June 4, 2019

ANALYSIS 2, 2012
LECTURE NOTES

JOHN E. HUTCHINSON

Abstract. The Notes indicate what we do in the lectures.

Additional remarks and proofs, often in response to student questions, are on the Wattle site under "Supplementary Remarks, Comments and Extensions".

Dates in boxes indicate approximately the beginning of the corresponding lecture

These notes are not a replacement for the lectures as they often lack detail, and they do not contain much motivation, overview, or indication of what are the main ideas in proofs and definitions. That is done in the lectures.

The text for the topology is Chapters 11 and 12 plus a little of Chapter 10, from "Real Analysis" by Royden and Fitzpatrick, edition 4. For the measure, integration and Hilbert spaces I mostly follow the first three chapters in the text *Real Analysis* by Stein and Shakarchi. There is also a little additional material on the axiom of choice and equivalent notions.

Assertions marked as *Exercise, why?*, etc., should be proved or explained. They are relatively straightforward and will help you understand the ideas. You will learn a lot from doing them all.

I do not necessarily write out complete definitions or proofs. For that, see the material from texts.

Contents

Part 1. TOPOLOGY

1. Background

Mon 20/2

1.1. A little History. The idea of a metric space was introduced by Fréchet in 1906. He applied this to obtain useful metrics on the set of continuous functions on a closed bounded interval, Euclidean space $\mathbb{R}^n$, holomorphic (i.e. analytic) functions $f(z)$ defined for $|z| < 1$, and the set of continuous curves in $\mathbb{R}^2$.

Hausdorff introduced the idea of a general topological space in 1914 (using "neighbourhoods" rather than "open sets" as the fundamental idea. He developed some consequence but did not realise the potential of the concept.

Von Neumann in 1934 began to realise the utility of these notions and then others up through the 1950's developed the applications in analysis that you will see in this course and in Analysis 3.

1.2. Connections. Topology, also called point set topology or general topology, is a fundamental tool throughout most of contemporary mathematics.

For an overview of mathematics see the mathematical atlas. For example, and for this part of the notes, click on "Topology" and then on "General Topology".

1.3. "Philosophical" Preliminaries.

(1) If (X, d) is a metric space then we have a definition of continuity for functions $f : X \to \mathbb{R}$.
    Now think of X as a surface in $\mathbb{R}^3$ (for example). Just using our informal understanding of "continuity", if we deform X in a continuous manner to $\overline{X}$ then the corresponding new function $\overline{f} : \overline{X} \to \mathbb{R}$ will also be continuous. But the metric will be changed.
    Draw a diagram.
(2) For this and other reasons we would like a notion of continuity which does not directly depend on a metric, but agrees with the usual notion of continuity if there is a metric. In fact you saw in MATH2320 that continuity can be defined in terms of open sets, and that different metrics can give the same class of open sets and hence the same notion of continuity.
(3) Motivated by the above we introduce the idea of a "topology" $\mathcal{T}$ on a set X. $\mathcal{T}$ is a collection of subsets of X, called the collection of "open sets", which satisfies certain conditions. There will be many different topologies on the same set X.
(4) Topological notions occur naturally throughout mathematics (analysis, algebraic topology, differential geometry), and in most fields of science and quantitative economics.

2. **Open and Closed Sets, Bases**

2.1. **Metric Spaces.** Recall the definition of a metric space (X, d)

(1) One example is $\mathbb{R}^n$ with the usual metric.
(2) Another is $\mathbb{C}[a, b]$, or more generally $\mathbb{C}(E)$ where E is a compact metric space, with the sup (here the max) metric.
(3) In many cases we may have the same set X, with *different* metrics, but the *same* open sets and the same notion of continuity.
(4) For this, and other reasons, it is natural and useful to study the idea of open sets, and continuity, without referring to a metric.

2.2. **Topological Space.** We use some of the properties of open sets in the case of metric spaces in order to define what is meant in general by a class of open sets and by a topology.

Definition 2.1. Let X be a nonempty set. A *topology* $\mathcal{T}$ for X is a collection of subsets of X such that $\emptyset, X \in \mathcal{T}$, and $\mathcal{T}$ is closed under arbitrary unions and finite intersections.

We say $(X, \mathcal{T})$ is a *topological space*. Members of $\mathcal{T}$ are called *open* sets.[1]

If $x \in X$ then a *neighbourhood* of x is an open set containing x.

Proposition 2.2. *E is open iff for every $x \in E$ there is a neighbourhood of x which is contained in E.*

Proof. Easy. The idea is motivated by thinking of what happens in $\mathbb{R}^2$. But make sure that you only use the definition of a topological space and not other properties of $\mathbb{R}^2$! □

Remark 2.3 (Examples of topological spaces).

(1) metric topology
(2) discrete topology (this is a metric topology)
(3) trivial topology
(4) see "other material" for all topologies on a three element set.

□

Tues 21/2

Definition 2.4. If $(X, \mathcal{T})$ is a topological space and $E \subset X$ then the *induced* (or *inherited* or *relative*) topology is the collection $\mathcal{S}$ of those subsets of E of the form $E \cap U$ where $U \in \mathcal{T}$.

We say $(E, \mathcal{S})$ is a *(topological) subspace* of $(X, \mathcal{T})$.

In the case of metric spaces the induced topology is consistent with the induced metric. More precisely:

Proposition 2.5. *Suppose X is a metric space, and also a topological space with the topology given by the metric. Suppose $E \subset X$. Then the induced metric on E gives the same topology on E as the topology induced from the topology on X.*

Proof. Exercise □

Example 2.6. Let $X = \mathbb{R}$ with the usual topology, $E = [a, b]$. Then $[a, x)$ is open in the induced topology for every $x \in [a, b]$.

---

[1]Since there may be more than one topology of interest on the set X, you may need to specify which topology is being used if there is likely to be any ambiguity.

2.3. **Base for a Topology.** In today's lecture I made a mistake in the definition of a base.[2] It should be as in the following discussion.

The difference between the following and the material in the text is that here I am not first discussing the notion of *a base at a point.*

Definition 2.7. Suppose $(X, \mathcal{T})$ is a topological space. A collection $\mathcal{B}$ of sets from $\mathcal{T}$ is a *base for* $\mathcal{T}$ if every nonempty set in $\mathcal{T}$ is a union of sets from $\mathcal{B}$.

It follows from the definition that

$$B_1 \in \mathcal{B},\ B_2 \in \mathcal{B},\ x \in B_1 \cap B_2 \implies \exists B \in \mathcal{B}\ (x \in B \subset B_1 \cap B_2)\,. \tag{1}$$

(Exercise)

Example 2.8. $X = \mathbb{R}^n$, $\mathcal{B} = \{B_r(x) : x \in \mathbb{R}^n, r > 0\}$. By $B_r(x) = B(x, r)$ we mean the usual open ball $\{y : |x - y| < r\}$.

Another base is the set of open n-rectangles $(a_1, b_1) \times \cdots \times (a_n, b_n)$, where $a_n < b_n$. *Why?*

Yet another base is the set of open n-cubes $(a_1, a_1 + h) \times \cdots \times (a_n, a_n + h)$, where $h > 0$. *Why?*

Moreover, every collection of sets satisfying (1) is the base for *some* topology. More precisely, we have the following definition and proposition.

Definition 2.9. Suppose X is any non-empty set. A collection $\mathcal{B}$ of subsets of X is a *base* if X is a union of sets from $\mathcal{B}$ and

$$B_1 \in \mathcal{B},\ B_2 \in \mathcal{B},\ x \in B_1 \cap B_2 \implies \exists B \in \mathcal{B}\ (x \in B \subset B_1 \cap B_2)\,. \tag{2}$$

It follows that every base (in the sense of Definition 2.9) is indeed the base for *some* topology (in the sense of Definition 2.7). More precisely:

Proposition 2.10. *If X is a set and $\mathcal{B}$ is a base, then $\mathcal{B}$ is a base for the topology $\mathcal{T}$ which consists of the empty set together with all unions of sets from $\mathcal{B}$.*

Proof. The main point is to check that $\mathcal{T}$ defined in this manner is indeed a topology.

In particular, show that $\mathcal{T}$ is closed under arbitrary unions and finite intersections. For the latter, it is sufficient to show that the intersection of just*two* sets from $\mathcal{T}$ is also a set in $\mathcal{T}$. *(Why?)* This is where (2) is needed. *(Do it)* □

Definition 2.11. Suppose $(X, \mathcal{T})$ is a topological space and $x \in X$.

A *base for the topology at x* is a collection $\mathcal{B}_x$ of neighbourhoods of x such that if U is a neighbourhood of x then $B \subset U$ for some $B \in \mathcal{B}_x$.

~~A *base for the topology $\mathcal{T}$* is a collection $\mathcal{B}$ of open sets that contains a base for the topology at each x.~~

(The second paragraph is "deleted" since we have already defined "base for a topology". As an exercise, show the "deleted" version is equivalent to the previous definition of a base.)

Example 2.12. $X = \mathbb{R}^n$, $\mathcal{B}_x = \{B_r(x) : r > 0\}$, $\mathcal{B} = \{B_r(x) : x \in \mathbb{R}^n, r > 0\}$. By $B_r(x) = B(x, r)$ we mean the usual open ball $\{y : |x - y| < r\}$.

