Companion to Real Analysis

John M. Erdman
Portland State University

Version November 20, 2012

E-mail address: erdman@pdx.edu

Contents

PREFACE

Paul Halmos famously remarked in his beautiful *Hilbert Space Problem Book* [22] that "The only way to learn mathematics is to do mathematics." Halmos is certainly not alone in this belief. The current set of notes is an activity-oriented companion to the study of real analysis. It is intended as a pedagogical companion for the beginner, an introduction to some of the main ideas in real analysis, a compendium of problems I think are useful in learning the subject, and an annotated reading/reference list.

The great majority of the results in beginning real analysis are straightforward and can be verified by the thoughtful student. Indeed, that is the main point of these notes—to convince the beginner that the subject is accessible. In the material that follows there are numerous indicators that suggest activity on the part of the reader: words such as "proposition", "example", "exercise", and "corollary", if not followed by a proof or a reference to a proof, are invitations to verify the assertions made. Of course, the proofs of some theorems, the *Hahn-Banach theorem* for example, are, in my opinion, too hard for the average student to (re)invent. In such cases when I have no improvements to offer to the standard proofs, instead of repeating them I give references.

These notes were written for a year long course in real analysis for seniors and first year graduate students at Portland State University. During the year students are asked to choose, in addition to these notes, other sources of information for the course, either printed texts or online documents, which suit their individual learning styles. As a result the material that is intended to be covered during the Fall quarter, the first 10–15 chapters, is relatively complete. After that, when students have found other sources they like, the notes become sketchier.

I have made some choices that may seem unusual. I have emphasized the role played by order relations in function spaces by trying to make explicit the vector lattice structure of many important examples. This material usually seems to have some "poor relation" status in most real analysis texts. To highlight fundamental concepts and avoid, on a first acquaintance, distracting technical details I have treated compact spaces more thoroughly than locally compact ones and real measures more thoroughly than extended real valued ones.

There are of course a number of advantages and disadvantages in consigning a document to electronic life. One advantage is the rapidity with which links implement cross-references. Hunting about in a book for *lemma 3.14.23* can be time-consuming (especially when an author engages in the entirely logical but utterly infuriating practice of numbering lemmas, propositions, theorems, corollaries, and so on, separately). A perhaps more substantial advantage is the ability to correct errors, add missing bits, clarify opaque arguments, and remedy infelicities of style in a timely fashion. The correlative disadvantage is that a reader returning to the web page after a short time may find everything (pages, definitions, theorems, sections) numbered differently. (LaTeXis an amazing tool.) I will change the date on the title page to inform the reader of the date of the last nontrivial update (that is, one that affects numbers or cross-references).

The most serious disadvantage of electronic life is impermanence. In most cases when a web page vanishes so, for all practical purposes, does the information it contains. For this reason (and the fact that I want this material to be freely available to anyone who wants it) I am making use of a "Share Alike" license from *Creative Commons*. It is my hope that anyone who finds this material useful will correct what is wrong, add what is missing, and improve what is clumsy. For more information on creative commons licenses see http://creativecommons.org/. Concerning the text itself, please send corrections, suggestions, complaints, and all other comments to the author at

erdman@pdx.edu

Greek Letters

Upper case	Lower case	English name (approximate pronunciation)
A	α	Alpha (AL-fuh)
B	β	Beta (BAY-tuh)
Γ	γ	Gamma (GAM-uh)
Δ	δ	Delta (DEL-tuh)
E	ϵ or ε	Epsilon (EPP-suh-lon)
Z	ζ	Zeta (ZAY-tuh)
H	η	Eta (AY-tuh)
Θ	θ	Theta (THAY-tuh)
I	ι	Iota (eye-OH-tuh)
K	κ	Kappa (KAP-uh)
Λ	λ	Lambda (LAM-duh)
M	μ	Mu (MYOO)
N	ν	Nu (NOO)
Ξ	ξ	Xi (KSEE)
O	o	Omicron (OHM-ih-kron)
Π	π	Pi (PIE)
P	ρ	Rho (ROH)
Σ	σ	Sigma (SIG-muh)
T	τ	Tau (TAU)
Υ	υ	Upsilon (OOP-suh-lon)
Φ	ϕ	Phi (FEE or FAHY)
X	χ	Chi (KHAY)
Ψ	ψ	Psi (PSEE or PSAHY)
Ω	ω	Omega (oh-MAY-guh)

Fraktur Fonts

In these notes Fraktur fonts are used (most often for families of sets and families of operators). Below are the Roman equivalents for each letter. When writing longhand or presenting material on a blackboard it is usually best to substitute script English letters.

Fraktur Upper case	Fraktur Lower case	Roman Lower Case
$\mathfrak{A}$	$\mathfrak{a}$	a
$\mathfrak{B}$	$\mathfrak{b}$	b
$\mathfrak{C}$	$\mathfrak{c}$	c
$\mathfrak{D}$	$\mathfrak{d}$	d
$\mathfrak{E}$	$\mathfrak{e}$	e
$\mathfrak{F}$	$\mathfrak{f}$	f
$\mathfrak{G}$	$\mathfrak{g}$	g
$\mathfrak{H}$	$\mathfrak{h}$	h
$\mathfrak{I}$	$\mathfrak{i}$	i
$\mathfrak{J}$	$\mathfrak{j}$	j
$\mathfrak{K}$	$\mathfrak{k}$	k
$\mathfrak{L}$	$\mathfrak{l}$	l
$\mathfrak{M}$	$\mathfrak{m}$	m
$\mathfrak{N}$	$\mathfrak{n}$	n
$\mathfrak{O}$	$\mathfrak{o}$	o
$\mathfrak{P}$	$\mathfrak{p}$	p
$\mathfrak{Q}$	$\mathfrak{q}$	q
$\mathfrak{R}$	$\mathfrak{r}$	r
$\mathfrak{S}$	$\mathfrak{s}$	s
$\mathfrak{T}$	$\mathfrak{t}$	t
$\mathfrak{U}$	$\mathfrak{u}$	u
$\mathfrak{V}$	$\mathfrak{v}$	v
$\mathfrak{W}$	$\mathfrak{w}$	w
$\mathfrak{X}$	$\mathfrak{x}$	x
$\mathfrak{Y}$	$\mathfrak{y}$	y
$\mathfrak{Z}$	$\mathfrak{z}$	z

CHAPTER 1

SETS

1.1. Set Notation

Everything in these notes is defined ultimately in terms of two primitive (some say, undefined) concepts: set and set membership. We assume that these are already familiar. In particular, we take it as understood that individual *elements* (or *members*, or *points*) can be regarded collectively as a single *set* (or *family*, or *collection*). It is occasionally useful to have all these different names. Many find a discussion concerning "a collection of families of sets" somewhat less confusing than trying to keep track of the identities of the members of the cast in "a set of sets of sets".

To indicate that x belongs to a set A we write $x \in A$; to indicate that it does not belong to A we write $x \notin A$. Synonymous with "x belongs to A" are "x is an element of A", "x is a member of A", and "x is a point in A".

We specify a set by listing its members between braces (for example, $\{1, 2, 3, 4, 5\}$ is the set of the first five natural numbers), by listing some of its members between braces with an ellipsis (three dots) indicating the missing members (for example, $\{1, 2, 3, \dots\}$ is the set of all natural numbers), or by writing $\{x \colon P(x)\}$ where $P(x)$ specifies what property the variable x must satisfy in order to be included in the set (for example, $\{x \colon 0 \leq x \leq 1\}$ is the closed unit interval $[0, 1]$).

1.1.1. CAUTION. Be aware that notational conventions for sets vary according to context. For example, if the statement of a theorem starts, "Let $\{s_1, s_2, \dots, s_n\}$ be a set," the author most often intends the set to contain n distinct members. But if in a proof we read, "Let $\{r_1, r_2, \dots, r_n\}$ be the roots of the n^{th} degree polynomial p", we do not conclude that the elements of the set are necessarily unique. (An n^{th} degree polynomial has n roots only when we count multiplicity.)

In a similar vein I adopt a convention in these notes: If I write, "Let a, b, and c be points in a topological space ...," I am not assuming that they are distinct, but if I write, "Let a, b, and c be three points in a topological space ...," I *am* assuming them to be distinct.

1.1.2. CAUTION. In attempting to prove a theorem which has as a hypothesis "Let S be a set" do not include in your proof something like "Suppose $S = \{s_1, s_2, \dots, s_n\}$" or "Suppose $S = \{s_1, s_2, \dots\}$". In the first case you are tacitly assuming that S is finite and in the second that it is countable. Neither is justified by the hypothesis.

1.2. Families of Sets

In these notes we devote much attention to such objects as σ-algebras and topologies, which are families of sets (that is, sets of sets). Ordinarily we will denote a family of sets by a single letter, as in "Let $\mathfrak{S}$ be a family of sets."

1.2.1. CAUTION. A single letter $\mathfrak{S}$ (an S in fraktur font) is an acceptable symbol in printed documents. Don't try to imitate it in hand-written work or on the blackboard. Use script letters instead.

Later in section 7.2 of chapter 7 we will consider a closely related, but quite different, concept of an *indexed family* of sets.

1.2.2. Example. Let $\mathfrak{S}$ be the family $\big\{\{x \in \mathbb{R} \colon |x - a| < b\} \colon a, b \in \mathbb{R}\big\}$ and $\mathfrak{T}$ be the family $\big\{\{x \in \mathbb{R} \colon a < x < b\} \colon a, b \in \mathbb{R}\big\}$. Then $\mathfrak{S} = \mathfrak{T}$. Characterize this family of sets in ordinary language (without the use of symbols).

1.3. Subsets

1.3.1. Definition. Let S and T be sets. We say that S is a SUBSET of T and write $S \subseteq T$ (or $T \supseteq S$) if every member of S belongs to T. If $S \subseteq T$ we also say that S is CONTAINED IN T or that T CONTAINS S. If $S \subseteq T$ but $S \neq T$, then we say that S is a PROPER SUBSET of T (or that S is PROPERLY CONTAINED IN T, or that T PROPERLY CONTAINS S) and write $S \subsetneqq T$.

1.3.2. Example. The EMPTY SET (or NULL SET), which is denoted by $\emptyset$, is defined to be the set which has no elements. (Or, if you like, define it to be $\{x \colon x \neq x\}$.) It is regarded as a subset of every set, so that $\emptyset \subseteq S$ is always true. (Note: "$\emptyset$" is a letter of the Danish alphabet, not the Greek letter "phi".)

1.3.3. Example. If S is a set, then the POWER SET of S, which we denote by $\mathfrak{P}(S)$, is the set of all subsets of S.

1.3.4. Remark. In these notes equality is used in the sense of identity. We write $x = y$ to indicate that x and y are two names for the same object. For example, we write $\{5, 7, a, b\} = \{b, 7, 5, a\}$ because sets are entirely determined by their members and the indicated sets have exactly the same members. Likewise we write $0.5 = 1/2 = 3/6 = 1/\sqrt{4}$ because 0.5, 1/2, 3/6, and $1/\sqrt{4}$ are different names for the same real number. You have probably encountered other uses of the term *equality*. In many high school geometry texts, for example, one finds statements to the effect that a triangle is isosceles if it has two equal sides (or two equal angles). What is meant of course is that a triangle is isosceles if it has two sides of *equal length* (or two angles of *equal angular measure*). We also make occasional use of the symbol := to indicate *equality by definition*. Thus when we write $a := b$ we are giving a new name a to an object b with which we are presumably already familiar.

1.4. Unions and Intersections

1.4.1. Definition. If $\mathfrak{A}$ is a family of subsets of a set S, then we define the UNION of the family $\mathfrak{A}$ to be the set of all x belonging to S such that $x \in A$ for at least one set A in $\mathfrak{A}$. We denote the union of the family $\mathfrak{A}$ by $\bigcup \mathfrak{A}$ (or by $\bigcup_{A \in \mathfrak{A}} A$, or by $\bigcup\{A \colon A \in \mathfrak{A}\}$). Thus $x \in \bigcup \mathfrak{A}$ if and only if there exists $A \in \mathfrak{A}$ such that $x \in A$.

If $\mathfrak{A}$ is a finite family of sets $A_1, \dots, A_n$, then we may write $\bigcup_{k=1}^{n} A_k$ or $A_1 \cup A_2 \cup \dots \cup A_n$ for $\bigcup \mathfrak{A}$. Similarly, if $\mathfrak{A}$ is a countably infinite family of sets $A_1, A_2, \dots$, then we may write $\bigcup_{k=1}^{\infty} A_k$ or $A_1 \cup A_2 \cup \dots$ for $\bigcup \mathfrak{A}$.

In a similar fashion, if $\mathfrak{A}$ is a family of subsets of a set S, then we define the INTERSECTION of the family $\mathfrak{A}$ to be the set of all x belonging to S such that $x \in A$ for every set A in $\mathfrak{A}$. We denote the intersection of the family $\mathfrak{A}$ by $\bigcap \mathfrak{A}$ (or by $\bigcap_{A \in \mathfrak{A}} A$, or by $\bigcap\{A \colon A \in \mathfrak{A}\}$). Thus $x \in \bigcap \mathfrak{A}$ if and only if $x \in A$ for every $A \in \mathfrak{A}$.

If $\mathfrak{A}$ is a finite family of sets $A_1, \dots, A_n$, then we may write $\bigcap_{k=1}^{n} A_k$ or $A_1 \cap A_2 \cap \dots \cap A_n$ for $\bigcap \mathfrak{A}$. Similarly, if $\mathfrak{A}$ is a countably infinite family of sets $A_1, A_2, \dots$, then we may write $\bigcap_{k=1}^{\infty} A_k$ or $A_1 \cap A_2 \cap \dots$ for $\bigcap \mathfrak{A}$.

The next two propositions are the familiar facts that union distributes over intersection and intersection distributes over union.

1.4.2. Proposition. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be nonempty families of sets and $A \in \mathfrak{A}$. Then*

(a) $A \cup \big(\bigcap \mathfrak{B}\big) = \bigcap\{A \cup B \colon B \in \mathfrak{B}\}$, *and*

(b) $\big(\bigcap \mathfrak{A}\big) \cup \big(\bigcap \mathfrak{B}\big) = \bigcap\{A \cup B \colon A \in \mathfrak{A} \text{ and } B \in \mathfrak{B}\}$.

1.4.3. Proposition. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be nonempty families of sets and $A \in \mathfrak{A}$. Then*

(a) $A \cap \big(\bigcup \mathfrak{B}\big) = \bigcup\{A \cap B \colon B \in \mathfrak{B}\}$, *and*

(b) $\big(\bigcup \mathfrak{A}\big) \cap \big(\bigcup \mathfrak{B}\big) = \bigcup\{A \cap B \colon A \in \mathfrak{A} \text{ and } B \in \mathfrak{B}\}$.

1.4.4. Definition. Sets S and T are said to be DISJOINT if $S \cap T = \emptyset$. More generally, a family $\mathcal{S}$ of sets is a DISJOINT FAMILY (or a PAIRWISE DISJOINT or a MUTUALLY DISJOINT FAMILY) if $S \cap T = \emptyset$ whenever S and T are distinct (that is, not equal) sets which belong to $\mathcal{S}$.

1.4.5. CAUTION. Let $\mathfrak{S}$ be a family of sets. Do not confuse the following two statements.

(a) $\mathfrak{S}$ is a (pairwise) disjoint family.
(b) $\bigcap \mathfrak{S} = \emptyset$.

Certainly, if $\mathfrak{S}$ contains more than a single set, then (a) implies (b). But if $\mathfrak{S}$ contains three or more sets the converse need not hold. For example, let $S = \{0,1\}$, $T = \{3,4\}$, $U = \{0,2\}$, and $\mathfrak{S} = \{S, T, U\}$. Then $\mathfrak{S}$ is not a disjoint family (because $S \cap U$ is nonempty), but $\bigcap \mathfrak{S} = \emptyset$.

1.4.6. Definition. Let A and B be sets. When A and B are disjoint we will often use the notation $A \uplus B$ instead of $A \cup B$ (to emphasize the disjointness of A and B). When $C = A \uplus B$ we say that C is the DISJOINT UNION of A and B. Similarly we frequently choose to write the union of a *pairwise disjoint* family $\mathfrak{A}$ of sets as $\biguplus \mathfrak{A}$. And the notation $\biguplus_{\lambda \in \Lambda} A_\lambda$ may be used to denote the union of a *pairwise disjoint* indexed family $\{A_\lambda \colon \lambda \in \Lambda\}$ of sets. When $C = \biguplus \mathfrak{A}$ or $C = \biguplus_{\lambda \in \Lambda} A_\lambda$ we say that C is the DISJOINT UNION of the appropriate family. (Indexed families of sets are discussed in more detail in section 7.2.)

1.4.7. Proposition. *Let $A_1, A_2, A_3, \dots$ be sets. Then there exist sets $B_1, B_2, B_3, \dots$ such that $\bigcup_{k=1}^{n} A_k = \biguplus_{k=1}^{n} B_k$ for each $n \in \mathbb{N}$ and, consequently, $\bigcup_{k=1}^{\infty} A_k = \biguplus_{k=1}^{\infty} B_k$.*

1.4.8. Definition. A family $\mathfrak{S}$ of subsets of a set S COVERS (or is a COVER FOR, or is a COVERING FOR) a set A if $A \subseteq \bigcup \mathfrak{S}$.

1.4.9. Exercise. Find a family of open intervals which covers the set $\mathbb{N}$ of natural numbers and has the property that the sum of the lengths of the intervals is 1. *Hint.* $\sum_{k=1}^{\infty} 2^{-k} = 1$.

1.4.10. CAUTION. It is a good idea to be alert to some differing attitudes towards "the" empty set. In particular, is the empty family of subsets of a nonempty set S the same thing as "the" empty set? Or does it depend on S? There appears to be general agreement that the union of the empty family of subsets of S is empty. (For an element x to belong to the union of the empty family there would have to be a member of that family to which x belongs, which cannot be.) But what about the intersection of the empty family? For x to belong to that intersection it must belong to each member of the empty family, which it vacuously does because *there are no* members of the empty family. So, some argue, *every* x belongs to the intersection of the empty family. But every x in what universe of discourse? Many take the view that since we are talking about subsets of a fixed set S, only those x in S qualify. Thus $\bigcap \emptyset = S$ when $\emptyset$ is the empty family of subsets of S. This is all quite reasonable, but it does have some odd consequences. One is that there are lots of different empty families of sets (since they have differing intersections), and this runs counter to our usual agreement that a set is entirely determined by its elements. (Perhaps we should be more modest and assert only that nonempty sets are determined by their elements.) Also it is rather strange to have the union of a family be a proper subset of its intersection.

1.5. Complements

We ordinarily regard the sets with which we work in a particular context as being subsets of some appropriate *universal* set. Think of "universal" in the sense of delimiting a *universe of discourse* rather than being *all-encompassing*. For each set A contained in this universal set S we define the COMPLEMENT of A , denoted by A^c, to be the set of all members of S which do not belong to A. That is, $x \in A^c$ if and only if $x \in S$ and $x \notin A$.

1.5.1. Example. Let A be the closed interval $(-\infty, 3]$. If nothing else is specified, we think of this interval as being a subset of the real line $\mathbb{R}$ (our universal set). Thus A^c is the set of all x in $\mathbb{R}$ such that x is not less than or equal to 3. That is, A^c is the interval $(3, \infty)$.

1.5.2. Example. Let A be the set of all points (x, y) in the plane such that $x \geq 0$ and $y \geq 0$. Then A^c is the set of all points (x, y) in the plane such that either $x < 0$ or $y < 0$. That is,

$$A^c = \{(x, y) \colon x < 0\} \cup \{(x, y) \colon y < 0\}.$$

The two following propositions are De Morgan's laws for sets. They say that the complement of the union of a family is the intersection of the complements (proposition 1.5.3), and the complement of the intersection of a family is the union of the complements (proposition 1.5.4). As you may expect, they are obtained by translating into the language of sets the facts of logic which go under the same name.

1.5.3. Proposition. *Let $\mathfrak{S}$ be a family of subsets of a set S. Then*

$$(\cup\mathfrak{S})^c = \cap\{A^c \colon A \in \mathfrak{S}\}.$$

1.5.4. Proposition. *Let $\mathfrak{S}$ be a family of subsets of a set S. Then*

$$(\cap\mathfrak{S})^c = \cup\{A^c \colon A \in \mathfrak{S}\}.$$

1.5.5. Definition. If A and B are subsets of a set S, we define the complement of B relative to A, (or the set difference of A and B), denoted by $A \setminus B$, to be the set of all $x \in S$ which belong to A but not to B. That is,

$$A \setminus B := A \cap B^c.$$

The operation $\setminus$ is called set subtraction.

1.5.6. Example. Let $A = [0, 5]$ and $B = [3, 10]$. Then $A \setminus B = [0, 3)$.

It is an elementary but frequently useful fact that the union of two sets can be written as a disjoint union (that is, the union of two disjoint sets).

1.5.7. Proposition. *Let A and B be subsets of a set S. Then $A \setminus B$ and B are disjoint sets whose union is $A \cup B$.*

1.5.8. Exercise. Let $A = \{x \in \mathbb{R} \colon x > 5\}$, $B = (-4, 8)$, $C = (-2, 12]$, $D = [10, 15]$, and $\mathfrak{S} = \{A^c, B, C \setminus D\}$.

(a) Find $\bigcup\mathfrak{S}$.
(b) Find $\bigcap\mathfrak{S}$.

1.5.9. Proposition. *If A, B, and C are subsets of a set S, then*

$$(A \cap B) \setminus C = (A \setminus C) \cap (B \setminus C) = A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C).$$

1.6. Symmetric Difference

1.6.1. Definition. Let A and B be subsets of a set S. Then the symmetric difference $A \triangle B$ of A and B is defined by

$$A \triangle B := (A \setminus B) \cup (B \setminus A).$$

1.6.2. Proposition. *Let A and B be subsets of a set S. Then $A \triangle B = (A \cup B) \setminus (A \cap B)$.*

1.6.3. Proposition. *If A, B, and C are subsets of a set S, then $A \triangle C \subseteq (A \triangle B) \cup (B \triangle C)$.*

1.6.4. Proposition. *Let A, B, and C be subsets of a set S. Then*

(a) *$A \triangle (B \triangle C) = (A \triangle B) \triangle C$;*
(b) *$A \triangle B \cap C = A \cap C \triangle B \cap C$;*
(c) *$A \triangle S = A^c$;*
(d) *$A \triangle \emptyset = A$; and*
(e) *$A \triangle A = \emptyset$.*

1.7. Notation for Sets of Numbers

Here is a list of fairly standard notations for some sets of numbers which occur frequently in these notes:

$\mathbb{C}$ is the set of complex numbers

$\mathbb{R}$ is the set of real numbers

$\mathbb{R}^n$ is the set of all n-tuples $(r_1, r_2, \dots, r_n)$ of real numbers

$\mathbb{R}^+ = \{x \in \mathbb{R} \colon x \geq 0\},$ the positive real numbers

$\mathbb{Q}$ is the set of rational numbers

$\mathbb{Q}^+ = \{x \in \mathbb{Q} \colon x \geq 0\},$ the positive rational numbers

$\mathbb{Z}$ is the set of integers

$\mathbb{Z}^+ = \{0, 1, 2, \dots\},$ the positive integers

$\mathbb{N} = \{1, 2, 3, \dots\}$, the set of natural numbers

$\mathbb{N}_n = \{1, 2, 3, \dots, n\}$ the first n natural numbers

$[a, b] = \{x \in \mathbb{R} \colon a \leq x \leq b\}$

$[a, b) = \{x \in \mathbb{R} \colon a \leq x < b\}$

$(a, b] = \{x \in \mathbb{R} \colon a < x \leq b\}$

$(a, b) = \{x \in \mathbb{R} \colon a < x < b\}$

$[a, \infty) = \{x \in \mathbb{R} \colon a \leq x\}$

$(a, \infty) = \{x \in \mathbb{R} \colon a < x\}$

$(-\infty, b] = \{x \in \mathbb{R} \colon x \leq b\}$

$(-\infty, b) = \{x \in \mathbb{R} \colon x < b\}$

$\mathbb{D} = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 < 1\},$ the open unit disc

$\mathbb{T} = \mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1\},$ the unit circle

CHAPTER 2

FUNCTIONS

2.1. Cartesian Products

Ordered pairs are familiar objects. They are used among other things for coordinates of points in the plane. In the first sentence of chapter 1 it was promised that all subsequent mathematical objects would be defined in terms of sets. So here just for the record is a formal definition of "ordered pair".

2.1.1. Definition. Let x and y be elements of arbitrary sets. Then the ORDERED PAIR (x, y) is defined to be $\{\{x, y\}, \{x\}\}$. This definition reflects our intuitive attitude: an ordered pair is a set $\{x, y\}$ with one of the elements, here x, designated as being "first". Thus we specify two things: $\{x, y\}$ and $\{x\}$.

Ordered pairs have only one interesting property: two of them are equal if and only if both their first coordinates and their second coordinates are equal. As you will discover in the next exercise, this fact follows easily from the definition.

2.1.2. Proposition. *Let x, y, u, and v be elements of arbitrary sets. Then $(x, y) = (u, v)$ if and only if $x = u$ and $y = v$.*

Hint for proof. Do not assume that the set$\{x, y\}$ has two elements. If $x = y$, then $\{x, y\}$ has only one element.

In section 8.3 we define the general notion of *product*. For the moment we concentrate on the most important of these—the *Cartesian product* of sets.

2.1.3. Definition. Let S and T be sets. The CARTESIAN PRODUCT of S and T, denoted by $S \times T$, is defined to be $\{(x, y) : x \in S \text{ and } y \in T\}$. The set $S \times S$ is often denoted by S^2.

In section 8.3 we define the general notion of *product*. For the moment we concentrate on the most important of these—the *Cartesian product* of sets.

2.1.4. Proposition. *Let S, T, U, and V be sets. Then*

(a) $(S \times T) \cap (U \times V) = (S \cap U) \times (T \cap V)$;
(b) $(S \times T) \cup (U \times V) \subseteq (S \cup U) \times (T \cup V)$; *and*
(c) *equality need not hold in (b).*

2.2. Relations

2.2.1. Definition. A RELATION from a set S to a set T is a subset of the Cartesian product $S \times T$. A relation from the set S to itself is often called a relation *on* S or a relation *among members of* S.

There is a notational oddity concerning relations. To indicate that an ordered pair (a, b) belongs to a relation $R \subseteq S \times T$, we almost always write something like $a\,R\,b$ rather than $(a, b) \in R$, which we would expect from the definition. For example, the relation "less than" is a relation on the real numbers. Technically then, since $<$ is a subset of $\mathbb{R} \times \mathbb{R}$, we could (correctly) write expressions such as $(3, 7) \in <$. Of course we don't. We write $3 < 7$ instead. And we say, "3 is less than 7", not "the pair $(3, 7)$ belongs to the relation *less than*". This is simply a matter of convention; it has no mathematical or logical content.

In definition 2.2.1 we defined a relation R on a set S as a subset of the Cartesian product $S \times S$. We now define some conditions which a relation may satisfy.

2.2.2. Definition. Let R be a relation on a set nonempty S. We will, as usual, write aRb for $(a, b) \in R$. The relation

(a) R is REFLEXIVE if aRa holds for every $a \in S$;
(b) R is TRANSITIVE if aRc holds whenever both aRb and bRc do;
(c) R is SYMMETRIC if aRb implies bRa; and
(d) R is ANTISYMMETRIC if aRb and bRa together imply $a = b$

Notice that the relation of equality on elements of a set is both symmetric and antisymmetric.

2.3. Functions

Functions are familiar from beginning calculus. Informally, a function consists of a pair of sets and a "rule" which associates with each member of the first set (the *domain*) one and only one member of the second (the *codomain*). While this informal "definition" is certainly adequate for most purposes and seldom leads to any misunderstanding, it is nevertheless sometimes useful to have a more precise formulation. This is accomplished by defining a function to be a special type of relation between two sets.

2.3.1. Definition. A FUNCTION f is an ordered triple (S, T, G) where S and T are sets and G is a subset of $S \times T$ satisfying:

(1) for each $s \in S$ there is a $t \in T$ such that $(s, t) \in G$, and
(2) if (s, t_1) and (s, t_2) belong to G, then $t_1 = t_2$.

In this situation we say that f is a *function from S into T* (or that *f maps S into T*) and write $f \colon S \to T$. The set S is the DOMAIN (or the INPUT SPACE) of f. The set T is the CODOMAIN (or TARGET SPACE, or the OUTPUT SPACE) of f. And the relation G is the GRAPH of f. In order to avoid explicit reference to the graph G it is usual to replace the expression "$(x, y) \in G$" by "$y = f(x)$"; the element $f(x)$ is the IMAGE of x under f. In these notes (but not everywhere!) the words "transformation", "map", and "mapping" are synonymous with "function". The domain of f is denoted by $\operatorname{dom} f$.

2.3.2. Notation. If S and T are sets, then the family of all functions with domain S and codomain T is denoted by $\mathcal{F}(S, T)$. Since real valued functions play such a large role in real analysis, we will shorten the notation for the family of real valued functions on a set S from $\mathcal{F}(S, \mathbb{R})$ to $\mathcal{F}(S)$.

2.3.3. Example. There are many ways of specifying a function. Statements (1)–(4) below define exactly the same function. We will use these (and other similar) notations interchangeably.

(1) For each real number x we let $f(x) = x^2$.
(2) Let $f = (S, T, G)$ where $S = T = \mathbb{R}$ and $G = \{(x, x^2) \colon x \in \mathbb{R}\}$.
(3) Let $f \colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$.
(4) Consider the function $f \colon \mathbb{R} \to \mathbb{R} \colon x \mapsto x^2$.

2.3.4. Convention. A REAL VALUED function is a function whose codomain lies in $\mathbb{R}$. A function OF A REAL VARIABLE is a function whose domain is contained in $\mathbb{R}$. Some real valued functions of a real variable may be specified simply by writing down a formula. When the domain and codomain are not specified, the understanding is that the domain of the function is the largest set of real numbers for which the formula makes sense and the codomain is taken to be $\mathbb{R}$.

2.3.5. Example. Let $f(x) = (x^2 + x)^{-1}$. Since this formula is meaningful for all real numbers except -1 and 0, we conclude that the domain of f is $\mathbb{R} \setminus \{-1, 0\}$.

2.4. Images and Inverse Images

2.4.1. Definition. If $f\colon S \to T$ and $A \subseteq S$, then $f^{\rightarrow}(A)$, the IMAGE of A under f, is $\{f(x)\colon x \in A\}$. It is common practice to write $f(A)$ for $f^{\rightarrow}(A)$. The set $f^{\rightarrow}(S)$ is the RANGE (or IMAGE) of f; usually we write $\operatorname{ran} f$ for $f^{\rightarrow}(S)$.

2.4.2. Definition. Let $f\colon S \to T$ and $B \subseteq T$. Then $f^{\leftarrow}(B)$, the INVERSE IMAGE of B under f, is $\{x \in S\colon f(x) \in B\}$. In many texts $f^{\leftarrow}(B)$ is denoted by $f^{-1}(B)$. This notation may cause confusion by suggesting that a function always has an inverse.

2.4.3. Exercise. Let $f(x) = \dfrac{x}{1-x}$.

(a) Find $f^{\leftarrow}([0,a])$ for $a > 0$.
(b) Find $f^{\leftarrow}([-\frac{3}{2}, -\frac{1}{2}])$.

2.4.4. Exercise. Let

$$f(x) = \begin{cases} x+1, & \text{for } x < 1 \\ 8 + 2x - x^2, & \text{for } x \geq 1. \end{cases}$$

Let $A = (-2, 3)$ and $B = [0, 1]$. Find $f^{\rightarrow}(A)$ and $f^{\leftarrow}(B)$.

2.5. Composition of Functions

2.5.1. Definition. Let $f\colon S \to T$ and $g\colon T \to U$. The COMPOSITE of g and f, denoted by $g \circ f$, is the function taking S to U defined by

$$(g \circ f)(x) = g(f(x))$$

for all x in S. The operation $\circ$ is COMPOSITION. We again make a special convention for real valued functions of a real variable: The domain of $g \circ f$ is the set of all x in $\mathbb{R}$ for which the expression $g(f(x))$ makes sense.

2.5.2. Proposition. *Composition of functions is associative but not necessarily commutative. That is, if $f\colon S \to T$, $g\colon T \to U$, and $h\colon U \to V$, then $h \circ (g \circ f) = (h \circ g) \circ f$. Also, it need not be true that $f \circ g$ and $g \circ f$ are equal.*

2.5.3. Exercise. For real valued functions of a real variable does composition distribute over addition? That is, is it always true that $f \circ (g + h) = (f \circ g) + (f \circ h)$ (assuming that both sides are defined)? How about $(f + g) \circ h = (f \circ h) + (g \circ h)$?

2.5.4. Proposition. *If $f\colon S \to T$ and $g\colon T \to U$, then*

(a) $(g \circ f)^{\rightarrow}(A) = g^{\rightarrow}(f^{\rightarrow}(A))$ *for every* $A \subseteq S$; *and*
(b) $(g \circ f)^{\leftarrow}(B) = f^{\leftarrow}(g^{\leftarrow}(B))$ *for every* $B \subseteq U$.

2.6. The Identity Function

2.6.1. Definition. If S is a set one member of the family $\mathcal{F}(S, S)$ of functions mapping S into itself is particularly noteworthy, the IDENTITY FUNCTION on S. It is defined by

$$I_S\colon S \to S\colon x \mapsto x.$$

Depending on context we may denote the identity function on a set S by I_S, or I, or id_S, or id.

It is easy to see that the identity function is characterized algebraically by the conditions:

$$\text{if } f\colon R \to S, \text{ then } I_S \circ f = f$$

and

$$\text{if } g\colon S \to T, \text{ then } g \circ I_S = g.$$

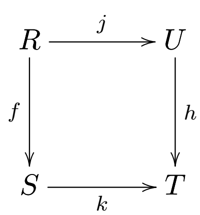
2.6.2. Definition. More general than the identity function are the inclusion maps. If $A \subseteq S$, then the INCLUSION MAP taking A into S is defined by

$$\iota_{A,S} \colon A \to S \colon x \mapsto x.$$

When no confusion is likely to result, we abbreviate $\iota_{A,S}$ to ι. Notice that $\iota_{S,S}$ is just the identity map I_S.

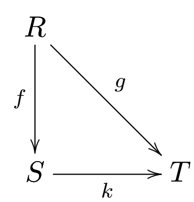
2.7. Diagrams

It is frequently useful to think of functions as arrows in diagrams. For example, the situation $j \colon R \to U$, $f \colon R \to S$, $k \colon S \to T$, $h \colon U \to T$ may be represented by the following diagram.



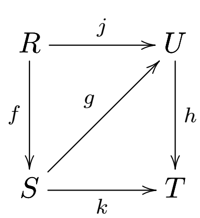
The diagram is said to COMMUTE (or to be a COMMUTATIVE DIAGRAM) if $h \circ j = k \circ f$.

Diagrams need not be rectangular. For instance,



is a commutative diagram if $g = k \circ f$.

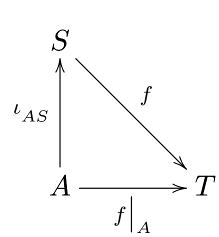
2.7.1. Example. Here is a diagrammatic way of stating the associative law for composition of functions. If the triangles in the diagram



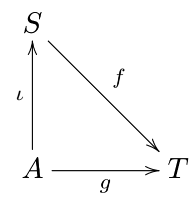
commute, then so does the rectangle.

2.8. Some Special Functions

2.8.1. Definition. If $f \colon S \to T$ and $A \subseteq S$, then the RESTRICTION of f to A, denoted by $f\big|_A$, is the function $f \circ \iota_{A,S}$. That is, it is the mapping from A into T whose value at each x in A is $f(x)$.



2.8.2. Definition. Suppose that $g\colon A \to T$ and $A \subseteq S$. A function $f\colon S \to T$ is an EXTENSION of g to S if $f\big|_A = g$, that is, if the diagram



commutes.

There are many ways of combining functions. Composition we have already encountered (section 2.5). Here are two more ways.

2.8.3. Notation. Suppose $f\colon R \to T$ and $g\colon S \to U$ are mappings between sets. Then $f \times g$ is the function mapping $R \times S$ into $T \times U$ defined by

$$(f \times g)(r, s) = \big(f(r), g(s)\big)$$

for all $r \in R$ and $s \in S$.

2.8.4. Notation. Suppose $f\colon T \to U$ and $g\colon T \to V$ are mappings between sets. Then (f, g) is the function mapping T into $U \times V$ defined by

$$(f, g)(t) = \big(f(t), g(t)\big)$$

for all $t \in T$.

2.8.5. Example. Let $T = [0, 2\pi]$, $f(t) = \cos t$, and $g(t) = \sin t$ for all $t \in T$. Then (f, g) is the usual (counterclockwise) parametrization of the unit circle in $\mathbb{R}^2$.

2.8.6. Definition. A function $f\colon S \to T$ between sets is a CONSTANT FUNCTION if $\operatorname{ran} f$ contains exactly one element of T. It is conventional to use the *value* of the constant function as its *name*. Thus, for example, the constant function on a set S whose value at each $s \in S$ is 2 is denoted by 2. (Thus of course the notation fails to tell us the domain of the function; we must deduce that from context.)

2.8.7. Definition. If A is a subset of a nonempty set S, we define $\chi_A\colon S \to \mathbb{R}$, the CHARACTERISTIC FUNCTION of A by

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

2.9. Injections, Surjections, and Bijections

2.9.1. Definition. A function f is INJECTIVE (or ONE-TO-ONE) if $x = y$ whenever $f(x) = f(y)$. That is, f is injective if no two distinct elements in its domain have the same image. For a real valued function of a real variable this condition may be interpreted graphically: A function is one-to-one if and only if each horizontal line intersects the graph of the function at most once. An injective map is called an INJECTION.

2.9.2. Exercise. Find an injective mapping from $\{x \in \mathbb{Q} \colon x > 0\}$ into $\mathbb{N}$.

2.9.3. Example. The function $x \mapsto \dfrac{2x-5}{3x+4}$ is injective.

2.9.4. Example. The function $x \mapsto 2x^2 - x - 15$ is *not* injective.

2.9.5. Definition. A function is SURJECTIVE (or ONTO) if its range is equal to its codomain.

2.9.6. Exercise. Find a surjection (that is, a surjective map) from $[0, 1]$ onto $[0, \infty)$.

2.9.7. Definition. A function is BIJECTIVE if it is both injective and surjective. A bijective map is called a BIJECTION or a ONE-TO-ONE CORRESPONDENCE.

2.9.8. Exercise. Give an explicit formula for a bijection between $\mathbb{Z}$ and $\mathbb{N}$.

2.9.9. Exercise. Give an explicit formula for a bijection between $\mathbb{R}$ and the open interval $(0,1)$.

2.9.10. Exercise. Give a formula for a bijection between the interval $[0,1)$ and the unit circle $x^2+y^2=1$ in the plane.

2.9.11. Proposition. *Let $f\colon S\to T$ and $g\colon T\to U$.*

(a) *If f and g are injective, so is $g\circ f$.*
(b) *If f and g are surjective, so is $g\circ f$.*
(c) *If f and g are bijective, so is $g\circ f$.*

It is important for us to know how the image $f^{\rightarrow}$ and the inverse image $f^{\leftarrow}$ of a function f behave with respect to unions, intersections, and complements of sets. The basic facts are given in the next six propositions. Although these results are quite elementary, they are quite useful.

2.9.12. Proposition. *Let $f\colon S\to T$ and $B\subseteq T$. Then*

(a) *$f^{\rightarrow}(f^{\leftarrow}(B))\subseteq B$;*
(b) *equality need not hold in (a); but*
(c) *equality does hold in (a) if f is surjective.*

2.9.13. Proposition. *Let $f\colon S\to T$ and $A\subseteq S$. Then*

(a) *$A\subseteq f^{\leftarrow}(f^{\rightarrow}(A))$;*
(b) *equality need not hold in (a); but*
(c) *equality does hold in (a) if f is injective.*

2.9.14. Exercise. Prove the converse of part (c) of proposition 2.9.13. That is, show that if $f\colon S\to T$ and $f^{\leftarrow}(f^{\rightarrow}(A))=A$ for all $A\subseteq S$, then f is injective. *Hint.* Suppose $f(x)=f(y)$. Let $A=\{x\}$. Show that $y\in f^{\leftarrow}(f^{\rightarrow}(A))$.

2.9.15. Proposition. *Let $f\colon S\to T$ and $D\subseteq T$. Then*

$$f^{\leftarrow}(D^c)=\big(f^{\leftarrow}(D)\big)^c.$$

2.9.16. Proposition. *Let $f\colon S\to T$ and $A\subseteq S$.*

(a) *If f is injective, then $f^{\rightarrow}(A^c)\subseteq\big(f^{\rightarrow}(A)\big)^c$.*
(b) *If f is surjective, then $f^{\rightarrow}(A^c)\supseteq\big(f^{\rightarrow}(A)\big)^c$.*
(c) *If f is bijective, then $f^{\rightarrow}(A^c)=\big(f^{\rightarrow}(A)\big)^c$.*

2.9.17. Proposition. *Let $f\colon S\to T$ and $\mathfrak{A}\subseteq\mathfrak{P}(S)$. Then*

(a) *$f^{\rightarrow}(\bigcap\mathfrak{A})\subseteq\bigcap\{f^{\rightarrow}(A)\colon A\in\mathfrak{A}\}$;*
(b) *equality need not hold in (a); but*
(c) *if f is injective, then equality holds in (a); and furthermore*
(d) *$f^{\rightarrow}(\bigcup\mathfrak{A})=\bigcup\{f^{\rightarrow}(A)\colon A\in\mathfrak{A}\}$.*

2.9.18. Proposition. *Let $f\colon S\to T$ and $\mathfrak{B}\subseteq\mathfrak{P}(T)$. Then*

(a) *$f^{\leftarrow}(\bigcap\mathfrak{B})=\bigcap\{f^{\leftarrow}(B):B\in\mathfrak{B}\}$; and*
(b) *$f^{\leftarrow}(\bigcup\mathfrak{B})=\bigcup\{f^{\leftarrow}(B):B\in\mathfrak{B}\}$.*

2.10. Inverse Functions

Let $f\colon S\to T$ and $g\colon T\to S$. If $g\circ f=I_S$, then g is a LEFT INVERSE of f and, equivalently, f is a RIGHT INVERSE of g. We say that f is INVERTIBLE if there exists a function from T into S which is both a left and a right inverse for f. Such a function is denoted by f^{-1} and is called the INVERSE of f. (Notice that the last "the" in the preceding sentence requires justification. See proposition 2.10.1 below.) A function is INVERTIBLE if it has an inverse. According to the definition just given, the inverse f^{-1} of a function f must satisfy

$$f \circ f^{-1} = I_T \qquad \text{and} \qquad f^{-1} \circ f = I_S .$$

A simple, but important, consequence of this is that for an invertible function, $y = f(x)$ if and only if $x = f^{-1}(y)$. (Proof: if $y = f(x)$, then $f^{-1}(y) = f^{-1}(f(x)) = I_S(x) = x$. Conversely, if $x = f^{-1}(y)$, then $f(x) = f(f^{-1}(y)) = I_T(y) = y$.)

2.10.1. Proposition. *A function can have at most one inverse. In fact, if a function has both a left inverse and a right inverse, then these inverses are equal (and therefore the function is invertible).*

2.10.2. Exercise. The *arcsine* function is defined to be the inverse of what function? (*Hint.* The answer is not *sine.*) What about *arccosine*? *arctangent*?

The next two propositions tell us that a necessary and sufficient condition for a function to have right inverse is that it be surjective and that a necessary and sufficient condition for a function to have a left inverse is that it be injective. Thus, in particular, a function is invertible if and only if it is bijective. In other words, the invertible members of $\mathcal{F}(S, T)$ are the bijections.

2.10.3. Proposition. *Let $S \neq \emptyset$. A function $f \colon S \to T$ has a right inverse if and only if it is surjective.*

PROOF. Suppose that f has a right inverse f_r. For each $y \in T$ it is clear that $y = I_T(y) = f\big(f_r(y)\big) \in \operatorname{ran} f$; so $\operatorname{ran} f = T$ and f is surjective.

Conversely, suppose that f is surjective. Then for every $y \in T$ the set $f^{\leftarrow}(\{y\})$ is nonempty. For each $y \in T$ let x_y be a member of $f^{\leftarrow}(\{y\})$ and define

$$f_r \colon T \to S \colon y \mapsto x_y\,.$$

Then $f\big(f_r(y)\big) = f(x_y) = y$, showing that f_r is a right inverse of f. □ □

2.10.4. Remark. In the preceding proof we have made use of an important set theoretic axiom called *the axiom of choice.* This axiom says, roughly, that from an arbitrary family of nonempty sets it is possible simultaneously to extract one member from each set. Here is a more formal statement of the axiom:

2.10.5. Axiom (Axiom of Choice I)**.** *If $\mathfrak{A}$ is a family of nonempty sets, then there exists a function $f \colon \mathfrak{A} \to \bigcup \mathfrak{A}$ such that $f(A) \in A$ for every $A \in \mathfrak{A}$.*

Such a function is called a CHOICE FUNCTION. Notice that in the preceding proof f_r is a choice function; it "chooses" for each y one member x_y of the set $f^{\leftarrow}(\{y\})$. Not every mathematician approves of the use of the *axiom of choice*; some complain that it should not be invoked because it is not "constructive". However, most mathematicians accept it, and in these notes it will be used freely.

Another assertion that is equivalent to the *axiom of choice* says that a nonempty product of nonempty sets is nonempty.

2.10.6. Axiom (Axiom of choice II)**.** *If $(A_\lambda)_{\lambda \in \Lambda}$ is a nonempty indexed family of pairwise disjoint nonempty sets, then the Cartesian product $\prod_{\lambda\in\Lambda} A_\lambda$ is nonempty. (For a definition of the Cartesian product see 8.3.6.)*

Here is yet another version of the *axiom of choice.*

2.10.7. Axiom (Zermelo's postulate)**.** *If $\mathfrak{A}$ is a disjoint family of nonempty sets, then there exists a set C such that $C \cap A$ has exactly one member for each $A \in \mathfrak{A}$.*

In chapter 5 we will introduce another axiom, *Zorn's lemma* 5.3.1, which is equivalent to the *axiom of choice.* For proofs that all these versions of the *axiom of choice* are indeed equivalent, consult Hewitt and Stromberg [23], section 3; Brown and Page [7], section 1.5; Kelley [27], pages 31–36; or any standard text on set theory.

2.10.8. Proposition. *Let $S \neq \emptyset$. A function $f\colon S \to T$ has a left inverse if and only if it is injective.*

2.10.9. Proposition. *If a function f is bijective, then $\left(f^{-1}\right)^{\leftarrow} = f^{\rightarrow}$.*

2.10.10. Exercise. Let $f(x) = \dfrac{ax+b}{cx+d}$ where a, b, c, $d \in \mathbb{R}$ and not both c and d are zero.

(a) Under what conditions on the constants a, b, c, and d is f injective?
(b) Under what conditions on the constants a, b, c, and d is f its own inverse?

2.11. Equivalence Relations and Quotients

2.11.1. Definition. A relation R on a set S is an EQUIVALENCE RELATION if it is reflexive, symmetric, and transitive. If s is an element of S, then the EQUIVALENCE CLASS containing s is the set of all the elements in S which are equivalent to s. We usually denote the equivalence class containing s by $[s]$. Any member t of the equivalence class $[s]$ is called a REPRESENTATIVE of that equivalence class. The family of all equivalence classes of members of S is called the QUOTIENT of S by R (or S modulo R) and is denoted by S/R. The map $\pi\colon S \to S/R\colon s \mapsto [s]$ is the QUOTIENT MAP.

2.11.2. Exercise. In each item of the following list S is a set and R is a relation on S. Determine if each relation is reflexive, if it is symmetric, if it is transitive, and if it is an equivalence relation.

(a) $S = \mathbb{R}$; xRy if and only if $x \neq y$.
(b) $S = \mathbb{N} \times \mathbb{N}$; $(m,n)R(p,q)$ if and only if $mq = np$.
(c) S is the set of all humans; xRy if and only if x is a sister of y.
(d) $S = \mathbb{Z}$; mRn if and only if $m^2 - n^2$ is divisible by 5.
(e) $S = \mathbb{R}$; xRy if and only if $x^2 + y^2 = 1$.
(f) $S = \mathbb{R}^2$; $(u,v)R(x,y)$ if and only if $u^2 + v^2 = x^2 + y^2$.

2.11.3. Example. Define a relation $\sim$ on the set $S = \{1, 2, \dots, 12\}$ by

$$x \sim y \text{ if and only if } x^2 - y^2 \text{ is divisible by } 6.$$

The relation $\sim$ is an equivalence relation on S. How many equivalence classes are there? What elements are in each one?

2.11.4. Example. Define a relation $\sim$ on the set $D = \{(x,y) \in \mathbb{R}^2 \colon (x,y) \neq (0,0)\}$ by

$$(a,b) \sim (c,d) \qquad \text{if and only if} \qquad c^2 b = a^2 d.$$

Show that $\sim$ is an equivalence relation. Identify (that is, describe geometrically) the set $D/\sim$ of equivalence classes of D with respect to $\sim$.

2.11.5. Definition. A family $\mathfrak{A}$ of nonempty subsets of a set S is a PARTITION of the set S if the family $\mathfrak{A}$ is pairwise disjoint and its union is all of S.

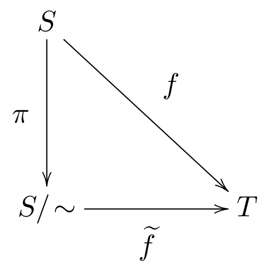
2.11.6. Example. Let $\{A_i \colon i \in I\}$ be a partition of a set A and $\{B_j \colon j \in J\}$ be a partition of a set B. Then $\{A_i \times B_j \colon (i,j) \in I \times J\}$ is a partition of $A \times B$.

2.11.7. Proposition. *If $\sim$ is an equivalence relation on a nonempty set S then the family $S/{\sim}$ of equivalence classes generated by $\sim$ is a partition of S and the quotient map $\pi\colon S \to S/{\sim}\colon s \mapsto [s]$ is surjective.*

The word "partition" is used as a verb as well as a noun: in light of the previous proposition, one may say that *the equivalence relation $\sim$ partitions the set S.*

2.11.8. Exercise. Prove the converse of proposition 2.11.7: Every partition $\mathfrak{F}$ of S is induced by some equivalence relation. (What *exactly* does this last assertion mean?)

2.11.9. Theorem (Fundamental quotient theorem for sets)**.** *Let $\sim$ be an equivalence relation on a set S and $\pi\colon S \to S/\!\sim$ be the quotient map. If $f\colon S \to T$ is a function from S into another set T which is constant on the equivalence classes of S (that is, if $x \sim y$ then $f(x) = f(y)$), then there exists a unique function $\widetilde{f}\colon S/\!\sim\, \to T$ which makes the following diagram commute.*



Furthermore, $\widetilde{f}$ is injective if and only if f takes different values on different equivalence classes (that is, if $x \not\sim y$, then $f(x) \neq f(y)$); and $\widetilde{f}$ is surjective if and only if f is.

When $\sim$ is an equivalence relation on a set S, we often have occasion to work with the equivalence classes generated by $\sim$. This is made slightly tricky by the fact that if elements x and y of S are equivalent, then the equivalence class containing them may denoted equally well by $[x]$ or $[y]$. That is, if $x \sim y$, then $[x] = [y]$. When we define some operation, say, on $[x]$ we must be sure that the definition we give does not depend on the representative we choose from the equivalence class, that is, that it does not depend on the *name* we choose for the equivalence class. In other words we must show that the operation is *well-defined*. The next exercise shows how things may go horribly wrong if we are not careful about such matters.

2.11.10. Exercise. The standard construction of the real numbers proceeds by first producing the positive integers and then using these to construct the set $\mathbb{Z}$ of all integers and to define addition and multiplication on them. The next step is to get the rational numbers. To do this let R be the set of all ordered pairs (m, n) of integers such that $n \neq 0$. Define a relation $\sim$ on R as follows:

$$(m, n) \sim (p, q) \quad \text{if} \quad mq = pn.$$

(a) Show that $\sim$ is an equivalence relation on R.

Now according to proposition 2.11.7 the relation $\sim$ partitions the set R into equivalence classes. We denote the equivalence class containing (m, n) by $\dfrac{m}{n}$. The set of these equivalence classes is the set $\mathbb{Q}$ of *rational numbers*.

(b) Now suppose we wish to define an operation $\oplus$ on $\mathbb{Q}$ as follows:

$$\frac{m}{n} \oplus \frac{p}{q} := \frac{m+p}{n+q}$$

Explain clearly why this definition makes no sense whatever; that is, explain why $\oplus$ is *not* well-defined.

(c) Prove that the corresponding (usual) definition for multiplication *is* well-defined.

$$\frac{m}{n} \cdot \frac{p}{q} := \frac{mp}{nq}$$

CHAPTER 3

CARDINALITY

3.1. Finite and Infinite Sets

There are a number of ways of comparing the "sizes" of sets. In this chapter we examine perhaps the simplest of these, cardinality. Roughly speaking, we say that two sets have the "same number of elements" if there is a one-to-one correspondence between the elements of the sets. In this sense the open intervals $(0, 1)$ and $(0, 2)$ have the same number of elements. (The map $x \mapsto 2x$ is a bijection.) Clearly this is only one sense of the idea of "size". It is certainly also reasonable to regard $(0, 2)$ as being bigger than $(0, 1)$ because it is twice as long.

3.1.1. Definition. Two sets S and T are CARDINALLY EQUIVALENT if there exists a bijection from S onto T, in which case we write $S \sim T$.

3.1.2. Proposition. *On any family of sets cardinal equivalence is an equivalence relation. That is, if S, T, and U are sets, then*

(a) $S \sim S$*;*
(b) *if* $S \sim T$*, then* $T \sim S$*; and*
(c) *if* $S \sim T$ *and* $T \sim U$*, then* $S \sim U$*.*

3.1.3. Definition. A set S is FINITE if it is empty or if there exists $n \in \mathbb{N}$ such that S is cardinally equivalent to some (unique) initial segment $\mathbb{N}_n = \{1, \dots, n\}$ of the natural numbers. A set is INFINITE if it is not finite.

If S is a finite set, then it is in one-to-one correspondence with some $\mathbb{N}_n$. We say that n is the CARDINAL NUMBER of the set S and write $n = \operatorname{card} S$. We take the cardinal number of the empty set, $\operatorname{card} \emptyset$, to be 0. Thus for a finite set S the assertion that "the cardinality of S is n" is just another way of saying that "the set S has n elements".

3.1.4. Proposition. *Let* $S \subseteq T$*. If* T *is finite, then* S *is finite and* $\operatorname{card} S \leq \operatorname{card} T$*.*

Hint for proof. The case $T = \emptyset$ is trivial. Suppose $T \neq \emptyset$. Let $\iota \colon S \to T$ be the inclusion map of S into T. There exist $n \in \mathbb{N}$ and a bijection $f \colon T \to \{1, \dots, n\}$.

The preceding proposition "subsets of finite sets are finite" has a useful contrapositive: *Sets which contain infinite sets are themselves infinite.*

3.1.5. Proposition. *If* S *and* T *are disjoint finite sets, then* $S \uplus T$ *is finite and*

$$\operatorname{card}(S \uplus T) = \operatorname{card} S + \operatorname{card} T\,.$$

3.1.6. Corollary. *If* S *is a set and* T *is a finite set, then*

$$\operatorname{card}(T \setminus S) = \operatorname{card} T - \operatorname{card}(T \cap S).$$

In the preceding proposition notice that if $S \subseteq T$ (where T is finite), then

$$\operatorname{card}(T \setminus S) = \operatorname{card} T - \operatorname{card} S\,.$$

3.1.7. Corollary. *If* S *and* T *are finite sets, then* $S \cup T$ *is finite and*

$$\operatorname{card}(S \cup T) = \operatorname{card} S + \operatorname{card} T - \operatorname{card}(S \cap T)\,.$$

How do we show that a set S is infinite? If our only tool were the definition, we would face the prospect of proving that there does *not* exist a bijection from S onto an initial segment of the natural numbers. It would be pleasant to have a more direct approach than establishing the nonexistence of maps. This is the point of the next proposition.

3.1.8. Proposition. *A set is infinite if and only if it is cardinally equivalent to a proper subset of itself.*

Proof. Suppose that S is infinite. We prove that there exists a proper subset T of S and a bijection f from S onto T. We choose a sequence of distinct elements a_k in S, one for each $k \in \mathbb{N}$. Let a_1 be an arbitrary member of S. Then $S \setminus \{a_1\} \neq \emptyset$. (Otherwise $S \sim \{a_1\}$ and S is finite.) Choose $a_2 \in S \setminus \{a_1\}$. Then $S \setminus \{a_1, a_2\} \neq \emptyset$. (Otherwise $S \sim \{a_1, a_2\}$ and S is finite.) In general, if distinct elements $a_1, \dots, a_n$ have been chosen, then $S \setminus \{a_1, \dots, a_n\}$ cannot be empty; so we may choose $a_{n+1} \in S \setminus \{a_1, \dots, a_n\}$. Let $T = S \setminus \{a_1\}$, and define $f\colon S \to T$ by

$$f(x) = \begin{cases} a_{k+1}, & \text{if } x = a_k \text{ for some } k \\ x, & \text{otherwise.} \end{cases}$$

Then f is a bijection from S onto the proper subset T of S.

For the converse construct a proof by contradiction. Suppose that $S \sim T$ for some proper subset $T \subseteq S$, and assume further that S is finite, so that $S \sim \{1, \dots, n\}$ for some $n \in \mathbb{N}$. Then by proposition 3.1.4 the set $S \setminus T$ is finite and, since it is nonempty, is therefore cardinally equivalent to $\{1, \dots, p\}$ for some $p \in \mathbb{N}$. Thus

$$\begin{aligned} n &= \operatorname{card} S \\ &= \operatorname{card} T \\ &= \operatorname{card}(S \setminus (S \setminus T)) \\ &= \operatorname{card} S - \operatorname{card}(S \setminus T) \qquad \text{(by corollary 3.1.6)} \\ &= n - p. \end{aligned}$$

Therefore $p = 0$, which contradicts the earlier assertion that $p \in \mathbb{N}$. ☐

3.1.9. Example. The set $\mathbb{N}$ of natural numbers is infinite.

3.1.10. Example. The interval $(0, 1)$ is infinite.

3.1.11. Example. The set $\mathbb{R}$ of real numbers is infinite.

The next two results tell us that functions take finite sets to finite sets and that for injective functions finite sets come from finite sets.

3.1.12. Proposition. *If T is a set, S is a finite set, and $f\colon S \to T$ is surjective, then T is finite.*

3.1.13. Proposition. *If S is a set, T is a finite set, and $f\colon S \to T$ is injective, then S is finite.*

3.1.14. Exercise. If S is a finite set with cardinal number n, what is the cardinal number of $\mathfrak{P}(S)$?

3.2. Countable and Uncountable Sets

There are many sizes of infinite sets—infinitely many in fact. In our subsequent work we need only distinguish between countably infinite and uncountable sets. A set is countably infinite if it is in one-to-one correspondence with the set of positive integers; if it is neither finite nor countably infinite, it is uncountable. In this section we present some basic facts about and examples of both countable and uncountable sets. This is all we will need. Except for exercise 3.2.25, which is presented for general interest, we ignore the many intriguing questions which arise concerning various sizes of uncountable sets. For a very readable introduction to such matters see [26], chapter 2.

3.2.1. Definition. A set is COUNTABLY INFINITE (or DENUMERABLE) if it is cardinally equivalent to the set $\mathbb{N}$ of natural numbers. A bijection from $\mathbb{N}$ onto a countably infinite set S is an ENUMERATION of the elements of S. A set is COUNTABLE if it is either finite or countably infinite. If a set is not countable it is UNCOUNTABLE.

3.2.2. Example. The set $\mathbb{E}$ of even integers in $\mathbb{N}$ is countable.

The first proposition of this section establishes the fact that the "smallest" infinite sets are the countable ones.

3.2.3. Proposition. *Every infinite set contains a countably infinite subset.*

If we are given a set S which we believe to be countable, it may be extremely difficult to prove this by exhibiting an explicit bijection between $\mathbb{N}$ and S. Thus it is of great value to know that certain constructions performed with countable sets result in countable sets. The next five propositions provide us with ways of generating new countable sets from old ones. In particular, we show that each of the following is countable.

(1) Any subset of a countable set.
(2) The range of a surjection with countable domain.
(3) The domain of an injection with countable codomain.
(4) The product of any finite collection of countable sets.
(5) The union of a countable family of countable sets.

3.2.4. Proposition. *If $S \subseteq T$ where T is countable, then S is countable.*

The preceding has an obvious corollary: *If $S \subseteq T$ and S is uncountable, then so is T.*

3.2.5. Proposition. *If $f : S \to T$ is injective and T is countable, then S is countable.*

3.2.6. Proposition. *If $f : S \to T$ is surjective and S is countable, then T is countable.*

3.2.7. Example. The set $\mathbb{N} \times \mathbb{N}$ is countable.

3.2.8. Example. The set $\{x \in \mathbb{Q} : x > 0\}$ is countable.

3.2.9. Proposition. *If S and T are countable sets, then so is $S \times T$.*

3.2.10. Corollary. *If $S_1, \dots, S_n$ are countable sets, then $S_1 \times \cdots \times S_n$ is countable.*

Finally we show that a countable union of countable sets is countable.

3.2.11. Proposition. *Suppose that $\mathcal{A}$ is a countable family of sets and that each member of $\mathcal{A}$ is itself countable. Then $\bigcup \mathcal{A}$ is countable.*

3.2.12. Example. The set $\mathbb{Q}$ of rational numbers is countable.

By virtue of 3.2.4–3.2.11 we have a plentiful supply of countable sets. We now look at an important example of a set which is not countable.

3.2.13. Example. The set $\mathbb{R}$ of real numbers is uncountable.

PROOF. We take it to be known that if we exclude decimal expansions which end in an infinite string of 9's, then every real number has a unique decimal expansion. By (the corollary to) proposition 3.2.4 it will suffice to show that the open unit interval $(0, 1)$ is uncountable. Argue by contradiction: assume that $(0, 1)$ is countably infinite. (We know, of course, from example 3.1.10 that it is not finite.) Let $r_1, r_2, r_3, \dots$ be an enumeration of $(0, 1)$. For each $j \in \mathbb{N}$ the number r_j has a unique decimal expansion

$$0.r_{j1}\, r_{j2}\, r_{j3}\, \dots .$$

Construct another number $x = 0.x_1\, x_2\, x_3\, \dots$ as follows. For each k choose $x_k = 1$ if $r_{kk} \neq 1$ and $x_k = 2$ if $r_{kk} = 1$. Then x is a real number between 0 and 1, and it cannot be any of the numbers r_k in our enumeration (since it differs from r_k at the k^{th} decimal place). But this contradicts the assertion that $r_1, r_2, r_3, \dots$ is an enumeration of $(0, 1)$. □

3.2.14. Example. The set of irrational numbers is uncountable.

3.2.15. Proposition. *If S is a countable set and T is uncountable, then $T \setminus S \sim T$.*

3.2.16. Exercise. A FINITE WORD is an n-tuple of upper case letters of the English language written without parentheses or commas. (Examples: DOG, BFTLZXP, ARRRRGHHHH.) An INFINITE WORD is an (infinite) sequence of upper case English letters.

(a) Prove or disprove: the set of finite words is countable.
(b) Prove or disprove: the set of infinite words is countable.

3.2.17. Notation. Let A and B be sets. Denote by A^B the family of all functions from B into A.

3.2.18. Proposition. *If S is a set, then $\{0,1\}^S$ is cardinally equivalent to $\mathfrak{P}(S)$, the power set of S.*

3.2.19. Exercise. Your good friend Fred R. Dim argues that there are only countably many real numbers. After all, he says, there are only ten choices for each position in a number's decimal expansion; there are only countably many positions; and a countable union of finite sets is certainly countable. Explain carefully to Fred what is wrong with his argument.

3.2.20. Exercise. Let ϵ be an arbitrary number greater than zero. Show that the rationals in $[0,1]$ can be covered by a countable family of open intervals the sum of whose lengths is no greater than ϵ. (Recall that a family $\mathfrak{U}$ of sets is said to COVER a set A if $A \subseteq \bigcup \mathfrak{U}$.) Is it possible to cover the set $\mathbb{Q}$ of all rationals in $\mathbb{R}$ by such a family?

3.2.21. Example. (Definition: The OPEN DISK in $\mathbb{R}^2$ with radius $r > 0$ and center (p,q) is defined to be the set of all points (x,y) in $\mathbb{R}^2$ such that $(x-p)^2 + (y-q)^2 < r^2$.) The family of all open disks in the plane whose centers have rational coordinates and whose radii are rational is countable.

3.2.22. Example. (Definition: A real number is ALGEBRAIC if it is a root of some polynomial of degree greater than 0 with integer coefficients. A real number which is not algebraic is TRANSCENDENTAL. It can be shown that the numbers π and e, for example, are transcendental.) The set of all transcendental numbers in $\mathbb{R}$ is uncountable. *Hint.* Start by showing that the set of polynomials with integer coefficients is countable.

3.2.23. Example. The set of all sequences each of whose terms is either 0 or 1 is uncountable.

3.2.24. Example. Let $\mathfrak{J}$ be a disjoint family of intervals in $\mathbb{R}$ each with length strictly greater than 0. Then $\mathfrak{J}$ is countable.

3.2.25. Example. The family $\mathcal{F}(\mathbb{R},\mathbb{R})$ of all real valued functions of a real variable is an uncountable set that is not cardinally equivalent to $\mathbb{R}$. *Hint.* Let $\mathcal{F} = \mathcal{F}(\mathbb{R},\mathbb{R})$. Assume there exists a bijection $\phi\colon \mathbb{R} \to \mathcal{F}$. What about the function f defined by

$$f(x) = 1 + \big(\phi(x)\big)(x)$$

for all $x \in \mathbb{R}$?

CHAPTER 4

GROUPS, VECTOR SPACES, AND ALGEBRAS

4.1. Operations

4.1.1. Definition. Let S be a set. A function from S into itself is sometimes called a UNARY OPERATION. A function from $S \times S$ into S is a BINARY OPERATION. Usually when the word "operation" is used without a modifier "binary operation" is meant.

4.1.2. Example. If S is a nonempty set, then the map $A \mapsto A^c$, which takes a subset A of S to its complement (in S), is a unary operation on $\mathfrak{P}(S)$, the power set of S.

4.1.3. Example. If S is a nonempty set, the map which takes a pair of subsets A and B to their union $A \cup B$ is a binary operation on $\mathfrak{P}(S)$. The notational convention here, putting the name of the function between the two operands, is standard. It is almost always the case that if $*$ is a binary operation we write $a * b$ or just ab for the value of $*$ at (a, b)—rather than something like $*(a, b)$.

4.1.4. Definition. Let S be a set, $*$ a binary operation on S, and $A \subseteq S$. The set A is CLOSED UNDER $*$ if $a * b$ belongs to A whenever a and b are members of A. Similarly, we say that the set A is closed under a unary operation $s \mapsto s'$ on S if a' belongs to A whenever a does. In chapter 10 we will encounter the notion of a set in a topological space being *closed*. One should be careful to distinguish this (topological) use of the word "closed" from the (algebraic) concept of a set being closed under some binary operation.

4.1.5. Example. The set of real numbers $r > 0$ is closed under the usual operations of addition, multiplication, and division, but not under subtraction.

4.1.6. Definition. Let S be a set and $*$ a binary operation on S. The operation $*$ is

(a) ASSOCIATIVE if $(s * t) * u = s * (t * u)$ for all s, t, $u \in S$; and
(b) COMMUTATIVE if $s * t = t * s$ for all s, $t \in S$.

If $*$ and $\diamond$ are both binary operations on S and

$$s * (t \diamond u) = (s * t) \diamond (s * u) \tag{4.1}$$

$$(t \diamond u) * s = (t * s) \diamond (u * s) \tag{4.2}$$

hold for all s, t, $u \in S$, we say that $*$ DISTRIBUTES over $\diamond$. Equation (4.1) is a LEFT DISTRIBUTIVE LAW, and (4.2) is a RIGHT DISTRIBUTIVE LAW.

4.1.7. Example. If S is a nonempty set the operations of union $\cup$ and intersection $\cap$ are both associative and commutative. Furthermore, union distributes over intersection and intersection distributes over union. In the real number system multiplication distributes over addition but addition does *not* distribute over multiplication.

4.1.8. Definition. If S is a set and $*$ is an associative binary operation on S, then the pair $(S, *)$ is a SEMIGROUP. It is illogical, but standard, to say *"let S be a semigroup"* instead of *"let $(S, *)$ be a semigroup"* and to say that the semigroup S is commutative when we mean that its operation $*$ is.

4.1.9. Definition. Let S be a set and $*$ a binary operation on S. An element $\mathbf{1}_S$ of S is an IDENTITY ELEMENT with respect to $*$ if

$$\mathbf{1}_S * s = s * \mathbf{1}_S = s$$

for all $s \in S$.

4.1.10. Definition. Let S be a set which has an identity element $\mathbf{1}_S$ with respect to a binary operation $*$ on S and let s be an element of S. An element s' of S which satisfies $s' * s = \mathbf{1}_S$ is a LEFT INVERSE of s (with respect to $*$). If it satisfies $s * s' = \mathbf{1}_S$ it is a RIGHT INVERSE of s (with respect to $*$). It is an INVERSE of s if it is both a left and right inverse of s. An element is INVERTIBLE if it has an inverse.

4.1.11. Definition. A semigroup $(S, *)$ with an identity is called a MONOID.

4.1.12. Example. If S is a nonempty set, the family $\mathcal{F}(S, S)$ of all functions from S into itself is a monoid under composition.

4.1.13. Example. The real numbers under addition form a monoid; so do the real numbers under multiplication.

4.1.14. Example. Monoids are used extensively in computer science. Search the Wikipedia [46] for examples such as *syntactic, transformation, trace,* and *history monoids.*

4.1.15. Proposition. *The identity element of a monoid is unique.*

4.1.16. Proposition. *An element of a monoid can have at most one inverse.*

4.1.17. Proposition. *If a is an invertible element of a monoid, then a^{-1} is also invertible and*

$$\left(a^{-1}\right)^{-1} = a\,.$$

4.1.18. Proposition. *If a and b are invertible elements of a monoid, then their product ab is also invertible and*

$$(ab)^{-1} = b^{-1}a^{-1}\,.$$

We can improve on proposition 4.1.16 a bit.

4.1.19. Proposition. *If an element c of a monoid has both a left inverse c^l and a right inverse c^r, then it is invertible and $c^{-1} = c^l = c^r$.*

Hint for proof. Look at your proof of proposition 2.10.1.

4.1.20. Corollary. *Let a and b be elements of a monoid. If both ab and ba are invertible, then so are a and b.*

4.2. Groups

4.2.1. Definition. A monoid $(G, *)$ is a GROUP if every element in G has an inverse with respect to $*$. It is customary to say, *"let G be an group"* rather than, *let $(G, *)$ be a group"*. Usually the identity element of an arbitrary group G is denoted by $\mathbf{1}_G$ or simply by $\mathbf{1}$ if no confusion is likely to arise. The standard notation for the inverse of an element a in a group is a^{-1}.

A group whose operation $*$ is commutative is usually called an ABELIAN GROUP. In most cases the preferred notation for the binary operation of an Abelian group G is $+$. Its identity is then denoted by $\mathbf{0}_G$ (or just by $\mathbf{0}$) and the inverse of an element a by $-a$. The notation $a-b$ is shorthand for $a + (-b)$. In brief, arbitrary groups are written multiplicatively and (usually) Abelian groups additively.

4.2.2. Example. In the real numbers $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ itself are Abelian groups under ordinary addition. The set $\mathbb{N}$ of natural numbers (under addition) is a commutative semigroup but not a group.

4.2.3. Example. Let G be the set of nonzero real numbers and $\cdot$ be ordinary multiplication. Then $(G,\,\cdot\,)$ is an Abelian group. The identity element is the number 1 and the inverse of $x \in G$ is its reciprocal. This is one of the few situations in which the operation of an Abelian group is (for obvious reasons) written multiplicatively (using $\cdot$) rather than additively (using $+$).

4.2.4. Example. Let $\mathbb{Z}_2 = \{0, 1\}$ and define addition on $\mathbb{Z}_2$ to be *addition modulo* 2; that is, define $0 + 0 = 1 + 1 = 0$ and $0 + 1 = 1 + 0 = 1$. This makes $\mathbb{Z}_2$ into an Abelian group.

4.2.5. Example. Recall that an $m \times n$ MATRIX of real numbers is a rectangular array of real numbers with m rows and n columns. If a is such a matrix, then a_{ij} is the element in the i^{th} row and the j^{th} column. It is common to use the notation $[a_{ij}]$ for the matrix a itself. If you find the notion of "rectangular array" uncongenial, define an $m \times n$ matrix of real numbers to be a function from $\mathbb{N}_m \times \mathbb{N}_n$ into $\mathbb{R}$. We denote by M_n the set of all $n \times n$ matrices of real numbers. Define addition of two $n \times n$ matrices coordinatewise; that is, if a, $b \in M_n$, define $a + b$ to be the $m \times n$ matrix whose entry in the i^{th} row and the j^{th} column is $a_{ij} + b_{ij}$. It is easy to see that this makes M_n into an Abelian group.

4.2.6. Example. If S is a set and G is an Abelian group, then $\mathcal{F}(S, G)$ (see 2.3.2) is an Abelian group under POINTWISE ADDITION, which is defined for all f, $g \in \mathcal{F}(S, G)$ by

$$(f + g)(s) = f(s) + g(s)$$

for all $s \in S$. Notice that in the displayed equation the $+$ on the left side is the addition that is being defined in $\mathcal{F}(S, G)$ while the one on the right is addition in G. It would be correct, but perhaps overly pedantic, to write

$$\big(f +_{\mathcal{F}(S,G)} g\big)(s) = f(s) +_G g(s).$$

4.2.7. Example. If S is a nonempty set, then $(\mathfrak{P}(S), \triangle)$, the power set of S under the operation of symmetric difference (see section 1.6), is an Abelian group.

4.2.8. Example. A real valued function f defined on a subset of the real line belongs to $\mathcal{F}_0$ if its domain contains some open interval which contains 0. For f, $g \in \mathcal{F}_0$ define $f + g$ pointwise; that is, let

$$(f + g)(x) = f(x) + g(x)$$

for all x common to the domains of f and g.

(a) Explain why $\mathcal{F}_0$ is not an Abelian group under this operation.
(b) Suppose we say that two functions f and g in $\mathcal{F}_0$ are "equivalent" if $f(x) = g(x)$ for all x in some open interval containing 0. Show that this relation really is an equivalence relation and explain carefully how the set of equivalence classes can be made into an Abelian group under "pointwise addition". (These equivalence classes are called GERMS of functions.)

4.2.9. Proposition. *Let a be an element of an arbitrary group. If $a^2 = a$, then $a = \mathbf{1}$.*

4.2.10. Definition. Let G be a group and H be a subset of G containing the identity element of G. Then H is a SUBGROUP of G if it is closed under the group operation of G and is a group under that operation.

4.2.11. Proposition. *The intersection of a family of subgroups of a group G is itself a subgroup of G.*

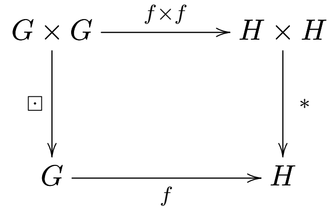
4.3. Homomorphisms of Semigroups and Groups

4.3.1. Definition. Let $(G, \boxdot\,)$ and $(H, *\,)$ be semigroups. A map $f\colon G \to H$ is a SEMIGROUP HOMOMORPHISM if

$$f(x \boxdot y) = f(x) * f(y)$$

for all x, $y \in G$. We will denote by $\operatorname{Hom}(G, H)$ the set of all homomorphisms from G into H and will abbreviate $\operatorname{Hom}(G, G)$ to $\operatorname{Hom}(G)$. A GROUP HOMOMORPHISM is just a homomorphism of the corresponding semigroups.

Here is another, diagrammatic, way of saying the same thing: the function f is a homomorphism if the diagram



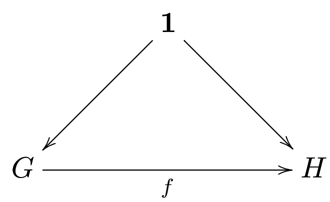
commutes.

4.3.2. Definition. Let $(G, \boxdot)$ and $(H, *)$ be semigroups. An ISOMORPHISM from G to H is a homomorphism $f\colon G \to H$ such that f^{-1} exists and is also a homomorphism. If an isomorphism from G to H exists, then G and H are ISOMORPHIC.

4.3.3. Example. Let G be the Abelian group of real numbers under addition and H be the Abelian group $\{x \in \mathbb{R}\colon x > 0\}$ of strictly positive real numbers under multiplication. The map $\exp\colon G \to H\colon x \mapsto e^x$ is an isomorphism. (We take the most basic properties of the elementary functions as prerequisite material. For more on this see the disclaimer in section 6.3.)

4.3.4. Proposition. *Every bijective homomorphism of semigroups is an isomorphism.*

4.3.5. Exercise. Let $\mathbf{1}$ be the one element group and $f\colon G \to H$ be a homomorphism between arbitrary groups G and H. Show that the following diagram of homomorphisms commutes.



What does this result say in ordinary English?

The preceding result need not hold in monoids. It is common practice to define a MONOID HOMOMORPHISM as a semigroup homomorphism which preserves the identity element.

4.3.6. Example. Regard $\mathbb{R}$ as a monoid under ordinary multiplication and $\mathbb{R}^2$ as a monoid under pointwise multiplication (where $(a,b)(c,d) = (ac,bd)$). Then the inclusion map $a \mapsto (a,0)$ of $\mathbb{R}$ into $\mathbb{R}^2$ is a semigroup homomorphism which does not preserve the identity element.

In addition to preserving identity elements group homomorphisms also automatically preserve inverses.

4.3.7. Proposition. *Let $f\colon G \to H$ be a group homomorphism. Then $f\big(x^{-1}\big) = \big(f(x)\big)^{-1}$ for each $x \in G$.*

4.3.8. Definition. Let G and H be Abelian groups. For f and g in $\operatorname{Hom}(G,H)$ define

$$f + g\colon G \to H\colon x \mapsto f(x) + g(x).$$

4.3.9. Proposition. *Let G and H be Abelian groups. With addition as defined above* $\operatorname{Hom}(G,H)$ *is an Abelian group.*

4.3.10. Notation. Let G, H, and J be Abelian groups and $f\colon G \to H$ and $g\colon H \to J$ be homomorphisms. Usually the composite of g with f is denoted by gf (rather than by $g \circ f$). That is,

$$gf\colon G \to J\colon x \mapsto g(f(x)).$$

4.3.11. Proposition. *Let G, H, and J be Abelian groups. If $f \in \operatorname{Hom}(G,H)$ and $g \in \operatorname{Hom}(H,J)$, then the composite of g with f belongs to* $\operatorname{Hom}(G,J)$.

4.3.12. Definition. Let G and H be Abelian groups and $f: G \to H$ be a homomorphism. Then $\{x \in G: f(x) = \mathbf{0}\}$ is the KERNEL of f and is denoted by $\ker f$.

4.3.13. Proposition. *Let $f: G \to H$ be a homomorphism of Abelian groups. Then $\ker f$ is a subgroup of G and $\operatorname{ran} f$ is a subgroup of H.*

4.4. Vector Spaces

4.4.1. Definition. Let $\mathbb{F}$ be a field. A VECTOR SPACE OVER $\mathbb{F}$ is an Abelian group V together with a mapping $m: \mathbb{F} \times V \to V: (\alpha, v) \mapsto \alpha v$, called SCALAR MULTIPLICATION, which satisfies the following axioms for all α, $\beta \in \mathbb{F}$ and x, $y \in V$:

(i) $\alpha(x + y) = \alpha x + \alpha y$;
(ii) $(\alpha + \beta)x = \alpha x + \beta x$;
(iii) $(\alpha\beta)x = \alpha(\beta x)$; and
(iv) $1\,x = x$.

The elements of V are called VECTORS and, in this context, the elements of $\mathbb{F}$ are called SCALARS.

Definitions of this sort, especially if used thoughtlessly, can make mathematicians unpopular in some circles. Suppose an engineer comes up to a mathematician and asks in all good faith, *You're an expert at this linear algebra stuff. What is a* vector*, really?* It is unhelpful and perhaps a bit insensitive—but of course entirely accurate—to reply, *Well, it's just an element of a vector space.*

A vector space over the field $\mathbb{R}$ of real numbers (see section 6.1) is a REAL VECTOR SPACE and one over the field $\mathbb{C}$ of complex numbers (see section 6.6) is a COMPLEX VECTOR SPACE. In these notes the only vector spaces which occur are the ones with real or complex scalars. Since the real vector spaces occur most frequently we make the following agreement.

4.4.2. Convention. In the sequel the unmodified term "vector space" will mean "real vector space". When we want a vector space V to have complex scalars, we will say so: we will say *Let V be a complex vector space.* We will ordinarily follow the usual convention and write "Let V be a vector space," when to be correct we should write "Let $(V, +, m)$ be a vector space," or even more precisely, "Let $(V, +, m, \mathbb{R}))$ be a real vector space."

4.4.3. Example. The vector space $V = \{\mathbf{0}\}$ comprising a single element is the ZERO VECTOR SPACE. When we say that a vector space is NONTRIVIAL we mean that it is not the zero vector space.

4.4.4. Example. The simplest nontrivial example of a (real) vector space is the set $\mathbb{R}$ of real numbers itself. Here both vectors and scalars are real numbers and scalar multiplication is ordinary multiplication of real numbers.

4.4.5. Example. Euclidean n-space $\mathbb{R}^n$ is a vector space under coordinatewise operations of addition and scalar multiplication:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) := (x_1 + y_1, \dots, x_n + y_n)$$

and

$$\alpha\,(x_1, \dots, x_n) := (\alpha\, x_1, \dots, \alpha\, x_n)$$

for $\alpha \in \mathbb{R}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$.

4.4.6. Example. The Abelian group M_n of $n \times n$ matrices of real numbers (see example 4.2.5) becomes a vector space when we define scalar multiplication coordinatewise:

$$\alpha\, a := [\alpha\, a_{ij}]$$

for $\alpha \in \mathbb{R}$ and $a = [a_{ij}] \in M_n(\mathbb{R})$.

4.4.7. Example. If S is a set and V is a vector space, then the Abelian group $\mathcal{F}(S,V)$ (see example 4.2.6) becomes a vector space under pointwise scalar multiplication:

$$(\alpha f)(s) := \alpha f(s)$$

for $\alpha \in \mathbb{R}$ and $f \in \mathcal{F}(S,V)$. Recall that for the special case of real valued functions we write $\mathcal{F}(S)$ instead of $\mathcal{F}(S,\mathbb{R})$.

4.4.8. Proposition. *Let x be a vector in some vector space. Then $(-1)\,x = -x$.*

4.4.9. Proposition. *Let x be a vector in some vector space and α be a scalar. Then $\alpha x = \mathbf{0}$ if and only if $\alpha = 0$ or $x = \mathbf{0}$.*

4.4.10. Notation. If x and y are vectors in some vector space, we define $x - y$ to be $x + (-y)$. If A and B are subsets of a vector space, we define

$$A + B := \{a + b : a \in A, b \in B\};$$

and if $\alpha \in \mathbb{R}$,

$$\alpha A := \{\alpha a : a \in A\}.$$

We also take $-A$ to be $(-1)A$, which, according to exercise 4.4.8, is the same as $\{-a\colon a \in A\}$. And of course, $A - B$ means $A + (-B)$.

4.4.11. Definition. A nonempty subset W of a vector space V is a SUBSPACE of V if it is a vector space under the operations of addition and scalar multiplication it inherits from V.

4.4.12. Proposition. *A nonempty subset of a vector space V is a subspace of V if it is closed under the operations of addition and scalar multiplication of V.*

4.4.13. Definition. A real valued function f on a set S is BOUNDED if there exists a number $M > 0$ such that $|f(s)| \le M$ for all $s \in S$. We denote the family of all bounded real valued functions on S by $\mathcal{B}(S,\mathbb{R})$ or, more briefly, by $\mathcal{B}(S)$.

4.4.14. Example. If S is a nonempty set, then the family $\mathcal{B}(S) = \mathcal{B}(S,\mathbb{R})$ of bounded real valued functions on S is a subspace of the vector space $\mathcal{F}(S)$ (see example 4.4.7).

4.5. Linear Transformations

4.5.1. Definition. A map $T\colon V \to W$ between vector spaces is LINEAR if

$$T(x + y) = Tx + Ty \qquad \text{for all } x,\, y \in V \tag{4.3}$$

and

$$T(\alpha x) = \alpha Tx \qquad \text{for all } x \in V \text{ and } \alpha \in \mathbb{R}. \tag{4.4}$$

A linear map is frequently called a LINEAR TRANSFORMATION, and, in case the domain and codomain are the same, it is usually called a LINEAR OPERATOR. The family of all linear transformations from V into W is denoted by $\mathfrak{L}(V,W)$. We shorten $\mathfrak{L}(V,V)$ to $\mathfrak{L}(V)$.

Two oddities of notation concerning linear transformations deserve comment. First, the value of T at x is usually written Tx rather than $T(x)$. Naturally the parentheses are used whenever their omission would create ambiguity. For example, in (4.3) above $Tx + y$ is not an acceptable substitute for $T(x+y)$. Second, as with group homomorphisms, the symbol for composition of two linear transformations is ordinarily omitted. If $S \in \mathfrak{L}(U,V)$ and $T \in \mathfrak{L}(V,W)$, then the composite of T and S is denoted by TS (rather than by $T \circ S$). As a consequence of this convention when $T \in \mathfrak{L}(V)$ the linear operator $T \circ T$ is written as T^2, $T \circ T \circ T$ as T^3, and so on.

The first condition (4.3) defining linearity says that a linear map must be a homomorphism of Abelian groups. We have already seen (at the beginning of section 4.3) how to view this as a commutative diagram. Condition (4.4) of the definition can similarly be thought of in terms of a diagram. For each scalar α define the function M_α, MULTIPLICATION BY α, from a vector space into itself by

$$M_\alpha(x) = \alpha x\,.$$

(We use the same symbol for multiplication by α in both of the spaces V and W.) Then condition (4.4) holds if and only if for every scalar α the following diagram commutes.

$$\begin{array}{ccc} V & \xrightarrow{\;T\;} & W \\ {\scriptstyle M_\alpha}\big\downarrow & & \big\downarrow{\scriptstyle M_\alpha} \\ V & \xrightarrow[\;T\;]{} & W \end{array}$$

4.5.2. Example. If $T\colon \mathbb{R}^3 \to \mathbb{R}^2\colon x \mapsto (x_1 + x_3, x_1 - 2x_2)$, then T is linear.

4.5.3. Proposition. *Let $T\colon V \to W$ be a linear transformation between vector spaces. Then*

(a) $T(\mathbf{0}) = \mathbf{0}$.
(b) $T(x - y) = Tx - Ty$ *for all* x, $y \in V$.

4.5.4. Example. Each coordinate projection defined on $\mathbb{R}^n$

$$\pi_k\colon \mathbb{R}^n \to \mathbb{R}\colon x \mapsto x_k$$

is linear.

4.5.5. Example. Let $\mathcal{F} = \mathcal{F}\big((a,b),\mathbb{R}\big)$ be the family of all real valued functions defined on the open interval (a,b) and let $\mathcal{D} = \mathcal{D}\big((a,b),\mathbb{R}\big)$ be the set of all members of $\mathcal{F}$ which are differentiable at each point of (a,b). Then $\mathcal{D}$ is a vector subspace of $\mathcal{F}$ and that the differentiation operator

$$D\colon \mathcal{D} \to \mathcal{F}\colon f \mapsto f'$$

(where f' is the derivative of f) is linear.

An important observation is that the composite of two linear transformations is linear.

4.5.6. Proposition. *Let U, V, and W be vector spaces. If $S \in \mathfrak{L}(U,V)$ and $T \in \mathfrak{L}(V,W)$, then $TS \in \mathfrak{L}(U,W)$.*

4.5.7. Definition. If $T\colon V \to W$ is a linear transformation between vector spaces, then the KERNEL (or NULL SPACE) of T, denoted by $\ker T$, is its kernel as a group homomorphism (see 4.3.12). That is,

$$\ker T = T^{\leftarrow}\{\mathbf{0}\} = \{x \in V : Tx = \mathbf{0}\}.$$

4.5.8. Exercise. In $\mathbb{R}^3$ we call the vectors $e^1 = \mathbf{i} = (1,0,0)$, $e^2 = \mathbf{j} = (0,1,0)$, and $e^3 = \mathbf{k} = (0,0,1)$ the STANDARD BASIS VECTORS for $\mathbb{R}^3$. Suppose that $T \in \mathfrak{L}(\mathbb{R}^3,\mathbb{R}^3)$ satisfies

$$\begin{aligned} Te^1 &= (1,-2,3) \\ Te^2 &= (0,0,0) \\ Te^3 &= (-2,4,-6)\,. \end{aligned}$$

Find and describe geometrically the kernel of T and the range of T.

It is useful to know that the kernel of a linear transformation is always a vector subspace of its domain, that its image is a vector subspace of its codomain, and that a necessary and sufficient condition for a linear transformation to be injective is that its kernel contain only the zero vector.

4.5.9. Proposition. *If $T\colon V \to W$ is a linear transformation between vector spaces, then* $\ker T$ *is a subspace of V and* $\operatorname{ran} T$ *is a subspace of W.*

4.5.10. Proposition. *A linear transformation T is injective if and only if* $\ker\ T\ =\ \{0\}$.

4.5.11. Proposition. *Every bijective linear map between vector spaces is an isomorphism.*

Hint for proof. Part of this has already been done in exercise 4.3.4.

4.5.12. Definition. Let V be a vector space. A linear map from a (real) vector space into its field $\mathbb{R}$ of scalars is a LINEAR FUNCTIONAL. We will denote the family of all linear functionals on V by $V^\dagger$ rather than by $\mathfrak{L}(V,\mathbb{R})$. This space is the ALGEBRAIC DUAL SPACE of V.

4.5.13. Example. Let S be a nonempty set and $V = \mathcal{F}(S)$ be the vector space of all real valued functions on S. For every $x \in S$ the function

$$E_x \colon V \to \mathbb{R} \colon f \mapsto f(x)$$

is a linear functional on V. The functional E_x belongs to $V^\dagger$ and is called the EVALUATION FUNCTIONAL at x.

4.5.14. Example. We know from beginning calculus that the family $\mathcal{R} = \mathcal{R}\big([a,b],\mathbb{R}\big)$ of Riemann integrable functions is closed under the pointwise operations of addition and multiplication by real scalars. Also recall that

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

where $\alpha \in \mathbb{R}$ and $f \in \mathcal{R}$, and that

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

where $f, g \in \mathcal{R}$. It is then clear that $\mathcal{R}$ is a vector subspace of $\mathcal{F}$ and that the function $f \mapsto \int_a^b f(x)\,dx$ is a linear functional on $\mathcal{R}$.

4.6. Rings and Algebras

4.6.1. Definition. Let $(R,+)$ be an Abelian group and $\cdot$ a binary operation on R (which we call *multiplication*). Then $(R,+,\cdot)$ is a RING if the multiplication operation is associative and distributes over addition. If, in addition, there exists an identity element $\mathbf{1}_R$ in R with respect to multiplication (sometimes called a *unit*), then we say that $(R,+,\cdot)$ is a UNITAL RING. A ring is said to be COMMUTATIVE if the multiplication operation is commutative.

Since a unital ring has an identity element with respect to two different operations we distinguish them by referring to the *additive identity* and to the *multiplicative identity*. We also refer to *additive inverses* and (when they exist) *multiplicative inverses*. Most often we write ab for $a \cdot b$.

We customarily say, "Let R be a ring," rather than, "Let $(R,+,m)$ be a ring." As a matter of notation it is usually harmless to denote the multiplicative identity of a unital ring by $\mathbf{1}$. If a and b are elements of a ring, then ab is called the *product* of a and b and is occasionally written as $a \cdot b$. We write a^2 for $a \cdot a$, and a^3 for $a \cdot a \cdot a$, and so on.

4.6.2. CAUTION. Many authors define rings to be both unital and commutative.

4.6.3. Example. Under the additional operation of ordinary multiplication the Abelian groups $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ (see example 4.2.2) become commutative unital rings.

4.6.4. Example. Let $\mathbb{Z}_2$ be the Abelian group given in example 4.2.4. If we define $0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0$ and $1 \cdot 1 = 1$, then $\mathbb{Z}_2$ becomes a commutative unital ring.

4.6.5. Example. Let S be a nonempty set. According to example 4.2.7 $(\mathfrak{P}(S), \triangle)$ is an Abelian group. With intersection as "multiplication" $(\mathfrak{P}(S), \triangle, \cap)$ is a commutative unital ring.

4.6.6. Example. Let G be an Abelian group. Then $\operatorname{Hom}(G)$ is a unital ring under pointwise addition and composition. (See exercise 4.3.9.)

4.6.7. Definition. Let $(A, +, \cdot\,)$ be a ring which is additionally equipped with a scalar multiplication m such that $(A, +, m)$ is a vector space. If

$$\alpha(xy) = (\alpha x)y = x(\alpha y) \tag{4.5}$$

for each x, $y \in A$ and $\alpha \in \mathbb{R}$, then $(A, +, \cdot\,, m)$ is a (real) ALGEBRA (sometimes referred to as a LINEAR ASSOCIATIVE ALGEBRA). Of course, if A is a complex vector space, condition (4.5) should hold for all complex numbers α. We abuse notation in the standard way by writing such things as, "Let A be an algebra." We say that an algebra A is UNITAL if its underlying ring $(A, +, \cdot\,)$ is. And it is COMMUTATIVE if its ring is.

4.6.8. Convention. Algebras, like vector spaces, may have either real or complex scalars. In these notes algebras will be assumed to have real scalars unless we specify otherwise. Later on when we want the scalars of an algebra A to come from the field $\mathbb{C}$ of complex numbers, we will say that A *is a complex algebra.*

4.6.9. Example. If S is a nonempty set, then the vector space $\mathcal{F}(S)$ (see example 4.4.7) is a commutative unital algebra under POINTWISE MULTIPLICATION, which is defined for all f, $g \in \mathcal{F}(S)$ by

$$(f \cdot g)(s) = f(s) \cdot g(s)$$

for all $s \in S$. The constant function $\mathbf{1}$ (that is, the function whose value at each $s \in S$ is 1) is the multiplicative identity.

4.6.10. Example. We have seen in example 4.4.6 that the set M_n of $n \times n$ matrices of real numbers is a vector space. If $a = \begin{bmatrix} a_{ij} \end{bmatrix}$ and $b = \begin{bmatrix} b_{kl} \end{bmatrix}$ are $n \times n$ matrices, then the PRODUCT of a and b is the $n \times n$ matrix $c = ab$ whose entry in the i^{th} row and k^{th} column is $c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$. This definition makes M_n into a unital algebra.

4.6.11. Example. If V is a vector space, then the set $\mathfrak{L}(V)$ of linear operators on V is a unital algebra under pointwise addition, pointwise scalar multiplication, and composition.

