an integrable function. Suppose that g is real-valued so that □

Dominated Convergence Theorem

This is the second important tool for interchanging the order of taking limits and integrals.

A function f is said to be dominated by a function g if $|f| \leq g$; note that $g \geq 0$ necessarily. A sequence of functions (f_n) is said to be *dominated* by g if $|f_n| \leq g$ for each n. If g can be taken to be a finite constant, the (f_n) is said to be bounded.

27.13 THEOREM. *Suppose that* $(f_n) \subset \mathcal{E}$ *is dominated by an integrable function* g. *If* $\lim f_n$ *exists, then it is integrable and*

$$\mu(\lim_n f_n) = \lim_n \mu f_n.$$

PROOF. By assumption, $-g \leq f_n \leq g$ for every n, and g and $-g$ are both integrable. Thus, μf_n exists and is sandwiched between the finite numbers $-\mu g$ and μg. Now, both statements of the last corollary apply and we get

$$\mu(\liminf f_n) \leq \liminf \mu f_n \leq \limsup \mu f_n \leq \mu(\limsup f_n).$$

If $\lim f_n$ exists, then $\liminf f_n = \limsup f_n = \lim f_n$, and $\lim f_n$ is integrable since it is dominated by g. Hence, the extreme members of the preceding expression are finite and equal, which means that equality holds throughout. □

If $(f_n) \subset \mathcal{E}$ is bounded, say by the constant b, and if the measure μ is finite, then we can take $g = b$ in the preceding theorem. The resulting corollary is called the *bounded convergence theorem:*

27.14 THEOREM. *Let $(f_n) \subset \mathcal{E}$ be bounded. Suppose that μ is finite. If $\lim f_n$ exists, then*

$$\mu(\lim_n f_n) = \lim_n \mu f_n.$$

27.15 EXAMPLE. Let $(E, \mathcal{E}) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ and let f_n be the sequence of functions shown in Figure **??**. Note that the functions are not monotone and there is no integrable function that dominates them. Also, $\mu f_n = 1$ for all n and so $\lim \mu f_n = 1$, whereas, $\lim f_n = 0$ and so $\mu \lim f_n = 0$.