Another base is the set of open n-rectangles $(a_1, b_1) \times \cdots \times (a_n, b_n)$, where $a_n < b_n$. *Why?*

Yet another base is the set of open n-cubes $(a_1, a_1 + h) \times \cdots \times (a_n, a_n + h)$, where $h > 0$. *Why?*

We often use a base to *define* a topology, as in the following definition. Think of the case $X = Y = \mathbb{R}$. *Why does this give the usual topology on $\mathbb{R}^2$?*

Definition 2.13. Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces. Then

$$\mathcal{B} := \{U \times V : U \in \mathcal{T},\ V \in \mathcal{S}\}$$

is a base for a topology. This topology is called the *product topology* on $X \times Y$.

The set $\mathcal{B}$ *is* a base for some topology on $X \times Y$. *Prove this using Definition 2.9.*

$Wed\ 22/2$

[2]Thanks to Don McKinnon for pointing it out.

2.4. **Closed sets.** In the following we assume X is a topological space with a topology $\mathcal{T}$. Unless stated otherwise, $E \subset X$ is an arbitrary set, not necessarily open.

As usual, members of $\mathcal{T}$ are called open sets.

Think of $X = \mathbb{R}^2$ with the standard topology,

$$E = \{x \in \mathbb{R}^2 : |x| < 1\} \cup \{x = (x_1, x_2) \in \mathbb{R}^2 : |x| = 1, x_1 \geq 0\},$$

and draw a picture.

Definition 2.14. Suppose $E \subset X$ where X is a topological space. Then $x \in X$ is a *point of closure* of E if every neighbourhood of x meets E.[3]

The *closure* $\overline{E}$ is the set of points of closure of E.

Clearly $E \subset \overline{E}$. We say E is *closed* if $E = \overline{E}$.

Proposition 2.15. *If $E \subset X$ where X is a topological space, then $\overline{E}$ is closed.*

$\overline{E}$ is the smallest closed set containing E, in the sense that if $E \subset F$ where F is closed, then $E \subset \overline{E} \subset F$.

Proof. The main point is to show that the closure $\overline{\overline{E}}$ of $\overline{E}$ is a subset of $\overline{E}$. The proof requires some care that only the previous definitions are used. I will do it. □

It is not true that an arbitrary set E has a smallest open set containing it. Think of $\{0\} \subset \mathbb{R}^2$.

Proposition 2.16. *If X is a topological space then $E \subset X$ is open iff E^c [4] is closed.*

Proof. Needs a little care that only the definitions are used. It is done in two steps:

1. (1) E open $\implies E^c$ closed;
2. (2) E^c closed $\implies E$ open.

□

It follows that a set F is closed off F^c is open. *Why?* This is often used as the definition of *closed*.

Thurs 23/2

Closed sets have properties analogous to open sets but with "∪" replaced by "∩".

Proposition 2.17. *The collection of closed sets in a topological space X contains $\emptyset$ and X, and it is closed under arbitrary intersections and finite unions.*

Proof. Essentially immediate by de Morgan's laws. □

Definition 2.18. Suppose $E \subset X$ where X is a topological space. The *interior* E° of E is the union of all open subsets of E.

Analogous to Proposition 2.15 is the following.

Proposition 2.19. *If $E \subset X$ where X is a topological space, then E° is open.*

E° is the largest open subset of E, in the sense that if $U \subset E$ where U is open, then $U \subset E^\circ \subset E$.

Proof. Straightforward. *Do it.* □

It is not true that an arbitrary set E has a largest closed set contained in it. Think of the interval $(0, 1) \subset \mathbb{R}$.

The following are not surprising.

---

[3] U *meets* E means $U \cap E \neq \emptyset$.

[4] E^c is the *complement* $X \setminus E$ of E.

Proposition 2.20. *Suppose $E \subset X$ where X is a topological space. Then*

$$(E^\circ)^c = \overline{E^c}, \qquad E^\circ = \left(\overline{E^c}\right)^c, \tag{3}$$
$$\left(\overline{E}\right)^c = (E^c)^\circ, \qquad \overline{E} = \left((E^c)^\circ\right)^c.$$

Proof. The first equality is because E° is the largest open subset of E. Taking complements gives that $(E^\circ)^c$ is the smallest closed set containing E^c and so equals $\overline{E^c}$. *Make this idea into a proof, or give another proof.*

Taking complements of both sides of the first equality gives the second. Replacing E by E^c in the second gives the third. Taking complements of both sides of the third gives the fourth. □

Think of the example E at the beginning of this section in the following definition.

Definition 2.21. Suppose $E \subset X$ where X is a topological space. The *boundary* ∂E of E is the set of points $x \in X$ such that every neighbourhood of x meets both E and E^c. That is,

$$\partial E = \overline{E} \cap \overline{E^c}.$$

Clearly, *∂E is closed*, and $\partial E = \partial E^c$.

Finally we have the following nice decomposition of X given by any set $E \subset X$.

Proposition 2.22. *Suppose $E \subset X$ where X is a topological space. Then the sets E°, ∂E and $(E^c)^\circ$ are disjoint. Moreover,*

$$X = E^\circ \cup \partial E \cup (E^c)^\circ, \quad \overline{E} = E^\circ \cup \partial E, \quad \overline{E^c} = \partial E \cup (E^c)^\circ.$$

Proof. The disjointedness of E° and $(E^c)^\circ$ follows from the fact that E and E^c are disjoint.

The fact E° and ∂E are disjoint follows from assuming otherwise, and then using the definition of ∂E to deduce that E° meets E^c, which is clearly false.

To see the second equality suppose that $x \in \overline{E}$ and $x \notin E^\circ$.

Assume some open set $U \ni x$ and U does not meet E, then $U \subset E^c$ and so $U \subset (E^c)^\circ$ (*why?*). This implies $x \in \left(\overline{E}\right)^c$ by the third equality in (3), which contradicts $x \in \overline{E}$. Hence the *assumption* is false and so every open set U such that $U \ni x$ must meet E.

It follows $x \in \partial E$.

The third equality follows from the second by replacing E by E^c.

The first equality follows from the second by the fourth equality in (3). □

3. Separation Properties

In interesting applications we usually require further properties of the topology $\mathcal{T}$ on a set X, other than those that follow from the axioms . And usually there will be more than one "natural" topology that reflects the various properties associated with X. (For us, X will often be a set of functions.)

Separation properties are important in this respect. Here are the main ones (*Pictures in class*):

Tychonoff or T_1 topology: For every 2 points $x, y \in X$ with $x \neq y$, there is a neighbourhood U of x which avoids y.

Proposition 3.1. *$(X, \mathcal{T})$ is Tychonoff iff every singleton is closed.*

Proof. Exercise. (straightforward) □

Hausdorff or T_2 topology: For every 2 points $x, y \in X$ with $x \neq y$ there exist disjoint open neighbourhoods of x and y respectively.

$\boxed{Mon\ 27/2}$

Regular or T_3 topology: Tychonoff (i.e. singletons are closed) + for every point x and closed set A which are disjoint there exist disjoint open neighbourhoods of x and A respectively.

Normal or T_4 topology: Tychonoff + for every two disjoint closed sets A and B there exist disjoint open neighbourhoods of A an B respectively.

Proposition 3.2.

$$\textit{metric topology} \Longrightarrow \textit{normal} \Longrightarrow \textit{regular} \Longrightarrow \textit{Hausdorff} \Longrightarrow \textit{Tychonoff.}$$

Proof. They are all immediate except the firsst.

To prove this, suppose (X, ρ) is a metric space. It is standard, and easy, to show every singleton is closed.

Next suppose A and B are disjoint closed sets.

Let

$$U = \{x : \rho(x, A) < \rho(x, B)\}, \quad V = \{x : \rho(x, A) > \rho(x, B)\}.$$

Check that

1. Since A is closed, $\rho(x, A) > 0$ if $x \notin A$.
2. $\therefore A \subset U$ and $B \subset V$ and $A \cap B = \emptyset$.
3. using the triangle inequality, U and V are open.

This completes the proof. □

The following gives an important property of normal tooplogies.

Proposition 3.3. *Suppose $(X, \mathcal{T})$ is Tychonoff (i.e. singletons are closed). Then $(X, \mathcal{T})$ is normal iff for every closed F and open U with $F \subset U$, there is an open O such that*

$$F \subset O \subset \overline{O} \subset U.$$

Proof. Fairly straightforward using the fact that the complement of a closed set is open and conversely. Draw a diagram. Proof discussed in class. □

4. **Countability and Separability**

We only briefly treat this section from Royden.

It begins with the natural definition of convergence of a sequence in terms of neighbourhoods. That is: $x_n \to x$ *if, for any neighbourhood* U *of* x*, the sequence eventually belongs to* U.

We do not usually work with sequences, unless the topological space is a metric space. The reason is that sequences do not contain enough information to characterise concepts such as continuity. It is not true in general that f is continuous at x if for every sequence $x_n \to x$, $f(x_n) \to f(x)$.[5]

But there is a more general notion of "nets", which is used instead of sequences, and this does characterise continuity. However, even this is not used that often,

Tues 28/2

Definition 4.1. Suppose X is a topological space. Then $E \subset X$ is *dense* in X if every non empty open set contains a point in E.