4.6.12. Definition. An element a of a ring R is an ANNIHILATOR of R if $ax = xa = \mathbf{0}$ for every $x \in R$.

4.6.13. Proposition. *The additive identity of a ring is an annihilator.*

4.6.14. Proposition. *If a and b are elements of a ring, then $(-a)b = a(-b) = -ab$ and $(-a)(-b) = ab$.*

4.6.15. Proposition. *Let a and b be elements of a unital ring. Then $\mathbf{1} - ab$ is invertible if and only if $\mathbf{1} - ba$ is.*

Hint for proof. Look at the product of $\mathbf{1} - ba$ and $\mathbf{1} + bca$ where c is the inverse of $1 - ab$.

4.6.16. Definition. A nonzero element a of a ring is a ZERO DIVISOR (or DIVISOR OF ZERO) if there exists a nonzero element b of the ring such that $ab = \mathbf{0}$ or $ba = \mathbf{0}$.

4.6.17. Definition. An element a of a ring (or, more generally, a semigroup) is LEFT CANCELLABLE if $ab = ac$ implies that $b = c$. It is RIGHT CANCELLABLE if $ba = ca$ implies that $b = c$. A ring (or a semigroup) has the CANCELLATION PROPERTY if every nonzero element of the ring is both left and right cancellable.

4.6.18. Proposition. *A ring has the cancellation property if and only if it has no zero divisors.*

4.6.19. Definition. A FIELD is a commutative unital ring R in which the additive and multiplicative identities are distinct ($\mathbf{1}_R \neq \mathbf{0}_R$) and every nonzero element x has a multiplicative inverse x^{-1}.

4.6.20. Example. The rings $\mathbb{R}$ and $\mathbb{Q}$ are fields; $\mathbb{Z}$ is not.

4.6.21. Example. The ring $\mathbb{Z}_2$ (see example 4.6.4) is a field.

4.6.22. Definition. An element a of a ring (or, more generally, a semigroup) is IDEMPOTENT if $a^2 = a$. A ring R is a BOOLEAN RING if every element of the ring is idempotent.

4.6.23. Example. The ring $\mathbb{Z}_2$ (see example 4.6.4) is a unital Boolean ring.

4.6.24. Example. Let S be a nonempty set. The family $\mathcal{F}(S, \mathbb{Z}_2)$ of all functions from S into the field $\mathbb{Z}_2$ (see example 4.6.21) under pointwise addition and multiplication is a unital Boolean ring.

4.6.25. Example. Let S be a nonempty set. The ring $(\mathfrak{P}(S), \triangle, \cap)$ (see example 4.6.5) is a Boolean ring.

4.6.26. Proposition. *For every element a of a Boolean ring $a + a = \mathbf{0}$.*

4.6.27. Proposition. *Every Boolean ring is commutative.*

4.6.28. Definition. A subset of an algebra A which is closed under the operations of addition, multiplication, and scalar multiplication is a SUBALGEBRA of A. If A is a unital algebra and B is a subalgebra of A which contains the multiplicative identity of A, then B is a UNITAL SUBALGEBRA of A.

4.6.29. CAUTION. Be very careful with the preceding definition. It is possible for B to be a subalgebra of an algebra A and to be a unital algebra but still not be a unital subalgebra of A! The definition requires that for B to be a unital subalgebra of A the identity of B must be the same as the identity of A. *Example:* Under pointwise operations $A = \mathbb{R}^2$ is a unital algebra. The set $B = \{(x, 0) \colon x \in \mathbb{R}\}$ is a subalgebra of A. And certainly B is unital (the element $(1, 0)$ is the multiplicative identity of B). But B is *not* a unital subalgebra of A because it does not contain the multiplicative identity $(1, 1)$ of A.

4.6.30. Example. Let S be a nonempty set. The family $\mathcal{B}(S)$ of all bounded real valued functions on S is a unital subalgebra of the algebra $\mathcal{F}(S)$ of all real valued functions on S.

4.6.31. Definition. A LEFT IDEAL in an algebra A is a vector subspace J of A such that $AJ \subseteq J$. (For RIGHT IDEALS, of course, we require $JA \subseteq J$.) We say that J is an IDEAL if it is both a left and a right ideal. A PROPER ideal is an ideal which is a proper subset of A. A MAXIMAL ideal is a proper ideal that is properly contained in no other proper ideal. We denote the family of all maximal ideals in an algebra A by $\operatorname{Max} A$.

4.6.32. Proposition. *No proper ideal in a unital algebra can contain an invertible element.*

4.6.33. Proposition. *Let a be an element of a commutative unital algebra A. If a is not invertible, then aA is a proper ideal in A. (It is called the* PRINCIPAL IDEAL *generated by a.)*

4.7. Ring and Algebra Homomorphisms

4.7.1. Definition. A function $f \colon R \to S$ between rings is a (RING) HOMOMORPHISM if it is a homomorphism of Abelian groups and preserves multiplication as well as addition; that is, if

$$f(x + y) = f(x) + f(y)$$

and

$$f(xy) = f(x)f(y) \tag{4.6}$$

for all x and y in R. If additionally R and S are unital rings and

$$f(\mathbf{1}_R) = \mathbf{1}_S \tag{4.7}$$

then f is a UNITAL (RING) HOMOMORPHISM. Obviously a ring homomorphism $f \colon R \to S$ is a group homomorphism of R and S regarded as Abelian groups. The KERNEL of f as a ring homomorphism is the kernel of f as a homomorphism of Abelian groups; that is $\ker f = \{x \in R \colon f(x) = \mathbf{0}\}$.

If f^{-1} exists and is also a ring homomorphism, then f is an ISOMORPHISM from R to S. If an isomorphism from R to S exists, then R and S are ISOMORPHIC.

4.7.2. Definition. A map $f\colon A \to B$ between algebras is an (ALGEBRA) HOMOMORPHISM if it is a linear map between A and B as vector spaces which preserves multiplication (in the sense of equation (4.6). In other words, an algebra homomorphism is a linear ring homomorphism. It is a UNITAL (ALGEBRA) HOMOMORPHISM if it preserves identities (as in (4.7)). The KERNEL of an algebra homomorphism $f\colon A \to B$ is, of course, $\{a \in A\colon f(a) = \mathbf{0}\}$.

If f^{-1} exists and is also an algebra homomorphism, then f is an ISOMORPHISM from A to B. If an isomorphism from A to B exists, then A and B are ISOMORPHIC.

Here are four essentially obvious facts about algebra homomorphisms.

4.7.3. Proposition. *Every bijective algebra (or ring) homomorphism is an isomorphism.*

4.7.4. Proposition. *If $f\colon A \to B$ is an isomorphism between algebras (or rings) and A is unital, then so is B and f is a unital homomorphism.*

4.7.5. Proposition. *Let A, B, and C be algebras (or rings). If $f\colon A \to B$ and $g\colon B \to C$ are homomorphisms, so is $gf\colon A \to C$. (As is the case with group homomorphism and linear maps, gf denotes the composite function $g \circ f$.) If f and g are unital, so is gf.*

4.7.6. Proposition. *If $f\colon A \to B$ is a bijection, A is an algebra, and B is any set, then we may equip B with algebraic operations in such a way that B is an algebra isomorphic to A.*

4.7.7. Remark. It is clear that the preceding proposition remains true if we substitute for the word "algebra" the name of nearly any type of algebraic object—semigroup, group, ring, field, vector space. It is worth remarking on a slight variation of this proposition, which can also be quite useful.

Suppose that $f\colon A \to B$ is a bijection where A is one of the algebraic structures we have discussed. For concreteness let's take A to be an algebra. Suppose further that B is a set equipped with corresponding algebraic operations (in this case, addition, multiplication, and scalar multiplication) and that f is operation preserving. Then B is itself an algebra and f is an (algebra) isomorphism.

4.7.8. Example. The algebras M_n (see 4.6.10) and $\mathfrak{L}(\mathbb{R}^n)$ (see 4.4.5 and 4.6.11) are isomorphic. If you found it irksome to prove that matrix multiplication was associative in example 4.6.10, explain how to use remark 4.7.7 to simplify that proof.

4.7.9. Example. Let S be a nonempty set. Notice that each element of $\mathcal{F}(S, \mathbb{Z}_2)$ is "essentially" a characteristic function on S except that it is $\mathbb{Z}_2$ valued rather than real valued. Thus we will denote a member f of $\mathcal{F}(S, \mathbb{Z}_2)$ by χ_A if $A = f^{\leftarrow}(\{1\})$. The map

$$\chi\colon \mathfrak{P}(S) \to \mathcal{F}(S, \mathbb{Z}_2)\colon A \mapsto \chi_A$$

is an isomorphism between the Boolean rings $\mathfrak{P}(S)$ and $\mathcal{F}(S, \mathbb{Z}_2)$ (see examples 4.6.24 and 4.6.25). If you found it clumsy to write a proof of the associative property for symmetric difference in proposition 1.6.4(a), use remark 4.7.7 to write a neater one.

CHAPTER 5

PARTIALLY ORDERED SETS

5.1. Partial and Linear Orderings

5.1.1. Definition. Let $\leq$ be a relation on a nonempty set S.

(a) If the relation $\leq$ is reflexive and transitive, it is a PREORDERING.
(b) If $\leq$ is a preordering and is also antisymmetric, it is a PARTIAL ORDERING.
(c) Elements x and y in a preordered set are COMPARABLE if either $x \leq y$ or $y \leq x$.
(d) If $\leq$ is a partial ordering with respect to which any two elements are comparable, it is a LINEAR ORDERING (or a TOTAL ORDERING).
(e) If the relation $\leq$ is a preordering (respectively, partial ordering, linear ordering) on S, then the pair $(S, \leq)$ is a PREORDERED SET (respectively, PARTIALLY ORDERED SET, LINEARLY ORDERED SET).
(f) A linearly ordered subset of a partially ordered set $(S, \leq)$ is a CHAIN in S.

We may write $b \geq a$ as a substitute for $a \leq b$. The notation $a < b$ (or equivalently, $b > a$) means that $a \leq b$ and $a \neq b$.

5.1.2. Example. The real line $\mathbb{R}$ is a linearly ordered set under its usual ordering $\leq$.

5.1.3. CAUTION. The most important of the preceding concepts is *partial ordering*. The notation $\leq$ for partial orderings tends to be somewhat misleading. Since the usual $\leq$ relation on the real numbers is a linear ordering it is an easy (and disastrous) mistake to treat a general partial ordering as a linear ordering, which usually it is not. Keep firmly in mind that two elements of a partially ordered set need not be comparable. A space of functions (under pointwise ordering, see example 5.1.8 below) is a much more reliable mental picture of a partially ordered set than the set $\mathbb{R}$ of real numbers.

5.1.4. Example. On $\mathbb{R}^2$ define a relation $\preceq$ by

$$(a, b) \preceq (c, d) \quad \text{if} \quad a \leq c.$$

Then $\preceq$ is a preordering which is not a partial ordering.

5.1.5. Example. On $\mathbb{R}^2$ define COORDINATEWISE ORDERING $\preceq$ by

$$(a, b) \preceq (c, d) \quad \text{if} \quad a \leq c \text{ and } b \leq d.$$

Coordinatewise ordering is a partial ordering on $\mathbb{R}^2$ which is not a linear ordering.

5.1.6. Example. On $\mathbb{R}^2$ define LEXICOGRAPHIC ORDERING $\preceq$ by

$$(a, b) \preceq (c, d) \quad \text{if} \quad \text{either } a < c \text{ or else } a = c \text{ and } b \leq d.$$

Lexicographic ordering is a linear ordering on $\mathbb{R}^2$. (Where do you think the name *lexicographic ordering* comes from? *Hint:* look up the word "lexicographic" in a dictionary.)

5.1.7. Example. Let S be a set which contains at least two elements. Then set inclusion is a partial ordering on the power set $\mathfrak{P}(S)$ which is not a linear ordering. (Why did we require S to have at least two elements?)

5.1.8. Example. If S is a nonempty set, then $\mathcal{F}(S) = \mathcal{F}(S, \mathbb{R})$ can be given a POINTWISE ORDERING by setting $f \leq g$ if $f(s) \leq g(s)$ for every $s \in S$. Under this ordering $\mathcal{F}(S)$ is a partially ordered set. If S contains more than one point, the ordering is *not* linear.

5.2. Infima and Suprema

5.2.1. Definition. Let $\leq$ be a partial ordering on a set S and $A \subseteq S$. An element $l \in S$ is a LOWER BOUND for A if $l \leq a$ for all $a \in A$. In this case we also say that A is BOUNDED BELOW (or MINORIZED) by l. Similarly, an element $u \in S$ is an UPPER BOUND for A if $u \geq a$ for all $a \in A$. In this case we also say that A is BOUNDED ABOVE (or MAJORIZED or DOMINATED) by u.

The element l is the GREATEST LOWER BOUND (or INFIMUM) of A if

(i) l is a lower bound for A, and
(ii) if m is any lower bound for A, then $l \geq m$.

In general we denote the infimum of a set A by either $\inf A$ or $\bigwedge A$. When $A = \{a_1, \dots, a_n\}$ is finite we usually write $a_1 \wedge a_2 \wedge \dots \wedge a_n$ or $\bigwedge_{k=1}^{n} a_k$ for $\bigwedge A$. If $A = \{a_1, a_2, \dots\}$ is denumerable, we may write $\bigwedge_{k=1}^{\infty} a_k$ for $\bigwedge A$.

Similarly, the element l is the LEAST UPPER BOUND (or SUPREMUM) of A if l is an upper bound for A, and if whenever M is any upper bound for A, then $l \leq M$. We denote the supremum of a set A by either $\sup A$ or $\bigvee A$. When $A = \{a_1, \dots, a_n\}$ is finite we usually write $a_1 \vee a_2 \vee \dots \vee a_n$ or $\bigvee_{k=1}^{n} a_k$ for $\bigvee A$. If $A = \{a_1, a_2, \dots\}$ is denumerable, we may write $\bigvee_{k=1}^{\infty} a_k$ for $\bigvee A$.

The "sup" in "$\sup A$" is pronounced like the English word "soup" not like the English word "sup" (to dine).

The plural of "infimum" is "infima"; the plural of "supremum" is "suprema".

A subset of a partially ordered set may have many upper bounds, but the definite article in the definition of "supremum" requires justification.

5.2.2. Proposition. *A subset of a partially ordered set can have at most one supremum and at most one infimum.*

5.2.3. Proposition. *Let A and B be nonempty subsets of a partially ordered set S with $B \subseteq A$. If both A and B have least upper bounds, then $\sup B \leq \sup A$. Also, if both A and B have greatest lower bounds, then $\inf B \geq \inf A$.*

5.2.4. Definition. Let $\leq$ be a partial ordering on a set S and $A \subseteq S$. An element $m \in A$ is a MAXIMAL element of A if there is no element of A strictly larger than m; that is, if $a = m$ whenever $a \in A$ and $a \geq m$. Similarly, $m \in A$ is a MINIMAL element of A if $a = m$ whenever $a \in A$ and $a \leq m$. An element $m \in A$ is the LARGEST (or GREATEST) element of A if $m \geq a$ for all $a \in A$. It is the SMALLEST (or LEAST) element of A if $m \leq a$ for all $a \in A$.

Clearly if m is the largest element of a set A it is a maximal element of A; in fact, it is the one and only maximal element of A.

5.2.5. Example. Let $S = \{a, b, c, d\}$ and $\mathcal{Q}$ be the family of all proper subsets of S. Then $\{a, b, c\}$, $\{a, b, d\}$, $\{a, c, d\}$, and $\{b, c, d\}$ are the maximal elements of $\mathcal{Q}$, but $\mathcal{Q}$ has no largest element.

5.2.6. Definition. A partially ordered set is ORDER COMPLETE if every nonempty subset has a supremum and an infimum. It is DEDEKIND COMPLETE if every nonempty subset which has an upper bound has a least upper bound. Clearly every order complete partially ordered set is Dedekind complete. Occasionally it is useful to have one more closely related concept. A partially ordered set is σ-DEDEKIND COMPLETE (or DEDEKIND σ-COMPLETE or COUNTABLY DEDEKIND COMPLETE) if every nonempty countable subset which is bounded above has a least upper bound. (The prefix "σ" is the Greek letter "sigma". It usually denotes something to do with countability.)

5.2.7. Example. The power set of a nonempty set (ordered by set inclusion) is order complete.

5.2.8. Example. The set of real numbers (under its usual ordering) is Dedekind complete (this is just the *least upper bound axiom*—see 6.1.6) but not order complete.

5.2.9. Proposition. *If S is a Dedekind complete partially ordered set, then every nonempty subset of S which has a lower bound has a greatest lower bound.*

Hint for proof. Suppose B is a nonempty subset of S which is bounded below. Notice that the set of all lower bounds of B is bounded above.

5.2.10. Definition. A function $f\colon S \to T$ between preordered sets is ORDER PRESERVING if $f(x) \leq f(y)$ in T whenever $x \leq y$ in S. It is ORDER REVERSING if $f(x) \geq f(y)$ in T whenever $x \leq y$ in S. We say that f is an ORDER ISOMORPHISM if it is bijective and $x \leq y$ in S if and only if $f(x) \leq f(y)$ in T.

5.2.11. Exercise. Give an example to show that a bijective order preserving map between partially ordered sets need not be an order isomorphism.

It is helpful to have a mental picture of objects with which one is working. Because of our familiarity with the ordering of the real line it may be tempting to use it as a visual model for a typical partially ordered set. In many situations this turns out to be a bad idea. The fact that the line is linearly ordered makes it too special to be a good guide. A much better visualization is a standard garden trellis or the family of continuous functions on the unit interval $[0, 1]$ (with pointwise ordering—see example 5.1.8). The next exercise shows the sort of thing that can go wrong when $\mathbb{R}$ is used as a visual model for a partially ordered set.

5.2.12. Exercise. Your good friend Fred R. Dimm is trying to prove a result about a partially ordered set S. He is able to show (correctly) that for a particular subset A of S, there exists an element $c \in S$ such that $c < \sup A$. He then concludes from this that there must be at least one element a of A such that $c < a$. Help Fred by finding a concrete example where this conclusion is false.

5.3. Zorn's Lemma

An axiom of set theory which is equivalent to the *axiom of choice* (see axiom 2.10.5 and 2.10.6) is the curiously named *Zorn's lemma.*

5.3.1. Axiom (Zorn's lemma)**.** *A partially ordered set in which every chain has an upper bound has a maximal element.*

One standard and elementary use of *Zorn's lemma* is to show that every vector space has a basis. Before proving this result we give a little background, with which you may be already familiar.

5.3.2. Definition. A vector $\mathbf{y}$ is a LINEAR COMBINATION of the (distinct) vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ if there exist scalars $\alpha_1, \dots \alpha_n$ such that $\mathbf{y} = \sum_{k=1}^{n} \alpha_k \mathbf{x}_k$. *Note:* linear combinations *are* finite sums. The linear combination $\sum_{k=1}^{n} \alpha_k \mathbf{x}_k$ is TRIVIAL if all the coefficients $\alpha_1, \dots \alpha_n$ are zero. If at least one α_k is different from zero, the linear combination is NONTRIVIAL.

5.3.3. Example. In $\mathbb{R}^2$ the vector $(8, 2)$ is a linear combination of the vectors $(1, 1)$ and $(1, -1)$.

5.3.4. Example. In $\mathbb{R}^3$ the vector $(1, 2, 3)$ is *not* a linear combination of the vectors $(1, 1, 0)$ and $(1, -1, 0)$.

5.3.5. Definition. Suppose that A is a subset (finite or not) of a vector space V. The SPAN of A is the set of all linear combinations of elements of A. For a nonempty set A there is another way of saying the same thing: the SPAN of A is the smallest subspace of V which contains A. We denote the span of A by $\operatorname{span} A$. If $U = \operatorname{span} A$, we say that A SPANS U or that U is SPANNED BY A.

5.3.6. Remark. There are many instances in mathematics when we have a set A of elements belonging to some sort of object and we say that "A generates a doohickey of type X". Alternatively, we speak of, "the smallest doohickey of type X which contains A". There are frequently two ways of defining the appropriate doohickey—a *constructive* way and, for the lack of a better term, an *abstract* way. The constructive way gives you a more or less explicit recipe for constructing the desired doohickey from the elements of A. The abstract approach observes that the intersection of a family of doohickeys of type X is again a doohickey of type X, and so the smallest type X

doohickey containing A is the intersection of *all* the type X doohickeys which contain A. It turns out to be *very* convenient to have both of these approaches available. But, unfortunately, it is not always possible. Whenever we do have both approaches to the "doohickey generated by A", it is important to prove that they are equivalent, that is, to show that the two approaches generate exactly the same thing. The next exercise 5.3.7) gives a concrete example of how this works in practice.

5.3.7. Exercise. Show that the assertion made in the preceding definition makes sense and is correct. That is, show first that there *is* a "smallest" subspace containing a given nonempty set A of vectors. (*Hint.* Try the intersection of the family of all subspaces containing A.) Then show that the "constructive" definition of the *span* of A (the set of all linear combinations of elements of A) is the same as the "abstract" characterization (the smallest subspace containing A).

5.3.8. Remark. Occasionally one must consider the trivial question of what is meant by the *span* of the empty set. According to the "abstract" definition above it is the intersection of all the subspaces which contain the empty set. That is, $\operatorname{span}\emptyset = \{\mathbf{0}\}$. (Had we adopted the "constructive" definition, the set of all linear combinations of elements in $\emptyset$, the span of the empty set would have been just $\emptyset$ itself.)

5.3.9. Example. For each $n = 0, 1, 2, \dots$ define a function p_n on $\mathbb{R}$ by $\mathbf{p}_n(x) = x^n$. Let $\mathcal{P}$ be the set of polynomial functions on $\mathbb{R}$. It is a subspace of the vector space of continuous functions on $\mathbb{R}$. Then $\mathcal{P} = \operatorname{span}\{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2 \dots\}$. The sine function, whose value at x is $\sin x$, is not in the span of the set $\{p_0, p_1, p_2 \dots\}$.

5.3.10. Definition. A subset A (finite or not) of a vector space is LINEARLY DEPENDENT if the zero vector $\mathbf{0}$ can be written as a nontrivial linear combination of elements of A; that is, if there exist vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in A$ and scalars $\alpha_1, \dots, \alpha_n$, **not all zero,** such that $\sum_{k=1}^{n} \alpha_k \mathbf{x}_k = \mathbf{0}$. A subset of a vector space is LINEARLY INDEPENDENT if it is not linearly dependent.

Technically, it is a *set* of vectors that is linearly dependent or independent. Nevertheless, these terms are frequently used as if they were properties of the vectors themselves. For instance, if $S = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is a finite set of vectors in a vector space, you may see the assertions "the set S is linearly independent" and "the vectors $\mathbf{x}_1, \dots \mathbf{x}_n$ are linearly independent" used interchangeably.

5.3.11. Example. For each $n = 0, 1, 2, \dots$ define a function p_n on $\mathbb{R}$ by $p_n(x) = x^n$. Then the set $\{p_0, p_1, p_2, \dots\}$ is a linearly independent subset of the vector space of continuous functions on $\mathbb{R}$.

5.3.12. Definition. A set B of vectors in a vector space V is a (HAMEL) BASIS for V if it is linearly independent and spans V. For Hilbert spaces we will encounter another type of basis, the *orthonormal basis* (see definition 28.1.14). For such spaces it must be made clear whether we are talking about a Hamel basis or an orthonormal basis. In the context of vector spaces, however, only Hamel bases make sense; so the modifier "Hamel" is usually omitted.

In a vector space V subsets of linearly independent sets are themselves linearly independent; and supersets of spanning sets are spanning sets. In order for a linearly independent set to be a basis for V it must be "large enough" to span V; and in order for a spanning set to be a basis for V it must be "small enough" to be linearly independent. Thus in order for a set B to be a basis for V it must be "just the right size". We formalize this loose language in propositions 5.3.13 and 5.3.17.

5.3.13. Proposition. *Let V be a nonzero vector space and B be a nonempty subset of V. Then the following assertions are equivalent.*

(a) *B is a maximal linearly independent subset of V.*
(b) *B is a minimal spanning set for V.*
(c) *B is a basis for V.*

5.3.14. Proposition. *Every vector space has a basis.*

Proof. First, notice that the empty set is a basis for the zero vector space $\{\mathbf{0}\}$. Reason: $\emptyset$ is linearly independent and its span is $\{\mathbf{0}\}$ (see remark 5.3.8).

If V is a nonzero vector space, let $\mathfrak{L}$ be the set of all linearly independent subsets of V. Partially order this set by inclusion. The union of any chain in $\mathfrak{L}$ is itself linearly independent (see the next exercise) and is therefore an upper bound in $\mathfrak{L}$ for the chain. Thus according to *Zorn's lemma* $\mathfrak{L}$ has a maximal element, and by proposition 5.3.13 a maximal linearly independent subset of V is a basis for V. □

5.3.15. Exercise. Verify the assertion made in the preceding proof that the union of any chain of linearly independent subsets of a vector space is itself linearly independent.

5.3.16. Exercise. Modify the proof of proposition 5.3.14 to prove that if A is a nonempty linearly independent subset of a vector space V, then there exists a basis for V which contains A.

Hamel bases are not of great utility if they are not finite. As a consequence many linear algebra texts prove only that if one basis for a vector space is finite then so is any other basis and both will have the same cardinal number. This turns out to be true whether or not the space has a finite basis.

5.3.17. Proposition. *Any two bases for a vector space are cardinally equivalent.*

The proof of this general result is slightly technical. If you are interested, proofs can be found in [35] (theorem 1.12), [23] (theorem 4.58), [9] (page 11, theorem 2.12), or [24] (page 185, theorem 2.7).

5.3.18. Definition. If a vector space V has a finite basis B, then V is said to be FINITE DIMENSIONAL and the cardinality of B is the DIMENSION of V. If B is infinite, then V is an INFINITE DIMENSIONAL vector space. Depending on the cardinality of B we may distinguish between *countably infinite* and *uncountably infinite dimensional* spaces.

5.3.19. Example. The so-called standard basis vectors $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ (see 4.5.8) constitute a (Hamel) basis for the vector space $\mathbb{R}^3$. So $\mathbb{R}^3$ is a 3-dimensional vector space.

5.3.20. Example. The set $\{p_0, p_1, p_2, \dots\}$ of polynomials defined in example 5.3.11 is a basis for the vector space $\mathcal{P}$ of polynomial functions on $\mathbb{R}$. So $\mathcal{P}$ is a (countably) infinite dimensional vector space.

5.3.21. Proposition. *Every proper ideal in a unital algebra A is contained in a maximal ideal. Thus, in particular, if $A \neq 0$, then* $\operatorname{Max} A$ *is not empty.*

5.4. Lattices

5.4.1. Definition. A partially ordered set $(S, \leq)$ is a LATTICE if each pair of elements has both an infimum and a supremum in S.

5.4.2. Example. Any linearly ordered set is a lattice. If $a \leq b$, then $a \wedge b = a$ and $a \vee b = b$. In particular, any nonempty subset of $\mathbb{R}$ (under its usual ordering) is a lattice.

5.4.3. Example. We observed in example 5.2.7 that if S is a nonempty set, then $\mathfrak{P}(S)$ is an order complete lattice under the partial ordering of set inclusion $\subseteq$. The infimum of two members of $\mathfrak{P}(S)$ is their intersection; the supremum is their union.

5.4.4. Example. Under coordinatewise ordering (see example 5.1.5) $\mathbb{R}^2$ is a lattice. We have $(a, b) \wedge (c, d) = (a \wedge c, b \wedge d)$ and $(a, b) \vee (c, d) = (a \vee c, b \vee d)$ for all a, b, c, $d \in \mathbb{R}$.

5.4.5. Example. Let S be a nonempty set. Then the family $\mathcal{F}(S)$ of real valued functions on S under pointwise ordering (see example 5.1.8) is a lattice. For all f, $g \in \mathcal{F}(S)$ we have

$$f \wedge g = \tfrac{1}{2}[f + g - |f - g|\,]$$

and

$$f \vee g = \tfrac{1}{2}[f + g + |f - g|\,]\,.$$

Hint for proof. The verification of these identities is immediate from two (fairly obvious) properties of the real numbers (see 6.4.1).

We regard $\wedge$ and $\vee$ as binary operations on a lattice L. That is,

$$\wedge\colon L \times L \to L\colon (a,b) \mapsto a \wedge b$$

and

$$\vee\colon L \times L \to L\colon (a,b) \mapsto a \vee b\,.$$

Lattice theorists often call the operation $\wedge$ and $\vee$, *meet* and *join*, respectively.

5.4.6. Proposition. *In a lattice the operations $\wedge$ and $\vee$ are associative and commutative.*

5.4.7. Example. If m, $n \in \mathbb{N}$ we say that m DIVIDES n and write $m|n$ if there exists $p \in \mathbb{N}$ such that $pm = n$. The relation $|$ is a partial ordering on $\mathbb{N}$. Under this partial ordering $\mathbb{N}$ is a lattice. What are the standard names for $m \wedge n$ and $m \vee n$?

5.4.8. Proposition. *If a and b are elements of a lattice, then the following are equivalent:*

(i) $a \leq b$;
(ii) $a \wedge b = a$; *and*
(iii) $a \vee b = b$

5.4.9. Proposition. *If $a \leq b$ and $c \leq d$ in some lattice, then $a \vee c \leq b \vee d$ and $a \wedge c \leq b \wedge d$.*

5.4.10. Notation. If A and B are subsets of a lattice, then $A \vee B$ denotes $\{a \vee b\colon a \in A \text{ and } b \in B\}$; we write $A \vee b$ for $A \vee \{b\}$. The definitions of $A \wedge B$ and $A \wedge b$ are similar.

5.4.11. Proposition. *Let A and B be subsets of a lattice. If* $\sup A$ *and* $\sup B$ *exist, then* $\sup(A \vee B)$ *exists and*

$$\sup(A \vee B) = (\sup A) \vee (\sup B)\,.$$

Similarly, if $\inf A$ *and* $\inf B$ *exist, then* $\inf(A \wedge B)$ *exists and*

$$\inf(A \wedge B) = (\inf A) \wedge (\inf B)\,.$$

5.4.12. Proposition. *In a lattice every pair of elements a and b satisfies the* laws of absorption

$$a \wedge (a \vee b) = a \qquad \text{and} \qquad a \vee (a \wedge b) = a\,.$$

5.4.13. Definition. A lattice L is DISTRIBUTIVE if

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$

and

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)\,.$$

5.4.14. Example. If S is a nonempty set, then $\mathfrak{P}(S)$ (partially ordered by inclusion) is a distributive lattice.

Of course, not every lattice is distributive.

5.4.15. Example. On the set $S = \{l, a, b, c, m\}$ define a partial order $\leq$ so that $l < a < m$, $l < b < m$, $l < c < m$, and no two of the elements in the set $\{a, b, c\}$ are comparable. Then $(S, \leq)$ is a lattice but is not distributive. (*Proof:* $a \wedge (b \vee c) = a$ while $(a \wedge b) \vee (a \wedge c) = l$.)

5.5. Lattice Homomorphisms

5.5.1. Definition. A function $f\colon L \to M$ between lattices is a LATTICE HOMOMORPHISM if $f(x \vee y) = f(x) \vee f(y)$ and $f(x \wedge y) = f(x) \wedge f(y)$ for all $x,\, y \in L$.

5.5.2. Proposition. *Every lattice homomorphism $f\colon L \to M$ is order preserving.*

The converse of this proposition does not hold.

5.5.3. Example. Let $S = \{0, 1\}$ and L be the lattice $\mathfrak{P}(S)$ (see example 5.4.3). Define a partial ordering on the set $M = \{a, b\}$ by setting $a \leq b$. Clearly M is also a lattice. The function $f\colon L \to M$ defined by $\emptyset \mapsto a$, $\{0\} \mapsto a$, $\{1\} \mapsto a$, and $S \mapsto b$ is order preserving but is not a lattice homomorphism.

5.5.4. Definition. A point c in a set S is a FIXED POINT of a function $f\colon S \to S$ if $f(c) = c$.

5.5.5. Proposition (Knaster-Tarski fixed point theorem)**.** *Every order preserving map from an order complete lattice into itself has a fixed point.*

Hint for proof. Let L be an order complete lattice, $\phi\colon L \to L$ be order preserving, and s be the supremum of $\{x \in L \colon x \leq \phi(x)\}$. Show that $\phi(s) = s$.

5.5.6. Proposition (Banach Decomposition Theorem)**.** *If S and T are sets, $f\colon S \to T$, and $g\colon T \to S$, then there exist subsets A of S and B of T such that $f^{\rightarrow}(A) = B$ and $g^{\rightarrow}(B^c) = A^c$.*

Hint for proof. Recall that we have seen previously (in example 5.4.3) that partially ordered by $\subseteq$ the set $\mathfrak{P}(S)$ is an order complete lattice. Corresponding to the function $f\colon S \to T$ is the mapping

$$f^{\rightarrow}\colon \mathfrak{P}(S) \to \mathfrak{P}(T)\colon A \mapsto \{f(a) \colon a \in A\}$$

which takes each subset S to its image under f in T. We also define the mapping

$$\complement_S\colon \mathfrak{P}(S) \to \mathfrak{P}(S)\colon A \mapsto A^c$$

which takes each subset of S to its complement. The maps $g^{\rightarrow}$ and $\complement_T$ are defined similarly. Now consider the composite function ϕ defined by

$$\phi := \complement_S \circ g^{\rightarrow} \circ \complement_T \circ f^{\rightarrow}\,.$$

Use the preceding proposition.

5.5.7. Theorem (Cantor-Schröder-Bernstein Theorem)**.** *If S and T are sets and both $f\colon S \to T$ and $g\colon T \to S$ are injective, then there exists a bijection $h\colon S \to T$.*

5.5.8. Definition. We will now extend the notion of *cardinality* introduced in 3.1.3 to general sets. We will say that sets A and B *have the same cardinality*, and will write $\operatorname{card} A = \operatorname{card} B$, if there exists a bijection from A onto B. And we will say that *the cardinality of A is less than or equal to the cardinality of B*, and write $\operatorname{card} A \leq \operatorname{card} B$, if there exists an injective map from A into B. Theorem 5.5.7 provides us with the appropriate antisymmetry result.

5.5.9. Proposition. *Let A and B be sets. If* $\operatorname{card} A \leq \operatorname{card} B$ *and* $\operatorname{card} B \leq \operatorname{card} A$*, then* $\operatorname{card} A = \operatorname{card} B$*.*

Thus it is clear that on any given family of sets cardinality induces an equivalence relation.

5.6. Boolean Algebras

5.6.1. Definition. A function $f\colon S \to S$ from a set to itself is an INVOLUTION if it is its own inverse; that is if $f(f(s)) = s$ for all $s \in S$.

5.6.2. Definition. A BOOLEAN ALGEBRA is a nonempty distributive lattice L together with an involution $a \mapsto a'$ on L such that for all $a,\, b \in L$

(a) if $a \leq b$, then $b' \leq a'$, and

(b) $a \wedge a' \leq b \leq a \vee a'$.

The element a' is often called the COMPLEMENT of a. Notice that condition (b) says that (for any element a whatever) $a \wedge a'$ is the smallest element of a Boolean algebra; we denote it by $\mathbf{0}$. Similarly, $a \vee a'$ is the largest element of the algebra; it is denoted by $\mathbf{1}$. If in a Boolean algebra $\mathbf{0} = \mathbf{1}$, then the algebra consists of a single element. This is the TRIVIAL Boolean algebra. Any other Boolean algebra is NONTRIVIAL.

5.6.3. Proposition. *In a Boolean algebra the mapping $a \mapsto a'$ is both injective and surjective. The converse of condition (a) above holds in any Boolean algebra.*

5.6.4. Example. The set $\mathbb{Z}_2 = \{0, 1\}$ is the smallest nontrivial Boolean algebra. Here the ordering $0 \leq 1$ and complements $0' = 1$ are just what you would expect.

5.6.5. Example. In example 5.4.14 we saw that the power set $\mathfrak{P}(S)$ of a set S is a lattice. Under the involution $A \mapsto A^c$ which takes a subset of S to its complement (in S), the power set becomes a Boolean algebra. The element $\mathbf{0}$ is the empty set and the element $\mathbf{1}$ is S itself.

5.6.6. Example. Let S be the set of sentences which make up the so-called *propositional calculus*. For our purposes all we need to know (or assume) about the propositional calculus is that it is a set of sentences closed under the usual logical operations: $\wedge$ ("and"), $\vee$ ("or"), $\sim$ ("not"), $\Longrightarrow$ ("implies"), and $\Longleftrightarrow$ ("if and only if" or "logical equivalence"). A good place to find background material on the propositional calculus is the beautiful, slim, and readable book by Halmos and Givant, *Logic as Algebra* [21].

It is clear that $\Longrightarrow$ is a preordering on S but not a partial ordering. We remedy this situation by using a quotient construction. The operation $\Longleftrightarrow$ of logical equivalence is indeed an equivalence relation on the set S of sentences. The resulting equivalence classes, which partition S (see proposition 2.11.7), we will call PROPOSITIONS. As we saw in section 2.11 the set $P = S/\!\Longleftrightarrow$ of propositions (equivalence classes) is the *quotient* of S *modulo* the relation $\Longleftrightarrow$. As usual, we denote the proposition containing a sentence s by $[s]$. The relation $\Longrightarrow$ on S induces a relation, which we also denote by $\Longrightarrow$, on P in the obvious fashion: $[s] \Longrightarrow [t]$ whenever $s \Longrightarrow t$. This relation on P is clearly well-defined since it does not depend on the choice of representatives from the equivalence classes (see the discussion just before exercise 2.11.10); that is, if $s \Longrightarrow t$, if s' is logically equivalent to s, and if t' is logically equivalent to t, then $s' \Longrightarrow t'$. It is quite clear that the new relation $\Longrightarrow$ on P is not only a preordering but in fact a partial ordering. (If $[s] \Longrightarrow [t]$ and $[t] \Longrightarrow [s]$, then $s \Longrightarrow t$ and $t \Longrightarrow s$, so that $s \Longleftrightarrow t$. Since s and t are equivalent, $[s] = [t]$.)

In a similar fashion the other logical connectives $\sim$ ("not"), $\wedge$ ("and"), and $\vee$ ("or") may be unambiguously extended to the set P of propositions. Then, under the partial ordering $\Longrightarrow$, the proposition $p \wedge q$ really is the infimum of the propositions p and q, and $p \vee q$ really is the supremum of p and q. This tells us that $(P, \Longrightarrow)$ is a lattice. It is not difficult to check that it is in fact a distributive lattice. Furthermore, with the addition of $\sim$ ("not") as an involution on P, the set of propositions becomes a Boolean algebra. The $\mathbf{0}$ element of this Boolean algebra is the equivalence class of all self-contradictory sentences—sometimes denoted by $\mathbf{F}$ (for *false*); and the $\mathbf{1}$ element is the set of all tautologies—sometimes denote by $\mathbf{T}$ (for *true*).

Hint for proof. There are quite a few details to be verified. To avoid confusion in the crucial arguments that $p \wedge q$ really is the infimum of p and q and that $p \vee q$ really is the supremum of p and q be careful with notation: for example, use $p \wedge q$ to mean "p and q" and use $\inf\{p, q\}$ for the greatest lower bound of p and q. Since the point of the argument is to prove that these two are the same, it would not be a good idea initially to use the same notation for both.

The power set of a set and the propositional calculus are "good" examples of Boolean algebras in the sense that many familiar results from set theory and logic turn out to be true in general Boolean algebras. Here is the Boolean algebra version of *De Morgan's laws*.

5.6.7. Proposition. *Let S be a Boolean algebra with partial ordering $\leq$ and involution $a \mapsto a'$. Then*

(a) $(a \vee b)' = a' \wedge b'$

and, dually,

(b) $(a \wedge b)' = a' \vee b'$

for all a, $b \in S$.

Definition 5.6.2 is one standard approach to Boolean algebras. Another, equivalent, way is to specify that a Boolean algebra is a set S together with two binary operations $\wedge$ and $\vee$, which are associative and commutative, an involution $s \mapsto s^c$, and two distinguished elements $\mathbf{0}$ and $\mathbf{1}$ such that the following conditions are satisfied:

(a) $\wedge$ distributes over $\vee$, and $\vee$ distributes over $\wedge$,
(b) $s \vee (s \wedge t) = s \wedge (s \vee t) = s$ for all s, $t \in S$, and
(c) $s \wedge s^c = \mathbf{0}$ and $s \vee s^c = \mathbf{1}$ for all $s \in S$.

5.6.8. Exercise. Prove that the definition in the preceding paragraph is equivalent to the one given in definition 5.6.2.

A third approach to Boolean algebras is through Boolean rings (see definition 4.6.22). In fact, a Boolean algebra is nothing more than a unital Boolean ring. Do the next two exercises to see how this comes about.

5.6.9. Exercise. Let $(R, +, \cdot)$ be a unital Boolean ring with multiplicative identity $\mathbf{1}$. For all a, $b \in R$ let $a \leq b$ if $ab = a$ and let $a' = \mathbf{1} + a$. Then $\leq$ is a partial ordering on R and the map $a \mapsto a'$ is an involution on R. Equipped with this partial ordering and involution the ring R is a Boolean algebra.

5.6.10. Exercise. Let R be a Boolean algebra with partial ordering $\leq$ and involution $a \mapsto a'$. For all a, $b \in R$ define

$$a + b := (a \wedge b') \vee (a' \wedge b)$$

and

$$a \cdot b := a \wedge b\,.$$

Then $(R, +, \cdot)$ is a unital Boolean ring.

5.6.11. Definition. "Algebra" is a seriously overworked term. Classically an ALGEBRA (or a FIELD) of subsets of a set S is defined as a family $\mathfrak{A}$ of subsets of S which is closed under the operations of union and complementation and which contains S. The next proposition shows that this makes $\mathfrak{A}$ into a Boolean algebra—which helps explain why the term "algebra of sets" is used for $\mathfrak{A}$ even though it is not an algebra (or a field, for that matter) in the usual algebraic sense. Notice that (by a trivial inductive argument) if $A_1, \dots, A_n$ all belong to an algebra $\mathfrak{A}$ of sets, then so does $\bigcup_{k=1}^{n} A_k$.

5.6.12. Proposition. *Let $\mathfrak{A}$ be an algebra of sets; that is, a nonempty family of subsets of a set S such that*

(a) *$A \cup B \in \mathfrak{A}$ whenever A, $B \in \mathfrak{A}$; and*
(b) *$A^c \in \mathfrak{A}$ whenever $A \in \mathfrak{A}$.*

Then the family $\mathfrak{A}$ is closed under the binary operations of intersection, set subtraction, and symmetric difference; it contains $\emptyset$ and S; and it is a Boolean algebra under set inclusion (as a partial ordering) and complementation (as involution). Note also that $(\mathfrak{A}, \triangle, \cap)$ is a unital Boolean ring.

5.6.13. Proposition. *Let S be a set and Ξ be a family of algebras of subsets of S. Then $\bigcap \Xi$ is an algebra of subsets of S.*

5.6.14. Exercise. Let $\mathfrak{F}$ be a family of subsets of a nonempty set S. Explain in careful detail what we mean by "the smallest algebra of sets containing $\mathfrak{F}$." Prove that such a thing exists. (If you have trouble with this look back at remark 5.3.6.)

5.6.15. Example. Let S be an infinite set and $\mathfrak{F} = \{\{x\} : x \in S\}$. The smallest algebra of sets containing $\mathfrak{F}$ comprises all finite and cofinite subsets of S. (A subset A of a set S is COFINITE if its complement is finite.)

CHAPTER 6

THE REAL NUMBERS

6.1. Axioms Defining the Real Numbers

The real numbers are typically approached axiomatically. That is, the set $\mathbb{R}$ of real numbers is defined to be a set together with two operations and an ordering which satisfy a list of four conditions (also called *axioms*) we impose. (Naturally these conditions are not arbitrary inventions, but have a long and complicated history behind them.) This leads to three rather big questions:

(1) Is there anything out there that actually satisfies these axioms? In other words, are the axioms *consistent*?
(2) Why do we speak of *the* real numbers? How do we know that there are not two essentially different things that both satisfy the axioms?
(3) How do we get all the standard "facts" about the real numbers from the axioms?

We sketch an answer to the first question in section 6.2 by describing very briefly the process through which we can construct a *model* for the axioms, that is, we discuss the creation of a set theoretic object which satisfies the axioms. The second question we answer without proof: the axiom system for $\mathbb{R}$ is *categorical*; that is, there is essentially only one model for the axioms defining the reals. (Two models are "essentially" the same if they are isomorphic.) Finally, we will assume that the most elementary parts of an answer to the third question (the business we usually call *arithmetic*) have been adequately covered in earlier courses.

6.1.1. Definition. Let $\mathbb{R}$ be a set, let $+$ and $\cdot$ be binary operations on $\mathbb{R}$, and $\mathbb{P} \subseteq \mathbb{R}$. The quadruple $(R, +, \cdot, \mathbb{P})$ is the system of REAL NUMBERS if the axioms I–IV below are satisfied.

6.1.2. Axiom (I)**.** $(\mathbb{R}, +, \cdot)$ *is a field.*

6.1.3. Axiom (II)**.** $\mathbb{P}$ *is closed under both* $+$ *and* $\cdot$*.*

6.1.4. Axiom (III)**.** $0 \in \mathbb{P} \cap (-\mathbb{P})$ *and if* $x \neq 0$*, then* x *belongs to either* $\mathbb{P}$ *or* $-\mathbb{P}$*, but not both.*

The set $\mathbb{P}$ is the POSITIVE CONE in $\mathbb{R}$. Its members are the POSITIVE real numbers. It is worth emphasizing that in these notes, a number x is *positive* if $x \geq 0$. If we want $x > 0$, we say that x is STRICTLY POSITIVE. Similar remarks hold for *negative* and *strictly negative*. In axiom (III), as in 4.4.10, $-\mathbb{P}$ is the set of additive inverses of members of $\mathbb{P}$. In subsequent chapters of these notes we will use the more common notation $\mathbb{R}^+$ for $\mathbb{P}$. Axiom (III) is often called the *law of trichotomy*. It says that every real number x satisfies exactly one of three conditions: it is 0, or it is strictly positive ($x > 0$), or it is strictly negative ($x < 0$).

Define an ordering on $\mathbb{R}$ by

$$x \leq y \quad \text{if and only if} \quad y - x \in \mathbb{P}\,.$$

We call this the ordering on $\mathbb{R}$ *induced by* the positive cone $\mathbb{P}$.

6.1.5. Proposition. *The ordering* $\leq$ *defined above is a linear ordering on* $\mathbb{R}$*.*

6.1.6. Axiom (IV)**.** *Under the ordering* $\leq$ *defined above,* $\mathbb{R}$ *is Dedekind complete.*

As was mentioned above we will take for granted the most elementary consequences of these axioms (grade school arithmetic through college algebra). For an elegant synopsis of this material written for grownups, see chapter 1 of Stromberg's *An Introduction to Classical Real Analysis* [43]. Below are are few points that are perhaps worthy of review.

6.1.7. Proposition. *The partial ordering $\leq$ induced on $\mathbb{R}$ by its positive cone respects the algebraic operations on $\mathbb{R}$ in the sense that*

$$x + z \leq y + z \quad \text{whenever} \quad x \leq y$$

and

$$xz \leq yz \quad \text{whenever} \quad x \leq y \text{ and } 0 \leq z\,.$$

6.1.8. Proposition. *Let A be a nonempty subset of $\mathbb{R}$. If A has a greatest lower bound, then $-A$ has a least upper bound and*

$$\sup(-A) = -\inf A\,.$$

If you wish you may skip the proof of this result for the moment. We will see in proposition 9.1.7 that it holds more generally in all ordered vector spaces. The next proposition also holds in an arbitrary ordered vector space (see proposition 9.1.5).

6.1.9. Proposition. *Let A and B be nonempty subsets of $\mathbb{R}$. If A and B both have least upper bounds, then so does $A + B$ and*

$$\sup(A + B) = \sup A + \sup B\,.$$

6.1.10. Proposition. *Let A and B be nonempty subsets of $\mathbb{P}$. If A and B both have least upper bounds, then so does $AB = \{ab \colon a \in A \text{ and } b \in B\}$ and*

$$\sup(AB) = (\sup A)(\sup B)\,.$$

6.1.11. Exercise. State and prove the *principle of mathematical induction.*

Hint for proof. Here is a little background that you may find useful. First, make a definition: A collection J of real numbers is INDUCTIVE if

(i) $1 \in J$, and
(ii) $x + 1 \in J$ whenever $x \in J$.

Then prove that the intersection of a family of inductive subsets of $\mathbb{R}$ is inductive. Use this to show that there exists a smallest inductive set. Call this the set of *natural numbers* and denote it by $\mathbb{N}$. Now here is an elegant statement of the *principle of mathematical induction*: Every inductive subset of $\mathbb{N}$ equals $\mathbb{N}$. (Is the proof of this obvious?) Use this to prove the statement of *mathematical induction* with which you are familiar.

6.1.12. Proposition (The Archimedean Property of $\mathbb{R}$)**.** *If $a \in \mathbb{R}$ and $\epsilon > 0$, then there exists $n \in \mathbb{N}$ such that $n\epsilon > a$.*

The following important fact is an easy consequence of the *Archimedean property* of $\mathbb{R}$.

6.1.13. Corollary. *For every pair a and b of real numbers with $a < b$ there exists a rational number r such that $a < r < b$.*

6.2. Construction of the Real Numbers

An important question to ask whenever we write down a definition (or, put differently, a set of axioms) for some mathematical object is whether any such object actually exists. In other words, *is the set of axioms consistent?* This turns out in general to be too difficult to answer. The best we can ordinarily do is to show that a set of axioms (a definition) is *relatively consistent.* That is, we try to show that if some other bit of mathematics (usually set theory or category theory) is consistent, then so is the bit we are interested in. Thus confronted, for example, with the axioms defining the real numbers, one may well ask, "But is there anything that satisfies these axioms?" That is, "Do these axioms have a *model*? The answer in this case is a qualified *yes*; the axioms for the real numbers are relatively consistent—if we believe (or are willing to assume) that some set of axioms (Zermelo-Fraenkel, for instance) governing sets is consistent, then we can construct a set-theoretic model for the real numbers. For more information on the axioms for Zermelo-Fraenkel set theory consult any text on axiomatic set theory or visit a good web site (for example, Mathworld [45] or Wikipedia [46]) and search for something like "Zermelo-Fraenkel axioms."

The construction of the real number system from sets is a fairly intricate (and, for some tastes, a rather tedious) business. In some universities a one-term or one-semester course is devoted to doing just that. Here we will simply outline a few of the major steps in the program, paying close attention to only one of them—the so-called *Grothendieck construction*, which produces an Abelian group from a commutative semigroup. The reason for singling out this one tool is its amazing usefulness in many fields of more advanced mathematics.

First Steps. The process may be started by defining the natural numbers. We take the natural number 1 to be $\{\emptyset\}$. [1] For any natural number n already defined we define its SUCCESSOR $S(n)$ to be $\{n\}$. We amalgamate all the objects 1, $S(1)$, $S(S(1))$, $S(S(S(1)))$, and so on, into a set $\mathbb{N}$. (Naturally, this intuitive amalgamation process requires for its justification an explicit set-theoretic axiom, usually called the *axiom of infinity*. For details on this consult [45] or [46].) Then we give these objects their conventional names: $2 := S(1)$, $3 := S(S(1)) = S(2)$, $4 := S(S(S(1))) = S(3)$, and so on. Thus we have the familiar set $\mathbb{N} = \{1, 2, 3, 4, \dots\}$ of natural numbers. The successor function $n \mapsto S(n)$ on this set gives us an obvious way of ordering this set: $1 < 2$, $2 < 3$, ... One verifies that this creates a linear ordering on $\mathbb{N}$. One shows, in fact, that this is a well-ordering of $\mathbb{N}$. (An ordered set is WELL-ORDERED if every nonempty subset of the set las a least element.)

The next step is to create an arithmetic for the natural numbers by defining addition and multiplication on $\mathbb{N}$. This is done inductively. For every m define

$$m + 1 := S(m)$$

and, supposing that $m + n$ has already been defined, let

$$m + S(n) := S(m + n)\,.$$

One now verifies that under this operation $\mathbb{N}$ is a commutative semigroup having the cancellation property. This is the point where we make our first use of the *Grothendieck construction*, which we now examine in detail.

The Grothendieck Construction. Throughout this subsection $(S, +)$ is a commutative semigroup. On the Cartesian product $S \times S$ define a relation $\sim$ by declaring $(a, b) \sim (c, d)$ if an only if there exists $z \in S$ such that

$$a + d + z = b + c + z\,.$$

6.2.1. Proposition. *The relation $\sim$ is an equivalence relation on $S \times S$.*

Our notation for the equivalence class containing the pair (a, b) would normally be $[(a, b)]$. For economy we will shorten it here to $\langle a, b\rangle$. Let $G(S)$ be $(S \times S)/\sim$, the quotient of $S \times S$ by $\sim$. On $G(S)$ define a binary operation (also denoted by $+$) by:

$$\langle a, b\rangle + \langle c, d\rangle := \langle a + c, b + d\rangle\,.$$

6.2.2. Proposition. *The operation $+$ defined above is well-defined.*

6.2.3. Proposition. *Under the operation $+$ defined above $G(S)$ is an Abelian group.*

The resulting Abelian group is called the GROTHENDIECK GROUP generated by S.

6.2.4. Definition. If $\phi\colon S \to G$ is an injective homomorphism from a semigroup into a group (so that S is isomorphic to its image under ϕ), say that ϕ is an (ALGEBRAIC) EMBEDDING and that S is EMBEDDED in G.

Now for an arbitrary $a \in S$ define a mapping

$$\phi\colon S \to G(S)\colon s \mapsto \langle a, a + s\rangle\,.$$

[1]Of course, we could start with $0 := \emptyset$, but in these notes 0 belongs to $\mathbb{Z}^+$ (the set of positive integers) but not to $\mathbb{N}$ (the set of natural numbers).

6.2.5. Proposition. *The mapping ϕ above is well-defined. That is, the definition is independent of the choice of a.*

6.2.6. Proposition. *The mapping ϕ is a semigroup homomorphism.*

This homomorphism is the GROTHENDIECK MAP associated with S.

6.2.7. Proposition. *The Grothendieck mapping associated with a semigroup S is an embedding if and only if S has the cancellation property.*

6.2.8. Exercise. Take S to be $\mathbb{N}$. Explain why it is reasonable to regard the resulting Abelian group as the group $\mathbb{Z}$ of integers. Explain the process for assigning old names (like 1, 2, 3, ...) to the new objects living in $G(\mathbb{N})$. Then identify the objects in $G(\mathbb{N})$ to which we assign new names (like 0, -1, -2, ...). Also explain how we can extend the ordering on $\mathbb{N}$ to a linear ordering on $\mathbb{Z}$.

6.2.1. The Rational Numbers. The point of the last subsection was to demonstrate how we can construct the Abelian group of integers from the commutative semigroup of natural numbers. Before proceeding to create the rational numbers it is convenient to define multiplication on the integers in such a way that they become a ring. One can use an inductive technique virtually identical to the one in the last paragraph of subsection 6.2 to define multiplication on $\mathbb{N}$ and then extend the operation to all of $\mathbb{Z}$. Of course, it must be proved that these rather obvious definitions do indeed create a ring structure on $\mathbb{Z}$.

Now to produce the rational numbers we apply the Grothendieck technique to the commutative semigroup $\mathbb{Z}$ *under multiplication*—with one slight change. Instead of working with $\mathbb{Z}\times\mathbb{Z}$ we work with $\mathbb{Z}\times\mathbb{Z}_0$ where $\mathbb{Z}_0$ is the set of *nonzero* integers. As before say that two ordered pairs in $\mathbb{Z}\times\mathbb{Z}_0$ are equivalent, $(m,n)\sim(p,q)$, if there exists $z\in\mathbb{Z}$ such that $mqz = npz$. The $\sim$ turns out to be an equivalence relation on $\mathbb{Z}\times\mathbb{Z}_0$. Let $\mathbb{Q}$ be the family $(\mathbb{Z}\times\mathbb{Z}_0)/\sim$ of equivalence classes. We adopt familiar notation by designating the equivalence class containing the pair (m,n) by $\frac{m}{n}$. Define multiplication on $\mathbb{Q}$ by

$$\frac{m}{n}\cdot\frac{p}{q} := \frac{mp}{nq}$$

and addition by

$$\frac{m}{n}+\frac{p}{q} := \frac{mq+np}{nq}\,.$$

Also define a map $\psi\colon\mathbb{Z}\to\mathbb{Q}\colon n\mapsto\frac{n}{1}$.

6.2.9. Exercise. Explain how the presence of an identity element in the multiplicative semigroup $\mathbb{Z}$ allows us to simplify the definition of the equivalence relation $\sim$ above.

6.2.10. Proposition. *Addition and multiplication are well-defined on $\mathbb{Q}$, and under these operations $\mathbb{Q}$ is a field.*

6.2.11. Proposition. *The mapping $\psi\colon\mathbb{Z}\to\mathbb{Q}$ defined above is an embedding of the commutative (multiplicative) semigroup $\mathbb{Z}$ into the Abelian (multiplicative) group $\mathbb{Q}$. The mapping ψ also preserves addition; that is, $\psi(m+n)=\psi(m)+\psi(n)$ for all m, $n\in\mathbb{Z}$.*

6.2.12. Exercise. Explain how to extend the linear ordering on $\mathbb{Z}$ to a linear ordering on $\mathbb{Q}$.

6.2.2. From $\mathbb{Q}$ to $\mathbb{R}$. Once we have the ordered field $\mathbb{Q}$, it is just one more step to the reals. The central idea here is that $\mathbb{Q}$ becomes a metric space when we define the *distance* between to rational numbers r and s to be $|r-s|$, but this metric space is not complete. It turns out to be possible to complete an arbitrary metric space, and the completion of the rationals is the set $\mathbb{R}$ of real numbers. (We will define and discuss metric spaces in section 10.4. Complete spaces and the process for completing a space appear in section 20.2. Of particular interest in the present context is proposition 20.2.4.)

Having produced the set $\mathbb{R}$ we next extend the operations of addition and multiplication and the linear ordering relation from $\mathbb{Q}$ to $\mathbb{R}$. The final, but crucial, thing that must be proved is that the object we have so laboriously constructed really does satisfy the axioms for the real numbers given in section 6.1.

6.3. Elementary Functions

In analysis the next logical step after developing the real number system is to derive in detail the most important properties of the elementary functions: the algebraic (including polynomial, rational, and power) functions, the trigonometric functions, the exponential functions, the inverses of all these (including radical, inverse trigonometric, and logarithmic functions), and all possible composites of the preceding. Pursuing this, I feel, would be far too lengthy a diversion from the principal subject material of these notes. So the standard properties of the elementary functions will, for the most part, simply be assumed. Here is a (very short) list indicating the sort of thing you may expect to see used without proof in the sequel.

(a) The n^{th}-root function $x \mapsto x^{1/n}$ is strictly increasing on the interval $(0, \infty)$ for every $n \in \mathbb{N}$.
(b) The arctangent function is strictly increasing on $\mathbb{R}$. Its derivative is $(1+x^2)^{-1}$ for every x. It is concave (down) on the interval $[0, \infty)$.
(c) The exponential function $x \mapsto e^x$ is strictly increasing and convex (up) on $\mathbb{R}$.
(d) If ln is the natural logarithm function on $(0, \infty)$ and x, $y > 0$, then $\ln xy = \ln x + \ln y$.
(e) $\arcsin \frac{1}{2} = \pi/6$.

What should you do if you are nervous about agreeing to some (or all) of the assumptions made in these notes concerning real numbers and elementary functions? My advice is to consult Karl Stromberg's amazing text, *An Introduction to Classical Real Analysis* [43]. It is a beautifully written, nearly flawless, essentially self-contained introduction to, among other topics, the real (and complex) number systems and the elementary functions.

6.4. Absolute Value

There are a few facts about the absolute value function which are used so frequently that it seems a good idea to write them down explicitly. In section 6.3 we avoided a discussion of the elementary functions by agreeing to assume knowledge of their standard properties. If, in particular, we are familiar with the square root function $\sqrt{\ } : [0, \infty) \to \mathbb{R}$, then it makes sense to define $|x|$ as $\sqrt{x^2}$ for each real number x. It is clear that if $x \geq 0$, then $|x| = x$; and if $x < 0$, then $|x| = -x$. The next exercise is quite simple to prove and parts of it crop up so often that they will usually be used without explicit reference.

6.4.1. Exercise. If a, $b \in \mathbb{R}$, then

(a) $|ab| = |a|\,|b|$;
(b) $|a + b| \leq |a| + |b|$; (the *triangle inequality*)
(c) $\big|\,|a| - |b|\,\big| \leq |a - b|$; and
(d) $|ab| \leq \frac{1}{2}(a^2 + b^2)$.

Hint for proof of (d). Consider the squares of $a - b$ and $a + b$.

It is interesting and somewhat surprising that the "algebraic" function *absolute value* and the "order" operations *sup* and *inf* are in a sense interchangeable. If we have one, we get the other. Much use is made of these relationships in the setting of Riesz spaces (beginning in chapter 14).

6.4.2. Proposition. *If x is a real number,then*

$$|x| = x \vee (-x) = -(x \wedge (-x))\,. \tag{6.1}$$

It is also true that if $x \in \mathbb{R}$, then

$$x \vee y = \tfrac{1}{2}(x + y + |x - y|\,) \tag{6.2}$$

and

$$x \wedge y = \tfrac{1}{2}(x + y - |x - y|\,)\,. \tag{6.3}$$

6.5. Some Useful Inequalities

Inequalities involving real numbers are often essential ingredients of proofs and derivations in nearly every field of mathematics. In this section we give a (very) few of the most useful. The first is known as the *Schwarz inequality* (or as the *Cauchy inequality*, or as the *Bunyakovskii inequality*, or as any combination of the preceding).

6.5.1. Proposition (Schwarz's Inequality)**.** *Let* $u_1, \dots, u_n, v_1, \dots, v_n \in \mathbb{R}$. *Then*

$$\left(\sum_{k=1}^{n} u_k v_k\right)^2 \le \left(\sum_{k=1}^{n} {u_k}^2\right)\left(\sum_{k=1}^{n} {v_k}^2\right).$$

Proof. To simplify notation make some abbreviations: let $a = \sum_{k=1}^{n} {u_k}^2$, $b = \sum_{k=1}^{n} {v_k}^2$, and $c = \sum_{k=1}^{n} u_k v_k$. Then

$$\begin{aligned} 0 &\le \sum_{k=1}^{n} \left(\sqrt{b}\, u_k - \frac{c}{\sqrt{b}}\, v_k\right)^2 \\ &= ab - 2c^2 + c^2 \\ &= ab - c^2 . \end{aligned}$$

□

A consequence of *Schwarz's inequality* is the *Minkowski inequality.*

6.5.2. Proposition (Minkowski's Inequality)**.** *Let* $u_1, \dots, u_n, v_1, \dots, v_n \in \mathbb{R}$. *Then*

$$\left(\sum_{k=1}^{n} (u_k + v_k)^2\right)^{\frac{1}{2}} \le \left(\sum_{k=1}^{n} {u_k}^2\right)^{\frac{1}{2}} + \left(\sum_{k=1}^{n} {v_k}^2\right)^{\frac{1}{2}}.$$

Proof. Let a, b, and c be as in proposition 6.5.1. Then

$$\begin{aligned} \sum_{k=1}^{n} (u_k + v_k)^2 &= a + 2c + b \\ &\le a + 2|c| + b \\ &\le a + 2\sqrt{ab} + b \qquad \text{(by 6.5.1)} \\ &= \left(\sqrt{a} + \sqrt{b}\right)^2 . \end{aligned}$$

□

Before we state a version of *Bernoulli's inequality* 6.5.5 we develop some background.

6.5.3. Definition. If n is a positive integer we define $n!$ (read "n factorial) inductively by $0\,! := 1$ and $n! := n(n-1)!$ whenever $n \ge 1$. If k and n are positive integers and $k \le n$, then the binomial coefficient $\binom{n}{k}$ is defined by

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

The coefficient $\binom{n}{k}$ is read "*n by k*" or "*n choose k*".

6.5.4. Lemma. *If k and n are positive integers and $1 \le k \le n$, then*

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

6.5.5. Proposition (Binomial Theorem)**.** *If a and b are real numbers and n is a positive integer, then*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k .$$

6.5.6. Exercise. Find the sum of the binomial coefficients $\binom{n}{k}$ from $k = 0$ to $k = n$.
Hint. $1 + 1 = 2$.

6.5.7. Proposition (Bernoulli's Inequality)**.** *Let $x \in \mathbb{R}$ and $n \in \mathbb{N}$. If $x > 0$ and $n > 1$, then*

$$(1 + x)^n > 1 + nx\,.$$

6.5.8. Exercise. Use the *Schwarz inequality* to prove part (a) and part (a) to prove part (b).

(a) If a_1, a_2, ..., a_n are strictly positive real numbers, then

$$\left(\sum_{k=1}^{n} a_k\right)\left(\sum_{k=1}^{n} \frac{1}{a_k}\right) \geq n^2.$$

(b) If a, b, and c are strictly positive real numbers such that $a + b + c = 1$, then

$$\left(\frac{1}{a} - 1\right)\left(\frac{1}{b} - 1\right)\left(\frac{1}{c} - 1\right) \geq 8.$$

6.6. Complex Numbers

Later in these notes we will encounter spaces of complex valued functions. The complex numbers are numbers of the form $a + bi$ where a and b are real numbers and i is a complex number whose square is -1. The arithmetic of the complex numbers is easy. Addition is defined in the obvious way

$$(a + bi) + (c + di) := (a + c) + (b + d)i$$

and so is multiplication

$$(a + bi)(c + di) := (ac - bd) + (ad + bc)i\,.$$

It is routine to verify that under these operations the set $\mathbb{C} = \{a + bi \colon a, b \in \mathbb{R}\}$ is a field.

Let $z = a + bi$ be a complex number (with a, $b \in \mathbb{R}$). Then a is the REAL PART of z and b is its IMAGINARY PART. The ABSOLUTE VALUE of z is $\sqrt{a^2 + b^2}$. The DISTANCE between two complex numbers $z = a + bi$ and $w = c + di$ is $|z - w|$. Another routine argument shows that this defines a metric (see section 10.4) on $\mathbb{C}$. The map $(a, b) \mapsto a + bi$ is an isometry (distance preserving mapping) from $\mathbb{R}^2$ (with its usual Euclidean distance onto the complex plane $\mathbb{C}$. Thus the complex plane $\mathbb{C}$ can be viewed as just the usual Euclidean plane $\mathbb{R}^2$ with additional arithmetic operations making it into a field.

The only problem with the foregoing discussion is the second sentence of this section where the "complex numbers" are defined in terms of a particular "complex number" i. It is certainly reasonable to be deeply suspicious of this circularity. What would lead us to believe that there is any object whatever whose square is -1? There is certainly no *real* number with this property. If you don't already know the answer to this question and are interested in finding out what sort of "number" has square -1, read definition 6.6.1 below. Otherwise, it can be skipped.

6.6.1. Definition. A COMPLEX NUMBER is an ordered pair of real numbers, that is, an element of $\mathbb{R}^2$. We define *addition* of complex numbers pointwise

$$(a, b) + (c, d) := (a + c, b + d)$$

and *multiplication* of complex numbers by

$$(a, b)(c, d) := (ac - bd, ad + bc)\,.$$

Notice that the map $\phi \colon \mathbb{R} \to \mathbb{C} \colon a \mapsto (a, 0)$ is an injective homomorphism. The fields $\mathbb{R}$ and $\operatorname{ran} \phi$ are isomorphic. We will use this isomorphism to identify the field of real numbers with the subfield of complex numbers of the form $(a, 0)$. In this way every real number becomes a (special kind of) complex number. We will abuse notation and write $a = (a, 0)$ whenever it seems convenient.

Next we will define $i := (0,1)$. Notice that since $i^2 = (0,1)(0,1) = (-1,0) = -1$, we have succeeded in finding a complex number whose square is -1. Notice also that if a and b are real numbers, then

$$\begin{aligned} a + bi &= (a,0) + (b,0)(0,1) \\ &= (a,0) + (0,b) \\ &= (a,b)\,. \end{aligned}$$

Thus $a + bi$ is just another way of writing the complex number (a,b). Now the first paragraph of this section should make sense.

6.6.2. Exercise. Show that the field $\mathbb{C}$ of complex numbers cannot be linearly ordered. That is, show that there is no subset $\mathbb{P}$ of $\mathbb{C}$ such that

(i) the sets $\mathbb{P}$, $-\mathbb{P}$, and $\{0\}$ are pairwise disjoint and their union is all of $\mathbb{C}$; and
(ii) $\mathbb{P}$ is closed under addition and multiplication.

CHAPTER 7

SEQUENCES AND INDEXED FAMILIES

There are many, many ways of thinking about functions. And sometimes these ways of thinking about functions take on a life of their own. We may find ourselves working with objects that are in fact functions without being conscious of the fact. Consider as an example a function $x\colon \mathbb{N}_4 \to \mathbb{R}$. To completely specify the function we need to know its values at each point in its domain $\{1, 2, 3, 4\}$. So suppose we know the values $x(1)$, $x(2)$, $x(3)$, and $x(4)$. There is no harm in rewriting these numbers as x_1, x_2, x_3, and x_4. Since these values completely specify the function x it is common practice to say that the function *is* the list of these values. That is, $x = (x_1, x_2, x_3, x_4)$. Thus our function has become an ordered quadruple of real numbers or, we may perhaps say, a point in Euclidean 4-space. If, however, we arrange these four values to look like $\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$ our function has become a 2×2 matrix of real numbers. So in one sense a 2×2 matrix of real numbers is the same thing as a quadruple of real numbers: each is just a function from $\mathbb{N}_4$ into $\mathbb{R}$. On the other hand they are normally regarded as quite different objects. Imagine how perplexed someone would be if you asked them to find the determinant of some ordered quadruple, or to calculate its trace. It is a mark, I think, of mathematical maturity to feel comfortable with mathematical objects no matter what disguise they choose to wear to a particular party.

7.1. Sequences

Everyone learns in elementary calculus (or before) to distinguish between the ordered pair $(1, 2)$ and the set $\{1, 2\}$. The sets $\{1, 2\}$ and $\{2, 1\}$ are equal but the ordered pairs $(1, 2)$ and $(2, 1)$ are not. The ordered pair $(3, 3)$ has two coordinates while the set $\{3, 3\}$ has exactly one member. The same distinction must be made between a countable set $\{x_1, x_2, x_3, \dots\}$ and the sequence $(x_1, x_2, x_3, \dots)$. The first is a set and the second is a function. It is perverse, I think, in some elementary texts to carefully distinguish notationally between ordered pairs and sets (parentheses for pairs, braces for sets) and then use set notation $\{x_1, x_2, \dots\}$ for sequences.

7.1.1. Definition. A SEQUENCE is a function whose domain is the set $\mathbb{N}$ of natural numbers. If S is a set then a function $a\colon \mathbb{N} \to S$ is called a sequence of elements of S. If a is a sequence, it is conventional to write its value at a natural number n as a_n rather than as $a(n)$. The sequence itself may be written in a number of ways:

$$a = (a_n)_{n=1}^{\infty} = (a_n) = (a_1, a_2, a_3, \dots).$$

The element a_n in the range of a sequence is the n^{th} TERM of the sequence.

It is important to distinguish in one's thinking between a sequence and its range, which is a countable set. Think of a sequence $(x_1, x_2, x_3, \dots)$ as an *ordered* set: there is a first element x_1, and a second element x_2, and so on. The range $\{x_1, x_2, x_3, \dots\}$ is just a set. There is no "first" element. For example, the sequences $(1, 2, 3, 4, 5, 6, \dots)$ and $(2, 1, 4, 3, 6, 5, \dots)$ are different whereas the sets $\{1, 2, 3, 4, 5, 6, \dots\}$ and $\{2, 1, 4, 3, 6, 5, \dots\}$ are exactly the same (both are $\mathbb{N}$). The sequence $(5, 5, 5, 5, \dots)$ has infinitely many terms; the set $\{5, 5, 5, 5, \dots\}$ has exactly one member.

Nevertheless, some terminology is carried over from the range of a sequence to the sequence itself. In particular, a sequence of real numbers is BOUNDED ABOVE if its range is; it is BOUNDED BELOW if its range is; and it is BOUNDED if its range is.

7.1.2. Remark. Occasionally it is convenient to alter the preceding definition a bit to allow the domain of a sequence to be some other set such as, for example, $\mathbb{N} \cup \{0\}$ the set of all positive integers.

Studying convergence (that is, the "long term" behavior) of sequences and nets is one of the most important topics in analysis. The words "eventually" and "frequently" are introduced to reduce the number of quantifiers which appear in discussions of convergence.

7.1.3. Definition. Let (x_n) be a sequence of elements of a set S and P be some property that members of S may possess. We say that the sequence (x_n) EVENTUALLY has property P if there exists $n_0 \in \mathbb{N}$ such that x_n has property P for every $n \geq n_0$. (Another way to say the same thing: x_n has property P for all but finitely many n.) For example, if A is a subset of S, then the sequence (x_n) is eventually in A if there is an index n_0 such that $x_n \in A$ whenever $n \geq n_0$. Sometimes instead of saying that a sequence (x_n) *eventually* has property P we say that (x_n) has property P *for sufficiently large* n or, more briefly, that (x_n) has property P *for large* n.

7.1.4. Example. Let S be a set, (A_n) be a sequence of subsets of S, and c an element of S. Now (A_n) is a sequence in the set $\mathfrak{P}(S)$ and the property in question is *containing the point* c. So we may say that the sequence (A_n) eventually contains c (or, equivalently, that the point c is EVENTUALLY IN (A_n); or that $c \in A_n$ for all but finitely many n) if there exists $k \in \mathbb{N}$ such that for all $n \geq k$ we have $c \in A_n$.

7.1.5. Example. The sequence (n^{-2}) of real numbers is eventually less than 0.001.

7.1.6. Definition. Let (x_n) be a sequence of elements of a set S and P be some property that members of S may possess. We say that the sequence (x_n) FREQUENTLY has property P if for every $k \in \mathbb{N}$ there exists $n \geq k$ such that x_n has property P. (An equivalent formulation: x_n has property P for infinitely many n.) For example, if $A \subseteq S$, we say that the sequence (x_n) is FREQUENTLY IN the set A if for every $k \in \mathbb{N}$ there exists $n \geq k$ such that $x_n \in A$.

7.1.7. Example. Let S be a set, (A_n) be a sequence of subsets of S, and c an element of S. Now (A_n) is a sequence in the set $\mathfrak{P}(S)$ and the property in question is *containing the point* c. So we may say that the sequence (A_n) frequently contains c (or equivalently, that the point c is FREQUENTLY IN (A_n); or $c \in A_n$ for infinitely many n) if for every $k \in \mathbb{N}$ there exists $n \geq k$ such that $c \in A_n$.

7.2. Indexed Families of Sets

In section 1.2 we discussed families of sets. We now introduce *indexed families* of sets. The distinction between families and indexed families is very similar to the distinction between countable sets and sequences.

7.2.1. Definition. Let Λ and S be sets. An INDEXED FAMILY of elements of S is a function f from Λ into S. We should be quite clear about this: *indexed family* is just another term for *function.* (See the first paragraph of this chapter.) The domain Λ of the indexed family f is referred to as the *index set* of f. When thinking of the function f as an indexed family it is conventional to write its value at $\lambda \in \Lambda$ as f_λ rather than $f(\lambda)$. The indexed family itself may be written in a number of ways:

$$f = \big(f_\lambda\big)_{\lambda \in \Lambda} = (f_\lambda).$$

(Compare this with 7.1.1.) The range of f is, of course, a subset $\{f_\lambda \colon \lambda \in \Lambda\}$ of S, just as the range of a sequence is a countable set.

7.2.2. Example. Suppose in definition 7.2.1 that S is the family $\mathfrak{P}(T)$ of all subsets of a nonempty set T. Then a function $A \colon \Lambda \to \mathfrak{P}(T)$ would be an *indexed family of subsets of* T and, thought of that way, would usually be written as $\big(A_\lambda\big)_{\lambda \in \Lambda}$.

Just as we use different notations to distinguish a sequence (a_n) from its range the countable set $\{a_n : n \in \mathbb{N}\}$, we should be careful to distinguish between an indexed family $\big(A_\lambda\big)_{\lambda\in\Lambda}$ of sets and its range the family $\{A_\lambda : \lambda \in \Lambda\}$. Not everyone does this. One argument for not doing so is linguistic: *$\{A_\lambda : \lambda \in \Lambda\}$ is a family of sets and each set A_λ has an index λ, so why shouldn't I call $\{A_\lambda : \lambda \in \Lambda\}$ an indexed family?*

You may in your reading run across the hybrid notation $\big\{A_\lambda\big\}_{\lambda\in\Lambda}$. There are at least two sources for this. One is a failure to make the distinction we have just been discussing, in which case $\big(A_\lambda\big)_{\lambda\in\Lambda}$, $\big\{A_\lambda\big\}_{\lambda\in\Lambda}$, and $\{A_\lambda : \lambda \in \Lambda\}$ are likely to be used interchangeably. Another is the use of this notation to denote multisets in some current work. (A *multiset* is much like a set but members may have multiple occurrences. For example, the prime factorization of 360 is $2 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 5$. As a set $\{2, 2, 2, 3, 3, 5\}$ has cardinality 3 while as a multiset it is regarded as having cardinality 6.) Multisets will not be discussed in these notes. And the notation $\big\{A_\lambda\big\}_{\lambda\in\Lambda}$ will not be used.

7.2.3. Definition. Recall from definition 1.4.4 that a family $\mathfrak{A}$ of sets is (PAIRWISE) DISJOINT if $A \cap B = \emptyset$ whenever A and B are distinct members of $\mathfrak{A}$. An *indexed* family $\big(A_\lambda\big)_{\lambda\in\Lambda}$ of sets is PAIRWISE DISJOINT if $A_\lambda \cap A_\mu = \emptyset$ whenever λ and μ are distinct members of Λ.

7.2.4. Example. Suppose $J = \{1, 2, 3\}$, $A_1 = A_2 = [0, 1]$ and $A_3 = [2, 3]$. Then $\{A_j : j \in J\}$ is pairwise disjoint while $\big(A_j\big)_{j\in J}$ is not.

7.2.5. Exercise. Let $\Lambda = \mathbb{N}$ and $A_\lambda = \mathbb{R}$ for every $\lambda \in \Lambda$. Explain carefully why the family $\{A_\lambda : \lambda \in \Lambda\}$ of sets is pairwise disjoint while the indexed family $\big(A_\lambda\big)_{\lambda\in\Lambda}$ is not.

There are many situations where the indexed family notation is both appropriate and useful. If associated with each set in some family of sets there is a real number, then $(S_a)_{a\in\mathbb{R}}$ may well be the most natural notation for the family. If, on the other hand, an index set plays no role whatever in a theorem or its proof, it is usually simpler to denote a family of sets by a single letter. You need not do as many authors do and compulsively index everything in sight!

7.3. Limit Inferior and Limit Superior (for Sets)

7.3.1. Definition. If $\big(A_n\big)_{n=1}^{\infty}$ is a sequence of sets, then the LIMIT INFERIOR (or LOWER LIMIT) of the sequence is defined by

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\,.$$

Similarly, the LIMIT SUPERIOR (or UPPER LIMIT) of the sequence is defined by

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\,.$$

If the sets $\liminf A_n$ and $\limsup A_n$ are equal, then we say that the LIMIT of the sequence (A_n) exists and is equal to this set. If $L = \liminf A_n = \limsup A_n$, then we write

$$\lim_{n\to\infty} A_n = L$$

or

$$A_n \to L \quad \text{as } n \to \infty\,.$$

Most people say (and write) "lim inf" for "limit inferior" and "lim sup" (pronounced *lim-soup*) for "limit superior". Usually we write $\liminf A_n$ for $\liminf\limits_{n\to\infty} A_n$, $\limsup A_n$ for $\limsup\limits_{n\to\infty} A_n$, and $\lim A_n$ for $\lim\limits_{n\to\infty} A_n$.

Many students find the preceding definitions uninviting. The next two propositions give more appealing characterizations of the lim inf and lim sup of a sequence of sets. Nothing is lost if you choose to regard these propositions (once proved) as the *definitions* of lim inf and lim sup.

7.3.2. Proposition. *If (A_n) is a sequence of sets, then an element x belongs to* $\liminf A_n$ *if and only if x is eventually in the sequence (A_n).*

7.3.3. Proposition. *If (A_n) is a sequence of sets, then an element x belongs to* $\limsup A_n$ *if and only if x is frequently in the sequence (A_n).*

7.3.4. Proposition. *If (A_n) is a sequence if sets, then* $\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \subseteq \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$.

7.3.5. Exercise. For each $n \in \mathbb{N}$ let A_n be the interval $\left(-\frac{1}{n}, 1\right]$ whenever n is odd and the interval $\left(-1, \frac{1}{n}\right]$ whenever n is even. Find $\liminf A_n$ and $\limsup A_n$.

7.3.6. Exercise. For each $n \in \mathbb{N}$, let A_n be the open ball in $\mathbb{R}^2$ of radius 1 centered at the point $\left(\frac{(-1)^n}{n}, 0\right)$. Find $\liminf A_n$ and $\limsup A_n$.

7.3.7. Exercise. Let $a_n = \displaystyle\sum_{k=1}^{n} \frac{1}{k}$, $I_n = \big(a_n - [a_n], a_{n+1} - [a_n]\big)$, and $A_n = \big((0,1] \setminus I_n\big) \cup \big(1 + I_n\big)$ for all $n \in \mathbb{N}$. (Here $[x]$ denotes the greatest integer that is less than or equal to x.) Find $\liminf A_n$ and $\limsup A_n$.

7.3.8. Proposition. *If (A_n) is a sequence of sets, then*

$$\liminf {A_n}^c = (\limsup A_n)^c$$

and

$$\limsup {A_n}^c = (\liminf A_n)^c \,.$$

7.3.9. Definition. A sequence (A_k) of sets is INCREASING if $A_k \subseteq A_{k+1}$ for each k. It is DECREASING if $A_k \supseteq A_{k+1}$ for each k. It is MONOTONE if it is either increasing or decreasing.

7.3.10. Proposition. *If (A_k) is a monotone sequence of sets, then* $\lim\limits_{k\to\infty} A_k$ *exists.*

7.3.11. Exercise. Explain why $\limsup A_k$ is an appropriate name for $\displaystyle\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. Also, explain why $\liminf A_k$ is an appropriate name for $\displaystyle\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

In the following proposition $\mathop{\triangle}\limits_{k=1}^{\infty} A_k$ denotes $\lim\limits_{n\to\infty} \mathop{\triangle}\limits_{k=1}^{n} A_k$ where $\mathop{\triangle}\limits_{k=1}^{n} A_k = A_1 \triangle A_2 \triangle \cdots \triangle A_n$.

7.3.12. Proposition. *Suppose that (A_k) is a sequence of sets. Then* $\mathop{\triangle}\limits_{k=1}^{\infty} A_k$ *exists if and only if* $\lim_{k\to\infty} A_k = \emptyset$.

7.4. Limit Inferior and Limit Superior (for Real Numbers)

In this section we introduce "lim inf" and "lim sup" for sequences of real numbers. Notice how closely the definitions and some of the results here parallel those in the preceding section.

7.4.1. Convention. In section 5.2 we introduced the notation $\bigwedge A$ for the infimum of a subset A of a partially ordered set. In order for this infimum to exist it is necessary that A be bounded below. If A is a subset of $\mathbb{R}$ and is *not* bounded below, we will write $\bigwedge A = -\infty$ as a notational convention. In, particular, if $\{a_n\}$ is a denumerable set of real numbers which is not bounded below we write $\displaystyle\bigwedge_{n=1}^{\infty} a_n = -\infty$ or $\displaystyle\bigwedge_{n} a_n = -\infty$ or just $\bigwedge a_n = -\infty$. Similarly, if A is a subset of $\mathbb{R}$ which is not bounded above we will write $\bigvee A = \infty$. In, particular, if $\{a_n\}$ is a denumerable set of real numbers which is not bounded above we write $\displaystyle\bigvee_{n=1}^{\infty} a_n = \infty$ or $\displaystyle\bigvee_{n} a_n = \infty$ or $\bigvee a_n = \infty$.

7.4.2. Definition. If $(a_n)_{n=1}^\infty$ is a sequence of real numbers, then the LIMIT INFERIOR (or LOWER LIMIT) of the sequence is defined by

$$\liminf_{n\to\infty} a_n := \bigvee_{n=1}^\infty \bigwedge_{k=n}^\infty a_k\,.$$

Similarly, the LIMIT SUPERIOR (or UPPER LIMIT) of the sequence is defined by

$$\limsup_{n\to\infty} a_n := \bigwedge_{n=1}^\infty \bigvee_{k=n}^\infty a_k\,.$$

If $\liminf a_n$ and $\limsup a_n$ are equal, then we say that the LIMIT of the sequence exists and is equal to their common value. If $l = \liminf a_n = \limsup a_n$, then we write

$$\lim_{n\to\infty} a_n = l$$

or

$$a_n \to l \quad \text{as } n \to \infty$$

or if no confusion seems likely just

$$a_n \to l\,.$$

Whenever there exists a real number l such that $a_n \to l$ we say that the sequence (a_n) CONVERGES. Otherwise it DIVERGES.

As with sets we usually write $\liminf a_n$ for $\liminf\limits_{n\to\infty} a_n$, $\limsup a_n$ for $\limsup\limits_{n\to\infty} a_n$, and $\lim a_n$ for $\lim\limits_{n\to\infty} a_n$.

7.4.3. Convention. When one says that the supremum (or infimum) of a set of real numbers *exists* or that the limit of a sequence of real numbers *exists* one always means *exists as a real number.* Thus even though it is correct to write $\bigwedge \mathbb{Z} = -\infty$, we still say that the infimum of the set of integers *does not exist.* Similarly, while it is correct to write $\lim_{n\to\infty} n^2 = \infty$, we nevertheless say that *the limit of the sequence* $\left(n^2\right)_{n=1}^\infty$ *does not exist.*

There are alternative characterizations of the lim inf and lim sup of a sequence of real numbers.

7.4.4. Proposition. *If (a_n) is a sequence of real numbers, then a real number l is the* $\liminf a_n$ *if and only if for every $\epsilon > 0$ the sequence (a_n) is eventually greater than $l - \epsilon$ and is frequently less than $l + \epsilon$. Furthermore,* $\liminf a_n = -\infty$ *if and only if the sequence (a_n) is not bounded below.*

7.4.5. Proposition. *If (a_n) is a sequence of real numbers, then a real number l is the* $\limsup a_n$ *if and only if for every $\epsilon > 0$ the sequence (a_n) is eventually less than $l + \epsilon$ and is frequently greater than $l - \epsilon$. Furthermore,* $\limsup a_n = \infty$ *if and only if (a_n) is not bounded above.*

7.4.6. Proposition. *If (a_n) is a bounded sequence of real numbers then* $\liminf a_n \le \limsup a_n$.

7.4.7. Convention. Notice that in the preceding exercise the hypothesis that the sequence be bounded is not really necessary—provided that we extend the partial ordering on the real numbers to the EXTENDED REAL NUMBERS $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ by agreeing that $-\infty < \infty$ and that both $-\infty < a$ and $a < \infty$ hold for every real number a. It is conventional also to extend the algebraic operations on $\mathbb{R}$ to these new objects $-\infty$ and ∞. Addition is extended the way you would expect: it is assumed to be commutative and if a is a real number, then $a + \infty = \infty$ and $a + (-\infty) = -\infty$. (The expression $\infty + (-\infty)$ is left undefined.) If a is a real number and $a > 0$, then we take multiplication to be commutative and associative and set

$$a \cdot \infty = \infty \cdot \infty = (-\infty) \cdot (-\infty) = (-a) \cdot (-\infty) = \infty\,.$$

Multiplication of the new objects by negative numbers can be taken care of by agreeing that $(-1) \cdot \infty = -\infty$ and $(-1) \cdot (-\infty) = \infty$. The only surprising (and, if we are not careful, treacherous) convention is setting $0 \cdot \infty = 0$. The rationale for this is, speaking very roughly, that in measure theory we wish the "area" of a line in the plane (considered as an "infinitely long rectangle of zero width") to be zero. It should be emphasized that these conventions do not turn the extended reals into any sort of interesting algebraic object. Addition on $\overline{\mathbb{R}}$ is not even a binary operation (since $\infty + (-\infty)$ is not defined).

The lim inf and lim sup functions (from sequences of real numbers to the extended reals) are monotone. We make this precise in the next proposition.

7.4.8. Proposition. *Let (a_n) and (b_n) be sequences of real numbers. If eventually $a_n \le b_n$, then $\liminf a_n \le \liminf b_n$ and $\limsup a_n \le \limsup b_n$.*

7.4.9. Remark. It is worthwhile to take time out to notice the role played in stating the preceding proposition by the conventions adopted in 7.4.7 concerning the symbols $-\infty$ and ∞. With the conventions in place it is completely irrelevant whether either sequence (a_n) or (b_n) is bounded. Were it not for these conventions, communicating the same information would require something like the following.

7.4.10. Proposition. *Let (a_n) and (b_n) be sequences of real numbers such that eventually $a_n \le b_n$.*

(a) *If (a_n) is bounded below then so is (b_n) and $\liminf a_n \le \liminf b_n$.*
(b) *If (b_n) is bounded above then so is (a_n) and $\limsup a_n \le \limsup b_n$.*

It is a matter of individual taste whether or not the economy achieved by adopting these conventions for algebraically challenged "numbers" like $-\infty$ and ∞ is worth the price. (After all, $\mathbb{R}$ is a field; $\overline{\mathbb{R}}$ is (algebraically) not much of anything.)

7.4.11. Proposition. *If (a_n) is a bounded sequence of real numbers, then $\limsup(-a_n) = -\liminf a_n$ and $\liminf(-a_n) = -\limsup a_n$.*

Hint for proof. Use proposition 6.1.8.

7.4.12. Exercise. We can say slightly more: exercise 7.4.11 remains true even if the word "bounded" is deleted. Why? How can we communicate this extra information if we choose not to make use of the conventions in 7.4.7?

7.4.13. Proposition. *If (a_n) and (b_n) are bounded sequences of real numbers, then*

$$\liminf a_n + \liminf b_n \le \liminf(a_n + b_n) \le \limsup(a_n + b_n) \le \limsup a_n + \limsup b_n\,.$$

7.4.14. Example. In the preceding proposition each of the inequalities may be strict. (We say that inequalities of the form $x < y$ and $x > y$, which fail to be reflexive, are *strict* inequalities. We distinguish these from *weak* inequalities of the form $x \le y$ and $x \ge y$.)

7.4.15. Example. Show by example that proposition 7.4.13 is *not* true if the word "bounded" is removed from the hypothesis. Can you think of a way to fix up the statement of the result so that it holds for (at least some) not necessarily bounded sequences?

7.4.16. Proposition. *If (a_n) and (b_n) are sequences of real numbers such that $a_n \to l \in \mathbb{R}$ and $b_n \to m \in \mathbb{R}$ as $n \to \infty$, then*

$$\lim(a_n + b_n) \to l + m \quad \text{as } n \to \infty.$$

7.4.17. Proposition. *If (a_n) and (b_n) are bounded sequences of positive real numbers, then*

$$(\liminf a_n)(\liminf b_n) \le \liminf(a_n b_n) \le \limsup(a_n b_n) \le (\limsup a_n)(\limsup b_n)\,.$$

7.4.18. Example. In the preceding proposition each of the inequalities may be strict.

7.4.19. Example. Show that proposition 7.4.17 is *not* true if the word "bounded" is removed from the hypothesis. Can you think of a way to fix up the statement of the result so that it holds for (at least some) not necessarily bounded sequences?

7.4.20. Proposition. *If (a_n) and (b_n) are sequences of positive real numbers such that $a_n \to l \in \mathbb{R}$ and $b_n \to m \in \mathbb{R}$ as $n \to \infty$, then*

$$\lim(a_n b_n) \to lm \quad \text{as } n \to \infty.$$

7.4.21. Definition. If a is a real number and $\epsilon > 0$, then the open interval $(a - \epsilon, a + \epsilon)$ is the ϵ-NEIGHBORHOOD of the point a.

7.4.22. Proposition. *If (a_n) is a sequence of real numbers and $c \in \mathbb{R}$, then $\lim a_n = c$ if and only if the sequence (a_n) is eventually in every ϵ-neighborhood of c.*

7.4.23. Example. If $c > 0$ in $\mathbb{R}$, then $c^{1/n} \to 1$ as $n \to \infty$.

Hint for proof. Treat the cases $c \geq 1$ and $0 < c < 1$ separately.

7.4.24. Proposition. *If $a_n > 0$ for every $n \in \mathbb{N}$, then*

$$\liminf \frac{a_{n+1}}{a_n} \leq \liminf \sqrt[n]{a_n} \leq \limsup \sqrt[n]{a_n} \leq \limsup \frac{a_{n+1}}{a_n}\,.$$

Hint for proof. For the first inequality let $p = \liminf a_{n+1}/a_n$. If $p > 0$ choose α between 0 and p. Argue that for some natural number n_0 we have $a_{k+1}/a_k > \alpha$ whenever $k \geq n_0$. Let $n > n_0$. Show that $a_n/a_{n_0} > \alpha^{n-n_0}$. Use example 7.4.23 to show that $\liminf \sqrt[n]{a_n} \geq \alpha$.

7.4.25. Example. Each of the inequalities in the preceding exercise may be strict.

7.4.26. Exercise. Recall (or look up) statements of the *ratio test* and the *root test* from beginning calculus. Explain carefully what is meant by, and justify, the assertion that *the root test is stronger than the ratio test.*

7.4.27. Example. If $a_n = \dfrac{n}{\sqrt[n]{n!}}$, then $a_n \to e$ as $n \to \infty$.

Hint for proof. You may assume without proof that $e = \lim\bigl(1 + \frac{1}{n}\bigr)^n$.

7.4.28. Definition. A sequence (a_k) of real numbers is INCREASING if $a_k \leq a_{k+1}$ for each k. It is DECREASING if $a_k \geq a_{k+1}$ for each k. It is MONOTONE if it is either increasing or decreasing.

7.4.29. Proposition. *Every bounded monotone sequence of real numbers converges.*

7.4.30. Exercise. Explain why $\limsup a_k$ is an appropriate name for $\displaystyle\bigwedge_{n=1}^{\infty} \bigvee_{k=n}^{\infty} a_k$. Also, explain why $\liminf a_k$ is an appropriate name for $\displaystyle\bigvee_{n=1}^{\infty} \bigwedge_{k=n}^{\infty} a_k$.

7.4.31. Example. The sequence $\left(\sqrt{2}, \sqrt{2+\sqrt{2}}, \sqrt{2+\sqrt{2+\sqrt{2}}}, \dots\right)$ converges.

7.4.32. Example. Let $x_1 = -10$ and $x_{n+1} = 1 - \sqrt{1 - x_n}$ for every $n \geq 1$. Then (x_n) converges and so does $\displaystyle\lim_{n\to\infty} \frac{x_{n+1}}{x_n}$.

In the next proposition the expressions $\limsup \chi_{A_n}$, $\liminf \chi_{A_n}$, and $\lim \chi_{A_n}$ are to be taken pointwise. If g is a real valued function on S and (f_n) is a sequence of real valued functions on S, we write $\limsup f_n = g$ if $\limsup f_n(x) = g(x)$ for every $x \in S$.