Note that E is dense in X iff $\overline{E} = X$. *Why?*

Definition 4.2. A topological space $(X, \mathcal{T})$ is *first countable* if it has a countable base at every point. (This means that for every $x \in X$ there is a countable collection $\mathcal{B}_x$ of open neighbourhoods of x with the property that for every open neighbourhood U of x there exists $V \in \mathcal{B}_x$ such that $x \in V \subset U$.)

$(X, \mathcal{T})$ is *second countable* if it has a countable base.

$(X, \mathcal{T})$ is *separable*[6] if it has a countable dense subset.

In a *topological space*:

$$\text{first countable} \underset{\not\Longrightarrow}{\Longleftarrow} \text{second countable} \underset{\not\Longleftarrow}{\Longrightarrow} \text{separable.}$$

The first implication is immediate.

The second implication is obtained by selecting one point from each set in a countable base. *Show* this gives a countable dense subset.

Problem 21 gives an example of a space that is both first countable and separable, but not second countable.

In a *metric space*:

$$\text{first countable always,} \quad \text{second countable} \Longleftrightarrow \text{separable.}$$

For the first statement suppose $x \in X$ and consider the collection

$$\mathcal{B}_x = \{B_{1/n}(x) : n = 1, 2, 3, \dots\}$$

Show $\mathcal{B}_x$ is a countable base at x.

We previously saw $\Longrightarrow$ for *any* topological space. For the converse in a metric space let E be a countable dense subset and consider the collection

$$\mathcal{B} = \{B_{1/n}(x) : x \in E, n = 1, 2, 3, \dots\}$$

Show $\mathcal{B}$ is a countable base. It is a bit tricky!

An important example of a metric space that is not separable, and hence not second countable, is the space $\mathcal{BC}(\mathbb{R})$ of bounded continuous functions defined on $\mathbb{R}$ with the sup metric. The space $\mathcal{BC}([0, 1])$ with the sup metric is separable. Argument outline in the first assignment.

[5] We give the definition of continuity in a general topological space in the next section.

[6] Separability has nothing to do with the separation properties in the previous section.

5. **Continuous Maps**

$\boxed{Wed\ 29/2}$ *We assume X and Y are topological spaces, unless stated otherwise.*

In a metric space, continuity can be defined either in terms of sequences or in terms of open sets. In a general topological space we have to use the open set definition.

Definition 5.1. A mapping $f : X \to Y$ between topological spaces is continuous at x_0 if for any open neighbourhood O of $f(x_0)$ there is an open neighbourhood U of x_0 such that $f(U) \subset O$. The mapping is continuous if it is continuous at every point in X.

In class I drew a diagram to show why this definition is reasonable — it illustrated what can go wrong if f is *not* continuous at x_0.

Proposition 5.2. *$f : X \to Y$ is continuous iff for any open $O \subset Y$, $f^{-1}(O)$ is open in X.*

Proof. Straightforward. □

Proposition 5.3. *If $f : X \to Y$ is continuous, then $f_A : A \to Y$ is continuous for any $A \subset Y$.*

Proof. Straightforward. Here A is given the induced topology, and f_A is the restriction of f to A. □

Proposition 5.4. *If $f : X \to Y$ and $g : Y \to Z$ are continuous, then so is $g \circ f : X \to Z$.*

Proof. Straightforward using Proposition 5.2. □

Remark 5.5. I will discuss weaker and stronger topologies, from Proposition 11 to Proposition 13 in the text, when and if we need it. □

Definition 5.6. If there is a one to one and onto map f between two topological spaces X and Y which is continuous and whose inverse is continuous, then X and Y are *homeomorphic*.

Remark 5.7. This is a natural definition since it says that that not only is f a one to one correspondence between X and Y, but it also sends open maps to open sets in both directions. As topological spaces, X and Y are indistinguishable.

It is not sufficient to assume that f is a continuous one to one correspondence. We also need that f^{-1} is continuous. In class we see this for the map

$$f : [0, 2\pi) \to S^1, \quad f(t) = (\cos t, \sin t).$$

□

Remark 5.8. Suppose X and Y are topological spaces and $E \subset X$. We say $f : E \to Y$ is continuous if it is continuous in the induced topology on E. This, or something equivalent, is the only reasonable definition. But you may need to realign your intuition.

In the following examples, when considering continuity you may work with either the induced topology or the induced metric. See Proposition 2.5. But you should understand the examples and the continuity or lack of it, *both* from the point of view of the metric *and* the point of view of the topology.

First consider

$$f[0,1] \cup [2,3] \to \mathbb{R}, \quad f(x) = \begin{cases} 0 & \text{if } 0 \le x \le 1, \\ 1 & \text{if } 2 \le x \le 3. \end{cases}$$

Draw the graph. Is f continuous?[7] (Think before you look at the footnote.) *Why?*

Does f have a continuous extension to $[0, 3]$? [8] *Why?*

Next consider

$$f[0,1) \cup (1,2] \to \mathbb{R}, \quad f(x) = \begin{cases} 0 & \text{if } 0 \le x < 1, \\ 1 & \text{if } 1 < x \le 2. \end{cases}$$

[7] Yes

[8] We say $\tilde{f} : [0, 3] \to \mathbb{R}$ is a continuous extension of f if $\tilde{f}$ is continuous and $\tilde{f}$ agrees with f on $[0, 1] \cup [2, 3]$.

Draw the graph. Is f continuous?[9] (Think before you look at the footnote.) *Why?*

Does f have a continuous extension to $[0, 2]$? *Why?* □

Remark 5.9. We discussed:

(1) Why do $[0, 1]$ and $[0, 1]^2$ have the same cardinality? (modulo problem with some points having two decimal expansions the correspondence is

$$.a_1b_1a_2b_2 \cdots \leftrightarrow (a = .a_1a_2 \ldots,\ b = .b_1, b_2, \ldots).$$

(2) The Peano space-filling curve: a continuous map $f : [0, 1] \to [0, 1]^2$ which is onto. (What is a continuous map $f : [0, 1]^2 \to [0, 1]$ which is onto?

(3) *Question for contemplation:* Why is $[0, 1]$ not homeomorphic to$[0, 1]^2$?

□

[9] Yes

6. **Compactness**

Thurs 1/3

When you studied metric spaces you would have seen the notion of compactness.

For subsets of $\mathbb{R}^n$ compactness is equivalent to "closed + bounded".

For subsets of an arbitrary metric space, compactness is defined in any of the following 3 equivalent ways:

(1) "every sequence has a convergent subsequence",
(2) "totally bounded + completeness",
(3) 'every open cover has a finite subcover" (see below)

The notion of total boundedness needs a metric and does not make sense in an arbitrary topological space. Moreover the notion of a sequence is not very useful in an arbitrary topological space, as we have discussed a couple of times. But the third notion does make sense in an arbitrary topological space, and turns out to be very useful.

In the following $(X, \mathcal{T})$ *is always a topological space.*

Definition 6.1. Suppose $E \subset X$. A *cover* of E is a collection $\{E_\lambda : \lambda \in \Lambda\}$ of sets $E_\lambda \subset X$ such that[10]

$$E \subset \{E_\lambda : \lambda \in \Lambda\}, \text{ i.e. } E \subset \bigcup_{\lambda \in \Lambda} E_\lambda.$$

The cover is *open* if all the E_λ are open.

We now use this to define compactness.

Definition 6.2. $(X, \mathcal{T})$ is compact if every open cover[11] of X contains a finite subcover.

A subset E of X is compact if it is compact in the topology induced from X.

The definition of compactness for a *subset* E of X depends on the induced topology on E. This shows that the notion of compactness only depends on the induced topology on E.[12] But a more useful criterion for compactness of a subset of X, which does not use the induced topology, is given by the following.

Proposition 6.3. *A set* $E \subset X$ *is compact if every open cover*[13] *of* E *has a finite subcover.*

Proof. Done in class in one direction. Do the other. They are straightforward, but it is important that you do them. □

An equivalent version of compactness is in the next proposition. But first a definition.

Definition 6.4. A collection of subsets of X satisfies the *finite intersection property* if every finite subcollection has non-empty intersection.

Example 6.5. If $\mathcal{C} = \{(0, 1/n] : n = 1, 2, \dots\}$, or $\mathcal{C} = \{[0, 1/n] : n = 1, 2, \dots\}$, or $\mathcal{C} = \{[n, \infty) : n = 1, 2, \dots\}$ then $\mathcal{C}$, has the finite intersection property.

But note that only the second has non empty intersection.

Proposition 6.6. *A topological space* X *is compact iff every collection of closed subsets of* X *that has the finite intersection property has non-empty intersection.*

Proof. Straightforward using de Morgan's law. (Ask in tutorial!) □

Remark 6.7. Why does the previous example show that $(0, 1]$ and $\mathbb{R}$ are not compact? □

---

[10] Λ is a a set of indices, which may be finite or infinite. For example $\Lambda = \mathbb{R}$.

[11] Since the sets covering X are open, it is implicit that they are subsets of X. Hence X *equals* the union of the sets in the cover.

[12] Note the difference with "closed". A set $E \subset X$ may or may not be closed in X, but it is always closed in the topology induced on E from X. *Why?* We say the notion of being compact is an *absolute* notion but the notion of being closed is a *relative* notion.

[13] Here it is implicit that "open" means open in X.

Mon 5/3

The next two propositions together give a useful way of checking when a subset of a compact hausdoff space is compact. (Recall that in analysis, all spaces are usually Hausdorff).

Proposition 6.8. *Suppose X is compact. If $K \subset X$ is closed it is compact.*

Proof Sketch. (Done in class) The idea is to take an open cover $\mathcal{C}$ of K, add K^c to get an open cover of X, and use the compactness of X to get the required subcover of K. □

Proposition 6.9. *Suppose X is Hausdorff. If $K \subset X$ is compact it is closed.*

Proof Sketch. (Done in class) The idea is to show that K^c is open by showing each point $x \in K^c$ has an open neighbourhood $U \subset K^c$.

This is done by using the fact X is Hausdorff to get disjoint open sets $V_y \ni y$ and $U_y \ni x$ for each $y \in k$. Using compactness a finite number of the V_y cover K. The intersection of the corresponding U_y gives the reqired neighbourhood U. □

We omit the material on sequential compactness and Proposition 17 in the text.

Proposition 6.10. *A compact Hausdorff space X is normal. That is, any two disjoint closed sets have disjoint open neighbourhoods.*

Proof Sketch. (Done in class) Suppose the 2 disjoint closed sets are E and F.

From Proposition 6.8 both E and F are compact. From the proof of Proposition 6.9, for each $x \in F$ there are disjoint open neighbourhoods V_x of x and U_x of E.

Now use compactness of F to get a subcover of F by a *finite* number of the V_x. Let V be the union of these V_x and let U be the intersection of the corresponding finite number of U_x. Then U and V are the required open neighbourhoods of E and F respectively. □

Remark 6.11. It is not true that the continuous image of a closed set is necessarily closed, nor that the continuous image of an open set is necessarily open. Give counterexamples for continuous functions $f : \mathbb{R} \to \mathbb{R}$. (Try $f(x) = \arctan(x)$ and $f(x) = x^2$.)

But we do have the following result. □

Proposition 6.12. *The continuous image of a compact set is compact.*

Proof. Suppose $f : X \to Y$, $K \subset X$ and K is compact. Let $\{U_\lambda : \lambda \in \Lambda\}$ be an open cover of $f(K)$.

Because f is continuous $\{f^{-1}(U_\lambda) : \lambda \in \Lambda\}$ is an *open* cover of K. By compactness, there is a finite subcover $\{f^{-1}(U_{\lambda_1}), \dots, f^{-1}(U_{\lambda_N})\}$, say, of $f^{-1}(K)$. But

$$f^{-1}(K) \subset f^{-1}(U_{\lambda_1}) \cup \dots \cup f^{-1}(U_{\lambda_N}) \Longrightarrow K \subset U_{\lambda_1} \cup \dots \cup U_{\lambda_N}.$$

Why? Hence $f(K)$ is compact. □

Compare the following with Remark 5.7.

Proposition 6.13. *Suppose $f : X \to Y$ where X is compact, Y is Hausdorff, f is continuous, one to one and onto. Then f^{-1} is continuous and so f is a homeomorphism.*

Proof. Note that the inverse f^{-1} of f certainly exists.

We need to show that E open implies $\left(f^{-1}\right)^{-1}(E)$, the pullback of E via f^{-1}, is open. Equivalently, we need to show that $f(E)$ is open.

Equivalently (*why?*), by taking complements, we need to show that A closed implies $f(A)$ is closed.

But A is compact by Proposition 6.8 and so $f(A)$ is compact by Proposition 6.12. Hence $f(A)$ is closed by Proposition 6.9.

This completes the proof. □

Proposition 6.14. *Suppose $(X, \mathcal{T})$ is a topological space, K is a compact subset of X and $f : K \to \mathbb{R}$ is continuous. Then f takes a maximum and a minimum value.*

Proof. By proposition 6.12, $f(K) \subset R$ is compact and hence closed and bounded. But a bounded subset of $\mathbb{R}$ has a supremum, and being closed it must contain its supremum. This proves the result for the maximum, *why?*

Similarly for the minimum, or apply the maximum result to $-f$. $\square$

7. Connectedness

Tues 6/3

We assume $(X,\mathcal{T})$ *is a topological space.*

Definition 7.1. Two open sets *separate*[14] X if they are non empty, disjoint and their union is X.

The space X is *connected* if there is no pair of open sets which separate X.

A set $E \subset X$ is *connected* if it is connected in the induced topology.

The previous definition gives a definition for E to be connected in terms of the induced topology. Just as happens in the case of compactness, there is a criterion for connectedness which does not use the induced topology.

Proposition 7.2. *A set* $E \subset X$ *is connected if there do not exist open sets* U *and* V *in* X *such that*

$$U \cap E \neq \emptyset,\ V \cap E \neq \emptyset,\ U \cap V \cap E = \emptyset,\ E \subset U \cup V. \tag{4}$$

Proof. Suppose there *do* exist open sets U and V in X such that (4) is true. Then $E \cap U$ and $E \cap V$ are open sets in the topology of E which separate E.

Suppose O_1 and O_2 are open sets in the topology of E which separate E. Then there exist open sets U and V in the sense of X, such that $O_1 \cap E = U$ and $O_2 \cap E = V$. It follows that (4) is true. *Why?*

The theorem follows. □

Remark 7.3. If U and V are disjoint open sets separating X, then it follows that U and V are also *closed.*

It also follows that X can be separated iff there exists a non empty proper subset[15] which is both open and closed. *Why?* □

Proposition 7.4. *A continuous image of a connected space is connected.*

Proof. Easy. Done in class. □

Definition 7.5. A non empty set $I \subset \mathbb{R}$ is an *interval* if whenever $a, b \in I$ and $a < t < b$ then $t \in I$.

That is, $I \subset \mathbb{R}$ is an interval iff I is convex.

Proposition 7.6. *A non empty set* $A \subset \mathbb{R}$ *is connected iff it is an interval.*

Proof. ($\Longrightarrow$): Suppose $\emptyset \neq A \subset \mathbb{R}$.

Suppose A is *not* an interval. Then there exist $a, b \in I$ with $a < t < b$ and $t \notin I$.

Let $U = (-\infty, t)$ and $V = (t, \infty)$. Then U and V are nonempty open sets in $\mathbb{R}$ which separate A, and so A is not connected.

($\Longleftarrow$): Suppose $\emptyset \neq A \subset \mathbb{R}$. Suppose A *is* an interval.

Assume A is not connected (in order to obtain a contradiction).

Then there exist (disjoint) open sets U and V separating A and w.l.o.g.[16] we may assume there exist $a \in U$ and $b \in V$ with $a < b$. Moreover, $[a, b] \subset A$ since A is an interval.

Let $E = [a, b] \cap U$ and let $c = \sup E$. Either $c \in V$ or $c \in U$.

Suppose $c \in V$. (Note $c \neq a$ but perhaps $c = b$.) Since V is open, $(c - \epsilon, c] \subset V$ for some $\epsilon > 0$. Hence

$$t \in (c - \epsilon, c] \Longrightarrow t \notin U \Longrightarrow t \notin U \cap [a, b].$$

This contradicts $c = \sup(U \cap [a, b])$. *Why?*

Suppose $c \in U$. Then $c \neq b$ but perhaps $c = a$. Since U is open, $[c, c+\epsilon) \subset U$ for some $\epsilon > 0$. By decreasing ϵ if necessary we also have $[c, c+\epsilon) \subset U \cap [a, b]$. This contradicts $c = \sup(U \cap [a, b])$.

Since we have a contradiction either way, it follows that A *is* connected. □

---

[14] "Separate" here has nothing to do with "separation" properties from Section 3, and neither have anything to do with "separable" from Section 4.

[15] $E \subset X$ is a *proper* subset if $E \neq X$.

[16] "without loss of generality"

Definition 7.7. X is *arcwise* or *pathwise* connected if for all $u, v \in X$ there is a continuous function $f : [0,1] \to \mathbb{R}$ such that $f(0) = u$ and $f(1) = v$.

Proposition 7.8. *Pathwise connected implies connected, but connected does not imply pathwise connected.*

Proof. Suppose X pathwise connected.

Assume X is not connected (in order to obtain a contradiction).

Then there exist open sets U and V separating X. Since U and V are non empty, there exist points $u \in U$ and $v \in V$.

Since X is arcwise connected, there exists a continuous $f : [0,1] \to \mathbb{R}$ such that $f(0) = u$ and $f(1) = v$. Since $[0,1]$ is connected by Proposition 7.6, it follows that the image set $f[0,1]$ is connected by Proposition 7.4.

However, U and V separate $f[0,1]$ (*why?*), and so $f[0,1]$ is not connected.

This contradiction implies the *assumption* is false and so X is connected.

For a counterexample in the other direction let

$$X = \left\{(x,y) : x = 0,\ -1 \le y \le 1\right\} \cup \left\{(x,y) : x > 0,\ y = \sin\frac{1}{x}\right\}.$$

This is connected but not pathwise connected. Discussed briefly in class. See Q 49, p 238. ☐

Proposition 7.9. *If X is connected then the image of any continuous $f : X \to \mathbb{R}$ is an interval.*

If X is not connected there is a continuous function $f : X \to \mathbb{R}$ whose image is $\{0,1\}$.

Proof. If X is connected then $f(X)$ is connected by Proposition 7.4 and hence an interval by Proposition (7.6).