7.4.33. Proposition. *Let (A_n) be a sequence of subsets of a nonempty set S. Show that*

(a) $\limsup \chi_{A_n} = \chi_{\limsup A_n}$.

(b) $\liminf \chi_{A_n} = \chi_{\liminf A_n}$.

(c) $\lim A_n$ *exists if and only if* $\lim \chi_{A_n}$ *exists. If these limits exist then*

$$\lim \chi_{A_n} = \chi_{\lim A_n}.$$

7.5. Subsequences and Cluster Points

7.5.1. Definition. Let (a_n) be a sequence of members of a set S. If $(n_k)_{k=1}^{\infty}$ is a strictly increasing sequence in $\mathbb{N}$, then the sequence

$$(a_{n_k}) = (a_{n_k})_{k=1}^{\infty} = (a_{n_1}, a_{n_2}, a_{n_3}, \dots)$$

is a SUBSEQUENCE of the sequence (a_n).

7.5.2. Example. If $a_k = 2^{-k}$ and $b_k = 4^{-k}$ for all $k \in \mathbb{N}$, then (b_k) is a subsequence of the sequence (a_k). Intuitively, this is clear, since the second sequence $(\frac{1}{4}, \frac{1}{16}, \frac{1}{64}, \dots)$ just picks out the even-numbered terms of the first sequence $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots)$. This "picking out" is implemented by the strictly increasing function $n(k) = 2\,k$ (for $k \in \mathbb{N}$). Thus $b = a \circ n$ since

$$a_{n_k} = a(n(k)) = a(2k) = 2^{-2k} = 4^{-k} = b_k$$

for all k in $\mathbb{N}$.

7.5.3. Definition. Let (a_n) be a sequence of real numbers. A point $c \in \mathbb{R}$ is a CLUSTER POINT of the sequence (a_n) if the sequence is frequently in every ϵ-neighborhood of c.

There are two observations which indicate how useful cluster points can be. One is that cluster points of sequences of real numbers are just the limits of subsequences. The other is that the lim inf of a sequence is the smallest cluster point of that sequence and the lim sup is its largest. This is the content of the next two proposition.

7.5.4. Proposition. *If (a_n) is a sequence of real numbers and $c \in R$, then c is a cluster point of the sequence if and only if there is a subsequence of (a_n) which converges to c.*

7.5.5. Example. The sequence $\left(\frac{1}{n} + \cos\frac{n\pi}{2}, \frac{1}{n^2} + \sin\frac{n\pi}{2}\right)_{n=1}^{\infty}$ has four cluster points in $\mathbb{R}^2$.

7.5.6. Proposition. *Let (a_n) be a bounded sequence of real numbers and C be the set of cluster points of (a_n). Then $\liminf a_n \in C$, $\limsup a_n \in C$, and for every $c \in C$*

$$\liminf a_n \le c \le \limsup a_n\,.$$

Many authors choose to extend the notion of cluster points to $\overline{\mathbb{R}}$. Thus when a sequence is not bounded above we may regard ∞ as a cluster point of the sequence; and when it is not bounded below we take $-\infty$ to be a cluster point. The convenience of adopting this convention is that then the hypothesis that (a_n) be bounded may be removed from proposition 7.5.6.

7.5.7. Proposition. *Every bounded sequence of real numbers has a convergent subsequence.*

CHAPTER 8

CATEGORIES

In mathematics we study things (objects) and certain mappings between them (morphisms). To mention just a few, sets and functions, groups and homomorphisms, topological spaces and continuous maps, vector spaces and linear transformations, and Hilbert spaces and bounded linear maps. These examples come from different branches of mathematics—set theory, group theory, topology, linear algebra, and functional analysis, respectively. But these different areas have many things in common: products, coproducts, quotients, and the like. Category theory is an attempt to formalize some of these common constructions. Very loosely speaking, category theory is the study of what different branches of mathematics have in common.

In these notes just as we use the language of sets without pursuing the study of set theory, we will cheerfully use the language of categories without first embarking on a preliminary study of category theory (which itself is a large and important area of research). It can sometimes be a bit challenging to learn even the language of categories because many of the introductory treatments lean heavily on one's algebraic background. For a very gentle entree to the world of categories try Chapter 3 of Semadeni's beautiful book [41] on *Banach Spaces of Continuous Functions*. Another standard text is the book *Categories for the Working Mathematician* by one of the founders of the subject, Saunders Mac Lane.

By pointing to unifying principles the language of categories often provides striking insight into "the way things work" in mathematics. Equally importantly, one gains in efficiency by not having to perform essentially the same construction over and over in slightly different contexts.

8.1. Objects and Morphisms

8.1.1. Definition. Let $\mathfrak{A}$ be a class, whose members we call OBJECTS. For every pair (S,T) of objects we associate a set $\mathfrak{Mor}(S,T)$, whose members we call MORPHISMS from S to T. We assume that $\mathfrak{Mor}(S,T)$ and $\mathfrak{Mor}(U,V)$ are disjoint unless $S=U$ and $T=V$.

We suppose further that there is an operation $\circ$ (called COMPOSITION) that associates with every $\alpha \in \mathfrak{Mor}(S,T)$ and every $\beta \in \mathfrak{Mor}(T,U)$ a morphism $\beta\circ\alpha \in \mathfrak{Mor}(S,U)$ in such a way that:

(1) $\gamma\circ(\beta\circ\alpha) = (\gamma\circ\beta)\circ\alpha$ whenever $\alpha \in \mathfrak{Mor}(S,T)$, $\beta \in \mathfrak{Mor}(T,U)$, and $\gamma \in \mathfrak{Mor}(U,V)$;
(2) for every object S there is a morphism $I_S \in \mathfrak{Mor}(S,S)$ satisfying $\alpha\circ I_S = \alpha$ whenever $\alpha \in \mathfrak{Mor}(S,T)$ and $I_S\circ\beta = \beta$ whenever $\beta \in \mathfrak{Mor}(R,S)$.

Under these circumstances the class $\mathfrak{A}$, together with the associated families of morphisms, is a CATEGORY.

We will reserve the notation $S \xrightarrow{\alpha} T$ for a situation in which S and T are objects in some category and α is a morphism belonging to $\mathfrak{Mor}(S,T)$. As is the case with groups and vector spaces we usually omit the composition symbol $\circ$ and write $\beta\alpha$ for $\beta\circ\alpha$.

8.1.2. Example. The category **SET** has sets for objects and functions (maps) as morphisms.

8.1.3. Example. The category **AbGp** has Abelian groups for objects and group homomorphisms as morphisms. (See proposition 4.3.11.)

8.1.4. Example. The category **VEC** has vector spaces for objects and linear transformations as morphisms. (See proposition 4.5.6.)

8.1.5. Example. The category **ALG** has algebras for objects and algebra homomorphisms for morphisms. (See proposition 4.7.5.

The preceding examples are examples of *concrete categories*—that is, categories in which the objects are sets (together, usually, with additional structure) and the morphism are functions (usually preserving this extra structure). In these notes the categories of interest to us are concrete ones. Here (for those who are curious) is a more formal definition of *concrete category*.

8.1.6. Definition. A category $\mathfrak{A}$ together with a function $\Box$ which assigns to each object A in $\mathfrak{A}$ a set $\Box A$ is a CONCRETE category if the following conditions are satisfied:

(i) every morphism $A \xrightarrow{f} B$ is a function from $\Box A$ to $\Box B$;
(ii) each identity morphism I_A in $\mathfrak{A}$ is the identity function on $\Box A$; and
(iii) composition of morphisms $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ agrees with composition of the functions $f \colon \Box A \to \Box B$ and $g \colon \Box B \to \Box C$.

If A is an object in a concrete category $\mathfrak{A}$, then $\Box A$ is the UNDERLYING SET of A.

Although it is true that the categories of interest in these notes are concrete categories, it may nevertheless be interesting to see an example of a category that is *not* concrete.

8.1.7. Example. Let G be a monoid. Consider a category **C** having exactly one object, which we call $\star$. Since there is only one object there is only one family of morphisms $\mathfrak{Mor}(\star,\star)$, which we take to be G. Composition of morphisms is defined to be the monoid multiplication. That is, $a \circ b := ab$ for all a, $b \in G$. Clearly composition is associative and the identity element of G is the identity morphism. So **C** is a category.

8.1.8. Definition. In any concrete category we will call an injective morphism a MONOMORPHISM and a surjective morphism an EPIMORPHISM.

8.1.9. CAUTION. The definitions above reflect the original Bourbaki use of the term and are the ones most commonly adopted by mathematicians outside of category theory where "monomorphism" means "left cancellable" and "epimorphism" means "right cancellable". (Notice that the terms *injective* and *surjective* may not make sense when applied to morphisms in a category that is not concrete.)

A morphism $B \xrightarrow{g} C$ is LEFT CANCELLABLE if whenever morphisms $A \xrightarrow{f_1} B$ and $A \xrightarrow{f_2} B$ satisfy $gf_1 = gf_2$, then $f_1 = f_2$. Mac Lane suggested calling left cancellable morphisms MONIC morphisms. The distinction between monic morphisms and monomorphisms turns out to be slight. In these notes almost all of the morphisms we encounter are monic if and only if they are monomorphisms. As an easy exercise prove that any injective morphism in a (concrete) category is monic. The converse sometimes fails.

In the same vein Mac Lane suggested calling a *right cancellable* morphism (that is, a morphism $A \xrightarrow{f} B$ such that whenever morphisms $B \xrightarrow{g_1} C$ and $B \xrightarrow{g_2} C$ satisfy $g_1 f = g_2 f$, then $g_1 = g_2$) an EPIC morphism. Again it is an easy exercise to show that in a (concrete) category any epimorphism is epic. The converse, however, fails in some rather common categories.

8.1.10. Definition. The terminology for inverses of morphisms in categories is essentially the same as for functions (see section 2.10). Let $S \xrightarrow{\alpha} T$ and $T \xrightarrow{\beta} S$ be morphisms in a category. If $\beta \circ \alpha = I_S$, then β is a LEFT INVERSE of α and, equivalently, α is a RIGHT INVERSE of β. We say that the morphism α is an ISOMORPHISM (or is INVERTIBLE) if there exists a morphism $T \xrightarrow{\beta} S$ which is both a left and a right inverse for α. Such a function is denoted by α^{-1} and is called the INVERSE of α.

In any concrete category one can inquire whether every bijective morphism (that is, every map which is both a monomorphism and an epimorphism) is an isomorphism. The answer is often a trivial *yes* (as in propositions 4.3.4, 4.5.11, 4.7.3, and the next two examples) or a trivial *no* (for example, in the category **POSET** of partially ordered sets and order preserving maps, see exercise 5.2.11). But on occasion the answer turns out to be a fascinating and deep result (see section 29.4).

8.1.11. Example. In the category **SET** every bijective morphism is an isomorphism.

8.1.12. Example. The category **LAT** has lattices as objects and lattice homomorphisms as morphisms. In this category bijective morphisms are isomorphisms.

8.1.13. Example. If in the category **C** of example 8.1.7 the monoid G is a group then every morphism in **C** is an isomorphism.

8.2. Quotients

Proposition 2.11.7 illustrates an important class of categorical constructions—quotients. The following definition evinces a very general attitude towards quotients. It specifies what quotients *do*—not what they *are*. This definition, like many very abstract definitions, may at first seem rather unmotivated and devoid of intuitive content. Throughout these notes you will encounter many quotient constructions. My advice is to learn the definition first, and then check it against the examples as they come along. A comfortable understanding of the idea of quotient will come at the *end* of this process. Don't demand it up front. For those with serious allergies to abstract notions an alternative strategy is to file the definition away until you have seen enough examples that you feel you are ready for the general notion.

8.2.1. Definition. Let A be an object in a category **C**. An epimorphism $A \xrightarrow{\pi} B$ in **C** is a QUOTIENT MAP for A if a map $g\colon B \to C$ (in **SET**) is a morphism (in **C**) whenever $g \circ \pi$ is a morphism. An object B in **C** is a QUOTIENT OBJECT for A if it is the range of some quotient map for A.

8.2.2. Example. In the category **SET** every epimorphism is a quotient map.

8.2.3. Example. In the category **AbGp** every epimorphism is a quotient map.

In the category **SET** there is a standard quotient object for a set S created whenever an equivalence relation is specified on S. That is the point of proposition 2.11.7. Similarly, in the category **AbGp** a standard quotient object for a group G is created whenever a subgroup of G is specified. Here is how the construction goes.

8.2.4. Definition. Let H be a subgroup of an Abelian group G. Define an equivalence relation $\sim$ on G by

$$x \sim y \qquad \text{if and only if} \qquad y - x \in H.$$

For each $x \in G$ let $[x]$ be the equivalence class containing x. Let G/H be the set of all equivalence classes of elements of G. For $[x]$ and $[y]$ in G/H define

$$[x] + [y] := [x + y]\,.$$

Under this operation G/H becomes an Abelian group. It is the QUOTIENT GROUP (or FACTOR GROUP) of G by H. The notation G/H is usually read "G mod H". The linear map

$$\pi\colon G \to G/H\colon x \mapsto [x]$$

is called the QUOTIENT MAP.

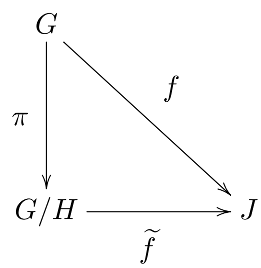
8.2.5. Exercise. Several things in the preceding definition require proof. Show that:

(a) $\sim$ is an equivalence relation.
(b) Addition of equivalence classes is well defined.
(c) G/H is an Abelian group.
(d) The map π is a quotient map (in the sense of definition 8.2.1).

(e) The space G/H is a quotient object.

The most important theorem about quotients in the category **AbGp** is frequently called the *fundamental homomorphism theorem* or the *first isomorphism theorem.*

8.2.6. Theorem (Fundamental quotient theorem for **AbGp**)**.** *Let G and J be Abelian groups and H be a subgroup of G. If f is a homomorphism from G to J and $\ker f \supseteq H$, then there exists a unique homomorphism $\widetilde{f}\colon G/H \to J$ which makes the following diagram commute.*



Furthermore, $\widetilde{f}$ is injective if and only if $\ker f = H$; and $\widetilde{f}$ is surjective if and only if f is.

Everything works in vector spaces almost exactly as it does in Abelian groups. There are just a few more details.

8.2.7. Example. Prove that in the category **VEC** every epimorphism is a quotient map.

8.2.8. Definition. Let M be a subspace of a vector space V. Define an equivalence relation $\sim$ on V by

$$x \sim y \qquad \text{if and only if} \qquad y - x \in M.$$

For each $x \in V$ let $[x]$ be the equivalence class containing x. Let V/M be the set of all equivalence classes of elements of V. For $[x]$ and $[y]$ in V/M define

$$[x] + [y] := [x + y]$$

and for $\alpha \in \mathbb{R}$ and $[x] \in V/M$ define

$$\alpha[x] := [\alpha x]\,.$$

Under these operations V/M becomes a vector space. It is the QUOTIENT SPACE (or FACTOR SPACE) of V by M. The notation V/M is usually read "V mod M". The linear map

$$\pi\colon V \to V/M \colon x \mapsto [x]$$

is a quotient map.

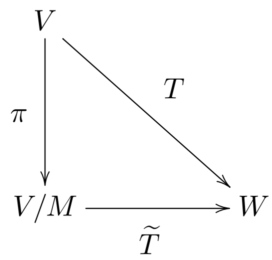
8.2.9. Exercise. Prove the facts necessary for the preceding definition to make sense. In particular, show that:

(a) $\sim$ is an equivalence relation.
(b) Addition of equivalence classes is well defined.
(c) Multiplication of an equivalence class by a scalar is well defined.
(d) V/M is a vector space.
(e) The map π is a quotient map (in the sense of definition 8.2.1).
(f) The space V/M is a quotient object.

8.2.10. Exercise. In the definition of a *quotient vector space* above take $V = \mathbb{R}^3$ and M to be the z-axis. Describe the resulting equivalence classes and exhibit an isomorphism between V/M and $\mathbb{R}^2$. Next, take M to be the xy-plane. Describe the resulting equivalence classes and explain why, in this case, V/M is isomorphic to $\mathbb{R}$.

Theorems 2.11.9 and 8.2.6 have an obvious analogue in the category of vector spaces and linear maps.

8.2.11. Theorem (Fundamental quotient theorem for **VEC**)**.** *Let V and W be vector spaces and M be a subspace of V. If T is a linear map from V to W and* $\ker T \supseteq M$*, then there exists a unique linear map $\widetilde{T}\colon V/M \to W$ which makes the following diagram commute.*



Furthermore, $\widetilde{T}$ is injective if and only if $\ker T = M$*; and $\widetilde{T}$ is surjective if and only if T is.*

Do not be tempted on the basis of 8.2.2, 8.2.3, and 8.2.7 to conjecture that in every category surjective morphisms are automatically quotient maps. We will encounter a simple counterexample in example 11.6.6.

8.2.12. Definition. Let J be a proper ideal in an algebra A. Define an equivalence relation $\sim$ on A by

$$a \sim b \qquad \text{if and only if} \qquad b - a \in J.$$

For each $a \in A$ let $[a]$ be the equivalence class containing a. Let A/J be the set of all equivalence classes of elements of A. For $[a]$ and $[b]$ in A/J define

$$[a] + [b] := [a + b] \qquad \text{and} \qquad [a][b] := [ab]$$

and for $\alpha \in \mathbb{C}$ and $[a] \in A/J$ define

$$\alpha[a] := [\alpha a]\,.$$

Under these operations A/J becomes an algebra. It is the QUOTIENT ALGEBRA of A by J. The notation A/J is usually read "A mod J". The surjective algebra homomorphism

$$\pi\colon A \to A/J\colon a \mapsto [a]$$

is called the QUOTIENT MAP.

8.2.13. Exercise. Verify the assertions made in the preceding definition.

8.3. Products

Products and coproducts, like quotients, are best described in terms of what they *do*.

8.3.1. Definition. Let A_1 and A_2 be objects in a category **C**. We say that the triple (P, π_1, π_2), where P is an object and $\pi_k\colon P \to A_k$ $(k = 1, 2)$ are morphisms, is a PRODUCT of A_1 and A_2 if for every object B and every pair of morphisms $f_k\colon B \to A_k$ $(k = 1, 2)$ there exists a unique morphism $g\colon B \to P$ such that $f_k = \pi_k \circ g$ for $k = 1, 2$.

It is conventional to say, "Let P be a product of ..." for, "Let (P, π_1, π_2) be a product of ...". The product of A_1 and A_2 is often written as $A_1 \times A_2$ or as $\prod_{k=1,2} A_k$.

In a particular category products may or may not exist. It is an interesting and elementary fact that whenever they exist they are unique (up to isomorphism), so that we may unambiguously speak of *the* product of two objects. When we say that a categorical object satisfying some condition(s) is *unique up to isomorphism* we mean, of course, that any two objects satisfying the condition(s) must be isomorphic. We will often use the phrase "essentially unique" for "unique up to isomorphism."

8.3.2. Proposition. *In any category products (if they exist) are essentially unique.*

8.3.3. Example. In the category **SET** the product of two sets A_1 and A_2 exists and is in fact the usual Cartesian product $A_1 \times A_2$ together with the usual coordinate projections $\pi_k\colon A_1 \times A_2 \to A_k\colon (a_1, a_2) \mapsto a_k$.

8.3.4. Example. If V_1 and V_2 are vector spaces we make the Cartesian product $V_1 \times V_2$ into a vector space as follows. Define addition by

$$(u, v) + (w, x) := (u + w, v + x)$$

and scalar multiplication by

$$\alpha(u, v) := (\alpha u, \alpha v)\,.$$

This makes $V_1 \times V_2$ into a vector space and that this space together with the usual coordinate projections $\pi_k \colon V_1 \times V_2 \to V_k \colon (v_1, v_2) \mapsto v_k$ $(k = 1, 2)$ is a product in the category **VEC**. It is usually called the DIRECT SUM of V and W and is denoted by $V \oplus W$.

It often possible and desirable to take the product of an arbitrary family of objects in a category. Following is a generalization of definition 8.3.1.

8.3.5. Definition. Let $\bigl(A_\lambda\bigr)_{\lambda\in\Lambda}$ be an indexed family of objects in a category **C**. We say that the object P together with an indexed family $\bigl(\pi_\lambda\bigr)_{\lambda\in\Lambda}$ of morphisms $\pi_\lambda \colon P \to A_\lambda$ is a PRODUCT of the objects A_λ if for every object B and every indexed family $\bigl(f_\lambda\bigr)_{\lambda\in\Lambda}$ of morphisms $f_\lambda \colon B \to A_\lambda$ there exists a unique map $g \colon B \to P$ such that $f_\lambda = \pi_\lambda \circ g$ for every $\lambda \in \Lambda$.

A category in which arbitrary products exist is said to be PRODUCT COMPLETE. Many of the categories we encounter in these notes are product complete.

8.3.6. Definition. Let $\bigl(S_\lambda\bigr)_{\lambda\in\Lambda}$ be an indexed family of sets. The CARTESIAN PRODUCT of the indexed family, denoted by $\prod_{\lambda\in\Lambda} S_\lambda$ or just $\prod S_\lambda$, is the set of all functions $f \colon \Lambda \to \bigcup S_\lambda$ such that $f(\lambda) \in S_\lambda$ for each $\lambda \in \Lambda$. The maps $\pi_\lambda \colon \prod S_\lambda \to S_\lambda \colon f \mapsto f(\lambda)$ are the canonical COORDINATE PROJECTIONS. In many cases the notation f_λ is more convenient than $f(\lambda)$. (See, for example, example 8.3.8 below.)

8.3.7. Example. A very important special case of the preceding definition occurs when all of the sets S_λ are identical: say $S_\lambda = A$ for every $\lambda \in \Lambda$. In this case the Cartesian product comprises all the functions which map Λ into A. That is, $\prod_{\lambda\in\Lambda} S_\lambda = A^\Lambda$ (see exercise 3.2.18). Notice also that in this case the coordinate projections are *evaluation maps*. For each λ the coordinate projection π_λ takes each point f in the product (that is, each function f from Λ into A) to $f(\lambda)$ its value at λ. Briefly, each coordinate projection is an evaluation map at some point.

8.3.8. Example. What is $\mathbb{R}^n$? It is just the set of all n-tuples of real numbers. That is, it is the set of functions from $\mathbb{N}_n = \{1, 2, \dots, n\}$ into $\mathbb{R}$. In other words $\mathbb{R}^n$ is just shorthand for $\mathbb{R}^{\mathbb{N}_n}$. In $\mathbb{R}^n$ one usually writes x_j for the j^{th} coordinate of a vector x rather than $x(j)$.

8.3.9. Example. In the category **SET** the product of an indexed family of sets exists and is in fact the Cartesian product of these sets together with the canonical coordinate projections.

8.3.10. Example. If $\bigl(V_\lambda\bigr)_{\lambda\in\Lambda}$ is an indexed family of vector spaces we make the Cartesian product $\prod V_\lambda$ into a vector space as follows. Define addition and scalar multiplication pointwise: for f, $g \in \prod A_\lambda$ and $\alpha \in \mathbb{R}$

$$(f + g)(\lambda) := f(\lambda) + g(\lambda)$$

and

$$(\alpha f)(\lambda) := \alpha f(\lambda)\,.$$

This makes $\prod V_\lambda$ into a vector space, which is sometimes called the DIRECT PRODUCT of the spaces V_λ, and this space together with the canonical coordinate projections (which are certainly linear maps) is a product in the category **VEC**.

8.3.11. Exercise. Prove or disprove:

(a) The Cartesian product of a finite family of countable sets is countable.
(b) The Cartesian product of a countable family of finite sets is countable.

Not every category is product complete.

8.3.12. Exercise. Show that in the category of all nonempty subsets of $\mathbb{R}$ and inclusion maps ($\mathfrak{Mor}(A, B) = \emptyset$ if $A \nsubseteq B$) products need not exist.

8.4. Coproducts

Coproducts are like products—except all the arrows are reversed.

8.4.1. Definition. Let A_1 and A_2 be objects in a category **C**. The triple (P, j_1, j_2), (P is an object and $j_k\colon A_k \to P$, $k = 1, 2$ are morphisms) is a COPRODUCT of A_1 and A_2 if for every object B and every pair of morphisms $g_k\colon A_k \to B$ (k = 1,2) there exists a unique morphism $h\colon P \to B$ such that $g_k = h \circ j_k$ for $k = 1, 2$.

8.4.2. Proposition. *In any category coproducts (if they exist) are essentially unique.*

8.4.3. Definition. In definition 1.4.6 we introduced the notation $A \uplus B$ for the union of two disjoint sets A and B. We now extend this notation somewhat and allow ourselves to take the *disjoint union* of sets A and B even if they are not disjoint. We "make them disjoint" by identifying the set A with the set $A' = \{(a, 1)\colon a \in A\}$ and B with the set $B' = \{(b, 2)\colon b \in B\}$. Then we denote the union of A' and B', which *are* disjoint, by $A \uplus B$ and call it the DISJOINT UNION of A and B.

In general, if $\big(A_\lambda\big)_{\lambda\in\Lambda}$ is an indexed family of sets, its DISJOINT UNION, $\biguplus_{\lambda\in\Lambda} A_\lambda$ is defined to be $\bigcup\{(A_\lambda, \lambda)\colon \lambda \in \Lambda\}$.

8.4.4. Example. In the category **SET** the coproduct of two sets A_1 and A_2 exists and is their disjoint union $A_1 \uplus A_2$ together with the obvious inclusion maps $\iota_k\colon A_k \to A_1 \uplus A_2$ $(k = 1, 2)$.

It is interesting to observe that while the product and coproduct of a finite collection of objects in the category **SET** are quite different, in the more complex category **VEC** they turn out to be exactly the same thing.

8.4.5. Example. If V_1 and V_2 are vector spaces make the Cartesian product $V_1 \times V_2$ into a vector space as in example 8.3.4. This space together with the obvious injections is a coproduct in the category **VEC**.

It often possible and desirable to take the coproduct of an arbitrary family of objects in a category. Following is a generalization of definition 8.4.1.

8.4.6. Definition. Let $\big(A_\lambda\big)_{\lambda\in\Lambda}$ be an indexed family of objects in a category **C**. We say that the object C together with an indexed family $\big(\iota_\lambda\big)_{\lambda\in\Lambda}$ of morphisms $\iota_\lambda\colon A_\lambda \to C$ is a COPRODUCT of the objects A_λ if for every object B and every indexed family $\big(f_\lambda\big)_{\lambda\in\Lambda}$ of morphisms $f_\lambda\colon A_\lambda \to B$ there exists a unique map $g\colon C \to B$ such that $f_\lambda = g \circ \iota_\lambda$ for every $\lambda \in \Lambda$. The usual notation for the coproduct of the objects A_λ is $\coprod_{\lambda\in\Lambda} A_\lambda$.

8.4.7. Example. In the category **SET** the coproduct of an indexed family of sets exists and is the disjoint union of these sets.

8.4.8. Definition. Let S be a set and V be a vector space. The SUPPORT of a function $f\colon S \to V$ is $\{s \in S\colon f(s) \neq \mathbf{0}\}$.

8.4.9. Example. If $\big(V_\lambda\big)_{\lambda\in\Lambda}$ is an indexed family of vector spaces we make the Cartesian product into a vector space as in example 8.3.10. The set of functions f belonging to $\prod V_\lambda$ which have finite support (that is, which are nonzero only finitely often) is clearly a subspace of $\prod V_\lambda$. This subspace is the DIRECT SUM of the spaces V_λ. It is denoted by $\bigoplus_{\lambda\in\Lambda} V_\lambda$. This space together with the obvious injective maps (which are linear) is a coproduct in the category **VEC**.

CHAPTER 9

ORDERED VECTOR SPACES

9.1. Partially Orderings on Vector Spaces

9.1.1. Definition. A partial ordering $\leq$ on a real vector space V is COMPATIBLE with (or RESPECTS) the operations (addition and scalar multiplication) on V if for all x, y, $z \in V$

(a) $x \leq y$ implies $x + z \leq y + z$, and
(b) $x \leq y$, $\alpha \geq 0$ imply $\alpha x \leq \alpha y$.

A real vector space equipped with a partial ordering which is compatible with the vector space operations is an ORDERED VECTOR SPACE.

9.1.2. Example. The real line $\mathbb{R}$ (with its usual operations and partial ordering) is an ordered vector space.

9.1.3. Exercise. The usual ordering on $\mathbb{R}$ cannot be extended to a partial ordering on the complex plane $\mathbb{C}$ which respects both addition and multiplication. (A partial ordering on a ring *respects multiplication* if ab is positive whenever both a and b are.)

9.1.4. Example. If S is a nonempty set the family $\mathcal{F}(S)$ of real valued functions on S is an ordered vector space under the pointwise operations defined in example 4.4.7 and the partial ordering defined in example 5.1.8. This is the USUAL ORDERING on $\mathcal{F}(S)$. Explain how this makes $\mathbb{R}^n$ and $s = s(\mathbb{R})$, the set of all sequences of real numbers, into ordered vector spaces.

Proposition 6.1.9 is a special case of the next result. The proof however may be quite different from the one you gave in 6.1.9.

9.1.5. Proposition. *Let A and B be nonempty subsets of an ordered vector space. If A and B both have least upper bounds, then so does $A + B$ and*

$$\sup(A + B) = \sup A + \sup B\,.$$

9.1.6. Proposition. *Let A be a nonempty subset of an ordered vector space. If A has a least upper bound and $\alpha \geq 0$, then αA has a least upper bound and*

$$\sup(\alpha A) = \alpha \sup A\,.$$

The result stated for sets of real numbers in exercise 6.1.8 holds in ordered vector spaces.

9.1.7. Proposition. *Let A be a nonempty subset of an ordered vector space. If A has a greatest lower bound, then $-A$ has a least upper bound and*

$$\sup(-A) = -\inf A\,.$$

9.1.8. Exercise. Show how to use the preceding exercise to prove that if A and B are nonempty subsets of an ordered vector space and if A and B both have greatest lower bounds, then so does $A + B$ and

$$\inf(A + B) = \inf A + \inf B\,.$$

It is a good idea to be alert for uses of special cases of the preceding very general results. For example,

$$(a + b) \wedge (a + c) = a + (b \wedge c)$$

is an identity which occurs frequently. It is just exercise 9.1.8 where $A = \{a\}$ and $B = \{b, c\}$.

9.2. Convexity

9.2.1. Definition. Let V be a vector space. Recall that a *linear combination* of a finite set $\{x_1, \dots, x_n\}$ of vectors in V is a vector of the form $\sum_{k=1}^{n} \alpha_k x_k$ where $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. If $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$, then the linear combination is *trivial*; if at least one α_k is different from zero, the linear combination is *nontrivial*. A linear combination $\sum_{k=1}^{n} \alpha_k x_k$ of the vectors $x_1, \dots, x_n$ is a CONVEX COMBINATION if $\alpha_k \geq 0$ for each k $(1 \leq k \leq n)$ and if $\sum_{k=1}^{n} \alpha_k = 1$.

9.2.2. Definition. If a and b are vectors in the vector space V, then the CLOSED SEGMENT between a and b, denoted by $[a, b]$, is $\{(1-t)a + tb \colon 0 \leq t \leq 1\}$.

9.2.3. CAUTION. Notice that there is a slight conflict between this notation, when applied to the vector space $\mathbb{R}$ of real numbers, and the notation introduced in section 1.7. In $\mathbb{R}$ the closed segment $[a, b]$ is the same as the closed interval $[a, b]$ provided that $a \leq b$. If $a > b$, however, the closed segment $[a, b]$ is the same as the segment $[b, a]$, it contains all numbers c such that $b \leq c \leq a$, whereas the closed interval $[a, b]$ is empty.

9.2.4. Definition. A subset C of a vector space V is CONVEX if the closed segment $[a, b]$ is contained in C whenever a, $b \in C$.

9.2.5. Example. A disk is a convex subset of $\mathbb{R}^2$. The set $\{(x, y) \colon 1 \leq x^2 + y^2 \leq 2\}$ is not a convex subset of $\mathbb{R}^2$.

9.2.6. Example. Every subspace of a vector space is convex.

9.2.7. Definition. Let A be a subset of a vector space V. The CONVEX HULL of A is the smallest convex subset of V which contain A.

9.2.8. Exercise. Show that the preceding definition makes sense by showing that the intersection of a family of convex subsets of a vector space is itself convex. Then show that a "constructive characterization" is equivalent; that is, prove that the convex hull of A is the set of all convex combinations of elements of A. (See remark 5.3.6.)

9.3. Positive Cones

9.3.1. Definition. Let V be a vector space. A subset C of V is a CONE in V if $\alpha C \subseteq C$ for every $\alpha \geq 0$. A cone C in V is PROPER if $C \cap (-C) = \{\mathbf{0}\}$.

9.3.2. Proposition. *A cone C in a vector space is convex if and only if $C + C \subseteq C$.*

9.3.3. Example. If V is an ordered vector space, then the set

$$V^+ := \{x \in V \colon x \geq \mathbf{0}\}$$

is a proper convex cone in V. This is the POSITIVE CONE of V and its members are the POSITIVE ELEMENTS of V.)

9.3.4. Proposition. *Let V be a vector space and C be a proper convex cone in V. Define $x \leq y$ if $y - x \in C$. The relation $\leq$ is a partial ordering on V and is compatible with the vector space operations on V. This relation is the* partial ordering induced by *the cone C. The positive cone V^+ of the resulting ordered vector space is just C itself.*

9.3.5. Definition. Let V be an ordered vector space. Its positive cone V^+ is GENERATING if $V = V^+ - V^+$; that is, if every element of V can be written (in at least one way) as the difference of positive elements.

9.3.6. Example. Let $C = \{(x, y) \in \mathbb{R}^2 \colon x \geq 0 \text{ and } y \geq 0\}$. Then C is a proper convex cone in $\mathbb{R}^2$. If $\leq$ is the partial ordering on $\mathbb{R}^2$ induced by C then $(u, v) \leq (x, y)$ if and only if $u \leq x$ and $v \leq y$. The cone C is generating. This is the USUAL ORDERING on $\mathbb{R}^2$.

9.3.7. Example. Let $C = \{(x, y) \in \mathbb{R}^2 : x > 0\} \cup \{(0, y) \in \mathbb{R}^2 : y \geq 0\}$. Then C is a proper convex cone in $\mathbb{R}^2$. If $\leq$ is the partial ordering on $\mathbb{R}^2$ induced by C then $(u, v) \leq (x, y)$ if and only if either $u < x$ or else $u = x$ and $v \leq y$. The cone C is generating. This is LEXICOGRAPHIC ORDERING on $\mathbb{R}^2$. (See example 5.1.6.)

9.3.8. Example. Let $C = \{(x, 0) \in \mathbb{R}^2 : x \geq 0\}$. Then C is a proper convex cone in $\mathbb{R}^2$. If $\leq$ is the partial ordering on $\mathbb{R}^2$ induced by C then $(u, v) \leq (x, y)$ if and only $u \leq x$ and $v = y$. The cone C is *not* generating.

9.3.9. Remark. It is quite easy to see that if the positive cone V^+ in an ordered vector space V is not generating, then $W = V^+ - V^-$ is a partially ordered vector space contained in V and its positive cone, which is V^+, *is* generating.

9.3.10. Example. In example 9.3.8 the positive cone C (the positive x-axis) is not generating. But the ordered vector space $C - C$ generated by C is clearly the (entire) x-axis in $\mathbb{R}^2$.

9.3.11. Example. Let S be a nonempty set and $\mathcal{F}(S)$ be the set of all real valued functions on S. In example 9.1.4 we found that under pointwise operations and partial ordering $\mathcal{F}(S)$ is an ordered vector space. In this space the set $C = \{f \in \mathcal{F}(S) : f(x) \geq 0 \text{ for all } x \in S\}$ is a proper convex cone. The partial ordering this cone induces on $\mathcal{F}(S)$ is exactly the (pointwise) ordering introduced in example 5.1.8.

9.4. Finitely Additive Set Functions

9.4.1. Definition. Let $\mathfrak{A}$ be an algebra of subsets of a set S. A real valued function μ on $\mathfrak{A}$ is a REAL FINITELY ADDITIVE SET FUNCTION on $\mathfrak{A}$ if

$$\mu(A \uplus B) = \mu(A) + \mu(B)$$

whenever A and B are disjoint members of $\mathfrak{A}$.

A real valued finitely additive set function on an algebra $\mathfrak{A}$ of sets is BOUNDED if there exists a number $M > 0$ such that

$$|\mu(A)| \leq M$$

for all $A \in \mathfrak{A}$. We denote by $\mathrm{ba}(S, \mathfrak{A})$ the set of all bounded real finitely additive set functions on $\mathfrak{A}$. (If no confusion will result we shorten the notation to $\mathrm{ba}(S)$.)

9.4.2. Proposition. *If μ is a real finitely additive set function on an algebra of sets, then $\mu(\emptyset) = 0$.*

9.4.3. Proposition. *If μ is a real finitely additive set function on an algebra $\mathfrak{A}$ of sets and A, $B \in \mathfrak{A}$, then*

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)\,.$$

9.4.4. Proposition. *If μ is a real finitely additive set function on an algebra $\mathfrak{A}$ of sets and A, $B \in \mathfrak{A}$, then*

$$\mu(B \setminus A) = \mu(B) - \mu(A \cap B)\,.$$

If, additionally, $A \subseteq B$, then the conclusion is

$$\mu(B \setminus A) = \mu(B) - \mu(A)\,.$$

9.4.5. Definition. We say that a real valued function f on any set S is POSITIVE if $f(s) \geq 0$ for all $s \in S$. In particular, a real finitely additive set function μ on an algebra $\mathfrak{A}$ of sets is POSITIVE if $\mu(A) \geq 0$ for all $A \in \mathfrak{A}$.

The next result says that positive finitely additive set functions are COUNTABLY SUBADDITIVE.

9.4.6. Proposition. *If μ is a positive finitely additive set function on an algebra $\mathfrak{A}$ of sets and if $A_1, \dots, A_n \in \mathfrak{A}$, then*

$$\mu\Big(\bigcup_{k=1}^{n} A_k\Big) \leq \sum_{k=1}^{n} \mu(A_k)\,.$$

9.4.7. Example. Let S be a finite set and for every $A \subseteq S$ let $\mu(A) = \operatorname{card} A$. Then μ is a bounded real finitely additive set function on $\mathfrak{P}(S)$.

It is a little hard to think of examples of unbounded finitely additive set functions, but here is one. (Another appears as a consequence of the *Hahn-Banach theorem* in example 29.1.17.)

9.4.8. Example. Let $\mathfrak{S}$ be the set of all intervals in $\mathbb{R}$ of the form $(a, b\,]$, $(-\infty, b\,]$, or (a, ∞). The family $\mathfrak{A}$ of all finite unions of members of $\mathfrak{S}$ is an algebra of subsets of $\mathbb{R}$. Define μ on $\mathfrak{S}$ by setting $\mu\big((a, b]\big) = b - a$ if $a \leq b$, $\mu\big((-\infty, a)\big) = a$, and $\mu\big((b, \infty)\big) = -b$; then use finite additivity to extend μ to all of $\mathfrak{A}$. This makes μ into a real finitely additive set function on $\mathfrak{A}$. It is not bounded.

9.4.9. Exercise. Give an example of an infinite set S and a bounded real finitely additive set function μ on $\mathfrak{P}(S)$. Make your example nontrivial by arranging things so that $\mu(\{s\})$ is different from zero for every $s \in S$.

9.4.10. Example. Let $\mathfrak{A}$ be an algebra of subsets of a set S. For μ, $\nu \in \operatorname{ba}(S)$ and $\alpha \in \mathbb{R}$ define

$$(\mu + \nu)(A) = \mu(A) + \nu(A) \qquad \text{and}$$
$$(\alpha\mu)(A) = \alpha\mu(A)$$

for all $A \in \mathfrak{A}$. Also let $\mu \leq \nu$ if $\mu(A) \leq \nu(A)$ for all A in $\mathfrak{A}$. Under these definitions $\operatorname{ba}(S)$ is an ordered vector space.

CHAPTER 10

TOPOLOGICAL SPACES

Most of (real) analysis is a study of the interplay of algebra and topology. Thus far in these notes there has been quite a bit of algebra, but no topology. Now we begin to remedy this unhappy situation.

The classical part of topology which can be studied without the invocation of elaborate algebraic machinery is usually referred to as *point-set topology.* One excellent reference for this material is *General topology* [48] by Stephen Willard. A remarkable (and unusual) feature of this text is its thoughtfully designed and useful index. Another book I recommend highly is Albert Wilansky's *Topology for Analysis* [47]. The tables in the appendix, although a bit tricky to decipher until you get used to them, are invaluable in finding the relationships between an amazing 42 different topological properties that spaces may possess. I know of nothing like this in any other text. One of the first and best written topology texts is John L. Kelly's *General Topology* [27]. And, of course, there are many, many others—some quite good.

10.1. Definition of Topology

10.1.1. Definition. A family $\mathfrak{T}$ of subsets of a set X is a TOPOLOGY on X if

(i) $\emptyset$ and X belong to $\mathfrak{T}$;
(ii) if $\mathfrak{S} \subseteq \mathfrak{T}$, then $\bigcup \mathfrak{S} \in \mathfrak{T}$; and
(iii) if $\mathfrak{S}$ is a *finite* subfamily of $\mathfrak{T}$, then $\bigcap \mathfrak{S} \in \mathfrak{T}$.

The pair $(X, \mathfrak{T})$ is a TOPOLOGICAL SPACE. As is by now familiar, we will abuse language and say things like, "Let X be a topological space." If $(X, \mathfrak{T})$ is a topological space, the members of $\mathfrak{T}$ are OPEN SETS. We often use the notation $U \overset{\circ}{\subseteq} X$ to indicate that U is an open subset of X.

10.1.2. Remark. It is not unreasonable to regard condition (i) in the preceding definition as redundant. It depends entirely on the conventions you choose to make about the union and intersection of an empty family of sets. Recall the discussion of this matter in the cautionary note 1.4.10. According to the arguments given there condition (iii) guarantees that the empty set automatically belongs to every topology on S and condition (ii) guarantees that S itself belongs to every such topology.

The virtue of looking at things in this fashion is that it streamlines the definition of topology: a topology on a set S is just a family of subsets of S which is closed under unions and finite intersections. The price one pays for this slight convenience, as was mentioned in 1.4.10, is having to distinguish between the empty family $\mathfrak{F}$ of subsets of the set S and the empty family $\mathfrak{G}$ of subsets of a different set T. They are different because $\bigcap \mathfrak{F} = S$ while $\bigcap \mathfrak{G} = T$. Most authors avoid dealing explicitly with the problem by including condition (i) in their definition of *topology.*

If you have not previously encountered topological spaces, the preceding definition may make little sense to you at first reading. Don't fret. As was remarked in section 8.2 this kind of abstract definition, although easy enough to remember, can be irritatingly difficult to understand. Staring at it doesn't help. It appears that a bewildering array of entirely different things might turn out to be topologies, even on relatively simple sets. And this is indeed the case. Some turn out to be interesting and useful and some do not. An understanding and appreciation of the definition come only gradually. You will notice as you advance through this material that many important concepts such as continuity, compactness, and connectedness are defined by (or characterized by) open sets. Thus theorems which involve these ideas will rely on properties of open sets, that is, on topology, for their proofs.

10.1.3. Example. The most rudimentary topology one can place on a nonempty set X is the INDISCRETE TOPOLOGY in which the only open sets are the empty set and X itself. (Notice the spelling: there is nothing indiscreet about the indiscrete topology.)

10.1.4. Example. At the opposite extreme is the largest possible topology on X, the DISCRETE TOPOLOGY in which *every* set is open; that is, the topology on X is $\mathfrak{P}(X)$.

10.1.5. Example. Let $X = \{0, 1\}$ and $\mathfrak{T} = \{\emptyset, \{0\}, X\}$. The topological space $(X, \mathfrak{T})$, which is sometimes useful as a counterexample, is called the SIERPIŃSKI SPACE.

10.1.6. Example. Let X be an infinite set and $\mathfrak{T}$ be the family consisting of $\emptyset$ and of all sets $U \subseteq X$ such that U^c is finite. This is the COFINITE TOPOLOGY on X.

10.1.7. Definition. A subset of a topological space is CLOSED if its complement is open.

10.1.8. CAUTION. It is a common mistake to treat subsets of a topological space as if they were doors or windows, and to conclude, for example, that a set is closed because it is not open, or that it cannot be closed because it is open. These "conclusions" are wrong! When the set $\{1, 2, 3\}$ is given the indiscrete topology the subset $\{1, 2\}$ is neither open nor closed. When the same set is given the discrete topology $\{1, 2\}$ is both open and closed.

10.1.9. Example. In the Sierpiński space (see example 10.1.5) the closed sets are $\emptyset$, $\{1\}$, and X itself.

10.1.10. Example. If X is an infinite set with the cofinite topology (see example 10.1.6), then the closed subsets of X are the finite sets together with X itself.

10.1.11. Proposition. *If $\mathfrak{A}$ is a family of closed subsets of a topological space, then $\bigcap \mathfrak{A}$ is closed.*

10.1.12. Exercise. Let S be a subset of a topological space X. What, precisely, do we mean by "the smallest closed set containing S"? (If this isn't obvious, consult remark 5.3.6.)

10.1.13. Proposition. *If $\mathfrak{A}$ is a finite family of closed subsets of a topological space, then $\bigcup \mathfrak{A}$ is closed.*

10.1.14. Definition. If $\mathfrak{S}$ and $\mathfrak{T}$ are topologies on a set S and $\mathfrak{S} \subseteq \mathfrak{T}$, then we say that $\mathfrak{S}$ is a SMALLER (or WEAKER, or COARSER) topology than $\mathfrak{T}$. Similarly, $\mathfrak{T}$ is said to be a LARGER (or STRONGER, or FINER) topology than $\mathfrak{S}$. We WEAKEN a topology if we replace it by a smaller one; we STRENGTHEN a topology by replacing it with a larger one.

10.1.15. Exercise. Let $X = \{a, b, c, d, e, f, g\}$. What is the smallest topology on X under which the sets $\{a\}$, $\{a, c\}$, and $\{e, g\}$ are open and the set $\{a, c, e, g\}$ is closed? List all of the closed sets in the resulting topological space.

10.1.16. Exercise. Let $X = \{a, b, c\}$. How many distinct topologies can X be given? (If you are interested in a somewhat more challenging problem let $X = \{a, b, c, d\}$.

The next three examples assume that you are already acquainted with the usual topology on the real line. If this is not the case, defer them until you have read example 10.2.6.

10.1.17. Example. The set $\mathbb{Z}$ of integers is a closed subset of $\mathbb{R}$.

10.1.18. Example. The set $\{\frac{1}{n} \colon n \in \mathbb{N}\}$ is *not* a closed subset of $\mathbb{R}$.

10.1.19. Example. The set $\{\frac{1}{n} \colon n \in \mathbb{N}\} \cup \{0\}$ is a closed subset of $\mathbb{R}$.

10.1.20. Exercise. Exhibit an example to show that the intersection of an arbitrary family of open sets in a topological space need not be open. Also give an example to show that an arbitrary union of closed sets need not be closed.

10.1.21. Example. Any subset S of a topological space X is itself a topological space in a natural way. The family of all sets of the form $U \cap S$ where $U \overset{\circ}{\subseteq} X$ is a topology on S. In this case we say that S is a (TOPOLOGICAL) SUBSPACE of X. The family $\{U \cap S \colon U \overset{\circ}{\subseteq} X\}$ is the topology on S INDUCED BY the topology on X. Subsets of topological spaces are thus ordinarily regarded as topological space in their own right without further explanation.

10.1.22. Example. It is a routine matter to "transfer" a topology from a topological space X to a set S by means of a bijection. Suppose that $f \colon S \to X$ is a bijection. Then the set of all subsets of S of the form $f^{\leftarrow}(U)$ where $U \overset{\circ}{\subseteq} X$ is a topology on S. (Compare this with proposition 4.7.6.)

10.1.23. Example. It is frequently convenient to regard the set $\overline{\mathbb{R}}$ as a topological space. An easy way of doing this is to use the technique in the preceding example. Let f be the arctangent function on $\mathbb{R}$ and extend it to $\overline{\mathbb{R}}$ by setting $f(-\infty) = -\frac{\pi}{2}$ and $f(\infty) = \frac{\pi}{2}$. We will observe in example 11.1.9 that with this topology on $\overline{\mathbb{R}}$ the set $\mathbb{R}$ of ordinary real numbers is a subspace of $\overline{\mathbb{R}}$.

10.2. Base for a Topology

10.2.1. Definition. Let $(X, \mathfrak{T})$ be a topological space. A family $\mathfrak{B} \subseteq \mathfrak{T}$ is a BASE for $\mathfrak{T}$ if each member of $\mathfrak{T}$ is a union of members of $\mathfrak{B}$. In other words, a family $\mathfrak{B}$ of open sets is a base for $\mathfrak{T}$ if for each open set U there exists a subfamily $\mathfrak{B}'$ of $\mathfrak{B}$ such that $U = \bigcup \mathfrak{B}'$.

In practice it is often more convenient to specify a base for a topology than to specify the topology itself. It is important to realize, however, that there may be many different bases for the same topology. Once a particular base has been chosen we refer to its members as *basic open sets.*

10.2.2. Example. Let X be a nonempty set and $\mathfrak{B}$ be the collection of all one-point sets in X. That is, let $\mathfrak{B} = \big\{\{x\} \colon x \in X\big\}$. Then $\mathfrak{B}$ is a base for the discrete topology on X.

10.2.3. Proposition. *Let X be a topological space. A necessary and sufficient condition for a family $\mathfrak{B}$ of open subsets of X to be a base for the topology on X is that for every $U \overset{\circ}{\subseteq} X$ and every $x \in U$ there exists $B \in \mathfrak{B}$ such that $x \in B \subseteq U$.*

10.2.4. Corollary. *Let X be a set, $\mathfrak{B}$ be a base for a topology $\mathfrak{T}$ on X, and $\mathfrak{B}'$ be a base for a topology $\mathfrak{T}'$ on X. Then $\mathfrak{T}'$ is stronger than $\mathfrak{T}$ if and only if for every set $B \in \mathfrak{B}$ and every $x \in B$ there exists a set $B' \in \mathfrak{B}'$ such that $x \in B' \subseteq B$.*

Proposition 10.2.3 gives a necessary and sufficient condition for a family of open sets to be a base for some particular topology on a set. It is quite natural, however, to ask a somewhat more interesting question. Suppose we are given a collection $\mathfrak{B}$ of subsets of a set X: under what conditions is $\mathfrak{B}$ a base for *some* topology on X? An answer is given in the next proposition.

10.2.5. Proposition. *A family $\mathfrak{B}$ of subsets of a set X is a base for some topology on X if and only if*

(i) *$\bigcup \mathfrak{B} = X$ and*
(ii) *for every U, $V \in \mathfrak{B}$ and for every $x \in U \cap V$, there exists a set W in $\mathfrak{B}$ such that $x \in W \subseteq U \cap V$.*

10.2.6. Example. Since the family of all bounded open intervals (a, b) in the real line satisfies conditions (i) and (ii) of proposition 10.2.5, it is a base for some topology on the real line $\mathbb{R}$. This is the USUAL TOPOLOGY on $\mathbb{R}$. Under this topology a subset of $\mathbb{R}$ is open if and only if it is a union of a family of bounded open intervals. Thus, happily, open intervals (even ones of infinite length) turn out to be open sets, and closed intervals are closed sets. When we say that this is *the usual topology* for the real numbers we mean that unless the contrary is explicitly stated, whenever we regard $\mathbb{R}$ as a topological space it will be with respect to this topology.

10.2.7. Example. If $a \in \mathbb{Z}$ and $d \in \mathbb{N}$ then the set $\{a + nd \colon n \in \mathbb{Z}\}$ is an ARITHMETIC PROGRESSION. (For example, if $a = 2$ and $d = 5$, then the corresponding arithmetic progression is $\{\dots, -8, -3, 2, 7, 12, \dots\}$.)

(a) The family of all arithmetic progressions is a base for a topology on $\mathbb{Z}$.
(b) Every arithmetic progression is a closed set as well as an open set.
(c) It is possible and entertaining to give a topological proof that there are infinitely many prime numbers. *Hint.* Consider the set $A = \bigcup_p A_p$ where A_p is the arithmetic progression consisting of all integer multiples of a prime number p and the union is taken over the set of all prime numbers.

10.2.8. Proposition. *If A is a nonempty closed subset of $\mathbb{R}$ which is bounded above, then* $\sup A$ *is a member of A.*

10.3. Some Elementary Topological Properties

10.3.1. Definition. A topological space is SECOND COUNTABLE if there exists a countable base for its topology.

10.3.2. Example. The topological space $\mathbb{R}$ (with its usual topology) is second countable. *Hint.* Consider bounded open intervals with rational endpoints. Use example 3.2.12 and proposition 3.2.9.

10.3.3. Example. A discrete topological space is second countable if and only if it is countable.

10.3.4. Definition. Two nonempty subsets A and B of a topological space X can be SEPARATED BY OPEN SETS if there exists open subsets U and V of X such that $A \subseteq U$, $B \subseteq V$, and $U \cap V = \emptyset$. (And of course we say that two points a and b can be *separated by open sets* if $\{a\}$ and $\{b\}$ can be.)

10.3.5. Example. In the real line $\mathbb{R}$ the sets $[-1, 0)$ and $(0, 1]$ can be separated by open sets; the sets $[-1, 0]$ and $(0, 1]$ can not.

10.3.6. Definition. A HAUSDORFF topological space is one in which every pair of distinct points can be separated by open sets.

10.3.7. Definition. If U is an open subset of a topological space X and $x \in U$ we say that U is a NEIGHBORHOOD of x.

Thus a topological space X is Hausdorff if for any two distinct points in X we can find disjoint neighborhoods.

10.3.8. Example. Every discrete topological space is Hausdorff.

10.3.9. Example. No space with the indiscrete topology having at least two points is Hausdorff.

10.3.10. Example. An infinite set with the cofinite topology is *not* Hausdorff.

10.3.11. Example. The real line (with its usual topology)is Hausdorff.

10.3.12. Example. The Sierpiński space 10.1.5 is *not* Hausdorff.

10.3.13. Exercise. Is the space in exercise 10.1.15 Hausdorff?

10.4. Metric Spaces

Most of the topological spaces appearing in these notes are metric spaces—that is, spaces whose topology is determined by a notion of *distance*. Although there is a bewildering variety of metric spaces, it is also true that they share some simple and pleasant properties, properties so striking that it seems natural to regard them collectively as a single class of examples of topological spaces. Thus the present section is little more than an extended list of examples of an example. Metric spaces turn out to be among the nicest inhabitants of the unruly world of general topological spaces. More details on these spaces can be found in my ProblemText [17].

10.4.1. Definition. Let M be a nonempty set. A function $d\colon M \times M \to \mathbb{R}$ is a METRIC (or DISTANCE FUNCTION) on M if for all x, y, $z \in M$

(1) $d(x,y) = d(y,x)$,
(2) $d(x,y) \le d(x,z) + d(z,y)$, (the *triangle inequality*)
(3) $d(x,x) = 0$, and
(4) $d(x,y) = 0$ only if $x = y$.

If d is a metric on a set M, then we say that the pair (M,d) is a METRIC SPACE. The notation "$d(x,y)$" is read, "the distance between x and y". (As usual we substitute the phrase, "Let M be a metric space ..." for the correct formulation "Let (M,d) be a metric space ...".

If $d\colon M \times M \to \mathbb{R}$ satisfies conditions (1)–(3), but not (4), then d is a PSEUDOMETRIC on M.

10.4.2. Definition. If A and B are subsets of a metric space M, then the DISTANCE BETWEEN A AND B, denoted by $d(A,B)$, is defined to be $\inf\{d(a,b)\colon a \in A \text{ and } b \in B\}$.

10.4.3. Proposition. *If d is a metric on a set M, then $d(x,y) \ge 0$ for all x, $y \in M$.*

10.4.4. Proposition. *For all points x, y, and z in a metric space (M,d)*

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

10.4.5. Definition. For each point a in a metric space (M,d) and each number $r > 0$ we define $B_r(a)$, the OPEN BALL about a of radius r, to be the set of all those points whose distance from a is strictly less than r. That is,

$$B_r(a) := \{x \in M\colon d(x,a) < r\}\,.$$

We also define $C_r(a)$, the CLOSED BALL about a of radius r, by

$$C_r(a) := \{x \in M\colon d(x,a) \le r\}\,.$$

and $S_r(a)$, the SPHERE about a of radius r, by

$$S_r(a) := \{x \in M\colon d(x,a) = r\}\,.$$

When $r = 1$ we have the OPEN UNIT BALL (or CLOSED UNIT BALL or UNIT SPHERE) about a. When the underlying set M is also a vector space we sometimes refer to *the* unit ball in M, by which we mean *the open unit ball about* $\mathbf{0}$ in M.

10.4.6. Example. The absolute value of the difference of two numbers is a metric on $\mathbb{R}$. We will call this the *usual metric* on $\mathbb{R}$. In this metric the open ball about a of radius r is just the open interval $(a - r, a + r)$.

10.4.7. Exercise. Define $d(x,y) = |\arctan x - \arctan y|$ for all $x, y \in \mathbb{R}$.

(a) Show that d is a metric on $\mathbb{R}$.
(b) Find $d(-1, \sqrt{3})$.
(c) Solve the equation $d(x,0) = d(x,\sqrt{3})$.

10.4.8. Example. Define $d(x,y) = |x^2 - y^2|$ for all x, $y \in \mathbb{R}$. Then d is a pseudometric but *not* a metric on $\mathbb{R}$.

10.4.9. Example. Let M be any nonempty set. For x, $y \in M$ define

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

It is easy to see that d is a metric; this is the DISCRETE METRIC on M. Although the discrete metric is rather trivial it proves quite useful in constructing counterexamples.

10.4.10. Proposition. *If M is a metric space, $a \in M$, and $0 < r < s$, then $B_r(a) \subseteq B_s(a)$.*

10.4.11. Example. For points $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$ let

$$d(x, y) := \left(\sum_{k=1}^{n} (x_k - y_k)^2 \right)^{\frac{1}{2}} .$$

Then d is the USUAL (or EUCLIDEAN) METRIC on $\mathbb{R}^n$. The only nontrivial part of the proof that d is a metric is the verification of the *triangle inequality* (that is, condition (2) of the definition):

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

To accomplish this let $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, and $z = (z_1, \dots, z_n)$ be points in $\mathbb{R}^n$. Apply *Minkowski's inequality* (6.5.2) with $u_k = x_k - z_k$ and $v_k = z_k - y_k$ for $1 \le k \le n$ to obtain

$$\begin{aligned} d(x, y) &= \left(\sum_{k=1}^{n} (x_k - y_k)^2 \right)^{\frac{1}{2}} \\ &= \left(\sum_{k=1}^{n} \big((x_k - z_k) + (z_k - y_k) \big)^2 \right)^{\frac{1}{2}} \\ &\le \left(\sum_{k=1}^{n} (x_k - z_k)^2 \right)^{\frac{1}{2}} + \left(\sum_{k=1}^{n} (z_k - y_k)^2 \right)^{\frac{1}{2}} \\ &= d(x, z) + d(z, y) . \end{aligned}$$

The Euclidean metric is by no means the only metric on $\mathbb{R}^n$ which is useful. Two more examples follow (10.4.12 and 10.4.13).

10.4.12. Example. For points $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$ let

$$d_1(x, y) := \sum_{k=1}^{n} |x_k - y_k|.$$

It is easy to see that d_1 is a metric on $\mathbb{R}^n$. When $n = 2$ this is frequently called the TAXICAB METRIC. (Why?)

10.4.13. Example. For $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$ let

$$d_u(x, y) := \max\{|x_k - y_k| : 1 \le k \le n\}.$$

Then d_u is a metric on $\mathbb{R}^n$. The triangle inequality is verified as follows:

$$\begin{aligned} |x_k - y_k| &\le |x_k - z_k| + |z_k - y_k| \\ &\le \max\{|x_i - z_i| : 1 \le i \le n\} + \max\{|z_i - y_i| : 1 \le i \le n\} \\ &= d_u(x, z) + d_u(z, y) \end{aligned}$$

whenever $1 \le k \le n$. Thus

$$\begin{aligned} d_u(x, y) &= \max\{|x_k - y_k| : 1 \le k \le n\} \\ &\le d_u(x, z) + d_u(z, y) . \end{aligned}$$

The metric d_u is called the UNIFORM METRIC on $\mathbb{R}^n$. The reason for this name will become clear later (see the comment after the verification of example 10.4.32).

Notice that on the real line the three immediately preceding metrics agree; the distance between points is just the absolute value of their difference. That is, when $n = 1$ the metrics given in examples 10.4.11, 10.4.12, and 10.4.13 reduce to the one given in example 10.4.6.

10.4.14. Remark. A crucial observation is that *every metric space is a topological space*. More precisely, we use proposition 10.2.5 to show that the family of open balls in a metric space M is a base for a topology on M. Unless the contrary is explicitly stated we will always regard metric spaces as being topological spaces under this topology, which we refer to as the *topology induced by the metric*. Since condition (i) of proposition 10.2.5 obviously holds for the family of all open balls in M, we need only verify condition (ii).

10.4.15. Proposition. *Let a and b be points in a metric space M and r, $s > 0$. If c belongs to $B_r(a) \cap B_s(b)$, then there exists a number $t > 0$ such that $B_t(c) \subseteq B_r(a) \cap B_s(b)$.*

10.4.16. Corollary. *Every metric space (M, d) is a topological space $(M, \mathfrak{T})$. The open balls of (M, d) are a base for the topology $\mathfrak{T}$.*

10.4.17. Remark. Assertions like the preceding corollary are very common in mathematics. Literally the claim made there is nonsense: a set together with a distance function is surely not the same thing as a set together with a family of subsets closed under certain operations. On the other hand it does say slightly more than, "Every metric space (M, d) induces a topology on M." It is taken to mean, additionally, that whenever we encounter a metric space M we will identify it with its induced topological space in the sense that references to topological properties—open, closed, compact, connected, and so on—of subsets of M will be regarded as meaningful without any explanation. In circumstances when you wish to consider a topology on a metric space other than the one induced by the metric, you must be quite explicit about what you are doing.

10.4.18. Example. It is clear from the preceding proposition 10.4.15 that every open ball in a metric space is an open set. Furthermore, every closed ball is a closed set and every sphere is a closed set.

10.4.19. Example. The space $\mathbb{R}^n$ with its usual (Euclidean) metric is a second countable topological space.

10.4.20. Example. The topology induced by the discrete metric on a set M (see example 10.4.9) is the discrete topology on M (see example 10.1.4).

10.4.21. Example. The topology induced by the usual metric on the set $\mathbb{R}$ (see example 10.4.11) is the usual topology on $\mathbb{R}$ (see example 10.2.6).

The next proposition shows that as a topological space every metric space is Hausdorff.

10.4.22. Proposition. *If a and b are distinct points in a metric space, then there exists a number $r > 0$ such that $B_r(a)$ and $B_r(b)$ are disjoint.*

Ahead of us lie many situations in which it will be possible to replace a computationally complicated metric on some space by a simpler one without affecting any topological property of the space. Below (in proposition 10.4.26) we give a sufficient condition for this process to work.

10.4.23. Definition. Two metrics d_1 and d_2 on a set M are EQUIVALENT if they induce the same topology on M.

10.4.24. Definition. Two metrics d_1 and d_2 on a set M are STRONGLY EQUIVALENT if there exist numbers $\alpha, \beta > 0$ such that

$$\begin{aligned} d_1(x, y) &\le \alpha\, d_2(x, y) \quad \text{and} \\ d_2(x, y) &\le \beta\, d_1(x, y) \end{aligned}$$

for all x and y in M.

10.4.25. Example. On $\mathbb{R}^2$ the three metrics d, d_1, and d_u, defined in examples 10.4.11, 10.4.12, and 10.4.13, are all strongly equivalent.

Hint for proof. First prove that if a, $b \ge 0$, then

$$\max\{a, b\} \le a + b \le \sqrt{2}\,\sqrt{a^2 + b^2} \le 2 \max\{a, b\}\,.$$

10.4.26. Proposition. *If d and ρ are strongly equivalent metrics on a set M, then they are equivalent.*

Hint for proof. Use corollary 10.2.4.

10.4.27. Definition. A metric d on a set M is BOUNDED if $\sup\{d(x,y)\colon x,y\in M\}<\infty$.

10.4.28. Example. Let (M,d) be a metric space. The function ρ defined on $M\times M$ by

$$\rho(x,y)=\frac{d(x,y)}{1+d(x,y)}$$

is a bounded metric on M.

Hint for proof. Show first that $\dfrac{u}{1+u}\le\dfrac{v}{1+v}$ whenever $0\le u\le v$.

10.4.29. Proposition. *Every metric is equivalent to a bounded metric.*

Hint for proof. Use corollary 10.2.4.

10.4.30. Example. In example 10.4.28 take M to be the real line $\mathbb{R}$ and d to be the usual metric on $\mathbb{R}$ (see 10.4.6). Then the metrics d and ρ are equivalent but *not* strongly equivalent on $\mathbb{R}$.

10.4.31. Definition. A subset B of a metric space M is BOUNDED if it is contained in some open ball. A function $f\colon S\to M$ from a set into a metric space is BOUNDED if its range is a bounded subset of M. (In particular, a sequence (x_n) of points in a metric space is bounded if its range $\{x_n\colon n\in\mathbb{N}\}$ is a bounded subset of the space.) The family of all bounded functions from S into M is denoted by $\mathcal{B}(S,M)$.

10.4.32. Example. Let S be a nonempty set and (M,ρ) be a metric space. For f, $g\in\mathcal{B}(S,M)$ define

$$d(f,g)=\sup\{\rho\big(f(x),g(x)\big)\colon x\in S\}\,.$$

Then d is a metric on $\mathcal{B}(S,M)$.

PROOF. The verification of this is not difficult. Establishing the triangle inequality however does require some care. To this end suppose that f, g, and h belong to $\mathcal{B}(S,M)$. The inequality

$$\rho\big(f(x),h(x)\big)\le\rho\big(f(x),g(x)\big)+\rho\big(g(x),h(x)\big)$$

obviously holds for all $x\in S$, and therefore

$$\sup\{\rho\big(f(x),h(x)\big)\colon x\in S\}\le\sup\{\rho\big(f(x),g(x)\big)+\rho\big(g(x),h(x)\big)\colon x\in S\}\,. \tag{10.1}$$

Next notice that

$$\{\rho\big(f(x),g(x)\big)+\rho\big(g(x),h(x)\big)\colon x\in S\}\subseteq\{\rho\big(f(x),g(x)\big)+\rho\big(g(y),h(y)\big)\colon x,y\in S\}\,.$$

Then from exercise 5.2.3 we see that

$$\begin{aligned}\sup\{\rho\big(f(x),g(x)\big)&+\rho\big(g(x),h(x)\big)\colon x\in S\}\\&\le\sup\{\rho\big(f(x),g(x)\big)+\rho\big(g(y),h(y)\big)\colon x,y\in S\}\,.\end{aligned} \tag{10.2}$$

Finally, it follows from exercise 6.1.9 that

$$\begin{aligned}\sup\{\rho\big(f(x),g(x)\big)&+\rho\big(g(y),h(y)\big)\colon x,y\in S\}\\&=\sup\{\rho\big(f(x),g(x)\big)\colon x\in S\}+\sup\{\rho\big(g(y),h(y)\big)\colon y\in S\}\end{aligned} \tag{10.3}$$

Putting inequalities (10.1), (10.2), and (10.3) together produces the desired result

$$d(f,h)\le d(f,g)+d(g,h)\,.$$

□

This example is particularly important in the special cases where $M=\mathbb{R}$ or $M=\mathbb{C}$ and $\rho(p,q)=|p-q|$ for p, $q\in\mathbb{R}$ (or $\mathbb{C}$). Then d is the UNIFORM METRIC on $\mathcal{B}(S,\mathbb{R})$ (or $\mathcal{B}(S,\mathbb{C})$). Notice that example 10.4.13 is a special case of this. (An n-tuple may be regarded as a function on the set $\mathbb{N}_n=\{1,\dots,n\}$. Thus $\mathbb{R}^n=\mathcal{B}(\mathbb{N}_n,\mathbb{R})$.)

10.4.33. Definition. A sequence (x_n) in a topological space X is CONVERGENT if it is eventually in every neighborhood of some point a in X. In this case we say that the sequence (x_n) CONVERGES TO a and write $x_n \to a$ as $n \to \infty$. In cases where limits are unique we say that a is *the* LIMIT of the sequence and we use the notation $a = \lim_{n\to\infty} x_n$. These notations are often shortened to $x_n \to a$ or $a = \lim x_n$.

10.4.34. Proposition. *Limits of sequences in metric spaces are unique (if they exist).*

10.4.35. Example. Limits of sequences in topological spaces need not be unique (even when they exist).

10.5. Interiors and Closures

The preceding section was concerned with an important class of examples of topological spaces—the metric spaces. We now return to general topological spaces and investigate a few of their most elementary properties.

10.5.1. Definition. Let A be a subset of a topological space X. A point x is an INTERIOR POINT of A if some neighborhood of x lies entirely in A. The INTERIOR of A, denoted by A°, is the set of all interior points of A. That is,

$$A^\circ := \{x \in M \colon \text{ there exists } U \overset{\circ}{\subseteq} M \text{ such that } x \in U \subseteq A\}.$$

10.5.2. Example. If $\mathbb{R}$ has its usual topology and A is the closed interval $[0,1]$, then $A^\circ = (0,1)$.

10.5.3. Example. Let A be the closed unit disk $\{(x,y) \colon x^2 + y^2 \leq 1\}$ in $\mathbb{R}^2$. Then the interior of A is the open disk $\{(x,y) \colon x^2 + y^2 < 1\}$.

10.5.4. Example. Let $A = \mathbb{Q}^2 \subseteq \mathbb{R}^2$. Then $A^\circ = \emptyset$.

10.5.5. Proposition. *A subset U of a topological space is open if and only if $U = U^\circ$.*

10.5.6. Proposition. *Let A and B be subsets of a topological space.*

(a) *If $A \subseteq B$, then $A^\circ \subseteq B^\circ$.*
(b) $A^{\circ\circ} = A^\circ$. (As you would expect, $A^{\circ\circ}$ means $\big(A^\circ\big)^\circ$.)
(c) $(A \cap B)^\circ = A^\circ \cap B^\circ$.

10.5.7. Proposition. *Let $\mathcal{A}$ be a family of subsets of a topological space. Then*

$$\big(\bigcap \mathcal{A}\big)^\circ \subseteq \bigcap\{A^\circ \colon A \in \mathcal{A}\} \subseteq \bigcup\{A^\circ \colon A \in \mathcal{A}\} \subseteq \big(\bigcup \mathcal{A}\big)^\circ .$$

10.5.8. Example. Give examples to show that in the preceding proposition each containment may be proper.

Next we give a sequential characterization of *interior* in metric spaces.

10.5.9. Proposition. *A point a in a metric space M is in the interior of a subset S of M if and only if every sequence in M that converges to a is eventually in S.*

10.5.10. Definition. A point p in a topological space X is an ACCUMULATION POINT of a set $A \subseteq X$ if every neighborhood of p contains a point of A distinct from p. (We do *not* require that p belong to A.) We denote the set of all accumulation points of A by A'. This is sometimes called the DERIVED SET of A. The CLOSURE of the set A, denoted by $\overline{A}$, is $A \cup A'$.

10.5.11. CAUTION. Some authors prefer to work with *limit points* rather than accumulation points. The concepts are similar; a point p is a LIMIT POINT of a set A if every neighborhood of p contains a point of A. Thus if, for example, $A = \{0\} \cup [1,2]$ and $p = 0$, then p is a limit point of A but is not an accumulation point of A. The matter is further complicated by the fact that some authors, see [23] for example, define *limit point* as I have defined *accumulation point*. Additionally. the notation A' is occasionally (again see [23]) to denote the complement of a set A.

10.5.12. Example. Let $\mathbb{R}^2$ have its usual metric and A be $\big[(0,1)\times(0,1)\big]\cup\{(2,3)\}\subseteq\mathbb{R}^2$. Then $A'=[0,1]\times[0,1]$ and $\overline{A}=\big([0,1]\times[0,1]\big)\cup\{(2,3)\}$.

The following proposition is a slight refinement of 10.2.8.

10.5.13. Proposition. *If $A\subseteq\mathbb{R}$ is bounded above, then* $\sup A$ *belongs to* $\overline{A}$.

10.5.14. Proposition. *A subset F of a topological space is closed if and only if $F=\overline{F}$.*

10.5.15. Definition. A subset A of a topological space X is DENSE in X if $\overline{A}=X$.

10.5.16. Example. The rational numbers are dense in the reals.

Hint for proof. Use exercise 6.1.13.

10.5.17. Example. The set $\mathbb{Q}^2$ is a subset of the topological space $\mathbb{R}^2$. Every ordered pair of real numbers is an accumulation point of $\mathbb{Q}^2$ since every open ball in $\mathbb{R}^2$ contains (infinitely many) points with both coordinates rational. So the closure of $\mathbb{Q}^2$ is all of $\mathbb{R}^2$. That is, $\mathbb{Q}^2$ is dense in $\mathbb{R}^2$.

10.5.18. Proposition. *Let A be a subset of a topological space. Then*

(a) $\big(A^\circ\big)^c=\overline{A^c}$.
(b) $\big(A^c\big)^\circ=\big(\overline{A}\big)^c$.

The relationship between interiors and closures given in proposition 10.5.18 makes the next two results quite routine.

10.5.19. Proposition. *Let A and B be subsets of a topological space. Then*

(a) *If $A\subseteq B$, then $\overline{A}\subseteq\overline{B}$.*
(b) $\overline{\overline{A}}=\overline{A}$.
(c) $\overline{A\cup B}=\overline{A}\cup\overline{B}$.

10.5.20. Proposition. *Let $\mathcal{A}$ be a family of subsets of a topological space. Then*

$$\overline{\bigcap\mathcal{A}}\subseteq\bigcap\{\overline{A}\colon A\in\mathcal{A}\}\subseteq\bigcup\{\overline{A}\colon A\in\mathcal{A}\}\subseteq\overline{\bigcup\mathcal{A}}.$$

10.5.21. Example. Give examples to show that in the preceding proposition each containment may be proper.

As is the case with interiors, closures can be characterized in metric spaces by means of sequences. (See proposition 10.5.9.)

10.5.22. Proposition. *A point b in a metric space M belongs to the closure of a subset S of M if and only if some sequence in S converges to b.*

10.5.23. Proposition. *In a metric space the closure $\overline{B_r(a)}$ of an open ball is always contained in the corresponding closed ball $C_r(a)$.*

10.5.24. Example. The closure of an open ball and the corresponding closed ball need not be equal. For example in the reals with the discrete metric $\overline{B_1(0)}$ is not the same as $C_1(0)$.

10.5.25. Definition. If A is a subset of a topological space X, then the BOUNDARY of A denoted by ∂A is defined to be $\overline{A}\cap\overline{A^c}$.

10.5.26. Proposition. *If A is a subset of a topological space X, then*

(a) $\overline{A}=A^\circ\uplus\partial A=A\cup\partial A$ *and*
(b) $A^\circ=A\setminus\partial A=\overline{A}\setminus\partial A$.

CHAPTER 11

CONTINUITY AND WEAK TOPOLOGIES

In this chapter we begin an examination of the category **TOP** whose objects are topological spaces and whose morphisms are continuous functions.

11.1. Continuity—the Global Property

Continuity has come a long way. It appeared in the philosophical arguments of Zeno, who was born about 490 BC in Elea (now in southern Italy). By the seventeenth century the concept had taken on a quite definite mathematical flavor. Leibniz (born in 1646 in Leipzig) formulated his *lex continui* (law of continuity) in quite a few different ways. One version is, "nature never makes leaps". Another, more sophisticated, version is (see [29],p. 539)

> When the difference between two instances in a given series or that which is presupposed can be diminished until it becomes smaller than any given quantity whatever, the corresponding difference in what is sought or in their results must of necessity also be diminished or become less than any given quantity whatever.

It is surely not a great leap from this formulation (from 1687) to the familiar ϵ-δ definition in beginning calculus (see remark 11.2.2). Another two and a half centuries of reflection on and tinkering with the concept lead to the currently favored (and quite general) definition: a function is *continuous* if the inverse image of every open set is open.

11.1.1. Definition. A function $f\colon X \to Y$ between topological spaces is CONTINUOUS if $f^{\leftarrow}(U)$ is an open subset of X whenever U is an open subset of Y. We denote by $\mathcal{C}(X,Y)$ the family of all continuous functions mapping the topological space X into the space Y. For real valued functions we usually abbreviate $\mathcal{C}(X,\mathbb{R})$ to $\mathcal{C}(X)$.

In order to see that we obtain a category by declaring topological spaces to be objects and continuous maps to be morphisms, it is necessary to know that the identity function on a topological space is continuous (which is obvious) and that the composite of continuous functions is continuous.

11.1.2. Proposition. *If $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous mappings between topological spaces, then the composite function $g \circ f\colon X \to Z$ is continuous.*

PROOF. If $U \overset{\circ}{\subseteq} Z$, then, since g is continuous, the set $g^{\leftarrow}(U)$ is open in Y. Then, similarly, $(g \circ f)^{\leftarrow} = f^{\leftarrow}(g^{\leftarrow}(U))$ is an open subset of X (see exercise 2.5.4(b)). □

Thus the class of topological spaces together with the continuous functions between them constitute a category. We denote it by **TOP**. The isomorphisms in this category are traditionally called HOMEOMORPHISMS.

11.1.3. Example. Every constant function between topological spaces is continuous.

11.1.4. Example. Let Y be a subspace of a topological space X. The inclusion map $\iota\colon Y \to X$ is continuous.

11.1.5. Example. Any restriction of a continuous function is continuous.

Hint for proof. Use example 11.1.4 and proposition 11.1.2.

11.1.6. Example. Let X be a nonempty set with the discrete topology and Y be an arbitrary topological space. Then $\mathcal{C}(X,Y) = \mathcal{F}(X,Y)$. That is, every function from X to Y is continuous.

11.1.7. Example. Let X be a nonempty set with the indiscrete topology and Y be a Hausdorff topological space. Then the only continuous functions from X to Y are the constant functions.

11.1.8. Example. Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$. Then f is not a continuous function.

11.1.9. Example. Assume that we know that the arctangent function and the tangent function (restricted to the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$) are both continuous. Then it is clear that our construction of the topology on the extended real numbers (see example 10.1.23 makes the set of ordinary reals $\mathbb{R}$ a subspace of $\overline{\mathbb{R}}$. The collection $\mathfrak{B}$ of all open intervals in $\overline{\mathbb{R}}$ is a base for the topology on $\overline{\mathbb{R}}$. (The open intervals in $\overline{\mathbb{R}}$ are of four types: $[-\infty, b)$, (a, b), $(a, \infty]$, and, of course, $\overline{\mathbb{R}}$ itself.)

11.1.10. Example. The identity map $\mathrm{id}_{\mathbb{R}}$ from the real line with the discrete topology to the real line with its usual topology is an example which shows that a bijective morphism in the category **TOP** need not be an isomorphism. That is, continuous bijections between topological spaces need not be homeomorphisms.

11.1.11. Definition. Let $(X, \mathfrak{T})$ be a topological space. A subfamily $\mathfrak{S} \subseteq \mathfrak{T}$ is a SUBBASE for the topology $\mathfrak{T}$ if the family of all finite intersections of members of $\mathfrak{S}$ is a base for $\mathfrak{T}$.

We saw in proposition 10.2.5 that not every family of subsets of a set S can be used as a base for a topology on S. The situation is much more agreeable in the case of subbases.

11.1.12. Proposition. *Any nonempty family of subsets of a set S is a subbase for some topology on S. (If you do not approve of the conventions mentioned in 1.4.10 and 10.1.2, add the hypothesis that the family contains $\emptyset$ and S.)*

11.1.13. Proposition. *Let X and Y be topological spaces and $f\colon X \to Y$.*

(a) *Let $\mathfrak{B}$ be a base for the topology on Y. Then f is continuous if and only if $f^{\leftarrow}(B) \overset{\circ}{\subseteq} X$ for every $B \in \mathfrak{B}$.*
(b) *Let $\mathfrak{S}$ be a subbase for the topology on Y. Then f is continuous if and only if $f^{\leftarrow}(S) \overset{\circ}{\subseteq} X$ for every $S \in \mathfrak{S}$.*

11.1.14. Example. The function $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto -2x + 3$ is continuous.

Hint for proof. Recall from example 10.2.6 that the set of bounded open intervals is a base for the usual topology on $\mathbb{R}$.

11.1.15. Example. The map $x \mapsto x^{1/3}$ from $\mathbb{R}$ into $\mathbb{R}$ is continuous.

11.1.16. Example. The map $x \mapsto |x|$ from $\mathbb{R}$ into $\mathbb{R}$ is continuous.

11.1.17. Proposition. *A function $f\colon X \to Y$ between topological spaces is continuous if and only if the inverse image of every closed subset of Y is closed in X.*

11.1.18. Proposition. *A function $f\colon X \to Y$ between topological spaces is continuous if and only if $f^{\rightarrow}(\overline{A}) \subseteq \overline{f^{\rightarrow}(A)}$ for every subset A of X.*

11.2. Continuity—the Local Property

11.2.1. Definition. Let X and Y be topological spaces and $a \in X$. A function $f\colon X \to Y$ is CONTINUOUS AT a if $f^{\leftarrow}(V)$ contains a neighborhood of a whenever V is a neighborhood of $f(a)$.

11.2.2. Remark. If in the preceding definition $\mathfrak{B}$ is a base for the topology on Y, it is easy to see that it is sufficient to require that $f^{\leftarrow}(V)$ contain a neighborhood of a for each set V in $\mathfrak{B}$ which contains $f(a)$. It follows that if (M, d) is a metric space, then a function $f\colon X \to M$ is continuous at a if and only if $f^{\leftarrow}\big(B_\epsilon(f(a))\big)$ contains a neighborhood of a for each $\epsilon > 0$. Put differently: f is continuous at a if and only if for every $\epsilon > 0$ there exists a neighborhood U of a such that $x \in U$ implies $d(f(x), f(a)) < \epsilon$.

If $f\colon M_1 \to M_2$ where both M_1 and M_2 are metric spaces, then f is continuous at a if and only if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$f^{\leftarrow}\big(B_\epsilon(f(a))\big) \supseteq B_\delta(a)\,.$$

Another way of saying exactly the same thing: f is continuous at a if and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that $d_2(f(x), f(a)) < \epsilon$ whenever $d_1(x, a) < \delta$.

Specializing further to the case where f is a real valued function of a real variable we have the usual beginning calculus definition of continuity at a point: f is continuous at a point a in its domain if and only if for every $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$.

Of course, the relationship between the local and global properties of continuity is exactly what one would expect.

11.2.3. Proposition. *A function $f\colon X \to Y$ between topological spaces is continuous if and only if it is continuous at each point of X.*

11.2.4. Example. Clearly $\mathbb{C} \setminus \{0\}$ is a metric space when the distance between two nonzero numbers w and z is $|w - z|$. The map $z \mapsto 1/z$ of $\mathbb{C} \setminus \{0\}$ into itself is continuous.

11.2.5. Example. The map $z \mapsto \overline{z}$ from $\mathbb{C}$ into itself is continuous.

Here is a sequential characterization of *continuity* for functions between metric spaces. An improved version holding for general topological spaces will appear in proposition 16.4.24.

11.2.6. Proposition. *Let M and N be metric spaces. A function $f\colon M \to N$ is continuous at a point a in M if and only if $f(x_n) \to f(a)$ in N whenever (x_n) is a sequence converging to a in M.*

11.2.7. Proposition. *Let X and Y be topological spaces and suppose that Y is Hausdorff. Suppose also that f, $g\colon X \to Y$ are continuous functions which agree on a dense subset D of X. (That is, suppose that $f(x) = g(x)$ for all $x \in D$.) Then $f = g$.*

11.2.8. Proposition. *Let $f\colon X \to \mathbb{R}$ be a continuous function on a topological space X and $a \in X$. If $f(a) > 0$, then there exist a neighborhood U of a and a number $b > 0$ such that $f(x) > b$ for every $x \in U$.*

11.3. Uniform Continuity

11.3.1. Definition. Let (M, d) and (N, ρ) be metric spaces. A function $f\colon M \to N$ is UNIFORMLY CONTINUOUS if for every $\epsilon > 0$ there exists $\delta > 0$ such that $\rho(f(x), f(y)) < \epsilon$ whenever x, $y \in M$ and $d(x, y) < \delta$.

11.3.2. Example. The function $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto 3x - 4$ is uniformly continuous.

11.3.3. Example. The function $f\colon [1, \infty) \to \mathbb{R}\colon x \mapsto x^{-1}$ is uniformly continuous.

11.3.4. Example. The function $g\colon (0, 1] \to \mathbb{R}\colon x \mapsto x^{-1}$ is not uniformly continuous.

11.3.5. Example. Let M be an arbitrary positive number. The function

$$f\colon [0, M] \to \mathbb{R}\colon x \mapsto x^2$$

is uniformly continuous. Prove this assertion using only the definition of "uniform continuity". Then show that the function

$$g\colon [0, \infty) \to \mathbb{R}\colon x \mapsto x^2$$

is *not* uniformly continuous.

11.3.6. Example. The function $g\colon [0, \infty) \to \mathbb{R}\colon x \mapsto \sqrt{x}$ is uniformly continuous.

11.3.7. Proposition. *Every uniformly continuous function is continuous.*

11.4. Weak Topologies

Here is a situation which frequently crops up in analysis. We are given an indexed family $(f_\alpha)_{\alpha\in A}$ of functions from a set S into a corresponding family $(X_\alpha)_{\alpha\in A}$ of topological spaces . What we want to do is put a topology on S in such a way that

(1) each of the functions f_α is continuous, and
(2) if g is a function from a topological space Y into S such that the composite functions $f_\alpha \circ g$ are continuous, then g is continuous.

$$Y \xrightarrow{\;g\;} S \xrightarrow{\;f_\alpha\;} X_\alpha$$

Certainly condition (1) by itself is easy enough to satisfy: just give S the discrete topology. The trouble with this solution to our problem is that the discrete topology may be too large to satisfy (2). To see this look at a simple example where S is the *set* of real numbers and there is only one function f, the identity map from the set $\mathbb{R}$ to the topological space $\mathbb{R}$ (with its usual topology). Give $S = \mathbb{R}$ the discrete topology (and denote it by $\mathbb{R}_d$). Now consider what happens if we take $Y = \mathbb{R}$ (again with its usual topology) and g to be the identity map.

$$\mathbb{R} \xrightarrow{\;\text{id}\;} \mathbb{R}_d \xrightarrow{\;\text{id}\;} \mathbb{R}$$

Clearly the composite of the two identity functions is continuous, but the one on the left of the diagram is not.

So the problem confronting us is to choose a topology for S which is large enough to make each f_α continuous and at the same time small enough to allow g to be continuous. The solution to this problem turns out to be pleasantly simple: use the *smallest* topology on S which makes the f_α's continuous. How is this accomplished? Easily: all we need to guarantee is that for every α the inverse image under f_α of each open subset of X_α is open in S. So take the family of these sets as a subbase for a topology on S.

11.4.1. Definition. Suppose that S is a set, that for every $\alpha \in A$ (where A is an arbitrary index set) X_α is a topological space, and that $f_\alpha\colon S \to X_\alpha$ for every $\alpha \in A$. Let

$$\mathfrak{S} = \{{f_\alpha}^{\leftarrow}(U_\alpha)\colon U_\alpha \overset{\circ}{\subseteq} X_\alpha \text{ and } \alpha \in A\}\,.$$

Use the family $\mathfrak{S}$ as a subbase for a topology on S. This topology is called the WEAK TOPOLOGY induced by (or determined by) the functions f_α.

11.4.2. Exercise. In definition 11.4.1 the family $\{{f_\alpha}^{\leftarrow}(U_\alpha)\colon U_\alpha \overset{\circ}{\subseteq} X_\alpha \text{ and } \alpha \in A\}$ was used for a subbase for the weak topology on S. Explain briefly why we could have used a smaller family of sets to generate the same topology. For example, show that if $\mathfrak{B}_\alpha$ is a base for the topology on X_α, then $\{{f_\alpha}^{\leftarrow}(B_\alpha)\colon B_\alpha \in \mathfrak{B}_\alpha \text{ and } \alpha \in A\}$ is a subbase for the weak topology on S. Show that the preceding sentence remains true even if the word "subbase" is substituted for "base".

The next two propositions verify that we have actually solved the problem that we started out to solve.

11.4.3. Proposition. *Under the weak topology (defined above) on a set S each of the functions f_α is continuous; in fact the weak topology is the weakest topology on S under which these functions are continuous.*

11.4.4. Proposition. *Let X be a topological space with the weak topology determined by a family $\mathcal{F}$ of functions. Prove that a function $g\colon W \to X$, where W is a topological space, is continuous if and only if $f \circ g$ is continuous for every $f \in \mathcal{F}$.*

11.4.5. Example. Let S be a nonempty set and $\mathcal{F}(S) = \mathcal{F}(S,\mathbb{R}) = \mathbb{R}^S$ be the vector space of all real valued functions on S (see example 4.4.7). The weak topology induced on $\mathcal{F}(S)$ by the family of all evaluation functionals E_x where $x \in S$ (see definition 4.5.13) is called the *topology of pointwise convergence*. The reason for this terminology will become clear in example 11.4.8 below.

11.4.6. Notation. If S is a set we denote by $\operatorname{Fin} S$ the family of all finite subsets of S.

11.4.7. Example. Let S be a nonempty set and $\mathcal{F}(S)$ be the vector space of all real valued functions on S. Then the family of all sets of the form

$$U(g, F, \epsilon) := \{f \in \mathcal{F}(S) \colon |f(x) - g(x)| < \epsilon \text{ whenever } x \in F\},$$

where $g \in \mathcal{F}(S)$, $F \in \operatorname{Fin} S$ and $\epsilon > 0$, is a base for the topology of pointwise convergence (see example 11.4.5).

11.4.8. Example. Let S be a nonempty set. A sequence (f_n) of functions in $\mathcal{F}(S)$ converges to a function $g \in \mathcal{F}(S)$ in the topology of pointwise convergence if and only if $f_n(x) \to g(x)$ for every $x \in S$.

11.4.9. Definition. Let $\big(A_\lambda\big)_{\lambda \in \Lambda}$ be an indexed family of topological spaces. The weak topology on the Cartesian product $\prod A_\lambda$ induced by the family of coordinate projections π_λ (see definition 8.3.6) is called the PRODUCT TOPOLOGY.

11.4.10. Proposition. *The product of a family of Hausdorff spaces is Hausdorff.*

11.4.11. Remark. Let $(X_\alpha)_{\alpha \in A}$ be an indexed family of topological spaces and $g \colon Y \to \prod X_\alpha$ be a function from a topological space Y into the product of the X_α's. If $\prod X_\alpha$ has the product topology, then according to proposition 11.4.4 g is continuous if and only if $\pi_\alpha \circ g$ is continuous for each $\alpha \in A$. Thus, for example, a curve $c \colon \mathbb{R} \to \mathbb{R}^n \colon t \mapsto \big(c_1(t), \dots, c_n(t)\big)$ in $\mathbb{R}^n$ is continuous if and only if each of its coordinate functions c_k is.

11.4.12. Corollary. *Let $f \colon W \to X$ and $g \colon W \to Y$ where W, X, and Y are topological spaces, and let $h = (f, g)$. Then h is continuous if and only if f and g are.*

11.4.13. Example. The product $X \times Y$ of two topological spaces (with the topology defined above) is actually a product in the category **TOP** of topological spaces and continuous maps.

11.4.14. Example. Let X and Y be topological spaces. Declare a subset U of their disjoint union $X \uplus Y$ to be *open* if $U \cap X$ is open in X and $U \cap Y$ is open in Y. The family of such sets is a topology on $X \uplus Y$ called the DISJOINT UNION TOPOLOGY. The resulting space is a coproduct in the category **TOP**.

11.5. Subspaces

In example 10.1.21 we introduced the notion of a *subspace* of a topological space. Recall that a subset Y of a topological space X becomes a subspace of X if we equip it with the topology induced by X, that is, with the topology whose members are intersections of Y with open subsets of X. An equivalent description of this induced topology on Y is that it is the smallest topology on Y which makes the inclusion mapping $\iota \colon Y \to X$ continuous. Yet another way of saying exactly the same thing: it is the weak topology on Y determined by the inclusion map $\iota \colon Y \to X$.

11.5.1. CAUTION. Be aware that being open (or closed) is not what I will call an *intrinsic* property of a set. A set (with a fixed topology) can be open in one space but fail to be open in another. Let's take an obvious example. It is nonsense to ask out of context if the interval $(1, 2]$ is an open set. As a subset of $[0, 2]$ (regarded as a subspace of $\mathbb{R}$) it is indeed open. But as a subset of $\mathbb{R}$ it is certainly not open. It is standard to say *U is an open set* without further qualification when it is clear what topological space is being referred to. But if in some discussion U is contained in a subspace Y of a topological space X, it must be clear whether the intended meaning is that *U is open in Y* of that it is *open in X*.

In a similar vein, when Y is a subspace of a topological space X and $A \subseteq Y$, some of the notation introduced earlier becomes ambiguous. In particular, the notation for closure and interior are inadequate. If for example $X = \mathbb{R}$, $Y = (0, 2)$, and $A = (1, 2)$, what would $\overline{A}$ be? If we intend the closure of A in the subspace Y, the answer would be $[1, 2)$, while if closure in X were intended the answer would be $[1, 2]$. When this sort of ambiguity threatens it is wise to adopt a more explicit notation.

11.5.2. Notation. Let Y be a subspace of a topological space X and $A \subseteq Y$. We will write $\operatorname{cl}_Y(A)$ for the closure of A in the space Y and $\operatorname{cl}_X(A)$ for its closure in X. Similarly we write $\operatorname{int}_Y(A)$ for the interior of A as a subset of Y and $\operatorname{int}_X(A)$ for its interior in X.

11.5.3. Proposition. *Let $Y \subseteq X$ where X is a topological space. A subset A of Y is closed in Y if and only if there exists a closed set F in X such that $A = Y \cap F$.*

11.5.4. Proposition. *If $A \subseteq Y \subseteq X$ where X is a topological space, then*

$$\operatorname{cl}_Y(A) = Y \cap \operatorname{cl}_X(A)\,.$$

Thus if Y is a closed subset of X, then the closure of a set in Y is the same as its closure in X.

11.5.5. Proposition. *If X is a topological space, V is an open subset of X, and U is an open subset of the subspace V, then U is open in X.*

11.5.6. Proposition. *If $A \subseteq Y \subseteq X$ where X is a topological space, then $\operatorname{int}_X(A) \subseteq \operatorname{int}_Y(A)$. If, additionally, Y is an open subset of X, then equality holds.*

11.5.7. Example. Letting $A = (0, 1]$, $Y = [0, 1]$, and $X = \mathbb{R}$ shows that equality need not hold in general in the preceding proposition.