If X is not connected let U and V be a separating open pair. Define $f(x) = 0$ if $x \in U$ and $f(x) = 1$ if $x \in V$. ☐

8. Urysohn's Lemma and Tietze Extension Theorem

Thurs 8/3

We discussed:

(1) The definition of connectedness in terms of closed (instead of open) sets; and the definition in terms of clopen sets.
(2) Every open $U \subset \mathbb{R}$ is a countable union of disjoint open intervals I_n, i.e. $U = \bigcup_{n\geq 1} I_n$.
(3) There exists an open set $U \subset \mathbb{R}$ as above such that $\mathbb{Q} \subset U$ and $\sum_{n\geq 1} |I_n| < 10^{-28}$.
(4) The Cantor set C, it is a uncountable, every $x \in C$ has a unique "address" consisting of an infinite sequence of 0's and 1' s, for every pair $x \neq y$ of points from C there exist separating open sets for$C$ containing x and y respectively — we say that C is *totally disconnected.*

To understand the following ideas it would be helpful to look at your copies of the diagrams I drew in the lectures.

Remark 8.1.

(1) We will be working in normal spaces. Important examples are metric spaces, and compact Hausdorff spaces.
(2) In the case of a metric space, the function f given in Urysohn's Lemma can be taken to be *(check this!)*

$$f(x) = \frac{d(x, A)}{d(x, A) + d(x, B)}.$$

(3) The interval $[0, 1]$ can be replaced by any interval $[a, b]$. Just compose f with a linear map from $[0, 1]$ to $[a, b]$.

□

Theorem 8.2 (Urysohn's Lemma)**.** *Suppose A and B are disjoint closed subsets of a normal space X. Then there is a a real-valued continuous function $f : X \to [0, 1]$ such that $f = 0$ on A and $f = 1$ on B.*

Proof. **Step A:** By the definition of a normal space, there exists an open set $O_{1/2}$ such that

$$A \subset O_{1/2} \subset \overline{O_{1/2}} \subset B^c.$$

Similarly, there exist open sets $O_{1/4}$ and $O_{3/4}$ such that

$$A \subset O_{1/4} \subset \overline{O_{1/4}} \subset O_{1/2} \subset \overline{O_{1/2}} \subset O_{3/4} \subset \overline{O_{3/4}} \subset B^c.$$

Iterating this, let $\Lambda = \{m/2^n : n = 1, 2, \dots\ ;\ m = 1, 2, \dots, 2^n - 1\}$. Then there exist open sets O_r for each $r \in \Lambda$ such that if $r < s$ then

$$A \subset O_r \subset \overline{O_r} \subset O_s \subset \overline{O_s} \subset B^c. \tag{5}$$

Step B: Define

$$f(x) = \begin{cases} \inf\{r \in \Lambda : x \in O_r\} & \text{if } x \in \bigcup_r O_r, \\ 1 & \text{if } x \in X \setminus \bigcup_r O_r. \end{cases} \tag{6}$$

Clearly $f : X \to [0, 1]$, $f = 0$ on A and $f = 1$ on B.

Step C: We claim that f is continuous.

To do this it is sufficient to show that if $0 < a, b < 1$ then

(1) $f^{-1}[0, b)$ is open,
(2) $f^{-1}(a, 1]$ is open.

Why does it follow from this that f is continuous?

But

$$f(x) < b \iff x \in \bigcup_{r<b} O_r,$$
$$f(x) > a \iff x \in \bigcup_{s>a} \overline{O_s}^{\,c}.$$

This follows from (6) and (5). *Exercise.* □

Tues 13/3

Remark 8.3.

(1) In the following theorem, taking $F = (-1,0) \cup (0,1)$ with $f = 0$ on $(-1,0)$ and $f = 1$ on $(0,1)$ shows the need for F to be closed. Or just take $F = (0,1)$ with $f(x) = x^{-1}$ for $x \in F$.
(2) On the other hand, $F = [-1,0] \cup [\delta,1]$ with small $\delta > 0$ is OK.
(3) The closed set F may be quite complicated. For example, it may be a Cantor style closed set. Or F could be a closed subset of $\mathbb{R}$ containing no rational numbers, where F^c is a countable union of disjoint intervals such that the sum of their lengths is $< 10^{-28}$.
(4) We will use two important results from your previous courses, whose proof you should look up, and which we recall in two footnotes.[17]

□

Theorem 8.4 (Tietze Extension Theorem)**.** *Suppose $F \subset X$ where X is normal and F is closed. Then any continuous $f : F \to [a,b]$ can be extended to a continuous function $g : X \to [a,b]$.*

The result is also true with $[a,b]$ replaced by $\mathbb{R}$.

Proof. W.l.o.g. take $[a,b] = [-1,1]$, so

$$\forall x \in F \ |f(x)| \le 1.$$

Step 1: Let

$$A = \Big\{x \in F : f(x) \le -\frac{1}{3}\Big\}, \quad B = \Big\{x \in F : f(x) \ge \frac{1}{3}\Big\},$$

Then A and B are disjoint closed subsets of X. By Urysohn's Lemma there exists a continuous function g_1 defined on X, such that $g_1 = -1/3$ on A, $g_1 = 1/3$ on B, and

$$\forall x \in X \ |g_1(x)| \le \frac{1}{3}.$$

It follows (diagram in class)

$$\forall x \in F \ |f(x) - g_1(x)| \le \frac{2}{3}.$$

By the same argument as before, now applied to $f - g_1$, there exists a continuous function g_2 defined on X such that

$$\forall x \in X \ |g_2(x)| \le \frac{1}{3}\cdot\frac{2}{3}, \quad \forall x \in F \ \big|f(x) - g_1(x) - g_2(x)\big| \le \left(\frac{2}{3}\right)^2.$$

---

[17] "A uniform limit of continuous functions is continuous." More precisely:

Theorem: ("A uniform limit of continuous functions is continuous.")

Suppose X is a topological space and (Y,d) is a metric space. Suppose (f_n) is a sequence of continuous functions $f_n : X \to Y$ such that $f_n \to f$ uniformly for some function $f : X \to Y$.

(That is, for every $\epsilon > 0$ there exists an integer N such that $\forall n \ge N \ \forall x \in X \ d\big(f_n(x).f(x)\big) < \epsilon$.)

Then f is continuous.

Note that $f_n \to f$ pointwise $\not\Longrightarrow$ $f_n \to f$ uniformly. For example, let

$$f_n(x) = \begin{cases} 0 & x \le 0, \\ nx & 0 \le x \le 1/n \\ 1 & x \ge 1/n. \end{cases}; \quad f(x) = \begin{cases} 0 & x \le 0 \\ 1 & x > 0. \end{cases}$$

The other result is the "Wierstrass M-test" for uniform convergence of a series of functions.

Theorem: Suppose X is an arbitrary set and $g_n : X \to \mathbb{R}$ for $n = 1, 2, \ldots$. Suppose for each n, $|g_n(x)| \le M_n$ for all $x \in X$, and suppose $\sum_{n\ge1} M_n < \infty$. Then $\sum g_n$ converges uniformly to some $g : X \to \mathbb{R}$.

Repeating the argument, for each $n \geq 1$ there exists a continuous function g_n defined on X such that

$$\forall x \in X \; |g_n(x)| \leq \frac{1}{3} \cdot \left(\frac{2}{3}\right)^{n-1}, \quad \forall x \in F \; \big|f(x) - g_1(x) - \ldots - g_n(x)\big| \leq \left(\frac{2}{3}\right)^{n}. \tag{7}$$

Step 2: From (7) and the Wierstrass M-test (or just directly by checking from (7) that the sequence of partial sums from $\sum_{n\geq 1} g_n(x)$ is uniformly Cauchy) it follows that $\sum_n g_n(x)$ converges uniformly on X to a function $g : X \to \mathbb{R}$. Since the g_n are continuous it follows from Footnote 17 that g is continuous. It also follows from (7) that $f = g$ on F.

This completes the proof for the case $f : F \to [a, b]$.

Step 3: Now suppose that $f : F \to \mathbb{R}$.

Let $\psi : \mathbb{R} \to (-1, 1)$ be the homeomorphism with inverse (*check!*) $\psi^{-1} : (-1, 1) \to \mathbb{R}$, given by

$$\psi(x) = \frac{x}{1 + |x|}, \quad \psi^{-1}(y) = \frac{y}{1 - |y|}.$$

Consider the (continuous) composition $\widehat{f}$ of f with $\psi : \mathbb{R} \to (-1, 1)$, that is

$$\widehat{f} := \psi \circ f = \frac{f}{1 + |f|} : F \to (-1, 1).$$

By the Tietze extension theorem for closed bounded intervals, there is a continuous function $\widehat{g} : X \to [-1, 1]$ (not to $(-1, 1)$ unfortunately) such that $\widehat{g}(x) = \widehat{f}(x)$ for $x \in F$.

By Urysohn's lemma there is a continuous function $\phi : X \to$ [-1,1] such that $\phi(x) = 1$ if $x \in F$, and $\phi(x) = 0$ if $\widehat{g}(x) = \pm 1$. (Note that F and $\{x : \widehat{g}(x) = \pm 1\}$ are disjoint and closed, *why?*). The function $\phi\widehat{g}$ agrees with $\widehat{g}$ on F, and hence agrees with $\widehat{f}$ on F. Moreover, it never takes the values ± 1, *why?*

It follows that the function

$$\psi^{-1} \circ (\phi\widehat{g}) = \frac{\phi\widehat{g}}{1 - |\phi\widehat{g}|} : X \to \mathbb{R}$$

is always finite and hence continuous, and it agrees with $\psi^{-1} \circ \widehat{f}$, i.e. with f, on F. □

9. **Baire Category Theorem**

Thurs 15/3

In this section (X, d) *is a* complete *metric space.*

The *very rough* idea is to first think of dense open sets as being "large", in the sense of being "disperse, i.e. spread out" (dense) and "locally fat" (open). Examples are any subset of $\mathbb{R}^2$ obtained by removing a finite number of "nice" (e.g. C^1) curves, not the Peano space filing curve.

The Baire Category Theorem says that a countable intersection of dense open subsets of a complete metric space X is not empty, and in fact is dense.

By passing to complements it follows that a complete metric space X is not a countable union of "small" sets, where "small" means *nowhere dense (in* X*)*.[18]

Example 9.1.

(1) $\mathbb{Q} \subset \mathbb{R}$ is a countable union of nowhere dense sets.
(2) $\mathbb{I} \subset \mathbb{R}$ is *not* a countable union of nowhere dense sets. (Because if it were, then so would $\mathbb{R}$, which is not the case by the Baire Category Theorem.)
(3) The Cantor set $\mathbb{C}$ is a countable union of nowhere dense sets, it fact it is itself nowhere dense. *Why?*

There is quite a bit of terminology, which I find easy to forget and so I will avoid.[19]

The Baire Category Theorem is used to prove the Open Mapping Theorem and the Closed Graph Theorem in Functional Analysis, see Analysis III, MATH3325.

The Baire Category Theorem is often used to show the existence in a complete metric space of objects with a property P. For example, the existence of a nowhere differentiable function — see the "Problems and Solution". See also examples below. The idea is to describe the set S of objects with the property P as the countable intersection of a collection of dense open sets. Then by the Baire Category Theorem $S \neq \emptyset$ and so there must be some object satisfying P, in fact a dense subset of such objects.

Theorem 9.2. *Let* (X, d) *be a complete metric space.*

(1) *If* $\{U_n\}_{n\geq 1}$ *is a sequence of dense open sets, then* $\bigcap_n U_n$ *is dense in* X*, and in particular is non-empty.*
(2) X *is not a countable union of nowhere dense sets.*

Proof. I explain the idea.[20] *The diagram drawn in class should clarify matters.*

Suppose U is an arbitrary open subset of X and $B_{r_0}(x_0) \subset U$.

Then U_1 meets $B_{r_0}(x_0)$ and so there exists $B_{r_1}(x_1) \subset U_1 \cap B_{r_0}(x_0)$. By decreasing the radius r_1, in fact there exists some $\overline{B_{r_1}}(x_1) \subset U_1 \cap B_{r_0}(x_0)$.[21]

Similarly U_2 meets $B_{r_1}(x_1)$ and so there exists $\overline{B_{r_2}}(x_2) \subset U_2 \cap B_{r_1}(x_1)$.

Similarly U_3 meets $B_{r_2}(x_2)$ and so there exists $\overline{B_{r_3}}(x_3) \subset U_3 \cap B_{r_2}(x_2)$.

Etc.

By decreasing the r_n if necessary (e.g. take $r_n < 1/n$) we can assume $r_n \to 0$.

If $m \geq n$ then $x_m \in B_{r_n}(x_n)$. It follows the sequence x_n is Cauchy and so $x_n \to x$ for some $x \in X$ (in fact if $m > n$ then $d(x_n, x_m) < 2r_n$, *why?*).

[18] *Definition*: If X is a topological space the set $E \subset X$ is *nowhere dense in* X if $\overline{E}$ has empty interior. Note that

$$E \text{ nowhere dense} \implies \overline{E} \text{ nowhere dense} \implies \overline{E}^c \text{ open and dense.}$$

[19] Countable union of nowhere dense sets = first category = meagre, not a countable union of nowhere dense sets = second category = non meagre, complement of a first category set = residual = non meagre, and even empty interior = hollow.

[20] See Royden p211 for details.

[21] *Prove* that if $r < s$ then $\overline{B_r}(x) \subset B_s(x)$.

So far we have not used the fact that we took *closed* balls at each point. We now use it to show that $x \in \bigcap_n U_n$, that is $x \in U_n$ for each n.[22]

The point is that if $m \geq n$ then by construction $x_m \in \overline{B_{r_n}}(x_n) \subset U_n$. Hence $x \in \overline{B_{r_n}}(x_n) \subset U_n$ for each n, and so $x \in \bigcap_n U_n$. □

[22]Give an example of a ("nested") sequence of open non empty intervals $I_n = (a_n, b_n) \subset X = [0,1]$, $I_1 \supset I_2 \supset I_3 \supset \cdots$, $b_n - a_n \to 0$, $x_n \in I_n$, $x_n \to x \notin I_1$.

Part 2. AXIOM OF CHOICE AND EQUIVALENTS

Mon 19/3

We only discussed the main ideas.

(1) *Sections 1 and 2.*
(2) *Read and think about Section 3, up to the end of Proposition 3.2.*
(3) *Read and think about Section 4, don't worry about the proofs — though they are not that difficult)*
(4) *Section 5 is optimal, but note for future reference.*

AC may seem obvious, but it implies WO which is quite counter intuitive, and ZL which is often used and not so easy to understand. In fact all three are equivalent.

The Axiom of Choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma? Jerry Bona

The Axiom of Choice is necessary to select a set from an infinite number of socks, but not an infinite number of shoes. Bertrand Russell

We follow the treatment in "Real analysis and probability" by R. M. Dudley, 2002, pp 12–15, 19–21.

1. Interesting Consequences

We will see that the axiom of choice, the well ordering principle and Zorn's lemma each imply one another. In particular, the following surprising results are true.

- There is a well ordering of the real numbers and of $\mathbb{R}^n$ for any n.
- Every vector space has a basis in the linear algebra sense. For example,
  (1) There is a a set of continuous functions
  $$S \subset \mathcal{C}(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\},$$
  such that every $f \in \mathcal{C}(\mathbb{R})$ can be written uniquely as a finite sum of the form
  $$f = \sum_{i=1}^{n} a_i \phi_i$$
  for some $n \geq 1$, $a_i \in \mathbb{R}$ and $\phi_i \in S$.
  (2) For any Hilbert space $\mathcal{H}$ there is a set $S \subset \mathcal{H}$ such that every $f \in \mathcal{H}$ can be written uniquely as a finite sum of the form
  $$f = \sum_{i=1}^{n} a_i \phi_i$$
  for some $n \geq 1$, $a_i \in \mathbb{R}$ (or $\mathbb{C}$) and $\phi_i \in S$.
  (3) There is a set $S \subset \mathbb{R}$ such that every $x \in \mathbb{R}$ is a unique finite linear sum
  $$x = \sum_{i=1}^{n} r_i y_i$$
  for some $n \geq 1$, $r_i \in \mathbb{Q}$ and $y_i \in S$.

We won't have time for Propositions 3.2 and 3.3, but they are included for completeness and later possible use. Also there will not be time to prove all the equivalences in Theorem 5.1.

2. Different Types of Orderings

Definition 2.1. (See Figure 1.) $(X, \leqslant)$ is a *partial ordering* if $\leqslant$ is a binary relation on X such that

(1) $a \leqslant b$ and $b \leqslant c$ implies $a \leqslant c$ (transitive),
(2) $a \leqslant a$ (reflexive),
(3) $a \leqslant b$ and $b \leqslant a$ implies $a = b$ (antisymmetric).

We write $a < b$ if $a \leqslant b$ and $a \neq b$.

Example 2.2. The following are partial orderings.

(1) For any set A, $(\mathcal{P}(A), \subset)$ (where "$\subset$" allows equality of sets).
   More generally, $(\mathcal{A}, \subset)$ for any subcollection $\mathcal{A} \subset \mathcal{P}(A)$, is a partial ordering.
(2) An example we will use later is, with V a vector space, the partial ordering $(X, \subset)$ where
$$X = \{S \subset V \mid \text{every finite subset of } S \text{ is linearly independent}\}.$$
(3) The usual $\leqslant$ ordering on $\mathbb{Z}$, $\mathbb{N}$ or $\mathbb{R}$.
(4) $X = \{a, b, c, d, e, f, g\}$ with the partial ordering described in Figure 1.

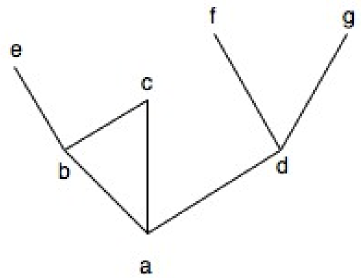

Figure 1. $x \leqslant y$ if $x = y$ or there is a "rising" path from x to y. For example, $a \leqslant b$, $a \leqslant c$ and $a \leqslant e$.

Definition 2.3. $(X, \leqslant)$ is a *linear ordering* if it is a partial ordering such that $a, b \in X$ implies $a \leqslant b$ or $b \leqslant a$ (note both are true precisely when $a = b$).

$L \subset X$ is an *initial segment* of X if $(y \in L \;\&\; x < y) \Longrightarrow x \in L$.

Example 2.4. Only (3) in Example 2.2 is a linear ordering.

Definition 2.5. If $(X, \leqslant)$ is a partial ordering then $a \in X$ is a *maximal* element if there is no b such that $a < b$. We say a is a *greatest* element if $b \leqslant a$ for every $b \in X$.