It is an interesting problem to decide which properties of a topological space are inherited by its subspaces. A property P of topological spaces which holds in every subspace whenever it holds in the "parent" space is a HEREDITARY property. A property P is F-HEREDITARY if whenever it holds in a topological space X it holds for every closed subset of X (regarded as a subspace of X), and it is G-hereditary if it holds for every open subset of X whenever it holds in X. (Memory aid: associate F with the French word *fermé*, which means *closed*.)

11.5.8. Example. The property of being Hausdorff is hereditary as is the property of being second countable. We will see later that compactness is an F-hereditary property (see proposition 17.1.7) and that being a Baire space is G-hereditary (see proposition 29.3.16).

11.6. Quotient Topologies

A weak topology on a set S is induced by a family of functions $f_\alpha\colon S \to X_\alpha$ mapping into topological spaces. It is the weakest topology on S which makes all these functions continuous (see exercise 11.4.3). A dual notion is that of a STRONG TOPOLOGY, a topology induced by a family of functions mapping *from* topological spaces X_α *into* S. It is the strongest topology under which all these functions are continuous. The only strong topology for which we will have much use is the *quotient topology.*

11.6.1. Definition. Let X be a topological space, S be a set, and $\pi\colon X \to S$ be a surjection. The family

$$\mathfrak{T} = \{U \subseteq S\colon \pi^{\leftarrow}(U) \overset{\circ}{\subseteq} X\}$$

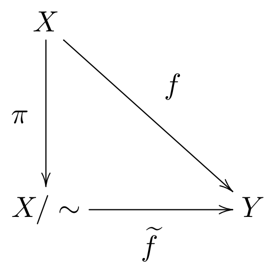
is a topology on S. It is the QUOTIENT TOPOLOGY determined by π.

11.6.2. Proposition. *The family $\mathfrak{T}$ defined above is a topology and the function π is continuous. In fact, $\mathfrak{T}$ is the strongest topology under which π is continuous.*

11.6.3. Example. With its quotient topology (as defined above) a set S is a quotient object in the category **TOP** and the mapping π is a quotient map (see definition 8.2.1).

11.6.4. Exercise. Let $\sim$ be an equivalence relation on a topological space X. Explain carefully how to place a topology on the family $X/\!\sim$ of equivalence classes of X in such a way that $X/\!\sim$ becomes a quotient object in the category of topological spaces and continuous maps.

11.6.5. Theorem (Fundamental Quotient Theorem for **TOP**)**.** *Let $\sim$ be an equivalence relation on a topological space X and $\pi\colon X \to X/\!\sim$ be the associated quotient map. Then for every topological space Y and every continuous function $f\colon X \to Y$ which is constant on the equivalence classes of X there exists a unique continuous function $\widetilde{f}\colon X/\!\sim\, \to Y$ which makes the following diagram commute.*



Furthermore, $\widetilde{f}$ is injective if and only if $f(x)$ is different from $f(y)$ whenever x and y are not equivalent; and $\widetilde{f}$ is surjective if and only if f is.

11.6.6. Example. In the category **TOP** not every epimorphism is a quotient map.

Hint for proof. Consider the identity map on the reals with different topologies on the domain and codomain.

11.6.7. Definition. A mapping $f\colon X \to Y$ is said to be OPEN if it takes open sets to open sets; that is, if $f^{\to}(U) \overset{\circ}{\subseteq} Y$ whenever $U \overset{\circ}{\subseteq} X$. It is CLOSED if it takes closed sets to closed sets.

11.6.8. Example. The mapping $\pi\colon [0,1] \to \mathbb{S}^1\colon t \mapsto e^{2\pi i t}$ shows that a quotient map need not be open.

11.6.9. Exercise. Your pal Fred R. Dimm has been considering two topological spaces: the interval $[0,1)$ and the unit circle $\mathbb{S}^1$ in the complex plane, each with its usual topology. He has decided that the latter is a quotient object of the former and that the function $\pi\colon [0,1) \to \mathbb{S}^1\colon t \mapsto e^{2\pi i t}$ is the quotient map. To prove this he considers a function f mapping $\mathbb{S}^1$ to a topological space X and assumes that $f \circ \pi$ is continuous. Now, he argues, if U is an open subset of X, then $\pi^{\leftarrow}(f^{\leftarrow}(U)) = (f \circ \pi)^{\leftarrow}(U)$ is an open subset of $[0,1)$. But according to the definition of the quotient topology this means that $f^{\leftarrow}(U)$ is an open subset of $\mathbb{S}^1$. Thus by definition 11.1.1 f is continuous. Help Fred to see the error of his ways and show him that π is not a quotient map. (What is the principal difference between this exercise and the preceding example?)

11.6.10. Example. The usual coordinate projection $\pi_1\colon \mathbb{R}^2 \to \mathbb{R}\colon (x,y) \mapsto x$ shows that a quotient map need not be closed.

11.6.11. Proposition. *In the category* **TOP** *every open epimorphism is a quotient map.*

11.6.12. Proposition. *In the category* **TOP** *every closed epimorphism is a quotient map.*

CHAPTER 12

NORMED LINEAR SPACES

12.1. Norms

In the world of analysis the predominant denizens are function spaces, vector spaces of real or complex valued functions. To be of interest to an analyst such a space should come equipped with a topology. Often the topology is a metric topology, which in turn frequently comes from a norm. We begin this section with the definition of "norm" and show how norms lead to metrics on vector spaces.

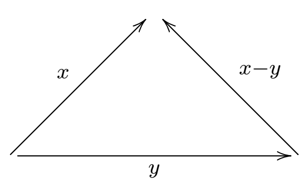
12.1.1. Definition. Let V be a vector space. A function $\| \ \| \colon V \to \mathbb{R} \colon x \mapsto \|x\|$ is a NORM on V if

(1) $\|x + y\| \le \|x\| + \|y\|$ for all x, $y \in V$;
(2) $\|\alpha x\| = |\alpha| \, \|x\|$ for all $x \in V$ and $\alpha \in \mathbb{R}$; and
(3) if $\|x\| = 0$, then $x = \mathbf{0}$.

(Of course, when V is a complex vector space, (2) should hold for all $\alpha \in \mathbb{C}$.) The expression $\|x\|$ may be read as "the *norm* of x" or "the *length* of x". If the function $\| \ \|$ satisfies (1) and (2) above (but perhaps not (3) it is a SEMINORM on V. A vector space on which a norm has been defined is a NORMED LINEAR SPACE (or NORMED VECTOR SPACE). A vector in a normed linear space which has norm 1 is a UNIT VECTOR.

12.1.2. Exercise. Show why it is clear from the definition that $\|\mathbf{0}\| = 0$ and that norms (and seminorms) can take on only positive values.

The most important fact is that *every normed linear space is a metric space.* More precisely, a norm on a vector space induces a metric d, which is defined by $d(x, y) = \|x - y\|$. That is, the distance between two vectors is the length of their difference.



If no other metric is specified we always regard a normed linear space as a metric space under this induced metric. Thus every metric (and hence every topological) concept makes sense in a (semi)normed linear space.

12.1.3. Example. Let V be a normed linear space. Define $d \colon V \times V \to \mathbb{R}$ by $d(x, y) = \|x - y\|$. Then d is a metric on V. If V is only a seminormed space, then d is a pseudometric.

12.1.4. Example. The absolute value function is a norm on $\mathbb{R}$. Obviously, it induces the usual metric on the real line (see example 10.4.6).

12.1.5. Example. For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\|x\| = \left(\sum_{k=1}^n {x_k}^2\right)^{1/2}$. The only nonobvious part of the proof that this defines a norm on $\mathbb{R}^n$ is the verification of the *triangle inequality.* But we have already done this: it is just Minkowski's inequality 6.5.2. This is the USUAL NORM (or EUCLIDEAN NORM) on $\mathbb{R}^n$; unless the contrary is explicitly stated, $\mathbb{R}^n$ when regarded as a normed linear space will always be assumed to possess this norm. It is clear that this norm induces the usual (Euclidean) metric on $\mathbb{R}^n$ (see example 10.4.11).

12.1.6. Example. For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\|x\|_1 = \sum_{k=1}^n |x_k|$. The function $x \mapsto \|x\|_1$ is easily seen to be a norm on $\mathbb{R}^n$. It is sometimes called the 1-NORM on $\mathbb{R}^n$. It induces the taxicab metric on $\mathbb{R}^n$ (see example 10.4.12).

The next four examples are actually a single example. The first two are special cases of the third, which is in turn a special case of the fourth.

12.1.7. Example. For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\|x\|_u = \max\{\,|x_k| \colon 1 \le k \le n\}$. This defines a norm on $\mathbb{R}^n$; it is the UNIFORM NORM on $\mathbb{R}^n$. It induces the uniform metric on $\mathbb{R}^n$ (see example 10.4.13).

12.1.8. Example. The set l_∞ of all bounded sequences of real numbers is clearly a vector space under pointwise operations of addition and scalar multiplication. For $x = (x_1, x_2, \dots,) \in l_\infty$ let $\|x\|_\infty = \sup\{\,|x_k| \colon 1 \le k\}$. This defines a norm on l_∞; it is the UNIFORM NORM on l_∞.

12.1.9. Example. Let S be a nonempty set. If f is a bounded real valued function on S, let

$$\|f\|_u := \sup\{|f(x)| \colon x \in S\}\,.$$

This is a norm on $\mathcal{B}(S)$ and is called the UNIFORM NORM. Clearly this norm gives rise to the uniform metric on $\mathcal{B}(S)$ (see example 10.4.32). Notice that examples 12.1.7 and 12.1.8 are special cases of this one. (Let $S = \mathbb{N}_n$ or $S = \mathbb{N}$.)

12.1.10. Example. It is easy to make a substantial generalization of the preceding example by replacing the real line $\mathbb{R}$ by an arbitrary normed linear space. Suppose then that S is a nonempty set and V is a normed linear space. For f in $\mathcal{B}(S, V)$ let

$$\|f\|_u := \sup\{\|f(x)\| \colon x \in S\}\,.$$

Then this is a norm and it is called the UNIFORM NORM on $\mathcal{B}(S, V)$. (Why is the word "easy" in the first sentence of this example appropriate?)

12.1.11. Definition. Let S be a set, V be a normed linear space, and (f_n) be a sequence of functions in $\mathcal{F}(S, V)$. If there is a function g in $\mathcal{F}(S, V)$ such that

$$\sup\{\|f_n(x) - g(x)\| \colon x \in S\} \to 0 \quad \text{as } n \to \infty,$$

then we say that the sequence (f_n) CONVERGES UNIFORMLY to g and write $f_n \to g\,(\text{unif})$. The function g is the UNIFORM LIMIT of the sequence (f_n). Notice that if g and all the f_n's belong to $\mathcal{B}(S, V)$, then uniform convergence of (f_n) to g is just convergence of (f_n) to g with respect to the uniform metric.

There are many ways in which sequences of functions converge. There is a detailed discussion of several of these in chapter 27. Arguably the two most common modes of convergence are uniform convergence, which we have just discussed, and pointwise convergence which we have encountered in examples 11.4.5 and 11.4.8.

12.1.12. Definition. Let S be a set, V be a normed linear space, and (f_n) be a sequence in $\mathcal{F}(S, V)$. If there is a function g such that

$$f_n(x) \to g(x) \qquad \text{for all } x \in S,$$

then (fn) CONVERGES POINTWISE to g. In this case we write

$$f_n \to g\,(\text{ptws}).$$

The function g is the POINTWISE LIMIT of the f_n's.

If (f_n) is an increasing sequence of real (or extended real) valued functions and $f_n \to g\,(\text{ptws})$, we write $f_n \uparrow g\,(\text{ptws})$. And if (f_n) is decreasing and has g as a pointwise limit, we write $f_n \downarrow g\,(\text{ptws})$.

The most obvious connection between these two types of convergence is that uniform convergence implies pointwise convergence.

12.1.13. Proposition. *If a sequence (f_n) in $\mathcal{F}(S,V)$ (where S is a set and V is a normed linear space) converges uniformly to a function g in $\mathcal{F}(S,V)$, then (f_n) converges pointwise to g.*

The converse is not true.

12.1.14. Example. For each $n \in \mathbb{N}$ let $f_n \colon [0,1] \to \mathbb{R} \colon x \mapsto x^n$. Then the sequence (f_n) converges pointwise on $[0,1]$, but not uniformly.

12.1.15. Example. For each $n \in \mathbb{N}$ and each $x \in \mathbb{R}^+$ let $f_n(x) = \frac{1}{n}x$. Then on each interval of the form $[0,a]$ where $a > 0$ the sequence (f_n) converges uniformly to the constant function 0. On the interval $[0,\infty)$ it converges pointwise to 0, but not uniformly.

A useful observation is that a uniform limit of bounded functions must itself be bounded.

12.1.16. Proposition. *Let S be a set, V be a normed linear space, and (f_n) be a sequence in $\mathcal{B}(S,V)$ and g be a member of $\mathcal{F}(S,V)$. If $f_n \to g$ (unif), then g is bounded.*

And a uniform limit of continuous functions is continuous.

12.1.17. Proposition. *Let X be a topological space, V be a normed linear space, and (f_n) be a sequence in $\mathcal{C}(X,V)$ and g be a member of $\mathcal{F}(X,V)$. If $f_n \to g$ (unif), then g is continuous.*

Spaces of bounded real finitely additive set functions provide us with more examples of normed linear spaces.

12.1.18. Example. Let $\mathfrak{A}$ be an algebra of subsets of a nonempty set S. For $\mu \in \operatorname{ba}(S)$ let

$$\|\mu\| := \sup\{|\mu(A)| \colon A \in \mathfrak{A}\}\,.$$

This function is a norm on $\operatorname{ba}(S)$, and thus $\operatorname{ba}(S)$ is a normed linear space.

12.1.19. Proposition. *If x and y are elements of a vector space V equipped with a seminorm $\|\ \ \|$, then*

$$\big|\,\|x\| - \|y\|\,\big| \le \|x - y\|\,.$$

12.1.20. Corollary. *A seminorm on a vector space is uniformly continuous.*

12.1.21. Proposition. *If V is a normed linear space, if $x \in V$, and if r, $s > 0$, then*

(a) *$B_r(\mathbf{0}) = -B_r(\mathbf{0})$;*
(b) *$B_{rs}(\mathbf{0}) = rB_s(\mathbf{0})$;*
(c) *$x + B_r(\mathbf{0}) = B_r(x)$; and*
(d) *$B_r(\mathbf{0}) + B_r(\mathbf{0}) = 2\,B_r(\mathbf{0})$. (Is it true in general that $A + A = 2A$ when A is a subset of a vector space?)*

12.1.22. Proposition. *Prove that in a normed linear space every open ball is a convex set. And so is every closed ball.*

12.1.23. Proposition. *Let V be a normed linear space. For each $a \in V$ the map $T_a \colon V \to V \colon x \mapsto x + a$ (called* TRANSLATION *by a) is a homeomorphism.*

12.1.24. Corollary. *If U is a nonempty open subset of a normed linear space V, then $U - U$ contains a neighborhood of 0.*

12.1.25. Proposition. *If (x_n) is a sequence in a normed linear space and $x_n \to a$, then $\displaystyle\frac{1}{n}\sum_{k=1}^{n} x_k \ \to\ a$.*

12.1.26. Proposition. *Let V be a vector space with two norms $\|\cdot\|_1$ and $\|\cdot\|_2$. Let $\mathfrak{T}_k$ be the topology induced on V by $\|\cdot\|_k$ (for $k = 1, 2$). If there exists a constant $\alpha > 0$ such that $\|x\|_1 \le \alpha\|x\|_2$ for every $x \in V$, then $\mathfrak{T}_1 \subseteq \mathfrak{T}_2$.*

In definition 10.4.24 we said what it means for two metrics to be strongly equivalent. The corresponding condition for norms is called simply *equivalence.*

12.1.27. Definition. Two norms on a vector space V are EQUIVALENT if there exist constants α, $\beta > 0$ such that for all $x \in V$

$$\alpha\|x\|_1 \le \|x\|_2 \le \beta\|x\|_1 \,.$$

12.1.28. Proposition. *If $\|\ \|_1$ and $\|\ \|_2$ are equivalent norms on a vector space V, then they induce the same topology on V.*

Exercise 12.1.28 gives us an easily verifiable sufficient condition for two norms to induce identical topologies on a vector space. Thus, if we are trying to show, for example, that some subset of a normed linear space is open it may very well be the case that the proof can be simplified by replacing the given norm with an equivalent one.

Similarly, suppose that we are attempting to verify that a function f between two normed linear spaces is continuous. Since continuity is defined in terms of open sets and equivalent norms produce exactly the same open sets (see proposition 12.1.28), we are free to replace the norms on the domain of f and the codomain of f with any equivalent norms we please. This process can sometimes simplify arguments significantly. (This possibility of simplification, incidentally, is one major advantage of giving a topological definition of continuity in the first place.)

12.1.29. Definition. A sequence $a = (a_n)$ of vectors in a real or complex normed linear space V is SUMMABLE if there exists a vector $b \in V$ such that $\|b - s_n\| \to 0$ as $n \to \infty$, where $s_n = \sum_{k=1}^{n} a_k$ is the n^{th} partial sum of the sequence a. The sequence a is ABSOLUTELY SUMMABLE if $\sum_{k=1}^{\infty}\|a_k\| < \infty$.

12.1.30. Exercise. Let l_1 be the set of all absolutely summable sequences of real numbers. Make l_1 into a vector space with the usual pointwise definition of addition and scalar multiplication. For every $a \in l_1$ define

$$\|a\|_1 := \sum_{k=1}^{\infty} |a_k|$$

and

$$\|a\|_u := \sup\{|a_k| \colon k \in \mathbb{N}\}.$$

(These are, respectively, the 1-norm and the *uniform* norm.) Show that both $\|\ \|_1$ and $\|\ \|_u$ are norms on l_1. Then prove or disprove:

(a) If a sequence (a^i) of vectors in the normed linear space $\big(l_1, \|\ \|_1\big)$ converges, then the sequence also converges in $\big(l_1, \|\ \|_u\big)$.

(b) If a sequence (a^i) of vectors in the normed linear space $\big(l_1, \|\ \|_u\big)$ converges, then the sequence also converges in $\big(l_1, \|\ \|_1\big)$.

12.2. Bounded Linear Maps

Analysts work with objects having both algebraic and topological structure. In the preceding section we examined vector spaces endowed with a topology derived from a norm. If these are the objects under consideration, what morphisms are likely to be of greatest interest? The plausible, and correct, answer is *maps which preserve both topological and algebraic structures*, that is, continuous linear maps. The resulting category is denoted by $\mathbf{NLS_\infty}$. The isomorphisms in this category are, topologically, homeomorphisms. It is clear that one might choose instead to study a category in which the mappings satisfy a stronger condition: the isomorphisms are *isometries*, that is they preserve distances (equivalently, they preserve norms). In this category, denoted by $\mathbf{NLS_1}$, the morphisms are CONTRACTIVE LINEAR MAPS, that is, linear maps $T \colon V \to W$ between normed linear spaces such that $\|Tx\| \le \|x\|$ for all $x \in V$. It seems appropriate to refer to $\mathbf{NLS_\infty}$ as the *topological category* of normed linear spaces and to $\mathbf{NLS_1}$ as the *geometric category* of normed linear spaces. Many authors use the unmodified term "isomorphism" for an isomorphism in the topological category $\mathbf{NLS_\infty}$ and "isometric isomorphism" for an isomorphism in the geometric category $\mathbf{NLS_1}$. In these notes we will focus on the topological category. The isometric theory of normed linear spaces although important is somewhat more specialized.

12.2.1. Exercise. Verify the unproved assertions in the preceding paragraph. In particular, prove that

(a) $\mathbf{NLS_\infty}$ is a category.
(b) An isomorphism in $\mathbf{NLS_\infty}$ is both a vector space isomorphism and a homeomorphism.
(c) A linear map between normed linear spaces is an isometry if and only if it is norm preserving. (*Definitions.* For $k = 1, 2$ let V_k be a normed linear space, $\| \quad \|_k$ be the norm on V_k, d_k be the metric on V_k induced by $\| \quad \|_k$, and $f \colon V_1 \to V_2$. Then f is an ISOMETRY (or an ISOMETRIC map) if $d_2(f(x), f(y)) = d_1(x, y)$ for all x, $y \in V_1$. It is NORM PRESERVING if $\|f(x)\|_2 = \|x\|_1$ for all $x \in V_1$.)
(d) $\mathbf{NLS_1}$ is a category.
(e) An isomorphism in $\mathbf{NLS_1}$ is both an isometry and a vector space isomorphism.
(f) The category $\mathbf{NLS_1}$ is a subcategory of $\mathbf{NLS_\infty}$ in the sense that every morphism of the former belongs to the latter.

Next we consider a very useful condition which, for linear maps, turns out to be equivalent to continuity.

12.2.2. Definition. A linear transformation $T \colon V \to W$ between normed linear spaces is BOUNDED if $T(B)$ is a bounded subset of W whenever B is a bounded subset of V. In other words, a bounded linear map takes bounded sets to bounded sets. We denote by $\mathfrak{B}(V, W)$ the family of all bounded linear transformations from V into W. A bounded linear map from a space V into itself is often called a (bounded linear) OPERATOR. The family of all operators on a normed linear space V is denoted by $\mathfrak{B}(V)$. The class of normed linear spaces together with the bounded linear maps between them constitute a category. We verify below that this is just the category $\mathbf{NLS_\infty}$.

12.2.3. CAUTION. It is extremely important to realize that a bounded linear map will *not*, in general, be a bounded function in the sense of definitions 4.4.13 or 10.4.31. The use of "bounded" in these two conflicting senses may be unfortunate, but it is well established.

12.2.4. Example. The linear map $T \colon \mathbb{R} \to \mathbb{R} \colon x \mapsto 3x$ is a bounded linear map (since it maps bounded subsets of $\mathbb{R}$ to bounded subsets of $\mathbb{R}$), but, regarded just as a function, T is not bounded (since its range is not a bounded subset of $\mathbb{R}$).

The following observation my help reduce confusion.

12.2.5. Proposition. *A linear transformation, unless it is the constant map that takes every vector to zero,* cannot *be a bounded function.*

One of the most striking aspects of linearity is that for linear maps the concepts of continuity, continuity at a single point, and uniform continuity coalesce. And in fact they are exactly the same thing as boundedness.

12.2.6. Theorem. *Let $T \colon V \to W$ be a linear transformation between normed linear spaces. Then the following are equivalent:*

(i) *T is continuous at* 0.
(ii) *T is continuous on V.*
(iii) *T is uniformly continuous on V.*
(iv) *The image of the closed unit ball under T is bounded.*
(v) *T is bounded.*
(vi) *The exists a number $M > 0$ such that $\|Tx\| \le M\|x\|$ for all $x \in V$.*

12.2.7. Proposition. *Let $T \colon V \to W$ be a bounded linear transformation between normed linear spaces. Then the following four numbers (exist and) are equal.*

(i) $\sup\{\|Tx\| : \|x\| \le 1\}$
(ii) $\sup\{\|Tx\| : \|x\| = 1\}$
(iii) $\sup\{\|Tx\|\,\|x\|^{-1} : x \ne \mathbf{0}\}$
(iv) $\inf\{M > 0 : \|Tx\| \le M\|x\| \ \ \textit{for all } x \in V\}$

12.2.8. Definition. If T is a bounded linear map, then $\|T\|$, called the NORM of T, is defined to be any one of the four expressions in the previous exercise.

12.2.9. Example. If V and W are normed linear spaces, then the function

$$\|\ \ \| : \mathfrak{B}(V, W) \to \mathbb{R} : T \mapsto \|T\|$$

is, in fact, a norm.

12.2.10. Example. The family $\mathfrak{B}(V, W)$ of all bounded linear maps between normed linear spaces is itself a normed linear space.

12.2.11. Example. Let $T : \mathbb{R}^2 \to \mathbb{R}^3 : (x, y) \mapsto (3x, x + 2y, x - 2y)$. Then $\|T\| = \sqrt{11}$.

12.3. Products of Normed Linear Spaces

12.3.1. Example. Let V and W be normed linear spaces. On the Cartesian product $V \times W$ define

$$\|(x, y)\|_2 = \sqrt{\|x\|^2 + \|y\|^2},$$
$$\|(x, y)\|_1 = \|x\| + \|y\|,$$

and

$$\|(x, y)\|_u = \max\{\|x\|, \|y\|\}\,.$$

The first of these is the EUCLIDEAN NORM (or 2-NORM) on $V \times W$, the second is the 1-NORM, and the last is the UNIFORM NORM. Verifying that these are all norms on $V \times W$ is quite similar to the arguments required in examples 12.1.5, 12.1.6, and 12.1.7. That they are equivalent norms is a consequence of the following inequalities and exercise 12.1.28.

$$\|x\| + \|y\| \le \sqrt{2}\sqrt{\|x\|^2 + \|y\|^2} \le 2\max\{\|x\|, \|y\|\} \le 2(\|x\| + \|y\|)$$

(The first inequality is an easy consequence of exercise 6.4.1(d).)

12.3.2. Convention. In the preceding example we defined three (equivalent) norms on the product space $V \times W$. We will take the first of these $\|\ \ \|_1$ as the PRODUCT NORM on $V \times W$. Thus whenever V and W are normed linear spaces, unless the contrary is specified we will regard the product $V \times W$ as a normed linear space under this norm. Usually we write just $\|(x, y)\|$ instead of $\|(x, y)\|_1$. The product of the normed linear spaces V and W is usually denoted by $V \oplus W$ and is called the DIRECT SUM of V and W. (In the special case where $V = W = \mathbb{R}$, what metric does the product norm induce on $\mathbb{R}^2$?)

12.3.3. Proposition. *If V and W are normed linear spaces, then the product topology on $V \times W$ is the same as the topology induced by the product norm on $V \oplus W$.*

12.3.4. Example. Show that product $V \oplus W$ of normed linear spaces is in fact a product in the category $\mathbf{NLS_\infty}$ of normed linear spaces and bounded linear maps. What are the coproducts in this category?

12.3.5. Exercise. Identify the product and coproduct of two spaces V and W in the category $\mathbf{NLS_1}$ of normed linear spaces and linear contractions.

12.3.6. Exercise. Let (x_n, y_n) be a sequence in the direct sum $V \oplus W$ of two normed linear spaces. Prove that (x_n, y_n) converges to a point (a, b) in $V \oplus W$ if and only if $x_n \to a$ in V and $y_n \to b$ in W.

12.3.7. Proposition. *Addition is a continuous operation on a normed linear space V. That is, the map*

$$A\colon V \oplus V \to V\colon (x, y) \mapsto x + y$$

is continuous.

12.3.8. Exercise. Give a *very* short proof (no ϵ's or δ's or open sets) that if (x_n) and (y_n) are sequences in a normed linear space which converge to a and b, respectively, then $x_n + y_n \to a + b$.

12.3.9. Proposition. *Scalar multiplication is a continuous operation on a normed linear space V in the sense that the map*

$$S\colon \mathbb{R} \times V \to V\colon (\alpha, x) \mapsto \alpha x$$

is continuous.

12.3.10. Proposition. *If B and C are subsets of a normed linear space and α is a scalar, then*

(a) $\overline{\alpha B} = \alpha \overline{B}$*; and*
(b) $\overline{B} + \overline{C} \subseteq \overline{B + C}$.

12.3.11. Example. If B and C are closed subsets of a normed linear space, then it does not necessarily follow that $B + C$ is closed (and therefore $\overline{B} + \overline{C}$ and $\overline{B + C}$ need not be equal).

12.3.12. Example. Let X be a topological space. Then the family $\mathcal{C}(X)$ of all continuous functions on X is a vector space under the usual pointwise operations of addition and scalar multiplication.

12.3.13. Example. Let X be a topological space. We denote by $\mathcal{C}_b(X)$ the family of all bounded continuous functions on X. It is a normed linear space under the uniform norm. In fact, it is a subspace of $\mathcal{B}(X)$ (see example 12.1.9).

12.3.14. Definition. Let A be a (real) algebra on which a norm has been defined. Suppose that additionally the SUBMULTIPLICATIVE property

$$\|xy\| \leq \|x\|\,\|y\|$$

is satisfied for all x, $y \in A$. Then A is a (REAL) NORMED ALGEBRA. (Much later in these notes we will also consider *complex normed algebras*, where the scalars come from $\mathbb{C}$ rather than $\mathbb{R}$.) We make one further requirement: if the algebra A is unital, then

$$\|\mathbf{1}\| = 1\,.$$

In this case A is a UNITAL NORMED ALGEBRA.

12.3.15. Example. In example 12.1.9 we showed that the family $\mathcal{B}(S)$ of bounded real valued functions on a set S is a normed linear space under the uniform norm. It is also a commutative unital normed algebra.

12.3.16. Proposition. *Multiplication is a continuous operation on a normed algebra V in the sense that the map*

$$M\colon V \times V \to V\colon (x, y) \mapsto xy$$

is continuous.

Hint for proof. If you can prove that multiplication on the real numbers is continuous you can almost certainly prove that it is continuous on arbitrary normed algebras.

12.3.17. Example. The family $\mathcal{C}_b(X)$ of bounded continuous real valued functions on a topological space X is, under the usual pointwise operations and the uniform norm, a commutative unital normed algebra.

12.3.18. Example. The family $\mathfrak{B}(V)$ of all (bounded linear) operators on a normed linear space V is a unital normed algebra.

12.3.19. Definition. In 4.5.12 we discussed the algebraic dual $V^\dagger$ of a vector space V. Of great importance in normed spaces is the study of the *continuous* members of $V^\dagger$, the *continuous linear functionals*, also known as *bounded linear functionals*. The set V^* of all such functionals on the normed space V is the DUAL SPACE of V. To distinguish it from the algebraic dual and the order dual, it is sometimes called the NORM DUAL or the TOPOLOGICAL DUAL of V.

12.3.20. Example. If S is a set and $a \in S$, then the evaluation functional

$$E_a \colon \mathcal{B}(S) \to \mathbb{R} \colon f \mapsto f(a)$$

is a bounded linear functional on the space $\mathcal{B}(S)$ of all bounded real valued functions on S. If S happens to be a topological space then we may also regard E_a as a member of the dual space of $\mathcal{C}_b(S)$. (In either case, what is $\|E_a\|$?)

12.4. Quotients of Normed Linear Spaces

In discussing quotients in the category $\mathbf{NLS_\infty}$ we use the notation introduced for quotient vector spaces in definition 8.2.8.

12.4.1. Proposition. *Let V be a normed linear space and M be a closed subspace of V. Then the map*

$$\|\ \ \| \colon V/M \to \mathbb{R} \colon [x] \mapsto \inf\{\|u\| \colon u \sim x\}$$

is a norm on V/M. Furthermore, the quotient map

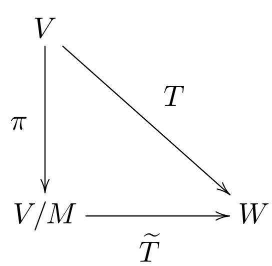
$$\pi \colon V \to V/M \colon x \mapsto [x]$$

is a bounded linear surjection with $\|\pi\| \le 1$.

12.4.2. Exercise. Revisit exercise 8.2.10 and explain how the example given there might motivate the definition of the norm on the quotient space V/M in proposition 12.4.1.

Next we give the analogue of theorems 2.11.9, 8.2.6, and 8.2.11 in the category of normed linear spaces and bounded linear maps.

12.4.3. Theorem (Fundamental quotient theorem for $\mathbf{NLS_\infty}$). *Let V and W be normed linear spaces and M be a closed subspace of V. If T is a bounded linear map from V to W and $\ker T \supseteq M$, then there exists a unique bounded linear map $\widetilde{T} \colon V/M \to W$ which makes the following diagram commute.*



Furthermore: $\|\widetilde{T}\| = \|T\|$; $\widetilde{T}$ is injective if and only if $\ker T = M$; and $\widetilde{T}$ is surjective if and only if T is.

CHAPTER 13

DIFFERENTIATION

We now pause for a *very* brief review of differential calculus. The central concept here is differentiability. A function f between normed linear spaces is said to be *differentiable* at a point p if (when the point $(p, f(p))$ is translated to the origin) the function is tangent to some continuous linear map. In this chapter we make this precise and record a few important facts about differentiability. A more detailed treatment can be found in my *ProblemText in Advanced Calculus* [17], chapters 25–29.

There are two sorts of textbooks on differential calculus: concept oriented and computation oriented. It is my belief that students who understand the concepts behind differentiation can do the calculations, while students who study calculations only often get stuck. Among the most masterful presentations of concept oriented differential calculus are [13] (volume I, chapter 8) and [30] (chapter 3). As of this writing the latter book is available without charge at the website of one of the authors:

`http://www.math.harvard.edu/~shlomo/docs/Advanced_Calculus.pdf`

The material in this chapter will benefit primarily those whose only encounter with multivariate calculus has been through partial derivatives and a *chain rule* that looks something like

$$\frac{\partial w}{\partial u} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial u} \tag{13.1}$$

The approach here is intended to be more geometric, emphasizing the role of *tangency*.

13.1. Tangency

13.1.1. Notation. Let V and W be normed linear spaces and $a \in V$. We denote by $\mathcal{F}_a(V,W)$ the family of all functions defined on a neighborhood of a taking values in W. That is, f belongs to $\mathcal{F}_a(V,W)$ if there exists a set U such that $a \in U \overset{\circ}{\subseteq} \operatorname{dom} f \subseteq V$ and if the image of f is contained in W. We shorten $\mathcal{F}_a(V,W)$ to $\mathcal{F}_a$ when no confusion will result. Notice that for each $a \in V$, the set $\mathcal{F}_a$ is closed under addition and scalar multiplication. (As usual, we define the sum of two functions f and g in $\mathcal{F}_a$ to be the function $f+g$ whose value at x is $f(x)+g(x)$ whenever x belongs to $\operatorname{dom} f \cap \operatorname{dom} g$.) Despite the closure of $\mathcal{F}_a$ under these operations, $\mathcal{F}_a$ is *not* a vector space. (Why not?)

13.1.2. Definition. Let V and W be normed linear spaces. A function f in $\mathcal{F}_0(V,W)$ belongs to $\mathfrak{O}(V,W)$ if there exist numbers $c > 0$ and $\delta > 0$ such that

$$\|f(x)\| \leq c\,\|x\|$$

whenever $\|x\| < \delta$.

A function f in $\mathcal{F}_0(V,W)$ belongs to $\mathfrak{o}(V,W)$ if for every $c > 0$ there exists $\delta > 0$ such that

$$\|f(x)\| \leq c\,\|x\|$$

whenever $\|x\| < \delta$. Notice that f belongs to $\mathfrak{o}(V,W)$ if and only if $f(0) = 0$ and

$$\lim_{h\to 0} \frac{\|f(h)\|}{\|h\|} = 0\,.$$

When no confusion seems likely we will shorten $\mathfrak{O}(V,W)$ to $\mathfrak{O}$ and $\mathfrak{o}(V,W)$ to $\mathfrak{o}$.

13.1.3. Exercise. Here is a list summarizing the important facts about the families $\mathfrak{O}$ and $\mathfrak{o}$. State precisely what each of these says and give a proof. (Here $\mathfrak{B}$ is the set of bounded linear maps between normed linear spaces V and W, and $\mathcal{C}_0$ is the set of all functions in $\mathcal{F}_0(V,W)$ which are continuous at 0.)

$$\mathfrak{B} \cup \mathfrak{o} \subseteq \mathfrak{O} \subseteq \mathcal{C}_0\,. \tag{1}$$
$$\mathfrak{B} \cap \mathfrak{o} = 0\,. \tag{2}$$
$$\mathfrak{O} + \mathfrak{O} \subseteq \mathfrak{O}\,; \qquad \alpha\,\mathfrak{O} \subseteq \mathfrak{O}\,. \tag{3}$$
$$\mathfrak{o} + \mathfrak{o} \subseteq \mathfrak{o}\,; \qquad \alpha\,\mathfrak{o} \subseteq \mathfrak{o}\,. \tag{4}$$
$$\mathfrak{o} \circ \mathfrak{O} \subseteq \mathfrak{o}\,. \tag{5}$$
$$\mathfrak{O} \circ \mathfrak{o} \subseteq \mathfrak{o}\,. \tag{6}$$
$$\mathfrak{o}(V,\mathbb{R}) \cdot W \subseteq \mathfrak{o}(V,W)\,. \tag{7}$$
$$\mathfrak{O}(V,\mathbb{R}) \cdot \mathfrak{O}(V,W) \subseteq \mathfrak{o}(V,W)\,. \tag{8}$$

13.1.4. Definition. Let V and W be normed linear spaces. Two functions f and g in $\mathcal{F}_0(V,W)$ are TANGENT (AT ZERO), in which case we write $f \simeq g$, if $f - g \in \mathfrak{o}(V,W)$.

13.1.5. Proposition. *The relation of tangency enjoys the following properties.*

(a) *"Tangency at zero" is an equivalence relation on $\mathcal{F}_0$.*
(b) *Let S, $T \in \mathfrak{B}$ and $f \in \mathcal{F}_0$. If $S \simeq f$ and $T \simeq f$, then $S = T$.*
(c) *If $f \simeq g$ and $j \simeq k$, then $f + j \simeq g + k$, and furthermore, $\alpha f \simeq \alpha g$ for all $\alpha \in \mathbb{R}$.*
(d) *Let $\phi, \psi \in \mathcal{F}_0(V,\mathbb{R})$ and $w \in W$. If $\phi \simeq \psi$, then $\phi w \simeq \psi w$.*
(e) *Let f, $g \in \mathcal{F}_0(V,W)$ and $T \in \mathfrak{B}(W,X)$. If $f \simeq g$, then $T \circ f \simeq T \circ g$.*
(f) *Let $h \in \mathfrak{O}(V,W)$ and f, $g \in \mathcal{F}_0(W,X)$. If $f \simeq g$, then $f \circ h \simeq g \circ h$.*

13.2. The Differential

13.2.1. Definition. Let V and W be normed linear spaces, $a \in V$, and $f \in \mathcal{F}_a(V,W)$. Define the function Δf_a by

$$\Delta f_a(h) := f(a+h) - f(a)$$

for all h such that $a + h$ is in the domain of f. Notice that since f is defined in a neighborhood of a, the function Δf_a is defined in a neighborhood of 0; that is, Δf_a belongs to $\mathcal{F}_0(V,W)$. Notice also that $\Delta f_a(0) = 0$.

13.2.2. Proposition. *If V and W are normed linear spaces and $a \in V$, then the function Δ has the following properties.*

(a) *If $f \in \mathcal{F}_a(V,W)$ and $\alpha \in \mathbb{R}$, then*
$$\Delta(\alpha f)_a = \alpha\,\Delta f_a\,.$$
(b) *If f, $g \in \mathcal{F}_a(V,W)$, then*
$$\Delta(f+g)_a = \Delta f_a + \Delta g_a\,.$$
(c) *If $\phi \in \mathcal{F}_a(V,\mathbb{R})$ and $f \in \mathcal{F}_a(V,W)$, then*
$$\Delta(\phi f)_a = \phi(a) \cdot \Delta f_a \,+\, \Delta\phi_a \cdot f(a) \,+\, \Delta\phi_a \cdot \Delta f_a\,.$$
(d) *If $f \in \mathcal{F}_a(V,W)$, $g \in \mathcal{F}_{f(a)}(W,X)$, and $g \circ f \in \mathcal{F}_a(V,X)$, then*
$$\Delta(g \circ f)_a = \Delta g_{f(a)} \circ \Delta f_a\,.$$
(e) *A function $f\colon V \to W$ is continuous at the point a in V if and only if Δf_a is continuous at 0.*

(f) *If $f\colon U \to U_1$ is a bijection between subsets of arbitrary vector spaces, then for each a in U the function $\Delta f_a\colon U - a \to U_1 - f(a)$ is invertible and*

$$\big(\Delta f_a\big)^{-1} = \Delta\big(f^{-1}\big)_{f(a)}\,.$$

13.2.3. Definition. Let V and W be normed linear spaces, $a \in V$, and $f \in \mathcal{F}_a(V,W)$. We say that f is DIFFERENTIABLE AT a if there exists a bounded linear map which is tangent at 0 to Δf_a. If such a map exists, it is called the DIFFERENTIAL of f at a and is denoted by df_a. Thus df_a is just a member of $\mathfrak{B}(V,W)$ which satisfies $df_a \simeq \Delta f_a$. We denote by $\mathcal{D}_a(V,W)$ the family of all functions in $\mathcal{F}_a(V,W)$ which are differentiable at a. We often shorten this to $\mathcal{D}_a$.

We establish next that there can be at most one bounded linear map tangent to Δf_a.

13.2.4. Proposition. *Let V and W be normed linear spaces and $a \in V$. If $f \in \mathcal{D}_a(V,W)$, then its differential is unique.*

13.2.5. Exercise. Let

$$f\colon \mathbb{R}^3 \to \mathbb{R}^2\colon (x,y,z) \mapsto (x^2y - 7, 3xz + 4y)$$

and $a = (1,-1,0)$. Use the *definition* of "differential" to find df_a. *Hint.* Work with the matrix representation of df_a. Since the differential must belong to $\mathfrak{B}(\mathbb{R}^3,\mathbb{R}^2)$, its matrix representation is a 2×3 matrix $M = \begin{bmatrix} r & s & t \\ u & v & w \end{bmatrix}$. Use the requirement that $\|h\|^{-1}\,\|\Delta f_a(h) - Mh\| \to 0$ as $h \to 0$ to discover the identity of the entries in M.

13.2.6. Exercise. Let $\mathbf{F}\colon \mathbb{R}^2 \to \mathbb{R}^4$ be defined by $\mathbf{F}(x,y) = (y, x^2, 4 - xy, 7x)$, and let $\mathbf{p} = (1,1)$. Use the *definition* of "differentiable" to show that $\mathbf{F}$ is differentiable at $\mathbf{p}$. Find the (matrix representation of the) differential of $\mathbf{F}$ at p.

13.2.7. Proposition. *Let V and W be normed linear spaces and $a \in V$. If $f \in \mathcal{D}_a$, then $\Delta f_a \in \mathfrak{O}$; thus, every function which is differentiable at a point is continuous there.*

13.2.8. Proposition. *Let V and W be normed linear spaces and $a \in V$. Suppose that f, $g \in \mathcal{D}_a(V,W)$ and that $\alpha \in \mathbb{R}$. Then*

(1) *αf is differentiable at a and*

$$d(\alpha f)_a = \alpha\, df_a\,;$$

(2) *also, $f + g$ is differentiable at a and*

$$d(f+g)_a = df_a + dg_a\,.$$

Suppose further that $\phi \in \mathcal{D}_a(V,\mathbb{R})$. Then

(c) *$\phi f \in \mathcal{D}_a(V,W)$ and*

$$d(\phi f)_a = d\phi_a \cdot f(a) \,+\, \phi(a)\, df_a\,.$$

It seems to me that the version of the *chain rule* given in (13.1), although (under appropriate hypotheses) a correct equation, really says very little. The idea that should be conveyed is that the best linear approximation to the composite of two smooth functions is the composite of their best linear approximations.

13.2.9. Theorem (The Chain Rule)**.** *Let V, W, and X be normed linear spaces with $a \in V$. If $f \in \mathcal{D}_a(V,W)$ and $g \in \mathcal{D}_{f(a)}(W,X)$, then $g \circ f \in \mathcal{D}_a(V,X)$ and*

$$d(g\circ f)_a = dg_{f(a)} \circ df_a\,.$$

Proof. Our hypotheses are $\Delta f_a \simeq df_a$ and $\Delta g_{f(a)} \simeq dg_{f(a)}$. By proposition 13.2.7 $\Delta f_a \in \mathfrak{O}$. Then by proposition 13.1.5(f)

$$\Delta g_{f(a)} \circ \Delta f_a \simeq dg_{f(a)} \circ \Delta f_a \tag{13.2}$$

and by proposition 13.1.5(e)

$$dg_{f(a)} \circ \Delta f_a \simeq dg_{f(a)} \circ df_a\,. \tag{13.3}$$

According to proposition 13.2.2(d)

$$\Delta(g \circ f)_a \simeq \Delta g_{f(a)} \circ \Delta f_a\,. \tag{13.4}$$

From (13.2), (13.3),(13.4), and proposition 13.1.5(a) it is clear that

$$\Delta(g \circ f)_a \simeq dg_{f(a)} \circ df_a\,.$$

Since $dg_{f(a)} \circ df_a$ is a bounded linear transformation, the desired conclusion is an immediate consequence of proposition 13.2.4. □

13.2.10. Exercise. Derive (under appropriate hypotheses) equation (13.1) from theorem 13.2.9.

13.2.11. Exercise. Let $\mathbf{T}$ be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $\mathbf{p} \in \mathbb{R}^n$. Find $d\mathbf{T}_{\mathbf{p}}$.

13.2.12. Example. Let T be a symmetric $n \times n$ matrix and let $\mathbf{p} \in \mathbb{R}^n$. Define a function $f\colon \mathbb{R}^n \to \mathbb{R}$ by $f(\mathbf{x}) = \langle T\mathbf{x}, \mathbf{x}\rangle$. Then

$$df_{\mathbf{p}}(\mathbf{h}) = 2\langle T\mathbf{p}, \mathbf{h}\rangle$$

for every $h \in \mathbb{R}^n$.

CHAPTER 14

RIESZ SPACES

Ordered vector spaces are a bit too general for our purposes. A much richer theory results when we add the requirement that these spaces be lattices. The resulting objects are called *Riesz spaces*. Perhaps the best introduction to these spaces is the book [**49**] by Adriaan Zaanen, one of the pioneers in the field. Two other excellent introductory treatments are Chapter 7 of [**1**] and Chapter 35 of [**19**]. For more advanced treatments see [**3**], [**31**], [**33**], and [**40**].

14.1. Definition and Elementary Properties

14.1.1. Definition. A RIESZ SPACE (or VECTOR LATTICE) is an ordered vector space in which every pair of elements has a supremum and an infimum. As usual we use the notation $x \vee y$ for $\sup\{x, y\}$ and $x \wedge y$ for $\inf\{x, y\}$. If x is an element of a Riesz space, the element $x^+ := x \vee 0$ is the POSITIVE PART of x and $x^- := -(x \wedge 0)$ is the NEGATIVE PART of x. (Notice that just as the imaginary part of a complex number is a real number, the negative part of a Riesz space vector is a *positive* vector in the space.)

14.1.2. Proposition. *An ordered vector space V is a Riesz space if and only if $x \vee 0$ exists for each $x \in V$. Furthermore, in any Riesz space the following formulas always hold:*

(a) $x \vee y = x + (y - x)^+$*; and*
(b) $x \wedge y = y - (y - x)^+$.

14.1.3. Example. The real line $\mathbb{R}$ is a Riesz space.

14.1.4. Example. The ordered vector space $\mathbb{R}^2$ with its usual ordering (see example 9.3.6) is a Riesz space.

14.1.5. Example. The ordered vector space $\mathbb{R}^2$ with lexicographic ordering (see example 9.3.7) is a Riesz space.

14.1.6. Example. If S is a nonempty set the family $\mathcal{F}(S)$ of all real valued functions on S (see example 9.3.11) is a Riesz space under the usual pointwise partial ordering.

14.1.7. Example. If S is a nonempty set the family $\mathcal{B}(S)$ of bounded real valued functions on S (see example 4.4.14) is a Riesz space under the usual pointwise partial ordering.

14.1.8. Example. If X is a nonempty topological space the family $\mathcal{C}(X)$ of continuous real valued functions on X (see example 12.3.12) is a Riesz space under the usual pointwise partial ordering.

14.1.9. Theorem (Jordan Decomposition Theorem)**.** *The positive cone of a Riesz space is generating. In particular, if E is a Riesz space and $x \in E$, then*

$$x = x^+ - x^- .$$

Hint for proof. Add (a) and (b) of proposition 14.1.2.

14.1.10. Proposition. *If x is a vector in a Riesz space, then $x^+ \wedge x^- = \mathbf{0}$.*

The *Jordan decomposition* is not a unique way of decomposing an element of a Riesz space into the difference of two positive elements, but it is "minimal" in the following sense.

14.1.11. Proposition. *If an element x of a Riesz space can be written in the form $x = a - b$ where a and b are positive, then $x^+ \leq a$ and $x^- \leq b$.*

14.1.12. Proposition. *If an element x in a Riesz space is written in the form $x = a - b$ where $a \wedge b = \mathbf{0}$, then $a = x^+$ and $b = x^-$.*

14.1.13. Proposition. *For vectors x and y in a Riesz space the following hold:*

(a) $x^- = (-x)^+$ *and* $x^+ = (-x)^-$*;*
(b) $x \le y$ *if and only if* $x^+ \le y^+$ *and* $x^- \ge y^-$*;*
(c) $(x+y)^+ \le x^+ + y^+$ *and* $(x+y)^- \le x^- + y^-$*.*

14.1.14. Notation. If x is an element of a Riesz space let

$$|x| := x \vee (-x)\,.$$

14.1.15. Proposition. *If x is an element of a Riesz space, then $|x| = x^+ + x^-$.*

Interesting and important examples of Riesz spaces can be found among the spaces of bounded real finitely additive set functions. We examine these in items 14.1.16–14.1.20.

14.1.16. Example. Let $\mathfrak{A}$ be an algebra of subsets of a nonempty set S. Given $\mu \in \mathrm{ba}(S)$ define a function σ on $\mathfrak{A}$ by

$$\sigma(A) = \sup\{\mu(B) \colon A \supseteq B \in \mathfrak{A}\}\,.$$

Then σ is a bounded (real) finitely additive set function.

14.1.17. Example. If S is a nonempty set, then the ordered vector space $\mathrm{ba}(S)$ (see 9.4.10) is a Riesz space.

Hint for proof. Show that the function σ defined in the preceding example is the least upper bound of μ and $\mathbf{0}$.

14.1.18. CAUTION. An especially interesting aspect of the spaces $\mathrm{ba}(S)$ is that suprema, infima, and absolute values are in general *not* evaluated pointwise. (See, for example, 14.1.20.)

14.1.19. Proposition. *Let $\mathfrak{A}$ be an algebra of subsets of a set S. If μ, $\nu \in \mathrm{ba}(S)$, then*

(a) $(\mu \vee \nu)(A) = \sup\{\mu(B) + \nu(C) \colon A = B \uplus C\}$ *and*
(b) $(\mu \wedge \nu)(A) = \inf\{\mu(B) + \nu(C) \colon A = B \uplus C\}$

for every $A \in \mathfrak{A}$.

14.1.20. Example. Let $S = \mathbb{Z} \cap [-5, 5]$ (that is, the integers between -5 and 5) and $\mathfrak{A} = \mathfrak{P}(S)$. Then $\mathfrak{A}$ is an algebra of sets. Define $\mu \colon \mathfrak{A} \to \mathbb{R}$ by setting $\mu(A)$ to be the sum of the elements in A for each nonempty $A \subseteq S$. (Also take $\mu(\emptyset) = 0$.) Then μ belongs to the Riesz space $\mathrm{ba}(S)$. For $A = \{-1, 0, 1\}$ we see that $(\mu \vee 0)(A) \ne \mu(A) \vee 0(A)$ and $|\mu|(A) \ne |\mu(A)|$.

The following proposition gives a few formulas which are useful when working with Riesz spaces. You can find a much more extensive list (43 formulas, in fact) in [41], pages 55–57.

14.1.21. Proposition. *Let E be a Riesz space and x, $y \in E$. Then*

(a) $-|x| \le x \le |x| = |-x|$
(b) $|x + y| \le |x| + |y|$
(c) $\big|\,|x| - |y|\,\big| \le |x - y|$
(d) $x \vee y = \frac{1}{2}(x + y + |x - y|)$
(e) $x \wedge y = \frac{1}{2}(x + y - |x - y|)$
(f) $(x \vee y) + (x \wedge y) = x + y$
(g) $(x \vee y) - (x \wedge y) = |x - y|$
(h) $|x| \vee |y| = \frac{1}{2}(\,|x + y| + |x - y|\,)$
(i) $|x| \wedge |y| = \frac{1}{2}\,\big|\,|x + y| - |x - y|\,\big|$

14.1.22. Proposition. *Let A be a subset of a Riesz space E and suppose that* $\sup A$ *exists. Then for every $b \in E$*

$$(\sup A) \wedge b = \sup\{a \wedge b \colon a \in A\}\,.$$

14.1.23. Remark. In studying lattices one often is told that some proposition holds "by duality". This particular form of duality can be regarded as a metamathematical principle derived from the observation that the operations $\vee$ and $\wedge$ appear symmetrically in the definition of "lattice". Notice that if $(S, \leq)$ is a partially ordered set, then so is $(S^{\text{op}}, \preceq)$ where $x \preceq y$ holds in S^{op} if and only if $y \leq x$ in S. The partially ordered set $(S^{\text{op}}, \preceq)$ is the DUAL (or OPPOSITE) of $(S, \leq)$. Any proposition in $(S, \leq)$ becomes a proposition of $(S^{\text{op}}, \preceq)$ if we reverse the order relation (change every occurrence of $\leq$ to $\preceq$, which is more commonly written as $\geq$) and interchange $\vee$ and $\wedge$, *lower bound* and *upper bound*, *minimum* and *maximum*, sup and inf, and so on. In Riesz spaces and, more generally, lattices, where there is by definition a symmetry between $\leq$ and $\geq$, propositions remain true when the interchanges mentioned above are made. (And so do the proofs!) Thus for example the two assertions of proposition 5.4.11 are dual to one another. It is standard to prove one of them and say that *the other holds by duality.* Don't be surprised if in a text or paper you find the dual of a proposition used without the author even mentioning duality.

14.1.24. Exercise. Duality like any other tool is susceptible of misuse. Explain carefully why "duality" cannot be invoked to argue that proposition 5.2.9 is obviously true. What is the dual assertion of 5.2.9?

14.1.25. Exercise. What is the dual of proposition 14.1.22?

14.1.26. Definition. Let F be a vector subspace of a Riesz space E. If $x \vee y$ (and $x \wedge y$) belong to F whenever x and y do, then F is a RIESZ SUBSPACE of E. **BEWARE:** the condition is *not* that x and y have a supremum in F, but that their supremum in E belongs to F. (See the following example.)

14.1.27. Example. The family of polynomial functions on the interval $[0, 1]$ provide an example of a linear subspace of a Riesz space $\mathcal{C}([0, 1])$ which is not a Riesz subspace of $\mathcal{C}([0, 1])$.

14.1.28. Proposition. *Let F be a vector subspace of a Riesz space E. In order for F to be a Riesz subspace of E it is sufficient that $x \vee \mathbf{0}$ (the supremum of x and $\mathbf{0}$ in E) belong to F whenever x does.*

Hint for solution. This follows easily from exercise 14.1.2.

14.1.29. Proposition. *If a, b, and c are positive elements of a Riesz space, then*

$$a \wedge (b + c) \leq (a \wedge b) + (a \wedge c)\,.$$

Hint for proof. Let $x = a \wedge (b + c)$. Show that $x - c$ is dominated by both a and b. Then show that $x - (a \wedge b)$ is dominated by both a and c.

14.1.30. Proposition (Riesz decomposition lemma)**.** *If E is a Riesz space and x, y, u, and v are members of E^+ such that $x + y = u + v$, then there exist a, b, c, and $d \in E^+$ such that*

$$\begin{aligned} x &= a + b \\ y &= c + d \\ u &= a + c \\ v &= b + d \end{aligned}$$

In other words, given x, y, u, and v as above, we can find positive "summands" which satisfy the following addition table.

$+$	b	c
a	x	u
d	v	y

Hint for proof. Let $a = x \wedge u$. The hardest part is showing that d is positive. Show first that $b \wedge c = 0$. Then use the fact that $b \le v + c$ to write b as $b \wedge (c + v)$.

14.1.31. Notation. If a and b are elements of an ordered vector space V, let

$$[a, b] := \{x \in V : a \le x \le b\}.$$

(Obviously, $[a, b]$ is empty unless $a \le b$.) This is the ORDER INTERVAL between a and b.

14.1.32. CAUTION. Notice that we are using the same notation for order intervals that we use for closed intervals on the real line and closed segments in vector spaces (see the cautionary note 9.2.3). It should be clear from context which is intended.

14.1.33. Proposition. *If E is a Riesz space and x, $y \in E^+$, then*

$$[0, x] + [0, y] = [0, x + y]\,.$$

14.1.34. Definition. A sequence (x_n) in a partially ordered set S is INCREASING if $x_m \le x_n$ whenever $m \le n$. It is DECREASING if $x_m \ge x_n$ whenever $m \le n$. A sequence is MONOTONE if it is either increasing or decreasing.

If (x_n) is an increasing sequence in S and $a = \sup\{x_n\}$, we write $x_n \uparrow a$. Similarly, $x_n \downarrow a$ means that the sequence (x_n) is decreasing and $a = \inf\{x_n\}$. In either case we say that the sequence CONVERGES and that the convergence of (x_n) to a is MONOTONE.

14.1.35. Definition. A Riesz space E is ARCHIMEDEAN if $n^{-1}x \downarrow 0$ whenever $x \in E^+$.

14.1.36. Example. The lexicographically ordered plane is not Archimedean. Recall that the *lexicographic ordering* on $\mathbb{R}^2$ is defined by

$$(u, v) \le (x, y) \text{ if and only if either } u < x \text{ or else } u = x \text{ and } v \le y.$$

14.1.37. Example. Every σ-Dedekind complete Riesz space is Archimedean.

14.2. Riesz Homomorphisms and Positive Operators

14.2.1. Definition. A linear map $T : E \to F$ between Riesz spaces which is also a lattice homomorphism is a RIESZ HOMOMORPHISM.

14.2.2. Proposition. *For a linear map $T : E \to F$ between Riesz spaces the following are equivalent:*

(a) $T(x \vee y) = Tx \vee Ty$ *for all* $x, y \in E$.
(b) T *is a Riesz homomorphism.*
(c) $T(x^+) = (Tx)^+$ *for all* $x \in E$.
(d) $T(x \wedge y) = Tx \wedge Ty$ *for all* $x, y \in E$.
(e) $Tx \wedge Ty = 0$ *in* F *whenever* $x \wedge y = 0$ *in* E.
(f) $|Tx| = T|x|$ *for all* $x \in E$.

14.2.3. Example. The class of Riesz spaces and the accompanying Riesz homomorphisms form a category. It will be denoted by **RS**.

14.2.4. Proposition. *In the category* **RS** *of Riesz spaces and Riesz homomorphisms, every bijective morphism is an isomorphism.*

14.2.5. Definition. A linear map $T\colon V \to W$ between ordered vector spaces is POSITIVE if $Tx \in W^+$ whenever $x \in V^+$. (Notice that T is positive if and only if it is order preserving.) As one would expect positive members of the algebraic dual (that is, positive linear maps from V into $\mathbb{R}$) are called POSITIVE LINEAR FUNCTIONALS on V.

14.2.6. Proposition. *Every Riesz homomorphism between Riesz spaces is positive.*

But not every positive linear map is a Riesz homomorphism.

14.2.7. Example. On the Riesz space $\mathcal{B}([0,2])$ of bounded real valued functions on the interval $[0,2]$ consider the map $T\colon \mathcal{B}([0,2]) \to \mathbb{R}\colon f \mapsto \int_0^2 f(x)\,dx$. This map is positive but is not a Riesz homomorphism. (If $g = \chi_{[0,1]}$ and $h = \chi_{[1,2]}$, then $T(g \vee h) \neq Tg \vee Th$.)

14.2.8. Example. Together the class of Riesz spaces and the positive linear maps between them constitutes a category. We will denote it by **RSP**.

14.2.9. Example. In the category **RSP** of Riesz spaces and positive linear maps a bijective morphism need not be an isomorphism. For example, the identity mapping from $\mathbb{R}^2$ with its usual ordering (see examples 9.3.6 and 14.1.4) to $\mathbb{R}^2$ with lexicographic ordering (see examples 9.3.7 and 14.1.5) is a positive bijection but not an isomorphism in the category **RSP**.

By definition, a mapping $T\colon E \to F$ between Riesz spaces is a RIESZ ISOMORPHISM if it is bijective and both T and T^{-1} are Riesz homomorphisms.

14.2.10. Proposition. *A linear bijection between Riesz spaces is a Riesz isomorphism if and only if both T and T^{-1} are positive. In other words, the categories* **RS** *and* **RSP** *have the same isomorphisms.*

14.2.11. Proposition. *Let E and F be Riesz spaces and suppose that F is Archimedean. Then every additive map $S\colon E^+ \to F^+$ extends uniquely to a positive linear map $T\colon E \to F$.*

Hint for proof. Recall that a map f between vector spaces is ADDITIVE if $f(x+y) = f(x)+f(y)$ for all x and y. When you try to prove additivity of the extension (whose definition is probably obvious), keep in mind that $(x+y)^+$ is not the same thing as $x^+ + y^+$. Show that if $x = u - v$ where u, $v \in E^+$, then $Tx = Su - Sv$. For the homogeneity of the extension you may want to prove that $Sx \le Sy$ whenever $0 \le x \le y$ and that $S(rx) = rS(x)$ for positive rational numbers r and $x \in E^+$.

14.2.12. Definition. Let S, $T\colon V \to W$ be linear maps between ordered vector spaces. We say that T DOMINATES (or MAJORIZES) S on V if $S(x) \le T(x)$ for all $x \in V^+$.

14.2.13. Proposition. *Let E be a Riesz space, F a Dedekind complete Riesz space, and G a Riesz subspace of E. Suppose that $S_0\colon G \to F$ and $T\colon E \to F$ are positive and linear and that T dominates S_0 on G. Then S_0 can be extended to a positive linear operator S on all of E which is dominated by T.*

14.2.14. Definition. A linear map $T\colon V \to W$ between ordered vector spaces is REGULAR if it is the difference of positive linear maps. We denote by $\mathfrak{L}_r(V,W)$ the family of all regular (linear) maps from V to W.

14.2.15. Proposition. *A linear map $T\colon E \to F$ between ordered vector spaces is regular if and only if it is dominated by a positive linear map.*

14.3. The Order Dual of a Riesz Space

14.3.1. Definition. A subset of an ordered vector space V is ORDER BOUNDED if it is bounded above and bounded below by elements of V.

14.3.2. Proposition. *Let E be a Riesz space.*

(a) *An element x belongs to the order interval $[-u, u]$ (where $u \in E^+$) if and only if $|x| \leq u$.*
(b) *A subset A of E is order bounded if and only if there exists $u \in E^+$ such that $|a| \leq u$ for all $a \in A$.*

14.3.3. Definition. A linear map $T\colon V \to W$ between Riesz spaces is ORDER BOUNDED if it takes order bounded sets to order bounded sets; that is, if $T(A)$ is an order bounded subset of W whenever A is an order bounded subset of V. We denote by $\mathfrak{L}_b(V, W)$ the family of all order bounded (linear) maps from V to W.

14.3.4. Proposition. *If $T\colon V \to W$ is a linear map between ordered vector spaces, then the following are equivalent:*

(i) *T is order bounded;*
(ii) *T takes order intervals into order intervals; and*
(iii) *T takes order intervals of the form $[0, x]$ into order intervals.*

14.3.5. Proposition. *Every positive linear map between Riesz spaces is order bounded.*

14.3.6. Proposition. *Every regular linear map between ordered vector spaces is order bounded.*

14.3.7. Proposition. *If E and F are Riesz spaces and F is Dedekind complete, then $\mathfrak{L}_r(E, F) = \mathfrak{L}_b(E, F)$.*

14.3.8. Proposition. *If E and F are Riesz spaces and F is Dedekind complete, then $\mathfrak{L}_b(E, F)$ is a Dedekind complete Riesz space with the set of positive operators from V to W as its positive cone.*

14.3.9. Notation. We denote by $E^\sim$ the family of all order bounded linear functionals on a Riesz space E. This is the ORDER DUAL of E.

14.3.10. Proposition. *If E is a Riesz space, then so is its order dual $E^\sim$.*

Hint for proof. Use the set of all positive linear functionals as the positive cone. To show that the supremum of an order bounded linear functional f and the zero functional exists, define g_0 on E^+ by

$$g_0(x) = \sup\{f(u)\colon u \in [0, x]\}\,.$$

14.3.11. Corollary. *Every order bounded linear functional on a Riesz space is the difference of two positive linear functionals.*

14.3.12. Proposition. *Let E be a Riesz space and $f \in E^\sim$. Then for every $x \in E^+$*

(a) *$f^+(x) = \sup\{f(u)\colon u \in [0, x]\}$;*
(b) *$f^-(x) = \sup\{-f(u)\colon u \in [0, x]\}$; and*
(c) *$|f|(x) = \sup\{f(u)\colon |u| \leq x\}$;*

14.3.13. Proposition. *Let E be a Riesz space and $f \in E^\sim$. Then for every $x \in E$*

(a) *$|f|(|x|) = \sup\{f(u)\colon |u| \leq |x|\}$ and*
(b) *$|f(x)| \leq |f|(|x|)$.*

14.3.14. Proposition. *Let E be a Riesz space and $f, g \in E^\sim$. Then for every $x \in E^+$*

(a) *$(f \vee g)(x) = \sup\{f(u) + g(x - u)\colon u \in [0, x]\}$ and*
(b) *$(f \wedge g)(x) = \inf\{f(u) + g(x - u)\colon u \in [0, x]\}$.*

14.3.15. Proposition. *The order dual of a Riesz space is always Dedekind complete.*

Hint for proof. Start with a nonempty subset A of the order dual that is bounded above. Show that without loss of generality we may assume that all the elements of A are positive. For each positive element x of the Riesz space let $h(x)$ be the supremum of the set of all numbers $s(x)$ where s is a finite supremum of elements of A. Show $h = \sup A$.

CHAPTER 15

MEASURABLE SPACES

15.1. σ-Algebras of Sets

15.1.1. Definition. A nonempty family $\mathfrak{A}$ of subsets of a nonempty set S is a σ-ALGEBRA if it is closed under complements and countable unions. That is, we require that

(1) If $\mathfrak{B}$ is a countable subfamily of $\mathfrak{A}$, then $\bigcup \mathfrak{B} \in \mathfrak{A}$; and
(2) If $A \in \mathfrak{A}$, then $A^c \in \mathfrak{A}$.

If $\mathfrak{A}$ is a σ-algebra of subsets of a set S, then we say that the pair $(S, \mathfrak{A})$ is a MEASURABLE SPACE. When it is thought that no confusion will result, this is often shortened to "S is a measurable space." When A belongs to $\mathfrak{A}$, we may say that A is a *measurable subset* of S and write $A \overset{m}{\subseteq} S$ (rather than $A \in \mathfrak{A}$). Compare this with the definitions given for *algebras of sets* and *topologies* given in 5.6.11 and 10.1.1.

15.1.2. Exercise. We have just defined a σ-algebra to be a (nonempty) family of sets which is closed under taking of complements and countable unions. Show that this is equivalent to defining a σ-algebra to be a (nonempty) family of sets which is closed under taking of complements and countable intersections.

15.1.3. Exercise. Let $\mathfrak{F}$ be a family of subsets of a nonempty set S. Explain carefully what we mean by "the smallest σ-algebra of sets containing $\mathfrak{F}$" or "the σ-algebra generated by $\mathfrak{F}$. Prove that such a thing exists. (See remark 5.3.6.)

15.1.4. Notation. If S is a set and $\mathfrak{F}$ is a family of subsets of S, let $\sigma_S(\mathfrak{F})$ (or, if no confusion will result, $\sigma(\mathfrak{F})$) be the σ-algebra of subsets of S generated by $\mathfrak{F}$ (that is, the smallest σ-algebra of subsets of S which contains $\mathfrak{F}$).

It is an interesting, if somewhat unpleasant, aspect of σ-algebras that there appears to be no good *constructive* method of describing the smallest such algebra containing a given family of sets. In many cases it is possible to keep generating new sets by alternately taking countable unions and complements of sets we already have. As an illustration of the resulting complexity you may wish to consider the family $\mathfrak{F}$ of bounded open intervals in $\mathbb{R}$. The smallest σ-algebra containing $\mathfrak{F}$ is called the family of *Borel sets* in $\mathbb{R}$. It is entertaining to try and find a subset of $\mathbb{R}$ which is *not* a Borel set. Also notice how difficult it would be to describe an arbitrary Borel set in terms of successive unions and complements of the intervals in $\mathfrak{F}$. We will discuss Borel sets in more detail in the next section 15.2.

15.1.5. Example. Let S be an infinite set and $\mathfrak{F} = \{\{x\} \colon x \in S\}$. Then $\sigma_S(\mathfrak{F})$ comprises all countable and cocountable subsets of S. (A subset A of a set S is COCOUNTABLE if its complement is countable.)

15.1.6. Example. Let $S = [0, 1)$ and $\mathfrak{A}$ be the family of all subsets of S which are finite unions of sets of the form $[a, b)$ where $a, b \in S$. Then $\mathfrak{A}$ is an algebra of sets but not a σ-algebra.

15.1.7. Proposition. *Let $f \colon S \to T$ be a function from a measurable space $(S, \mathfrak{A})$ into a set T. If $\mathfrak{B} = \{B \subseteq T \colon f^{\leftarrow}(B) \in \mathfrak{A}\}$, then $(T, \mathfrak{B})$ is a measurable space.*

In the next result we make use of the notation $\mathfrak{F} \cap A$ where $\mathfrak{F}$ is a family of sets and A is a set. This means what you probably think it means:

$$\mathfrak{F} \cap A = \{F \cap A \colon F \in \mathfrak{F}\}.$$

15.1.8. Proposition. *Let $\mathfrak{F}$ be a family of subsets of a set S and $A \subseteq S$. Then*

$$\sigma_A(\mathfrak{F} \cap A) = \sigma_S(\mathfrak{F}) \cap A.$$

Hint for proof. It is clear that $\sigma_A(\mathfrak{F} \cap A) \subseteq \sigma_S(\mathfrak{F}) \cap A$. To establish the reverse inclusion let $\mathfrak{G} = \{B \in \sigma_S(\mathfrak{F}) \colon B \cap A \in \sigma_A(\mathfrak{F} \cap A)\}$. Show that $\mathfrak{G}$ is a σ-algebra which contains $\mathfrak{F}$. This is all we need, for then $\sigma_S(\mathfrak{F}) \subseteq \mathfrak{G}$. In his lovely text [4] Ash labels this the *good sets principle*. A set in $\sigma_S(\mathfrak{F})$ is "good" if its intersection with A belongs to $\sigma_A(\mathfrak{F} \cap A)$. So $\mathfrak{G}$ is the family of all "good" sets. All we are required to show is that every member of $\sigma_S(\mathfrak{F})$ is "good".

15.1.9. Proposition. *Let S and T be sets, $\Phi \colon S \to T$, and $\mathfrak{F} \subseteq \mathfrak{P}(T)$. Then*

$$\sigma_S(\Phi^{\leftarrow}(\mathfrak{F})) = \Phi^{\leftarrow}(\sigma_T(\mathfrak{F})).$$

15.1.10. Exercise. Show how to use proposition15.1.9 to give a *very* easy proof of proposition 15.1.8.

15.2. Borel Sets

In a topological space the most important σ-algebra is the family of *Borel sets*, the σ-algebra generated by the open sets.

15.2.1. Definition. Let $(X, \mathfrak{T})$ be a topological space. The BOREL SETS of X are the sets in $\sigma(\mathfrak{T})$. Denote the family of Borel sets in X by $\mathfrak{B}\mathrm{or}(X)$.

15.2.2. Example. Every countable and every cocountable subset of $\mathbb{R}$ is a Borel set.

15.2.3. Definition. A nonempty family $\mathfrak{A}$ of subsets of a set S is a DYNKIN SYSTEM if it satisfies the following conditions:

(i) $S \in \mathfrak{A}$;
(ii) if A, $B \in \mathfrak{A}$ and $A \subseteq B$, then $B \setminus A \in \mathfrak{A}$; and
(iii) if (A_n) is an increasing sequence of sets in $\mathfrak{A}$ with $B = \bigcup_{n=1}^{\infty} A_n$, then $B \in \mathfrak{A}$.

15.2.4. Proposition. *A nonempty family of subsets of a set S is a σ-algebra if and only if it is a Dynkin system which is closed under finite intersections.*

15.2.5. Exercise. Let $S = \{a, b, c, d\}$ be a four-element set. Find a Dynkin system in S which is not an algebra of sets.

15.2.6. Proposition. *Let $\mathfrak{A}$ be a Dynkin system in a set S and $\mathfrak{F}$ be a nonempty family of sets which is contained in $\mathfrak{A}$. If $\mathfrak{F}$ is closed under finite intersections, then $\sigma(\mathfrak{F}) \subseteq \mathfrak{A}$.*

15.2.7. Definition. A nonempty family $\mathfrak{M}$ of subsets of a set S is a MONOTONE CLASS provided that it satisfies the following condition: if (A_n) is a sequence of sets in $\mathfrak{M}$ such that either $A_n \uparrow B$ or $A_n \downarrow B$, then $B \in \mathfrak{M}$.

15.2.8. Proposition. *Every σ-algebra is a Dynkin system and every Dynkin system is a monotone class.*

15.2.9. Proposition. *If $\mathfrak{A}$ is an algebra of subsets of a set S, then $\sigma(\mathfrak{A})$ is the smallest monotone class in S which contains $\mathfrak{A}$.*

15.2.10. Corollary. *An algebra of sets is a σ-algebra if and only if it is a monotone class.*

15.2.11. Proposition. *If X is a topological space, then the family $\mathfrak{B}\mathrm{or}(X)$ of Borel sets in X is the smallest Dynkin system containing the topology of X.*

15.2.12. Definition. A subset of a topological space X is an F_σ SUBSET of X if it is a countable union of closed subsets of X. It is a G_δ SUBSET of X if it is a countable intersection of open subsets of X. One way of keeping these definitions straight is to associate the F in F_σ with the word "*fermé*", which is French for *closed* and the σ with "Summe", which is German for *sum*.

The connection here is that some older monographs use "sum" for "union" (or more specifically, "disjoint union") of sets and "product" for "intersection". Perhaps this makes the "algebra of sets" look more algebraic. Also, you may wish to associate the G in G_δ with the German word "*Gebiet*, which means *region* (suggesting *open*) and the δ with "Durchschnitt", which is German for *intersection*.

15.2.13. Example. Every G_δ subset of a topological space X is a Borel subset of X; so is every F_σ subset.

15.2.14. Proposition. *A subset of a topological space is an F_σ set if and only if its complement is a G_δ set.*

15.2.15. Example. In a metric space every closed set is a G_δ set; and every open set is an F_σ set.

Hint for proof. Let C be a closed subset of a metric space. Consider the function $x \mapsto d(x, C)$ defined on M.

15.2.16. Definition. Let f be a real valued function on a topological space X, a be a point in X, and $\mathfrak{U}_a$ be the family of all neighborhoods of a. We define the OSCILLATION $\omega_f(a)$ of f at a by

$$\omega_f(a) := \inf_{U \in \mathfrak{U}_a} \big\{ \sup_{x,y \in U} \{ |f(x) - f(y)| \} \big\}.$$

15.2.17. Proposition. *A real valued function f on a topological space is continuous at a point a in its domain if and only if $\omega_f(a) = 0$.*

15.2.18. Proposition. *Let f be a real valued function on a topological space X. The set $D(f)$ of points at which f is discontinuous is an F_σ subset of X.*

15.3. Measurable Functions

15.3.1. Definition. Let $(S, \mathfrak{A})$ and $(T, \mathfrak{B})$ be measurable spaces. A function $f \colon S \to T$ is $(\mathfrak{A}, \mathfrak{B})$-MEASURABLE if the inverse image under f of every measurable set is measurable; that is, if $f^{\leftarrow}(B) \in \mathfrak{A}$ whenever $B \in \mathfrak{B}$; equivalently, if $f^{\leftarrow}(B) \overset{m}{\subseteq} S$ whenever $B \overset{m}{\subseteq} T$. When no confusion will result, we say just that the function is MEASURABLE. The family of all measurable functions from S into T is denoted by $\mathcal{M}(S, T)$. In the frequently occurring case when $T = \mathbb{R}$ we shorten $\mathcal{M}(S, \mathbb{R})$ to $\mathcal{M}(S)$.

15.3.2. Convention. Whenever a topological space is referred to as a *measurable space*, without further specification, it is understood that the measurable sets are the Borel sets. Thus if $(S, \mathfrak{A})$ is a measurable space, a function $f \colon S \to \mathbb{R}$ is said to be *measurable* (or, for emphasis, *Borel measurable*) if it is $(\mathfrak{A}, \mathfrak{B}\mathrm{or}(\mathbb{R}))$-measurable. Thus for real valued functions *measurable* means *Borel measurable* unless the contrary is explicitly stated.

15.3.3. Example. On a measurable space the characteristic function of every measurable set is measurable.

15.3.4. Proposition. *The composite of any two measurable functions is measurable. (That is, if $f \colon R \to S$ and $g \colon S \to T$ are measurable functions between measurable spaces, then $g \circ f \colon R \to T$ is measurable.)*

15.3.5. Proposition. *Let $f \colon S \to X$ be a function from a measurable space S into a topological space X. Then f is measurable if and only if $f^{\leftarrow}(U)$ is measurable whenever $U \overset{\circ}{\subseteq} X$.*

Hint for proof. Proposition 15.1.9.

15.3.6. Corollary. *Every continuous function between topological spaces is (Borel) measurable. That is, if X and Y are topological spaces and $f \colon X \to Y$ is continuous, then the inverse image under f of every Borel set in Y is a Borel set in X.*

15.3.7. Proposition. *If S is a measurable space, then a real valued function f on S is measurable if and only if $f^{\leftarrow}\big((a,\infty)\big)$ is measurable for every $a \in \mathbb{R}$.*

15.3.8. Proposition. *If S is a measurable space and f and g are real valued measurable functions on S, then the function $(f,g)\colon S \to \mathbb{R}^2$ is measurable.*

In the next proposition the codomain of a function f is $\mathbb{R}^2$. For $k = 1, 2$ we define $f^k = \pi_k \circ f$ where π_k is the k^{th} projection function, $\pi_k\colon \mathbb{R}^2 \to \mathbb{R}\colon (x_1, x_2) \mapsto x_k$. (See definition 8.3.6.) We refer to f^1 and f^2 as the *coordinate functions* of f.

15.3.9. Proposition. *Let S be a measurable space. A function $f\colon S \to \mathbb{R}^2$ is measurable if and only if its coordinate functions f^1 and f^2 are.*

15.3.10. Proposition. *Let S be a measurable space. If $f\colon S \to \mathbb{R}$ is measurable and $p \geq 0$, then $|f|^p$ is measurable. If $f\colon S \to \mathbb{R}$ is measurable and is never zero, then the function $1/f$ is measurable.*

15.3.11. Example. The family $\mathcal{M}(S)$ of all real valued measurable functions on a measurable space S is a Riesz space (under the usual pointwise operations).

15.3.12. Example. The family $\mathcal{M}(S)$ of all real valued measurable functions on a measurable space S is an algebra (under the usual pointwise operations).

15.3.13. Proposition. *A complex valued function on a measurable space is measurable if and only if its real and imaginary parts are.*

15.3.14. Proposition. *Let f be an extended real valued function defined on a measurable space $(S, \mathfrak{A})$. If $f^{\leftarrow}\big((a,\infty]\big) \in \mathfrak{A}$ for every $a \in \mathbb{R}$, then f is measurable.*

Hint for proof. Use proposition 15.1.7 and example 11.1.9.

15.3.15. Proposition. *Let (f_n) be a sequence of measurable real valued functions defined on a measurable space S. Then $\bigvee f_n$, $\bigwedge f_n$, $\liminf f_n$, and $\limsup f_n$ are all measurable. Thus if $f_n \to g$ (ptws), then g is measurable.*

15.3.16. Exercise. Let $S = [0,1] \times [0,1]$ and $\mathfrak{A}$ be the σ-algebra of subsets of S generated by sets of the form $U \times V$ where either U or U^c is countable and either V or V^c is countable. Does the diagonal $D = \{(x,x)\colon 0 \leq x \leq 1\}$ belong to $\mathfrak{A}$?

Hint for solution. If E and F are subsets of S and F is uncountable, say that E is LARGE IN F if $F \setminus E$ is countable and that E is SMALL IN F if $F \cap E$ is countable. Let $A = \{(x, 1-x)\colon 0 \leq x \leq 1\}$. Consider the family $\mathfrak{E}$ of all subsets of S which are either large in both A and D or else small in both A and D.

15.4. Functors

15.4.1. Definition. If $\mathbf{A}$ and $\mathbf{B}$ are categories a COVARIANT FUNCTOR F from $\mathbf{A}$ to $\mathbf{B}$ (written $\mathbf{A} \xrightarrow{F} \mathbf{B}$) is a pair of maps: an OBJECT MAP F which associates with each object S in $\mathbf{A}$ an object $F(S)$ in $\mathbf{B}$ and a MORPHISM MAP (also denoted by F) which associates with each morphism $f \in \mathfrak{Mor}(S,T)$ in $\mathbf{A}$ a morphism $F(f) \in \mathfrak{Mor}(F(S), F(T))$ in $\mathbf{B}$, in such a way that

(1) $F(g \circ f) = F(g) \circ F(f)$ whenever $g \circ f$ is defined in $\mathbf{A}$; and
(2) $F(\mathrm{id}_S) = \mathrm{id}_{F(S)}$ for every object S in $\mathbf{A}$.

The definition of a CONTRAVARIANT FUNCTOR $\mathbf{A} \xrightarrow{F} \mathbf{B}$ differs from the preceding definition only in that, first, the morphism map associates with each morphism $f \in \mathfrak{Mor}(S,T)$ in $\mathbf{A}$ a morphism $F(f) \in \mathfrak{Mor}(F(T), F(S))$ in $\mathbf{B}$ and, second, condition (1) above is replaced by

(1′) $F(g \circ f) = F(f) \circ F(g)$ whenever $g \circ f$ is defined in $\mathbf{A}$.

15.4.2. Example. Let S be a nonempty set.

(a) The power set $\mathfrak{P}(S)$ of S partially ordered by $\subseteq$ is an order complete lattice.
(b) The class of order complete lattices and order preserving maps is a category.
(c) For each function f between sets let $\mathfrak{P}(f) = f^{\rightarrow}$. Then $\mathfrak{P}$ is a covariant functor from the category of sets and functions to the category of order complete lattices and order preserving maps.
(d) For each function f between sets let $\mathfrak{P}(f) = f^{\leftarrow}$. Then $\mathfrak{P}$ is a contravariant functor from the category of sets and functions to the category of order complete lattices and order preserving maps.
(e) In parts (c) and (d) can the phrase "order preserving maps" be replaced by "lattice homomorphisms"?

15.4.3. Example. Let X and Y be topological spaces and $\phi\colon X \to Y$ be continuous. Define $\mathcal{C}\phi$ on $\mathcal{C}(Y)$ by

$$\mathcal{C}\phi(g) = g \circ \phi$$

for all $g \in \mathcal{C}(Y)$. Then the pair of maps $X \mapsto \mathcal{C}(X)$ and $\phi \mapsto \mathcal{C}\phi$ is a contravariant functor from the category of topological spaces and continuous maps to the category of unital algebras and unital algebra homomorphisms.

15.4.4. Example. For every set S let S^σ be the family of all σ-algebras of subsets of S. If $\phi\colon S \to T$ where S and T are sets, define ϕ^σ on S^σ by

$$\phi^\sigma(\mathfrak{A}) = \{E \subseteq T \colon \phi^{\leftarrow}(E) \in \mathfrak{A}\}.$$

(a) S^σ is an order complete lattice.
(b) ϕ^σ maps S^σ into T^σ.
(c) The pair of maps $S \mapsto S^\sigma$, $\phi \mapsto \phi^\sigma$ is a covariant functor from the category of sets and maps to the category of order complete lattices and order preserving maps.

15.4.5. Exercise. Use example 15.4.4 to give a short proof of proposition 15.3.5.

15.4.6. Example. For each bounded linear map $T\colon V \to W$ between normed linear spaces define

$$T^*\colon W^* \to V^*\colon g \mapsto g \circ T.$$

The map T^* is the ADJOINT of T. Then the pair of maps $V \mapsto V^*$ and $T \mapsto T^*$ is a contravariant functor from the category $\mathbf{NLS_\infty}$ of normed linear spaces and bounded linear maps into itself. We will refer to it as the *duality functor* on $\mathbf{NLS_\infty}$.

15.4.7. Example. If $\mathbf{C}$ is a category let $\mathbf{C^2}$ be the category whose objects are ordered pairs of objects in $\mathbf{C}$ and whose morphisms are ordered pairs of morphisms in $\mathbf{C}$. Thus if $A \xrightarrow{f} C$ and $B \xrightarrow{g} D$ are morphisms in $\mathbf{C}$, then $(A, B) \xrightarrow{(f,g)} (C, D)$ (where $(f, g)(a, b) = \big(f(a), g(b)\big)$ for all $a \in A$ and $b \in B$) is a morphism in $\mathbf{C^2}$. Composition of morphism is defined in the obvious way: $(f, g) \circ (h, j) = (f \circ h, g \circ j)$. We define the DIAGONAL FUNCTOR $\mathbf{C} \xrightarrow{\mathsf{D}} \mathbf{C^2}$ by $\mathsf{D}(A) := (A, A)$. This is a covariant functor.

CHAPTER 16

THE RIESZ SPACE OF REAL MEASURES

Locally compact spaces (in topology), positive measures (in analysis), and algebras (in operator theory) present a common pedagogical problem. In the study of each there is an irresistibly attractive special case (compact spaces, real measures, and unital algebras) with properties so pleasant and structure so pellucid that one feels almost compelled to present the special case first. But here is the problem. By far the most important example of a locally compact space is the real line, which is not compact; the most important measure is Lebesgue measure, which is not a real measure; and many algebras central to operator theory, the compact operators and the continuous functions on $\mathbb{R}$ which vanish at infinity, for example, fail to be unital. My solution to the problem is to give in to temptation. In these notes we look at real measures in this chapter and Lebesgue measure later (in chapter 18). In chapter 17 compact spaces come first, locally compact spaces at the end. And in chapter 21 unital algebras are emphasized.

16.1. Real Measures

16.1.1. Definition. Let $(S, \mathfrak{A})$ be a measurable space. A complex (or real) valued function μ on $\mathfrak{A}$ is COUNTABLY ADDITIVE if

$$\mu(B) = \sum_{k=1}^{\infty} \mu(A_k) \qquad \text{whenever} \qquad B = \biguplus_{k=1}^{\infty} A_k\,. \tag{16.1}$$

A countably additive function $\mu\colon \mathfrak{A} \to \mathbb{R}$ is a REAL MEASURE and a countably additive function from $\mathfrak{A}$ into $\mathbb{C}$ is a COMPLEX MEASURE. Notice that the series $\sum_{k=1}^{\infty} \mu(A_k)$ is absolutely convergent, since any rearrangement of the series will have the same sum $\mu(B)$.

In case the function μ maps $\mathfrak{A}$ into either $(-\infty, \infty]$ or $[-\infty, \infty)$ we will say that it is COUNTABLY ADDITIVE if it satisfies (16.1) and also $\mu(\emptyset) = 0$. (This last condition excludes the uninteresting case of set functions whose value on every set is infinite.) A countably additive set function $\mu\colon \mathfrak{A} \to [0, \infty]$ is a POSITIVE MEASURE; and many authors call a countably additive function from $\mathfrak{A}$ into either $(-\infty, \infty]$ or $[-\infty, \infty)$ a SIGNED MEASURE. Care must be taken to distinguish between positive measures and positive real measures: the former may take on infinite values, the latter may not.

A REAL MEASURE SPACE is a triple $(S, \mathfrak{A}, \mu)$ where $(S, \mathfrak{A})$ is a measurable space and μ is a real measure on $\mathfrak{A}$. Similarly, a COMPLEX (respectively, POSITIVE or SIGNED MEASURE SPACE is a triple $(S, \mathfrak{A}, \mu)$ where $(S, \mathfrak{A})$ is a measurable space and μ is a complex (respectively, positive or signed) measure on $\mathfrak{A}$.

In an effort to simplify technical exposition and remove notational clutter authors introduce many variants of the preceding terminology. Instead of the expression, "$(S, \mathfrak{A}, \mu)$ is a real measure space", you will regularly find in the literature (including these notes) such replacements as "μ is a real measure on S", "S is a real measure space", and "the real measure space (S, μ)". It is not unreasonable to hope that these shortened forms, despite being utterly illogical, will cause no confusion. Also I should point out that in many treatments the word *measure* means what is here called *positive measure*.

We denote by $\operatorname{ca}(S, \mathfrak{A})$ the family of all real measures on a measurable space $(S, \mathfrak{A})$. When no confusion appears likely we write just $\operatorname{ca}(S)$. Similarly, for the family of complex measures on the measurable space S we write $\operatorname{ca}(S, \mathfrak{A}, \mathbb{C})$ or just $\operatorname{ca}(S, \mathbb{C})$.

16.1.2. Example. Let S be a set, $b \in S$, and $\mathfrak{A} = \mathfrak{P}(S)$. For every $A \in \mathfrak{A}$ define

$$\delta_b(A) = \begin{cases} 1, & \text{if } b \in A \\ 0, & \text{if } b \notin A. \end{cases}$$

Then $(S, \mathfrak{A}, \delta_b)$ is a positive real measure space and the positive real measure δ_b is called DIRAC MEASURE CONCENTRATED AT b.

16.1.3. Example. Let S be a nonempty set and $\mathfrak{A} = \mathfrak{P}(S)$. For every $A \in \mathfrak{A}$ define

$$\mu(A) = \begin{cases} \operatorname{card} A, & \text{if } A \text{ is finite} \\ \infty, & \text{if } A \text{ is infinite.} \end{cases}$$

Then $(S, \mathfrak{A}, \mu)$ is a positive measure space and the positive measure μ is called COUNTING MEASURE on S. If S is finite, then μ is a positive real measure.

16.1.4. Notation. If (A_n) is a sequence of subsets of a set S, then we write $A_n \uparrow B$ to indicate that the sequence (A_n) is increasing (that is, $A_n \subseteq A_{n+1}$ for each $n \in \mathbb{N}$) and that $\bigcup_{n=1}^{\infty} A_n = B$. Similarly, $A_n \downarrow B$ means that the sequence (A_n) is decreasing ($A_n \supseteq A_{n+1}$ for each $n \in \mathbb{N}$) and that $\bigcap_{n=1}^{\infty} A_n = B$. The next result says, roughly, that real measures are *σ-order continuous* (*sequentially order continuous*) from below and from above.

16.1.5. Proposition. *Let S be a measurable space, (A_n) be a sequence of measurable subsets of S, and μ be a real measure on S.*

(a) *If $A_n \uparrow B$, then $\mu(A_n) \to \mu(B)$.*
(b) *If $A_n \downarrow B$, then $\mu(A_n) \to \mu(B)$.*

The following partial converse of this result is occasionally useful in simplifying the task of proving that certain set functions are countably additive.

16.1.6. Proposition. *Let μ be a real finitely additive set function on a measurable space $(S, \mathfrak{A})$. Then the following are equivalent:*

(a) *μ is countably additive;*
(b) *if the sets A_n are measurable and $A_n \downarrow \emptyset$, then $\mu(A_n) \to 0$;*
(c) *if the sets A_n are measurable and $A_n \uparrow B$, then $\mu(A_n) \to \mu(B)$.*

16.1.7. Proposition. *Let μ be a positive real measure on a measurable space S. If $A_n \overset{m}{\subseteq} S$ for every $n \in \mathbb{N}$ and $B = \lim A_n$, then*

$$\mu(B) = \lim \mu(A_n).$$

Hint for proof. Consider the sets $B_n := \bigcap_{k=n}^{\infty} A_k$ and $C_n := \bigcup_{k=n}^{\infty} A_k$.

The next proposition proves that $\operatorname{ca}(S) \subseteq \operatorname{ba}(S)$. The definition of the function σ here will look formally like the definition in 14.1.16. The principal difference is that in the earlier example we were working with merely an algebra of sets and a function μ which was *assumed* to be finitely additive and bounded. Here the domain is a σ-algebra of sets; we assume that μ is countably additive and *prove* that it is bounded.

16.1.8. Proposition. *Every real measure is bounded. (That is, if μ is a real measure on a measurable space $(S, \mathfrak{A})$, then there exists a number $M > 0$ such that $|\mu(A)| \le M$ for all $A \overset{m}{\subseteq} S$.)*

Hint for proof. For each measurable subset A of S define

$$\sigma(A) := \sup\{\mu(B) \colon B \overset{m}{\subseteq} A\}.$$

Proving that $\sigma(S) < \infty$ shows that μ is bounded above. This may be accomplished in three steps. First show that if $\sigma(A) = \infty$ and if $A = B \uplus C$, then either $\sigma(B) = \infty$ or $\sigma(C) = \infty$. Next show that if $\sigma(A) = \infty$, then there exists a decomposition $A = B \uplus C$ such that $|\mu(B)| > 1$ and $\sigma(C) = \infty$.

(To do this show that if a subset of A is chosen to have measure greater than $|\mu(A)| + 1$, then its complement in A has measure greater than 1 in absolute value.) Finally, obtain inductively a disjoint family sets $B_1, B_2, \dots$ with $|\mu(B_n)| > 1$ for each n. Conclude that μ is bounded above.

16.1.9. Proposition. *If $(S, \mathfrak{A})$ is a measurable space, then* $\mathrm{ca}(S)$ *is a vector subspace of the Riesz space* $\mathrm{ba}(S)$ *(see example 14.1.17).*

We now turn to the problem of showing that the ordered vector space $\mathrm{ca}(S)$ is also a lattice (and therefore a Riesz space). According to exercise 14.1.2 it suffices to show that $\mu \vee 0$ exists in $\mathrm{ca}(S)$ for each real measure μ. The function σ defined in 16.1.8 above is a real valued set function. We must show next that it is countably additive and is, in fact, the supremum of $\{\mu, \mathbf{0}\}$ in $\mathrm{ca}(S)$. Once we have obtained the desired expression for $\mu^+ = \mu \vee 0$, it is an easy matter to derive similar expressions for μ^-, $\mu \vee \nu$, and $\mu \wedge \nu$. (Compare this with 14.1.19.)

16.1.10. Example. Let (S, μ) be a measurable space. Then $\mathrm{ca}(S)$ is a Riesz space. Furthermore, if μ, $\nu \in \mathrm{ca}(S)$, then the following formulas hold for all measurable subsets A of S.

(a) $\mu^+(A) = \sup\{\mu(B) \colon B \overset{m}{\subseteq} A\}$.
(b) $\mu^-(A) = -\inf\{\mu(B) \colon B \overset{m}{\subseteq} A\}$.
(c) $(\mu \vee \nu)(A) = \sup\{\mu(B) + \nu(C) \colon A = B \uplus C\}$.
(d) $(\mu \wedge \nu)(A) = \inf\{\mu(B) + \nu(C) \colon A = B \uplus C\}$.

Hint for proof. Let $A = \biguplus_{k=1}^{\infty} B_k$. Half of showing that the set function σ (defined in 16.1.8 above) is countably additive is proving $\sum_{k=1}^{\infty} \sigma(B_k) \le \sigma(A)$. This is facilitated by the following procedure: given $\epsilon > 0$, choose for each $k \in \mathbb{N}$ a set $C_k \overset{m}{\subseteq} B_k$ such that

$$\mu(C_k) > \sigma(B_k) - 2^{-k}\epsilon.$$

Look at $\mu\big(\biguplus_{k=1}^{\infty} C_k\big)$.

16.2. Ideals in Riesz Spaces

16.2.1. Definition. A subset A of a Riesz space E is SOLID if x belongs to A whenever $|x| \le |a|$ for some $a \in A$. A solid linear subspace of E is an ORDER IDEAL in E.