Remark 2.6.

(1) There is at most one greatest element in a partial ordering.
(2) Every greatest element is a maximal element.
(3) In Figure 1 the maximal elements are e, c, f, g. There is no greatest element.
(4) An important example is $(X, \subset)$ from (2) in Example 2.2. We say that $S \subset V$ is a *basis* for V if $S \in X$ (i.e. if every finite subset of S is linearly independent), and if every $v \in V$ is a *finite* linear combination of elements from S.[23]

Definition 2.7. $(X, \leqslant)$ is a *well ordering* if it is a linear ordering such that every non-empty subset of X contains a least element.[24]

[23] An orthonormal basis in the Hilbert space sense is *not* a basis in this sense. Here we are defining a basis in the usual algebraic sense, sometimes called a Hamel basis — there is no notion of a limit.

It follows $S \in X$ is maximal iff S is a basis for V *Why?*.

An orthonormal base in a separable Hilbert space is countable. A basis is always uncountable. *Exercise.*

[24] A *least* element of $A \subset X$ is of course an element $a \in A$ such that $a \leqslant b$ for all $b \in A$.

Example 2.8.

(1) The only example of a well ordering in Example 2.2 is $\mathbb{N}$ in (3).
(2) $X = \mathbb{N} \cup \{a\}$ is a well ordering if we use the standard ordering on $\mathbb{N}$ and set $n < a$ for all $n \in \mathbb{N}$.
(3) $\mathbb{N}$ followed by another copy of $\mathbb{N}$ is a well ordering.
(4) Many more examples can be constructed this way.

3. Properties of Well Orderings

Definition 3.1. Suppose $(X, \leqslant)$ is a well ordering.

The *first* or *initial* element is the least element of X.

If $x \in X$ is any element then its *successor* is the least element y such that $x < y$, and is denoted by $s(x)$. Thus $s(x)$ exists unless x is the greatest element of $(X, \leqslant)$.

A *limit* element $x \in X$ is an element which is not the first element and not a successor element. Equivalently, x is not the first element and for every $y < x$ there is an element z such that $y < z < x$.

For a well ordering, it follows from the definitions that every element is either the first element, a successor element, or a limit element.

In Example 2.8(3) the "first" 1 is the initial element. The "second" 1 is a limit element. All other elements are successor elements.

Proposition 3.2. *If $(A, \leqslant)$ and $(B, \preccurlyeq)$ are well orderings then one is isomorphic to an initial segment (perhaps all) of the other.*

Proof. We don't need the result, and the proof is like that for the following Proposition. So I leave it as an exercise. □

We often define a function by induction over the natural numbers by first defining $f(1)$ and then defining $f(n)$ in terms of $f(1), \dots, f(n-1)$. For example, the Fibonacci sequence is defined by

$$f(1) = 1,\ f(2) = 1,\ f(n) = f(n-2) + f(n-1) \text{ for } n \geq 3.$$

In general, we can define

$$f(n) = g\big(f\lfloor I(n)\big),$$

where g is a given function, $I(n)$ is the initial segment $\{k : k < n\}$, and $f\lfloor I(n)$ is the restriction of f to $I(n)$.

It is useful here to think of a function as a set of ordered pairs, so $f\lfloor I(n) = \{(k, f(k)) : k < n\}$.

We now generalise inductive definitions to well ordered sets.

Proposition 3.3 (Definition by Recursion)**.** *Suppose $(X, \leqslant)$ is a well ordered set. Let $I(x) = \{y : y < x\}$ for each $x \in X$. Then given g with range T and appropriate domain,*[25] *there is a unique function $f : X \to T$ such that*

$$f(x) = g\big(f\lfloor I(x)\big). \tag{8}$$

Proof. The idea is straightforward.

Let $J(x) = \{y : y \leqslant x\}$.[26]

Let G be the set of those $z \in X$ such that there is a function of the form $f : J(z) \to T$ which satisfies (8) for all $x \in J(z)$.

For each $z \in G$ there can be at most one such function. If not, consider the first $x \leqslant z$ where two such functions differ and obtain an immediate contradiction, since by (8) they must agree at x.

If $y_1 < y_2$ and there exist corresponding such functions f_1 and f_2, then $f_2\lfloor I(y_1)$ is of the required form and so equals f_1.

It follows that G is an initial segment of X. Moreover, f is uniquely defined and satisfies (8) for all $x \in G$.

If $G \neq X$ let y be the least element in $X \setminus G$. Then we can use (8) to extend f from G to $G \cup \{y\}$. This contradicts the definition of G.

Hence $G = X$ and we are done. □

---

[25]The domain of G is the set of functions whose domain is an initial segment of X and whose range is T.

[26]It is very convenient to work with initial segments of the form $J(x)$ as well as with $I(x)$.

4. Five equivalent versions of AC

See Figure 2 for the following.

Assertion 4.1 (AC: The Axiom of Choice)**.** Suppose $\{S_x : x \in I\}$ is a family of non-empty sets. Then there exists a function $f : I \to \bigcup_{x\in I} S_x$ such that $f(x) \in S_x$ for all $x \in I$.

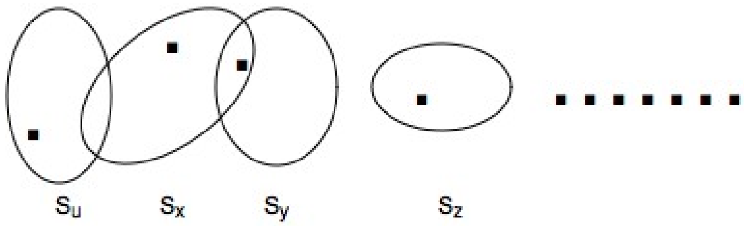

Figure 2. Axiom of Choice with choice function indicated by black dots.

Remark 4.2.

(1) A function is a set of ordered pairs, and so AC asserts the existence of a *set* with certain properties.

The axiom of choice is different from most other axioms of set theory which assert the existence of a set, in that other axioms give a criterion for membership of that set. For example, the axiom of set theory asserting the existence of the union of a family of sets, or the axiom asserting the existence of the power set of a set, gives a criterion for membership of the union or power set respectively.

(2) The axiom of choice as formulated above says that if S_x is non empty for every $x \in I$ then the cartesian product $\prod_{x\in I} S_x$ is also non empty. So it certainly seems a very harmless axiom! But it has some non obvious and even counter intuitive consequences, as we will see.

(3) If the index set I is finite, then the axiom of choice follows from the other axioms of set theory and the rules of logic (i.e. first order predicate calculus).

(4) If there is a "rule" for selecting an object from each S_x then one does not need the axiom of choice. For example, if $S_x = \{0, 1\}$ for each x then one can take $f(x) = 0$ for all x, or $f(x) = 1$ for all x.

The following is a useful alternative to AC.

Assertion 4.3 (AC*)**.** For any set X let $I = \{A \subset X : A \neq \emptyset\}$. Then there exists $f : I \to X$ such that $f(A) \in A$ for all $A \in I$.

Proposition 4.4. *AC is equivalent to AC*.*

Proof. (AC $\Longrightarrow$ AC*): AC* is the particular case of AC obtained by taking $S_A = A$.

(AC* $\Longrightarrow$ AC): Assume AC*. Suppose $\{S_x : x \in I\}$ is a family of non-empty sets. Let $X = \bigcup_{i\in I} S_x$. See Figure 3

By AC*

$$\exists g : \mathcal{P}(X) \setminus \{\emptyset\} \to X \text{ such that } g(A) \in A \quad \forall A \subset X, A \neq \emptyset.$$

Define

$$f(x) = g(S_x) \quad \forall x \in I.$$

□

Assertion 4.5 (WO: The Well-Ordering Principle)**.** Every set can be well-ordered.

Assertion 4.6 (HMP: Hausdorff's Maximal Principle)**.** Suppose $(X, \leqslant)$ is a partially ordered set. Then X contains a maximal linearly ordered subset.[27]

[27] L is a *maximal* linearly ordered subset if it is linearly ordered by $\leqslant$ and if there is no larger linearly ordered subset containing L.

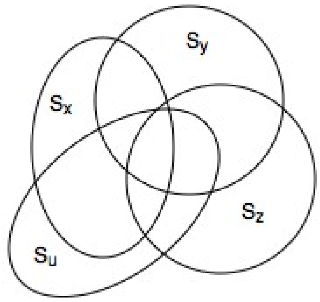

FIGURE 3. $X = \bigcup_{x\in I} S_x$. See Proposition 4.4.

Assertion 4.7 (ZL: Zorn's Lemma)**.** Suppose $(X, \leqslant)$ is a partially ordered set such that every linearly ordered subset L has an upper bound in X. (That is, $\exists y \in X$ such that $x \leqslant y$ for all $x \in L$.)

Then $(X, \leqslant)$ has a maximal element.

Here are two surprising propositions which follow from ZL and WO respectively.

First recall that a subset S of a vector space V is *linearly independent* if every finite subset of S is linearly independent in the usual sense. A subset S is a *basis* for V if it is linearly independent and if every $v \in V$ can be written as a finite linear combination of vectors in S.