16.2.2. Proposition. *If x belongs to a solid subset A of a Riesz space, then $|x|$, x^+, and x^- all belong to A.*

16.2.3. Proposition. *Every order ideal in a Riesz space E is a Riesz subspace of E.*

16.2.4. Example. Consider the Riesz space $\mathcal{B}([0,1])$ of bounded real valued functions on the interval $[0,1]$. The set $\mathcal{C}([0,1])$ of continuous functions on $[0,1]$ is a Riesz subspace of but not an order ideal in $\mathcal{B}([0,1])$.

16.2.5. Example. Let X be a topological space. A linear subspace $\mathcal{L}$ of $\mathcal{C}(X)$ is a Riesz subspace of $\mathcal{C}(X)$ if and only if $|f|$ belongs to $\mathcal{L}$ whenever f does.

We can improve on example 16.1.10: not only is $\mathrm{ca}(S)$ a Riesz space contained in $\mathrm{ba}(S)$, it is an order ideal in $\mathrm{ba}(S)$.

16.2.6. Example. The family $\mathrm{ca}(S)$ of all real measures on a measurable space $(S, \mathfrak{A})$ is an order ideal in the Riesz space $\mathrm{ba}(S)$.

Hint for proof. To show that $\mathrm{ca}(S)$ is solid it suffices to show that if $\nu \in \mathrm{ca}(S)$, if $0 \le \mu \le \nu$, and if $A_n \uparrow B$, then $\mu(A_n) \to \mu(B)$. Explain.

16.2.7. Proposition. *If A is a nonempty solid subset of a Riesz space E and A^+ is closed under addition, then A is an order ideal in E.*

16.2.8. Proposition. *If J_1 and J_2 are order ideals in a Riesz space E, then $J_1 + J_2$ is also an ideal and $(J_1 + J_2)^+ = {J_1}^+ + {J_2}^+$.*

Hint for proof. Use the *Riesz decomposition lemma* 14.1.30.

16.2.9. Proposition. *The kernel of every Riesz homomorphism between Riesz spaces is an order ideal.*

16.2.10. Proposition. *Let J be an order ideal in a Riesz space E. If α and β are elements of the quotient vector space E/J, then the following are equivalent:*

(a) *There exist vectors $x \in \alpha$ and $y \in \beta$ such that $x \leq y$;*
(b) *For every $x \in \alpha$ there exists $y \in \beta$ such that $x \leq y$; and*
(c) *For every $x \in \alpha$ and for every $y \in \beta$ there exists $j \in J$ such that $y - x \geq j$.*

16.2.11. Proposition. *Let J be an ideal in a Riesz space E. In the quotient vector space E/J define*

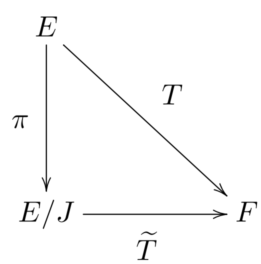
$$[x] \leq [y] \quad \text{if and only if} \quad \exists\, a \in [x] \text{ and } \exists\, b \in [y] \text{ such that } a \leq b\,.$$

This defines a partial ordering on E/J under which it becomes a Riesz space and the map $\pi\colon E \to E/J\colon x \mapsto [x]$ is a Riesz homomorphism.

The space E/J is the QUOTIENT RIESZ SPACE of E by J and π is the QUOTIENT MAP in the category **RS** of Riesz spaces and Riesz homomorphisms. The map π is also the QUOTIENT MAP in the category **RSP** of Riesz spaces and positive linear maps.

16.2.12. Example. The *quotient map* defined above is in fact a quotient map in the categorical sense (see definition 8.2.1) and, consequently, the quotient Riesz space is a quotient object in the category **RS**.

16.2.13. Theorem (Fundamental quotient theorem for **RS** and **RSP**)**.** *Let E and F be Riesz spaces and J be an order ideal in E. Show that if T is a Riesz homomorphism from E to F and $\ker T \supseteq J$, then there exists a unique Riesz homomorphism $\widetilde{T}\colon E/J \to F$ which makes the following diagram commute. Show also that if T is a positive linear map from E to F and $\ker T \supseteq J$, then there exists a unique positive linear map $\widetilde{T}\colon E/J \to F$ which makes the diagram commute.*



Furthermore, in either case, $\widetilde{T}$ is injective if and only if $\ker T = J$; and $\widetilde{T}$ is surjective if and only if T is.

16.2.14. Example. Let A be a nonempty subset of a Riesz space. The smallest order ideal containing A, also known as the *order ideal generated by A*, is the intersection of all order ideals containing A. It will be denoted by J_A. The order ideal generated by a single element a of a Riesz space is the PRINCIPAL ORDER IDEAL containing a. It will be denoted by J_a.

16.2.15. Definition. Let X be a topological space. A linear map $T\colon \mathcal{C}(X) \to \mathcal{C}(X)$ is a MULTIPLICATION OPERATOR if there exists a continuous function $\phi\colon X \to \mathbb{R}$ such that $(Tf)(x) = \phi(x)f(x)$ for every $f \in \mathcal{C}(X)$ and every $x \in X$.

16.2.16. Example. Let X be a topological space. If $T\colon \mathcal{C}(X) \to \mathcal{C}(X)$ is a linear transformation which maps every order ideal of $\mathcal{C}(X)$ into itself, then T is a multiplication operator.

Hint for proof. Consider the image under T of the constant function $\mathbf{1}$ on X.

16.3. Bands in Riesz spaces

16.3.1. Definition. An order ideal J in a Riesz space E is a BAND if $\ell \in J$ whenever $A \subseteq J$ and $\ell = \sup A$ exists in E.

16.3.2. Example. In the Riesz space $\mathcal{C}[a,b]$ the set $\{f : f(a) = 0\}$ is an order ideal but not a band.

16.3.3. Example. In the Riesz space $\mathcal{C}[a,b]$ the set of all functions f such that $f(x) = 0$ for every $x \in [0, \frac{1}{2}]$ is a band.

16.3.4. Example. If S is a measurable space, then the family $\mathrm{ca}(S)$ of real measures on S is a band in the Riesz space $\mathrm{ba}(S)$ of all bounded finitely additive set functions on S.

PROOF. See [1], theorem 9.55.

16.3.5. Example. Let A be a nonempty subset of a Riesz space. The smallest band containing A, also known as the *band generated by* A, is the intersection of all bands containing A. It will be denoted by B_A. The band generated by a single element a of a Riesz space is the PRINCIPAL BAND containing a. It will be denoted by B_a.

16.3.6. Proposition. *Let J be an order ideal in a Riesz E, then the band generated by J is the set of all $x \in E$ which satisfy*

$$|x| = \sup A \qquad \text{for some nonempty set } A \subseteq J^+ .$$

PROOF. See [49], Theorem 7.8. In [3], Theorem 3.4, the proposition is stated in terms of nets, which we discuss in the next section 16.4.

16.3.7. Example. If B_1 and B_2 are bands in a Riesz space, then $B_1 + B_2$ need not be a band. (Compare this with exercise 16.2.8.)

Hint for proof. Look in $\mathcal{C}([-1,1])$.

16.4. Nets

Nets (sometimes called *generalized sequences*) are useful for characterizing many properties of both topological spaces and Riesz spaces. They are used in fundamentally the same way as sequences are used in metric spaces.

16.4.1. Definition. A preordered set in which every pair of elements has an upper bound is a DIRECTED SET.

16.4.2. Example. If S is a set, then $\operatorname{Fin} S$ is a directed set under inclusion $\subseteq$.

16.4.3. Definition. Let S be a set and Λ be a directed set. A mapping $x \colon \Lambda \to S$ is a NET in S (or a *net of members of* S). The value of x at $\lambda \in \Lambda$ is usually written x_λ (rather than $x(\lambda)$) and the net x itself is often denoted by $\big(x_\lambda\big)_{\lambda\in\Lambda}$ or just $\big(x_\lambda\big)$.

16.4.4. Example. The most obvious examples of nets are sequences. These are functions whose domain is the (preordered) set $\mathbb{N}$ of natural numbers.

16.4.5. Definition. A net $x = \big(x_\lambda\big)$ is said to EVENTUALLY have some specified property if there exists $\lambda_0 \in \Lambda$ such that x_λ has that property whenever $\lambda \geq \lambda_0$; and it has the property FREQUENTLY if for every $\lambda_0 \in \Lambda$ there exists $\lambda \in \Lambda$ such that $\lambda \geq \lambda_0$ and x_λ has the property. (Compare these definitions with the ones given for sequences in 7.1.3 and 7.1.6.)

16.4.6. Definition. A net x in a topological space X CONVERGES to a point $a \in X$ if it is eventually in every neighborhood of a. In this case a is the LIMIT of x and we write $x_\lambda \xrightarrow[\lambda\in\Lambda]{} a$ or just $x_\lambda \to a$. When limits are unique we may also use the notation $\lim_{\lambda\in\Lambda} x_\lambda = a$ or, more simply, $\lim x_\lambda = a$.

The point a in X is a CLUSTER POINT of the net x if x is frequently in every neighborhood of a.

16.4.7. Example. The sequence $(1 - n^{-1})$ converges in $\mathbb{R}$.

16.4.8. Example. Let X be a topological space and $a \in X$. A BASE for the family of neighborhoods of the point a is a family $\mathfrak{B}_a$ of neighborhoods of a with the property that every neighborhood of a contains at least one member of $\mathfrak{B}_a$. In general, there are many choices for a neighborhood base at a point. Once such a base is chosen we refer to its members as *basic neighborhoods* of a.

Let $\mathfrak{B}_a$ be a base for the family of neighborhoods at a. Order $\mathfrak{B}_a$ by containment (the reverse of inclusion); that is, for U, $V \in \mathfrak{B}_a$ set

$$U \preceq V \text{ if and only if } U \supseteq V\,.$$

This makes $\mathfrak{B}_a$ into a directed set. Now (using the axiom of choice) choose one element x_U from each set $U \in \mathfrak{B}_a$. Then (x_U) is a net in X and $x_U \to a$.

The next two propositions assure us that in order for a net to converge to a point a in a topological space it is sufficient that it be eventually in every basic (or even subbasic) neighborhood of a.

16.4.9. Proposition. *Let $\mathfrak{B}$ be a base for the topology on a topological space X and a be a point of X. A net (x_λ) converges to a if and only if it is eventually in every neighborhood of a which belongs to $\mathfrak{B}$.*

16.4.10. Proposition. *Let $\mathfrak{S}$ be a subbase for the topology on a topological space X and a be a point of X. A net (x_λ) converges to a if and only if it is eventually in every neighborhood of a which belongs to $\mathfrak{S}$.*

16.4.11. Example. Let $J = [a, b]$ be a fixed interval in the real line, $\mathbf{x} = (x_0, x_1, \dots, x_n)$ be an $(n+1)$-tuple of points of J, and $\mathbf{t} = (t_1, \dots, t_n)$ be an n-tuple. The pair $(\mathbf{x}; \mathbf{t})$ is a PARTITION WITH SELECTION of the interval J if:

(a) $x_{k-1} < x_k$ for $1 \le k \le n$;
(b) $t_k \in [x_{k-1}, x_k]$ for $1 \le k \le n$;
(c) $x_0 = a$; and
(d) $x_n = b$.

The idea is that $\mathbf{x}$ *partitions* the interval into subintervals and t_k is the point *selected* from the k^{th} subinterval. If $\mathbf{x} = (x_0, x_1, \dots, x_n)$, then $\{\mathbf{x}\}$ denotes the *set* $\{x_0, x_1, \dots, x_n\}$. Let $P = (\mathbf{x}; \mathbf{s})$ and $Q = (\mathbf{y}; \mathbf{t})$ be partitions (with selections) of the interval J. We write $P \preccurlyeq Q$ and say that Q is a REFINEMENT of P if $\{\mathbf{x}\} \subseteq \{\mathbf{y}\}$. Under the relation $\preccurlyeq$ the family $\mathfrak{P}$ of partitions with selection on J is a directed (but *not* partially ordered) set.

Now suppose f is a bounded function on the interval J and $P = (\mathbf{x}; \mathbf{t})$ is a partition of J into n subintervals. Let $\Delta x_k := x_k - x_{k-1}$ for $1 \le k \le n$. Then define

$$S_f(P) := \sum_{k=1}^{n} f(t_k)\,\Delta x_k\,.$$

Each such sum $S_f(P)$ is a RIEMANN SUM of f on J. Notice that since $\mathfrak{P}$ is a directed set, S_f is a net of real numbers. If the net S_f of Riemann sums converges we say that the function f is RIEMANN INTEGRABLE. The limit of the net S_f (when it exists) is the RIEMANN INTEGRAL of f over the interval J and is denoted by $\int_a^b f(x)\,dx$.

16.4.12. Example. In a topological space with the indiscrete topology every net in the space converges to every point of the space.

As the preceding example shows nets in general topological spaces need not have unique limits. A space does not require very restrictive assumptions however to make this pathology go away. For limits to be unique it is sufficient to require that a space be Hausdorff. Interestingly, it turns out that this condition is also necessary.

16.4.13. Proposition. *If X is a topological space, then the following are equivalent:*

(a) *X is Hausdorff.*
(b) *No net in X can converge to more than one point.*
(c) *The diagonal in $X \times X$ is closed.*

(The DIAGONAL of a set S is $\{(s, s) : s \in S\} \subseteq S \times S$.)

16.4.14. Example. Recall from beginning calculus that the n^{th} PARTIAL SUM of an infinite series $\sum_{k=1}^{\infty} x_k$ is defined to be $s_n = \sum_{k=1}^{n} x_k$. The series is said to CONVERGE if the sequence (s_n) of partial sums converges and, if the series converges, its *sum* is the limit of the sequence (s_n) of partial sums. Thus, for example, the infinite series $\sum_{k=1}^{\infty} 2^{-k}$ converges and its sum is 1.

The idea of *summing* an infinite series depends heavily on the fact that the partial sums of the series are linearly ordered by inclusion. The "partial sums" of a net of numbers need not be partially ordered, so we need a new notion of *summation*.

16.4.15. Definition. Let $A \subseteq \mathbb{R}$. For every $F \in \operatorname{Fin} A$ define S_F to be the sum of the numbers in F, that is, $S_F = \sum F$. Then $S = \big(S_F\big)_{F \in \operatorname{Fin} A}$ is a net in $\mathbb{R}$. If this net converges, the set A of numbers is said to be SUMMABLE; the limit of the net is the SUM of A and is denoted by $\sum A$.

16.4.16. Definition. A net (x_λ) of real numbers is INCREASING if $\lambda \leq \mu$ implies $x_\lambda \leq x_\mu$. A net of real numbers is BOUNDED if its range is.

16.4.17. Example. Let $f(n) = 2^{-n}$ for every $n \in \mathbb{N}$, and let $x \colon \operatorname{Fin}(\mathbb{N}) \to \mathbb{R}$ be defined by $x(A) = \sum_{n \in A} f(n)$.

(a) The net x is increasing.
(b) The net x is bounded.
(c) The net x converges to 1.
(d) The set of numbers $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots\}$ is summable and its sum is 1.

16.4.18. Example. In a metric space every convergent sequence is bounded. Give an example to show that a convergent net in a metric space (even in $\mathbb{R}$) need not be bounded.

It is familiar from beginning calculus that bounded increasing sequences in $\mathbb{R}$ converge. The preceding example is a special case of a more general observation: bounded increasing nets in $\mathbb{R}$ converge.

16.4.19. Proposition. *Every bounded increasing net $\big(x_\lambda\big)_{\lambda \in \Lambda}$ of real numbers converges and*

$$\lim x_\lambda = \sup\{x_\lambda : \lambda \in \Lambda\}\,.$$

16.4.20. Example. Let (x_k) be a sequence of positive real numbers and $A = \{x_k : k \in \mathbb{N}\}$. Then $\sum A$ (as defined above) is equal to the sum of the series $\sum_{k=1}^{\infty} x_k$.

16.4.21. Example. The set $\big\{\frac{1}{k} : k \in \mathbb{N}\big\}$ is not summable and the infinite series $\sum_{k=1}^{\infty} \frac{1}{k}$ does not converge.

16.4.22. Example. The set $\big\{(-1)^k \frac{1}{k} : k \in \mathbb{N}\big\}$ is not summable but the infinite series $\sum_{k=1}^{\infty} (-1)^k \frac{1}{k}$ does converge.

16.4.23. Proposition. *Every summable subset of real numbers is countable.*

Hint for proof. If A is a summable set of real numbers and $n \in \mathbb{N}$, how many members of A can be larger than $1/n$?

In proposition 16.4.13 we have already seen one example of a way in which nets characterize a topological property: a space is Hausdorff if and only if limits of nets in the space are unique (when they exist). Here is another topological property—continuity at a point—that can be conveniently characterized in terms of nets.

16.4.24. Proposition. *Let X and Y be topological spaces. A function $f: X \to Y$ is continuous at a point a in X if and only if $f(x_\lambda) \to f(a)$ in Y whenever (x_λ) is a net converging to a in X.* (*Cf.* exercise 11.2.6.)

16.4.25. Example. Let Y be a topological space with the weak topology (see definition 11.4.1) determined by an indexed family $(f_\alpha)_{\alpha\in A}$ of functions where for each $\alpha \in A$ the codomain of f_α is a topological space X_α. Then a net $(y_\lambda)_{\lambda\in\Lambda}$ in Y converges to a point $a \in Y$ if and only if $f_\alpha(y_\lambda) \xrightarrow[\lambda\in\Lambda]{} f_\alpha(a)$ for every $\alpha \in A$.

16.4.26. Example. If $Y = \prod_{\alpha\in A} X_\alpha$ is a product of nonempty topological spaces with the product topology (see definition 11.4.9), then a net (y_λ) in Y converges to $a \in Y$ if and only if $(y_\lambda)_\alpha \to a_\alpha$ for every $\alpha \in A$.

16.4.27. Example. Let X and Y be nonempty topological spaces. Recall that the set $Y^X = \mathcal{F}(X, Y)$ of all functions from X into Y is a product (see example 8.3.7). Give it the usual product topology (see definition 11.4.9). A net (f_λ) of functions in Y^X converges to a function $g \in Y^X$ if and only if $f_\lambda(x) \to g(x)$ for each $x \in X$. Thus the product topology on Y^X is usually called the TOPOLOGY OF POINTWISE CONVERGENCE.

In metric spaces we characterized interiors and closures in terms of sequences. For general topological spaces we must replace sequences by nets.

16.4.28. Proposition. *A point a in a topological space X is in the interior of a subset S of X if and only if every net in X that converges to a is eventually in S.* Hint. *One direction is easy. In the other use exercise 16.4.9 and the sort of net constructed in example 16.4.8. (*Cf.* proposition 10.5.9.)*

16.4.29. Proposition. *A point b in a topological space X belongs to the closure of a subset S of X if and only if some net in S converges to b. (*Cf.* exercise 10.5.22.)*

16.4.30. Proposition. *A subset A of a topological space is closed if and only if the limit of every convergent net in A belongs to A.*

Nets (as well as sequences) are also useful in characterizing many properties of Riesz spaces.

16.4.31. Definition. A net $(x_\lambda)_{\lambda\in\Lambda}$ in a Riesz space E is INCREASING if $\lambda \le \mu$ in Λ implies $x_\lambda \le x_\mu$ in E. This is denoted by $(x_\lambda)\uparrow$. We write $x_\lambda \uparrow a$ to indicate that the net (x_λ) is increasing and $a = \sup\{x_\lambda : \lambda \in \Lambda\}$. Decreasing nets and the notation $x_\lambda \downarrow a$ are defined similarly.

16.4.32. Proposition. *Let (x_λ) and (y_λ) be nets (whose domains are the same directed set Λ) in a Riesz space. Show that if $x_\lambda \downarrow a$ and $y_\lambda \downarrow b$, then $x_\lambda + y_\lambda \downarrow a + b$.*

16.4.33. Proposition. *Let (x_λ) be a net in a Riesz space such that $x_\lambda \downarrow a$. If $\alpha > 0$, then $\alpha x_\lambda \downarrow \alpha a$ (and if $\alpha < 0$, then $\alpha x_\lambda \uparrow \alpha a$).*

16.4.34. Proposition. *Let (x_λ) and (y_λ) be nets (whose domains are the same directed set Λ) in a Riesz space. Show that if $x_\lambda \downarrow a$ and $y_\lambda \downarrow b$, then $x_\lambda \vee y_\lambda \downarrow a \vee b$ and $x_\lambda \wedge y_\lambda \downarrow a \wedge b$.*

16.4.35. Exercise. State and prove the facts about sequences which correspond to the facts about nets in the three preceding propositions. Also state and prove the dual of each of these propositions.

16.4.36. Proposition. *A Riesz space E is σ-Dedekind complete if and only if every increasing sequence (x_n) in E^+ which is bounded above has a least upper bound.*

16.4.37. Proposition. *A Riesz space E is Dedekind complete if and only if every increasing net (x_λ) in E^+ which is bounded above has a least upper bound.*

16.4.38. Example. Let $\mathfrak{A}$ be an algebra of subsets of a nonempty set S. Then the Riesz space $\mathrm{ba}(S)$ (see example 14.1.17) is Dedekind complete.

PROOF. See [31], example 23.3(v) or [49], example 12.5(iv).

16.4.39. Example. If $(S, \mathfrak{A})$ is a nonempty measurable space, then $\mathrm{ca}(S)$ is a band in the Riesz space $\mathrm{ba}(S)$.

PROOF. See [31], example 25.3 or [49], theorem 27.3.

16.4.40. Proposition. *An order ideal in a Dedekind complete Riesz space is itself Dedekind complete.*

16.4.41. Example. If $(S, \mathfrak{A})$ is a nonempty measurable space, then the Riesz space $\mathrm{ca}(S)$ is Dedekind complete.

16.4.42. Proposition. *Let v be an element in a Riesz space E. The principal order ideal J_v generated by v is given by*

$$J_v = \{x \in E \colon |x| \le \alpha|v| \text{for some } \alpha > \mathbf{0}\}.$$

16.4.43. Proposition. *Let v be an element in a Riesz space E. The principal band B_v generated by v is given by*

$$B_v = \{x \in E \colon |x| \wedge n|v| \uparrow |x|\}$$

(where $n \in \mathbb{N}$).

16.5. Disjointness in Riesz Spaces

16.5.1. Definition. In a Riesz space two elements x and y are DISJOINT (or ORTHOGONAL, or MUTUALLY SINGULAR) if $|x| \wedge |y| = 0$. In this case we write $x \perp y$.

16.5.2. Proposition. *Let x, y, and z be elements of a Riesz space.*

(a) *If $y \perp z$ and $|x| \le |y|$, then $x \perp z$.*
(b) *If $x \perp y$ and $\alpha \in \mathbb{R}$, then $\alpha x \perp y$.*

16.5.3. Proposition. *Let E be a Riesz space.*

(a) *If x, y, $z \in E$, then*

$$|(x \vee z) - (y \vee z)| + |(x \wedge z) - (y \wedge z)| = |x - y|\,.$$

(b) *If u, v, $w \in E^+$, then*

$$(u + v) \wedge w \le (u \wedge w) + (v \wedge w)\,.$$

(c) *If u, v, $w \in E^+$ and $u \perp v$, then*

$$(u + v) \wedge w = (u \wedge w) + (v \wedge w)\,.$$

(d) *If $x \perp z$ and $y \perp z$ in E, then $x + y \perp z$.*

16.5.4. Proposition. *Let E be a Riesz space.*

(a) *$x \perp y$ in E if and only if $x^+ \perp y$ and $x^- \perp y$.*
(b) *Let $D \subseteq E$ and suppose that $\ell = \bigvee D$ exists (in E). Then*

$$\ell^+ = \bigvee\{x^+ \colon x \in D\} \quad \text{and} \quad \ell^- = \bigwedge\{x^- \colon x \in D\}\,.$$

(c) *Let $D \subseteq E$ and suppose that $\ell = \bigvee D$ exists (in E). If $y \perp x$ for all $x \in D$, then $y \perp \ell$.*

16.5.5. Exercise. Let E be a Riesz space.

(a) If $x \perp y$ in E, then

$$|\,|x| - |y|\,| = |x + y| = |x - y| = |x| + |y| = |x| \vee |y|\,.$$

(b) If $|x + y| = |x - y|$ in E, then $x \perp y$.
(c) If $|x + y| = |x| \vee |y|$ in E, then $x \perp y$.
(d) If $|x - y| = |x| \vee |y|$ in E, then $x \perp y$.

16.5.6. Definition. Let A be a nonempty subset of a Riesz space E. The DISJOINT COMPLEMENT of A, denoted by A^d, is the set of all elements of E that are disjoint from every member of A. That is,

$$A^d := \{x \in E \colon x \perp a \text{ for every } a \in A\}\,.$$

16.5.7. Proposition. *Let $A \subseteq B \subseteq E$ where E is a Riesz space. Then*

(a) $A \cap A^d \subseteq \{\mathbf{0}\}$,
(b) $B^d \subseteq A^d$,
(c) $A \subseteq A^{dd}$, *and*
(d) $A^d = A^{ddd}$.

16.5.8. Example. If A is a subset of a Riesz space E, then A^d is a band in E.

16.5.9. Proposition. *If J_1 and J_2 are order ideals in a Riesz space E such that $E = J_1 \oplus J_2$, then $J_2 = {J_1}^d$ and $J_1 = {J_2}^d$. Thus J_1 and J_2 are bands.*

16.5.10. Definition. Two real measures μ and ν on a measurable space $(S, \mathfrak{A})$ are MUTUALLY SINGULAR if they are disjoint in the Riesz space $\mathrm{ca}(S)$; that is, if $|\mu| \wedge |\nu| = 0$. We extend this terminology to complex and positive measures: two such measures are mutually singular if $|\mu| \wedge |\nu| = 0$. If μ and ν are mutually singular measures we write $\mu \perp \nu$. It is possible to give a very simple necessary and sufficient condition for the mutual singularity of measures.

16.5.11. Proposition. *If μ and ν are real measures on a measurable space $(S, \mathfrak{A})$, then $\mu \perp \nu$ if and only if there exists a measurable subset D of S such that*

$$|\mu|(D) = |\nu|(D^c) = 0.$$

Hint for proof. It is enough to consider positive real measures. Suppose $\mu \wedge \nu = 0$. Then by example 16.1.10 it is possible to choose a measurable set A so that $\mu(A) + \nu(A^c)$ is arbitrarily small. Use this fact to show that for each $\epsilon > 0$ there exists $B \overset{m}{\subseteq} S$ such that $\mu(B) = 0$ and $\nu(B^c) < \epsilon$. Use this in turn to find a set D such that $\mu(D) = 0$ and $\nu(D^c) = 0$.

16.5.12. Definition. Let μ be a real measure on a measure space S and $P \overset{m}{\subseteq} S$. We say that P is a POSITIVE SET for μ if $\mu(A) \geq 0$ for every measurable subset A of P. Similarly, if $\mu(A) \leq 0$ for every $A \overset{m}{\subseteq} P$, then P is a NEGATIVE SET for μ.

If P is a positive set for the real measure μ and if P^c is a negative set for μ, then the pair (P, P^c) is a HAHN DECOMPOSITION of S. If a Hahn decomposition of a real measure space (S, μ) is known, then it becomes a very simple matter to compute the Jordan decomposition of μ. To determine the action of μ^+ on a measurable set A, intersect A with P and apply μ; to find $\mu^-(A)$, apply $-\mu$ to $A \cap P^c$.

16.5.13. Proposition. *If (P, P^c) is a Hahn decomposition of a real measure space (S, μ), then for every $A \overset{m}{\subseteq} S$*

$$\mu^+(A) = \mu(A \cap P)$$

and

$$\mu^-(A) = -\mu(A \cap P^c)\,.$$

Notice the following very simple corollary: If (P, P^c) is a Hahn decomposition of a real measure space, then $\mu^-(B) = 0$ whenever $B \overset{m}{\subseteq} P$ and $\mu^+(B) = 0$ whenever $B \overset{m}{\subseteq} P^c$.

16.5.14. Example. Let $(S, \mathfrak{A}, \mu)$ be a positive real measure space such that $\mu(S) = 1$ and let $\nu = \mu - \delta_b$ where δ_b is Dirac measure on $(S, \mathfrak{A})$ concentrated at some point $b \in S$.

(a) If $P = S \setminus \{b\}$, then the pair (P, P^c) is a Hahn decomposition of S induced by ν.
(b) The Jordan decomposition $\nu = \nu^+ - \nu^-$ is given by $\nu^+(A) = \mu(A \setminus \{b\}\,)$ and $\nu^-(A) = 1 - \mu(\{b\})$ if $b \in A$ and $\nu^-(A) = 0$ otherwise.

(c) Suppose further that c is a member of S different from b such that $\mu(\{c\}) = 0$. Then (Q, Q^c) is a second Hahn decomposition of S when $P^c = \{b, c\}$.

As the previous example indicates, Hahn decompositions are not, in general, unique. However, they are "almost" unique.

16.5.15. Proposition. *If (P, P^c) and (Q, Q^c) are Hahn decompositions of a real measure space (S, μ), then*

$$|\mu|(P \bigtriangleup Q) = 0\,.$$

In light of the convenience provided by a Hahn decomposition in computing μ^+ and μ^-, it is gratifying to know that such a decomposition always exists.

16.5.16. Proposition. *Every real measure space has a Hahn decomposition.*

16.5.17. Definition. If μ is a real measure on a measurable space S, then the corresponding positive real measure $|\mu| = \mu^+ + \mu^-$ is called the TOTAL VARIATION of μ. As is true of μ^+ and μ^-, it is possible to give a simple characterization of the total variation measure.

16.5.18. Proposition. *If μ is a real measure on a measurable space S, then for each measurable subset A of S*

$$|\mu|(A) = \sup\left\{\sum_{k=1}^{n} |\mu(B_k)| : A = \biguplus_{k=1}^{n} B_k\right\}.$$

Hint for proof. The inequality in one direction can be obtained by showing that:

$$|\mu|(A) = |\mu(A \cap P)| + |\mu(A \cap P^c)|$$

where $A \overset{m}{\subseteq} S$ and (P, P^c) is a Hahn decomposition of S.

16.5.19. Example. Let $\mathfrak{A}$ be an algebra of subsets of a set S. For $\mu \in \operatorname{ba}(S, \mathfrak{A}))$ define

$$\|\mu\| = |\mu|(S) \ (= \sup\{|\mu|(A) \colon A \in \mathfrak{A}\}).$$

This makes $\operatorname{ba}(S)$ into a real normed linear space. If $\mathfrak{A}$ is a σ-algebra of subsets of S, then $\operatorname{ca}(S, \mathfrak{A})$ is also a real normed linear space. Notice that this is *not* the same norm given in example 12.1.18.

16.6. Absolute Continuity

16.6.1. Definition. If J is any interval in $\mathbb{R}$ a real (or complex) valued function f on J is ABSOLUTELY CONTINUOUS if it satisfies the following condition:

> for every $\epsilon > 0$ there exists $\delta > 0$ such that if $(a_1, b_1), \dots, (a_n, b_n)$ are disjoint nonempty subintervals of J with $\sum_{k=1}^{n}(b_k - a_k) < \delta$, then $\sum_{k=1}^{n}|f(b_k) - f(a_k)| < \epsilon$.

16.6.2. Example. The function $x \mapsto \sqrt{x}$ is absolutely continuous on $[0, \infty)$.

It is obvious that an absolutely continuous function on an interval in $\mathbb{R}$ is uniformly continuous. But the converse is not true.

16.6.3. Example. The function $f \colon [0, 1] \to \mathbb{R}$ defined by $f(x) = x \sin\left(\frac{1}{x}\right)$ for $0 < x \leq 1$ and $f(0) = 0$ is uniformly continuous but not absolutely continuous.

Hint for proof. The easiest way to see that f is uniformly continuous is to make use of two results we have not yet proved, 17.1.29 and 17.1.21.

16.6.4. Definition. If μ and ν are real measures on a measurable space, we say that μ is ABSOLUTELY CONTINUOUS WITH RESPECT TO ν, and write $\mu \ll \nu$, if $|\nu|(A) = 0$ implies $\mu(A) = 0$.

16.6.5. Proposition. *If μ and ν are real measures on a measurable space S, then the following are equivalent:*

(a) $\mu \ll \nu$;
(b) $\mu^+ \ll |\nu|$ *and* $\mu^- \ll |\nu|$; *and*
(c) $|\mu| \ll |\nu|$.

16.6.6. Proposition. *Let* μ, ν, *and* ρ *be real measures on a measurable space* S.

(a) $\mu \ll \mu$.
(b) *If* $\rho \ll \nu$ *and* $\nu \ll \mu$, *then* $\rho \ll \mu$.
(c) *If* $0 \leq \nu \leq \mu$, *then* $\nu \ll \mu$.
(d) *If* $\mu \ll 0$, *then* $\mu = 0$.
(e) *If* $\rho \ll \mu$ *and* $\nu \ll \mu$, *then* $\rho + \nu \ll \mu$.
(e) *If* $\mu, \nu \geq 0$, *then* $\nu \ll \mu + \nu$.
(f) *If* $\nu \ll \mu$ *and* $\mu \perp \rho$, *then* $\nu \perp \rho$.

The next result may help explain the connection between absolute continuity of functions and absolute continuity of measures.

16.6.7. Proposition. *Suppose that* ν *is a measure and* μ *is a complex measure on a measurable space* S. *Then the following are equivalent:*

(a) $\mu \ll \nu$; *and*
(b) *For every* $\epsilon > 0$ *there exists* $\delta > 0$ *such that* $|\mu(A)| < \epsilon$ *whenever* $A \overset{m}{\subseteq} S$ *and* $\nu(A) < \delta$.

16.6.8. Proposition. *Let* μ *and* ν *be positive real measures on a measurable space* S. *Then the following are equivalent:*

(a) $\mu \ll \nu$;
(b) *if* (A_n) *is a sequence of measurable subsets of* S *such that* $\nu(A_n) \to 0$ *as* $n \to \infty$, *then* $\mu(A_n) \to 0$ *as* $n \to \infty$; *and*
(c) *for every* $\epsilon > 0$ *there exists* $\delta > 0$ *such that* $\mu(A) < \epsilon$ *whenever* $A \overset{m}{\subseteq} S$ *and* $\nu(A) < \delta$.

CHAPTER 17

COMPACT SPACES

17.1. Compactness

17.1.1. Definition. Let $\mathfrak{U}$ be a family of sets and S be a set. Recall that if $\bigcup\mathfrak{U} \supseteq S$, then we say that $\mathfrak{U}$ *covers* S, or that $\mathfrak{U}$ *is a cover for* S, or that $\mathfrak{U}$ *is a covering of* S. If a cover $\mathfrak{U}$ for a topological space X consists entirely of open subsets of M, then $\mathfrak{U}$ is an OPEN COVER for X. If $\mathfrak{U}$ is a family of sets which covers a set S and $\mathfrak{V}$ is a subfamily of $\mathfrak{U}$ which also covers S, then $\mathfrak{V}$ is a SUBCOVER of $\mathfrak{U}$ for S. A topological space X is COMPACT if *every* open cover of X has a finite subcover.

We have just defined what we mean by a compact *space*. It will be convenient also to speak of a compact *subset* of a topological space X. If $K \subseteq X$, we say that K is a COMPACT SUBSET of X if, regarded as a subspace of X, it is a compact topological space.

Suppose we wish to show that some particular subset K of a topological space X is compact. Is it necessary that we work with coverings made up of open subsets of K (as the definition demands) or can we just as well use coverings whose members are open subsets of X? The answer is that, happily, either will do. In other words compactness is an intrinsic property of a set. (Recall the cautionary note at the beginning of section 11.5.1.)

17.1.2. Proposition. *A subset K of a topological space X is compact if and only if every cover of K by open subsets of X has a finite subcover.*

The class of compact Hausdorff spaces together with the continuous maps between them constitute an important category. We denote this category by **CpH**.

Compactness in metric spaces can be characterized in terms sequences. A metric space is said to be SEQUENTIALLY COMPACT if every sequence in the space has a convergent subsequence.

17.1.3. Proposition. *A metric space is compact if and only if it is sequentially compact.*

PROOF. See [17], theorem 16.2.1.

17.1.4. Proposition. *In a metric space every compact set is bounded.*

17.1.5. Example. Every finite subset of a topological space is compact.

17.1.6. Example. Let x be a convergent sequence in a topological space X and $a = \lim x_n$. Then $\operatorname{ran} x \cup \{a\}$ is a compact subset of X.

An important fact about compactness is that it is F-hereditary.

17.1.7. Proposition. *Every closed subset of a compact topological space is compact.*

17.1.8. Proposition. *In a Hausdorff topological space every compact set is closed.*

17.1.9. Example. In the preceding proposition the hypothesis that the space be Hausdorff is necessary.

17.1.10. Definition. A family $\mathcal{F}$ of sets is said to have the FINITE INTERSECTION PROPERTY if every finite subfamily of $\mathcal{F}$ has nonempty intersection.

17.1.11. Proposition. *A topological space X is compact if and only if every family of closed subsets of X with the finite intersection property has nonempty intersection.*

17.1.12. Proposition. *Let (F_k) be a decreasing sequence of nonempty closed subsets of a compact topological space. Then $\bigcap_{k=1}^{\infty} F_k$ is nonempty.*

The next result is fundamental: the continuous image of a compact set is compact.

17.1.13. Proposition. *If $f\colon X \to Y$ is a continuous function between topological spaces and X is compact, then $\operatorname{ran} f$ is compact.*

17.1.14. Proposition. *In the category* **CpH** *every epimorphism is a quotient map.*

Compare the preceding proposition with example 11.6.6.

17.1.15. Proposition. *Every continuous bijection from a compact space to a Hausdorff space is a homeomorphism.*

Recall that we have seen in example 11.1.10 that in the category **TOP** bijective morphisms need not be homeomorphisms. However ...

17.1.16. Corollary. *In the category* **CpH** *bijective morphisms are isomorphisms.*

The preceding corollary and proposition 17.1.14 help to explain analysts' affection for compact Hausdorff spaces.

17.1.17. Example. let $X = [0, 2\pi]$ and let $\sim$ be the equivalence relation which identifies 0 and 2π. (That is, the only distinct elements which are equivalent are 0 and 2π.) Show that the resulting quotient space $X/\sim$ (see exercise 11.6.4) is (homeomorphic to) the unit circle $\mathbb{T}$.

17.1.18. Theorem (Extreme value theorem)**.** *Every continuous real valued function on a compact topological space assumes a maximum and a minimum value somewhere on the space.*

17.1.19. Theorem (Dini's Theorem)**.** *If a decreasing sequence (f_n) of real valued functions on an interval $[a, b]$ in $\mathbb{R}$ converges pointwise to zero, then it converges uniformly to zero.*

Hint for proof. Given $\epsilon > 0$ let $A_n := {f_n}^{-1}([\epsilon, \infty))$. Argue that since the family $\{A_n : n \in \mathbb{N}\}$ has empty intersection, it cannot have the finite intersection property.

17.1.20. Exercise. Generalize *Dini's theorem* to real valued functions on a compact Hausdorff space X. Also see if you can weaken the hypotheses: instead of assuming that a decreasing sequence (f_n) of continuous functions converges pointwise to zero, assume instead that $\mathcal{F}$ is a family of continuous functions such that

(i) the infimum of any two members of $\mathcal{F}$ dominates some third member of $\mathcal{F}$ and
(ii) the function $g\colon x \mapsto \inf\{f(x) : f \in \mathcal{F}\}$ (where $x \in X$) is real valued and continuous.

17.1.21. Proposition. *Every continuous function between metric spaces whose domain is compact is uniformly continuous.*

17.1.22. Exercise. Show that the interval $[0, \infty)$ is not compact using each of the following:

(a) the definition of compactness (see 17.1.1);
(b) the *extreme value theorem* (see 17.1.18);
(c) proposition 17.1.21;
(d) the *finite intersection property* (see 17.1.11);
(e) *Dini's theorem* (see!17.1.19).

17.1.23. Exercise. Using only the definition of compactness prove that the closed unit interval $[0, 1]$ is compact.

17.1.24. Exercise. Let A, $B \subseteq M$ where (M, d) is a metric space. Prove or disprove:

(a) If A and B are disjoint closed sets, then $d(A, B) > 0$.
(b) If A and B are disjoint sets, if A is compact, and if B is closed, then $d(A, B) > 0$.
(c) If A and B are disjoint compact sets, then $d(A, B) > 0$.

(d) If A and B are closed sets, then there exist points $a \in A$ and $b \in B$ such that $d(A, B) = d(a, b)$.
(e) If A is compact and B is closed, then there exist points $a \in A$ and $b \in B$ such that $d(A, B) = d(a, b)$.
(f) If A and B are compact sets, then there exist points $a \in A$ and $b \in B$ such that $d(A, B) = d(a, b)$.

(Recall that in a metric space the distance $d(A, B)$ between two sets A and B is defined to be $\inf\{d(a, b) \colon a \in A \text{ and } b \in B\}$.)

17.1.25. Proposition. *Let $\mathfrak{T}_1$ and $\mathfrak{T}_2$ be topologies on a set X with $\mathfrak{T}_1$ weaker than $\mathfrak{T}_2$. If the space $(X, \mathfrak{T}_1)$ is Hausdorff and the space $(X, \mathfrak{T}_2)$ is compact, then the two topologies are equal.*

17.1.26. Exercise. In each of the following sentences, fill in the blank with either the word "weakened" or the word "strengthened" to make a correct assertion.

(a) A Hausdorff space remains Hausdorff if its topology is ___________ .
(b) A continuous function $f \colon X \to Y$ remains continuous if the topology on X is ___________ .
(c) A continuous function $f \colon X \to Y$ remains continuous if the topology on Y is ___________ .
(d) A convergent sequence in a topological space remains convergent if the topology on the space is ___________ .
(e) A compact space remains compact if its topology is ___________ .

17.1.27. Theorem (Tychonoff)**.** *Let $\{X_\alpha \colon \alpha \in A\}$ be a family of topological spaces. Their product $\prod X_\alpha$ is compact if and only if each factor space X_α is compact.*

Tychonoff proved this theorem in 1930 for the special case where each X_α is the unit interval; Čech published the proof of the general case in 1937.

Proofs of *Tychonoff's theorem* highlight the important distinction between "elementary proofs" and "easy proofs". The short and easy proofs of this theorem all make use of somewhat advanced concepts: Aliprantis and Border [1] (theorem 2.57), Dugundji [16] (theorem XI.1.4(4)), and Wilansky [47] (theorem7.4.1) employ the machinery of *ultrafilters*, while Willard [48] (theorem 17.8) uses *universal nets*. An elementary (but longer) proof appears in Lang [28] (chapter 2, theorem 3.12). Perhaps the most approachable compromise is the route taken by Rudin [38] (theorem A.3) and Hewitt and Stromberg [23] (theorem 6.43) where the principal tool is the *Alexander subbase theorem*, which says that a topological space is compact if and only if every cover by subbasic open sets has a finite subcover. Whatever path you choose, somewhere you will need some version of the *axiom of choice*. (Kelley proved in 1950 that *Tychonoff's theorem* implies the *axiom of choice*.)

17.1.28. Proposition. *If K and L are compact subsets of a normed linear space, then $K + L$ is compact.*

17.1.29. Theorem (Heine-Borel theorem)**.** *A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

17.1.30. Proposition. *Any two norms on $\mathbb{R}^n$ are equivalent.*

17.1.31. Example (Cantor set)**.** Define a subset C of the unit interval $[0, 1]$, called *the Cantor set*, as follows. Let F_1 be $[0, 1]$ with its open middle third $\left(\frac{1}{3}, \frac{2}{3}\right)$ removed. So F_1 consists of two closed intervals. Let F_2 be F_1 with the open middle thirds of both intervals removed. Proceed inductively. Having defined the set F_n, which consists of 2^n closed intervals, let F_{n+1} be F_n with the open middle third of each of its constituent intervals removed. Finally, let $C = \cap_{n=1}^{\infty} F_n$. Then

(a) $C = \left\{ \sum_{k=1}^{\infty} a_k 3^{-k} \colon a_k \text{ is 0 or 2} \right\}$;
(b) C consists of all points of $[0, 1]$ which have at least one ternary expansion not containing the digit 1:
(c) C is a compact metric space;

(d) C is uncountable; and
(e) C is homeomorphic to the product space $\{0,1\}^{\mathbb{N}}$ (where $\{0,1\}$ is a two-point space with the discrete topology).

Hint for proof. Suppose that x is a number in the interval $[0,1]$ and $x = \sum_{k=1}^{\infty} a_k 3^{-k}$, where each a_k is 0, 1, or 2. Then we may adopt a shorthand notation $x = 0._3 a_1 a_2 a_3 a_4 \dots$. This is the TERNARY EXPANSION of x. As is the case with decimal expansions, ternary expansions need not be unique. For example, $\frac{1}{3} = 0._3 1000 \dots = 0._3 0222 \dots$. For part (d) consider the function from the Cantor set C into $[0,1]$ whose value at a point $\sum_{k=1}^{\infty} a_k 3^{-k}$ (each a_k being 0 or 2) is $\sum_{k=1}^{\infty} a_k 2^{-(k+1)}$. We will call this function the CANTOR FUNCTION. This function is surjective; but be careful—it is not quite one-to-one; its values at the endpoints of each interval removed in the construction of the Cantor set are equal. (You can find careful discussions of the elementary properties of the Cantor set in pages 26–29 of [10] and in section 2.5 of [32].)

17.1.32. CAUTION. Many authors assign the name *Cantor function* to the extension of the function defined above to the entire interval $[0,1]$. (This is accomplished by assigning the function a constant value on each of the open intervals removed in the construction of the Cantor set in such a way that the resulting function is increasing. See example 22.5.1 for details.) We will call this extension the *Lebesgue singular function*.

17.1.33. Definition. An ISOLATED POINT of a subset A of a topological space X is a point belonging to A which is not an accumulation point of A. A closed subset of X is said to be PERFECT if it contains no isolated points; that is, if $A = A'$.

17.1.34. Example. At each stage in the construction of the Cantor set we removed the middle third of each remaining closed interval. The fraction one-third is a bit arbitrary. We may if we wish pursue essentially the same procedure but (at each stage) remove an open interval of some other length from the center of each remaining closed interval. This will result in sets known as CANTOR-LIKE SETS. A detailed description of this process, together with a proof that such a set is compact, perfect and has empty interior, can be found in [23], pages 70–71.

In this context two more very interesting results are proved in [23]. The first generalizes the fact (see example 17.1.31(d)) that the Cantor set is uncountable.

17.1.35. Proposition. *If A is a nonempty perfect subset of a complete metric space, then* $\operatorname{card} A \geq \operatorname{card} \mathbb{R}$.

PROOF. See [23], theorem 6.65.

The second result says that closed subsets of second countable spaces are "almost" perfect.

17.1.36. Proposition (Cantor-Bendixson)**.** *If A is a closed subset of a second countable topological space X, then there exists a perfect subset P of X contained in A such that $A \setminus P$ is countable.*

PROOF. See [23], theorem 6.66.

17.2. Local Compactness

17.2.1. Definition. A topological space is LOCALLY COMPACT if every point in the space has a neighborhood whose closure is compact.

17.2.2. Example. Every compact space is locally compact.

17.2.3. Example. For each $n \in \mathbb{N}$ Euclidean n-space $\mathbb{R}^n$ is locally compact but not compact.

17.2.4. Example. The space $\mathbb{Q}$ of rational numbers is not locally compact.

17.2.5. Notation. Let A and B be subsets of a topological space. We write $A \prec B$ if $\overline{A} \subseteq B^{\circ}$.

17.2.6. Proposition. *The following hold for subsets A, B, and C of a topological space:*

(a) *if* $A \prec B$ *and* $B \prec C$*, then* $A \prec C$*;*
(b) *if* $A \prec C$ *and* $B \prec C$*, then* $A \cup B \prec C$*; and*
(c) *if* $A \prec B$ *and* $A \prec C$*, then* $A \prec B \cap C$*.*

Starting in section 21.4 we study so-called *normal* topological spaces. The class of all such spaces is important because it contains all metric spaces and all compact Hausdorff spaces. One very important result for these spaces is *Urysohn's lemma* concerning the separation of a pair of disjoint closed sets by a continuous function (see theorem 21.5.9). Unfortunately not every locally compact Hausdorff space is normal. The next three results 17.2.7, 17.2.8, and 17.2.10 give us a version of *Urysohn's lemma* for locally compact Hausdorff spaces.

It is not easy to dream up an example of a locally compact Hausdorff space which is not normal. There is, however, a standard example, the *Tychonoff plank*. You can find details in [16], page 145, example 4 and page 239, example 2 or [47], section 14.6.

17.2.7. Proposition. *Let* X *be a locally compact Hausdorff space,* $x \in X$*, and* V *be a neighborhood of* x*. Then there exists a neighborhood* U *of* x *with compact closure such that* $U \prec V$*.*

Proof. See [23], theorem 6.78 or [11], proposition 7.1.2. There is also a proof in [32], proposition 7.22(a), but it makes use of facts about normal topological spaces.

17.2.8. Proposition. *Let* X *be a locally compact Hausdorff space,* K *a compact subset of* X*, and* V *be an open subset of* X *such that* $K \subseteq V$*.Then there exists an open set* U *in* X *with compact closure such that* $K \prec U \prec V$*.*

Proof. See [23], theorem 6.79; [11], proposition 7.1.3; or [32], proposition 7.22(b).

17.2.9. Definition. We will say that disjoint sets A and B in topological space X can be functionally separated if there exists a continuous function $f\colon X \to [0,1]$ such that $f^{\to}(A) = \{0\}$ and $f^{\to}(B) = \{1\}$.

17.2.10. Theorem (Urysohn's lemma for locally compact Hausdorff spaces)**.** *Let* X *be a locally compact Hausdorff space,* K *be a compact subset of* X*, and* C *be a closed subset of* X*. If* K *and* C *are disjoint, then they can be functionally separated.*

Proof. See [23], theorem 6.80. Again there is a proof in [32], theorem 7.14, but it makes use of facts about normal spaces.

17.2.11. Definition. We say that a family $\mathcal{F}$ of continuous functions on a topological space X separates points of X (or is a separating family of functions on X) if for every pair of distinct points x, $y \in X$ there exists $f \in \mathcal{F}$ such that $f(x) \neq f(y)$. (Notice that there is a slight (and unimportant) difference between separating points x and y and separating the sets $\{x\}$ and $\{y\}$.)

17.2.12. Proposition. *Let* X *be a compact topological space. Then* $\mathcal{C}(X)$ *is separating if and only if* X *is Hausdorff.*

17.2.13. Definition. A topological space $(X, \mathfrak{T})$ is metrizable if its topology $\mathfrak{T}$ is the one induced by some metric on the set X. (See remark 10.4.14.)

17.2.14. Proposition. *Every second countable locally compact Hausdorff space is metrizable.*

Proof. See [32], theorem 7.16.

17.3. Compactifications

17.3.1. Definition. Let X and Y be topological spaces. A homeomorphism $f\colon X \to Y$ from X onto a subspace of Y is a (topological) embedding and we say that X is embedded in Y. A compactification of a Hausdorff topological space X is an ordered pair (K, ϕ) where K is a compact Hausdorff space and ϕ is an embedding of X into K.

17.3.2. CAUTION. This definition is not entirely standard. Many authors require that for (K, ϕ) to be a compactification of X the range of ϕ must be *dense* in K.

17.3.3. Proposition. *Every locally compact Hausdorff space X has a compactification $\widetilde{X}$ such that $\widetilde{X} \setminus X$ consists of a single point.*

PROOF. Let $\widetilde{X} := X \cup \{\infty\}$ where ∞ is any point which does not belong to X. Declare the open subsets of $\widetilde{X}$ to be the open subsets of X together with the complements in $\widetilde{X}$ of compact subsets of X. It is easy to show that this family of subsets does indeed comprise a topology on $\widetilde{X}$ and that with this topology $\widetilde{X}$ is a compact Hausdorff space. □

The space $\widetilde{X}$ defined above is the ONE-POINT COMPACTIFICATION (or the ALEXANDROFF COMPACTIFICATION) of X. The single point ∞ which we have adjoined to X is often called, somewhat fancifully, the POINT AT INFINITY. There is nothing particularly mysterious or infinite about this point. It is, as we have said, just any point not in X. Can we be sure that there is such a point? Certainly. We can use the set X itself. Wilansky in [47] repeats the perhaps apocryphal story that some mathematician once told his students in response to this question, "*I* am the point at infinity."

17.3.4. Definition. There is some disagreement about terminology here. If one defines *compactification* in a way that requires the original space to be dense in its compactification (see 17.3.2), then it seems inappropriate to call $\widetilde{X}$ a compactification of X when X is compact to begin with. For in this case the point ∞ is clearly an isolated point of $\widetilde{X}$ so that X is not dense in $\widetilde{X}$. Thus some writers restrict the terminology of *one-point compactification* to spaces X which are *noncompact* locally compact Hausdorff spaces.

17.3.5. Example. The interval $[0, 1]$ is the one-point compactification of $(0, 1]$.

17.3.6. Example. The unit circle $\mathbb{S}^1$ is the one-point compactification of $\mathbb{R}$. More generally, the n-sphere $\mathbb{S}^n$ is the one-point compactification of $\mathbb{R}^n$.

17.3.7. Definition. Here is a list of some of the *separation axioms* which may be enjoyed by a topological space X.

(i) X is a T_1 space if for every pair x and y of distinct points in X there exist a neighborhood of x which does not contain y and a neighborhood of y which does not contain x.
(ii) X is a T_2 space if it is Hausdorff.
(iii) X is a REGULAR space if closed sets and points not belonging to the sets can be separated by open sets.
(iv) X is a T_3 space if it is T_1 and regular.
(v) X is a COMPLETELY REGULAR space if closed sets and points not belonging to the sets can be functionally separated.
(vi) X is a TYCHONOFF SPACE (or a $T_{3\frac{1}{2}}$ space) if it is T_1 and completely regular.

17.3.8. Proposition. *A topological space X is T_1 if and only if every finite subset of X is closed.*

17.3.9. Proposition. *The following implications hold for topological spaces:*

$$T_{3\frac{1}{2}} \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1.$$

17.3.10. Proposition. *Every locally compact Hausdorff space is a Tychonoff space.*

PROOF. See [48], theorem 19.3.

17.3.11. Theorem (Stone-Čech compactification)**.** *Let X be a Tychonoff topological space. Then there exists a unique (up to homeomorphism) compactification $\beta(X)$ of X which satisfies the following condition: if K is a compact Hausdorff space, then every continuous function $f\colon X \to K$ has a continuous extension to $\beta(X)$.*

PROOF. See [1], section 2.17; [16], theorem 8.2; [34], proposition 4.3.18; or [48], theorem 19.12.

17.3.12. Remark. To appreciate how remarkable the Stone-Čech compactification $\beta(X)$ is, consider the problem of extending the function

$$f\colon (0,1] \to \mathbb{R}\colon x \mapsto \sin\left(\frac{1}{x}\right)$$

to a continuous function on some compact set in which $(0,1]$ is dense.

CHAPTER 18

LEBESGUE MEASURE

18.1. Positive Measures

Many authors refer to *positive measures* simply as *measures*. In these notes the word "measure" encompasses real measures, positive measures, complex measures, and signed measures, although little will be said about the last of these. Many facts about positive measures are quite similar to the facts about real measures presented in chapter 16, but often an additional hypothesis concerning the finiteness of the measure of some set is required.

First we observe that positive measures are monotone. Real measures, complex measures, and signed measures need not be monotone.

18.1.1. Proposition. *Let μ be a positive measure on a set S. If A, $B \overset{m}{\subseteq} S$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$.*

The next proposition says that positive measures are *countably subadditive.*

18.1.2. Proposition. *If μ is a positive measure on a measurable space S and if $A_1, A_2, \dots$ are measurable subsets of S, then*

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) \leq \sum_{k=1}^{\infty} \mu(A_k)\,.$$

18.1.3. Proposition. *Let μ be a positive measure on a measurable space S. If A, $B \overset{m}{\subseteq} S$, if $A \subseteq B$, and if $\mu(A) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.*

(Compare this with proposition 9.4.4.)

18.1.4. Proposition. *Let μ be a positive measure on a measurable space S. If A, $B \overset{m}{\subseteq} S$, and if $\mu(A \cap B) < \infty$, then*

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)\,.$$

(Compare this with proposition 9.4.3.)

18.1.5. Proposition. *Let S be a measurable space, (A_n) be a sequence of measurable subsets of S, and μ be a positive measure on S.*

(a) *If $A_n \uparrow B$, then $\mu(A_n) \to \mu(B)$.*
(b) *If $A_n \downarrow B$ and if $\mu(A_1) < \infty$, then $\mu(A_n) \to \mu(B)$.*

(Compare this with proposition 16.1.5.)

18.1.6. Example. Let S be a nonempty set and $\mathfrak{A} = \mathfrak{P}(S)$. For $A \subseteq S$ let $\mu(A) = \operatorname{card} A$ if A is finite and $\mu(A) = \infty$ if A is infinite. Then $(S, \mathfrak{A}, \mu)$ is a positive measure space. The positive measure μ is called COUNTING MEASURE on S.

18.1.7. Example. Give an example to show that the second hypothesis in part (b) of proposition 18.1.5 is necessary.

18.1.8. Example. Let S be an uncountable set and $\mathfrak{A} = \{A \subseteq S \colon A \text{ or } A^c \text{ is countable}\}$. For $A \in \mathfrak{A}$ let $\mu(A) = 0$ if A is countable and $\mu(A) = 1$ if A is uncountable. Then $(S, \mathfrak{A}, \mu)$ is a positive (real) measure space.

18.1.9. Proposition. *Suppose that* (A_n) *is a sequence of measurable subsets of a positive measure space* (S, μ)*. Then*

(a) $\mu(\liminf A_n) \le \liminf \mu(A_n)$.
(b) *If* $\mu(\cup_{n=1}^{\infty} A_n) < \infty$*, then* $\mu(\limsup A_n) \ge \limsup \mu(A_n)$.
(c) *If* $\mu(\cup_{n=1}^{\infty} A_n) < \infty$ *and* $\lim A_n$ *exists, then* $\lim \mu(A_n)$ *exists and* $\mu(\lim A_n) = \lim \mu(A_n)$.

18.1.10. Definition. If $(S, \mathfrak{A}, \mu)$ is any type of measure space we say that some property $P(x)$ of an element x in S holds *almost everywhere* or that property P holds for *almost all* $x \in S$ if

$$\mu(\{x \in S \colon P(x) \text{ does not hold}\}) = 0.$$

We use the abbreviation μ-a.e. (or just a.e.) for *almost everywhere.* Thus, for example, $f = g$ a.e. means that $f(x) = g(x)$ for almost all x (in the measure space S) and for a sequence (f_n) of functions on S, the notation $f_n \to g$ (a.e.) means that $f_n(x) \to g(x)$ for almost all x in S. Whenever we claim that something happens almost everywhere on a subset of $\mathbb{R}^n$ without specifying a measure, Lebesgue measure is understood.

18.1.11. Proposition (The Borel-Cantelli lemma)**.** *Let* (A_n) *be a sequence of measurable subsets of a positive measure space* (S, μ) *such that* $\sum_{k=1}^{\infty} \mu(A_k) < \infty$*. Then almost all* x *lie in only finitely many of the sets* A_k*.*

18.1.12. Definition. A point a in a positive measure space (X, μ) is an ATOM of X if $\mu(\{a\}) > 0$.

18.1.13. Proposition. *Let* (X, μ) *be a positive real measure space and* A *be the set of atoms of* X*. Then there exist positive real measures* μ_d *and* μ_c *on* X *such that* (X, μ_d) *is a discrete measure space,* (X, μ_c) *is a measure space with no atoms, and* $\mu = \mu_d + \mu_c$*. Show also that* μ *has at most countably many atoms and that*

$$\mu_d = \sum_{a \in A} \mu(\{a\})\, \delta_a$$

(where δ_a *is Dirac measure concentrated at* a*).*

The measure μ_d is the DISCRETE PART of μ and μ_c is the CONTINUOUS PART of μ.

18.1.14. Definition. A positive measure space (S, μ) (or its measure μ) is said to be σ-FINITE if S is the disjoint union of a countable collection of measurable subsets each having finite measure.

18.1.15. Exercise. Show that proposition 18.1.13 can be generalized to σ-finite positive measures.

18.2. Outer Measures

Positive measures are designed to be generalizations of such things as *length* on the real line, *area* in the plane, and *volume* in 3-space. Ideally we would like to have on $\mathbb{R}$, for example, a positive measure on $\mathfrak{P}(\mathbb{R})$ which is translation invariant (see definition 18.3.6) and has the property that the measure of the unit interval is 1. Unfortunately this cannot be achieved (see proposition 18.3.12). We can scarcely give up positivity, countable additivity, translation invariance, or the correct measure of interval length; so we have to settle for a positive measure defined on some proper subset of the power set $\mathfrak{P}(\mathbb{R})$. Lebesgue measure on $\mathbb{R}$ is the best we can do. It is defined on a quite large subset of $\mathfrak{P}(\mathbb{R})$, a set in fact so large that finding a set not in the domain of Lebesgue measure requires the use of the axiom of choice. The price we pay for this success is the appearance of a number of technical details, which, depending on one's proclivities, are either fascinating or annoying. The next two sections discuss the construction of Lebesgue measure.

In the current section we give one answer to the rather general question: *where do positive measures come from?* Briefly stated (in terms to be defined shortly): start with a *gauge* ℓ on a *sequential covering class* of subsets of a set S; this produces an *outer measure* m_ℓ on the power set $\mathfrak{P}(S)$, which in turn produces a positive measure on the σ-algebra of m_ℓ*-measurable* subsets of S.

18.2.1. Definition. A family $\mathfrak{C}$ of subsets of a set S is a SEQUENTIAL COVERING CLASS for S if

(i) $\emptyset \in \mathfrak{C}$ and
(ii) $S = \bigcup_{k=1}^{\infty} C_k$ for some sequence (C_k) of sets in $\mathfrak{C}$.

18.2.2. Example. The family of bounded open intervals in $\mathbb{R}$ is a sequential covering class for $\mathbb{R}$.

18.2.3. Definition. Let $\mathfrak{C}$ be a sequential covering class for a set S. A function $\ell\colon \mathfrak{C} \to [0,\infty]$ such that $\ell(\emptyset) = 0$ is a GAUGE on S. (This is clearly an abuse of language: we should say that ℓ is a gauge *on* $\mathfrak{C}$).

An *outer measure* is a positive, increasing, countably subadditive function on the power set of a set whose value at $\emptyset$ is 0.

18.2.4. Definition. Let S be a nonempty set. A function $m\colon \mathfrak{P}(S) \to [0,\infty]$ is an OUTER MEASURE on S if:

(i) $m(\emptyset) = 0$;
(ii) $m(A) \le m(B)$ whenever $A \subseteq B \subseteq S$; and
(iii) $m\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} m(A_k)$ whenever (A_k) is a sequence of subsets of S.

(Saying that m is an outer measure *on* S is again a standard abuse of language: we should say that it is an outer measure *on* $\mathfrak{P}(S)$.)

18.2.5. Example. Let $\mathfrak{C}$ be a sequential covering class of a set S and $\ell\colon \mathfrak{C} \to [0,\infty]$ be a gauge on S. For every $A \subseteq S$ define

$$m_\ell(A) := \inf\left\{\sum_{k=1}^{\infty} \ell(C_k)\colon (C_k) \text{ is a sequence in } \mathfrak{C} \text{ which covers } A.\right\}$$

(For convenience we will write $m(A)$ for $m_\ell(A)$ when no confusion seems likely.) Then m is an outer measure on S.

18.2.6. Example. The preceding example says that any gauge on an arbitrary sequential covering class of a set S generates an outer measure on S. It does *not* say that it generates an *interesting* outer measure. Let $S = \mathbb{R}$ and $\mathfrak{C}$ be the sequential covering class consisting of all bounded open intervals in $\mathbb{R}$. If $a < b$ define $\ell\big((a,b)\big) = (b-a)^{-1}$ and let $\ell(\emptyset) = 0$. Describe the outer measure generated by the gauge ℓ.

18.2.7. Definition. Let m be an outer measure on a set S. A subset E of S is m-MEASURABLE (or MEASURABLE WITH RESPECT TO THE OUTER MEASURE m) if

$$m(T) = m(T \cap E) + m(T \cap E^c)$$

for every $T \subseteq S$.

18.2.8. Exercise. Let $\mathfrak{C}$ be the family of open intervals in $\mathbb{R}$ of length 1 together with $\emptyset$. Let $\ell(\emptyset) = 0$ and $\ell(C) = 1$ if $C \in \mathfrak{C}$ and $C \neq \emptyset$. Then $\mathfrak{C}$ is a sequential covering class and ℓ is a gauge on $\mathfrak{C}$. Denote by m the outer measure induced on $\mathfrak{P}(\mathbb{R})$ by ℓ.

(a) Find $m(\{0\})$.
(b) Find $m\big((0,1)\big)$.
(c) Find $m\big([0,1]\big)$.
(d) Which subsets of $\mathbb{R}$ are m-measurable?
(e) What measure μ_m is induced by m on the σ-algebra of m-measurable subsets of $\mathbb{R}$?

The preceding exercise shows that the outer measure induced on a sequential covering class by a gauge may not be an extension of the gauge. What do we have to require of a gauge ℓ so that the outer measure m it induces *will* be an extension of ℓ? Since m is countably subadditive whether or not ℓ is, it is clear that in order for m to be an extension of ℓ it is necessary that ℓ itself be subadditive. This is also sufficient.

18.2.9. Proposition. *Let ℓ be a gauge on a sequential covering class $\mathfrak{C}$ of a nonempty set S and m be the outer measure induced by ℓ. Then $m(C) = \ell(C)$ for every $C \in \mathfrak{C}$ if and only if ℓ is countably subadditive (in the sense that $\ell(C) \leq \sum_{k=1}^{\infty} \ell(C_k)$ whenever $C \in \mathfrak{C}$ and (C_k) is a sequence in $\mathfrak{C}$ which covers C).*

18.2.10. Example. Arguably the most important example of a gauge is *length*

$$\ell\big((a,b)\big) := b - a \qquad \text{when } a \leq b$$

defined on the sequential covering class of bounded open intervals in $\mathbb{R}$. The outer measure m generated by this gauge is called Lebesgue outer measure. When m is Lebesgue outer measure on $\mathbb{R}$, an m-measurable subset of $\mathbb{R}$ is a Lebesgue measurable set.

18.2.11. Remark. Notice that in order to show that a set E is m-measurable it is necessary only to verify that $m(T) \geq m(T \cap E) + m(T \setminus E)$ for each subset T of S, since the reverse inequality holds by virtue of the countable subadditivity of m.

18.2.12. Proposition. *Let m be an outer measure on a set S. A subset A of S is m-measurable if and only if for every $\epsilon > 0$ there exists an m-measurable set $E \subseteq A$ such that $m(A \setminus E) < \epsilon$.*

18.2.13. Proposition. *Let m be an outer measure on a set S. A subset A of S is m-measurable if and only if for every $\epsilon > 0$ there exists an m-measurable set E such that $m(A \triangle E) < \epsilon$.*

18.2.14. Definition. Let m be an outer measure on a set S. A subset A of S is a null set with respect to the outer measure m if $m(A) = 0$. For clarity such a set is often referred to as an m-null set to avoid confusion with the empty set.

18.2.15. Proposition. *If m is an outer measure on a set S, then every m-null subset of S is m-measurable.*

18.2.16. Proposition. *If m is an outer measure on a set S, then the union of a countable family of m-null sets in S is itself an m-null set.*

18.2.17. Proposition. *Let m be an outer measure on a set S and $A \subseteq S$. If $E_1, \ldots, E_n$ are disjoint m-measurable subsets of S, then*

$$m\left(A \cap \left[\biguplus_{k=1}^{n} E_k\right]\right) = \sum_{k=1}^{n} m(A \cap E_k)\,.$$

18.2.18. Theorem (Carathéodory)**.** *Let m be an outer measure on a set S. Then the family $\mathfrak{M}_m$ of all m-measurable subsets of S is a σ-algebra of subsets of S. Thus $(S, \mathfrak{M}_m)$ is a measurable space. Furthermore, the restriction μ_m of m to the family $\mathfrak{M}_m$ is a positive measure, so $(S, \mathfrak{M}_m, \mu_m)$ is a positive measure space.*

Proof. The proof is standard, but somewhat complicated. It is easy to find. Some sources are [1], Theorem 9.19; [18], Theorem 1.11; [28], chapter 11, lemma 7.3; [32], propositions 4.15 and 4.16; and [36], chapter 12, theorem 1.

18.2.19. Proposition. *If an outer measure is finitely additive, then it must be countably additive.*

18.2.20. Proposition. *Let A be a subset of $\mathbb{R}$ with finite Lebesgue outer measure m. Define $F(x) = m\big(A \cap (-\infty, x]\big)$ for every $x \in \mathbb{R}$. Then F is uniformly continuous.*

18.3. Lebesgue Measure on $\mathbb{R}$

In this section we give some of the details concerning Lebesgue measure on the real line. Many authors carry out the details in the more general setting of Euclidean n-space. See for example [2], section 15; [18], chapter 2, section 6; [19], volume 1, section 115; [37], chapter 2, 2.19–2.20; and [44], chapter V.

18.3.1. Definition. Example 18.2.10 defines Lebesgue outer measure $m = m_\ell$ (where ℓ is the length function on bounded open intervals in $\mathbb{R}$) and theorem 18.2.18 asserts that the restriction λ of m to the σ-algebra $\mathfrak{M}_\lambda$ of Lebesgue measurable subsets of R is a positive measure. This measure is LEBESGUE MEASURE on $\mathbb{R}$. In the sequel when a measure is denoted by λ it is intended to be Lebesgue measure.

18.3.2. Proposition. *Every subset of a measurable set with Lebesgue measure zero is itself measurable and has Lebesgue measure zero.*

18.3.3. Proposition. *The Lebesgue outer measure of any interval in $\mathbb{R}$ is its length.*

HINT FOR PROOF. Before starting the proof it is good to observe that for a bounded open interval (a, b) the assertion that $m\big((a, b)\big) = \ell\big((a, b)\big)$ may not be entirely obvious. Certainly we know that $m\big((a, b)\big) \leq \ell\big((a, b)\big)$. But why does the reverse inequality hold? How do we know that the interval (a, b) cannot be covered by a family of open intervals whose total length is strictly less than $b - a$? One's intuition, of course, rebels at the idea, but in measure theory, as in much of mathematics, one's intuition is not an infallible guide. Prove first that the assertion holds for closed bounded intervals (using a compactness argument); then prove it for all bounded intervals. Finally deal with the case of unbounded intervals.

18.3.4. Example. The Lebesgue outer measure of the Cantor set is zero.

18.3.5. Theorem. *Every Borel set in $\mathbb{R}$ is Lebesgue measurable.*

Thus the collection of Lebesgue measurable subsets of $\mathbb{R}$ is quite large indeed. One may well wonder if there are any of $\mathbb{R}$ which are *not* Lebesgue measurable. We will see shortly (in examples 18.3.9 and 18.3.10) that such sets do exist. The proof will require the *axiom of choice*.

Another perfectly reasonable question: is every Lebesgue measurable subset of $\mathbb{R}$ a Borel set? A cardinality argument, which can be found in [23], remark 10.21, shows that there are in fact $2^{\mathfrak{c}}$ Lebesgue measurable sets which are not Borel sets (where $\mathfrak{c}$ is the cardinality of the continuum, that is, of the set of real numbers). The explicit construction of classes of Lebesgue measurable sets strictly containing the Borel sets is undertaken in [39] and [42].

18.3.6. Definition. A measure μ defined on a σ-algebra $\mathfrak{A}$ of subsets of an Abelian group G is TRANSLATION INVARIANT if $A + x \in \mathfrak{A}$ and $\mu(A + x) = \mu(A)$ whenever $A \in \mathfrak{A}$ and $x \in G$.

18.3.7. Proposition. *Lebesgue measure is translation invariant on $\mathbb{R}$ (regarded as an Abelian group under addition).*

18.3.8. Proposition. *Suppose that μ is a measure defined on a σ-algebra $\mathfrak{A}$ of subsets of a set S and that T is a member of $\mathfrak{A}$. Let $\mathfrak{B} = \{A \cap T \colon A \in \mathfrak{A}\}$ and $\nu = \mu\big|_{\mathfrak{B}}$. Then $(T, \mathfrak{B}, \nu)$ is a measure space.*

The preceding proposition makes it possible to regard any measurable subset of a measure space as a measure space in its own right. The measure ν in that proposition is the MEASURE INDUCED ON T BY μ. (Any restriction of Lebesgue measure λ will also be denoted by λ.)

18.3.9. Example. The *axiom of choice* guarantees the existence of a set A contained in the interval $(0, 1)$ which contains exactly one element from each of the equivalence classes which constitute the Abelian group $\mathbb{R}/\mathbb{Q}$. The set A is not Lebesgue measurable.

HINT FOR PROOF. Show that $(0, 1) \subseteq \bigcup\{A + r \colon r \in \mathbb{Q} \cap (-1, 1)\} \subseteq (-1, 2)$.

You can find a leisurely detailed treatment of this nonmeasurable set in [10], pages 289–291.

It is not terribly difficult to generalize example 18.3.9 from the open unit interval to an arbitrary subset of $\mathbb{R}$ with strictly positive measure.

18.3.10. Example. Every subset of the real line with strictly positive Lebesgue measure contains a nonmeasurable set.

Proof. See [23], theorem 10.28 or [37], (corollary to) theorem 2.22.

18.3.11. Proposition. *Lebesgue outer measure is not countably additive.*

18.3.12. Proposition. *There does not exist a positive measure on $\mathfrak{P}(\mathbb{R})$ which is translation invariant and such that the measure of the closed unit interval $[0, 1]$ is 1.*

18.3.13. Proposition. *If A is a Lebesgue measurable subset of $\mathbb{R}$, then the (Lebesgue) measure of A is the infimum of the measures of open sets containing A.*

18.3.14. Proposition. *If A is a Lebesgue measurable subset of $\mathbb{R}$, then the (Lebesgue) measure of A is the supremum of the measures of compact sets contained in A.*

18.3.15. Exercise. Find the largest number t having the following property:

> If any four Lebesgue measurable subsets of the unit interval are given, each having measure at least $1/3$, then two of them intersect in a set whose Lebesgue measure is at least t.

18.3.16. Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be Lebesgue measurable if it is $(\mathfrak{M}_\lambda, \mathfrak{B}\mathrm{or}(\mathbb{R}))$-measurable. A similar convention holds for functions between Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^n$.

18.3.17. Proposition. *Every Borel measurable function is Lebesgue measurable, but not conversely.*

18.3.18. Proposition. *If f be a function from $(\mathbb{R}, \mathfrak{M}_\lambda)$ to $(\mathbb{R}, \mathfrak{B}\mathrm{or}(\mathbb{R}))$. Then the following are equivalent:*

(a) *f is Lebesgue measurable.*
(b) *$f^{\leftarrow}\big((-\infty, a)\big) \overset{m}{\subseteq} \mathbb{R}$ for all $a \in \mathbb{R}$.*
(c) *$f^{\leftarrow}\big((a, \infty)\big) \overset{m}{\subseteq} \mathbb{R}$ for all $a \in \mathbb{R}$.*

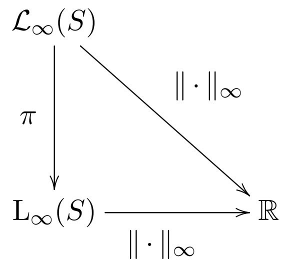
18.3.19. Exercise. Recall that proposition 15.3.4 asserts that the composite of two measurable functions is measurable.

(a) Prove that a *Lebesgue measurable* function is indeed a measurable function.
(b) Prove that the composite of two Lebesgue measurable functions need not be Lebesgue measurable.
(c) So proposition 15.3.4 says that the composite of two measurable functions must be measurable, but what you just proved in part (b) says that they need not be. Similarly, in [28] we find at the top of page 247, "If $f : X \to Y$ is measurable and $g : Y \to Z$ is measurable, then the composite $g \circ f$ is measurable," while in [23] in the paragraph immediately preceding theorem 11.7 we read, "Thus the composition of two measurable functions need not be measurable." *What* is going on? Explain *very* carefully.
(d) Restate exercise 15.3.4 so that no confusion can possibly result.

18.4. The Space $\mathrm{L}_\infty(S)$

18.4.1. Definition. Let (S, μ) be a (real or positive) measure space. A function $f \in \mathcal{M}(S)$ is essentially bounded if there exists $M > 0$ such that $|f| \le M$ almost everywhere. The infimum of all such numbers M is the essential supremum of $|f|$ and is denoted by $\|f\|_\infty$. The family of all essentially bounded functions on S is denoted by $\mathcal{L}_\infty(S)$.

18.4.2. Proposition. *Let (S, μ) be a (real or positive) measure space and $\mathcal{N}(S)$ be the family of all measurable functions f on S such that $f(x) = 0$ for almost all $x \in S$. Define $\mathrm{L}_\infty(S) := \mathcal{L}_\infty(S)/\mathcal{N}(S)$. Then there exists a unique norm (also denoted by $\|\cdot\|$) on $\mathrm{L}_\infty(S)$ that makes the* *following diagram commute.*



The norm $\|\cdot\|$ is the ESSENTIAL SUPREMUM NORM (or L-INFINITY NORM) on $\mathrm{L}_\infty(S)$.

18.4.3. Notation. If we wish to emphasize the role played by a measure μ we may write $\mathcal{L}_\infty(S,\mu)$ or $\mathrm{L}_\infty(S,\mu)$ for $\mathcal{L}_\infty(S)$ or $\mathrm{L}_\infty(S)$, respectively.

18.4.4. Example. If S is a (real or positive) measure space, then $\mathrm{L}_\infty(S)$ is a unital commutative normed algebra.

18.4.5. Example. If S is a (real or positive) measure space, then $\mathrm{L}_\infty(S)$ is a Riesz space.

18.4.6. Definition. Let (S,μ) be a measure space and $f \in \mathcal{M}(S)$. The ESSENTIAL RANGE of f, denoted by $\operatorname{essran} f$, is the set of all $\lambda \in \mathbb{R}$ such that $\mu\big(f^{\leftarrow}(B_\epsilon(\lambda))\big) > 0$ for every $\epsilon > 0$.

18.4.7. Example. .

(a) The essential range of the characteristic function $\chi_{\mathbb{Q}} \colon \mathbb{R} \to \mathbb{R}$ is $\{0\}$, which is a proper subset of its range $\{0,1\}$.
(b) Let $f(x) = \arctan x$ for all $x \in \mathbb{R}$. Then the essential range of f is $[-\frac{\pi}{2}, \frac{\pi}{2}]$, which properly contains its range $(-\frac{\pi}{2}, \frac{\pi}{2})$.

18.4.8. Proposition. *Let (S,μ) be a measure space and $f \in \mathcal{L}_\infty(S)$. The essential range of f is a compact subset of $\mathbb{R}$ and that $\|f\|_\infty = \sup\{|\lambda| \colon \lambda \in \operatorname{essran} f\}$.*

Note: This result also holds for $[f] \in \mathrm{L}_\infty(S)$ because $\operatorname{essran} f = \operatorname{essran} g$ whenever $f = g$ a.e.

18.4.9. Definition. Let A be a unital algebra and $a \in A$. The SPECTRUM of a, denoted by $\sigma(a)$, is the set of all $\lambda \in \mathbb{C}$ such that $a - \lambda 1$ is *not* invertible.

18.4.10. Example. If z is an element of the algebra $\mathbb{C}$ of complex numbers, then $\sigma(z) = \{z\}$.

18.4.11. Example. The family $\mathfrak{M}_3$ of 3×3 matrices of complex numbers is a unital algebra under the usual matrix operations. The spectrum of the matrix $\begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$ is $\{1,2\}$.

18.4.12. Example. Let X be a compact Hausdorff space. If f is an element of the algebra $\mathcal{C}(X)$ of continuous complex valued functions on X, then the spectrum of f is its range.

18.4.13. Example. Let $[f] \in \mathrm{L}_\infty(S,\mu)$ where (S,μ) is a measure space. Then $\sigma([f]) = \operatorname{essran} f$.

CHAPTER 19

THE LEBESGUE INTEGRAL

One matter that must be clear from the start is what the Lebesgue integral is *not*. It is not a creature dependent in some way on Lebesgue measure. It is *not* by definition an integral of a function with respect to Lebesgue measure. We will define the Lebesgue integral of functions with respect to *arbitrary* real measures, complex measures, and positive measures. Of course, one special case is integration with respect to Lebesgue measure. But in general Lebesgue integration is defined for arbitrary measures.

19.1. Integration of Simple Functions

19.1.1. Definition. Let $\mathfrak{A}$ be an algebra of subsets of a set S. An $\mathfrak{A}$-SIMPLE FUNCTION on S is a linear combination of characteristic functions of members of $\mathfrak{A}$. When the algebra $\mathfrak{A}$ is clear from context we will write *simple functions* for $\mathfrak{A}$-*simple functions*. We conclude easily from examples 15.3.3 and 15.3.11 that on a measurable space *every simple function is measurable.* A moment's reflection indicates that a function $\sigma\colon S \to \mathbb{R}$ is simple if and only if

(i) the range of σ is finite, and
(ii) $\sigma^{\leftarrow}(\{y\}) \in \mathfrak{A}$ for every $y \in \operatorname{ran}\sigma$.

Let $\mathcal{S}\mathrm{im}(S)$ be the family of all (real valued) simple functions on S and $\mathcal{S}\mathrm{im}(S)^+$ be the family of all positive members of $\mathcal{S}\mathrm{im}(S)$.

Suppose σ is an $\mathfrak{A}$-simple function on S and that $\operatorname{ran}\sigma = \{\alpha_1, \dots, \alpha_n\}$ (where the α_k's are distinct). Let $A_k = \sigma^{\leftarrow}(\{\alpha_k\})$ for $1 \le k \le n$. It is easy to see that

$$\sigma = \sum_{k=1}^{n} \alpha_k \chi_{A_k} \,. \tag{19.1}$$

When a simple function is written in the form (19.1) we say that it is written in STANDARD FORM.

When $(S,\mathfrak{A})$ is a measurable space we say that the A_k's form a MEASURABLE PARTITION (or a MEASURABLE DECOMPOSITION) of S. That is, $S = \biguplus_{k=1}^{n} A_k$ and each A_k is a measurable set.

19.1.2. Exercise. Define $\sigma\colon [0,10] \to \mathbb{R}$ by

$$\sigma = 2\,\chi_{[0,4]} - \chi_{(3,6)} + \chi_{[5,6]} - \chi_{[8,10]} + 2\,\chi_{\mathbb{Q}}\chi_{[7,9]} \,.$$

Write σ in standard form.

A pleasant and crucial observation is that on a measurable space *every* positive measurable function is the pointwise limit of an increasing sequence of positive simple functions. The proof is quite standard.

19.1.3. Proposition. *Let $f\colon S \to [0,\infty]$ be a positive measurable function on a measurable space S. Then there exists a sequence (σ_n) of simple functions on S such that*

$$0 \le \sigma_n \uparrow f \text{ (ptws)} \,.$$

PROOF. For each $n \in \mathbb{N}$ let

$$A_{n,k} = f^{\leftarrow}\left(\left[\tfrac{k-1}{2^n}, \tfrac{k}{2^n}\right)\right) \quad \text{whenever } 1 \le k \le n\,2^n$$

and

$$B_n = f^{\leftarrow}\left([n,\infty)\right).$$

Then define

$$\sigma_n = \sum_{k=1}^{n2^n} \tfrac{k-1}{2^n}\,\chi_{A_{n,k}} + n\,\chi_{B_n}.$$

It is not difficult to see that the sequence (σ_n) satisfies $0 \le \sigma_n \uparrow f$. (Notice, in particular, that if x is fixed, then $\sigma_n(x) \ge f(x) - 2^{-n}$ for all $n > f(x)$.) □

19.1.4. Exercise. Let $f\colon (0,3] \to \mathbb{R}\colon x \mapsto \frac{1}{x}$. Make a careful sketch of the corresponding simple functions σ_2 and σ_3 which appear in the proof of proposition 19.1.3.

19.1.5. Example. If $\mathfrak{A}$ is an algebra of subsets of a set S, then $\mathcal{S}\mathrm{im}(S)$ is a Riesz space. Notice that, in particular, if $(S,\mathfrak{A})$ is a measurable space, then $\mathcal{S}\mathrm{im}(S)$ is a Riesz subspace of $\mathcal{M}(S)$ (see example 15.3.11).

19.1.6. Example. If $\mathfrak{A}$ is an algebra of subsets of a set S, then $\mathcal{S}\mathrm{im}(S)$ is an algebra. Moreover, it is a normed algebra (under the uniform norm). If $(S,\mathfrak{A})$ is a measurable space, then $\mathcal{S}\mathrm{im}(S)$ is a subalgebra of $\mathcal{M}(S)$ (see example 15.3.12).

Now, supposing that our algebra of subsets is equipped with a finitely additive set function μ. Then we can define the *integral* of a simple function with respect to μ.

19.1.7. Definition. Let μ be a real finitely additive set function on an algebra $\mathfrak{A}$ of subsets of a set S and $\sigma = \sum \alpha_k \chi_{A_k}$ be an $\mathfrak{A}$-simple function on S written in standard form. Then the LEBESGUE INTEGRAL of σ (with respect to μ) is defined by

$$\int_S \sigma\,d\mu := \sum \alpha_k \mu(A_k)\,. \tag{19.2}$$

When no confusion seems likely we may simplify the notation to $\int \sigma\,d\mu$ or even $\int \sigma$. In these notes we are interested primarily in the Lebesgue integral. So whenever the unmodified term *integral* appears we take it to mean *Lebesgue integral* (unless, of course, context indicates otherwise).

The next two exercises are almost obvious. The first says that in order for equation (19.2) to hold it is not important that σ be written in standard form. Any decomposition of its domain into members of the algebra $\mathfrak{A}$ such that σ is constant on each member of the decomposition will do.

19.1.8. Proposition. *Let $\mathfrak{A}$ be an algebra of subsets of a set S, μ be a real finitely additive set function on $\mathfrak{A}$, σ be an $\mathfrak{A}$-simple function on S, and $\{A_1,\dots,A_n\}$ be a decomposition of S into members of $\mathfrak{A}$ such that σ is constant on each A_k. Then $\int \sigma = \sum \alpha_k \mu(A_k)$, where α_k is the value of σ on A_k $(1 \le k \le n)$.*

19.1.9. Proposition. *Let σ and τ be simple functions on the same algebra $\mathfrak{A}$ of subsets of a set S and μ be a real finitely additive set function on $\mathfrak{A}$. If $\mu(\{x \in S\colon \sigma(x) \ne \tau(x)\}) = 0$, then $\int \sigma\,d\mu = \int \tau\,d\mu$.*

19.1.10. Exercise. Let σ be the simple function defined in exercise 19.1.2 and λ be Lebesgue measure. Compute $\int \sigma\,d\lambda$.

19.1.11. Proposition. *Let $\mathfrak{A}$ be an algebra of subsets of a set S, μ be a real finitely additive set function on $\mathfrak{A}$, and σ be an $\mathfrak{A}$-simple function on S. Then*

$$\int_S \sigma\,d\mu = \int_S \sigma^+\,d\mu - \int_S \sigma^-\,d\mu\,.$$

19.1.12. Proposition. *Let μ be a real measure on a measurable space S and σ be a simple function on S. Then*

$$\int_S \sigma\,d\mu = \int_S \sigma\,d\mu^+ - \int_S \sigma\,d\mu^- \,.$$

19.1.13. Example. Let $\mathfrak{A}$ be an algebra of subsets of a set S and μ be a real finitely additive set function on $\mathfrak{A}$. Then the map

$$\widetilde{\mu}\colon\ \mathcal{S}\mathrm{im}(S) \to \mathbb{R}\colon \sigma \mapsto \int_S \sigma\,d\mu$$

is a linear functional on the vector space $\mathcal{S}\mathrm{im}(S)$; that is, $\widetilde{\mu} \in \mathcal{S}\mathrm{im}(S)^\dagger$.

19.1.14. Example. If μ is a *positive* finitely additive set function on an algebra $\mathfrak{A}$ of subsets of a set S, then the function

$$\widetilde{\mu}\colon\ \mathcal{S}\mathrm{im}(S) \to \mathbb{R}\colon \sigma \mapsto \int_S \sigma\,d\mu$$

is a positive (and therefore order bounded—see proposition 14.3.5) linear functional on the Riesz space $\mathcal{S}\mathrm{im}(S)$. That is, $\widetilde{\mu} \in \mathcal{S}\mathrm{im}(S)^\sim$.

19.1.15. Example. If μ is a bounded real finitely additive set function on an algebra $\mathfrak{A}$ of subsets of a set S, then the function

$$\widetilde{\mu}\colon\ \mathcal{S}\mathrm{im}(S) \to \mathbb{R}\colon \sigma \mapsto \int_S \sigma\,d\mu$$

is a bounded linear functional on the normed linear space $\mathcal{S}\mathrm{im}(S)$. That is, $\widetilde{\mu} \in \mathcal{S}\mathrm{im}(S)^*$.

19.1.16. Proposition. *Let $\mathfrak{A}$ be an algebra of subsets of a set S. Then the map*

$$\Phi\colon\ \mathrm{ba}\,S \to \mathcal{S}\mathrm{im}(S)^*\colon \mu \mapsto \widetilde{\mu}$$

(where $\widetilde{\mu}$ is defined in example 19.1.15) is an isometric isomorphism.

19.1.17. Definition. Let μ be a real finitely additive set function on an algebra $\mathfrak{A}$ of subsets of a set S, σ be an $\mathfrak{A}$-simple function on S, and $E \in \mathfrak{A}$. Then the LEBESGUE INTEGRAL of σ OVER E (WITH RESPECT TO μ) is defined by

$$\int_E \sigma\,d\mu := \int_S \chi_E \sigma\,d\mu. \tag{19.3}$$

19.1.18. Exercise. Alter definition 19.1.7 to define the *Lebesgue integral* of a positive simple function σ on a measurable space $(S, \mathfrak{A})$ with respect to a positive measure μ. (Explain why it is a good idea to restrict our attention to *positive* simple functions.) Verify that if σ and τ are positive simple functions and $\alpha \geq 0$, then $\int(\sigma+\tau)\,d\mu = \int \sigma\,d\mu + \int \tau\,d\mu$ and $\int \alpha\sigma\,d\mu = \alpha\int\sigma\,d\mu$.

19.2. Integration of Positive Functions

19.2.1. Definition. Let μ be a positive measure on a measurable space S and $f\colon S \to [0,\infty]$ be a positive measurable function on S. Then the LEBESGUE INTEGRAL of f (with respect to μ) is defined by

$$\int_S f\,d\mu = \sup\left\{\int_S \sigma\,d\mu\colon \sigma \in \mathcal{S}\mathrm{im}(S) \text{ and } 0 \leq \sigma \leq f\right\}.$$

As with simple functions if E is a measurable subset of S we define

$$\int_E f\,d\mu = \int_S \chi_E f\,d\mu.$$

19.2.2. Proposition. *Let (S,μ) be a positive measure space and f, $g\colon S \to [0,\infty]$ be positive measurable functions on S. If $f \leq g$, then $\int_S f\,d\mu \leq \int_S g\,d\mu$.*

19.2.3. Theorem (Monotone Convergence Theorem)**.** *Let* (S,μ) *be a positive measure space and* (f_n) *be a sequence of measurable functions from* S *into* $[0,\infty]$*. If* $0 \le f_n \uparrow g$ (ptws)*, then*

$$\int_S g\,d\mu = \lim_{n\to\infty}\int_S f_n\,d\mu.$$

Proof. See [10] theorem 18.7; [11], proposition 2.3.3; [32], theorem 4.6; or [37], theorem 1.26;

19.2.4. Proposition. *Let* f, $g\colon S \to [0,\infty]$ *be positive measurable functions on a positive measure space* (S,μ)*. Then* $\int_S (f+g)\,d\mu = \int_S f\,d\mu + \int_S g\,d\mu$*.*

19.2.5. Proposition. *Let* $f\colon S \to [0,\infty]$ *be a positive measurable function on a positive measure space* (S,μ) *and* $\alpha \ge 0$*. Then* $\int_S \alpha f\,d\mu = \alpha\int_S f\,d\mu$*.*

19.2.6. Proposition. *Let* $f\colon S \to [0,\infty]$ *be a positive measurable function on a positive measure space* (S,μ)*. Then* $f=0$ *a.e. if and only if* $\int_S f\,d\mu = 0$*.*

19.2.7. Proposition. *Let* $f\colon S \to [0,\infty]$ *be a positive measurable function on a positive measure space* (S,μ) *and* A *and* B *be measurable subsets of* S *with* $A \subseteq B$*. Then* $\int_A f\,d\mu \le \int_B f\,d\mu$*.*

19.2.8. Proposition. *For each* $n \in \mathbb{N}$ *let* $f_n\colon S \to [0,\infty]$ *be a positive measurable function on a positive measure space* (S,μ)*. Then*

$$\int_S\left(\sum_{n=1}^{\infty} f_n\right)d\mu = \sum_{n=1}^{\infty}\int_S f_n\,d\mu.$$

19.2.9. Corollary. *Suppose that* $a_{jk} \ge 0$ *for every* j, $k \in \mathbb{N}$*. Then*

$$\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_{jk} = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} a_{jk}.$$

19.2.10. Proposition (Fatou's lemma)**.** *For each* $n \in \mathbb{N}$ *let* $f_n\colon S \to [0,\infty]$ *be a positive measurable function on a positive measure space* (S,μ)*. Then*

$$\int_S (\liminf f_n)\,d\mu \le \liminf \int_S f_n\,d\mu.$$

19.2.11. Example. Give an example to show that the inequality in the preceding proposition may be strict.

19.2.12. Proposition. *Let* $f\colon S \to [0,\infty]$ *be a positive measurable function on a positive measure space* (S,μ)*. Define for every* $A \overset{m}{\subseteq} S$

$$\nu(A) = \int_A f\,d\mu. \tag{19.4}$$

Then ν *is a positive measure on* S *and*

$$\int_S g\,d\nu = \int_S gf\,d\mu \tag{19.5}$$

for every measurable $g \ge 0$ *on* S*.*

The equation (19.5) is often written in abbreviated form

$$d\nu = f\,d\mu$$

or even more fancifully

$$\frac{d\nu}{d\mu} = f.$$

For this reason f is often referred to as the *Radon-Nikodym derivative* of ν with respect to μ.

19.2.13. Definition. Let μ be a real measure on a measurable space S and $f\colon S \to [0,\infty]$ be a positive measurable function on S. Then we define

$$\int_S f\,d\mu = \int_S f\,d\mu^+ - \int_S f\,d\mu^-$$

whenever not both $\int_S f\,d\mu^+$ and $\int_S f\,d\mu^-$ are infinite.

19.3. Integration of Real and Complex Valued Functions

19.3.1. Definition. Let (S,μ) be a positive measure space. A measurable complex (or extended real) valued function is (Lebesgue) integrable if $\int_S |f|\,d\mu < \infty$. We denote the family of all complex valued Lebesgue integrable functions on S by $\mathcal{L}_1(\mu,\mathbb{C})$ or $\mathcal{L}_1(S,\mathbb{C})$ and the family of all extended real valued Lebesgue integrable functions on S by $\mathcal{L}_1(\mu)$ or $\mathcal{L}_1(S)$. If $f \in \mathcal{L}_1(\mu)$ we define

$$\int_S f\,d\mu = \int_S f^+\,d\mu - \int_S f^-\,d\mu$$

and if $f \in \mathcal{L}_1(\mu,\mathbb{C})$ we define

$$\int_S f\,d\mu = \int_S u\,d\mu + i\int_S v\,d\mu$$

where $u(x)$ is the real part of $f(x)$ and $v(x)$ is the imaginary part of $f(x)$ for each $x \in S$.

Alternative notations for the Lebesgue integral are $\int_S f(x)\,d\mu(x)$ and $\int f\,d\mu$. The notations $\int_S f(x)\,d\lambda(x)$ and $\int_S f\,d\lambda$ denote integration with respect to Lebesgue measure. Because of proposition 19.3.18 the integral with respect to Lebesgue measure is frequently written as $\int_S f(x)\,dx$. When f is a function defined on an interval $[a,b]$ it is common practice to write $\int_a^b f\,d\mu$ for $\int_{[a,b]} f\,d\mu$. A similar convention holds for other intervals $(a,b]$, $[a,\infty)$, etc.

19.3.2. Proposition. *If (S,μ) is a positive measure space, then $\mathcal{L}_1(\mu)$ is a Riesz space and the map*

$$p\colon \mathcal{L}_1(\mu) \to \mathbb{R}\colon f \mapsto \int_S f\,d\mu$$

is a positive linear functional on $\mathcal{L}_1(\mu)$.

19.3.3. Corollary. *If f is a real valued integrable function on a positive measure space (S,μ), then*

$$\left|\int_S f\,d\mu\right| \le \int_S |f|\,d\mu.$$

Proof. The result is follows immediately from propositions 19.3.2 and 14.3.13(b). □

19.3.4. Proposition. *If f is a complex valued integrable function on a positive measure space (S,μ), then*

$$\left|\int_S f\,d\mu\right| \le \int_S |f|\,d\mu.$$

19.3.5. Proposition. *If (S,μ) is a positive measure space then the map $f \mapsto \int_S |f|\,d\mu$ is a seminorm on the vector space $\mathcal{L}_1(\mu)$.*

19.3.6. Proposition. *Let (S,μ) be a positive measure space and f, $h \in \mathcal{L}_1(\mu)$. If g is a measurable function on S such that $f \le g \le h$ a.e., then $g \in \mathcal{L}_1(\mu)$.*

19.3.7. Theorem (Lebesgue Dominated Convergence Theorem)**.** *Let (S,μ) be a positive measure space and (f_n) be a sequence on measurable complex valued functions on S. If there exists an integrable function g such that $|f_n| \le g$ for every $n \in \mathbb{N}$ and $h(x) = \lim_n f_n(x)$ exists for every $x \in S$, then $h \in \mathcal{L}_1(\mu,\mathbb{C})$ and*

$$\int h\,d\mu = \lim_{n\to\infty} \int_S f_n\,d\mu.$$

Hint for proof. Apply *Fatou's lemma* to the sequences $(g + f_n)$ and $(g - f_n)$.

19.3.8. Proposition. *If f and g are measurable functions on a positive measure space (S, μ) such that $\int |f - g|\, d\mu = 0$, then $f = g$ a.e.*

19.3.9. Example. Let μ be counting measure on $\mathbb{N}$ and let $a_n = (-1)^n/n$ for each $n \in \mathbb{N}$. Then the infinite series $\sum a_n$ converges (by the *alternating series test*) but, as a function on $\mathbb{N}$, is not integrable with respect to μ.

19.3.10. Example. The function $x \mapsto \dfrac{1}{x}$ is not integrable (with respect to Lebesgue measure) on the interval $(0, 1]$.

19.3.11. Example. The function $x \mapsto \dfrac{1}{\sqrt{x}}$ is integrable (with respect to Lebesgue measure) on the interval $(0, 1]$.

Hint for proof. For each $n \in \mathbb{N}$ let

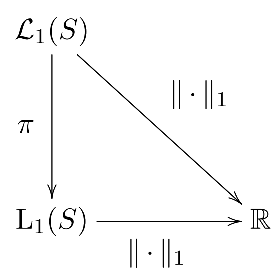
$$\tau_n = (n+1)\ \chi_{\left(0\,,\,n^{-2}\right]} + \sum_{k=1}^{n-1} (k+1)\ \chi_{\left((k+1)^{-2}\,,\,k^{-2}\right]}\,.$$

19.3.12. Proposition. *Let f be an integrable function on a positive measure space (S, μ) and (E_n) be a sequence of pairwise disjoint measurable subsets of S whose union is F. Then*

$$\int_F f\, d\mu = \sum_{k=1}^{\infty} \int_{E_k} f\, d\mu\,.$$

19.3.13. Proposition. *If (S, μ) is a positive measure space and f is an integrable function on S, then $\int_A f\, d\mu = 0$ for every measurable subset A of S if and only if $f = 0$ a.e. on S.*

19.3.14. Proposition. *Let (S, μ) be a positive measure space and $\mathcal{N}(S)$ be the family of all measurable functions f on S such that $f(x) = 0$ for almost all $x \in S$. Define $\mathrm{L}_1(S) := \mathcal{L}_1(S)/\mathcal{N}(S)$. An alternative notation for $\mathrm{L}_1(S)$ is $\mathrm{L}_1(\mu)$. Then there exists a unique norm on $\mathrm{L}_1(S)$ that makes the following diagram commute.*



19.3.15. Proposition. *On a positive measure space (S, μ) let $\mathrm{L}_1(S)$ be defined as in proposition 19.3.14. Then $\mathrm{L}_1(S)$ is a Riesz space.*

19.3.16. Proposition. *If S is a positive measure space, then the set of all integrable simple functions is dense in the normed linear space $\mathrm{L}_1(S)$.*

19.3.17. Proposition. *Suppose that (f_n) is a sequence of measurable functions defined on a measurable subset E of a positive measure space (S, μ) such that $\sum_{k=1}^{\infty} |f_k|$ is an integrable function on E. Then the function $\sum_{k=1}^{\infty} f_k$ is integrable and*

$$\int_E \sum_{k=1}^{\infty} f_k\, d\mu = \sum_{k=1}^{\infty} \int_E f_k\, d\mu\,.$$

19.3.18. Proposition. *Let f be a bounded real valued function on the interval $[a,b]$ in $\mathbb{R}$. If f is Riemann integrable, then it is (Lebesgue measurable and) Lebesgue integrable and its Lebesgue integral is the same as its Riemann integral; that is,*

$$\int_{[a,b]} f\,d\lambda = \int_a^b f(x)\,dx\,.$$

Proof. See [2], theorem 19.6; [4], theorem 1.7.1(b); [14], theorem on page 99; [18], theorem 2.28(a); or [32], theorem 3.23.

19.3.19. Example. Let $f(x) = \dfrac{\sin x}{x}$ for $x > 0$ and $f(0) = 1$. Then f has a finite improper Riemann integral but (while it is Lebesgue measurable) it is not Lebesgue integrable.

Proof. The somewhat complicated business of computing the integral of the function f can be found in [2], theorem 20.7, case II, where it is shown that $\int_0^\infty \dfrac{\sin x}{x}\,dx = \dfrac{\pi}{2}$.

19.3.20. Proposition. *Let f be a bounded real valued function on the interval $[a,b]$ in $\mathbb{R}$. Then f is Riemann integrable if and only if it is continuous almost everywhere (with respect to Lebesgue measure).*

Proof. See [2], theorem 19.7; [4], theorem 1.7.1(a); [14], theorem on page 99; or [18], theorem 2.28(b).

19.3.21. Exercise. Regard $S = (0,3]$ as a positive real measure space under Lebesgue measure λ. Define functions f and g on S by

$$f(x) = \sin \tfrac{\pi}{2}x \qquad \text{and} \qquad g(x) = \chi_{(0,1]} - \chi_{(1,2]} + 2\chi_{(2,3]}\,.$$

Define real measures μ and ν by

$$\nu(A) = \int_A f\,d\lambda \qquad \text{and} \qquad \mu(A) = \int_A g\,d\lambda$$

for every $A \overset{m}{\subseteq} S$.

(a) Find a Hahn decomposition of (S,μ).
(b) Find the Jordan decomposition of μ. (Express your answer in terms of Lebesgue measure.)
(c) Find $|\mu|(A)$ for $A \overset{m}{\subseteq} S$.
(d) Find $(\mu \vee \nu)\big((0,2]\big)$.

CHAPTER 20

COMPLETE METRIC SPACES

20.1. Cauchy Sequences

20.1.1. Definition. A sequence (x_n) in a metric space is a CAUCHY SEQUENCE if for every $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that $d(x_m, x_n) < \epsilon$ whenever $m, n \geq n_0$. This condition is often abbreviated as follows: $d(x_m, x_n) \to 0$ as $m, n \to \infty$ (or $\lim_{m,n\to\infty} d(x_m, x_n) = 0$).

20.1.2. Proposition. *In a metric space every convergent sequence is Cauchy.*

20.1.3. Proposition. *Every Cauchy sequence which has a convergent subsequence is itself convergent (and to the same limit as the subsequence).*

20.1.4. Proposition. *In a metric space every Cauchy sequence is bounded.*

20.1.5. Definition. A metric space M is COMPLETE if every Cauchy sequence in M converges.

Completeness is not a topological property of a metric space, as the next example shows.

20.1.6. Example. Consider the interval $J = (0, 1]$ under the usual metric ($d(x, y) = |x - y|$) and under the metric d' defined by $d'(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right|$. This produces two metric spaces with identical topologies; but one is complete and the other is not.

20.1.7. Definition. The DIAMETER of a subset A of a metric space is defined by

$$\operatorname{diam} A := \sup\{d(x, y) \colon x, y \in A\}$$

if this supremum exists. Otherwise $\operatorname{diam} A := \infty$.

20.1.8. Proposition. *If A is a subset of a metric space, then* $\operatorname{diam} \overline{A} = \operatorname{diam} A$.

20.1.9. Proposition. *In a metric space M the following are equivalent:*

(a) *M is complete.*
(b) *Every decreasing sequence of nonempty closed sets in M whose diameters approach* 0 *has nonempty intersection.*

Hint for proof. To show that (a) implies (b), let (F_k) be a decreasing sequence of nonempty closed subsets of M. For each k choose $x_k \in F_k$. Show that the sequence (x_k) is Cauchy. To show that (b) implies (a), let (x_k) be a Cauchy sequence in M. Define $A_n = \{x_k \colon k \geq n\}$ and $F_n = \overline{A_n}$. Show that (F_n) is a decreasing sequence of closed sets whose diameters approach 0. Choose a point a in $\cap F_n$. Find a subsequence (x_{n_k}) of (x_n) such that $d\left(a, x_{n_k}\right) < 2^{-k}$.

20.1.10. Example. Since $\mathbb{R}$ is complete, the preceding proposition tells us that every nested sequence of nonempty closed subsets of $\mathbb{R}$ whose diameters approach 0 has nonempty intersection.

(a) This statement is no longer correct if the words "whose diameters approach 0" are deleted.
(b) The statement is no longer correct if the word "closed" is deleted.

20.1.11. Proposition. *Every complete subspace of a metric space is closed and every closed subset of a complete metric space is complete.*

20.1.12. Proposition. *Every compact metric space is complete.*

20.1.13. Proposition. *The product of two complete metric spaces is complete.*

20.1.14. Proposition. *If d and ρ are strongly equivalent metrics on a set M, then (M,d) is complete if and only if (M,ρ) is.*

20.1.15. Example. The conclusion of the preceding proposition may not hold if the two metrics are merely equivalent.

20.1.16. Definition. A metric space (M,d) is TOTALLY BOUNDED if for every $\epsilon > 0$ there exists a finite subset F of M such that for every $a \in M$ there is a point $x \in F$ such that $d(x,a) < \epsilon$. The finite set F is frequently called an ϵ-NET for M. This definition has a more or less standard paraphrase: A space is totally bounded if it can be kept under surveillance by a finite number of arbitrarily near-sighted policemen.

The following equivalent formulations of total boundedness appear in [26].

20.1.17. Proposition. *For a metric space (M,d) the following are equivalent:*

(a) *Every sequence in M has a Cauchy subsequence.*
(b) *The space M is totally bounded.*
(c) *For every $\epsilon > 0$ the space M can be written as a union of a finite collection of sets each with diameter strictly less than ϵ.*
(d) *Suppose $\epsilon > 0$. Then every infinite subset of M contains an infinite set of diameter strictly less than ϵ.*

Hint for proof. To prove (a) implies (b), let $\epsilon > 0$ and $\mathfrak{F}$ be the family of all subsets F of M with the property that $d(x,y) \geq \epsilon$ for every pair of distinct points x and y in F. Partially order $\mathfrak{F}$ by inclusion. Use *Zorn's lemma*.

To prove (d) implies (a) let (x_k) be a sequence in M. Without loss of generality suppose $x_j \neq x_k$ whenever $j \neq k$. Choose a subsequence $(x_{1,k})$ of (x_k) such that the distance between any two elements in its range is strictly less than 1. Next choose a subsequence $(x_{2,k})$ of $(x_{1,k})$ such that the distance between any two elements in its range is strictly less than $\frac{1}{2}$. Next choose a subsequence $(x_{3,k})$ of $(x_{2,k})$ such that the distance between any two elements in its range is strictly less than $\frac{1}{3}$. Continue inductively. What can you say about the subsequence $(x_{k,k})$?

20.1.18. Proposition. *A metric space is compact if and only if it is complete and totally bounded.*

We extend the notion of an infinite series from the real numbers to an arbitrary normed linear space. The language is exactly the same.

20.1.19. Definition. Let (x_n) be a sequence in a normed linear space V. The sum $s_n = \sum_{k=1}^{n} x_k$ of the first n terms of the sequence is the n^{TH}-PARTIAL SUM of the sequence. The sequence (s_n) of partial sums is an INFINITE SERIES and is denoted by $\sum_{k=1}^{\infty} x_k$. The series is said to CONVERGE if the sequence (s_n) of partial sums converges and, if the series converges, its *sum* is the limit of the sequence (s_n) of partial sums. The notation for the sum of the series is, confusingly, the same as for the series itself $\sum_{k=1}^{\infty} x_k$. Notice that saying that the series $\sum_{k=1}^{\infty} x_k$ converges is exactly the same as saying that the sequence (x_n) is summable (see definition 12.1.29).

Similarly, the assertion that the series $\sum_{k=1}^{\infty} x_k$ converges absolutely is the same as claiming that the sequence (x_n) is absolutely summable. That is, the series $\sum_{k=1}^{\infty} x_k$ is ABSOLUTELY CONVERGENT if $\sum_{k=1}^{\infty} \|x_k\| < \infty$.

20.1.20. Proposition. *A normed linear space V is complete if and only if every absolutely summable sequence in V is summable.*

20.1.21. Proposition. *If M is a closed subspace of a complete normed linear space V, then V/M is complete.*

20.1.22. Example. The normed algebra $\mathcal{B}(S)$ (see example 12.1.9) is complete.

20.1.23. Example. If X is a topological space then the normed algebra $\mathcal{C}_b(X)$ (see example 12.3.17) is complete. A special case of this occurs when X is a compact topological space. Then $\mathcal{C}(X)$ is a complete normed algebra.

20.2. Completions and Universal Morphisms

20.2.1. Definition. Let M and N be metric spaces. We say that N is a COMPLETION of M if N is complete and M is isometric to a dense subset of N.

20.2.2. Example. If in the definition 20.2.1 of *completion* we were to substitute the word "homeomorphic" for "isometric" then the uniqueness result 20.2.11 would no longer be true. In fact, it would then be possible to have two "completions" of a metric space that are not even homeomorphic.

In the construction of the real numbers from the rationals one approach is to treat the set of rationals as a metric space and define the set of reals to be its completion. One standard way of producing the completion involves equivalence classes of Cauchy sequences of rational numbers. This technique extends to metric spaces: an elementary way of completing an arbitrary metric space starts with defining an equivalence relation on the set of Cauchy sequences of elements of the space. Here is a slightly less elementary but considerably simpler approach.

20.2.3. Proposition. *Every metric space M is isometric to subspace of $\mathcal{B}(M)$.*

Hint for proof. If (M, d) is a metric space fix $a \in M$. For each $x \in M$ define $\phi_x \colon M \to \mathbb{R}$ by $\phi_x(u) = d(x, u) - d(u, a)$. Show first that $\phi_x \in \mathcal{B}(M)$ for every $x \in M$. Then show that $\phi \colon M \to \mathcal{B}(M)$ is an isometry.

20.2.4. Corollary. *Every metric space has a completion.*

Much of mathematics involves the construction of new objects from old ones—things such as products, coproducts, quotients, completions, compactifications, and unitizations. More often than not it is possible to characterize such a construction by the existence of a unique morphism having some particular property. Because this morphism and its corresponding property characterize the construction in question, they are referred to as a *universal morphism* and a *universal property*, respectively. Here is one very common way in which such morphisms arise.

20.2.5. Definition. Let $\mathbf{A} \xrightarrow{\mathsf{F}} \mathbf{B}$ be a functor between categories $\mathbf{A}$ and $\mathbf{B}$ and B be an object in $\mathbf{B}$. A pair (A, u) with A an object in $\mathbf{A}$ and u a $\mathbf{B}$-morphism from B to $\mathsf{F}(A)$ is a UNIVERSAL MORPHISM for B (with respect to the functor F) if for every object A' in $\mathbf{A}$ and every $\mathbf{B}$-morphism $B \xrightarrow{f} \mathsf{F}(A')$ there exists a unique $\mathbf{A}$-morphism $A \xrightarrow{\tilde{f}} A'$ such that the following diagram commutes.

$$\begin{array}{ccccc} B & \xrightarrow{u} & \mathsf{F}(A) & & A \\ & {\scriptstyle f}\searrow & \Big\downarrow {\scriptstyle \mathsf{F}(\tilde{f})} & & \Big\downarrow {\scriptstyle \tilde{f}} \\ & & \mathsf{F}(A') & & A' \end{array} \tag{20.1}$$

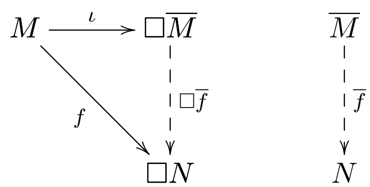
In this context the object A is often referred to as a UNIVERSAL OBJECT in $\mathbf{A}$.

20.2.6. Definition. As mentioned earlier the categories of interest in these notes are concrete categories (see 8.1.6). If A is an object in some concrete category $\mathbf{C}$, we denote by $\Box A$ its underlying set. And if $A \xrightarrow{f} B$ is a morphism in $\mathbf{C}$ we denote by $\Box f$ the map from $\Box A$ to $\Box B$ regarded simply as a function between sets. It is easy to see that $\Box$, which takes objects in $\mathbf{C}$ to objects in $\mathbf{Set}$ (the category of sets and maps) and morphisms in $\mathbf{C}$ to morphisms in $\mathbf{Set}$, is a functor. It is referred to as a FORGETFUL FUNCTOR. In the category $\mathbf{Vec}$ of vector spaces and linear maps, for example, $\Box$ causes a vector space V to "forget" about its addition and scalar multiplication ($\Box V$ is just a set). And if $T \colon V \to W$ is a linear transformation, then $\Box T \colon \Box V \to \Box W$ is just a map between sets—it has "forgotten" about preserving the operations.

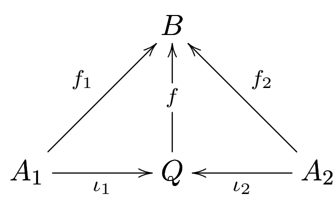
Occasionally it is useful to allow a forgetful functor to forget some of the structure of a category but not all. For example, $\Box$ might take the category $\mathbf{ALG}$ to the category $\mathbf{VEC}$ by forgetting about just multiplication or, perhaps, it might take the category **MET** of metric spaces and continuous maps to the category **TOP** by forgetting about the metrics but not the topologies they induce.

The next proposition shows that the completion of a metric space is universal. In this result the categories of interest are the category of metric spaces and uniformly continuous maps and its subcategory consisting of complete metric spaces and uniformly continuous maps. Here the forgetful functor $\Box$ just forgets about completeness.

20.2.7. Proposition. *Let M be a metric space and $\overline{M}$ be its completion. If N is a complete metric space and $f\colon M \to N$ is uniformly continuous, then there exists a unique uniformly continuous function $\overline{f}\colon \overline{M} \to N$ which makes the following diagram commute.*

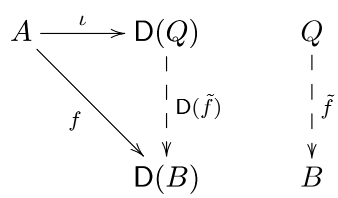


20.2.8. Example. Here is the usual presentation of the coproduct of two objects in a category. Let A_1 and A_2 be two objects in a category **C**. A COPRODUCT of A_1 and A_2 is a triple (Q, ι_1, ι_2) with Q an object in **C** and $A_k \xrightarrow{\iota_k} Q$ $(k = 1, 2)$ morphisms in **C** which satisfies the following condition: if B is an arbitrary object in **C** and $A_k \xrightarrow{f_k} B$ $(k = 1, 2)$ are arbitrary morphisms in **C**, then there exists a unique **C**-morphism $Q \xrightarrow{f} B$ which makes the following diagram commute.



(20.2)

It may not be obvious at first glance that this construction is *universal* in the sense of definition 20.2.5. To see that it in fact is, let D be the diagonal functor from a category **C** to the category of pairs $\mathbf{C^2}$ (see example 15.4.7). Suppose that (Q, ι_1, ι_2) is a coproduct of the **C**-objects A_1 and A_2 in the sense defined above. Then $A = (A_1, A_2)$ is an object in $\mathbf{C^2}$, $A \xrightarrow{\iota} \mathsf{D}(Q)$ is a $\mathbf{C^2}$-morphism, and the pair (Q, ι) is universal in the sense of 20.2.5. The diagram corresponding to diagram (20.1) is

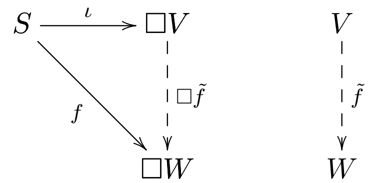


(20.3)

where B is an arbitrary object in **C** and (for $k = 1, 2$) $A_k \xrightarrow{f_k} B$ are arbitrary **C**-morphisms so that $f = (f_1, f_2)$ is a $\mathbf{C^2}$-morphism.

20.2.9. Example. Let S be a set and $\mathbf{VEC} \xrightarrow{\Box} \mathbf{SET}$ be the forgetful functor from the category **VEC** to the category **SET**. If there exists a vector space V and an injection $S \xrightarrow{\iota} \Box V$ which constitute a universal morphism for S (with respect to $\Box$), then V is the FREE VECTOR SPACE over S. Of course merely *defining* an object does not guarantee its existence. In fact, free vector spaces exist over arbitrary sets. Given the set S let V be the set of all real valued functions on S which have finite support. Define addition and scalar multiplication pointwise. The map

$\iota\colon s \mapsto \chi_{\{s\}}$ of each element $s \in S$ to the characteristic function of $\{s\}$ is the desired injection. To verify that V is free over S it must be shown that for every vector space W and every function $S \xrightarrow{\ f\ } \Box W$ there exists a unique linear map $V \xrightarrow{\ \widetilde{f}\ } W$ which makes the following diagram commute.



20.2.10. Proposition. *Universal objects in a category are essentially unique.*

The following consequence of propositions 20.2.7 and 20.2.10 allows us to speak of *the* completion of a metric space.

20.2.11. Corollary. *Metric space completions are unique (up to isometry).*

20.3. Compact Subsets of $\mathcal{C}(X)$

20.3.1. Definition. Let X be a topological space, (N, ρ) be a metric space, And $\mathcal{F}$ be a nonempty subset of $\mathcal{C}(X, N)$. The family $\mathcal{F}$ is EQUICONTINUOUS AT a point $a \in X$ if for every $\epsilon > 0$ there exists a neighborhood U of a such that $\rho\big(f(x), f(a)\big) < \epsilon$ for every $f \in \mathcal{F}$ and every $x \in U$. The family $\mathcal{F}$ is EQUICONTINUOUS ON a set $A \subseteq X$ if it is equicontinuous at every point in A.

If (M, d) and (N, ρ) are metric spaces and $\mathcal{F}$ is a nonempty subset of $\mathcal{C}(M, N)$ we say that the family $\mathcal{F}$ is UNIFORMLY EQUICONTINUOUS on M if for every $\epsilon > 0$ there exists $\delta > 0$ such that $\rho\big(f(x), f(y)\big) < \epsilon$ whenever $f \in \mathcal{F}$ and $d(x, y) < \delta$. (Where we use the expression *uniformly equicontinuous* many authors use just *equicontinuous*. We show below in proposition 20.3.4 that, at least on compact metric spaces, the two notions coincide.)

Careful and readable discussions of equicontinuity can be found in [10], pages 178–183; [20], pages 365–374; and [43], pages 164–169. You can find some instructive and interesting exercises on equicontinuity in [6], (4.5.7) and (4.5.9); [10], pages 179–180 and 182–183; [20], pages 372–374, and [43], pages 168–169.

20.3.2. Example. For each $n \in \mathbb{N}$ let $f_n\colon [0, 1] \to \mathbb{R}$ be defined by $f_n(x) = x^n$. The family $\{f_n\colon n \in \mathbb{N}\}$ is not equicontinuous at $x = 1$.

20.3.3. Example. The family $\{f_n\colon n \in \mathbb{N}\}$ of functions defined in the preceding exercise is equicontinuous at $x = \frac{99}{100}$.

20.3.4. Proposition. *A nonempty family $\mathcal{F}$ of continuous functions from a compact metric space (M, d) into a metric space (N, ρ) is equicontinuous on M if and only if it is uniformly equicontinuous on M.*

20.3.5. Definition. Let S be a nonempty set and M be a metric space. A family $\mathcal{F}$ of functions $f\colon S \to M$ is POINTWISE BOUNDED if $\{f(x)\colon f \in \mathcal{F}\}$ is a bounded subset of M for each $x \in S$. It is common to say that the family $\mathcal{F}$ is UNIFORMLY BOUNDED if it is a bounded subset of the metric space $\mathcal{B}(S, M)$.

Two things should be clear:

(1) uniform boundedness implies pointwise boundedness; and
(2) the word “uniformly” in *uniformly bounded* is, strictly speaking, redundant—it is there for emphasis.

20.3.6. Proposition. *Let $\mathcal{F}$ be an equicontinuous family of functions in $\mathcal{C}(X)$ where X is a compact topological space. If $\mathcal{F}$ is pointwise bounded, then it is (uniformly) bounded.*

20.3.7. Proposition. *A family $\mathcal{F}$ of functions in $\mathcal{C}(X)$, where X is a compact topological space, is totally bounded if and only if it is equicontinuous and pointwise bounded.*

PROOF. See [20], proposition 5.1.10.

20.3.8. Theorem (Ascoli-Arzelà theorem)**.** *Let X be a compact topological space. A family $\mathcal{F} \subseteq \mathcal{C}(X)$ is compact if and only if it is closed, pointwise bounded, and equicontinuous.*

The preceding theorem also goes under the names *Ascoli's theorem* and the *Arzelà-Ascoli theorem.*

A subset A of a topological space X is said to be RELATIVELY COMPACT if its closure is compact. Thus the Ascoli-Arzelà theorem may be given a slightly shorter form: *For a compact space X a subset $\mathcal{F} \subseteq \mathcal{C}(X)$ is relatively compact if and only if it is bounded and equicontinuous.*

20.4. Banach Spaces; $\mathrm{L_p}$-spaces

20.4.1. Definition. A BANACH SPACE is a complete normed linear space. We denote the category of Banach spaces and bounded linear maps by **BAN**. A BANACH ALGEBRA is a complete normed algebra.

20.4.2. Example. Under its usual operations $\mathbb{R}$ is a unital commutative Banach algebra.

20.4.3. Example. Let S be a nonempty set. Under the usual pointwise operations $\mathcal{B}(S)$ is a unital commutative Banach algebra. (See example 20.1.22.)

20.4.4. Example. Let X be a nonempty compact Hausdorff space. Under the usual pointwise operations $\mathcal{C}(X)$ is a unital commutative Banach algebra. (See example 20.1.23.) By the *extreme value theorem* 17.1.18 it is a unital (Banach) subalgebra of $\mathcal{B}(X)$.

20.4.5. Example. Let X be a nonempty topological space. Under the usual pointwise operations $\mathcal{C}_b(X)$ is a unital commutative Banach algebra. (See example 20.1.23.)

20.4.6. Example. Let X be a nonempty locally compact Hausdorff space. A complex valued function f on X is said to VANISH AT INFINITY if for every $\epsilon > 0$ there exists a compact set outside of which $|f(x)| < \epsilon$. The family of all continuous complex valued functions on X which vanish at infinity is denoted by $\mathcal{C}_0(X)$. Under the usual pointwise operations and the uniform norm $\mathcal{C}_0(X)$ is a (in general nonunital) commutative Banach algebra.

20.4.7. Example. Let V and W be a normed linear spaces. Then $\mathfrak{B}(V, W)$ is a Banach space if and only if W is complete. In particular the dual V^* of any normed linear space V is a Banach space. Furthermore, if V is complete, then the family $\mathfrak{B}(V)$ of operators on V is a unital (in general noncommutative) Banach algebra.

20.4.8. Example. The set M_n of $n \times n$ matrices of real numbers is a unital (in general noncommutative) Banach algebra. (See example 4.6.10.)

20.4.9. Example. If (S, μ) is a positive measure space, then (under pointwise multiplication of functions) $\mathrm{L}_\infty(S)$ is a unital commutative Banach algebra.

20.4.10. Example. If (S, μ) is a positive measure space, then $\mathrm{L}_1(S)$ is a Banach space. (See proposition 19.3.14.)

20.4.11. Definition. Suppose that (S, μ) is a positive measure space and $1 < p < \infty$. Let $\mathcal{L}_p(S)$ be the set of all measurable functions f on S such that $\int_S |f|^p \, d\mu < \infty$. The mapping

$$f \mapsto \|f\|_p := \left(\int_S |f|^p \, d\mu \right)^{1/p}$$

is a seminorm on $\mathcal{L}_p(S)$. Identifying functions that differ only on a set of measure zero produces a normed linear space $\mathrm{L_p}(S)$ (or $\mathrm{L_p}(\mu)$) in exactly the same way as the space $\mathrm{L}_1(S)$ was created in proposition 19.3.14. The resulting norm, also denoted by $\|\cdot\|_p$, is the LP NORM on $\mathrm{L_p}(S)$.

In these notes when S is a subset of $\mathbb{R}$ and no measure is specified, Lebesgue measure is understood; that is, $\mathrm{L_p}(S)$ is taken to mean $\mathrm{L_p}(S,\lambda)$.

20.4.12. Definition. If $1 \le p, q \le \infty$ we say that p and q are CONJUGATE EXPONENTS if $\dfrac{1}{p}+\dfrac{1}{q}=1$. In particular, we take 1 and ∞ to be conjugate exponents.

20.4.13. Proposition (Hölder's inequality)**.** *Let (S,μ) be a positive measure space and p and q be conjugate exponents where $1 \le p \le \infty$. If $f \in \mathrm{L_p}(S)$ and $g \in \mathrm{L_q}(S)$, then $fg \in \mathrm{L_1}(S)$ and*

$$\|fg\|_1 \le \|f\|_p \, \|g\|_q .$$

PROOF. See [2], theorem 25.3; [4], 2.4.5; [18] (6.2); [32], theorem 9.9; or [37], theorems 3.5 and 3.8.

20.4.14. Example. The space $\mathrm{L_2}([0,1])$ is contained in $\mathrm{L_1}([0,1])$.

The preceding example easily generalizes to all spaces of finite measure but, as the next example shows, not to spaces of infinite measure.

20.4.15. Example. The function $x \mapsto \frac{1}{x}$ belongs to $\mathrm{L_2}([1,\infty))$ but not to $\mathrm{L_1}([1,\infty))$.

20.4.16. Proposition (Minkowski's inequality)**.** *Let (S,μ) be a positive measure space and $1 \le p \le \infty$. If $f \in \mathrm{L_p}(S)$,then*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p .$$

PROOF. See [2], theorem 25.4; [4], 2.4.7; [18] (6.5); [32], theorem 9.10; or [37], theorems 3.5 and 3.9.

20.4.17. Example. Let (S,μ) be a positive measure space and $1 < p < \infty$. Then $\mathrm{L_p}(S)$ is a Banach space.

The next result generalizes proposition 19.3.16.

20.4.18. Proposition. *Let (S,μ) be a positive measure space and $1 \le p < \infty$. Then the set of all integrable simple functions is dense in $\mathrm{L_p}(S)$.*

The notation $\|\cdot\|_p$ for the norm on the space $\mathrm{L_p}(S)$ seems natural enough. But some people find the notation $\|\cdot\|_\infty$ for the essential supremum norm a trifle odd. Perhaps the next result will make it seem more plausible.

20.4.19. Proposition. *Let E be a Lebesgue measurable subset of $\mathbb{R}^n$ with nonzero finite Lebesgue measure and let f be a real (or extended real) valued function on E. Then as p runs over the natural numbers we have*

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty .$$

Hint for proof. Show that

(1) $\limsup_{p\to\infty} \|f\|_p \le \|f\|_\infty$ and
(2) $\liminf_{p\to\infty} \|f\|_p \ge M$ for every M such that $0 < M < \|f\|_\infty$.

Two other important normed linear spaces, which we have already encountered, are complete: $\mathrm{ba}(S)$ and $\mathrm{ca}(S)$. (See example 16.5.19.)

20.4.20. Example. If $\mathfrak{A}$ is an algebra of subsets of a set S, then under the norm defined in example 16.5.19 the space $\mathrm{ba}(S)$ is a Banach space. If $\mathfrak{A}$ is a σ-algebra of subsets of a set S, then $\mathrm{ca}(S)$ is also a Banach space.

PROOF. See [1], theorems 9.52 and 9.55.

It is also pleasant to discover that on a normed linear space, switching from the given norm to an equivalent one does not affect completeness of the space.

20.4.21. Proposition. *Suppose that $\|\cdot\|$ and $|||\cdot|||$ are equivalent norms on a vector space V. If V is a Banach space under $\|\cdot\|$, then it is also a Banach space under $|||\cdot|||$.*

20.5. Banach Algebras

20.5.1. Definition. A map $f\colon A \to B$ between Banach algebras is a (Banach algebra) homomorphism if it is both an algebraic homomorphism and a continuous map between A and B. We denote by **BALG** the category of Banach algebras and continuous algebra homomorphisms.

20.5.2. Proposition. *Let A and B be Banach algebras. The* direct sum *of A and B, denoted by $A \oplus B$, is defined to be the set $A \times B$ equipped with pointwise operations:*

(i) $(a, b) + (a', b') = (a + a', b + b')$;
(ii) $(a, b)\,(a', b') = (aa', bb')$;
(iii) $\alpha(a, b) = (\alpha a, \alpha b)$

for a,$a' \in A$, b, $b' \in B$, and $\alpha \in \mathbb{C}$. As in 12.3.2 define $\|(a, b)\| = \|a\| + \|b\|$. This makes $A \oplus B$ into a Banach algebra.

20.5.3. Corollary. *If A and B are Banach spaces, then so is their direct sum $A \oplus B$ (with addition, scalar multiplication, and norm defined as in the preceding proposition).*

20.5.4. Exercise. Identify products and coproducts (if they exist) in the category **BALG** of Banach algebras and continuous algebra homomorphisms and in the category of Banach algebras and contractive algebra homomorphisms. (Here *contractive* means $\|Tx\| \le \|x\|$ for all x.)

20.5.5. Proposition. *In a Banach algebra the operations of addition and multiplication (regarded as maps from $A \oplus A$ to A) are continuous and scalar multiplication (regarded as a map from $\mathbb{R} \oplus A$ to A) is continuous.*

20.5.6. Proposition. *Let a be an element of a unital Banach algebra A. If $\|a\| < 1$, then $\mathbf{1} - a \in \operatorname{inv} A$ and $(\mathbf{1} - a)^{-1} = \sum_{k=0}^{\infty} a^k$.*

Hint for proof. Show that the sequence $(\mathbf{1}, a, a^2, \dots)$ is summable. (In a unital algebra we take a^0 to mean $\mathbf{1}$.)

20.5.7. Proposition. *If A is a unital Banach algebra, then $\operatorname{inv} A \overset{\circ}{\subseteq} A$.*

Hint for proof. Let $a \in \operatorname{inv} A$. Show, for sufficiently small h, that $\mathbf{1} - a^{-1}h$ is invertible.

20.5.8. Proposition. *If a belongs to a unital Banach algebra and $\|a\| < 1$, then*

$$\|(\mathbf{1} - a)^{-1} - \mathbf{1}\| \le \frac{\|a\|}{1 - \|a\|}\,.$$

20.5.9. Proposition. *Let A be a unital Banach algebra. The map $a \mapsto a^{-1}$ from $\operatorname{inv} A$ into itself is continuous.*

20.5.10. Proposition. *Let A be a unital Banach algebra. The map $r\colon a \mapsto a^{-1}$ from $\operatorname{inv} A$ into itself is differentiable and at each invertible element a, we have $dr_a(h) = -a^{-1}ha^{-1}$ for all $h \in A$.*

20.5.11. Notation. Let f be a complex valued function on some set S. Denote by Z_f the set of all points x in S such that $f(x) = 0$. This is the zero set of f.

20.5.12. Example. The invertible elements in the Banach algebra $\mathcal{C}(X)$ of all continuous real or complex valued functions on a compact Hausdorff space X are the functions which vanish nowhere. That is,

$$\operatorname{inv} \mathcal{C}(X) = \{f \in \mathcal{C}(X) \colon Z_f = \emptyset\}\,.$$

This is not true in general of either $\mathcal{C}_b(X)$ or $\mathcal{C}_0(X)$ when X fails to be compact.

20.6. Hilbert spaces

We now make an abrupt change of convention. Heretofore we have taken vector spaces (with or without extra structure) to have real scalars unless the contrary is stated. In the study of Hilbert spaces however it is convenient to let the default case be complex scalars.

20.6.1. Definition. Let V be a complex vector space. A function which associates to each pair of vectors x and y in V a complex number $\langle x, y\rangle$ is an INNER PRODUCT (or a DOT PRODUCT) on V provided that the following four conditions are satisfied:

(a) If x, y, $z \in V$, then

$$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle.$$

(b) If x, $y \in V$, then

$$\langle \alpha x, y\rangle = \alpha\langle x, y\rangle.$$

(c) If x, $y \in V$, then

$$\langle x, y\rangle = \overline{\langle y, x\rangle}.$$

(d) For every nonzero x in V we have $\langle x, x\rangle > 0$.

Items (a)–(d) say that the inner product is a *positive definite conjugate symmetric sesquilinear form.* When a vector space has been equipped with an inner product we define the NORM (or LENGTH) of a vector x by

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

A HILBERT SPACE is a complete inner product space. (Here, of course, we intend completeness of the metric space induced by the norm induced by the inner product.)

20.6.2. Proposition. *If x, y, and z are vectors in an inner product space and $\alpha \in \mathbb{C}$, then*

(a) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$,
(b) $\langle x, \alpha y\rangle = \overline{\alpha}\langle x, y\rangle$, *and*
(c) $\langle x, x\rangle = 0$ *if and only if* $x = 0$.

20.6.3. Example. For vectors $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ belonging to $\mathbb{C}^n$ define

$$\langle x, y\rangle = \sum_{k=1}^{n} x_k\overline{y_k}\,.$$

Then $\mathbb{C}^n$ is an inner product space.

20.6.4. Example. Let l_2 be the set of all square summable sequences of complex numbers. (A sequence $x = (x_k)_{k=1}^{\infty}$ is SQUARE SUMMABLE if $\sum_{k=1}^{\infty}|x_k|^2 < \infty$.) For vectors $x = (x_1, x_2, \dots)$ and $y = (y_1, y_2, \dots)$ belonging to l_2 define

$$\langle x, y\rangle = \sum_{k=1}^{\infty} x_k\overline{y_k}\,.$$

Then l_2 is a Hilbert space. (It must be shown, among other things, that the series in the preceding definition actually converges.)

20.6.5. Example. For $a < b$ let $\mathcal{C}([a, b], \mathbb{C})$ be the family of all continuous complex valued functions on the interval $[a, b]$. For every f, $g \in \mathcal{C}([a, b])$ define

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\,dx.$$

Then $\mathcal{C}([a, b])$ is an inner product space.

20.6.6. Example. Let (S, μ) be a positive measure on space. A measurable function f is SQUARE INTEGRABLE if $\int_S |f(x)|^2 \, d\mu(x) < \infty$. As we have seen the family of (equivalence classes of) square integrable functions on S is denoted by $\mathrm{L}_2(S)$. For every $f,\, g \in \mathrm{L}_2(S)$ define

$$\langle f, g \rangle = \int_S f(x)\overline{g(x)} \, d\mu(x) \,.$$

This is an inner product under which $\mathrm{L}_2(S)$ becomes a Hilbert space.

20.6.7. Theorem. *In every inner product space the* Schwarz inequality

$$|\langle x, y \rangle| \le \|x\| \, \|y\|.$$

holds for all vectors x and y.

20.6.8. Proposition. *If (x_n) is a sequence in an inner product space V which converges to a vector $a \in V$, then $\langle x_n, y \rangle \to \langle a, y \rangle$ for every $y \in V$.*

Every inner product space is a normed linear space.

20.6.9. Proposition. *Let V be an inner product space. For every $x \in V$ define*

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

Then the map $x \mapsto \|x\|$ is a norm on V.

20.6.10. Definition. Vectors x and y in an inner product space V are ORTHOGONAL (or PERPENDICULAR) if $\langle x, y \rangle = 0$. In this case we write $x \perp y$. Subsets A and B of V are ORTHOGONAL if $a \perp b$ for every $a \in A$ and $b \in B$. In this case we write $A \perp B$.

20.6.11. Definition. If M and N are subspaces of an inner product space V we use the notation $V = M \oplus N$ to indicate not only that V is the (vector space) direct sum of M and N but also that M and N are orthogonal. Thus we say that V is the (INTERNAL) ORTHOGONAL DIRECT SUM of M and N.

20.6.12. Proposition. *Let a be a vector in an inner product space V. Then $a \perp x$ for every $x \in V$ if and only if $a = 0$.*

20.6.13. Proposition (The Pythagorean theorem)**.** *If $x \perp y$ in an inner product space, then*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 \,.$$

20.6.14. Example. In $\mathbb{R}^2$ let M be the x-axis and L be the line whose equation is $y = x$. If we think of $\mathbb{R}^2$ as a (real) vector space, then it is correct to write $\mathbb{R}^2 = M \oplus L$. If, on the other hand, we regard $\mathbb{R}^2$ as a (real) inner product space, then $\mathbb{R}^2 \ne M \oplus L$ (because M and L are not perpendicular).

20.6.15. Proposition. *Let V be an inner product space. The inner product on V, regarded as a map from $V \oplus V$ into $\mathbb{C}$, is continuous. So is the norm, regarded as a map from V into $\mathbb{R}$.*

Concerning the proof of the preceding proposition, notice that the maps $(v, v') \mapsto \|v\| + \|v'\|$, $(v, v') \mapsto \sqrt{\|v\|^2 + \|v'\|^2}$, and $(v, v') \mapsto \max\{\|v\|, \|v'\|\}$ are all norms on $V \oplus V$. Which one is induced by the inner product on $V \oplus V$? Why does it not matter which one we use in proving that the inner product is continuous?

20.6.16. Proposition (The parallelogram law)**.** *If x and y are vectors in an inner product space, then*

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

20.6.17. Example. Consider the space $\mathcal{C}([0, 1])$ of continuous complex valued functions defined on $[0, 1]$. Under the UNIFORM NORM

$$\|f\|_u := \sup\{|f(x)| \colon 0 \le x \le 1\}$$

$\mathcal{C}([0, 1])$ is a normed linear space. There is no inner product on $\mathcal{C}([0, 1])$ which induces this norm.

Hint for proof. Use the preceding proposition.

20.6.18. Proposition (The polarization identity). *If x and y are vectors in an inner product space, then*

$$\langle x, y\rangle = \tfrac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + i\,\|x+iy\|^2 - i\,\|x-iy\|^2)\,.$$

20.6.19. Notation. Let V be an inner product space, $x \in V$, and A, $B \subseteq V$. If $x \perp a$ for every $a \in A$, we write $x \perp A$; and if $a \perp b$ for every $a \in A$ and $b \in B$, we write $A \perp B$. We define $A^\perp$, the ORTHOGONAL COMPLEMENT of A, to be $\{x \in V : x \perp A\}$. We write $A^{\perp\perp}$ for $\left(A^\perp\right)^\perp$.

20.6.20. Proposition. *If A is a subset of an inner product space V, then $A^\perp$ is a closed linear subspace of V.*

20.6.21. Theorem (Riesz-Fréchet Theorem). *If $f \in V^*$ where V is an inner product space, then there exists a unique vector a in V such that*

$$f(x) = \langle x, a\rangle$$

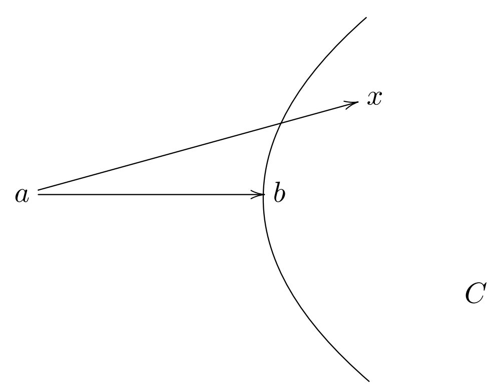
for all x in V.

20.6.22. Convention. In the context of Banach (and, in particular, Hilbert) spaces the word "subspace" will always mean *closed vector subspace*. To indicate that M is a subspace of a Banach space B we write $M \preccurlyeq B$. A (not necessarily closed) vector subspace of a Banach space is often called by other names such as *linear subspace* or *linear manifold*.

20.6.23. Definition. Let A be a nonempty subset of a Banach space B. We define the CLOSED LINEAR SPAN of A (denoted by $\bigvee A$) to be the intersection of the family of all subspaces of B which contain A. This is frequently referred to as *the smallest subspace of B containing A*.

20.6.24. Proposition. *The preceding definition makes sense. It is equivalent to defining $\bigvee A$ to be the closure of the (linear) span of A.*

20.6.25. Theorem (Minimizing Vector Theorem). *If C is a nonempty closed convex subset of a Hilbert space H and $a \in C^c$, then there exists a unique $b \in C$ such that $\|b-a\| \le \|x-a\|$ for every $x \in C$.*



20.6.26. Example. The vector space $\mathbb{R}^2$ under the uniform metric is a Banach space. To see that in this space the *minimizing vector theorem* does not hold take C to be the closed unit ball about the origin and a to be the point $(2, 0)$.

20.6.27. Example. The sets

$$C_1 = \left\{ f \in \mathcal{C}(\,[0,1]\,) : \int_0^{1/2} f - \int_{1/2}^1 f = 1 \right\} \quad \text{and} \quad C_2 = \left\{ f \in L_1(\,[0,1]\,) : \int_0^1 f = 1 \right\}$$

are examples that show that neither the existence nor the uniqueness claims of the *minimizing vector theorem* necessarily holds in a Banach space.

20.6.28. Theorem (Vector decomposition theorem)**.** *Let H be a Hilbert space, M be a subspace of H, and $x \in H$. Then there exist unique vectors $y \in M$ and $z \in M^\perp$ such that $x = y + z$.*

20.6.29. Proposition. *Let H be a Hilbert space. Then the following hold:*

(a) *if $M \subseteq H$, then $M \subseteq M^{\perp\perp}$;*
(b) *if $M \subseteq N \subseteq H$, then $N^\perp \subseteq M^\perp$;*
(c) *if M is a subspace of H, then $M = M^{\perp\perp}$; and*
(d) *if $M \subseteq H$, then $\bigvee M = M^{\perp\perp}$.*

20.6.30. Definition. Let V and W be inner product spaces. Recall (from 8.3.4) that if, for (v, w) and (v', w') in $V \times W$ and $\alpha \in \mathbb{C}$, we define

$$(v, w) + (v', w') = (v + v', w + w')$$

and

$$\alpha(v, w) = (\alpha v, \alpha w)$$

then $V \times W$ becomes a vector space, $V \oplus W$, which is the *(external) direct sum* of V and W. To make it into an inner product space define

$$\langle (v, w), (v', w') \rangle = \langle v, v' \rangle + \langle w, w' \rangle.$$

This makes the direct sum of V and W into an inner product space. It is the (EXTERNAL ORTHOGONAL) DIRECT SUM of V and W and is denoted by $V \oplus W$.

Notice that the same notation $\oplus$ is used for both internal and external direct sums and for both vector space direct sums (see definition 8.3.4) and orthogonal direct sums. So when we see the symbol $V \oplus W$ it is important to know which category we are in: vector spaces or inner product spaces, especially as it is common practice to omit the word "orthogonal" as a modifier to "direct sum" even in cases when it is intended. Notice also that, despite the fact that every inner product space "is" a normed linear space, the norm on the direct sum $V \oplus W$ of two inner product spaces is not the same as the norm on $V \oplus W$ when regarded as the product of two normed linear spaces (see 12.3.2).

20.6.31. Example. If H and K are Hilbert spaces so is their direct sum $H \oplus K$.

20.6.32. Definition. A mapping $T \colon V \to W$ between complex vector spaces is ANTILINEAR if $T(x + y) = Tx + Ty$ and $T(\alpha x) = \overline{\alpha}\, Tx$ for all x, $y \in V$ and $\alpha \in \mathbb{C}$. A bijective antilinear map is an ANTI-ISOMORPHISM.

20.6.33. Example. Every Hilbert space is isometrically anti-isomorphic to its dual space.

CHAPTER 21

ALGEBRAS AND LATTICES OF CONTINUOUS FUNCTIONS

21.1. Banach Lattices

21.1.1. Definition. A norm $\|\cdot\|$ on a Riesz space is a LATTICE NORM (or a RIESZ NORM) if $|x| \leq |y|$ implies $\|x\| \leq \|y\|$ for all x, $y \in E$. A Riesz space equipped with a lattice norm is a NORMED RIESZ SPACE. A complete normed Riesz space is a BANACH LATTICE.

21.1.2. Example. Under pointwise operations and partial ordering $\mathbb{R}^n$ is a Banach lattice for every $n \in \mathbb{N}$.

21.1.3. Example. If $\mathfrak{A}$ is an algebra of subsets of a set S, then $\mathcal{S}\mathrm{im}(S)$ is a Banach lattice. (See example 19.1.5.)

21.1.4. Example. Let X be a nonempty compact Hausdorff space. Under the usual pointwise operations $\mathcal{C}(X)$ is a Banach lattice. (See example 20.4.3.)

21.1.5. Example. Let S be a nonempty set. The family $\mathcal{B}(S)$ of bounded real valued functions on S is a Banach lattice (under the usual pointwise operations and uniform norm). (See example 20.4.4.)

21.1.6. Example. Let X be a nonempty topological space. Under the usual pointwise operations $\mathcal{C}_b(X)$ is a Banach lattice. (See example 20.4.5.)

21.1.7. Example. Let X be a nonempty locally compact Hausdorff space. Under the usual pointwise operations $\mathcal{C}_0(X)$ is a Banach lattice. (See example 20.4.6.)

21.1.8. Example. If (S, μ) is a positive measure space, then $\mathrm{L}_\infty(S)$ is a Banach lattice. (See example 20.4.9.)

21.1.9. Example. If (S, μ) is a positive measure space, then $\mathrm{L}_1(S)$ is a Banach lattice. (See proposition 20.4.10.)

21.1.10. Example. Let (S, μ) be a positive measure space and $1 < p < \infty$. Then $\mathrm{L}_\mathrm{p}(S)$ is a Banach lattice. (See example 20.4.17.)

21.1.11. Example. Let $\mathfrak{A}$ be an algebra of subsets of a set S. Notice that while the norm on $\mathrm{ba}(S)$ given in example 12.1.18 is not a lattice norm the one given in example 16.5.19 is. Under this latter norm $\mathrm{ba}(S)$ is a Banach lattice. So too is $\mathrm{ca}(S)$ whenever $\mathfrak{A}$ is a σ-algebra of subsets of S. (See example 20.4.20 and its proof.)

21.1.12. Proposition. *Every normed Riesz space is Archimedean.*

21.1.13. Proposition. *In a normed Riesz space every band is closed.*

21.1.14. Proposition. *Let E be a normed Riesz space. Then*

(a) $\big\||x|\big\| = \|x\|$ *for all* $x \in E$,
(b) *the map* $x \mapsto x^+$ *is uniformly continuous, and*
(c) *the map* $x \mapsto |x|$ *is uniformly continuous.*

Hint for proof. For (b) show first that $\|x^+ - y^+\| \leq \|x - y\|$ and for (c) show $\big\||x| - |y|\big\| \leq \|x - y\|$.

21.1.15. Proposition. *If E is a normed Riesz space, then its (norm) dual E^* is a Banach lattice.*

Hint for proof. Show that E^* is an order ideal in $E^\sim$.

21.1.16. Proposition. *If E is a Banach lattice, then its norm dual and order dual are identical (that is, $E^* = E^\sim$).*

Hint for proof. Show first that the norm dual of any normed Riesz space is an order ideal in its order dual. Then argue by contradiction. Let $f \in E^\sim \setminus E^*$. Choose unit vectors x_n in E so that $|f(x_n)| \geq n^3$ for every n. Let $y_n = \sum_{k=1}^n k^{-2}|x_k|$. Show that (y_n) converges to some element b and that $0 \leq y_n \leq b$ for every n. Deduce that $n \leq n^{-2}|f(x_n)| \leq |f|(b)$ for every n. (Or see [2], theorem 24.10.)

21.1.17. Proposition. *If J_1 and J_2 are closed order ideals in a Banach lattice, then $J_1 + J_2$ is also a closed order ideal.*

Hint for proof. First show that if $x_1 \in J_1$, $x_2 \in J_2$, and $|u| \leq |x_1 + x_2|$, then there exist $v_1 \in J_1$ and $v_2 \in J_2$ such that $u = v_1 + v_2$, and $|v_i| \leq |u|$ (for $i = 1$, 2). Next suppose that a sequence (x_n) in $J_1 + J_2$ converges to a point a in the Banach lattice. Without loss of generality we may assume that $\|x_n - x_{n-1}\| \leq n^{-2}$. (Why?) For each n let $u_n = x_n - x_{n-1}$ (with $x_0 = 0$). Now decompose each u_n as above.

21.2. The Stone-Weierstrass Theorems

We have seen in examples 20.4.4 and 20.4.3 that if X is a compact Hausdorff space, then the space $\mathcal{C}(X)$ of continuous real valued functions on X is a closed (unital) subalgebra of the Banach algebra $\mathcal{B}(X)$ of bounded real valued functions on X. Next we look at a normed subalgebra of $\mathcal{B}(X)$ (and $\mathcal{C}(X)$) that is *not* closed.

21.2.1. Example. The family $\mathcal{P}\big([-1,1]\big)$ of polynomial functions on the interval $[-1,1]$ is a subalgebra of the Banach algebra $\mathcal{C}\big([-1,1]\big)$ which is not closed.

Hint for proof. Let $s(t) = |t|$ for $-1 \leq t \leq 1$. Recursively define a sequence $(p_n)_{n=0}^\infty$ of polynomials on the interval $[-1,1]$ by $p_0(t) = 0$ and $p_{n+1}(t) = p_n(t) + \frac{1}{2}\big[t^2 - \big(p_n(t)\big)^2\big]$ for $-1 \leq t \leq 1$ and $n \in \mathbb{Z}^+$. Use induction to show that

$$0 \leq |t| - p_n(t) \leq \frac{2|t|}{2 + n|t|} \leq \frac{2}{2+n}$$

for $-1 \leq t \leq 1$ and $n \in \mathbb{Z}^+$. Conclude that $p_n \to s$ (unif).

The preceding example shows that the family $\mathcal{P}\big([-1,1]\big)$ of polynomials on $[-1,1]$ is not closed in $\mathcal{B}\big([-1,1]\big)$. This leads naturally to the question: What *is* the closure of $\mathcal{P}\big([-1,1]\big)$ in $\mathcal{B}\big([-1,1]\big)$? We know from proposition 12.1.17 that the uniform limit of a sequence of continuous functions is continuous, so $\overline{\mathcal{P}\big([-1,1]\big)} \subseteq \mathcal{C}\big([-1,1]\big)$. Since the closure of a normed subalgebra is again a subalgebra, we know that $\overline{\mathcal{P}\big([-1,1]\big)}$ is a closed subalgebra of $\mathcal{C}\big([-1,1]\big)$. That it is, in fact, equal to $\mathcal{C}\big([-1,1]\big)$ is by no means obvious (but true); this is the classical *Weierstrass approximation theorem*. We will obtain this result as a corollary to the substantially more general *Stone-Weierstrass theorem*. The first (and hardest) step in deriving this celebrated result will be to prove a lattice-theoretic version of the theorem (in 21.2.3).

21.2.2. Definition. Let X be a topological space. We will say that a family $\mathcal{F}$ of continuous functions on X possesses the TWO-POINT DUPLICATION PROPERTY WITH RESPECT TO a function $f \in \mathcal{C}(X)$ if for every pair of points $x \neq y$ in X there exists $h \in \mathcal{F}$ which agrees with f at x and y.

21.2.3. Theorem (Stone-Weierstrass theorem: lattice version)**.** *Let X be a compact topological space. If a sublattice $\mathcal{A}$ of $\mathcal{C}(X)$ has the two-point duplication property with respect to a function $f \in \mathcal{C}(X)$, then $f \in \overline{\mathcal{A}}$.*

Hint for proof. Given $\epsilon > 0$ we want to find a function $h \in \mathcal{A}$ such that $\|f - h\|_u < \epsilon$. For every pair p and q of distinct points in X choose a function $h_{pq} \in \mathcal{A}$ which agrees with f at p and q. Let

$$U(p,q) = \{x \in X : h_{pq}(x) < f(x) + \epsilon\}$$

and

$$V(p,q) = \{x \in X : h_{pq}(x) > f(x) - \epsilon\}.$$

For fixed $q \in X$ find $p_1, \dots p_m$ in X such that $\{U(p_1,q), \dots U(p_m,q)\}$ covers X. Let $h_q = \bigwedge_{j=1}^m h_{p_j q}$ and $V(q) = \bigcap_{j=1}^m V(p_j,q)$. Then find $q_1, \dots, q_n$ in X such that $\{V(q_1), \dots, V(q_n)\}$ covers X. Let $h = \bigvee_{k=1}^n h_{q_k}$.

21.2.4. Proposition. *Let X be a compact topological space. Every closed subalgebra $\mathcal{A}$ of $\mathcal{C}(X)$ is a sublattice.*

Hint for proof. Show that $|f|$ belongs to $\mathcal{A}$ whenever f does. Deal first with the case $\|f\|_u \le 1$. Use example 21.2.1.

21.2.5. Proposition. *Let X be a topological space. If $\mathcal{A}$ is a separating unital subalgebra of $\mathcal{C}(X)$, then $\mathcal{A}$ has the two-point duplication property with respect to every $f \in \mathcal{C}(X)$.*

21.2.6. Theorem (Stone-Weierstrass theorem: real version)**.** *If X is a compact topological space, then every separating unital subalgebra of $\mathcal{C}(X)$ is dense in $\mathcal{C}(X)$.*

21.2.7. Corollary (Weierstrass approximation theorem)**.** *Every continuous function on the closed bounded interval $[a,b]$ can be approximated uniformly by polynomials. (That is, every continuous function is the uniform limit of a sequence of polynomials.)*

21.2.8. Proposition. *The Banach algebra $\mathcal{C}\big([a,b]\big)$ is separable (and therefore second countable).*

21.2.9. Proposition. *Let $f \in \mathcal{C}\big([0,1]\big)$. If $\int_0^1 x^n f(x)\,dx = 0$ for $n = 0,1,2,\dots,$ then $f = \mathbf{0}$.*

21.2.10. Proposition. *Let $\mathcal{G}$ be the set of all functions f in $\mathcal{C}\big([0,1]\big)$ such that f is differentiable on $(0,1)$ and $f'(\frac{1}{2}) = 0$. Then $\mathcal{G}$ is dense in $\mathcal{C}\big([0,1]\big)$.*

21.2.11. Exercise. Let F be a closed subset of $[0,\infty)$. Give a useful necessary and sufficient condition on F for the following to hold: every continuous real valued function can be uniformly approximated on F by polynomials in x^2.

21.2.12. Exercise. Find $\displaystyle\lim_{n\to\infty} \frac{\int_0^1 x^n f(x)\,dx}{\int_0^1 x^n\,dx}$ for continuous real valued functions f on $[0,1]$.

21.2.13. Definition. A family of complex valued functions is SELF-ADJOINT if it is closed under complex conjugation.

21.2.14. Theorem (Stone-Weierstrass theorem: complex version)**.** *If X is a compact topological space, then every unital separating self-adjoint subalgebra of $\mathcal{C}(X,\mathbb{C})$ is dense in $\mathcal{C}(X,\mathbb{C})$.*

21.2.15. Theorem (Generalized Stone-Weierstrass theorem)**.** *If X is a compact topological space and $\mathcal{A}$ is a closed unital subalgebra of $\mathcal{C}(X)$, then there exists a compact Hausdorff space Y such that $\mathcal{A}$ is isometrically isomorphic to $\mathcal{C}(Y)$.*

21.3. Semicontinuous Functions

21.3.1. Definition. A real valued function f on a topological space X is LOWER SEMICONTINUOUS at a point a in X if for every $\epsilon > 0$ there exists a neighborhood U of a such that $f(x) > f(a) - \epsilon$ whenever $x \in U$. The function f is UPPER SEMICONTINUOUS at a if for every $\epsilon > 0$ there exists a neighborhood U of a such that $f(x) < f(a) + \epsilon$ whenever $x \in U$.

21.3.2. Proposition. *Let f be a real valued function on a topological space X. Then*

(a) *f is lower semicontinuous if and only if $f^{\leftarrow}(-\infty, a]$ is closed for every $a \in \mathbb{R}$;*
(b) *f is upper semicontinuous if and only if $f^{\leftarrow}[a, \infty)$ is closed for every $a \in \mathbb{R}$; and*
(c) *f is lower semicontinuous if and only if $-f$ is upper semicontinuous.*

21.3.3. Remark. It involves only routine computations to show that the sum of two lower semicontinuous functions is also lower semicontinuous and that the product of two positive lower semicontinuous functions is lower semicontinuous. The preceding sentence remains true if "lower" is replaced in each case by "upper".

21.3.4. Proposition. *A subset A of a topological space X is open if and only if the characteristic function χ_A is lower semicontinuous. The set A is closed if and only if χ_A is upper semicontinuous.*

21.4. Normal Topological Spaces

21.4.1. Definition. A topological space $(X, \mathfrak{T})$ is NORMAL if disjoint closed sets can be separated by open sets. A T_4 space is a normal space which is also T_1.

21.4.2. Example. The indiscrete topology on a set with at least two elements shows that a normal topological space need not even be T_1.

21.4.3. Proposition. *A Hausdorff topological space X is normal if and only if it satisfies the following condition: If A and C are subsets of X such that $A \prec C$, then there exists a set B in X such that $A \prec B \prec C$.*

We can improve on proposition 17.3.9.

21.4.4. Proposition. *The following implications hold for topological spaces:*

$$T_4 \Rightarrow T_{3\frac{1}{2}} \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1 .$$

A plentiful supply of normal topological spaces is guaranteed by the next two results.

21.4.5. Example. Every compact Hausdorff space is normal.

PROOF. Let A and B be disjoint closed subsets of a compact Hausdorff space X. Fix $a \in A$; then for each $b \in B$ choose disjoint neighborhoods U_b of a and V_b of b. Because B is compact there exist $b_1, \dots, b_n$ in B such that $\cup_{k=1}^{n} V_{b_k} \supseteq B$. Notice that $\cup_{k=1}^{n} V_{b_k}$ is disjoint from $\cap_{k=1}^{n} U_{b_k}$ (which is a neighborhood of a).

The preceding paragraph shows that for each $a \in A$ there exist disjoint open sets W_a containing a and Y_a containing B. Since A is compact, there exist $a_1, \dots, a_m \in A$ such that $\cup_{k=1}^{m} W_{a_k} \supseteq A$. Then $\cup_{k=1}^{m} W_{a_k}$ and $\cap_{k=1}^{m} Y_{a_k}$ are disjoint open sets containing A and B respectively. □

21.4.6. Example. Every metric space is normal.

21.4.7. Proposition. *Let (A_k) and (B_k) be sequences of subsets of a normal topological space X and let S and T be subsets of X. If for every $n \in \mathbb{N}$ we have $A_n \subseteq S \prec B_n$ and $A_n \prec T \subseteq B_n$, then there exists a closed set F such that $A_n \prec F \prec B_n$ for every $n \in \mathbb{N}$.*

PROOF. There exists $C_1 \subseteq X$ such that $A_1 \prec C_1 \prec T$ [since $A_1 \prec T$]. There also exists $D_1 \subseteq X$ such that $S \cup C_1 \prec D_1 \prec B_1$ [since $S \prec B_1$ and $C_1 \prec B_1$]. Proceeding inductively, suppose that $C_1, \dots, C_n, D_1, \dots, D_n$ have been chosen in such a fashion that $A_k \prec C_k \prec T$ and $S \prec D_k \prec B_k$ for $k \in \mathbb{N}_n$. Now there exists $C_{n+1} \subseteq X$ such that

$$A_{n+1} \prec C_{n+1} \prec T \cap \left(\bigcap_{k=1}^{n} D_k \right) \tag{21.1}$$

[since $A_{n+1} \prec T$ and $A_{n+1} \subseteq S \prec D_k$ for $k \in \mathbb{N}_n$]. Furthermore, there exists $D_{n+1} \subseteq X$ such that

$$S \cup \left(\bigcup_{k=1}^{n+1} C_k \right) \prec D_{n+1} \prec B_{n+1} \tag{21.2}$$

[since $S \prec B_{n+1}$ and $C_k \prec T \subseteq B_{n+1}$ for $k \in \mathbb{N}_{n+1}$]. Taking $F = \overline{\cup_{k=1}^{\infty} C_k}$, we see [using (21.1) and (21.2)] that for each $n \in \mathbb{N}$

$$A_n \prec C_n \subseteq F \subseteq \overline{D_n} \prec B_n. \qquad \square$$

21.5. The Hahn-Tong-Katětov Theorem

21.5.1. Theorem (Hahn-Tong-Katětov Theorem). *Let X be a normal topological space, $g\colon X \to \mathbb{R}$ be upper semicontinuous, and $h\colon X \to \mathbb{R}$ be lower semicontinuous with $g \leq h$. Then there exists a continuous function $f\colon X \to \mathbb{R}$ such that $g \leq f \leq h$.*

PROOF. The following is a somewhat expanded version of the proof found in [41], Theorem 6.4.4.

Define functions $H\colon \mathbb{Q} \to \mathfrak{P}(X)$ and $G\colon \mathbb{Q} \to \mathfrak{P}(X)$ by

$$H(r) = h^{\leftarrow}(-\infty, r] \qquad \text{and} \qquad G(r) = g^{\leftarrow}(-\infty, r).$$

Notice that $H(r)$ is closed and $G(r)$ is open for every rational number r. Notice also that if r and s are rational numbers such that $r < s$, then $H(r) \prec G(s)$.

Let $t\colon \mathbb{N} \to \mathbb{Q}$ be an enumeration of $\mathbb{Q}$. Since $\{H(r)\colon r \in \mathbb{Q} \text{ and } r < t_1\}$ and $\{G(s)\colon s \in \mathbb{Q} \text{ and } s > t_1\}$ are countable families of sets and since $H(r) \subseteq H(t_1) \prec G(s)$ and $H(r) \prec G(t_1) \subseteq G(s)$ whenever r and s are rational numbers satisfying $r < t_1 < s$, we may use proposition 21.4.7 to assert the existence of a closed set $F_1 \subseteq X$ such that $H(r) \prec F_1 \prec G(s)$ whenever r and s are rational numbers satisfying $r < t_1 < s$.

Proceeding inductively, we suppose that closed sets $F_1, \dots, F_n \subseteq X$ have been chosen in such a way that

(i) $H(r) \prec F_k \prec G(s)$ whenever $k \in \mathbb{N}_n$ and r and s are rational numbers satisfying $r < t_k < s$; and

(ii) $F_j \prec F_k$ whenever $j, k \in \mathbb{N}_n$ and $t_j < t_k$.

Now define

$$\begin{aligned}
\mathfrak{E}^- &= \{F_k\colon k \in \mathbb{N}_n \text{ and } t_k < t_{n+1}\},\\
\mathfrak{E}^+ &= \{F_k\colon k \in \mathbb{N}_n \text{ and } t_k > t_{n+1}\},\\
S &= H\big(t_{n+1}\big) \cup \left(\bigcup \mathfrak{E}^-\right),\\
T &= G\big(t_{n+1}\big) \cap \left(\bigcap \mathfrak{E}^+\right),\\
\mathfrak{M} &= \mathfrak{E}^- \cup \{H(r)\colon r \in \mathbb{Q} \text{ and } r < t_{n+1}\}, \text{ and}\\
\mathfrak{N} &= \mathfrak{E}^+ \cup \{G(s)\colon s \in \mathbb{Q} \text{ and } s > t_{n+1}\}.
\end{aligned}$$

It is clear that $\mathfrak{M}$ and $\mathfrak{N}$ are countable families of sets. A little patience—but no imagination—is required to establish that $M \subseteq S \prec N$ and $M \prec T \subset N$ whenever $M \in \mathfrak{M}$ and $N \in \mathfrak{N}$. Thus we may again make use of proposition 21.4.7 to obtain a closed set $F_{n+1} \subseteq X$ such that $M \prec F_{n+1} \prec N$ whenever $M \in \mathfrak{M}$ and $N \in \mathfrak{N}$.

It is easy to see that $H(r) \prec F_{n+1} \prec G(s)$ whenever r and s are rational numbers satisfying $r < t_{n+1} < s$ [since $r < t_{n+1}$ implies $H(r) \in \mathfrak{M}$, and $s > t_{n+l}$ implies $G(s) \in \mathfrak{N}$]. If $k \in \mathbb{N}_n$ and $t_k < t_{n+1}$, then $F_k \in \mathfrak{E}^- \subseteq \mathfrak{M}$; so $F_k \prec F_{n+1}$. If, on the other hand, $k \in \mathbb{N}_n$ and $t_k > t_{n+1}$, then $F_k \in \mathfrak{E}^+ \subseteq \mathfrak{N}$; so $F_{n+1} \prec F_k$. This completes the induction argument. Thus we conclude that (i) and (ii) above, hold for all $n \in \mathbb{N}$.

Now regard F as a function from $\mathbb{N}$ into $\mathfrak{P}(X)$; that is, $F\colon \mathbb{N} \to \mathfrak{P}(X)\colon k \mapsto F_k$. Define a new function $\breve{F}\colon \mathbb{Q} \to \mathfrak{P}(X)$ by $\breve{F} = F \circ t^{-1}$. In other words $\breve{F}(t_k) = F_k$ for each rational t_k. The preceding inductive construction gives us two important properties of $\breve{F}$:

(i) $H(r) \prec \breve{F}(u) \prec G(s)$ whenever r, u, and s are rational numbers such that $r < u < s$; and

(ii) $\breve{F}(r) \prec \breve{F}(s)$ whenever r and s are rational numbers such that $r < s$.

Let $x \in X$ and $r \in \mathbb{Q}$ and suppose $x \notin \breve{F}(r)$. Then whenever $s < r$, we have $x \notin H(s) = h^{\leftarrow}(-\infty, s]$. Thus for $s < r$ we have $h(x) > s$. Therefore

$$h(x) \geq r \quad \text{whenever } x \notin \breve{F}(r). \tag{21.3}$$

It follows from (21.3) that every x in X is in at least one $\breve{F}(r)$.

Similarly, if $x \in X$ and $r \in \mathbb{Q}$ and $x \in \breve{F}(r)$, then for $s > r$ we have $x \in G(s) = g^{\leftarrow}(-\infty, s)$ and thus $g(x) < s$. Therefore,

$$g(x) \leq r \quad \text{whenever } x \in \breve{F}(r). \tag{21.4}$$

It follows from (21.4) that every $x \in X$ fails to be a member of at least one $\breve{F}(r)$. We make one more simple observation:

$$\inf\{s \in \mathbb{Q} \colon x \in \breve{F}(s)\} = \sup\{r \in \mathbb{Q} \colon x \notin \breve{F}(r)\} \tag{21.5}$$

Now define a function $f(x) = \inf\{s \in \mathbb{Q} \colon x \in \breve{F}(s)\}$. That $\operatorname{dom}(f) = X$ and $\operatorname{ran} f \subseteq \mathbb{R}$ follows from (21.3) and (21.4)above. From (21.4) we also obtain $f(x) = \inf\{r \in \mathbb{Q} \colon x \in \breve{F}(r)\} \geq g(x)$. From (21.3) and (21.5) we obtain $f(x) = \sup\{r \in \mathbb{Q} \colon x \notin \breve{F}(r)\} \leq h(x)$. Thus $g \leq f \leq h$.

It remains to show that $f \colon X \to \mathbb{R}$ Is continuous. Let $a \in X$. Given $\epsilon > 0$ choose rational numbers r, s_1 and s such that

$$f(a) - \epsilon < r < f(a) < s_1 < s < f(a) + \epsilon.$$

Clearly $a \notin \breve{F}(r)$. Since $a \in \breve{F}(s_1) \prec \breve{F}(s)$, we see that $a \in \big(\breve{F}(s)\big)^{\circ}$. Let $U = \big(\breve{F}(s)\big)^{\circ} \setminus \breve{F}(r)$. Then U is a neighborhood of a. Suppose that $x \in U$. Then $x \in \breve{F}(s)$ and so $f(x) = \inf\{u \in \mathbb{Q} \colon x \in \breve{F}(u)\} \leq s$. Similarly, $x \notin \breve{F}(r)$ and so $f(x) = \sup\{u \in \mathbb{Q} \colon x \notin \breve{F}(u)\} \geq r$. Therefore, $f(a) - \epsilon < r \leq f(x) \leq s < f(a) + \epsilon$. This shows that f is continuous at a. □

21.5.2. Example. Give examples to show that the preceding theorem fails if we replace the assumptions that g is upper semicontinuous and h is lower semicontinuous with any of the following:

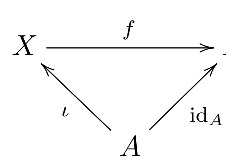
(a) g is lower semicontinuous and h is upper semicontinuous;
(b) both g and h are lower semicontinuous;
(c) both g and h are upper semicontinuous.

21.5.3. Theorem (Tietze extension theorem (real version))**.** *If A is a closed subset of a normal topological space X and if $f_0 \in \mathcal{C}(A)$ is bounded, then there exists a continuous extension f of f_0 to all of X such that* $\sup(\operatorname{ran} f) = \sup(\operatorname{ran} f_0)$ *and* $\inf(\operatorname{ran} f) = \inf(\operatorname{ran} f_0)$.

Tietze's theorem can be generalized to not necessarily bounded functions.

21.5.4. Theorem. *If A is a closed subset of a normal topological space X and if $f_0 \in \mathcal{C}(A)$, then there exists a continuous extension f of f_0 to all of X.*

21.5.5. Definition. Let A be a subset of a topological space X. A continuous map $f \colon X \to A$ is a RETRACTION if $f(x) = x$ for all $x \in A$. Thus a retraction is a continuous surjection from X to A which leaves A pointwise fixed. Another way of saying the same thing: f is a retraction (of X to A) if it is continuous and if the following diagram commutes.



When there exists a retraction from X to A we say that A is a RETRACT of X.

21.5.6. CAUTION. The preceding is the usual definition of a retraction from a topological space to a subspace: thus f is a retraction if the inclusion mapping ι of A into X is a right inverse for f. More generally, in arbitrary categories the word "retraction" often means "right invertible". And "section" and "coretraction" are synonyms for "left invertible".

21.5.7. Example. If $r \geq 0$, then $\{z \in \mathbb{C} : |z| \leq r\}$ is a retract of the complex plane $\mathbb{C}$.

21.5.8. Theorem (Tietze extension theorem (complex version))**.** *If A is a closed subset of a normal topological space X and if $f_0 \in \mathcal{C}(A, \mathbb{C})$ is bounded, then there exists a (complex valued) continuous extension f of f_0 to all of X such that $\|f\|_u = \|f_0\|_u$.*

21.5.9. Theorem (Urysohn's lemma)**.** *Any two nonempty disjoint closed subsets of a normal topological space can be functionally separated.*

21.5.10. Proposition. *Let X be a compact Hausdorff space and f be a positive lower semicontinuous function on X. Then*

$$f = \sup\{g \in \mathcal{C}(X) : 0 \leq g \leq f\}\,.$$

21.5.11. Proposition. *Let X be a compact topological space. Then $\mathcal{C}(X)$ is separating if and only if X is Hausdorff.*

21.5.12. Proposition. *Let X and Y be compact Hausdorff spaces and $\phi \colon X \to Y$ be continuous. Define $\mathcal{C}\phi$ on $\mathcal{C}(Y)$ by*

$$\mathcal{C}\phi(g) = g \circ \phi$$

for all $g \in \mathcal{C}(Y)$.

- (a) *Then $\mathcal{C}\phi$ maps $\mathcal{C}(Y)$ into $\mathcal{C}(X)$.*
- (b) *The pair of maps $X \mapsto \mathcal{C}(X)$ and $\phi \mapsto \mathcal{C}\phi$ is a contravariant functor from the category of compact Hausdorff spaces and continuous maps to the category of unital commutative Banach algebras and continuous unital algebra homomorphisms.*
- (c) *The homomorphism $\mathcal{C}\phi$ is injective if and only if the function ϕ is surjective.*
- (d) *The homomorphism $\mathcal{C}\phi$ is surjective if and only if the function ϕ is injective.*
- (e) *Suppose $X = Y$. Then $\mathcal{C}\phi$ is idempotent (with respect to composition) if and only if $\phi \colon X \to \operatorname{ran}\phi$ is a retraction.*

21.5.13. Corollary. *If X and Y are homeomorphic compact Hausdorff spaces, then the Banach algebras $\mathcal{C}(X)$ and $\mathcal{C}(Y)$ are isometrically isomorphic.*

21.5.14. Proposition. *Let X be a compact Hausdorff space, F be a closed subset of X, and $M = \{f \in \mathcal{C}(X) : f(x) = 0 \text{ for every } x \in F\}$. Then $\mathcal{C}(X)/M$ is isometrically isomorphic to $\mathcal{C}(F)$.*

21.6. Ideals in $\mathcal{C}(X)$

21.6.1. Proposition. *If J is a proper ideal in a unital Banach algebra, then so is its closure.*

21.6.2. Corollary. *Every maximal ideal in a unital Banach algebra is closed.*

21.6.3. Proposition. *Let J be a proper closed ideal in a Banach algebra A. On the quotient algebra A/J (see definition 8.2.12) define a function*

$$\|\ \ \| \colon A/J \to \mathbb{R} \colon \alpha \mapsto \inf\{\|u\| : u \in \alpha\}\,.$$

This function is a norm on A/J, under this norm A/J is a Banach algebra, and the quotient map is continuous with $\|\pi\| \leq 1$. The Banach algebra A/J is the QUOTIENT ALGEBRA *of A by J.*

21.6.4. Proposition. *Let I be a proper closed ideal in a unital commutative Banach algebra A. Then I is maximal if and only if A/I is a field.*

21.6.5. Example. For every subset C of a topological space X the set

$$J_C := \left\{ f \in \mathcal{C}(X) : f^{\rightarrow}(C) = \{0\} \right\}$$

is an ideal in $\mathcal{C}(X)$. Furthermore, $J_C \supseteq J_D$ whenever $C \subseteq D \subseteq X$. (In the following we will write J_x for the ideal $J_{\{x\}}$.)

21.6.6. Proposition. *Let X be a compact topological space and I be a proper ideal in $\mathcal{C}(X)$. Then there exists $x \in X$ such that $I \subseteq J_x$.*

21.6.7. Proposition. *Let x and y be points in a compact Hausdorff space. If $J_x \subseteq J_y$, then $x = y$.*

21.6.8. Proposition. *Let X be a compact Hausdorff space. A subset I of $\mathcal{C}(X)$ is a maximal ideal in $\mathcal{C}(X)$ if and only if $I = J_x$ for some $x \in X$.*

21.6.9. Corollary. *If X is a compact Hausdorff space, then the map $x \mapsto J_x$ from X to* $\operatorname{Max} \mathcal{C}(X)$ *is bijective.*

Compactness is an important ingredient in proposition 21.6.8.

21.6.10. Example. In the Banach algebra $\mathcal{C}_b\big((0,1)\big)$ of bounded continuous functions on the interval $(0,1)$ there exists a maximal ideal I such that for *no* point $x \in (0,1)$ is $I = J_x$. Let I be a maximal ideal containing the ideal S of all functions f in $\mathcal{C}_b\big((0,1)\big)$ for which there exists a neighborhood U_f of 0 in $\mathbb{R}$ such that $f(x) = 0$ for all $x \in U_f \cap (0,1)$.

21.6.11. Proposition. *Let X be a compact Hausdorff space, $\mathcal{C}$ the family of a;; closed subsets of X, and $\mathfrak{J}$ be the family of all closed ideals in $\mathcal{C}(X)$. Then the map*

$$J \colon \mathcal{C} \to \mathfrak{J} \colon C \mapsto J_C$$

is a bijection.

CHAPTER 22

FUNCTIONS OF BOUNDED VARIATION

In these notes a real valued function on a subset A of $\mathbb{R}$ is INCREASING if $f(x) \leq f(y)$ whenever $x \leq y$ and $x, y \in A$. If we want $f(x) < f(y)$ whenever $x < y$ we will say that f is STRICTLY INCREASING. (Of course we adopt a similar convention for the terms DECREASING and STRICTLY DECREASING.) A function is MONOTONE if it is either increasing or decreasing. And it is STRICTLY MONOTONE if it is strictly increasing or strictly decreasing.

It is not a profound observation that in the vector space of real valued functions on an interval $[a, b]$ (with $a < b$) the set $\mathcal{I} = \mathcal{I}([a, b])$ of increasing functions is a convex cone (see proposition 9.3.2). The cone is not proper: the only conclusion one can make from $f \in \mathcal{I} \cap (-\mathcal{I})$ is that f is constant. To produce a proper cone the most obvious move is to factor out the constant functions; that is, to regard two functions as equivalent if they differ by a constant. Equivalently, consider the family $\mathcal{I}_0 = \mathcal{I}_0([a, b])$ of increasing functions f on $[a, b]$ such that $f(a) = 0$. Clearly $\mathcal{I}_0$ is a proper convex cone in the vector space $\mathcal{F} = \mathcal{F}([a, b])$ of real valued functions on $[a, b]$. By proposition 9.3.4 this cone induces a partial ordering on $\mathcal{F}$ which makes $\mathcal{F}$ into an ordered vector space with $\mathcal{I}_0$ as its positive cone. It is obvious that it does not generate all of $\mathcal{F}$, since differences of functions in $\mathcal{I}_0$ would necessarily vanish at a. But we do know (see remark 9.3.9) that it generates *some* ordered vector space contained in $\mathcal{F}$. Could it be $\mathcal{F}_0$, the set of all functions on $[a, b]$ which vanish at a? Can *every* real valued function be expressed as the difference of two increasing functions? The answer to this turns out to be *no*. If a functions wriggles about too much it may be impossible to decompose it into the difference of increasing functions. Which functions then belong to $\mathcal{I}_0 - \mathcal{I}_0$? Answer: the *functions of bounded variation* for which $f(a) = 0$. We investigate these functions in this chapter.

There is more at stake here than simply answering the question just posed. The family of functions of bounded variation turns out to be quite important. In particular, as we shall see later in this chapter, it is a natural domain on which Riemann-Stieltjes integrals operate. We also show in this chapter that the family $\mathcal{I}_0 - \mathcal{I}_0$ is a Riesz space. It is an especially interesting Riesz space because the ordering on the space is not pointwise: a function f is "less than" a function g if (and only if) the function $g - f$ is increasing.

22.1. Preliminaries on Monotone Functions

22.1.1. Definition. Let f be a real valued function on an interval J in $\mathbb{R}$. (We are not committing ourselves to J's being closed or open or bounded or unbounded.) Suppose that $c \in \overline{J}$ and that $(c, \infty) \cap J \neq \emptyset$. This says that c is either an interior point of J or, in case J is bounded below, a left-hand endpoint of J. (The second supposition merely prohibits c from being a right-hand endpoint, for then $(c, \infty) \cap J$ would be empty.) In this situation we say that a number $f(c+)$ is the RIGHT-HAND LIMIT of f at c if for every $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(c+)| < \epsilon$ whenever $c < x < c + \delta$.

A real valued function f defined on an interval J in $\mathbb{R}$ is RIGHT-CONTINUOUS at a point c if c belongs to $J = \operatorname{dom} f$ and if for every $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(c)| < \epsilon$ whenever $x \in (c, c + \delta) \cap J$.

22.1.2. CAUTION. There is a not-very-interesting technical detail that should be noted about the terms just defined. As stated the definition does *not* imply that being right-continuous at a point and having a right-hand limit there are necessarily related. Take for example the identity function $x \mapsto x$ on the interval $(0,1]$. It has a right-hand limit at 0 but is not right-continuous there (because it is not defined at 0). On the other hand it is right-continuous at 1 (notice that no matter what δ is chosen, the set $(1,1+\delta)\cap(0,1]$ is empty) but has no right-hand limit there.

22.1.3. Exercise. Give a definition analogous to the preceding one for the *left-hand limit* $f(c-)$ of a function f at a point c and for the term *left-continuous*.

It follows from remark 11.2.2 that a real valued function on an interval is continuous at a point in its domain if and only if it is both left- and right-continuous there.

22.1.4. Proposition. *Let $f\colon [a,b]\to\mathbb{R}$ be an increasing function. At each $c\in(a,b]$ the function f has a left-hand limit and $f(c-)\le f(c)$. At each $c\in[a,b)$ the function f has a right-hand limit and $f(c+)\ge f(c)$.*

22.1.5. Proposition. *Let $f\colon [a,b]\to\mathbb{R}$ be an increasing function. If $a\le c<d\le b$, then $f(c+)\le f(d-)$.*

22.1.6. Corollary. *An increasing function on an interval in $\mathbb{R}$ has at most countably many discontinuities.*

22.2. Variation

22.2.1. Convention. Let f be a real valued function defined on the interval $[a,b]$. Suppose that $\mathbf{P}=\{t_0,t_1,\dots,t_n\}$ is an ordered $(n+1)$-tuple of points in $[a,b]$ such that $t_0=a$, $t_n=b$, and $t_{k-1}<t_k$ for all $k\in\mathbb{N}_n$. The $(n+1)$-tuple $\mathbf{P}$ determines a partition of $[a,b]$ which comprises the n subintervals $[t_0,t_1]$, $[t_1,t_2]$, …, $[t_{n-1},t_n]$. Since the correspondence between such $(n+1)$-tuples of points in $[a,b]$ and partitions of $[a,b]$ into subintervals is bijective we will abuse language by referring to the the set $\mathbf{P}$ as a PARTITION of $[a,b]$.

22.2.2. Definition. Let $\mathbf{P}=(s_0,s_1,\dots,s_m)$ and $\mathbf{Q}=(t_0,t_1,\dots,t_n)$ be partitions of the interval $[a,b]$. Write $\mathbf{P}\preccurlyeq\mathbf{Q}$ if the set associated with $\mathbf{P}$ is contained in the set associated with $\mathbf{Q}$, that is if $\{s_0,s_1,\dots,s_m\}\subseteq\{t_0,t_1,\dots,t_n\}$. In this case we say that $\mathbf{Q}$ is a REFINEMENT of $\mathbf{P}$.

22.2.3. Proposition. *The set of all partitions of the interval $[a,b]$ is a directed set under the relation $\preccurlyeq$ defined above. If $\mathbf{P}$ and $\mathbf{Q}$ are partitions of $[a,b]$ then their least upper bound $\mathbf{P}\vee\mathbf{Q}$ is called the* LEAST COMMON REFINEMENT *of $\mathbf{P}$ and $\mathbf{Q}$.*

22.2.4. Notation. Suppose that f is a fixed real valued function on $[a,b]$ where $a<b$. For every partition $\mathbf{P}=\{t_0,t_1,\dots,t_n\}$ of $[a,b]$ define

$$W_f(\mathbf{P}) := \sum_{k=1}^{n}\left|f(t_k)-f(t_{k-1})\right|.$$

22.2.5. Proposition. *If f is a real valued function on the interval $[a,b]$, then the function W_f defined above is an increasing net of real numbers whose domain is the family of partitions of $[a,b]$.*

22.2.6. Definition. Let f be a real valued function on $[a,b]$. If the net W_f defined in 22.2.4 is bounded, we say that the function f is of BOUNDED VARIATION. According to proposition 16.4.19 every bounded increasing net of real numbers converges. So if f is a function of bounded variation on $[a,b]$, the corresponding net W_f must converge. Its limit is denoted by $V_a^b f$ and is called the TOTAL VARIATION of f over $[a,b]$. It will be convenient to define the total variation of f over the one point interval $\{a\}$ to be 0; that is, $V_a^a f=0$. The family of all real valued functions of bounded variation on $[a,b]$ is denoted by $\mathcal{BV}([a,b])$.

22.2.7. Proposition. *Every increasing function f on a closed bounded interval $[a,b]$ is of bounded variation and $V_a^b f=f(b)-f(a)$.*

22.2.8. Proposition. *Let $a < c < b$ and $f: [a,b] \to \mathbb{R}$. Then f is of bounded variation on $[a,b]$ if and only if it of bounded variation on both $[a,c]$ and $[c,b]$. Furthermore, if f is of bounded variation on $[a,b]$, then $V_a^b f = V_a^c f + V_c^b f$.*

22.2.9. Example. Find the total variation of the following functions.

(a) $x \mapsto \sin x$ on the interval $[0, 2\pi]$.
(b) $x \mapsto \frac{5}{2}x^6 - 9x^5 - \frac{15}{4}x^4 + 15x^3$ on the interval $[-2, 2]$.

22.2.10. Example. A continuous function, even a differentiable function, on a closed and bounded interval need not be of bounded variation. For example, consider $f(x) = x^2 \sin \dfrac{\pi}{x^2}$ for $0 < x \leq 1$ and $f(0) = 0$.

22.2.11. Example. A function of bounded variation need not be piecewise monotone. (A function f is PIECEWISE MONOTONE on an interval if there exists a partition $(t_0, \dots, t_n)$ of the interval such that f is monotone on each subinterval $[t_{k-1}, t_k]$.) Consider $f(x) = x^2 \sin \dfrac{\pi}{2x}$ for $0 < x \leq 1$ and $f(0) = 0$.

22.3. The Variation Norm

On the family $\mathcal{BV}([a,b])$ of functions of bounded variation the function $f \mapsto V_a^b(f)$ is a seminorm. We can make this into a norm, which we will call the *variation norm*, by the simple expedient of adding $|f(a)|$ to the total variation $V_a^b(f)$ of f. Every function of bounded variation on an interval is bounded. So as a subspace of $\mathcal{B}([a,b])$ the set of functions of bounded variation have another norm, the uniform norm (see example 12.1.9), which turns out to be always smaller than the variation norm. We formalize this in the next proposition.

22.3.1. Proposition. *Let $a < b$. Then $\mathcal{BV}([a,b])$ is a vector subspace of $\mathcal{B}([a,b])$. For every $f \in \mathcal{BV}([a,b])$ define*

$$\|f\|_V := |f(a)| + V_a^b(f)\,.$$

The function $f \mapsto \|f\|_V$ is a norm on $\mathcal{BV}([a,b])$, which is called the VARIATION NORM. *Furthermore, for every function f of bounded variation on $[a,b]$*

$$\|f\|_u \leq \|f\|_V\,.$$

22.3.2. Example. Under the usual pointwise ordering $\mathcal{BV}([a,b])$ (where $a < b$) is a Riesz space.

The variation norm has an interesting multiplicative property.

22.3.3. Proposition. *If f and g are functions of bounded variation on the interval $[a,b]$, then the product fg is of bounded variation and*

$$\|fg\|_V \leq \|f\|_u \|g\|_V + \|f\|_V \|g\|_u\,.$$

Thus $\mathcal{BV}([a,b])$ is an algebra of functions.

It is fairly clear that if a function f gets arbitrarily close to zero, then its reciprocal $1/f$ becomes unbounded and consequently is not of bounded variation. If, however, the function f is of bounded variation and stays some fixed (but perhaps very small) distance from zero, then its reciprocal is itself of bounded variation.

22.3.4. Definition. A real valued function f on a set S is BOUNDED AWAY FROM ZERO if there exists $\delta > 0$ such that $|f(x)| \geq \delta$ for all $x \in S$.

22.3.5. Proposition. *If a function f is of bounded variation and is bounded away from zero on an interval $[a,b]$, then its reciprocal $1/f$ is also of bounded variation. In fact, if $|f(x)| \geq \delta > 0$ for $a \leq x \leq b$, then*

$$\left\|\frac{1}{f}\right\|_V \leq \frac{1}{\delta^2}\,\|f\|_V\,.$$

22.3.6. Definition. For a function f of bounded variation on the interval $[a,b]$ define

$$v_f\colon [a,b] \to \mathbb{R}\colon x \mapsto V_a^x\,.$$

This function is the VARIATION FUNCTION of f.

22.3.7. Proposition. *If f is a function of bounded variation on $[a,b]$, then its associated variation function v_f is right-continuous (or left-continuous, or continuous) at a point c in $[a,b]$ if and only if f is.*

22.3.8. Proposition. *If f is a function of bounded variation on $[a,b]$, then the function $v_f + f$ and $v_f - f$ are increasing.*

22.3.9. Corollary. *Every function of bounded variation on an interval $[a,b]$ is the difference of two increasing functions.*

22.3.10. Proposition. *If g and h are increasing functions on $[a,b]$, then so is $g + h - v_{g-h}$.*

Proposition 22.1.4and corollary 22.3.9 and make the following proposition clear.

22.3.11. Proposition. *If f is a function of bounded variation on the interval $[a,b]$, then it has both a left-hand and a right-hand limit at each point of (a,b). The limits $f(a+)$ and $f(b-)$ also exist.*

Corollaries 22.1.6 and 22.3.9 justify the following result.

22.3.12. Proposition. *A function of bounded variation has at most countably many discontinuities.*

22.4. Another Ordering on $\mathcal{BV}([a,b])$

We have already seen (in example 22.3.2) that using the family of positive functions on an interval $[a,b]$ as a positive cone makes the vector space $\mathcal{BV}([a,b])$ into a Riesz space. But recall that at the beginning of this chapter we set out to do something quite different (and more interesting). We sought to use the family of *increasing* functions on $[a,b]$ as a positive cone. We observed that in order to make the cone proper, we would need to work with the increasing functions that vanish at a. In this section we pursue the promised construction.

22.4.1. Notation. Let $\mathcal{I}_0([a,b])$ be the family of all increasing functions on the interval $[a,b]$ such that $f(a) = 0$, and let $\mathcal{BV}_0([a,b])$ be the family of all functions of bounded variation on the interval $[a,b]$ with $f(a) = 0$.

22.4.2. Example. If $a < b$, then $\mathcal{BV}_0([a,b])$ (under the induced algebraic operations, pointwise ordering, and variation norm) is a normed linear space and a Riesz space. Notice that the variation norm is quite simple for such functions: $\|f\|_V = V_a^b(f)$.

22.4.3. Proposition. *The family $\mathcal{I}_0([a,b])$ is a proper convex cone in the vector space $\mathcal{BV}[a,b]$.*

By virtue of the preceding proposition and proposition 9.3.4 there exists a partial ordering $\preccurlyeq$ on $\mathcal{BV}_0([a,b])$ which is compatible with the vector space structure. (We use $\preccurlyeq$ to distinguish this partial ordering from the usual pointwise ordering $\leq$.) According to proposition 9.3.4 we know that $f \preccurlyeq g$ if and only if $g - f$ is increasing. Thus by taking $\mathcal{I}_0([a,b])$ as the positive cone we have made $\mathcal{BV}_0([a,b])$ into an ordered vector space. Is it in fact a Riesz space? The work we have done above shows that the answer is *yes*.

22.4.4. Example. Under the partial ordering $\preccurlyeq$ defined above $\mathcal{BV}_0([a,b])$ is a Riesz space whose positive cone is the set $\mathcal{I}_0([a,b])$ of increasing functions on $[a,b]$ that vanish at a.

Hint for proof. By proposition 14.1.2 it is enough to show that $f \vee 0$ exists for every $f \in \mathcal{BV}([a,b])$. Prove that $f \vee 0 = \frac{1}{2}(v_f + f)$ for every such f.

22.5. The Fundamental Theorem of Calculus

The basic facts concerning differentiable functions between normed linear spaces have, as we saw in chapter 13, rather uncomplicated proofs. It may be then somewhat surprising to discover, even in the simple case of real valued functions of a real variable, how technically demanding matters become when we examine the relationship between differentiation and Lebesgue integration. Eventually we determine conditions under which the equation $f(x) = f(a) + \int_a^x f'\,d\lambda$ holds for all points x in some interval $[a, b]$. We know from beginning calculus that a sufficient condition for this to hold is that f be differentiable with continuous derivative. But surely this condition is not necessary. (Consider the absolute value function on $[-1, 1]$.) Before attempting to find a condition that is both necessary and sufficient, let's look at an example which shows how badly things can go wrong.

22.5.1. Example. Recall from example 17.1.31 that the Cantor function (let's call it ψ) maps the Cantor set C into $[0, 1]$ by taking the point $\sum_{k=1}^{\infty} a_k 3^{-k}$ (each a_k being 0 or 2) to $\sum_{k=1}^{\infty} a_k 2^{-(k+1)}$. Recall also that if (a, b) is one of the open intervals removed during the construction of the Cantor set, then ψ takes on the same value at a and at b. Extend ψ to all of $[0, 1]$ by making it constant on the closure of each such interval. That is, set $\psi(x) = \psi(a)(= \psi(b))$ whenever $a \le x \le b$. More formally define $\psi(x) = \sup\{\psi(c) \colon c \in C \text{ and } c \le x\}$ for all $x \in [0, 1] \setminus C$. We will call this extension of the Cantor function the *Lebesgue singular function*. (Many authors call both functions the Cantor function.) The Lebesgue singular function is increasing and continuous. It is differentiable, with derivative zero, almost everywhere so that

$$\psi(1) - \psi(0) = 1 \ne 0 = \int_0^1 \psi\,d\lambda.$$

22.5.2. Definition. A function $f \colon [a, b][\to \mathbb{R}$ is SINGULAR if it is monotone, continuous, nonconstant, and has zero derivative almost everywhere. The definition of this term appears not to be well-standardized. In [10], for example, we find "A function satisfying $f' = 0$ a.e. is called **singular**." Of course this allows the constant functions to be singular, which seems a bit odd.

22.5.3. Example. It is an easy result that differentiable functions are continuous. While everyone knows that the converse is false, it may come as a surprise how badly it fails. In [23] (theorem 17.7) you can find the details of the construction of a bounded continuous function on the real line that has a derivative at no point whatever.

Even worse: it turns out that, in some appropriate sense, most continuous functions on the real line are nowhere differentiable. A full explanation and proof of the assertion are given in theorem 17.8 of [23]. (Since the proof depends crucially on the *Baire category theorem* you may want to postpone reading this until you are familiar with the material discussed in section 29.3.)

Monotone functions (continuous or not) are another matter altogether. They must be differentiable almost everywhere. Working through proofs of this fact can be a surprisingly complicated undertaking. Be prepared to learn about such technical matters as Vitali coverings and Dini derivates.

22.5.4. Theorem. *An increasing real valued function on the interval $[a, b]$ is differentiable almost everywhere.*

PROOF. See [2], theorem 29.9; [4], theorem 2.3.9; [10], theorem 20.6; [19], theorem 222A; [23], theorem 17.12; [32], theorem 6.2; and [36], chapter5, theorem 2.

22.5.5. Corollary. *A function of bounded variation on the interval $[a, b]$ is differentiable almost everywhere.*

PROOF. Proposition 22.3.9. □

22.5.6. Proposition. *Let $f \in \mathcal{L}_1([a, b])$ where $a < b$. Define $F(x) = \int_a^x f\,d\lambda$ for every $x \in [a, b]$. Then F is a continuous function of bounded variation on $[a, b]$ with $V_a^b F = \int_a^b |f|\,d\lambda$.*

Proof. See [10], theorem 20.9 and [32], proposition 6.2.

The *first fundamental theorem of calculus* says roughly that for integrable functions differentiation is almost everywhere a left inverse of indefinite integration.

22.5.7. Theorem (Fundamental Theorem of Calculus I)**.** *Let $f \in \mathcal{L}_1([a,b])$ where $a < b$. Define $F(x) = \int_a^x f\,d\lambda$ for every $x \in [a,b]$. Then F is differentiable almost everywhere on $[a,b]$ and $F' = f$ a.e.*

Proof. See [10], theorem 20.9; [19], theorem 222E; and [32], theorem 6.5.

While it is certainly gratifying to know that the indefinite integral F in theorem 22.5.7 is differentiable almost everywhere, we must realize that such a result gives us no information whatever about what happens at any particular point in the domain of f. So 22.5.7 should certainly not be regarded as a replacement for the much more elementary result from beginning calculus that guarantees differentiability of F at points of continuity of its integrand f.

22.5.8. Proposition. *Let $f \in \mathcal{L}_1([a,b])$ where $a < b$. Define $F(x) = \int_a^x f\,d\lambda$ for every $x \in [a,b]$. Then F is differentiable at every point of (a,b) where f is continuous and at such points $F' = f$.*

22.5.9. Proposition. *Every absolutely continuous function on the interval $[a,b]$ is continuous and of bounded variation.*

Proof. See [10], proposition 20.15 and [23], theorems 18.12 and 18.13.

22.5.10. Corollary. *Every absolutely continuous function on the interval $[a,b]$ is differentiable almost everywhere.*

It is often useful to know that many of the results in this section are also true for complex valued functions. The following proposition is the key to making possible such extensions.

22.5.11. Proposition. *If f is a complex valued absolutely continuous function on the interval $[a,b]$, then there exist real valued, increasing, absolutely continuous functions f_1, f_2, f_3, and f_4, such that $f = f_1 - f_2 + i(f_3 - f_4)$.*

Proof. See [23], theorem 18.13.

22.5.12. Proposition. *An absolutely continuous function defined on an interval in $\mathbb{R}$ which has zero derivative almost everywhere is constant.*

Proof. See [10], theorem 20.16 and [23], theorem 18.15.

22.5.13. Example. The Lebesgue singular function is not absolutely continuous.

22.5.14. Definition. A function f is an indefinite integral on an interval $[a,b]$ if there exists an integrable function g on $[a,b]$ such that $f(x) = f(a) + \int_a^x g\,d\lambda$ for all $x \in [a,b]$.

22.5.15. Theorem. *A function f is an indefinite integral on an interval if and only if it is absolutely continuous on the interval.*

Proof. See [32], propositions 6.6–6.7 and [36], chapter 5, theorem 13.

The *second fundamental theorem of calculus*, which says that any absolutely continuous function is the indefinite integral of its derivative, is an easy consequence of theorem 22.5.15.

22.5.16. Corollary (Fundamental theorem of calculus II)**.** *If f is an absolutely continuous function on $[a,b]$, then its derivative (exists a.e. and) is integrable on $[a,b]$ and*

$$f(x) - f(a) = \int_a^x f'\,d\lambda$$

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

22.5.17. Example. The family H of all absolutely continuous complex valued functions f on $[0,1]$ satisfying $f(0) = 0$ and $f' \in \mathcal{L}_2([0,1])$ is a Hilbert space when we define an inner product on H by setting

$$\langle f, g\rangle = \int_0^1 f' \, \overline{g'} \, d\lambda$$

whenever f and g belong to H.

Hint for proof. Show first that the map $(f,g) \mapsto \langle f, g\rangle$ is well-defined and is an inner product. Then to show that H is complete note that if (f_n) is a Cauchy sequence in H, then the sequence (f_n') is Cauchy and therefore convergent to some function g in $\mathrm{L}_2([0,1])$. Show that g is integrable and let h be its indefinite integral $x \mapsto \int_0^x g \, d\lambda$. Finally verify that h belongs to H and is the limit of the sequence (f_n).

22.5.18. Proposition. *Let f be a function of bounded variation on the interval $[a,b]$. Then $f' \in \mathcal{L}_1([a,b])$ and $\int_a^b |f'| \, d\lambda \le V_a^b f$.*

Hint for proof. We have already seen in corollary 22.5.5 that since f is of bounded variation f' exists almost everywhere in $[a,b]$. Suppose for the moment that f is increasing and has been extended to $[a,\infty)$ by setting $f(x) = f(b)$ for all $x > b$. For each $n \in \mathbb{N}$ and $x \in [a,b]$ let $g_n(x) := n[f(x+n^{-1}) - f(x)]$. Since $f'(x) = \lim_{n\to\infty} g_n(x)$ almost everywhere, it is clear that f' is positive and measurable. Use *Fatou's lemma* 19.2.10 to show that

$$\int_a^b f' \, d\lambda \le \liminf_{n\to\infty} \; n\left[\int_b^{b+n^{-1}} f \, d\lambda - \int_a^{a+n^{-1}} f \, d\lambda\right] \le f(b) - f(a).$$

Next assume only that f is of bounded variation (not necessarily increasing). Use the preceding argument to show that $\int_a^b |f'| \, d\lambda \le \int_a^b v_f' \, d\lambda \le V_a^b f$. □

22.6. The Riemann-Stieltjes Integral

22.6.1. Definition. Let $J = [a,b]$ be a fixed interval in the real line, $\mathbf{x} = (x_0, x_1, \dots, x_n)$ be an $(n+1)$-tuple of points of J, and $\mathbf{t} = (t_1, \dots, t_n)$ be an n-tuple. The pair $(\mathbf{x}; \mathbf{t})$ is a PARTITION WITH SELECTION of the interval J if:

(a) $x_{k-1} < x_k$ for $1 \le k \le n$;
(b) $t_k \in [x_{k-1}, x_k]$ for $1 \le k \le n$;
(c) $x_0 = a$; and
(d) $x_n = b$.

The family of all partitions with selections of $J = [a,b]$ is denoted by $\mathcal{P}(J)$ or $\mathcal{P}[a,b]$.

22.6.2. Notation. If $\mathbf{x} = (x_0, x_1, \dots, x_n)$, then $\{\mathbf{x}\}$ denotes the *set* $\{x_0, x_1, \dots, x_n\}$.

22.6.3. Definition. Let $P = (\mathbf{x}; \mathbf{s})$ and $Q = (\mathbf{y}; \mathbf{t})$ be partitions (with selections) of the interval J. We write $P \preccurlyeq Q$ and say that Q is a REFINEMENT of P if $\{\mathbf{x}\} \subseteq \{\mathbf{y}\}$.

22.6.4. Proposition. *Under the relation $\preccurlyeq$ defined above $\mathcal{P}(J)$ is a directed set.*

22.6.5. Definition. Let f and α be bounded functions on the interval $J = [a,b]$ and $P = (\mathbf{x}; \mathbf{t})$ be a partition of J into n subintervals. Let $\Delta\alpha_k := \alpha(x_k) - \alpha(x_{k-1})$ for $1 \le k \le n$. Then define

$$S_{f,\alpha}(P) := \sum_{k=1}^{n} f(t_k) \, \Delta\alpha_k \, .$$

A sum $S_{f,\alpha}(P)$ where $P \in \mathcal{P}(J)$ is a RIEMANN-STIELTJES SUM of f with respect to α on J. Notice that since $\mathcal{P}(J)$ is a directed set, $S_{f,\alpha}$ is a net of real numbers.

22.6.6. Example. Let $f(x) = 3x - 2$ and $\alpha(x) = x^2$ for $0 \le x \le 4$, $\mathbf{x} = (0,1,2,3,4)$, $\mathbf{t} = (1,2,2,3)$, and $P = (\mathbf{x}, \mathbf{t})$. Calculate $S_{f,\alpha}(P)$.

22.6.7. Definition. Let f and α be bounded functions on the interval $J = [a, b]$. If the net $S_{f,\alpha}$ of Riemann-Stieltjes sums converges we say that the function f is RIEMANN-STIELTJES INTEGRABLE with respect to α. We denote by $\mathcal{R}(\alpha)$ the set of all bounded functions f which are Riemann-Stieltjes integrable with respect to α.

The limit of the net $S_{f,\alpha}$ (when it exists) is the RIEMANN-STIELTJES INTEGRAL of f with respect to α over the interval J and is denoted by

$$\int_a^b f\,d\alpha \qquad \text{or} \qquad \int_a^b f(x)\,d\alpha(x)\,.$$

The function f is the INTEGRAND and α is the INTEGRATOR. Notice that i the special case where $\alpha(x) = x$ on J, the integral is just the Riemann integral (see example 16.4.11).

22.6.8. Example. Let f be a bounded real valued function on the interval $[0, 2]$ which is continuous at $x = 1$. Let $\alpha(x) = \begin{cases} 0, & \text{if } 0 \le x \le 1 \\ 1, & \text{if } 1 < x \le 2 \end{cases}$. Prove that f is Riemann-Stieltjes integrable with respect to α on $[0, 2]$ and calculate $\int_0^2 f\,d\alpha$.

22.6.9. Exercise. Using only the definition of the Riemann-Stieltjes integral prove that

$$\int_a^b d\alpha(x) = \alpha(b) - \alpha(a)\,.$$

22.6.10. Proposition. *If $f \in \mathcal{R}(\alpha)$ and $f \in \mathcal{R}(\beta)$ on $[a, b]$ and $c \in \mathbb{R}$, then f belongs to both $\mathcal{R}(\alpha + \beta)$ and to $\mathcal{R}(c\alpha)$ and*

$$\int_a^b f\,d(\alpha+\beta) = \int_a^b f\,d\alpha + \int_a^b f\,d\beta \qquad \textit{and} \qquad \int_a^b f\,d(c\alpha) = c\int_a^b f\,d\alpha\,.$$

22.6.11. Exercise. Prove that if $f \in \mathcal{R}(\alpha)$ on $[a, b]$, then $\alpha \in \mathcal{R}(f)$ on $[a, b]$ and

$$\int_a^b f\,d\alpha + \int_a^b \alpha\,df = f(b)\alpha(b) - f(a)\alpha(a)\,.$$

For a fixed integrator function α the family $\mathcal{R}(\alpha)$ of Riemann-Stieltjes integrable functions is a vector space.

22.6.12. Proposition. *If f, $g \in \mathcal{R}(\alpha)$ on $[a, b]$ and $c \in \mathbb{R}$, then $f + g$ and cf also belong to $\mathcal{R}(\alpha)$ and*

$$\int_a^b (f+g)\,d\alpha = \int_a^b f\,d\alpha + \int_a^b g\,d\alpha \qquad \textit{and} \qquad \int_a^b (cf)\,d\alpha = c\int_a^b f\,d\alpha\,.$$

22.6.13. Proposition (Change of Variables)**.** *Suppose that $f \in \mathcal{R}(\alpha)$ on the interval $[a, b]$. Also suppose that g is a strictly increasing function on the interval $[c, d]$ with $g(c) = a$ and $g(d) = b$. Define $h = f \circ g$ and $\beta = \alpha \circ g$. Then $h \in \mathcal{R}(\beta)$ on $[c, d]$ and*

$$\int_a^b f\,d\alpha = \int_c^d h\,d\beta\,.$$

22.6.14. Proposition (Reduction to Riemann Integral)**.** *Let α be a continuously differentiable function on the interval $[a, b]$. If $f \in \mathcal{R}(\alpha)$ on $[a, b]$, then the Riemann integral $\int_a^b f\alpha'$ exists and*

$$\int_a^b f(x)\,d\alpha(x) = \int_a^b f(x)\alpha'(x)\,dx\,.$$

Hint for proof. Let $g = f\alpha'$. Try to find partitions $P = (\mathbf{x}, \mathbf{t})$ of $[a, b]$ which make the distance between the Riemann sums $S_g(P) = \sum g(t_k)\,\Delta x_k$ and the Riemann-Stieltjes sums $S_{f,\alpha}(P) = \sum f(t_k)\,\Delta\alpha_k$ small. Apply the *mean value theorem* (see, for example, [17], theorem 8.4.26) to α' on each subinterval $[x_{k-1}, x_k]$ to find points u_k such that $\Delta\alpha_k = \alpha'(u_k)\,\Delta x_k$. Also use the boundedness of f and the uniform continuity of α' to make $|\alpha'(u_k) - \alpha'(t_k)| < \epsilon/K$ (for an appropriate K) whenever u_k and t_k lie in the same subinterval of P.

22.6.15. Definition. Let $P = (\mathbf{x}, \mathbf{t})$ be a partition of $[a, b]$. Define

$$M_k(f) := \sup\{f(x) \colon x_{k-1} \le x \le x_k\}\,.$$

The Riemann-Stieltjes sums

$$U_{f,\alpha}(P) := \sum_{k=1}^{n} M_k(f)\Delta\alpha_k$$

are the UPPER RS-SUMS of f with respect to α on $[a, b]$. Similarly, if

$$m_k(f) := \inf\{f(x) \colon x_{k-1} \le x \le x_k\}$$

then the Riemann-Stieltjes sums

$$L_{f,\alpha}(P) := \sum_{k=1}^{n} m_k(f)\Delta\alpha_k$$

are the LOWER RS-SUMS of f with respect to α on $[a, b]$.

22.6.16. Notation. If J is an interval in $\mathbb{R}$ we write "$f \uparrow$ on J" to mean that the function f is (defined and) increasing on J.

22.6.17. Proposition. *Let $\alpha \uparrow$ on $[a, b]$ and P, $Q \in \mathcal{P}[a, b]$ with $P \preccurlyeq Q$. Then*

$$U_{f,\alpha}(P) \ge U_{f,\alpha}(Q) \qquad \textit{and} \qquad L_{f,\alpha}(P) \le L_{f,\alpha}(Q)\,.$$

22.6.18. Proposition. *Let $\alpha \uparrow$ on $[a, b]$ and P, $Q \in \mathcal{P}[a, b]$. Then*

$$L_{f,\alpha}(P) \le U_{f,\alpha}(Q)\,.$$

22.6.19. Definition. Let $\alpha \uparrow$ on the interval $[a, b]$. The UPPER DARBOUX-STIELTJES INTEGRAL of f with respect to α is defined by

$$\overline{\int_a^b} f\,d\alpha := \inf\{U_{f,\alpha}(P) : P \in \mathcal{R}(P)[a, b]\}.$$

Similarly, the LOWER DARBOUX-STIELTJES INTEGRAL of f with respect to α is defined by

$$\underline{\int_a^b} f\,d\alpha := \sup\{L_{f,\alpha}(P) : P \in \mathcal{R}(P)[a, b]\}.$$

For simplicity we sometimes write $\overline{I}(f, \alpha)$ for $\overline{\int_a^b} f\,d\alpha$ and $\underline{I}(f, \alpha)$ for $\underline{\int_a^b} f\,d\alpha$.

22.6.20. Proposition. *If $\alpha \uparrow$ on the interval $[a, b]$, then $\underline{I}(f, \alpha) \le \overline{I}(f, \alpha)$.*

22.6.21. Example. In the preceding proposition the quantities $\underline{I}(f, \alpha)$ and $\overline{I}(f, \alpha)$ may differ.

22.6.22. Definition. A function f is said to satisfy the RIEMANN CONDITION with respect to α on $[a, b]$ if for every $\epsilon > 0$ there exists $P_0 \in \mathcal{R}(P)[a, b]$ such that

$$0 \le U_{f,\alpha}(P) - L_{f,\alpha}(P) < \epsilon$$

whenever $P_0 \preccurlyeq P$.

22.6.23. Proposition. *If $\alpha \uparrow$ on $[a, b]$, then the following are equivalent:*

(a) *$f \in \mathcal{R}(R)(\alpha)$ on $[a, b]$;*
(b) *f satisfies the Riemann condition with respect to α on $[a, b]$; and*
(c) *$\underline{I}(f, \alpha) = \overline{I}(f, \alpha)$.*

22.6.24. Proposition (First Mean Value Theorem)**.** *If $\alpha \uparrow$ and $f \in \mathcal{R}(\alpha)$ on $J = [a, b]$, then there exists a number k in $[\inf f(J), \sup f(J)]$ such that*

$$\int_a^b f\,d\alpha = k(\alpha(b) - \alpha(a))\,.$$

If, in addition, f is continuous on J, there exists $c \in J$ such that $k = f(c)$.

22.6.25. Proposition (Second Mean Value Theorem)**.** *If α is continuous and $f \uparrow$ on $[a, b]$, then there exists $c \in [a, b]$ such that*

$$\int_a^b f\,d\alpha = f(a)[\alpha(c) - \alpha(a)] + f(b)[\alpha(b) - \alpha(c)]\,.$$

22.6.26. Proposition. *Suppose $a \le c \le b$. Then a function f belongs to $\mathcal{R}(\alpha)$ on $[a, b]$ if and only if it belongs to $\mathcal{R}(\alpha)$ on both $[a, c]$ and $[c, b]$, and that in this case*

$$\int_a^b f(x)\,d\alpha(x) = \int_a^c f(x)\,d\alpha(x) + \int_c^b f(x)\,d\alpha(x)\,.$$

22.6.27. Proposition. *If f is a continuous function on the interval $[a, b]$ and $\alpha \uparrow$ on $[a, b]$, then $f \in \mathcal{R}(\alpha)$ on $[a, b]$.*

22.6.28. Proposition. *If f is a continuous function on the interval $[a, b]$ and α is of bounded variation on $[a, b]$, then $f \in \mathcal{R}(\alpha)$ on $[a, b]$. The same conclusion holds if f is of bounded variation and α is continuous.*

CHAPTER 23

PROBABILITY

Probability theory is a branch of measure theory. As a body mathematics it is straightforward, entirely clear, and enjoyable. Its applications to real world events, to the world of statistics are open to all manner of controversy. The fundamental principles governing *the applications of* probability theory are subject to endless controversy and are, in my opinion, poorly understood. This short chapter is nothing more than a brief introduction to the language of probability theory together with a few words of caution concerning some of the more thoughtless aspects of its applications.

23.1. The Language of Probability

Although, mathematically, probability theory is a branch of measure theory, some aspects of it may seem unfamiliar at first because the language it uses is quite different from the language of measure theory, and so is its notation.

23.1.1. Definition. A PROBABILITY SPACE is a positive measure space $(\Omega, \mathfrak{A}, P)$ such that $P(\Omega) = 1$. The set Ω is called a SAMPLE SPACE, its elements are called OUTCOMES, and its measurable subsets, that is, members of $\mathfrak{A}$ are called EVENTS. The measure P is a PROBABILITY MEASURE. Events A and B are MUTUALLY EXCLUSIVE if they are disjoint. More generally, the events in a family $\mathfrak{B} \subseteq \mathfrak{A}$ are PAIRWISE MUTUALLY EXCLUSIVE if every pair of distinct members of $\mathcal{B}$ are mutually exclusive. The family $\mathfrak{B}$ of events is EXHAUSTIVE if their union is all of Ω. Events A and B are INDEPENDENT if $P(A \cap B) = P(A)P(B)$.

23.1.2. Example. A DISCRETE PROBABILITY SPACE is a probability space $(\Omega, \mathfrak{A}, P)$ where the sample space $\Omega = \{\omega_i \colon i \in I\}$ is countable and $\mathfrak{A} = \mathfrak{P}(\Omega)$. The probabilities $p_i = P(\{\omega_i\})$ of the individual outcomes determine the probabilities of all events in the obvious way

$$P(A) = \sum_{\omega_i \in A} p_i\,.$$

23.1.3. Example. A special case of the preceding example is the subject of much classical probability theory. The subject of study is a finite probability space in which each outcome is equally likely; that is, $\Omega = \{\omega_1, \dots, \omega_n\}$ and $P(\{\omega_i\}) = \frac{1}{n}$ for each $i \in \mathbb{N}_n$. Thus for each event A

$$p(A) = \frac{\operatorname{card} A}{\operatorname{card} \Omega}\,.$$

Clearly problems in this classical probability setting make no use of measure theory. Everything is reduced to counting elements of finite sets.

23.1.4. Example. One frequently encounters phrases such as,

The probability of rolling any given number from 1 to 6 with a fair die is $\frac{1}{6}$. $\quad (*)$

Here, of course, we have left the world of mathematics for the world of applications and its attendant ambiguities. Although almost anyone would tend to agree with this assertion, it is not easy to explain exactly what has been said. What, we might ask, is a *fair* die? Sometimes it is defined to be a perfect cube made of perfectly homogeneous material with the numbers 1 through 6 painted on its sides. Since perfection is somewhat beyond the capacities of human engineering, we might enquire under what conditions a slightly imperfect die would be regarded for experimental (or gambling) purposes as fair. One fairly common response is to require the frequencies of the occurrence of each integer 1 through 6 on a "large" number of rolls to be "roughly" equal. This clearly changes the definition of "fair" and makes $(*)$ into a tautology. It also leaves open the questions of how, in practice, to interpret the words "large" and "roughly". This approach to probability is quite common and is called the *relative-frequency interpretation* of probability. It certainly cannot be the whole story. It is, for example, surely not a sensible way to interpret the assurance from a utility company that the risk of a meltdown of a nuclear power station they are building in your town is less than one in one hundred million.

23.1.5. Exercise. Explain how to set up a sample space to model a single throw of a pair of fair dice. What is the probability of throwing a pair of 6's? What is the probability of throwing a 7? What is the probability of throwing at least 7?

23.1.6. Exercise. If a word is drawn at random from the statement of this exercise, what is the probability it has at least three letters?

In case you are inclined to ask, while reading the preceding exercise, what the expression "at random" means, I must confess that I don't know. In the context of this particular exercise I would take it as an instruction to assign probability $\frac{1}{23}$ to the outcome of choosing any particular occurrence of any word in the sentence. Thus the probability of choosing the word "drawn" is $\frac{1}{23}$ while the probability of choosing the word "at" is $\frac{2}{23}$. A great deal has been written, much of it quite technical, by mathematicians, statisticians, philosophers, computer scientists, and others trying to tie down the meaning of the word "random". For an entertaining, informative, but nontechnical introduction to the subject I recommend the small book *Randomness* [5] by Deborah J. Bennett.

23.2. Conditional Probability

23.2.1. Definition. Let A and B be events in a probability space (Ω, P) with $P(A) > 0$. The CONDITIONAL PROBABILITY $P(B|A)$ OF THE EVENT B GIVEN A is defined by

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

What could be simpler? The (mathematical) definition of conditional probability is entirely straightforward. Applying the definition to concrete cases is another matter. Probability is often regarded as a measure of conviction. The degree of one's initial conviction concerning the occurrence of an event B is the probability $P(B)$ in the absence of information. (This is usually called the *prior probability* of the event.) On the basis of additional information (say the occurrence of event A) it is natural for one to alter one's degree of conviction to match the conditional probability $P(B|A)$. (This is the *posterior probability* of B given A.) There is however a serious problem here, one that is seldom addressed adequately in statistics texts. What exactly does the "given A" really mean? That A has occurred? That you know it has occurred? Or that you know in some special way that it has occurred? Pay special attention to exercises 23.2.9 and 23.2.10 to see how treacherous failure to be precise about this matter can be. It is difficult to find serious discussions of this matter. The most clear-headed treatment I know is Keith Hutchinson's article [25] in *The British Journal for the Philosophy of Science*. Many special cases have received a great deal of attention. (For an entertaining example search for "Monty Hall problem" in the Wikipedia [46].)

23.2.2. Proposition. *If A and B are independent events in a probability space with $P(A) > 0$, then*

$$P(B|A) = P(B)\,.$$

23.2.3. Proposition. *If $\{B_j\}$ is a countable family of mutually exclusive and exhaustive events in a probability space (Ω, P) and A is an event in Ω, then*

$$P(A) = \sum_j P(B_j)\,P(A|B_j)$$

where the sum is taken over all j such that $P(B_j) > 0$.

23.2.4. Proposition. *If A, B, and C are events of strictly positive probability (in some probability space), then*

$$P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)\,.$$

23.2.5. Proposition. *If A, B, and C are events of strictly positive probability (in some probability space), then*

$$P(A \cap B|C) = P(A|C)\,P(B|A \cap C)\,.$$

23.2.6. Exercise (Bayes Box Paradox)**.** A chest has three drawers. The first drawer contains 2 gold coins; the second contains 1 gold and 1 silver coin, and the third contains 2 silver coins. Randomly you choose a drawer and a coin from that drawer. It is gold. What is the probability that the remaining coin in the drawer you have chosen is gold?

23.2.7. Proposition (Bayes' Law)**.** *Let A and B be events in a probability space (Ω, P) with $P(A) > 0$. Then*

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}\,.$$

23.2.8. Exercise. Big Brother Corporation decides to do its part in making the community in which it is located a safer and more law-abiding place. It is going to accomplish this by testing all its employees (except of course for top management) for methamphetamine (meth) use. It makes use of an excellent new test which produces a positive result 99% of the time for meth users and a negative result 99% of the time for those who do not use meth. According to recent Government statistics approximately 731,000 Americans are current meth users, which is roughly 0.3% of the population over the age of 12. Assume that this is the percentage of meth users working for Big Brother.

(a) A Big Brother employee tests positive for meth use. Given the remarkable accuracy of the test, what would you guess is the probability that this employee is a meth user?
(b) Do the math. Use *Bayes' law* to calculate the probability that the employee is a meth user.

23.2.9. Exercise. (a) What is the probability that when two (fair) dice are rolled that the sum of the faces showing is 7?
(b) At a gambling table a croupier, whom you know to be honest, shakes two dice in a cup, puts the cup down on the table, and peeks under it. He says, "One of the dice is a two". What is the conditional probability that the sum of the two dice is 7 given the information you have received that one of them is a two?
(c) A few moments later the croupier again rolls the dice, looks under the cup and says, "one of the dice is a …". Someone coughs, causing you to miss the last word he said. You conclude that he must have specified some number between 1 and 6. In each of the six cases what would you conclude about the conditional probability of getting a 7? Does this result seem strange?

23.2.10. Exercise. Try to find a correct approach (and answer) to each of the following questions. Alternatively, give an argument that the question is ill-defined. (*Assume for this exercise that the probabilities of a child's being a boy and being a girl are equal.*)

(a) You know that Anastasia is the mother of two children of different ages. One day you observe Anastasia walking with a young boy. Someone tells you (correctly) that this is the older of her children. What is the probability that both of Anastasia's children are boys?
(b) You know that Anastasia is the mother of two children of different ages. One day you observe Anastasia walking with a young boy. Someone tells you (correctly) that this is one of her children. What is the probability that both of Anastasia's children are boys?
(c) You know that Anastasia is the mother of two children of different ages. Someone tells you (correctly) that the older of her children is a boy. What is the probability that both of Anastasia's children are boys?

(d) You know that Anastasia is the mother of two children of different ages. Someone tells you (correctly) that at least one of her children is a boy. What is the probability that both of Anastasia's children are boys?

(e) You know that Anastasia is the mother of two children of different ages. Someone tells you (correctly) that at least one of her children is a boy named Fred. What is the probability that both of Anastasia's children are boys?

Several versions of the preceding exercise are discussed (rather lightly) in the last chapter of [**5**]. In the same chapter the author discusses a number of other fascinating "paradoxes" of probability. For a deeper discussion see [**25**].

23.2.11. Exercise. In a particular region of the United States a study is conducted by a large tobacco company in which they find, to their delight, that the lung cancer mortality rate among cigarette smokers is actually *lower* than among nonsmokers. A public interest group does the same study among the urban population of the region and finds that the smokers there have a *higher* lung cancer mortality rate than the nonsmokers. A second public interest group does the study for the rural (that is, non-urban) population of the region and also finds the smokers to be at a *higher* risk of dying of lung cancer than the nonsmokers. Can you conclude from this that someone here is mistaken—or lying? Explain.

23.3. Random Variables

In probability and statistics one encounters *random variables*. The terminology is odd since they are neither random nor variables.

23.3.1. Definition. A real valued function on a probability space (Ω, P) is called a RANDOM VARIABLE if it is Borel measurable.

23.3.2. Notation. If X is a random variable on a probability space (Ω, P) and B is a Borel subset of $\mathbb{R}$, then it is usual to write the event $\{\omega \in \Omega \colon X(\omega) \in B\}$ as $X \in B$. Similarly, if $a \in \mathbb{R}$, then $X \le a$ denotes the event $\{\omega \in \Omega \colon X(\omega) \le a\}$.

23.3.3. Definition. Let X be a random variable on a probability space (Ω, P). The PROBABILITY DISTRIBUTION OF X (or the PROBABILITY MEASURE INDUCED BY X) is the function

$$P_X \colon \mathfrak{B}\mathrm{or}(\mathbb{R}) \to [0,1] \colon B \mapsto P(X \in B).$$

And the DISTRIBUTION FUNCTION of the random variable X is the function

$$F_X \colon \mathbb{R} \to [0,1] \colon x \mapsto P(X \le x).$$

23.3.4. Proposition. *In the preceding definition the function P_X is indeed a probability measure on the measurable space $(\mathbb{R}, \mathfrak{B}\mathrm{or}(\mathbb{R}))$.*

23.3.5. Proposition. *If two random variables on a probability space have the same distribution functions, then they induce the same probability measure on $\mathbb{R}$.*

23.3.6. Definition. Let X be a random variable on a probability space (Ω, P). The EXPECTATION (or EXPECTED VALUE) of X is defined by

$$E(X) := \int_\Omega X \, dP \, .$$

23.3.7. Proposition. *Let X be a random variable on a probability space (Ω, P) and $g \colon \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. Then*

$$E(g \circ X) = \int_{\mathbb{R}} g \, dP_X \, .$$

23.3.8. Definition. Let X be a random variable on a probability space (Ω, P). The VARIANCE of X is defined by

$$\operatorname{Var}(X) := E\big((X - E(X))^2\big)\,.$$

If $\operatorname{Var}(X) < \infty$, then X is of FINITE VARIANCE and $\sigma(X) := \sqrt{\operatorname{Var}(X)}$ is called the STANDARD DEVIATION of X.

23.3.9. Proposition. *If X is a random variable on a probability space, then*

$$\operatorname{Var}(X) = E(X^2) - \big(E(X)\big)^2\,.$$

CHAPTER 24

PRODUCT MEASURES AND ITERATED INTEGRALS

24.1. Product Measures

In this chapter we discuss the product and the coproduct of a pair of measurable spaces. On the product of two σ-finite positive measure spaces we introduce a "product measure". Then we examine the (somewhat fussy) relationship between integration with respect to this product measure and the usual "iterated integration" encountered in beginning calculus courses.

24.1.1. Definition. Let $(S, \mathfrak{A})$ and $(T, \mathfrak{B})$ be measurable spaces. If $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$, then $A \times B$ is a MEASURABLE RECTANGLE in $S \times T$. Let $\mathfrak{A} \otimes \mathfrak{B}$ be the σ-algebra generated by the family of all measurable rectangles in $S \times T$. A common notation for $\mathfrak{A} \otimes \mathfrak{B}$ is (the very misleading) $\mathfrak{A} \times \mathfrak{B}$. The measurable space $(S \times T, \mathfrak{A} \otimes \mathfrak{B})$ is the PRODUCT SPACE of the measurable spaces $(S, \mathfrak{A})$ and $(T, \mathfrak{B})$. When it appears convenient to avoid explicit mention of the σ-algebra $\mathfrak{A} \otimes \mathfrak{B}$ one may choose to write $E \overset{m}{\subseteq} S \times T$ for $E \in \mathfrak{A} \otimes \mathfrak{B}$.

24.1.2. Example. $\mathfrak{B}\text{or}(\mathbb{R}) \otimes \mathfrak{B}\text{or}(\mathbb{R}) = \mathfrak{B}\text{or}(\mathbb{R}^2)$.

24.1.3. Example. The *product* of measurable spaces as defined in24.1.1 (together with the two obvious projection functions) is in fact a product in the category of measurable spaces and measurable maps (see 8.3.1).

Hint for proof. In 24.1.1 $(S \times T, \mathfrak{A} \otimes \mathfrak{B})$ is a measurable space by definition. It is easy to verify that the usual projection functions π_1 and π_2 from $S \times T$ to S and T, respectively, are morphisms (that is, are measurable functions). Let $(U, \mathfrak{C})$ be an arbitrary measurable space. It should be clear that there is only one function h mapping U to $S \times T$ for which $f = \pi_1 \circ h$ and $g = \pi_2 \circ h$. The only part of the proof that requires some thought is showing that h is a morphism. It is a routine calculation to prove that $h^{\leftarrow}(R)$ belongs to $\mathfrak{C}$ whenever R is a measurable rectangle in $S \times T$. Use proposition 15.1.9 to prove that this is enough to establish the measurability of h.

24.1.4. Exercise. Identify the *coproduct* of two measurable spaces (if it exists). *Hint.* Recall example 8.4.4.

24.1.5. Proposition. *If $(S, \mathfrak{A})$ and $(T, \mathfrak{B})$ are measurable spaces, then the family of finite (pairwise) disjoint unions of measurable rectangles is an algebra of subsets of $S \times T$.*

Hint for proof. Instead of using the definition given in 5.6.11 for an *algebra of sets* show that the family in question is closed under (finite) intersections and taking of complements.

24.1.6. Notation. Let S and T be measurable spaces and $E \overset{m}{\subseteq} S \times T$. For each $x \in S$ we define ${}^xE := \{y \in T \colon (x, y) \in E\}$; this is a VERTICAL SECTION of E. Similarly, For each $y \in T$ we define $E^y := \{x \in S \colon (x, y) \in E\}$; this is a HORIZONTAL SECTION of E.

If $f \colon S \times T \to \overline{\mathbb{R}}$ is a measurable function, then for each $x \in S$ we define

$$ {}^xf \colon T \to \mathbb{R} \colon y \mapsto f(x, y) $$

and for each $y \in T$ we define

$$ f^y \colon S \to \mathbb{R} \colon x \mapsto f(x, y)\,. $$

24.1.7. Proposition. *Let S and T be measurable spaces.*

(a) *If $E \overset{m}{\subseteq} S \times T$, then for every $x \in S$ the section ${}^xE$ is a measurable subset of T and for every $y \in T$ the section E^y is a measurable subset of S.*
(b) *If $f\colon S \times T \to \overline{\mathbb{R}}$ is measurable, then so are the functions ${}^xf$ and f^y for $x \in S$ and $y \in T$.*

Hint for proof. For the first claim (that each ${}^xE$ is measurable) let

$$\mathfrak{G} = \{E \overset{m}{\subseteq} S \times T\colon {}^xE \overset{m}{\subseteq} T \text{ for every } x \in S\}$$

and use the *good sets principle.*

24.1.8. Proposition. *Let $(S, \mathfrak{A}, \mu)$ and $(T, \mathfrak{B}, \nu)$ be σ-finite positive measure spaces. For each set $E \in \mathfrak{A} \otimes \mathfrak{B}$ define*

$$f_E\colon S \to [0, \infty]\colon x \mapsto \nu({}^xE)$$

and

$$g_E\colon T \to [0, \infty]\colon y \mapsto \mu(E^y).$$

Then f_E is $\mathfrak{A}$-measurable, g_E is $\mathfrak{B}$-measurable, and

$$\int_S f_E \, d\mu = \int_T g_E \, d\nu. \tag{24.1}$$

Hint for proof. Notice first that for each $x \in S$ we can write $\nu({}^xE)$ as $\int_T \chi_E(x, y)\, d\nu(y)$.

Let $\mathfrak{E}$ be the family of all sets E in $\mathfrak{A} \otimes \mathfrak{B}$ for which (24.1) holds. Show that a subset E_0 of $S \times T$ belongs to $\mathfrak{E}$ if it satisfies any of the following four conditions:

(i) it is a measurable rectangle;
(ii) it is the limit of an increasing sequence of sets belonging to $\mathfrak{E}$;
(iii) it is the union of a countable disjoint family of sets belonging to $\mathfrak{E}$; or
(iv) it is the limit of a decreasing sequence of sets belonging to $\mathfrak{E}$ the first of which is contained in a measurable rectangle whose sides have finite measure.

By hypothesis $S = \biguplus_{k=1}^{\infty} S_k$ and $T = \biguplus_{k=1}^{\infty} T_k$ where each S_k and T_k has finite measure. (To accommodate the finite case we allow some of these sets to be empty.) For each j, $k \in \mathbb{N}$ and $E \in \mathfrak{E}$ let $E_{jk} = E \cap (S_j \times T_k)$. Then let $\mathfrak{M}$ be the family of all sets $E \in \mathfrak{A} \otimes \mathfrak{B}$ such that E_{jk} belongs to $\mathfrak{E}$ for every j and k. Deduce from (ii) and (iv) that $\mathfrak{M}$ is a monotone class. From (i) and (iii) conclude that the family $\mathfrak{R}$ of finite disjoint unions of measurable rectangles is contained in $\mathfrak{M}$. Use propositions 24.1.5 and 15.2.9 to show that $\mathfrak{A} \otimes \mathfrak{B} = \mathfrak{M}$. Then a simple application of (iii) shows that $\mathfrak{A} \otimes \mathfrak{B} = \mathfrak{E}$.

24.1.9. Proposition. *Let $(S, \mathfrak{A}, \mu)$ and $(T, \mathfrak{B}, \nu)$ be σ-finite positive measure spaces. On the family $\mathfrak{R}$ of finite disjoint unions of measurable rectangles in $S \times T$, define a function ρ by*

$$\rho(R) = \sum_{k=1}^{n} \mu(A_k)\nu(B_k)$$

where $R = \biguplus_{k=1}^{n} A_k \times B_k$ is a finite disjoint union of measurable rectangles $A_k \times B_k$. Then ρ extends to a positive measure on $\sigma(\mathfrak{R})$.

Hint for proof. For $E \in \mathfrak{A} \otimes \mathfrak{B}$ let $\rho(E) = \int_S \nu({}^xE)\, d\mu(x)$ or, equivalently (by proposition 24.1.8), $\rho(E) = \int_T \mu(E^y)\, d\nu(y)$. Use proposition 19.2.8 to establish countable additivity.

24.1.10. Definition. The measure ρ produced in the preceding proposition is the PRODUCT MEASURE on the measurable space $S \times T$. It is frequently denoted by $\mu \times \nu$. The space $(S \times T, \mathfrak{A} \otimes \mathfrak{B}, \mu \times \nu)$ is the PRODUCT MEASURER SPACE.

24.1.11. Example. If μ is counting measure on $\mathbb{N}$, then $\mu \times \mu$ is counting measure on $\mathbb{N} \times \mathbb{N}$.

24.2. Complete Measure Spaces

It is natural to find comfort in a result such as proposition 15.3.15, which tells us that the pointwise limit of a sequence of measurable functions is measurable. Our sense of well-being may, however, be somewhat diminished when it occurs to us that in practice pointwise convergence of sequences of functions in measure spaces is not all that common. Much more frequently we encounter *convergence almost everywhere*. Well then, does the convergence almost everywhere of a sequence of measurable functions to a function g imply that g is measurable? Unfortunately not. We give a simple example of how this failure may occur in 24.2.3. In preparation for the example we need to know that not every subset of the Cantor set is a Borel set.

24.2.1. Example. Let C be the Cantor set and f be the Lebesgue singular function (see example 22.5.1). The measure space we are considering is $([0,1], \mathfrak{Bor}([0,1]), \lambda)$, where λ is Lebesgue measure restricted to the Borel subsets of $[0,1]$. Since we are restricting our attention to the unit interval, we denote by C^c the complement of the Cantor set in $[0,1]$. On C^c the function f is locally constant. (A function is LOCALLY CONSTANT on a set if every point of the set has a neighborhood on which the function is constant.) It follows from this that $\lambda\big(f^{\rightarrow}(C)\big) = 1$.

Let $g(x) = x + f(x)$ for all $x \in [0,1]$. Then g is a homeomorphism between $[0,1]$ and $[0,2]$. (To prove that g^{-1} is continuous, notice that g takes open intervals to open intervals.) From this we deduce that $g^{\rightarrow}(C)$ has Lebesgue measure one. Thus by example 18.3.10 $g^{\rightarrow}(C)$ has a subset, call it F, which is not Lebesgue measurable. Let $E = g^{\leftarrow}(F)$. Then E is a subset of the Cantor set. If it were a Borel set, then F too would be a Borel set and would therefore be Lebesgue measurable. So E is a subset of the Cantor set which is not a Borel set.

24.2.2. Example. The set E in the preceding example shows that a subset of a set of measure zero need not be measurable.

24.2.3. Example. As in example 24.2.1 we consider the positive measure space $([0,1], \mathfrak{Bor}([0,1]), \lambda)$, where λ is Lebesgue measure restricted to the Borel subsets of $[0,1]$. For each $n \in \mathbb{N}$ let f_n be the function identically zero on $[0,1]$. Let E be a subset of the Cantor set which is not a Borel set (see 24.2.1) and $g = \chi_E$. Then each f_n is (Borel) measurable, $f_n \to g$ (λ-a.e.), but g is not (Borel) measurable.

If we contemplate the source of the unpleasantness in the preceding example we soon see that what our intuition tells us is to ignore sets of measure zero. They don't matter. But this is not going to work well if sets of measure zero can have nonmeasurable subsets. This, basically, is the whole problem. Fortunately its solution is fairly straightforward.

Among the badly overworked terms in mathematics is "complete". In these notes we have already encountered *metric completeness, order completeness, Dedekind completeness, σ-Dedekind completeness,* and *product completeness* (of categories). And shortly we will discussing *complete orthonormal sets*. Well, here is one more.

24.2.4. Definition. A positive measure space $(S, \mathfrak{A}, \mu)$ is COMPLETE if every subset of a measurable set of measure zero is itself measurable; that is, if $E \in \mathfrak{A}$ whenever $E \subseteq A$, $A \in \mathfrak{A}$, and $\mu(A) = 0$.

24.2.5. Example. The positive measure space $(\mathbb{R}, \mathfrak{M}_\lambda, \lambda)$ (where λ is Lebesgue measure and $\mathfrak{M}_\lambda$ is the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}$) is complete. (See example 18.3.2.)

24.2.6. Example. The positive measure space $([0,1], \mathfrak{Bor}([0,1]), \lambda)$, where λ is Lebesgue measure restricted to the Borel subsets of $[0,1]$ is not complete. (See example 24.2.2.)

24.2.7. Example. The product of two complete positive measure spaces need not be complete. Consider the complete positive measure space comprising the real line and Lebesgue measure on the Lebesgue measurable subsets of $\mathbb{R}$. The product of this space with itself is not complete.

Hint for proof. Consider the set $A \times \{0\} \subseteq \mathbb{R} \times \mathbb{R}$, where A is a nonmeasurable subset of $\mathbb{R}$.

In these notes we have discussed Lebesgue measure, which generalizes the notion of length, on $\mathbb{R}$. Of course, it is important to have corresponding positive measures, also called *Lebesgue measures*, on higher dimensional Euclidean spaces which generalize the notions of area, volume, *etc.* Many texts develop Lebesgue n-dimensional measure in detail—see, for example, [2], section 15; [11], Chapter 1, sections 3–4; [37], sections 2.19–2.21; or [44], chanpter V, section 1.

It turns out that for each $n \in \mathbb{N}$ Lebesgue n-dimensional measure produces a complete positive measure space on the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}^n$. It is worth noting that (as we discovered in example 24.2.7) if λ is Lebesgue measure on $\mathbb{R}$, then the product measure space with measure $\lambda \times \lambda$ on $\mathbb{R}^2$ is *not* complete. So Lebesgue measure on $\mathbb{R}^2$ is certainly not $\lambda \times \lambda$.

The good news here is that every positive measure space can be embedded in a slightly larger such space (called its *completion*), and, further, that Lebesgue measure on $\mathbb{R}^2$ is just the completion of the product space whose measure is $\lambda \times \lambda$ (see proposition 24.2.11 below). The fundamental idea of the completion of a measure space is to throw in the sets that are missing—all the subsets of measurable sets of measure zero. Of course, we must do that in such a way that we end up with a σ-algebra of measurable sets. Here is how it is done.

24.2.8. Notation. Let $(S, \mathfrak{A}, \mu)$ be a positive measure space. Define a new family $\widetilde{\mathfrak{A}}$ of sets as follows: a subset A of S belongs to $\widetilde{\mathfrak{A}}$ if and only if there exist sets E, $F \in \mathfrak{A}$ such that

(i) $E \subseteq A \subseteq F$ and
(ii) $\mu(F \setminus E) = 0$.

Also define $\widetilde{\mu} \colon \widetilde{\mathfrak{A}} \to [0, \infty]$ by

$$\widetilde{\mu}(A) = \mu(E) = \mu(F).$$

24.2.9. Proposition. *With notation as above $(S, \widetilde{\mathfrak{A}}, \widetilde{\mu})$ is a complete positive measure space.*

Hint for proof. Don't forget to show that $\widetilde{\mu}$ is well-defined. In particular, it must be shown that the value $\widetilde{\mu}(A)$ does not depend on the choice of the particular sets E and F satisfying conditions (i) and (ii). (If you get stuck, you can find a detailed and readable proof in [11], proposition 1.5.1.)

24.2.10. Definition. With notation as above the complete positive measure space $(S, \widetilde{\mathfrak{A}}, \widetilde{\mu})$ is the COMPLETION of the space $(S, \mathfrak{A}, \mu)$.

24.2.11. Proposition. *For each $p \in \mathbb{N}$ let λ_p be Lebesgue measure on $\mathbb{R}^p$ and $\mathfrak{M}_p$ be the family of Lebesgue measurable subsets of $\mathbb{R}^p$. Then $(\mathbb{R}^{p+q}, \mathfrak{M}_{p+q}, \lambda_{p+q})$ is the completion of the space $(\mathbb{R}^{p+q}, \mathfrak{M}_p \otimes \mathfrak{M}_q, \lambda_p \times \lambda_q)$.*

PROOF. See [37], theorem 8.11.

24.3. The Theorems of Fubini and Tonelli

When a mathematician claims in a proof to be making use of *Fubini's theorem* (s)he may be referring to any one of a dozen or so closely related theorems about changing the order of integration. Below are two typical versions of these theorems. Try using proposition 24.1.8 to prove them. Almost any good text on Measure Theory or Real Variables will have proofs of these (or very similar) results.

24.3.1. Theorem (Tonelli)**.** *Let (S, μ) and (T, ν) be σ-finite positive measure spaces. If f is a positive $\mu \times \nu$-measurable function, then the functions*

$$g \colon S \to [0, \infty] \colon x \mapsto \int_T {}^x\!f \, d\nu \qquad \textit{and} \qquad h \colon T \to [0, \infty] \colon y \mapsto \int_S f^y \, d\mu$$

are measurable and

$$\int_{S \times T} f \, d(\mu \times \nu) = \int_S \left[\int_T f(x, y) \, d\nu(y) \right] d\mu(x) = \int_T \left[\int_S f(x, y) \, d\mu(x) \right] d\nu(y)\,.$$

24.3.2. Theorem (Fubini)**.** *Let* (S,μ) *and* (T,ν) *be* σ*-finite positive measure spaces. If* $f \in \mathrm{L}_1(\mu\times\nu)$*, then* ${}^x f \in \mathrm{L}_1(\nu)$ *for almost all* $x \in S$*,* $f^y \in \mathrm{L}_1(\mu)$ *for almost all* $y \in T$*, and the functions*

$$g\colon S \to [0,\infty]\colon x \mapsto \int_T {}^x f\, d\nu \qquad \text{and} \qquad h\colon T \to [0,\infty]\colon y \mapsto \int_S f^y\, d\mu$$

(defined almost everywhere) belong, respectively, to $\mathrm{L}_1(\mu)$ *and* $\mathrm{L}_1(\nu)$*. Furthermore,*

$$\int_{S\times T} f\, d(\mu\times\nu) = \int_S \left[\int_T f(x,y)\, d\nu(y)\right] d\mu(x) = \int_T \left[\int_S f(x,y)\, d\mu(x)\right] d\nu(y)\,.$$

24.3.3. Exercise. Let $f(x,y) = \dfrac{x^2-y^2}{(x^2+y^2)^2}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Compute the integrals $\int_0^1 \left[\int_0^1 f(x,y)\,dx\right] dy$ and $\int_0^1 \left[\int_0^1 f(x,y)\,dy\right] dx$. How is this related to the theorems of Fubini and Tonelli? Try to explain clearly what is going on here by converting the integral to polar coordinates.

24.3.4. Exercise. What does *Fubini's theorem* say about the measure spaces $(\mathbb{N}, \mathfrak{P}(\mathbb{N}), \mu)$ and $(\mathbb{N}, \mathfrak{P}(\mathbb{N}), \nu)$ where μ and ν are both counting measure? What does *Tonelli's theorem* say in this case?

24.3.5. Proposition. *Let* (S,μ) *and* (T,ν) *be* σ*-finite measure spaces and the functions* $f\colon S \to \mathbb{R}$ *and* $g\colon T \to \mathbb{R}$ *be* μ*-integrable and* ν*-integrable, respectively. Then the function*

$$f \otimes g\colon S \times T \to \mathbb{R}\colon (s,t) \mapsto f(s)g(t)$$

is $\mu\times\nu$*-integrable and*

$$\int\limits_{S\times T} f\otimes g\, d(\mu\times\nu) = \left(\int\limits_S f\, d\mu\right)\cdot\left(\int\limits_T g\, d\nu\right).$$

24.3.6. Example. Define f on $(0,1)\times(0,1)$ by $f(x,y) = \dfrac{x-y}{(x+y)^3}$. Then f does not belong to the space $\mathcal{L}_1\big((0,1)\times(0,1)\big)$.

24.3.7. Exercise. Let μ be counting measure on $[0,1]$ and λ be Lebesgue measure on $[0,1]$. Define $f\colon [0,1]\times[0,1] \to \mathbb{R}$ by letting $f(x,y) = 1$ on the line $y = x$ $(0 \le x \le 1)$ and 0 elsewhere. Calculate $\int_0^1\int_0^1 f(x,y)\, d\mu(x)\, d\lambda(y)$ and $\int_0^1\int_0^1 f(x,y)\, d\lambda(y)\, d\mu(x)$. Explain why your result does not contradict the *Fubini* and *Tonelli theorems*.

24.3.8. Proposition. *For a function* f *defined on* $\mathbb{R}$ *let* F *be defined on* $\mathbb{R}^2$ *by* $F(x,y) = f(x-y)$*. If* f *is measurable, then so is* F*.*

24.3.9. Proposition. *For* $f,\ g \in \mathcal{L}_1(\mathbb{R})$ *define*

$$h(x) = \int_{-\infty}^{\infty} f(x-y)g(y)\, dy \tag{1}$$

whenever the function $y \mapsto f(x-y)g(y)$ *belongs to* $\mathcal{L}_1(\mathbb{R})$*. Then h(x) is defined and finite for almost all x. Set* $h(x) = 0$ *whenever (1) is undefined. Then* $h \in \mathcal{L}_1(\mathbb{R})$ *and* $\|h\|_1 \le \|f\|_1\, \|g\|_1$*. The function h is usually denoted by* $f * g$*; this is the* CONVOLUTION *of f and g.*

24.3.10. Example. The Banach space $\mathrm{L}_1(\mathbb{R})$ becomes a commutative Banach algebra when convolution is used as multiplication. (Why is it alright to let $[f]*[g]$ be $[f*g]$?)

24.3.11. Exercise. Let $t > 0$. Use the *Fubini-Tonelli theorems* to show that

$$\int_0^t \frac{\sin x}{x}\, dx = \int_0^\infty \left(\int_0^t e^{-xy}\sin x\, dx\right) dy\,.$$

Take the limit as $t \to \infty$ to derive the formula

$$\int_0^\infty \frac{\sin x}{x}\, dx = \frac{\pi}{2}\,.$$

24.3.12. Exercise. Let A be a measurable subset of $[a, b]$ and $h > 0$. Show that

$$\frac{1}{2h}\int_a^b \lambda\big(A \cap (x-h, x+h)\big)\,dx \le \lambda(A)\,.$$

Hint. Express the integrand as an integral.

24.3.13. Exercise. Evaluate $\int_0^1 \int_y^1 e^{x^2}\,dx\,dy$. (Justify each step carefully.)

CHAPTER 25

SOME REPRESENTATION THEOREMS

25.1. The Radon-Nikodym Theorem

25.1.1. Theorem (Radon-Nikodym)**.** *Let $(S, \mathfrak{A}, \mu)$ be a σ-finite positive measure space and ν be a complex (or real, or σ-finite positive) measure on $\mathfrak{A}$ such that $\nu \ll \mu$. Then there exists a unique $f \in \mathrm{L}_1(S, \mu)$ such that*

$$\nu(A) = \int_A f \, d\mu$$

for every $A \in \mathfrak{A}$.

Proof. See [2], theorem 27.6; [11], theorems 4.2.2 and 4.2.3; [23], theorems (19.23), (19.24), and (19.36); [32], theorem 6.10; and [36], chapter 11, proposition 23.

The preceding theorem is often referred to as the *Lebesgue-Radon-Nikodym* theorem. It is not difficult to extend it to other cases. Here for example is the theorem when both μ and ν are real measures.

25.1.2. Theorem (Radon-Nikodym)**.** *Let $(S, \mathfrak{A}, \mu)$ be a real measure space, let ν be another real measure on $\mathfrak{A}$, and suppose that $\nu \ll \mu$. Then there exists a unique $f \in \mathrm{L}_1(S, \mu)$ such that*

$$\nu(A) = \int_A f \, d\mu$$

for every $A \in \mathfrak{A}$.

25.1.3. Definition. The function f in the preceding two theorems is called the Radon-Nikodym derivative of ν with respect to μ and is often denoted by $\dfrac{d\nu}{d\mu}$. One may also write $d\nu = f \, d\mu$.

25.1.4. Theorem (Lebesgue Decomposition)**.** *Let $(X, \mathfrak{A}, \mu)$ be a σ-finite positive measure space and ν be a complex (or real, or σ-finite positive) measure on $\mathfrak{A}$. Then there exist unique measures ν_a and ν_s on $\mathfrak{A}$ such that*

(1) *$\nu_s \perp \mu$,*
(2) *$\nu_a \ll \mu$, and*
(3) *$\nu = \nu_a + \nu_s$.*

Proof. See [2], theorem 27.7; [11], theorem 4.7.1; [23], theorem (19.42); [32], theorem 6.13; and [36], chapter 11, proposition 24.

John von Neumann used Hilbert space techniques to give a lovely proof of the *Radon-Nikodym theorem* which at the same time proves the *Lebesgue decomposition theorem*. A proof along these lines can be found in [37], theorem 6.10.

25.1.5. Proposition. *Let μ, ν, and ρ be real measures on a measurable space S. If $\rho \ll \mu$ and $\nu \ll \mu$, then $\rho + \nu \ll \mu$ and*

$$\frac{d(\rho + \nu)}{d\mu} = \frac{d\rho}{d\mu} + \frac{d\nu}{d\mu} \qquad (\mu\text{-a.e.})$$

25.1.6. Proposition (Chain rule)**.** *Let μ, ν, and ρ be real measures on a measurable space S. If $\rho \ll \nu$ and $\nu \ll \mu$, then $\rho \ll \mu$ and*

$$\frac{d\rho}{d\mu} = \frac{d\rho}{d\nu}\frac{d\nu}{d\mu} \qquad (\mu\text{-a.e.})$$

25.1.7. Proposition. *Let μ and ν be real measures on a measurable space S. If $\mu \ll \nu$ and $\nu \ll \mu$, then*

$$\frac{d\nu}{d\mu}\frac{d\mu}{d\nu} = 1 \qquad (\mu\text{-a.e.})$$

25.2. Regular Borel Measures

25.2.1. Definition. Let X be a locally compact Hausdorff space and $(X, \mathfrak{A}, \mu)$ be a measure space. We say that μ is a BOREL MEASURE if $\mathfrak{A}$ contains the family $\mathfrak{Bor}(X)$ of all Borel subsets of X.

25.2.2. Definition. Let μ be a positive real Borel measure on a locally compact Hausdorff space X and $E \subseteq X$. The measure μ is INNER REGULAR ON E if $\mu(E) = \sup\{\mu(K) \colon K \subseteq E \text{ and } K \text{ is compact}\}$. We say that μ is INNER REGULAR if it is inner regular on every Borel subset of X. The measure μ is OUTER REGULAR ON E if $\mu(E) = \inf\{\mu(U) \colon E \subseteq U \overset{\circ}{\subseteq} X\}$. We say that μ is OUTER REGULAR if it is outer regular on every Borel subset of X. A positive real measure is REGULAR if it is both inner and outer regular. A real Borel measure is REGULAR if both μ^+ and $\mu^-$are and a complex Borel measure is REGULAR if both its real and imaginary parts are. A positive Borel measure is REGULAR if it is inner and outer regular and satisfies the additional condition that $\mu(K) < \infty$ for every compact set $K \subseteq X$.

25.2.3. Proposition. *A real or complex Borel measure on a compact Hausdorff space is inner regular if and only if it is outer regular.*

25.2.4. Proposition. *A real or complex measure on a locally compact Hausdorff space X is regular if and only if for every Borel set B in X and for every $\epsilon > 0$ there exist an open subset U and a compact subset K of X such that $K \subseteq B \subseteq U$ and $|\mu|(U \setminus K) < \epsilon$.*

25.2.5. Notation. If X is a compact Hausdorff space we denote by $\mathbf{M}(X)$ or by $\mathrm{rca}(X)$ the family of all regular real measures defined on the Borel subsets of X.

25.2.6. Proposition. *If X is a compact Hausdorff space, then $\boldsymbol{M}(X)$ is a Riesz space.*

Hint for proof. Show that $\mathbf{M}(X)$ is an order ideal in $\mathrm{ca}(X)$.

25.2.7. Proposition. *If X is a compact Hausdorff space, then $\boldsymbol{M}(X)$ is a normed linear space. (See example 16.5.19.)*

25.2.8. Proposition. *Let $(S, \mathfrak{A}, \mu)$ be a measure space, $(T, \mathfrak{B})$ be a measurable space, and $\phi \colon S \to T$ be a measurable function. Then the function $\boldsymbol{M}\phi(\mu) := \mu \circ \phi^{\leftarrow}$ is a measure on T.*

25.2.9. Proposition. *Let $\phi \colon X \to Y$ be a continuous function between compact Hausdorff spaces. Define*

$$\boldsymbol{M}\phi \colon \boldsymbol{M}(X) \to \boldsymbol{M}(Y) \colon \mu \mapsto \mu \circ \phi^{\leftarrow}.$$

Then $\mu \circ \phi^{\leftarrow} \in \boldsymbol{M}(Y)$ and $\boldsymbol{M}$ is a covariant functor from the category **CpH** *of compact Hausdorff spaces and continuous maps into the category* $\mathbf{NLS_1}$ *of normed linear spaces and contractive linear maps.*

25.2.10. Proposition. *Let $\phi \colon X \to Y$ be a continuous function between compact Hausdorff spaces. As above define*

$$\boldsymbol{M}\phi \colon \boldsymbol{M}(X) \to \boldsymbol{M}(Y) \colon \mu \mapsto \mu \circ \phi^{\leftarrow}.$$

Then $\boldsymbol{M}$ is a covariant functor from the category **CpH** *of compact Hausdorff spaces and continuous maps into the category* **RSP** *of Riesz spaces and positive maps.*

25.2.11. Theorem (Change of variables)**.** *Let (S,μ) be a real (or complex) measure space, $(T,\mathfrak{B})$ be a measurable space, and $\phi\colon S \to T$ be a measurable function. If $\nu = \boldsymbol{M}\phi(\mu)$ is defined as in proposition 25.2.8), then*

$$\int_S f \circ \phi \, d\mu = \int_T f \, d\nu$$

for every measurable function $f\colon T \to \mathbb{R}$ for which either of the integrals exists.

This is by no means the whole story concerning changing variables. For a deep and meticulous account see Fremlin [19], volume II, sections 235 and 263.

25.3. Natural Transformations

25.3.1. Definition. Let $\mathbf{A}$ and $\mathbf{B}$ be categories and F, $G\colon \mathbf{A} \to \mathbf{B}$ be covariant functors. A NATURAL TRANSFORMATION from F to G is a map τ which assigns to each object A in $\mathbf{A}$ a morphism $\tau_A\colon F(A) \to G(A)$ in $\mathbf{B}$ in such a way that for every morphism $\alpha\colon A \to A'$ in $\mathbf{A}$ the following diagram commutes.

$$\begin{array}{ccc} F(A) & \xrightarrow{F(\alpha)} & F(A') \\ {\scriptstyle \tau_A}\big\downarrow & & \big\downarrow{\scriptstyle \tau_{A'}} \\ G(A) & \xrightarrow[G(\alpha)]{} & G(A') \end{array}$$

We denote such a transformation by $\tau\colon F \to G$. (The definition of a natural transformation between two contravariant functors should be obvious: just reverse the horizontal arrows in the preceding diagram.

A natural transformation $\tau\colon F \to G$ is a NATURAL EQUIVALENCE if each morphism τ_A is an isomorphism in $\mathbf{B}$.

25.3.2. Example. Let V be a finite dimensional vector space and $x \in V$. Define

$$\widehat{x}\colon V^\dagger \to \mathbb{R}\colon f \mapsto f(x)\,.$$

Recall that $V^\dagger$ is the algebraic dual of V, that is, the vector space of all linear functionals on V. The second dual $\big(V^\dagger\big)^\dagger$ is denoted by $V^{\dagger\dagger}$. If $T\colon V \to W$ is a linear map between finite dimensional vector spaces, then its algebraic dual is defined as in 15.4.6: if $g \in W^\dagger$, then $T^\dagger(g)$ is the composite linear map gT. And of course $T^{\dagger\dagger}$ denotes $\big(T^\dagger\big)^\dagger$. The SECOND DUAL FUNCTOR $(\,\cdot\,)^{\dagger\dagger}$ is the pair of maps $V \mapsto V^{\dagger\dagger}$, $T \mapsto T^{\dagger\dagger}$ from $\mathbf{VEC}$ into itself. Because the dual functor is contravariant, the second dual functor is covariant.

On the category $\mathbf{VEC}$ of vector spaces and linear maps let I be the identity functor and $(\,\cdot\,)^{\dagger\dagger}$ be the second dual functor. The NATURAL MAP τ, where (for every vector space V)

$$\tau_V\colon V \to V^{\dagger\dagger}\colon x \mapsto \widehat{x}$$

is a natural transformation from I to $(\,\cdot\,)^{\dagger\dagger}$. It is in fact a natural equivalence.

25.3.3. Example. Recall that if B is a Banach space its *dual space* B^* is the set of all continuous linear functionals on B. On B^* addition and scalar multiplication are defined pointwise. The norm is the usual operator norm: if $f \in B^*$, then

$$\|f\| := \inf\{M > 0\colon |f(x)| \le M\|x\| \text{ for all } x \in B\} = \sup\{|f(x)|\colon \|x\| \le 1\}.$$

Under the metric induced by this norm the dual space B^* is complete and thus is itself a Banach space. If $T\colon B \to C$ is a continuous linear map between Banach spaces, $T^*\colon C^* \to B^*$ is defined as for vector spaces: $T^*(g) = g \circ T$ for every $g \in C^*$. The map T^*, called the ADJOINT of T, is a continuous linear map. The dual B^{**} of the dual B^* of a Banach space B is the SECOND DUAL of B and $T^{**} := \big(T^*\big)^*$ maps B^{**} into C^{**}. If B is a Banach space, then the map $\tau_B\colon B \to B^{**}\colon x \mapsto x^{**}$

where $x^{**}(f) = f(x)$ for all $f \in B^*$ is called the NATURAL INJECTION of B into B^{**}. (It is not obvious that this mapping is injective. We will show that it is in proposition 29.1.12.)

On the category $\mathbf{BAN}_\infty$ of Banach spaces and bounded linear transformations let I be the identity functor and $(\,\cdot\,)^{**}$ be the second dual functor. Then the *natural injection* τ is a natural transformation from I to $(\,\cdot\,)^{**}$.

25.3.4. Example. Explain exactly what is meant when people say that the isomorphism

$$\mathcal{C}(X \uplus Y) \cong \mathcal{C}(X) \times \mathcal{C}(Y)$$

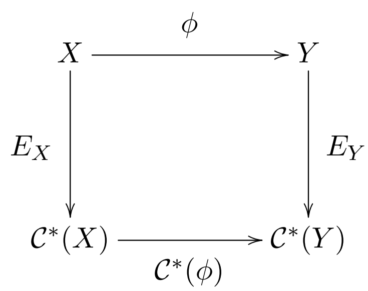
is a natural equivalence. The categories involved are **CpH** (compact Hausdorff spaces and continuous maps) and $\mathbf{BAN}_\infty$ (Banach spaces and bounded linear transformations).

25.3.5. Example. On the category **CpH** of compact Hausdorff spaces and continuous maps let I be the identity functor and let $\mathcal{C}^*$ be the functor which takes a compact Hausdorff space X to the Banach space $(C(X))^*$ and takes a continuous function $\phi\colon X \to Y$ between compact Hausdorff spaces to the contractive linear transformation $\mathcal{C}^*\phi = (\mathcal{C}\phi)^*$. Define the *evaluation map* E by

$$E_X\colon X \to \mathcal{C}^*(X)\colon x \mapsto E_X(x)$$

where $(E_X(x))(f) = f(x)$ for all $f \in \mathcal{C}(X)$.

(a) Notice that this definition does make sense. If $x \in X$, then $E_X(x)$ really does belong to $\mathcal{C}^*(X)$.
(b) Show that the map E makes the following diagram commute



The preceding example may seem a bit disappointing. It seems very much as if we are attempting to establish a natural transformation between the identity functor (on **CpH**) and the $\mathcal{C}^*$ functor. The problem, of course, is that the two functors have different target categories. This is something we will fix in chapter 30. The evaluation map $E_X\colon X \to \mathcal{C}^*(X)$ is certainly not surjective. Notice, for example, that every functional of the form $E_X(x)$ preserves multiplication (as well as addition and scalar multiplication). A consequence of this is that such functionals all live on the unit sphere of $\mathcal{C}^*(X)$ (see proposition 30.2.7). With a new topology (the so-called *weak-star topology*—see 29.2.5) on $\mathcal{C}^*(X)$ the range of E_X turns out to be a compact Hausdorff space which is naturally equivalent to X itself. Even more remarkably, we show that the range of E_X can be regarded as the maximal ideal space of the algebra $\mathcal{C}(X)$. See theorem 30.2.23.

25.3.6. Example. Let $\mathbf{M}$ be the functor defined in 25.2.9 (or the one defined in 25.2.10) and $\mathcal{C}^*$ be the functor defined in 25.3.5. Define the RADON MAP $\Theta\colon \mathbf{M} \to \mathcal{C}^*$ as follows: for every compact Hausdorff space X let

$$\Theta_X\colon \mathbf{M}(X) \to \mathcal{C}^*(X)\colon \mu \mapsto \phi_\mu$$

where $\phi_\mu(f) = \int_X f\,d\mu$ for every $f \in \mathcal{C}(X)$. Then the Radon map Θ is a natural transformation.

25.4. The Riesz Representation Theorems

25.4.1. Theorem (Riesz representation I)**.** *Let (S, μ) be a positive measure space, $1 < p < \infty$, and q be the conjugate exponent of p. If ϕ is a bounded linear functional on $\mathrm{L_p}(\mu)$ then there exists* *a unique* $g \in \mathrm{L_q}(\mu)$ *such that*

$$\phi(f) = \int_S fg\,d\mu$$

for all $f \in \mathrm{L_p}(\mu)$. *Additionally, this result holds for* $p = 1$ *provided that the positive measure space* S *is* σ*-finite.*

Proof. See [2], theorems 27.9 and 27.10; [32], theorem 9.12; and [37], theorem 6.16.

25.4.2. Corollary. *If* (S, μ) *is a* σ*-finite positive measure space,* $1 \le p < \infty$*, and* q *is the conjugate exponent of* p*, then* $\big(\mathrm{L_p}(\mu)\big)^*$ *is isometrically isomorphic to* $\mathrm{L_q}(\mu)$*.*

25.4.3. Theorem (Riesz representation II)**.** *Let* X *be a compact Hausdorff space. If* ϕ *is a positive linear functional on* $\mathcal{C}(X)$*, then there exists a unique positive regular Borel measure* μ *on* X *such that*

$$\phi(f) = \int_X f\,d\mu$$

for all $f \in \mathcal{C}(X)$*.*

Proof. See [4], theorem 4.2.10.

It would be pleasant if theorem 25.4.3 were true for locally compact Hausdorff spaces as well. And it is—*almost*. You can find an example of a (fairly pathological) locally compact Hausdorff space for which it fails in [37], chapter 2, exercise 17. One solution to the problem, the one we invoke here, is to require a little less of the representing measure than regularity. If, in the definition of "regularity", we restrict the demand of inner regularity to the open sets in the space (rather than requiring it of all Borel sets) we have what we will call a *Radon measure*. (Compare definitions 25.2.2 and 25.4.4.) Then 25.4.3 holds for all locally compact Hausdorff spaces if "regular" is replaced by "Radon" in the statement of the theorem.

A second approach is to require a little bit more of our topological space. If we adopt the attitude that all reasonably nice locally compact Hausdorff spaces (all Euclidean spaces, for example) are σ-compact (that is, a countable union of compact subspaces), and restrict our attention to those spaces, the theorem remains true. (Consult [37], theorem 2.18.)

A third approach (a disguised version of the first) is to change the definition of "regular". Hewitt and Stromberg [23] (definition 12.39) use the word "regular" (as applied to positive measures) for the concept we here call "Radon". The price paid for this strategy is a very different looking definition for *regularity of complex measures*. (Compare [23] definitions 12.39 and 20.40.) For complex measure spaces these two definitions do indeed turn out to be equivalent, but it requires a fairly substantial theorem to establish this (see [23], theorem 12.40).

25.4.4. Definition. A positive Borel measure μ on a locally compact Hausdorff space X is a Radon measure if

(i) $\mu(K) < \infty$ for every compact subset K of X,
(ii) μ is outer regular, and
(iii) μ is inner regular on every open subset of X.

It is worthwhile to keep in mind how seldom the distinction between *Radon measures* and *regular Borel measures* is required. On spaces which are countable unions of compact sets, the concepts are identical.

25.4.5. Proposition. *Let* X *be a locally compact Hausdorff space which is a countable union of compact sets. Then every Radon measure on* X *is regular.*

Proof. See [18], proposition 7.5 and corollary 7.6.

25.4.6. Theorem (Riesz representation III)**.** *Let X be a locally compact Hausdorff space. If ϕ is a positive linear functional on $\mathcal{C}_c(X)$, then there exists a unique positive Radon measure μ on X such that*

$$\phi(f) = \int_X f \, d\mu$$

for all $f \in \mathcal{C}_c(X)$.

Proof. See [2], theorem 28.3; [11], theorem 7.2.8; [18], theorem 7.2; [23], theorems 12.35 and 12.36; and [37], theorem 2.14.

25.4.7. Theorem (Riesz representation IV)**.** *Let X be a compact Hausdorff space. If ϕ is a bounded linear functional on $\mathcal{C}(X)$, then there exists a unique real regular Borel measure μ on X such that*

$$\phi(f) = \int_X f \, d\mu$$

for all $f \in \mathcal{C}(X)$ and $\|\mu\| = \|\phi\|$.

Proof. See [4], theorems 4.3.13 and 4.3.15; [8], theorems 18.3 and 18.4; and [32], theorem 9.15.

25.4.8. Corollary. *If X is a compact Hausdorff space, then $\big(\mathcal{C}(X)\big)^*$ is isometrically isomorphic to $\boldsymbol{M}(X)$.*

25.4.9. Theorem (Riesz representation V)**.** *Let X be a locally compact Hausdorff space. If ϕ is a bounded linear functional on $\mathcal{C}_0(X)$, then there exists a unique real regular Borel measure μ on X such that*

$$\phi(f) = \int_X f \, d\mu$$

for all $f \in \mathcal{C}_0(X)$ and $\|\mu\| = \|\phi\|$.

Proof. See [18], theorem 7.17; [23], theorems 20.47 and 20.48; [32], theorem 9.16;and [37], theorem 6.19.

25.4.10. Corollary. *If X is a locally compact Hausdorff space, then $\big(\mathcal{C}_0(X)\big)^*$ is isometrically isomorphic to $\boldsymbol{M}(X)$.*

25.4.11. Corollary. *If X is a locally compact Hausdorff space, then $\boldsymbol{M}(X)$ is a Banach space.*

25.4.12. Proposition. *The Radon map $\Theta\colon \boldsymbol{M} \to \mathcal{C}^*$ defined in example 25.3.6 is a natural equivalence.*

CHAPTER 26

CONNECTED SPACES

26.1. Connectedness and Path Connectedness

26.1.1. Proposition. *If X is a topological space, then the following conditions are equivalent:*

(a) *X is the union of two disjoint nonempty open sets;*
(b) *X has a nonempty proper subset which is both closed and open;*
(c) *there is a subset of X whose characteristic function is continuous but not constant; and*
(d) *the algebra $\mathcal{C}(X)$ of continuous real valued functions on X has a nontrivial idempotent.*

26.1.2. Definition. A topological space which satisfies any (hence all) of the conditions listed in the preceding proposition is DISCONNECTED. If U and V are nonempty disjoint open sets whose union is the topological space X, we say that U and V DISCONNECT X. A space is CONNECTED if it is not disconnected.

26.1.3. Example. Every discrete topological space with more than one point is disconnected.

26.1.4. Example. Every indiscrete topological space is connected.

26.1.5. Example. Every infinite set with the cofinite topology is connected. (Otherwise there would be a finite set with a finite complement.)

26.1.6. Example. The space $\mathbb{Q}$ of rational numbers (under the topology it inherits from $\mathbb{R}$) is disconnected.

26.1.7. Example. The connected subsets of $\mathbb{R}$ are the intervals.

The next proposition says that the continuous image of a connected space is connected.

26.1.8. Proposition. *If X and Y are topological spaces, X is connected, and $f\colon X \to Y$ is a continuous surjection, then Y is connected.*

26.1.9. Example. Let X be a connected topological space. Show that a function $f\colon X \to \mathbb{Q}$ is continuous if and only if it is constant.

26.1.10. Theorem (Intermediate value theorem)**.** *If $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, then the image under f of an interval is an interval.*

When is a subspace of a topological space connected? It is tempting—and hopelessly wrong—to conjecture that a subspace W of a topological space X is disconnected if and only if there exist disjoint open subsets U and V of X both of which intersect W and whose union contains W. Certainly the "if" part of this assertion is correct. The next example, however, shows that W may well be disconnected even when there do not exist such subsets U and V of X. This is the reason for proposition 26.1.12. It gives a criterion for a subspace of X to be disconnected *which does not rely on the relative topology of W*. The criterion is in terms of the topology on X.

26.1.11. Definition. Two subsets A and B of a topological space X are MUTUALLY SEPARATED if

$$\overline{A} \cap B = A \cap \overline{B} = \emptyset\,.$$

26.1.12. Example. Let $X = \{a, b, c\}$ and $\mathfrak{T} = \big\{\emptyset, \{b\}, \{a, b\}, \{b, c\}, X\big\}$. Then $\mathfrak{T}$ is a topology on X. There exists a subset W of X which is disconnected but which can*not* be disconnected by two open subsets of X. (That is, there do not exist disjoint sets U, $V \overset{\circ}{\subseteq} X$ both of which intersect W and whose union contains W.)

26.1.13. Proposition. *A subset C of a topological space X is disconnected if and only if it is the union of two nonempty mutually separated subsets of X.*

PROOF. See [48], theorem 26.5.

26.1.14. Proposition. *If $\mathfrak{S}$ is a family of connected subsets of a topological space X whose intersection is nonempty, then $\bigcup \mathfrak{S}$ is connected.*

26.1.15. Corollary. *If (C_k) is a sequence of connected subsets of a topological space and $C_n \cap C_{n+1}$ is nonempty for every $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} C_n$ is connected.*

26.1.16. Proposition. *In a topological space the closure of a connected set is connected. In fact, if $C \subseteq A \subseteq \overline{C} \subseteq X$ and C is connected, then so is A.*

26.1.17. Proposition. *A nonempty product $X = \prod_{\lambda \in \Lambda} X_\lambda$ of topological spaces is connected if and only if each X_λ is connected.*

PROOF. See [16], theorem V.1.7; [47], theorem 6.6.3; and [48], theorem 26.10.

26.1.18. Definition. Let a be a point in a topological space X. We define C_a, the COMPONENT (or CONNECTED COMPONENT) of X containing a to be the union of all connected subsets of X which contain a. Thus C_a is the largest connected subset of X containing a.

26.1.19. Proposition. *Every component of a topological space is a closed set.*

26.1.20. Proposition. *Let X be a topological space and x, $y \in X$. If $y \in C_x$, then $C_y = C_x$ and if $y \notin C_x$, then $C_y \cap C_x = \emptyset$.*

26.1.21. Example. No two of the intervals $(0, 1)$, $(0, 1]$, and $[0, 1]$ are homeomorphic.

26.1.22. Example. A closed bounded interval $[a, b]$ in $\mathbb{R}$, a closed disk in the plane, and the boundary of a disk (a circle) are all connected compact Hausdorff spaces, but no two of them are homeomorphic.

26.1.23. Proposition. *Let E be a Lebesgue measurable subset of $\mathbb{R}$ with Lebesgue measure one. For each $t \in \mathbb{R}$ let $E_t = \{x \in E : x < t\}$. Then for every $\alpha \in (0, 1)$ there exists a number $t_\alpha \in \mathbb{R}$ such that $\lambda\big(E_{t_\alpha}\big) = \alpha$.*

A concept closely related to (but stronger than) connectedness is *arcwise connectedness.*

26.1.24. Definition. A topological space X is ARCWISE CONNECTED (or PATH CONNECTED) if for every x, $y \in X$ there exists a continuous map $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$. Such a function f is an ARC (or PATH , or CURVE) connecting x to y. It is very easy to prove that arcwise connected spaces are connected (proposition 26.1.25). The converse is false (example 26.1.26). If, however, we restrict our attention to open subsets of $\mathbb{R}^n$, then the converse does hold (proposition 26.1.27).

26.1.25. Proposition. *If a topological space is arcwise connected, then it is connected.*

26.1.26. Example. Let $B = \{(x, \sin x^{-1}) : 0 < x \leq 1\}$. Then $\overline{B}$ is a connected subset of $\mathbb{R}^2$ but is not arcwise connected.

Hint for proof. Let $M = \overline{B}$. To show that M is not arcwise connected argue by contradiction. Assume there exists a continuous function $f : [0, 1] \to M$ such that $f(1) \in B$ and $f(0) \notin B$. Prove that $f^2 = \pi_2 \circ f$ is discontinuous at the point $t_0 = \sup f^{\leftarrow}(M \setminus B)$. To this end show that $t_0 \in f^{\leftarrow}(M \setminus B)$. Then, given $\delta > 0$, choose t_1 in $[0, 1]$ so that $t_0 < t_1 < t_0 + \delta$. Without loss of generality one may suppose that $f^2(t_0) \leq 0$. Show that $\big(f^1\big)^{\rightarrow}[t_0, t_1]$ is an interval containing 0 and $f^1(t_1)$ (where $f^1 = \pi_1 \circ f$). Find a point t in $[t_0, t_1]$ such that $0 < f^1(t) < f^1(t_1)$ and $f^2(t) = 1$.

26.1.27. Proposition. *Every connected open subset of* $\mathbb{R}^n$ *is arcwise connected.*

Proof. See [17], solution Q.17.7.

26.2. Disconnectedness Comes in Many Flavors

26.2.1. Definition. A topological space is TOTALLY DISCONNECTED if all its components are points. It is EXTREMALLY DISCONNECTED if the closure of every open set is open. It is ZERO-DIMENSIONAL if it has a base consisting of sets which are both closed and open.

26.2.2. Example. Recall that the indiscrete topology on a set X has only two open sets, $\emptyset$ and X. Let a set (with at least two elements) have the indiscrete topology. Is the resulting space discrete? disconnected? extremally disconnected? totally disconnected? zero-dimensional?

26.2.3. Example. Consider the three-point space of example 26.1.12. Is it T_1? Hausdorff? discrete? disconnected? extremally disconnected? totally disconnected? zero-dimensional?

26.2.4. Example. Is the set $\mathbb{Q}$ of rational numbers discrete? disconnected? extremally disconnected? totally disconnected? zero-dimensional?

26.2.5. Example. Let $V = \{1/n : n \in \mathbb{N}\}$, $W = \{0\} \cup V$, and $X = \{*\} \cup W$ (where $*$ is any element not in W). Let a subset U belong to $\mathfrak{T}$ if U is an open subset of W or if $(U \setminus \{*\}) \cup \{0\}$ is an open subset of W. Check that $\mathfrak{T}$ is a topology for X. Is the resulting topological space T_1? Hausdorff? discrete? disconnected? extremally disconnected? totally disconnected? zero-dimensional?

26.2.6. Example. The set of intervals of the form $[a, b)$ (where $a < b$) is a base for a topology on the set $\mathbb{R}$ of real numbers. The resulting topological space is sometimes called the *Sorgenfrey line*. In this space find the closure of the sets

$$\{1/n : n \in \mathbb{N}\} \qquad \text{and} \qquad \{-1/n : n \in \mathbb{N}\}\,.$$

Is the *Sorgenfrey line* T_1? Hausdorff? discrete? disconnected? extremally disconnected? totally disconnected? zero-dimensional?

26.2.7. Example. Let $\mathbb{N}$ have the cofinite topology. (The nonempty open sets are the ones with finite complement.) Is this space discrete? disconnected? extremally disconnected? totally disconnected? zero-dimensional?

26.2.8. Exercise. In topological spaces what relationships (if any) exist between the properties discrete, extremally disconnected, totally disconnected, and zero-dimensional?

26.2.9. Definition. A topological space X is SEPARABLE if it has a countable dense subset. It is FIRST COUNTABLE if each point in X has a countable neighborhood base. It is LINDELÖF if every open cover of X has a countable subcover.

26.2.10. Example. The Sorgenfrey line (see example 26.2.6 is a separable first countable, Lindelöf space which is not second countable.

26.2.11. Exercise (The Moore plane)**.** Consider $H = \{(x, y) \in \mathbb{R}^2 : y \geq 0\}$. Put a topology $\mathfrak{T}$ on H as follows. Basic neighborhoods of points (x, y) where $y > 0$ will be the usual open disks (restricted to those that lie in H, of course). A basic neighborhood of a point p on the x-axis will be a set of the form $\{p\} \cup D$ where D is an open disk in H tangent to the x-axis at p. Show that $\mathfrak{T}$ is a topology on H. Describe the operations of interior and closure in this space. Is this space separable? $\mathbb{T}_1$? Hausdorff? connected? regular? normal? first countable? Lindelöf? locally compact? Must a subspace of a separable topological space be separable? *Hint.* The question of normality of H is a little tricky. One approach is to suppose that for every irrational number p we are given an equilateral triangle T_p satisfying:

(i) one vertex of T_p is the point $(p, 0)$; and
(ii) the side of T_p opposite this vertex lies above the x-axis and is parallel to it.

For every irrational p let h_p be the height of the triangle T_p. For every natural number n let $D_n = \{p \in \mathbb{Q}^c \colon h_p \geq \frac{1}{n}\}$. Show that for some $n \in \mathbb{N}$ the closure of the set D_n (in the usual topology on $\mathbb{R}$) contains some interval $[a, b]$ (with $a < b$). Deduce from this that $\bigcup_{p \in \mathbb{Q}^c} T_p$ contains a rectangular region of the form $[a, b] \times (0, \frac{1}{n})$ (with $n \in N$ and $a < b$). What does this tell you about the possibility of separating $\mathbb{Q} \times \{0\}$ and $\mathbb{Q}^c \times \{0\}$ by open sets?

CHAPTER 27

MODES OF CONVERGENCE

27.1. Functions on Positive Measure Spaces

27.1.1. Definition. Let (S, μ) be a positive measure space, (f_n) be a sequence of measurable real valued functions on S, and $g\colon S \to \mathbb{R}$ be measurable.

(1) (i) The sequence (f_n) CONVERGES TO g ALMOST EVERYWHERE (written $f_n \to g$ (μ-a.e.) or just $f_n \to g$ (a.e.)) if $\mu(\{x \in S\colon f_n(x) \nrightarrow g(x)\}) = 0$.

(ii) The sequence (f_n) CONVERGES TO g ALMOST UNIFORMLY (written $f_n \to g$ (a.u.)) if for every $\epsilon > 0$ there exists $E \overset{m}{\subseteq} S$ such that $f_n \to g$ uniformly on E and $\mu(E^c) < \epsilon$.

(iii) The sequence (f_n) CONVERGES TO g IN MEASURE (written $f_n \to g$ (meas) or $f_n \xrightarrow{\mu} g$) if for every $\epsilon > 0$

$$\mu\big(\{x \in S\colon |f_n(x) - g(x)| \geq \epsilon\}\big) \to 0$$

as $n \to \infty$.

(iv) If (f_n) is a sequence of integrable functions and g is integrable, then the sequence (f_n) CONVERGES TO g IN MEAN (written $f_n \to g$ (mean)) if $\int_S |f_n - g|\, d\mu \to 0$ as $n \to \infty$. This is just the usual seminorm convergence in the space $\mathcal{L}_1(S, \mu)$.

27.1.2. Proposition. *Let (S, μ) be a positive measure space, (f_n) be a sequence of measurable functions on S, and g and h be measurable functions on S. If $f_n \to g$ (μ-a.e.) and $f_n \to h$ (μ-a.e.), then $g = h$ μ-a.e.*

27.1.3. Proposition. *Let S be a positive measure space. In the space of measurable functions on S almost uniform convergence implies almost everywhere convergence.*

27.1.4. Corollary. *Let (S, μ) be a positive measure space, (f_n) be a sequence of measurable functions on S, and g and h be measurable functions on S. If $f_n \to g$ (a.u.) and $f_n \to h$ (a.u.), then $g = h$ μ-a.e.*

27.1.5. Proposition. *Let (S, μ) be a positive measure space, (f_n) be a sequence of measurable functions on S, and g and h be measurable functions on S. If $f_n \xrightarrow{\mu} g$ and $f_n \xrightarrow{\mu} h$, then $g = h$ μ-a.e.*

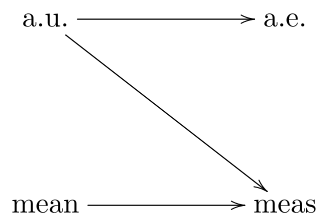
27.1.6. Proposition. *Let S be a positive measure space. In the space of measurable functions on S almost uniform convergence implies convergence in measure.*

27.1.7. Proposition. *Let S be a positive measure space. In the space of measurable functions on S convergence in mean implies convergence in measure.*

27.1.8. Corollary. *Let (S, μ) be a positive measure space, (f_n) be a sequence of measurable functions on S, and g and h be measurable functions on S. If $f_n \to g$ (mean) and $f_n \to h$ (mean), then $g = h$ μ-a.e.*

27.1.9. Example. Define a sequence (J_n) of subsets of the unit interval $[0, 1]$ as follows: $J_1 = [0, 1]$, $J_2 = [0, \frac{1}{2}]$, $J_3 = [\frac{1}{2}, 1]$, $J_4 = [\frac{2}{3}, 1]$, $J_5 = [\frac{1}{3}, \frac{2}{3}]$, $J_6 = [0, \frac{1}{3}]$, $J_7 = [0, \frac{1}{4}]$,$J_8 = [\frac{1}{4}, \frac{1}{2}]$, $J_9 = [\frac{1}{2}, \frac{3}{4}]$, $J_{10} = [\frac{3}{4}, 1]$, $J_{11} = [\frac{4}{5}, 1]$, $J_{12} = [\frac{3}{5}, \frac{4}{5}]$, $\dots$. For each $n \in \mathbb{N}$ let $f_n = \chi_{J_n}$. The functions (f_n) are sometimes called the *dancing functions.* With respect to Lebesgue measure on $\mathbb{R}$ the sequence converges in mean and in measure, but not almost uniformly or almost everywhere.

27.1.10. Example. In the following diagram the three arrows indicate the implications proved in propositions 27.1.3, 27.1.6, and 27.1.7. Altogether there are twelve possible implications. Give examples to show that none of the nine remaining implications hold.



27.1.11. Proposition. *Let S be a positive measure space. Suppose that in the space of measurable functions on S a sequence (f_n) converges in measure to a function g. Then there exists a subsequence $\big(f_{n_k}\big)$ which converges to g almost uniformly.*

Proof. See [4], theorem 2.5.3.

27.1.12. Corollary. *Let S be a positive measure space. Suppose that in the space of measurable functions on S a sequence (f_n) converges in measure to a function g. Then there exists a subsequence $\big(f_{n_k}\big)$ which converges to g almost everywhere.*

27.1.13. Corollary. *Let S be a positive measure space. Suppose that in the space of measurable functions on S a sequence (f_n) converges in mean to a function g. Then there exists a subsequence $\big(f_{n_k}\big)$ which converges to g almost uniformly.*

27.1.14. Corollary. *Let S be a positive measure space. Suppose that in the space of measurable functions on S a sequence (f_n) converges in mean to a function g. Then there exists a subsequence $\big(f_{n_k}\big)$ which converges to g almost everywhere.*

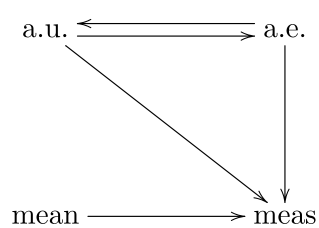
27.2. Functions on Finite Measure Spaces

27.2.1. Proposition (Egoroff's theorem)**.** *Let S be a positive real measure space. In the space of measurable functions on S convergence almost everywhere implies almost uniform convergence.*

Proof. See [2], theorem 13.7; [11], proposition 3.1.3; [18], theorem 2.33; and [23], theorem 11.32.

27.2.2. Corollary. *Let S be a positive real measure space. In the space of measurable functions on S convergence almost everywhere implies convergence in measure.*

27.2.3. Example. In the following diagram the five arrows indicate the implications proved in section 27.1, proposition 27.2.1, and corollary 27.2.2, all of which hold for finite positive measure spaces. Give examples to show that none of the seven remaining implications hold.



An interesting consequence of *Egoroff's theorem* 27.2.1 is that measurable functions on sets of finite measure are, in some sense, almost continuous. This is *Lusin's theorem.*

27.2.4. Theorem (Lusin)**.** *Let μ be a Radon measure on a locally compact Hausdorff space X and f be a measurable complex valued function whose support has finite measure. Then for every $\epsilon > 0$ there exists a continuous function $g\colon X \to \mathbb{C}$ with compact support which agrees with f except perhaps on a measurable set E such that $\mu(E) < \epsilon$. Furthermore, if f is bounded, then g may be chosen so that $\|g\|_u \le \|f\|_u$.*

PROOF. See [18], Theorem 7.10.

Lusin's theorem admits some rather far-reaching generalizations. For a sampling of these see [4], theorem 4.3.16 and corollaries 4.3.17; [11], theorem 7.4.3; [23], theorem 11.36; and [37], theorem 2.24.

27.2.5. Proposition. *Let μ be a Radon measure on a locally compact Hausdorff space X. Then $\mathcal{C}_c(X)$ is dense in $\mathrm{L_p}(X,\mu)$ for $1 \le p < \infty$.*

27.3. Dominated Convergence

27.3.1. Definition. We say that a sequence (f_n) of real (or complex) valued functions on a set S is DOMINATED by a positive function h if $|f_n| \le h$ on S.

The first result of this section is an easy consequence of the *Lebesgue dominated convergence theorem.*

27.3.2. Proposition. *Let (S,μ) be a measure space and (f_n) be a sequence of measurable functions on S which is dominated by an $\mathcal{L}_1$ function h. If (f_n) converges almost everywhere to a measurable function g, then g is integrable and (f_n) converges in mean to g.*

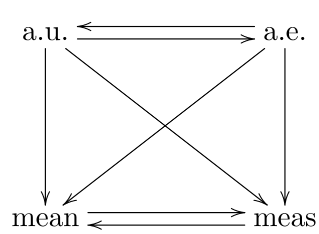
27.3.3. Proposition. *Let (S,μ) be a measure space and (f_n) be a sequence of measurable functions on S which is dominated by an $\mathcal{L}_1$ function h. If (f_n) converges almost everywhere to a measurable function g, then(f_n) converges to g almost uniformly.*

27.3.4. Proposition. *Let S be a measure space and (f_n) be a sequence of measurable functions on S which is dominated by an $\mathcal{L}_1$ function h. If (f_n) converges in measure to a measurable function g, then it converges in mean to g.*

27.3.5. Corollary. *Let S be a measure space and (f_n) be a sequence of measurable functions on S which is dominated by an $\mathcal{L}_1$ function h. If (f_n) converges almost uniformly to a measurable function g, then it converges in mean to g.*

27.3.6. Corollary. *Let S be a measure space and (f_n) be a sequence of measurable functions on S which is dominated by an $\mathcal{L}_1$ function h. If (f_n) converges to a measurable function g almost everywhere, then it converges to g in measure.*

27.3.7. Example. In the following diagram the eight arrows indicate the implications proved in this section and in section 27.1. Give examples to show that none of the four remaining implications hold (even if the space has finite measure).



27.4. Convergence in Measure is Topological

27.4.1. Proposition. *Let V be a vector space and $s\colon V \to \mathbb{R}$ be a function that satisfies:*

(i) $s(x+y) \le s(x) + s(y)$,
(ii) $s(-x) = s(x)$, *and*
(iii) $s(0) = 0$

for all x, $y \in V$. Define $\rho(x,y) = s(x-y)$ for x, $y \in V$. Then ρ is a pseudometric on V.

27.4.2. Proposition. *Let (S,μ) be a measure space and $\mathcal{M}(S)$ be the vector space of all measurable real valued functions on S. For each $f \in \mathcal{M}(S)$ and $\alpha > 0$ let $E(f,\alpha) = \{x \in S : |f(x)| \geq \alpha\}$. Define*

$$s\colon \mathcal{M}(S) \to \mathbb{R}^+ : f \mapsto \inf_{\alpha>0} \left\{\arctan\bigl(\alpha + \mu(E(f,\alpha))\bigr)\right\} .$$

Thus μ, acting through the function s, induces a topology on $\mathcal{M}(S)$.

Hint for proof. Show that $E(f+g,\alpha+\beta) \subseteq E(f,\alpha) \cup E(g,\beta)$. Use the preceding proposition 27.4.1.

The following result shows that *convergence in measure* is a topological notion.

27.4.3. Proposition. *Let (S,μ) be a measure space, (f_n) be a sequence in $\mathcal{M}(S)$, and $g \in \mathcal{M}(S)$. Then $f_n \to g$ in the topology induced on $\mathcal{M}(S)$ by μ (see the preceding proposition 27.4.2) if and only if $f_n \to g$ (meas).*

Hint for proof. Notation as in the two preceding propositions. For g, $h \in \mathcal{M}(S)$ and ϵ, $\eta > 0$ prove that

$$\rho(h,g) < \min\{\arctan\epsilon, \arctan\eta\} \implies \mu\bigl(E(h-g,\eta)\bigr) < \epsilon.$$

Prove the contrapositive. Assume that $\mu(E(h-g,\eta)) \geq \epsilon$. Show that for $\alpha > 0$

$$\arctan\bigl(\alpha + \mu(E(h-g,\alpha))\bigr) \geq \min\{\arctan\epsilon, \arctan\eta\}.$$

(Look at the cases $\alpha \leq \eta$ and $\alpha > \eta$ separately.)

CHAPTER 28

OPERATORS ON HILBERT SPACES

28.1. Orthonormal Bases

28.1.1. Definition. Let H be a Hilbert space and $A \subseteq H$. For every $F \in \operatorname{Fin} A$ define

$$s_F = \sum F.$$

Then $s = \big(s_F\big)_{F \in \operatorname{Fin} A}$ is a net in H. If this net converges, the set A is said to be SUMMABLE; the limit of the net is the SUM of A and is denoted by $\sum A$. Indexed sets require a slightly different notation. Suppose, for example, that $A = \{x_i \colon i \in I\}$ where I is an arbitrary index set. Then for each $F \in \operatorname{Fin} I$

$$s_F = \sum_{i \in F} x_i \quad \big(= \sum \{x_i \colon i \in F\}\big).$$

As above s is a net in H. If it converges $\{x_i \colon i \in I\}$ is summable and its sum is denoted by $\sum_{i \in I} x_i$ (or by $\sum \{x_i \colon i \in I\}$). An alternative way of saying that $\{x_i \colon i \in I\}$ is summable is to say that the "series" $\sum_{i \in I} x_i$ CONVERGES or that the "sum" $\sum_{i \in I} x_i$ EXISTS.

28.1.2. Remark. One frequently finds some version of the following "proposition" in textbooks.

If $\sum_{i \in I} x_i = u$ and $\sum_{i \in I} x_i = v$ in a Hilbert (or Banach) space, then $u = v$.

(I once saw a 6 line proof of this.) Of which result is this a trivial consequence?

28.1.3. Proposition. *If $\{x_i \colon i \in I\}$ and $\{y_i \colon i \in I\}$ are summable subsets of a Hilbert space and α is a scalar, then $\{\alpha x_i \colon i \in I\}$ and $\{x_i + y_i \colon i \in I\}$ are summable; and*

$$\sum_{i \in I} \alpha x_i = \alpha \sum_{i \in I} x_i$$

and

$$\sum_{i \in I} x_i + \sum_{i \in I} y_i = \sum_{i \in I} (x_i + y_i)\,.$$

28.1.4. Proposition. *If $\{x_i \colon i \in I\}$ is a summable subset of a Hilbert space H and y is a vector in H, then $\{\langle x_i, y\rangle : i \in I\}$ is a summable set of scalars and*

$$\sum_{i \in I} \langle x_i, y\rangle = \left\langle \sum_{i \in I} x_i, y \right\rangle.$$

28.1.5. Proposition. *Let H be a Hilbert space and $M \preccurlyeq H$. For every $x \in H$ there exists a unique vector $y \in M$ such that $x - y \in M^\perp$.*

28.1.6. Definition. Let H be a Hilbert space and $M \preccurlyeq H$. Define a function

$$P_M \colon H \to H \colon x \mapsto y$$

where y is the vector whose existence and uniqueness was asserted in the preceding proposition.

28.1.7. Proposition. *Let H be a Hilbert space and $M \preccurlyeq H$. Then the map P_M defined above is a (bounded linear) operator on H and $\|P_M\| \leq 1$.*

28.1.8. Proposition. *Let H be a Hilbert space and $M \preccurlyeq H$. Then ${P_M}^2 = P_M$, $\ker P_M = M^\perp$, and $\operatorname{ran} P_M = M$.*

28.1.9. Definition. A nonempty subset E of a Hilbert space is ORTHONORMAL if $e \perp f$ for every pair of distinct vectors e and f in E and $\|e\| = 1$ for every $e \in E$.

28.1.10. Proposition (Gram-Schmidt orthonormalization)**.** *If (v^k) is a linearly independent sequence in an inner product space V, then there exists an orthonormal sequence (e^k) in V such that $\operatorname{span}\{v^1, \dots, v^n\} = \operatorname{span}\{e^1, \dots, e^n\}$ for every $n \in \mathbb{N}$.*

28.1.11. Proposition. *In a Hilbert space H let P be the orthogonal projection onto the subspace M which is the closed linear span of an orthonormal set E of vectors in H. Then for every $x \in H$ we have $Px = \sum\{\langle x, e\rangle\, e \colon e \in E\}$.*

28.1.12. Proposition (Bessel's inequality)**.** *Let H be a Hilbert space and E be an orthonormal subset of H. For every $x \in H$*

$$\sum_{e \in E} \|\langle x, e\rangle\|^2 \le \|x\|^2 .$$

28.1.13. Proposition. *Let E be an orthonormal set in a Hilbert space H. Then the following are equivalent.*

(i) *E is a maximal orthonormal set.*
(ii) *If $x \perp E$, then $x = 0$ (E is* total*).*
(iii) *$\bigvee E = H$ (E is a* complete orthonormal set*).*
(iv) *$x = \displaystyle\sum_{e \in E} \langle x, e\rangle e$ for all $x \in H$.* (Fourier expansion)
(v) *$\langle x, y\rangle = \displaystyle\sum_{e \in E} \langle x, e\rangle\langle e, y\rangle$ for all $x, y \in H$.* (Parseval's identity.)
(vi) *$\|x\|^2 = \displaystyle\sum_{e \in E} |\langle x, e\rangle|^2$ for all $x \in H$.* (Parseval's identity.)

28.1.14. Definition. If H is a Hilbert space, then an orthonormal set E satisfying any one (hence all) of the conditions in proposition 28.1.13 is called an ORTHONORMAL BASIS for H (or a COMPLETE ORTHONORMAL SET in H, or a HILBERT SPACE BASIS for H). Notice that this is a very different thing from the usual (Hamel) basis for a vector space (see 5.3.12).

28.1.15. Proposition. *If E is an orthonormal subset of a Hilbert space H, then there exists an orthonormal basis for H which contains E.*

28.1.16. Example. For each $n \in \mathbb{N}$ let $\mathbf{e}^n$ be the sequence in l_2 whose n^{th} coordinate is 1 and all the other coordinates are 0. Then $\{\mathbf{e}^n \colon n \in \mathbb{N}\}$ is an orthonormal basis for l_2. This is the USUAL ORTHONORMAL BASIS for l_2.

28.1.17. Exercise. A function $f \colon \mathbb{R} \to \mathbb{C}$ of the form

$$f(t) = a_0 + \sum_{k=1}^{n} (a_k \cos kt + b_k \sin kt)$$

where $a_0, \dots, a_n, b_1, \dots, b_n$ are complex numbers is a TRIGONOMETRIC POLYNOMIAL.

(a) Explain why complex valued 2π-periodic functions on the real line $\mathbb{R}$ are often identified with complex valued functions on the unit circle $\mathbb{T}$.
(b) Some authors say that a *trigonometric polynomial* is a function of the form

$$f(t) = \sum_{k=-n}^{n} c_k e^{ikt}$$

where $c_1, \dots, c_n$ are complex numbers. Explain carefully why this is exactly the same as the preceding definition.

(c) Justify the use of the term *trigonometric polynomial.*
(d) Prove that every continuous 2π-periodic function on $\mathbb{R}$ is the uniform limit of a sequence of trigonometric polynomials.

28.1.18. Proposition. *Consider the Hilbert space $L_2([0, 2\pi], \mathbb{C})$ with inner product*

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(x)\overline{g(x)}\,dx.$$

For every integer n (positive, negative, or zero) let $\mathbf{e}^n(x) = e^{inx}$ for all $x \in [0, 2\pi]$. Then the set of these functions $\mathbf{e}^n$ for $n \in \mathbb{Z}$ is an orthonormal basis for $L_2([0, 2\pi], \mathbb{C})$.

28.1.19. Exercise. Calculate $\displaystyle\sum_{k=1}^{\infty}\frac{1}{k^2}$. *Hint.* Let $f(x) = x$ for $0 \le x \le 2\pi$. Regard f as a member of $L_2([0, 2\pi], \mathbb{C})$. Find $\|f\|$ in two ways.

28.2. Adjoints of Hilbert Space Operators

28.2.1. Proposition. *Let $S, T \in \mathfrak{B}(H, K)$ where H and K are Hilbert spaces. If $\langle Sx, y\rangle = \langle Tx, y\rangle$ for every $x \in H$ and $y \in K$, then $S = T$.*

28.2.2. Proposition. *If H is a complex Hilbert space and $T \in \mathfrak{B}(H)$ satisfies $\langle Tz, z\rangle = 0$ for all $z \in H$, then $T = 0$. The corresponding result fails in real Hilbert spaces.*

Hint for proof. In the hypothesis replace z first by $x + y$ and then by $x + iy$.

Despite our convention made at the beginning of section 20.6 that in these notes all Hilbert spaces are complex, the word "complex" was added to the hypotheses of the preceding proposition to draw attention to the fact that it is one of the few facts which holds *only* for complex spaces. While proposition 28.2.1 holds for both real and complex Hilbert spaces 28.2.2 does not. (Consider the operator which rotates the plane $\mathbb{R}^2$ by 90 degrees.)

28.2.3. Definition. A function $T\colon V \to W$ between complex vector spaces is CONJUGATE LINEAR if it is additive ($T(x + y) = Tx + Ty$) and satisfies $T(\alpha x) = \overline{\alpha}Tx$ for all $x \in V$ and $\alpha \in \mathbb{C}$. A complex valued function of two variables $\phi\colon V \times W \to \mathbb{C}$ is a SESQUILINEAR FUNCTIONAL if it is linear in its first variable and conjugate linear in its second variable.

If H and K are normed linear spaces, a sesquilinear functional ϕ on $H \times K$ is BOUNDED if there exists a constant $M > 0$ such that

$$|\phi(x, y)| \le M\|x\|\|y\|$$

for all $x \in H$ and $y \in K$.

28.2.4. Proposition. *If $\phi\colon H \oplus K \to \mathbb{C}$ is a bounded sesquilinear functional on the direct sum of two Hilbert spaces, then the following numbers (exist and) are equal:*

- $\sup\{|\phi(x, y)|\colon \|x\| \le 1, \|y\| \le 1\}$
- $\sup\{|\phi(x, y)|\colon \|x\| = \|y\| = 1\}$
- $\sup\left\{\dfrac{|\phi(x, y)|}{\|x\|\,\|y\|}\colon x, y \ne 0\right\}$
- $\inf\{M > 0\colon |\phi(x, y)| \le M\|x\|\,\|y\|$ *for all* $x,\ y \in H\}$.

Hint for proof. The proof is virtually identical to the corresponding result for linear maps (see 12.2.6).

28.2.5. Definition. Let $\phi\colon H \oplus K \to \mathbb{C}$ be a bounded sesquilinear functional on the direct sum of two Hilbert spaces. We define $\|\phi\|$, the *norm* of ϕ, to be any of the (equal) expressions in the preceding result.

28.2.6. Proposition. *Let $T\colon H \to K$ be a bounded linear map between Hilbert spaces. Then $\phi\colon H \oplus K \to \mathbb{C}\colon (x,y) \mapsto \langle Tx, y\rangle$ is a bounded sesquilinear functional on $H \oplus K$ and $\|\phi\| = \|T\|$.*

28.2.7. Proposition. *Let $\phi\colon H \times K \to \mathbb{C}$ be a bounded sesquilinear functional on the product of two Hilbert spaces. Then there exist unique bounded linear maps $T \in \mathfrak{B}(H,K)$ and $S \in \mathfrak{B}(K,H)$ such that*

$$\phi(x,y) = \langle Tx, y\rangle = \langle x, Sy\rangle$$

for all $x \in H$ and $y \in K$. Also, $\|T\| = \|S\| = \|\phi\|$.

Hint for proof. Show that for every $x \in H$ the map $y \mapsto \overline{\phi(x,y)}$ is a bounded linear functional on K.

28.2.8. Definition. Let $T\colon H \to K$ be a bounded linear map between Hilbert spaces. The mapping $(x,y) \mapsto \langle Tx, y\rangle$ from $H \oplus K$ into $\mathbb{C}$ is a bounded sesquilinear functional. By the preceding proposition there exists an unique bounded linear map $T^*\colon K \to H$ called the ADJOINT of T such that

$$\langle Tx, y\rangle = \langle x, T^*y\rangle$$

for all $x \in H$ and $y \in K$. Also, $\|T\| = \|T^*\|$.

28.2.9. Example. Recall from example 20.6.4 that the family l_2 of all square summable sequences of complex numbers is a Hilbert space. Let

$$S\colon l_2 \to l_2\colon (x_1, x_2, x_3, \dots) \mapsto (0, x_1, x_2, \dots)\,.$$

Then S is an operator on l_2,called the UNILATERAL SHIFT OPERATOR, and $\|S\| = 1$. The adjoint S^* of the unilateral shift is given by

$$S^*\colon l_2 \to l_2\colon (x_1, x_2, x_3, \dots) \mapsto (x_2, x_3, x_4, \dots).$$

In 16.2.15 we defined multiplication operators on $\mathcal{C}(X)$ where X is a topological space. Multiplication operators can be defined on many spaces; there is nothing special in this respect about $\mathcal{C}(X)$. For example, we can just as well consider multiplication operators on spaces of square integrable functions.

28.2.10. Example. Let (S,μ) be a sigma-finite measure space and $L_2(\mu)$ be the Hilbert space of all (equivalence classes of) complex valued functions on S which are square integrable with respect to μ. Let ϕ be an essentially bounded complex valued μ-measurable function on S. Define M_ϕ on $L_2(\mu)$ by $M_\phi(f) := \phi f$. Then M_ϕ is an operator on $L_2(S)$; it is called a MULTIPLICATION OPERATOR. Its norm is given by $\|M_\phi\| = \|\phi\|_\infty$ and its adjoint by ${M_\phi}^* = M_{\overline{\phi}}$.

28.2.11. Example. Let (S,μ) be a sigma-finite measure space and $L_2(\mu)$ be the Hilbert space of all (equivalence classes of) complex valued functions on S which are square integrable with respect to μ. If $k\colon S \times S \to \mathbb{C}$ is square integrable with respect to the product measure $\mu \times \mu$ on $S \times S$, define K on $L_2(\mu)$ by

$$(Kf)(x) = \int_0^1 k(x,y) f(y)\, dy\,.$$

Then K is an INTEGRAL OPERATOR and its KERNEL is the function k. The adjoint K^* of K is also an integral operator and its kernel is the function k^* defined on $S \times S$ by $k^*(x,y) = \overline{k(y,x)}$.

28.2.12. Exercise. Let V be the integral operator on $L_2([0,1])$ whose kernel is the characteristic function of $\{(x,y) \in [0,1] \times [0,1] : y < x\}$.

(a) Find a simple expression for $Vf(x)$.
(b) Show that V is an operator on H.
(c) Show that V is injective.
(d) Compute V^*.
(e) What is $V + V^*$? What is its range?

28.2.13. Proposition. *If S and T are operators on a Hilbert space H and $\alpha \in \mathbb{C}$, then*

(i) $(S+T)^* = S^* + T^*$;
(ii) $(\alpha T)^* = \overline{\alpha} T^*$;
(iii) $T^{**} = T$; *and*
(iv) $(TS)^* = S^* T^*$.

Notice that, despite the fact that every Hilbert space "is" a Banach space, the adjoint of a Hilbert space operator is defined differently from the adjoint of a Banach space operator (see 15.4.6). For an operator T between Banach spaces (ii) above becomes $(\alpha T)^* = \alpha T^*$.

28.2.14. Proposition. *Let T be an operator on a Hilbert space H. Then*

$$\|T^*T\| = \|T\|^2 .$$

28.2.15. Proposition. *If T is an operator on a Hilbert space, then*

(i) $\ker T^* = (\operatorname{ran} T)^\perp$,
(ii) $\overline{\operatorname{ran} T^*} = (\ker T)^\perp$,
(iii) $\ker T = (\operatorname{ran} T^*)^\perp$, *and*
(iv) $\overline{\operatorname{ran} T} = (\ker T^*)^\perp$.

28.2.16. Proposition. *An operator T on a Hilbert space H is self-adjoint if and only if $\langle Tx, x\rangle \in \mathbb{R}$ for every $x \in H$.*

28.2.17. Proposition. *The pair of maps $H \mapsto H$ and $T \mapsto T^*$ taking every Hilbert space to itself and every bounded linear map between Hilbert spaces to its adjoint is a contravariant functor from the category $\mathbf{HIL}_\infty$ of Hilbert spaces and bounded linear maps to itself.*

28.2.18. Example. Define an operator V on the Hilbert space $H = L_2([0,\infty))$ by

$$Vf(t) = \begin{cases} f(t-1), & \text{if } t \geq 1 \\ 0, & \text{if } 0 \leq t < 1 \end{cases} .$$

Then V is an isometry but not an isomorphism in the category $\mathbf{HIL_1}$ of Hilbert spaces and contractive linear maps. What changes if $H = L_2(\mathbb{R})$?

CHAPTER 29

OPERATORS ON BANACH SPACES

29.1. The Hahn-Banach Theorems

29.1.1. Definition. A real valued function f on a real vector space V is a SUBLINEAR FUNCTIONAL if $f(x+y) \le f(x) + f(y)$ and $f(\alpha x) = \alpha f(x)$ for all x, $y \in V$ and $\alpha \ge 0$.

29.1.2. Example. A norm on a real vector space is a sublinear functional.

29.1.3. Theorem (Hahn-Banach theorem I)**.** *Let M be a subspace of a real vector space V and p be a sublinear functional on V. If f is a linear functional on M such that $f \le p$ on M, then f has an extension $\widehat{f}$ to all of V such that $\widehat{f} \le p$ on V.*

PROOF. See [12], theorem III.6.2 and lemma III.6.9; [23], theorem 14.9; and [32], theorem 10.1.

29.1.4. Theorem (Hahn-Banach theorem II)**.** *Let M be a vector subspace of a (real or complex) normed linear space V. Then every continuous linear functional f on M has an extension $\widehat{f}$ to all of V such that $\|\widehat{f}\| = \|f\|$.*

PROOF. See [23], corollary 14.11 and theorem 14.12; [32], theorem 10.2 and corollary 10.1; and [37], theorem 5.16 and proposition 5.17.

29.1.5. Exercise. Your buddy Fred R. Dimm thinks he has discovered a really simple direct proof of the version of the Hahn-Banach theorem given in proposition 29.1.4: A continuous linear functional f defined on a (linear) subspace M of a normed linear space V can be extended to a continuous linear functional $\hat{f}$ on all of V without increasing its norm. He says that all one needs to do is let N be any vector space complement of M—that is, a linear subspace N such that, as vector spaces, $V = M \oplus N$—and let B be a (Hamel) basis for N. Define $\hat{f}(e) = 0$ for every vector $e \in B$ and $\hat{f} = f$ on M. Then extend $\hat{f}$ by linearity to all of V. Surely, Fred says, $\hat{f}$ is linear by construction and its norm has not increased since we have made it zero where it was not previously defined. Show Fred the error of his ways.

29.1.6. Theorem (Hahn-Banach theorem III)**.** *Let V be a (real or complex) normed linear space, $M \preccurlyeq V$, and $x \in V$. Then there exists a functional $f \in V^*$ such that $f^{\rightarrow}(M) = \{0\}$, $f(x) = d(x, M)$, and $\|f\| \le 1$.*

PROOF. See [23], corollary 14.13; and [32], theorem 10.3.

The most useful special case of the preceding is when $M = \{\mathbf{0}\}$ and $x \ne \mathbf{0}$. This tells us that the only way for $f(x)$ to vanish for every $f \in V^*$ is for x to be the zero vector.

29.1.7. Corollary. *Let V be a normed linear space and $x \in V$. If $f(x) = 0$ for every $f \in V^*$, then $x = \mathbf{0}$.*

29.1.8. Corollary. *If V is a normed linear space, then V^* separates points of V.*

29.1.9. Corollary. *Let V be a normed linear space and $x \in V$. Then there exists $f \in V^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$.*

29.1.10. Proposition. *Let V be a normed linear space, $\{x_1, \dots, x_n\}$ be a linearly independent subset of V, and $\alpha_1, \dots, \alpha_n$ be scalars. Then there exists an f in V^* such that $f(x_k) = \alpha_k$ for $1 \le k \le n$.*

Hint for proof. For $x = \sum_{k=1}^n \beta_k x_k$ let $g(x) = \sum_{k=1}^n \alpha_k \beta_k$.

29.1.11. Proposition. *Let V be a normed linear space and $\mathbf{0} \neq x \in V$. Then*

$$\|x\| = \max\{|f(x)|\colon f \in V^* \text{ and } \|f\| \le 1\}.$$

The use of max instead of sup in the preceding proposition is intentional. There exists an f in the closed unit ball of V^* such that $f(x) = \|x\|$.

29.1.12. Proposition. *If B is a Banach space the map $\tau_B \colon B \to B^{**} \colon x \mapsto x^{**}$ (defined in example 25.3.3) is an isometric linear map.*

29.1.13. Definition. A Banach space B is REFLEXIVE if the map τ_B is onto.

Example 25.3.3 shows that in the category of Banach spaces and bounded linear maps, the mapping $\tau \colon B \mapsto \tau_B$ is a natural transformation from the identity functor to the second dual functor. With the *Hahn-Banach theorem* we can say more.

29.1.14. Proposition. *In the category of reflexive Banach spaces and bounded linear maps the mapping $\tau \colon B \mapsto \tau_B$ is a natural equivalence between the identity and second dual functors.*

29.1.15. Example. Every Hilbert space is reflexive.

29.1.16. Exercise. Let V and W be normed linear spaces, U be a nonempty convex subset of V, and $f \colon U \to W$. Without using any form of the *mean value theorem* show that if $df = \mathbf{0}$ on U, then f is constant on U.

29.1.17. Example. Let l_∞ be the Banach space of all bounded sequences of real numbers and c be the subspace of l_∞ comprising all sequences $x = (x_n)$ such that $\lim_{n\to\infty} x_n$ exists. Define

$$f \colon c \to \mathbb{R} \colon x \mapsto \lim_{n\to\infty} x_n.$$

Then f is a continuous linear functional on c and $\|f\| = 1$. By the *Hahn-Banach theorem* f has an extension to $\widehat{f} \in {l_\infty}^*$ with $\|\widehat{f}\| = 1$. For every subset A of $\mathbb{N}$ define a sequence $\big(x_n^A\big)$ by $x_n^A = \begin{cases} 1, & \text{if } n \in A; \\ 0, & \text{if } n \notin A. \end{cases}$ Next define $\mu \colon \mathfrak{P}(\mathbb{N}) \to \mathbb{R} \colon A \mapsto \widehat{f}\big(x_n^A\big)$. Then μ is a finitely additive set function but is not countably additive.

29.2. Banach Space Duality

29.2.1. Notation. Let B be a normed linear space, $M \subseteq B$, and $F \subseteq B^*$. Then we define

$$M^\perp := \{f \in B^* \colon f(x) = 0 \text{ for all } x \in M\}$$

$$F_\perp := \{x \in B \colon f(x) = 0 \text{ for all } f \in F\}$$

The set $M^\perp$ is the ANNIHILATOR of M and $F_\perp$ is the PRE-ANNIHILATOR of F.

29.2.2. Proposition (Annihilators)**.** *Let B be a Banach space, $M \subseteq B$, and $F \subseteq B^*$. Then*

(a) *If $M \subseteq N \subseteq B$, then $N^\perp \subseteq M^\perp$.*
(b) *If $F \subseteq G \subseteq B^*$, then $G_\perp \subseteq F_\perp$.*
(c) *$M^\perp$ is a closed linear subspace of B^*.*
(d) *$F_\perp$ is a closed linear subspace of B.*
(e) *$M \subseteq {M^\perp}_\perp$.*
(f) *$F \subseteq {F_\perp}^\perp$.*
(g) *${M^\perp}_\perp = \bigvee M$.*
(h) *$M^\perp = B^*$ if and only if $M = \{0\}$.*

(i) $F_\perp = B$ *if and only if* $F = \{0\}$.
(j) *If* M *is a linear subspace of* B*, then* $M^\perp = \{0\}$ *if and only if* M *is dense in* B.
(k) *If* B *is reflexive (that is, if the natural embedding* $\tau\colon B \to B^{**}$ *is an isomorphism), then* ${F_\perp}^\perp = \bigvee F$. Hint. *Show that* $\tau(F_\perp) = F^\perp$ *and that therefore* ${F_\perp}^\perp = {F^\perp}_\perp$.

29.2.3. Proposition. *Let* $T \in \mathfrak{B}(B,C)$ *where* B *and* C *are Banach spaces. Then*

(a) $\ker T^* = (\operatorname{ran} T)^\perp$.
(b) $\ker T = (\operatorname{ran} T^*)_\perp$.
(c) $\overline{\operatorname{ran} T} = (\ker T^*)_\perp$.
(d) $\overline{\operatorname{ran} T^*} \subseteq (\ker T)^\perp$.

29.2.4. Example. The closed unit ball in l_2 is not compact.

29.2.5. Definition. Let B be a Banach space. The WEAK TOPOLOGY on B is the weak topology induced by the elements of B^*. (So, of course, the weak topology on the dual space B^* is the weak topology induced by the elements of B^{**}.) When a net (x_λ) converges weakly to a vector a in B we write $x_\lambda \xrightarrow{w} a$. The w^*-TOPOLOGY (pronounced *weak star topology*) on the dual space B^* is the weak topology induced by the elements of $\operatorname{ran}\tau$ where τ is the natural injection of B into B^{**}. When a net (f_λ) converges weakly to a vector g in B^* we write $f_\lambda \xrightarrow{w^*} g$. Notice that when B is reflexive (in particular, for Hilbert spaces) the weak and weak star topologies on B^* are identical.

29.2.6. Proposition. *Let* $f \in B^*$ *where* B *is a Banach space. For every* $\epsilon > 0$ *and every finite subset* $F \subseteq B$ *define*

$$U(f;F;\epsilon) = \{g \in B^* \colon |f(x) - g(x)| < \epsilon \text{ for all } x \in F\}.$$

The family of all such sets is a base for the w^**-topology on* B^*.

29.2.7. Theorem (Alaoglu's theorem)**.** *The closed unit ball of a Banach space is compact in the* w^**-topology.*

PROOF. See [8], theorem 15.11; [12], theorem V.3.1; and [15], theorem 1.23.

29.3. The Baire Category Theorem

As a matter of logical organization this section should probably appear in chapter 20 because of its close connection with complete metric spaces. But because the material is a bit complicated and several new concepts are introduced, I thought it better to postpone introducing the *Baire category theorem* until it is actually needed, which it is in the next section for the proof of the *open mapping theorem*.

The word "category" in the title *Baire category theorem* has nothing to do with the modern algebraic sense of the word that we introduced earlier (in chapter 8). In the older usage of the word there are two types of topological spaces: those of *first category* that can be written as a countable union of nowhere dense sets, and everything else, known as sets of *second category*. This language is neither attractive nor helpful, but it is classical. Quite a few authors (Semadeni [41], for example), apparently disliking these unsuggestive terms, have replaced *first category* and *second category* with *meager* and *non-meager*, respectively. Central to the discussion is the topological notion of *nowhere dense*.

29.3.1. Definition. A subset of a topological space is NOWHERE DENSE (or RARE) if its closure has empty interior.

29.3.2. Example. The set $\mathbb{Z}$ of integers is nowhere dense in the real line $\mathbb{R}$. The set $\mathbb{Q}$ of rational numbers fails to be nowhere dense in $\mathbb{R}$.

Notice that being nowhere dense is not an intrinsic property of a set. (Recall the discussion of *open* and *closed* in section 11.5.) A set (with a fixed topology) may be nowhere dense in one space and fail to be nowhere dense in another. If there is only one topological space under discussion, it is unambiguous to say *A is a nowhere dense set*. If, however, there is more than one space that contains A, be explicit and say *A is nowhere dense in X*, if that is what you intend.

29.3.3. Example. Let X be the x-axis in $\mathbb{R}^2$ and $A = \{(r, 0)\colon r \in \mathbb{Q}\}$. Then A is nowhere dense in $\mathbb{R}^2$ but fails to be nowhere dense in the subspace X.

Despite the preceding example we do have the following useful positive result.

29.3.4. Proposition. *Let Y be a subspace of a topological space X and $A \subseteq Y$. If A is nowhere dense in Y, then it is nowhere dense in X. Furthermore, if, additionally, Y is an open subset of X, then A is nowhere dense in Y if and only if it is nowhere dense in X.*

29.3.5. Example. The set $\{f \in \mathcal{C}([0,1])\colon |f(x) - f(y)| \le |x - y| \text{ for all } x, y \in [0,1]\}$ is nowhere dense in $\mathcal{C}([0,1])$.

Here are some elementary observations.

29.3.6. Proposition. *The following hold in any topological space.*

(a) *A set is nowhere dense if and only if its closure is.*
(b) *Every subset of a nowhere dense set is nowhere dense.*
(c) *The union of a finite family of nowhere dense sets is nowhere dense.*

The key to understanding many arguments related to the *Baire category theorem* is a realization that in general topological spaces the closed nowhere dense sets are the complements of open dense sets.

29.3.7. Proposition. *A subset of a topological space is open and dense if and only if its complement is closed and nowhere dense.*

29.3.8. Corollary. *A topological space is a countable union of nowhere dense sets if and only if there exists a countable family of open dense sets in the space whose intersection is empty.*

29.3.9. Definition. A subset of a topological space is of FIRST CATEGORY (or is MEAGER) in the space if it is a countable union of nowhere dense sets. It is of SECOND CATEGORY (or is NONMEAGER) if it is not of first category. A subset is RESIDUAL if its complement is of first category.

29.3.10. Example. The set of rational numbers, while dense in the reals, is a countable union of nowhere dense sets, so is of first category in $\mathbb{R}$.

With this new language we may rephrase corollary 29.3.8.

29.3.11. Corollary. *A topological space X is of second category (in itself) if and only if every countable family of dense open sets in X has nonempty intersection.*

Again, a reminder. Neither of these concepts is intrinsic. If, in some particular context, a set A is contained in two topological spaces, say X and Y, it is necessary to be explicit and specify that A is, say, of first category *in X* or of first category *in Y*.

29.3.12. Proposition. *Let Y be a subspace of a topological space X and $A \subseteq Y$. If A is of first category in Y, then it is of first category in X. In particular, any subspace of X which is of first category in itself is also of first category in X.*

29.3.13. Example. It is easy to see that the preceding result does not hold for second category sets. The set $\mathbb{N}$ is of second category in itself, but not in $\mathbb{R}$.

Here is a fundamental fact.

29.3.14. Proposition. *If X is a nonempty topological space, then the following are equivalent:*

(a) *Every residual subset of X is dense.*
(b) *Every nonempty open subset of X is of second category.*
(c) *Every first category subset of X has empty interior.*
(d) *The intersection of every countable family of open dense subsets of X is dense in X.*

29.3.15. Definition. A topological space is a BAIRE SPACE if the intersection of every countable family of dense open subsets is dense. Obviously, had we so chosen, we could have used (a), (b), or (c) of the preceding proposition 29.3.14 instead of (d) for this definition.

First we observe that the property of being a Baire space is G-hereditary.

29.3.16. Exercise. Every nonempty open subset of a Baire space is itself a Baire space.

29.3.17. Theorem (Baire category theorem I)**.** *Every complete metric space is a Baire space.*

PROOF. Let M be a complete metric space and (U_n) be a sequence of dense open sets. What we want to show is that if V is a nonempty open set in M, then the set $V \cap \bigcap_{n=1}^{\infty} U_n$ is nonempty. We will generate a decreasing sequence of closed sets $\overline{B_n}$, each contained in U_n, with nonempty intersection.

Since V is nonempty and open and U_1 is open and dense, we may choose an open ball B_1 whose closure is contained in $V \cap U_1$ and whose diameter is strictly less than 1. Now, since B_1 is nonempty and open and U_2 is open and dense, we may choose an open ball B_2 whose closure is contained in $V \cap U_2$ and whose diameter is strictly less than $\frac{1}{2}$. Proceed in this fashion to produce a decreasing sequence of closed sets $\overline{B_n}$ each with diameter strictly less than $\frac{1}{2^n}$ and contained in both V and U_n. By proposition 20.1.9 $\bigcap_{n=1}^{\infty} \overline{B_n}$ is nonempty. This shows that $V \cap \left(\bigcap_{n=1}^{\infty} U_n\right)$ is nonempty, as desired. □

Recall that by definition a Baire space is one in which the intersection of a sequence of dense open sets is dense and by corollary 29.3.11 a second category space is one in which the intersection of a sequence of dense open sets is nonempty. So, clearly, every nonempty Baire space is of second category in itself. The converse need not be true, however, even in metric spaces.

29.3.18. Example. The space $X = \mathbb{Q} \cup (0,1)$ (with the metric topology it inherits from $\mathbb{R}$) is of second category (in itself) but is not a Baire space.

Hint for proof. Let r_1, r_2, …be an enumeration of the rational numbers which do not lie in $(0,1)$. Consider the sets $A_n = (\mathbb{Q} \setminus \{r_n\}) \cup (0,1)$.

Making use of proposition 29.3.14 we can give another statement of the theorem.

29.3.19. Theorem (Baire category theorem II)**.** *In a complete metric space every nonempty open set is of second category.*

One of the simplest and most common statements of the theorem is entirely adequate for most purposes.

29.3.20. Theorem (Baire category theorem III)**.** *Every nonempty complete metric space is of second category.*

Clearly, propositions 29.3.7 and 29.3.14 make it possible to state Baire's theorem in many different guises. Here is one more.

29.3.21. Theorem (Baire category theorem IV)**.** *If a nonempty complete metric space is the union of a sequence of closed sets, then one of these sets has nonempty interior.*

29.3.22. Example. The set $\mathbb{Q}$ of rational numbers is not a G_δ subset of R.

Hint for proof. Consider sets of the form $\mathbb{R} \setminus \{r\}$ where r is rational.

29.3.23. Example. Let $\mathcal{F} = \{f \in \mathcal{C}([0,1]) \colon f(r) \text{ is irrational for every rational number } r \text{ in } [0,1]\}$. Then $\mathcal{F}$ is of second category in $\mathcal{C}([0,1])$.

Hint for proof. For each pair (r,s) of rational numbers with $r \in [0,1]$ and s arbitrary, consider the set $\mathcal{F}_{r,s}$ of all $f \in \mathcal{C}([0,1])$ such that $f(r) = s$.

29.3.24. Example. There exists a subset of the unit interval which is of first category but whose Lebesgue measure is 1.

Hint for proof. Can you find a closed nowhere dense subset of $[0,1]$ whose measure is $1 - 1/n$? Try Cantor-like sets.

29.3.25. Proposition. *Every proper finite dimensional subspace of a normed linear space is closed and nowhere dense.*

29.3.26. Proposition. *A Hamel basis for an infinite dimensional Banach space is uncountable.*

Hint for proof. Use the preceding proposition 29.3.25.

29.3.27. Exercise. Let l be the set of all sequences of real numbers which are eventually zero. Give l the usual pointwise vector space operations and the uniform norm. What is the cardinality of a Hamel basis for l? Explain why your answer doesn't contradict the result in the preceding proposition 29.3.26.

29.4. The Open Mapping Theorem

29.4.1. Definition. A mapping $f \colon X \to Y$ between topological spaces is OPEN if it takes open sets to open sets; that is, if $f^{\rightarrow}(U)$ is open in Y whenever U is open in X.

29.4.2. Theorem (Open mapping theorem)**.** *Every bounded linear surjection between Banach spaces is an open mapping.*

29.4.3. Example. Let $\mathcal{C}_u$ be the set of continuous real valued functions on $[0,1]$ under the uniform norm and $\mathcal{C}_1$ be the same set of functions under the L_1 norm ($\|f\|_1 = \int_0^1 |f|\, d\lambda$). Then the identity map $I \colon \mathcal{C}_u \to \mathcal{C}_1$ is a continuous linear bijection but is not open. (Why does this not contradict the *open mapping theorem*?)

29.4.4. Corollary. *Every bounded linear bijection between Banach spaces is an isomorphism.*

29.4.5. Corollary. *Every bounded linear surjection between Banach spaces is a quotient map in* **BAN**.

29.4.6. Example. Let l be the set of sequences of real numbers that are eventually zero, V be l equipped with the l_1 norm, and W be l with the uniform norm. The identity map $I \colon V \to W$ is a bounded linear bijection but is not an isomorphism.

29.4.7. Example. The mapping $T \colon \mathbb{R}^2 \to \mathbb{R} \colon (x,y) \mapsto x$ shows that although bounded linear surjections between Banach spaces map open sets to open sets they need not map closed sets to closed sets.

29.4.8. Proposition. *Let V be a vector space and suppose that $\|\cdot\|_1$ and $\|\cdot\|_2$ are norms on V whose corresponding topologies are $\mathfrak{T}_1$ and $\mathfrak{T}_2$. If V is complete with respect to both norms and if $\mathfrak{T}_1 \supseteq \mathfrak{T}_2$, then $\mathfrak{T}_1 = \mathfrak{T}_2$.*

29.4.9. Exercise. Does there exist a sequence (a_n) of complex numbers satisfying the following condition: a sequence (x_n) in $\mathbb{C}$ is absolutely summable if and only if $(a_n x_n)$ is bounded? *Hint.* Consider $T \colon l_\infty \to l_1 \colon (x_n) \mapsto (x_n/a_n)$.

29.4.10. Example. Let $\mathcal{C}^2([a,b])$ be the family of all twice continuously differentiable real valued functions on the interval $[a,b]$. Regard it as a vector space in the usual fashion. For each $f \in \mathcal{C}^2([a,b])$ define

$$\|f\| = \|f\|_u + \|f'\|_u + \|f''\|_u\,.$$

This is in fact a norm and under this norm $\mathcal{C}^2([a,b])$ is a Banach space.

29.4.11. Example. Let $\mathcal{C}^2([a,b])$ be the Banach space given in exercise 29.4.10, and p_0 and p_1 be members of $\mathcal{C}([a,b])$. Define

$$T\colon C^2([a,b]) \to C([a,b])\colon f \mapsto f'' + p_1 f' + p_0 f.$$

Then T is a bounded linear map.

29.4.12. Exercise. Consider a differential equation of the form

$$y'' + p_1\, y' + p_0\, y = q \qquad (*)$$

where p_0, p_1, and q are (fixed) continuous real valued functions on the interval $[a,b]$. Make precise, and prove, the assertion that *the solutions to* $(*)$ *depend continuously on the initial values.* You may use without proof a standard theorem: For every point c in $[a,b]$ and every pair of real numbers a_0 and a_1, there exists a unique solution of $(*)$ such that $y(c) = a_0$ and $y'(c) = a_1$. *Hint.* For a fixed $c \in [a,b]$ consider the map

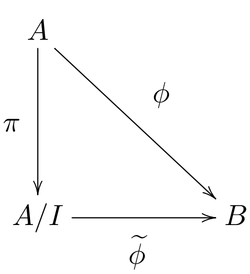
$$S\colon \mathcal{C}^2([a,b]) \to \mathcal{C}([a,b]) \times \mathbb{R}^2\colon f \mapsto \big(Tf, f(c), f'(c)\big)$$

where $\mathcal{C}^2([a,b])$ is the Banach space of example 29.4.10 and T is the bounded linear map of example 29.4.11.

29.4.13. Definition. An bounded linear map $T\colon V \to W$ between normed linear spaces is BOUNDED AWAY FROM ZERO (or BOUNDED BELOW) if there exists a number $\delta > 0$ such that $\|Tx\| \geq \delta\|x\|$ for all $x \in V$. (Recall that the word "bounded" means one thing when modifying a linear map and and something quite different when applied to a general function; so does the expression "bounded away from zero". See definition 22.3.4.)

29.4.14. Proposition. *A bounded linear map between Banach spaces is bounded away from zero if and only if it has closed range and zero kernel.*

29.4.15. Theorem (Fundamental quotient theorem for **BALG**)**.** *Let A and B be Banach algebras and I be a proper closed ideal in A. If ϕ is a homomorphism from A to B and* $\ker\phi \supseteq I$*, then there exists a unique homomorphism $\widetilde{\phi}\colon A/I \to B$ which makes the following diagram commute.*



Furthermore, $\widetilde{\phi}$ is injective if and only if $\ker\phi = I$*; and $\widetilde{\phi}$ is surjective if and only if ϕ is.*

29.5. The Closed Graph Theorem

29.5.1. Theorem (Closed graph theorem)**.** *Let $T\colon B \to C$ be a linear map between Banach spaces. If the graph of T is closed in $B \oplus C$, then T is bounded.*

HINT FOR PROOF. Let $G = \{(x, Tx)\colon x \in B\}$ be the graph of T. Apply (corollary 29.4.4 of) the *open mapping theorem* to the map

$$\pi\colon G \to B\colon (x, Tx) \mapsto x\,.$$

29.5.2. Proposition. *Let $T\colon B \to C$ be a linear mapping between Banach spaces. Then T is continuous if and only if the following condition is satisfied:*

$$\text{if } x_n \to 0 \text{ in } B \text{ and } Tx_n \to c \text{ in } C, \text{ then } c = 0.$$

29.5.3. Proposition. *Suppose $T\colon B \to C$ is a linear map between Banach spaces such that if $x_n \to 0$ in B then $(g \circ T)(x_n) \to 0$ for every $g \in C^*$. Then T is bounded.*

29.5.4. Proposition. *Let $T\colon B \to C$ be a linear function between Banach spaces. Define $T^*f = f \circ T$ for every $f \in C^*$. If T^* maps C^* into B^*, then T is continuous.*

29.5.5. Definition. A bounded linear map from a Banach space into itself is a PROJECTION if it is idempotent. If P is such a map, we say that it is a projection ALONG $\ker P$ ONTO $\operatorname{ran} P$.

29.5.6. Proposition. *An idempotent linear map from a Banach space into itself is a projection if and only if it has closed kernel and range.*

29.5.7. Proposition. *Let H be a Hilbert space and S and T be functions from H into itself such that*

$$\langle Sx, y \rangle = \langle x, Ty \rangle$$

for all x, $y \in H$. Then S and T are bounded linear operators. (Recall that in this case the operator T is the (Hilbert space) adjoint *of S and $T = S^*$.)*

29.5.8. Definition. Let $a < b$ in $\mathbb{R}$. A function $f\colon [a,b] \to \mathbb{R}$ is continuously differentiable if it is differentiable on (an open set containing) $[a,b]$ and its derivative f' is continuous on $[a,b]$. The set of all continuously differentiable real valued functions on $[a,b]$ is denoted by $\mathcal{C}^1([a,b])$.

29.5.9. Example. Let $D\colon \mathcal{C}^1([0,1]) \to C([0,1])\colon f \mapsto f'$ where both $C^1([0,1])$ and $C([0,1])$ are equipped with the uniform norm. The mapping D is linear and has closed graph but is not continuous. (Why does this not contradict the *closed graph theorem*?)

29.6. Projections and Complemented Subspaces

29.6.1. Definition. Let M be a closed linear subspace of a Banach space B. If there exists a closed linear subspace N of B such that $B = M \oplus N$, then we say that M is a COMPLEMENTED subspace, that N is its (BANACH SPACE) COMPLEMENT, and that the subspaces M and N are COMPLEMENTARY.

29.6.2. Proposition. *Let M and N be complementary subspaces of a Banach space B. Clearly, each element in B can be written uniquely in the form $m + n$ for some $m \in M$ and $n \in N$. Define a mapping $P\colon B \to B$ by $P(m + n) = m$ for $m \in M$ and $n \in N$. Then P is the projection along N onto M.*

29.6.3. Proposition. *If P is a projection on a Banach space B, then its kernel and range are complementary subspaces of B.*

29.6.4. Proposition. *Let B be a Banach space. If $P\colon B \to B$ is a projection onto a subspace M along a subspace N, then $I - P$ is a projection onto N along M.*

29.6.5. Proposition. *If M and N are complementary subspaces of a Banach space B, then B/M and N are isomorphic.*

29.6.6. Proposition. *A bounded linear map $T\colon B \to C$ between Banach spaces has a left inverse if and only if T is injective and* $\operatorname{ran} T$ *is complemented.*

29.6.7. Proposition. *Let $T\colon B \to C$ and $S\colon C \to B$ be bounded linear maps between Banach spaces. If $ST = I$, then*

(a) *S is surjective;*
(b) *T is injective;*

(c) *TS is a continuous projection along* $\ker S$ *onto* $\operatorname{ran} T$*; and*
(d) $C = \operatorname{ran} T \oplus \ker S$.

29.6.8. Exercise. Let $S\colon C \to B$ be a bounded linear map between Banach spaces. Find a necessary and sufficient condition for S to have a right inverse.

29.7. The Principle of Uniform Boundedness

29.7.1. Definition. Let S be a set and V be a normed linear space. A family $\mathcal{F}$ of functions from S into V is POINTWISE BOUNDED if for every $x \in S$ there exists a constant $M_x > 0$ such that $\|f(x)\| \le M_x$ for every $f \in \mathcal{F}$.

29.7.2. Definition. Let V and W be normed linear spaces. A family $\mathcal{T}$ of bounded linear maps from V into W is UNIFORMLY BOUNDED if there exists $M > 0$ such that $\|T\| \le M$ for every $T \in \mathcal{T}$.

29.7.3. Proposition. *Let V and W be normed linear spaces. If a family $\mathcal{T}$ of bounded linear maps from V into W is uniformly bounded, then it is pointwise bounded.*

The *principle of uniform boundedness* is the assertion that the converse of 29.7.3 holds whenever V is complete. To prove this we will make use of a "local" analog of this for continuous functions on a complete metric space.

29.7.4. Proposition. *Let M be a complete metric space and $\mathcal{F} \subseteq \mathcal{C}(M, \mathbb{R})$. If for every $x \in M$ there exists a constant $N_x > 0$ such that $|f(x)| \le N_x$ for every $f \in \mathcal{F}$, then there exists a nonempty open set U in M and a number $N > 0$ such that $|f(u)| \le N$ for every $f \in \mathcal{F}$ and every $u \in U$.*

Hint for proof. For each natural number k let $A_k = \bigcap\{f^{\leftarrow}\big([-k,k]\big)\colon f \in \mathcal{F}\}$. Conclude from the *Baire category theorem* 29.3.21 that at least one A_N has nonempty interior. Let $U = {A_N}^o$.

29.7.5. Theorem (Principle of uniform boundedness). *Let B be a Banach space and W be a normed linear space. If a family $\mathfrak{T}$ of bounded linear maps from B into W is pointwise bounded, then it is uniformly bounded.*

Hint for proof. For every $T \in \mathfrak{T}$ define

$$f_T\colon B \to \mathbb{R}\colon x \mapsto \|Tx\|.$$

Use the preceding proposition to show that there exist a nonempty open subset U of B and a constant $M > 0$ such that $f_T(x) \le M$ for every $T \in \mathfrak{T}$ and every $x \in U$. Choose a point $a \in B$ and a number $r > 0$ so that $\overline{B_r(a)} \subseteq U$. Then verify that

$$\|Ty\| \le r^{-1}\big(\, f_T(a + ry) + f_T(a)\,\big) \le 2Mr^{-1}$$

for every $T \in \mathfrak{T}$ and every $y \in B$ such that $\|y\| \le 1$.

29.7.6. Theorem (Banach-Steinhaus). *Let B be a Banach space, W be a normed linear space, and (T_n) be a sequence of bounded linear maps from B into W. If the pointwise limit $\lim_{n\to\infty} T_n x$ exists for every x in B, then the map $S\colon B \to W\colon x \mapsto \lim_{n\to\infty} T_n x$ is a bounded linear transformation.*

29.7.7. Definition. A subset A of a normed linear space V is WEAKLY BOUNDED (w-bounded) if $f^{\rightarrow}(A)$ is a bounded subset of the scalar field of V for every $f \in V^*$.

29.7.8. Proposition. *A subset of a normed linear space is bounded if and only if it is weakly bounded.*

29.7.9. Exercise. Your good friend Fred R. Dimm needs help again. This time he is worried about the assertion (see 29.7.8) that a subset of a normed linear space is bounded if and only if it is weakly bounded. He is considering the sequence (a_n) in the Hilbert space l^2 defined by $a_n = \sqrt{n}\,\mathbf{e}^n$ for each $n \in \mathbb{N}$, where $\{\mathbf{e}^n\colon n \in \mathbb{N}\}$ is the usual orthonormal basis for l_2 (see example 28.1.16). He sees that it is obviously not bounded in the usual (norm) topology on l^2. But it looks to him as if it is weakly bounded. He argues that there is no sequence $x \in l^2$ such that the set $\{|\langle a_n, x\rangle|\colon n \in \mathbb{N}\}$

is unbounded, because such a sequence would have to decrease more slowly than the sequence $\big(n^{-1/2}\big)$, which already decreases too slowly to belong to l^2. Put Fred's mind to rest by finding him a suitable sequence.

29.7.10. Definition. A sequence (x_n) in a normed linear space V is WEAKLY CAUCHY if the sequence $(f(x_n))$ is Cauchy for every f in V^*. The sequence CONVERGES WEAKLY to a point $a \in V$ if $f(x_n) \to f(a)$ for every $f \in V^*$ (that is, if it converges in the weak topology induced by the members of V^*). In this case we write $x_n \xrightarrow{w} a$.

29.7.11. Proposition. *In a normed linear space every weakly Cauchy sequence is bounded.*

29.7.12. Proposition. *Let $\mathcal{E}$ be an orthonormal basis for a Hilbert space H and (x_n) be a sequence in H. Then $x_n \xrightarrow{w} \mathbf{0}$ if and only if (x_n) is bounded and $\langle x_n, e\rangle \to 0$ for every $e \in \mathcal{E}$.*

29.7.13. Definition. Let H and K be Hilbert spaces. A sequence (T_n) in $\mathfrak{B}(H,K)$ CONVERGES WEAKLY (or CONVERGES IN THE WEAK OPERATOR TOPOLOGY) to $B \in \mathfrak{B}(H,K)$ if

$$\langle T_n x, y\rangle \to \langle Bx, y\rangle$$

for every $x \in H$ and every $y \in K$. In this case we write $T_n \xrightarrow{\textbf{WOT}} B$.

A sequence (T_n) in $\mathfrak{B}(H,K)$ CONVERGES STRONGLY (or CONVERGES IN THE STRONG OPERATOR TOPOLOGY) to $B \in \mathfrak{B}(H,K)$ if

$$T_n x \to Bx$$

for every $x \in H$. In this case we write $T_n \xrightarrow{\textbf{SOT}} B$.

A family $\mathfrak{T} \subseteq \mathfrak{B}(H,K)$ is WEAKLY BOUNDED (or BOUNDED IN THE WEAK OPERATOR TOPOLOGY) if for every $x \in H$ and every $y \in K$ there exists a positive constant $\alpha_{x,y}$ such that

$$|\langle Tx, y\rangle| \le \alpha_{x,y}$$

for every $T \in \mathfrak{T}$.

29.7.14. Proposition. *Let H and K be Hilbert spaces, (T_n) be a sequence in $\mathfrak{B}(H,K)$, and $B \in \mathfrak{B}(H,K)$. then the following implications hold:*

$$T_n \to B \quad \Longrightarrow \quad T_n \xrightarrow{\textbf{SOT}} B \quad \Longrightarrow \quad T_n \xrightarrow{\textbf{WOT}} B.$$

29.7.15. Proposition. *Let (S_n) and (T_n) be sequences of operators on a Hilbert space H. If $S_n \xrightarrow{\textbf{WOT}} A$ and $T_n \xrightarrow{\textbf{SOT}} B$, then $S_nT_n \xrightarrow{\textbf{WOT}} AB$.*

Hint for proof. You may find it helpful to prove first that every weakly convergent sequence of Hilbert space operators is bounded.

29.7.16. Example. In the preceding proposition the hypothesis $T_n \xrightarrow{\textbf{SOT}} B$ cannot be replaced by $T_n \xrightarrow{\textbf{WOT}} B$.

CHAPTER 30

THE GELFAND TRANSFORM

30.1. The Spectrum

In this chapter we will, for the first time, assume, unless otherwise stated, that vector spaces and algebras have complex, rather than real, scalars. A good part of the reason for this is the necessity of using *Liouville's theorem* 30.1.7 to prove proposition 30.1.8.

30.1.1. Proposition. *Let a be an element of a unital algebra such that $a^2 = \mathbf{1}$. Then either*

(i) *$a = \mathbf{1}$, in which case $\sigma(a) = \{1\}$, or*
(ii) *$a = -\mathbf{1}$, in which case $\sigma(a) = \{-1\}$, or*
(iii) *$\sigma(a) = \{-1, 1\}$.*

Hint for proof. In (iii) to prove $\sigma(a) \subseteq \{-1, 1\}$, consider $\dfrac{1}{1-\lambda^2}(a + \lambda\mathbf{1})$.

30.1.2. Proposition. *Recall that an element a of an algebra is* IDEMPOTENT *if $a^2 = a$. Let a be an idempotent element of a unital algebra. Then either*

(i) *$a = \mathbf{1}$, in which case $\sigma(a) = \{1\}$, or*
(ii) *$a = 0$, in which case $\sigma(a) = \{0\}$, or*
(iii) *$\sigma(a) = \{0, 1\}$.*

Hint for proof. In (iii) to prove $\sigma(a) \subseteq \{0, 1\}$, consider $\dfrac{1}{\lambda-\lambda^2}\big(a + (\lambda-1)\mathbf{1}\big)$.

30.1.3. Proposition. *Let a be an element of a unital Banach algebra A. Then the spectrum of a is compact and $|\lambda| \le \|a\|$ for every $\lambda \in \sigma(a)$.*

Hint for proof. Use the *Heine-Borel theorem.* To prove that the spectrum is closed notice that $(\sigma(a))^c = f^{\leftarrow}(\operatorname{inv} A)$ where $f(\lambda) = a - \lambda\mathbf{1}$ for every complex number λ. Also show that if $|\lambda| > \|a\|$, then $\mathbf{1} - \lambda^{-1}a$ is invertible.

30.1.4. Definition. Let a be an element of a unital Banach algebra. The RESOLVENT MAPPING for a is defined by

$$R_a \colon \mathbb{C} \setminus \sigma(a) \to A \colon \lambda \mapsto (a - \lambda\mathbf{1})^{-1}\,.$$

30.1.5. Definition. Let $U \overset{\circ}{\subseteq} \mathbb{C}$ and A be a unital Banach algebra. A function $f \colon U \to A$ is ANALYTIC on U if

$$f'(a) := \lim_{z\to a} \frac{f(z) - f(a)}{z - a}$$

exists for every $a \in U$. A complex valued function which is analytic on all of $\mathbb{C}$ is an ENTIRE function.

30.1.6. Proposition. *For a an element of a unital Banach algebra A and ϕ a bounded linear functional on A let $f := \phi \circ R_a \colon \mathbb{C} \setminus \sigma(a) \to \mathbb{C}$. Then*

(i) *f is analytic on its domain, and*
(ii) *$f(\lambda) \to 0$ as $|\lambda| \to \infty$.*

Hint for proof. For (i) notice that in a unital algebra $a^{-1} - b^{-1} = a^{-1}(b-a)b^{-1}$.

30.1.7. Theorem (Liouville)**.** *Every bounded entire function on $\mathbb{C}$ is constant.*

Comment on proof. The proof of this theorem requires a little background in the theory of analytic functions, material that is not covered in these notes. The theorem would surely be covered in any first course in complex variables. You can find a nice proof in [37], theorem 10.23.

30.1.8. Proposition. *The spectrum of every element of a unital Banach algebra is nonempty.*

Hint for proof. Argue by contradiction. Use *Liouville's theorem* 30.1.7 to show that $\phi \circ R_a$ is constant for every bounded linear functional ϕ on A. Then use (corollary 29.1.7 to) the *Hahn-Banach theorem* to prove that R_a is constant. Why must this constant be 0?

Note. It is important to keep in mind that we are working only with *complex* algebras. This result is *false* for real Banach algebras. An easy counterexample is the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ regarded as an element of the (real) Banach algebra M_2 of all 2×2 matrices of real numbers.

30.1.9. Theorem (Gelfand-Mazur)**.** *If A is a unital Banach algebra in which every nonzero element is invertible, then A is isometrically isomorphic to $\mathbb{C}$.*

Hint for proof. Let $B = \{\lambda\mathbf{1} \colon \lambda \in \mathbb{C}\}$. Use the preceding proposition 30.1.8) to show that $B = A$.

The following is an immediate consequence of the *Gelfand-Mazur theorem* 30.1.9 and proposition 21.6.4.

30.1.10. Corollary. *If I is a maximal ideal in a commutative unital Banach algebra A, then A/I is isometrically isomorphic to $\mathbb{C}$.*

30.1.11. Proposition. *Let a be an element of a unital algebra. Then*

$$\sigma(a^n) = [\sigma(a)]^n$$

for every $n \in \mathbb{N}$. (The notation $[\sigma(a)]^n$ means $\{\lambda^n \colon \lambda \in \sigma(a)\}$.)

30.1.12. Definition. Let a be an element of a unital algebra. The SPECTRAL RADIUS of a, denoted by $\rho(a)$, is defined to be $\sup\{|\lambda| \colon \lambda \in \sigma(a)\}$.

30.1.13. Theorem (Spectral radius formula)**.** *If a is an element of a unital Banach algebra, then*

$$\rho(a) = \inf\left\{\|a^n\|^{1/n} \colon n \in \mathbb{N}\right\} = \lim_{n\to\infty} \|a^n\|^{1/n}.$$

30.1.14. Proposition. *Let a be an invertible element of a unital Banach algebra. Then a complex number λ belongs to the spectrum of a if and only if $1/\lambda$ belongs to the spectrum of a^{-1}.*

30.1.15. Proposition. *Let A be a unital Banach algebra and B a closed subalgebra containing $\mathbf{1}_A$. Then*

1. $\operatorname{inv} B$ *is both open and closed in* $B \cap \operatorname{inv} A$*;*
2. $\sigma_A(b) \subseteq \sigma_B(b)$ *for every* $b \in B$*; and*
3. *if* $b \in B$ *and* $\sigma_A(b)$ *has no holes (that is, if its complement in* $\mathbb{C}$ *is connected), then* $\sigma_A(b) = \sigma_B(b)$.

30.2. Characters

30.2.1. Definition. A CHARACTER (or NONZERO MULTIPLICATIVE LINEAR FUNCTIONAL) on an algebra A is a nonzero homomorphism from A into $\mathbb{C}$. The set of all characters on A is denoted by ΔA.

30.2.2. Proposition. *Let A be a unital algebra and $\phi \in \Delta A$. Then*

(a) $\phi(\mathbf{1}) = 1$*;*
(b) *if* $a \in \operatorname{inv} A$*, then* $\phi(a) \neq 0$*;*
(c) *if a is* NILPOTENT *(that is, if $a^n = 0$ for some $n \in \mathbb{N}$), then $\phi(a) = 0$;*

(d) *if a is idempotent, then $\phi(a)$ is 0 or 1; and*
(e) *$\phi(a) \in \sigma(a)$ for every $a \in A$.*

Why is part (e) of the preceding proposition not an easy way to show that spectrum of an element of a unital algebra is nonempty?

30.2.3. Example. The algebra $M_2(\mathbb{C})$ of 2×2 matrices of complex numbers has no characters.

Hint for proof. Square the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

30.2.4. Exercise. Let A be the algebra of all 2×2 matrices $[a_{ij}]$ of complex numbers whose upper right entry a_{12} is zero. Find the set of characters and the set of maximal ideals of A.

Hint for solution. Notice that

$$\begin{bmatrix} a & 0 \\ b & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + (c-a) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Use the preceding proposition.

30.2.5. Proposition. *Let A be a unital algebra and ϕ be a linear functional on A. Then $\phi \in \Delta A$ if and only if $\ker \phi$ is closed under multiplication and $\phi(\mathbf{1}) = 1$.*

Hint for proof. For the converse apply ϕ to the product of $a - \phi(a)\mathbf{1}$ and $b - \phi(b)\mathbf{1}$ for $a, b \in A$.

30.2.6. Example. The only character on the Banach algebra $\mathbb{C}$ is the identity map on $\mathbb{C}$

30.2.7. Proposition. *Every multiplicative linear functional on a unital Banach algebra A is continuous. In fact, if $\phi \in \Delta(A)$, then $\|\phi\| = 1$.*

30.2.8. Proposition. *Let X be a topological space and $x \in X$. Then the evaluation functional at x*

$$E_X(x) \colon \mathcal{C}(X, \mathbb{C}) \to \mathbb{C} \colon f \mapsto f(x)$$

is a character on $\mathcal{C}(X, \mathbb{C})$.

30.2.9. Definition. Let A be a unital Banach algebra. We have shown that $\|\phi\| = 1$ for every character ϕ on A. That is, the set ΔA of characters is contained in the unit sphere $S_1(\mathbf{0})$ of the dual A^* of the Banach algebra A. Thus we may (and do) regard ΔA as possessing the relative w^*-topology inherited from A^*. This is the GELFAND TOPOLOGY on ΔA. The topological space which results from endowing the set ΔA with the Gelfand topology is called the CHARACTER SPACE or the CARRIER SPACE or (for reasons that will soon become apparent) the MAXIMAL IDEAL SPACE of A.

30.2.10. Proposition. *The character space of a unital Banach algebra is a compact Hausdorff space.*

Hint for proof. Use *Alaoglu's theorem* 29.2.7.

30.2.11. Proposition. *If $\phi \in \Delta A$ where A is a unital algebra, then $\ker \phi$ is a maximal ideal in A.*

Hint for proof. To show maximality, suppose I is an ideal in A which properly contains $\ker \phi$. Choose $z \in I \setminus \ker \phi$. Consider the element $x - \big(\phi(x)/\phi(z)\big)\, z$ where x is an arbitrary element of A.

30.2.12. Proposition. *A character on a unital algebra is completely determined by its kernel.*

Hint for proof. Let a be an element of the algebra and ϕ be a character. For how many complex numbers λ can $a^2 - \lambda a$ belong to the kernel of ϕ?

30.2.13. Corollary. *If A is a unital algebra, then the map $\phi \mapsto \ker \phi$ from ΔA to $\operatorname{Max} A$ is injective.*

30.2.14. Exercise. Let I be a maximal ideal in a commutative unital Banach algebra A. Then there exists a character on A whose kernel is I.

Hint for proof. Why can we think of the quotient map as a character?

30.2.15. Corollary. *If A is a unital algebra, then the map $\phi \mapsto \ker\phi$ is a bijection from ΔA onto $\operatorname{Max} A$.*

30.2.16. Definition. Let A be a unital commutative Banach algebra. In light of the preceding corollary we can give $\operatorname{Max} A$ a topology under which it is homeomorphic to the character space ΔA. This is the MAXIMAL IDEAL SPACE of A. Since ΔA and $\operatorname{Max} A$ are homeomorphic it is common practice to identify them.

30.2.17. Proposition. *Let X be a compact Hausdorff space. Then the evaluation map on X*

$$E_X\colon X \to \Delta\mathcal{C}(X,\mathbb{C})\colon x \mapsto E_X(x)$$

is a homeomorphism. Thus

$$X \approx \Delta\mathcal{C}(X,\mathbb{C}) \approx \operatorname{Max}\mathcal{C}(X,\mathbb{C})\,.$$

30.2.18. Proposition. *Let A and B be unital commutative Banach algebras and $T\colon A \to B$ be an algebra homomorphism. Define ΔT on ΔB by*

$$\Delta T(\psi) = \psi \circ T$$

for all $\psi \in \Delta B$. Then

(a) *ΔT maps ΔB into ΔA.*
(b) *The map $\Delta T\colon \Delta B \to \Delta A$ is continuous.*
(c) *If T is surjective, then ΔT is injective.*
(d) *If T is an (algebra) isomorphism, then ΔT is a homeomorphism.*
(e) *The pair of maps $A \mapsto \Delta A$ and $T \mapsto \Delta T$ is a contravariant functor from the category of unital commutative Banach algebras and algebra homomorphisms to the category of compact Hausdorff spaces and continuous maps.*
(f) *If A and B are (algebraically) isomorphic, then ΔA and ΔB are homeomorphic.*

30.2.19. Proposition. *Let X and Y be compact Hausdorff spaces. If $\mathcal{C}(X,\mathbb{C})$ and $\mathcal{C}(Y,\mathbb{C})$ are algebraically isomorphic, then X and Y are homeomorphic.*

30.2.20. Corollary. *Two compact Hausdorff spaces are homeomorphic if and only if their algebras of continuous complex valued functions are isomorphic.*

30.2.21. Corollary. *Let X and Y be compact Hausdorff spaces. If $\mathcal{C}(X,\mathbb{C})$ and $\mathcal{C}(Y,\mathbb{C})$ are algebraically isomorphic, then they are isometrically isomorphic.*

30.2.22. Example. In example 24.3.10 it was pointed out that the Banach space $L_1(\mathbb{R},\lambda)$ can be made into a Banach algebra using convolution as "multiplication". The sequence space $l_1(\mathbb{Z})$ is a Banach space. If we replace $\mathbb{R}$ by $\mathbb{Z}$ and Lebesgue measure λ with counting measure, we then make $l_1(\mathbb{Z})$ into a Banach algebra.

(a) Explain in more elementary terms (that is, without explicitly mentioning the word "measure") how to make the Banach space $l_1(\mathbb{Z})$ into a commutative Banach algebra. Verify your assertions.
(b) For each $n \in \mathbb{Z}$ let e^n be the sequence in $l_1(\mathbb{Z})$ that is 1 in the n^{th} coordinate and 0 elsewhere. What is $e^n * e^m$? Is $l_1(\mathbb{Z})$ a unital algebra?
(c) The maximal ideal space of the Banach algebra $l_1(\mathbb{Z})$ is the unit circle $\mathbb{T}$.

Hint for proof. To show that $\mathbb{T} \approx \Delta l_1(\mathbb{Z})$ define

$$\psi_z\colon l_1(\mathbb{Z}) \to \mathbb{C}\colon a \mapsto \sum_{k=-\infty}^{\infty} a_k z^k\,.$$

Show that $\psi_z \in \Delta l_1(\mathbb{Z})$. Then show that the map

$$\psi\colon \mathbb{T} \to \Delta l_1(\mathbb{Z})\colon z \mapsto \psi_z$$

is a homeomorphism.

Next we will fulfill a promise made in example 25.3.5 and show that the range of the evaluation map E_X is naturally homeomorphic to X itself whenever X is a compact Hausdorff space.

30.2.23. Theorem. *The mapping E_X which takes compact Hausdorff spaces to their corresponding evaluation maps is a natural equivalence between the identity functor and the functor $\Delta\mathcal{C}$ in the category of compact Hausdorff spaces and continuous maps. That is, the following diagram commutes and each vertical arrow is a homeomorphism.*

$$\begin{array}{ccc} X & \xrightarrow{\quad \phi \quad} & Y \\ {\scriptstyle E_X}\big\downarrow & & \big\downarrow{\scriptstyle E_Y} \\ \Delta\mathcal{C}(X) & \xrightarrow[\Delta\mathcal{C}(\phi)]{} & \Delta\mathcal{C}(Y) \end{array}$$

(Note: in the preceding diagram $\mathcal{C}(X)$ and $\mathcal{C}(Y)$ are algebras of complex valued continuous functions.)

30.2.24. Remark. Thus a compact Hausdorff space X and its character space $\Delta\mathcal{C}(X)$ are not only homeomorphic, they are *naturally homeomorphic*. This provides strong justification for the (very common) informal assertion that the maximal ideals of the Banach algebra $\mathcal{C}(X)$ “are” just the points of X.

30.3. C^*-algebras

30.3.1. Definition. An INVOLUTION on an algebra A is a map $x \mapsto x^*$ from A into A which satisfies

(i) $(x+y)^* = x^* + y^*$,
(ii) $(\alpha x)^* = \overline{\alpha}\, x^*$,
(iii) $x^{**} = x$, and
(iv) $(xy)^* = y^* x^*$

for all $x, y \in A$ and $\alpha \in \mathbb{E}$. An algebra on which an involution has been defined is a *-ALGEBRA (pronounced “star algebra”). An algebra homomorphism ϕ between *-algebras which preserves involution (that is, such that $\phi(a^*) = (\phi(a))^*$) is a *-HOMOMORPHISM (pronounced “star homomorphism”. A *-homomorphism $\phi\colon A \to B$ between unital algebras is said to be UNITAL if $\phi(\mathbf{1}_A) = \mathbf{1}_B$.

30.3.2. Example. Complex conjugation is an involution on the algebra $\mathbb{C}$ of complex numbers.

30.3.3. Example. Let X be a topological space. Then complex conjugation is an involution on the algebra $\mathcal{C}(X, \mathbb{C})$ of complex valued functions on X.

30.3.4. Example. On the algebra $M_n(\mathbb{C})$ of $n \times n$ matrices of complex numbers the operation of *conjugate transpose* is an involution. (The CONJUGATE TRANSPOSE of the matrix $[a_{ij}]$ is the matrix $[b_{ij}]$ where $b_{ij} = \overline{a_{ji}}$ for all i, $j \in \mathbb{N}_n$.)

30.3.5. Example. On the algebra $\mathfrak{B}(H)$ of operators on a Hilbert space the operation $T \mapsto T^*$ of taking adjoints is an involution.

30.3.6. Proposition. *Let a and b be elements of a *-algebra. Then a commutes with b if and only if a^* commutes with b^*.*

30.3.7. Proposition. *In a unital* **-algebra* $\mathbf{1}^* = \mathbf{1}$.

30.3.8. Proposition. *Let a be an element of a unital* **-algebra. Then a^* is invertible if and only if a is. And when a is invertible we have*

$$(a^*)^{-1} = \left(a^{-1}\right)^* .$$

30.3.9. Proposition. *Let a be an element of a unital* **-algebra. Then $\lambda \in \sigma(a)$ if and only if $\overline{\lambda} \in \sigma(a^*)$.*

30.3.10. Definition. An element a of a *-algebra A is SELF-ADJOINT (or HERMITIAN) if $a^* = a$. It is NORMAL if $a^*a = aa^*$. And it is UNITARY if $a^*a = aa^* = \mathbf{1}$. The set of all self-adjoint elements of A is denoted by $\mathcal{H}(A)$, the set of all normal elements by $\mathcal{N}(A)$, and the set of all unitary elements by $\mathcal{U}(A)$.

30.3.11. Proposition. *Let a be an element of a* **-algebra. Then there exist unique self-adjoint elements u and v such that $a = u + iv$.*

Hint for proof. Think of the algebra elements as being "like" complex numbers and self-adjoint elements as being "like" real numbers.

30.3.12. Definition. A C^*-ALGEBRA is a Banach algebra A with involution which satisfies

$$\|a^*a\| = \|a\|^2$$

for every $a \in A$.

30.3.13. Example. The set $\mathbb{C}$ of complex numbers is a C^*-algebra. (Use the usual algebraic operations, absolute value for norm, and complex conjugation for involution.)

30.3.14. Example. Let X be a nonempty locally compact Hausdorff space. Then, under the usual pointwise operations, uniform norm, and complex conjugation for involution, $\mathcal{C}_0(X, \mathbb{C})$ is a commutative C^*-algebra, which does not necessarily have an identity. If X is compact, then the C^*-algebra $\mathcal{C}_0(X, \mathbb{C}) = \mathcal{C}(X, \mathbb{C})$ is unital.

30.3.15. Example. Let H be a Hilbert space. Then $\mathfrak{B}(H)$ is a C^*-algebra. (Use composition for multiplication, the operator norm, and adjoints for involution. See proposition 28.2.14.)

30.3.16. Proposition. *In a C^*-algebra involution is an isometry. That is, $\|a^*\| = \|a\|$ for every element a of the algebra.*

30.3.17. Proposition. *In a unital C^*-algebra $\|\mathbf{1}\| = 1$.*

30.3.18. Proposition. *If a is a normal element of a unital C^*-algebra, then $\|a^2\| = \|a\|^2$ and consequently $\rho(a) = \|a\|$.*

30.3.19. Corollary. *There is at most one norm on a* **-algebra under which it is a C^*-algebra.*

30.3.20. Proposition. *A unital* **-homomorphism $\phi\colon A \to B$ between unital C^*-algebras takes invertible elements of A to invertible elements of B. Thus, $\sigma(\phi(a)) \subseteq \sigma(a)$ for every $a \in A$.*

30.3.21. Proposition. *Every unital* **-homomorphism between unital C^*-algebras is contractive (and therefore continuous).*

30.4. The Gelfand-Naimark Theorem

30.4.1. Definition. Let A be a commutative unital Banach algebra and $a \in A$. We define $\Gamma_A a\colon \Delta A \to \mathbb{C}$ by $\Gamma_A a(\tau) = \tau(a)$ for every $\tau \in \Delta(A)$. (Alternative notations: we frequently write Γa or $\hat{a}$ for $\Gamma_A a$.) Since $\Delta A \subseteq A^*$ it is clear that $\Gamma_A a$ is just the restriction of a^{**} to the character space of A. Recall that the Gelfand topology on ΔA is the relative w^*-topology; that is, the weakest topology under which a^{**} is continuous on ΔA for each $a \in A$. Thus $\Gamma_A a$ is a continuous function on ΔA. The map $\Gamma_A : A \to \mathcal{C}(\Delta A, \mathbb{C})$ is the GELFAND TRANSFORM OF A. (The element $\Gamma_A a = \Gamma a = \hat{a}$

is called the *Gelfand transform of a*—because trying repeatedly to say *the Gelfand transform of A evaluated at a* is unpleasant.) Since ΔA is a compact Hausdorff space $\mathcal{C}(\Delta A, \mathbb{C})$ is a unital Banach algebra.

30.4.2. Proposition. *If A is a unital commutative Banach algebra, then $\Gamma_A \colon A \to \mathcal{C}(\Delta A, \mathbb{C})$ is a contractive algebra homomorphism having norm one. Furthermore $\Gamma_A \mathbf{1}_A$ is the constant function $\mathbf{1}$ on ΔA and the range of Γ_A is a separating subalgebra of $\mathcal{C}(\Delta A, \mathbb{C})$.*

30.4.3. Proposition. *Let a be an element of a commutative unital Banach algebra. Then a is invertible in A if and only if $\hat{a}$ is invertible in $\mathcal{C}(\Delta A, \mathbb{C})$.*

30.4.4. Proposition. *Let a be an element of a commutative unital Banach algebra A. Then the following are equivalent:*

(a) $\lim \|a^n\|^{1/n} = 0$;
(b) $\rho(a) = 0$;
(c) $\sigma(a) = \{0\}$;
(d) $\Gamma_A a = 0$;
(e) $\tau(a) = 0$ *for every* $\tau \in \Delta A$;
(f) $a \in \bigcap \operatorname{Max} A$.

30.4.5. Proposition. *Let A be a commutative unital Banach algebra. Then the following are equivalent:*

(a) *if* $\lim \|a^n\|^{1/n} = 0$, *then* $a = 0$;
(b) *if* $\rho(a) = 0$, *then* $a = 0$;
(c) *if* $\sigma(a) = \{0\}$, *then* $a = 0$;
(d) Γ_A *is a monomorphism;*
(e) *if* $\tau(a) = 0$ *for every* $\tau \in \Delta A$, *then* $a = 0$;
(f) $\bigcap \operatorname{Max} A = \{0\}$.

30.4.6. Proposition. *Let X be a compact Hausdorff space. Then the Gelfand transform on the Banach algebra $\mathcal{C}(X, \mathbb{C})$ is an isometric isomorphism.*

Hint for proof. Let $A = \mathcal{C}(X, \mathbb{C})$. Since the evaluation map $E \colon X \to \Delta A$ on X is a homeomorphism, it follows that the induced map $\mathcal{C}E \colon \mathcal{C}(\Delta A, \mathbb{C}) \to A$ is an isometric isomorphism. Show that Γ is simply $(\mathcal{C}E)^{-1}$.

30.4.7. Proposition. *If a is an element of a commutative unital Banach algebra, then $\operatorname{ran} \hat{a} = \sigma(a)$ and $\|\hat{a}\|_u = \rho(a)$.*

30.4.8. Example. Let S be a positive measure space. The Gelfand transform on the unital commutative Banach algebra $L_\infty(S, \mathbb{C})$ is an isometric isomorphism.

Hint for proof. Use example 18.4.13 together with propositions 30.4.7 and 18.4.8 to show that $\|\Gamma f\|_u = \|f\|_\infty$. From this and proposition 30.4.2 it is easy to see that the range of Γ is a closed separating unital subalgebra of $\mathcal{C}\big(\Delta\big(L_\infty(S, \mathbb{C})\big)\big)$. To see that this subalgebra is self-adjoint use proposition 30.4.7 and example 18.4.13. Then the conclusion is immediate from the *Stone-Weierstrass theorem* 21.2.14.

Next we examine some of the peculiarities associated with the rather pathological maximal ideal space of the unital commutative Banach algebra $L_\infty\big([0,1], \mathbb{C}\big)$.

30.4.9. Example. Let $A = L_\infty\big([0,1], \mathbb{C}\big)$. The maximal ideal space ΔA is made into a probability space in a standard fashion. Since

$$\Gamma \colon A \to \mathcal{C}(\Delta A) \colon f \mapsto \Gamma f$$

is an isometric isomorphism (see 30.4.8), the map

$$\phi\colon \mathcal{C}(\Delta A) \to \mathbb{C}\colon \Gamma f \mapsto \int_0^1 f\, d\lambda$$

is well-defined and is a bounded linear functional with $\|\phi\| = 1$. By the *Riesz representation theorem* (25.4.7) there exists a unique regular Borel measure μ on ΔA such that $\|\mu\| = 1$ and

$$\int_{\Delta A} (\Gamma f)\, d\mu = \int_0^1 f\, d\lambda$$

for all Γf in $\mathcal{C}(\Delta A)$.

30.4.10. Lemma. *As in the preceding example let $A = \mathrm{L}_\infty\big([0,1],\mathbb{C}\big)$ and regard ΔA as a probability space. Every nonempty open subset of ΔA has strictly positive measure.*

30.4.11. Example. Let $A = \mathrm{L}_\infty\big([0,1],\mathbb{C}\big)$. If f is a complex valued Borel measurable function on the maximal ideal space ΔA, then there exists a continuous function h on ΔA such that $f = h$ μ-a.e. (where μ is the probability measure defined on ΔA in example 30.4.9). Roughly, we are saying that

$$\mathrm{L}_\infty(\Delta A,\mathbb{C}) = \mathcal{C}(\Delta A)\,. \tag{30.1}$$

Of course, this cannot be literally true since the objects on the left side of (30.1) are equivalence classes of functions whereas the objects on the right are functions. Interpret (30.1) as saying that each equivalence class in $\mathrm{L}_\infty(\Delta A,\mathbb{C})$ has a continuous member. Quite a remarkable property!

The next example may make one feel better about the last one.

30.4.12. Example. The maximal ideal space of $\mathrm{L}_\infty\big([0,1],\mathbb{C}\big)$ is extremally disconnected.

30.4.13. Proposition. *If u is a unitary element in a unital C^*-algebra, then $\sigma(u)$ is contained in $\mathbb{T}$, the unit circle in the complex plane.*

30.4.14. Proposition. *If h is a self-adjoint element in a commutative unital C^*-algebra, then $\sigma(h) \subseteq \mathbb{R}$.*

30.4.15. Corollary. *Every character on a unital commutative C^*-algebra is a *-homomorphism.*

30.4.16. Theorem (Gelfand-Naimark Theorem)**.** *Let A be a commutative unital C^*-algebra. Then the Gelfand transform $\Gamma_A\colon a \mapsto \hat{a}$ is an isometric *-isomorphism of A onto $\mathcal{C}(\Delta A,\mathbb{C})$.*

30.5. The Spectral Theorem

30.5.1. Definition. Let S be a subset of a *-algebra A. Then $S^* = \{s^*\colon s \in S\}$. The subset S is SELF-ADJOINT if $S^* = S$. (Note: this does *not* say that each element of S is self-adjoint.) A nonempty self-adjoint subalgebra of A is a *-SUBALGEBRA (or a SUB-*-ALGEBRA). Let A be a C^*-algebra. A nonempty subset of A is a C^*-SUBALGEBRA (or a SUB-C^*-ALGEBRA) of A if it is a C^*-algebra under the algebraic operations and norm it inherits from A.

30.5.2. Proposition. *Let A be a C^*-algebra. The closure of a *-subalgebra of A is a C^*-subalgebra of A.*

30.5.3. Definition. Let S be a nonempty subset of a C^*-algebra A. The intersection of the family of all C^*-subalgebras of A which contain S is the C^*-SUBALGEBRA GENERATED BY S. We denote it by $C^*(S)$. (It is easy to see that the intersection of a family of C^*-subalgebras is a C^*-algebra.) In some cases we shorten the notation slightly: for example, if $a \in A$ we write $C^*(a)$ for $C^*(\{a\})$.

30.5.4. Proposition. *Let S be a nonempty subset of a C^*-algebra A. For each natural number n define the set W_n to be the set of all elements a of A for which there exist x_1, x_2, …, x_n in $S \cup S^*$ such that $a = x_1 x_2 \cdots x_n$. Let $W = \cup_{n=1}^\infty W_n$. Then*

$$C^*(S) = \overline{\operatorname{span}(W)}.$$

30.5.5. Theorem (Abstract Spectral Theorem)**.** *If a is a normal element of a C^*-algebra, then the C^*-algebra $\mathcal{C}(\sigma(a),\mathbb{C})$ is isometrically *-isomorphic to $C^*(\mathbf{1},a)$.*

Hint for proof. Use the Gelfand transform of a to identify the maximal ideal space of $C^*(\mathbf{1},a)$ with the spectrum of a. Apply the functor $\mathcal{C}$. Compose the resulting map with Γ^{-1} where Γ is the Gelfand transform on the C^*-algebra $C^*(\mathbf{1},a)$.

30.5.6. Exercise. Suppose that in the preceding theorem $\psi\colon \mathcal{C}(\sigma(a),\mathbb{C}) \to C^*(\mathbf{1},a)$ implements the isometric *-isomorphism. What is the image under ψ of the constant function $\mathbf{1}$ on the spectrum of a? What is the image under ψ of the identity function $\lambda \mapsto \lambda$ on the spectrum of a?

30.5.7. Exercise. Let T be a normal operator on a Hilbert space H whose spectrum is contained in $[0,\infty)$. Suppose that $\psi\colon \mathcal{C}(\sigma(T),\mathbb{C}) \to C^*(\mathbf{1},T)$ implements the isometric *-isomorphism between these two C^*-algebras. By $\sqrt{T}$ we presumably mean an operator on H whose square is T. How do we know that there is at least one such operator? In case there are several such operators explain how the *abstract spectral theorem* singles out one of these. If f is a continuous function on the spectrum of T explain what we mean by the expression $f(T)$.

30.5.8. Definition. A projection operator on a Hilbert space is an ORTHOGONAL PROJECTION if its range is orthogonal to its kernel.

In 29.5.5 we defined projections on a Banach space B to be the idempotents in $\mathfrak{B}(B)$. We now characterize the orthogonal projections on a Hilbert space H as the self-adjoint idempotents in the C^*-algebra $\mathfrak{B}(H)$.

30.5.9. Proposition. *A projection $P\colon H \to H$ on a Hilbert space is an orthogonal projection if and only if $P = P^*$.*

30.5.10. Proposition. *If N is a normal Hilbert space operator whose spectrum is $\{0,1\}$, then N is an orthogonal projection.*

30.5.11. Proposition. *The Gelfand transform Γ is a natural equivalence between the identity functor and the $\mathcal{C}\Delta$ functor on the category of commutative unital C^*-algebras and *-homomorphisms.*

30.5.12. Theorem (Spectral Mapping Theorem)**.** *If a is a normal element of a C^*-algebra and $f \in \mathcal{C}(\sigma(a),\mathbb{C})$, then*

$$\sigma(f(a)) = f(\sigma(a))\,.$$

Hint for proof. Use the *abstract spectral theorem* 30.5.5. The proof should be *very* short.

Bibliography

1. Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer-Verlag, Berlin, 1999. 101, 119, 129, 133, 138, 157
2. Charalambos D. Aliprantis and Owen Burkinshaw, *Principles of Real Analysis*, Elsevier North Holland, New York, 1981. 138, 149, 157, 164, 175, 190, 193, 197, 198, 204
3. ______, *Positive Operators*, Academic Press, New York, 1985. 101, 119
4. Robert B. Ash, *Real Analysis and Probability*, Academic Press, New York, 1972. 110, 149, 157, 175, 197, 198, 204, 205
5. Deborah J. Bennett, *Randomness*, Harvard University Press, Cambridge, MA, 1998. 182, 184
6. Douglas S. Bridges, *Foundations of Real and Abstract Analysis*, Springer-Verlag, New York, 1998. 155
7. A. L. Brown and A. Page, *Elements of Functional Analysis*, Van Nostrand Reinhold Co., London, 1970. 13
8. Arlen Brown and Carl Pearcy, *Introduction to Operator Theory I: Elements of Functional Analysis*, Springer-Verlag, New York, 1977. 198, 215
9. William C. Brown, *A second Course in Linear Algebra*, John Wiley, New York, 1988. 37
10. N. L. Carothers, *Real Analysis*, Cambridge University Press, Cambridge, 2000. 130, 139, 146, 155, 175, 176
11. Donald L. Cohn, *Measure Theory*, Birkhäuser, Boston, 1980. 131, 146, 190, 193, 198, 204, 205
12. John B. Conway, *A Course in Functional Analysis*, second ed., Springer, New York, 1990. 213, 215
13. J. Dieudonné, *Treatise on Analysis, Volumes I and II*, Academic Press, New York, 1969, 1970. 97
14. J. L. Doob, *Measure Theory*, Springer-Verlag, New York, 1994. 149
15. Ronald G. Douglas, *Banach Algebra Techniques in Operator Theory*, Academic Press, New York, 1972. 215
16. James Dugundji, *Topology*, Allyn and Bacon, Boston, 1970. 129, 131, 133, 200
17. John M. Erdman, *A ProblemText in Advanced Calculus*, `http://web.pdx.edu/~erdman/PTAC/PTAClicensepage.html`. 74, 97, 127, 178, 201
18. Gerald B. Folland, *Real Analysis: Modern Techniques and their Applications*, John Wiley and Sons, New York, 1984. 138, 149, 157, 197, 198, 204, 205
19. D. H. Fremlin, *Measure Theory, Volumes I–IV*, 2003–2006, `http://www.essex.ac.uk/maths/staff/fremlin/mt.htm`. 101, 138, 175, 176, 195
20. Dzung Minh Ha, *Functional Analysis, Volume 1: A Gentle Introduction*, Matrix Editions, Ithaca, New York, 2006. 155, 156
21. Paul Halmos and Steven Givant, *Logic as Algebra*, Mathematical Associaton of America, 1998. 40
22. Paul R. Halmos, *A Hilbert Space Problem Book*, Springer-Verlag, New York, Heidelberg, Berlin, 1982. vii
23. Edwin Hewitt and Karl Stromberg, *Real and Abstract Analysis*, Springer-Verlag, New York, 1965. 13, 37, 79, 129, 130, 131, 139, 140, 175, 176, 193, 197, 198, 204, 205, 213
24. Thomas W. Hungerford, *Algebra*, Springer-Verlag, New York, 1974. 37
25. Keith Hutchinson, *What are Conditional Probabilities Conditional Upon?*, Brit. J. Phil. Sci. **50** (1999), 665–695. 182, 184
26. Irving Kaplansky, *Set Theory and Metric Spaces*, Chelsea Publ. Co., New York, 1977. 18, 152
27. John L. Kelley, *General Topology*, D. Van Nostrand, Princeton, 1955. 13, 71
28. Serge Lang, *Real Analysis*, second ed., Addison-Wesley, Reading, MA, 1983. 129, 138, 140
29. Gottfried Wilhelm Leibniz, *Philosophical Papers and Letters*, University of Chicago Press, Chicago, 1956, (transl. and ed. by Leroy E. Loemker), [Le]. 81
30. Lynn H. Loomis and Shlomo Sternberg, *Advanced calculus*, Jones and Bartlett, Boston, 1990. 97
31. W. A. J. Luxemburg and A. C. Zaanen, *Riesz Spaces, Vol. I*, North-Holland, Amsterdam, 1971. 101, 122, 123
32. John N. McDonald and Neil A. Weiss, *A Course in Real Analysis*, Academic Press, San Diego, 1999. 130, 131, 138, 146, 149, 157, 175, 176, 193, 197, 198, 213
33. Peter Meyer-Nieberg, *Banach Lattices*, Springer-Verlag, Berlin, 1991. 101
34. Gert K. Pedersen, *Analysis Now*, Springer-Verlag, New York, 1989. 133
35. Steven Roman, *Advanced Linear Algebra*, second ed., Springer-Verlag, New York, 2005. 37
36. H. L. Royden, *Real Analysis*, second ed., Macmillan, New York, 1968. 138, 175, 176, 193
37. Walter Rudin, *Real and Complex Analysis*, third ed., McGraw-Hill, New York, 1987. 138, 140, 146, 157, 190, 193, 197, 198, 205, 213, 224

38. ______, *Functional Analysis*, second ed., McGraw Hill, New York, 1991. 129
39. Stanislaw Saks, *Theory of the Integral*, second ed., Phoenix, Dover Publications, New York, 1937. 139
40. Helmut H. Schaefer, *Banach Lattices and Positive Operators*, Springer-Verlag, Berlin, 1974. 101
41. Zbigniew Semadeni, *Banach Spaces of Continuous Functions*, Polish Scientific Publishers, Warzawa, 1971. 59, 102, 167, 215
42. Waclaw Sierpiński, *General Topology*, second ed., Dover Publications, New York, 1956. 139
43. Karl R. Stromberg, *An Introduction to Classical Real Analysis*, Wadsworth, Belmont, 1981. 43, 47, 155
44. Alberto Torchinsky, *Real Variables*, Addison-Wesley, Redwood City, CA, 1988. 138, 190
45. Eric W. Weisstein, *MathWorld*, A Wolfram Web Resource, http://mathworld.wolfram.com. 45
46. Wikipedia, *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org. 22, 45, 182
47. Albert Wilansky, *Topology for Analysis*, Ginn, Waltham, MA, 1970. 71, 129, 131, 132, 200
48. Stephen Willard, *General Topology*, Addison-Wesley, Reading, MA, 1970. 71, 129, 132, 133, 200
49. Adriaan C. Zaanen, *Introduction to Operator Theory in Riesz Spaces*, Springer-Verlag, Berlin, 1997. 101, 119, 122, 123

Index