It follows that if S is a basis then each $v \in V$ can be written in a *unique* way as a finite linear combination of vectors in S, *why?* Moreover, a linearly independent set S is a basis iff it cannot be extended to a larger linearly independent set, *why?* Another way of expressing this is that a linearly independent set $S \subset V$ is a basis for V iff it is maximal in the collection of linearly independent subsets of V with the partial ordering given by set inclusion.

Proposition 4.8. *Every vector space V contains a basis S^*. More generally, every linearly independent set $S_0 \subset V$ can be extended to a basis.*

Proof. (The point here is that this includes the case where V is not finite dimensional).

We apply ZL. (One could alternatively use HMP.)

Let

$$\mathcal{X} = \{S \subset V \mid S_0 \subset S,\ S \text{ is linearly independent}\}.$$

Note that $(\mathcal{X}, \subset)$ is a partial ordering.

If $\mathcal{L}$ is a linearly ordered subset of $\mathcal{X}$ then $\bigcup\{S : S \in \mathcal{L}\}$ is a subset of V with the property that every finite subset is linearly independent, *why?* Hence $\bigcup\{S : S \in \mathcal{L}\} \in \mathcal{X}$ and is an upper bound for $\mathcal{L}$, *why?*

By Zorn's lemma, $\mathcal{X}$ has a maximal element S^* (in fact many). Every element in V is a finite linear combination of elements of S^*, as otherwise we could enlarge S^* and contradict its maximality. □

Proposition 4.9. $\mathbb{R}^n$ *can be well ordered.*

Proof. By the well ordering principle. □

5. Proof of the various equivalences

Theorem 5.1. *AC $\iff$ AC* $\iff$ HMP $\iff$ ZL $\iff$ WO.*

Proof. We have already seen AC $\iff$ AC*.

We will show AC* $\implies$ WO $\implies$ HMP $\implies$ ZL $\implies$ AC*.

1: (AC* $\implies$ WO): Let X be a non empty set. Suppose f is a "choice function" satisfying $f(A) \in A$ for all $\emptyset \neq A \subset X$. The informal idea is to define the well ordering $x_1, x_2, \ldots, x_n, \ldots, x_\omega, \ldots$ by

$$x_1 = f(X),\ x_2 = f(X \setminus \{x_1\}),\ x_3 = f\big(X \setminus \{x_1, x_2\}\big),\ \ldots\ ,\ x_\omega = f\big(X \setminus \{x_n : n \in \mathbb{N}\}\big),\ \ldots\ .$$

To make this precise consider all well orderings $(E, \leqslant)$ such that

(1) $E \subset X$,
(2) $x \in E \implies x = f\big(X \setminus \{y \in E : y < x\}\big)$.

We have just seen there are indeed some such well orderings.

We *claim* if $(A, \leqslant)$ and $(B, \preccurlyeq)$ are two such well orderings then one is an initial segment of the other (with the same ordering).

Suppose neither $(A, \leqslant)$ nor $(B, \preccurlyeq)$ is an initial segment of the other.

Then it follows that for some $x \in A$, $\{y \in A : y \leqslant x\}$[28] is not an initial segment of $(B, \preccurlyeq)$.[29]

Take the first $x \in A$ such that $\{y \in A : y \leqslant x\}$ is *not* an initial segment of $(B, \preccurlyeq)$. It follows that $D := \{y \in A : y < x\}$ is an initial segment of both $(A, \leqslant)$ and $(B, \preccurlyeq)$.[30]

We will show $x \in B$ and that by "adding" x to the end of D we get $\{y \in A : y \leqslant x\}$ is an initial segment of both A and B, contradicting the definition of x.

By property (2) for the well ordering $(A, \leqslant)$, we have $x = f(X \setminus D)$.

On the other hand, since $B \setminus D$ is non empty (as otherwise B is an initial segment of A) there is a least element x^* in $B \setminus D$. Then by property (2) for the well ordering $(B, \preccurlyeq)$, we have $x^* = f(X \setminus D)$.

Hence $x^* = x$ and so by "adding" $x = x^*$ to the end of D we see $\{y \in A : y \leqslant x\}$ is an initial segment of $(B, \preccurlyeq)$, contradicting the definition of x.

Thus for any two well orderings satisfying (1) and (2), one is an initial segment of the other. For this reason we can take the union of *all* such well orderings to get a well ordering of some $Y \subset X$. If $Y \neq X$ then we can enlarge the well ordering by adding $f(X \setminus Y)$ at the end, thus contradicting the fact we took the union of all such well orderings.

2: (WO $\implies$ HMP): Suppose $(X, \leqslant)$ is a partial ordering and assume there is a well ordering $(X, \preccurlyeq)$.

The idea is to use the well ordering $\preccurlyeq$ to build up the maximal linearly ordered subset L for $\leqslant$. Let the well ordering be $x_1, x_2, \ldots, x_n, \ldots, x_\omega, \ldots$. Set $x_1 \in L$. If $\leqslant$ gives a linear ordering on $\{x_1, x_2\}$ then include $x_2 \in L$, otherwise exclude it. If $\leqslant$ gives a linear ordering on $\{x_3\} \cup \{\text{elements already in } L\}$ then include $x_3 \in L$, otherwise exclude it. ... If $\leqslant$ gives a linear ordering on $\{x_\omega\} \cup \{\text{elements already in } L\}$ then include $x_\omega \in L$, otherwise exclude it. Etc.

To make this precise we define L, or more precisely the characteristic function $\mathcal{X}_L$, by recursion as in Proposition 3.3. That is

$$\mathcal{X}_L(x) = \begin{cases} 1 & \text{if } \leqslant \text{ is a linear ordering on } \{x\} \cup \{y : y \prec x \ \&\ \mathcal{X}_L(y) = 1\} \\ 0 & \text{otherwise} \end{cases}.$$

Then $\prec$ is a linear order on L and L is also maximal in this respect.[31]

3: (HMP $\implies$ ZL): Assume HMP.

---

[28] It is important for the proof to take $\{y \in A : y \leqslant x\}$ and not $\{y \in A : y < x\}$. Note that doing this gives us more information, since not every initial segment is of the form $\{y \in A : y \leqslant x\}$.

[29] Since otherwise, if $\{y \in A : y \leqslant x\}$ is an initial segment of $(B, \preccurlyeq)$ for every $x \in A$, then taking the union one can check that this implies A is an initial segment of $(B, \preccurlyeq)$.

[30] Since $\{y \in A : y < x\} = \bigcup_{\{y \in A : y < x\}} \{z \in A : z \leqslant y\}$, and a union of initial segments is an initial segment. Similarly for B.

[31] Since if $x \in X$ and $\prec$ is a linear order on $L \cup \{x\}$ then the definition of $\mathcal{X}_L$ implies $\mathcal{X}_L(x) = 1$, and so $x \in L$.

Suppose $(X, \leqslant)$ is a partially ordered set such that every linearly ordered subset has an upper bound in X. By HMP there is a *maximal* linearly ordered subset L. Let y be an upper bound for L.

Then y is clearly a maximal element for $(X, \leqslant)$.[32]

4: (ZL $\Longrightarrow$ AC*): Assume ZL.

Given a set X, let F be the set of all choice functions f whose domain is a *subfamily* of the family of all non empty subsets of X. By "choice" function it is meant as usual that $f(A) \in A$ for all A in the domain of f.

Define the partial order $\leqslant$ on F by $f \leqslant g$ if g is an extension of f. Then it is straightforward to check that $(F, \leqslant)$ is a partial ordering. Moreover, any linearly ordered subset has a greatest element, obtained by taking the union (regarding a function as a set of ordered pairs) of all functions in the linearly ordered subset.

By ZL there is a maximal element $f \in (F, \leqslant)$. The domain of f must be $\mathcal{P}(X) \setminus \{\emptyset\}$ as otherwise we could extend f by adding an ordered pair (x, A) for any non empty set $A \subset X$ and $x \in A$.[33] $\square$

---

[32] Since if $y < w$ then $L \cup \{w\}$ would be linearly ordered, contradicting the maximality of L.

[33] There is *not* a hidden application here of AC because of our "choosing" A and then choosing $x \in A$. *Why?*

Part 3. MEASURE THEORY

| *Tues* 20/3 − − − *Thurs* 29/3 | *We covered Sections 3.0 – 3.3, up to the end of the section on measurable sets, on page* *39**.*

0. Introduction

The text begins with 5 examples where we need to consider a more general notion of integration than Riemann integration.

(1) *Fourier Series*:
   (a) Work on $[-\pi, \pi]$ for simplicity of notation.
   (b) For "nice" (Riemann integrable) $f : [-\pi, \pi] \to \mathbb{R}$ we define the Fourier coefficients

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\cos nx\,dx,\;\; n \geq 0,$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\sin nx\,dx,\;\; n \geq 1.$$

   Or (better) used complex valued functions as in text.
   (c) Then

$$a_0^2 + \sum_{n\geq 1}(a_n^2 + b_n^2) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx.$$

   (d) Natural question: Given numbers a_n, b_n such that left sum is finite, is there a corresponding function f so equality holds? What form of integration is needed?
   The answer requires Lebesgue integration and Hilbert space theory.

(2) *Limits and Integrals:* If a sequence of continuous functions $f_n \to f$ (not necessarily uniformly) does $\int f_n \to \int f$? The answer requires Lebesgue integration.
   Related problems arise naturally when one looks for functions minimising various "energies".

(3) *Length of Curves:* When is the length of a curve in $\mathbb{R}^2$, given by $(x(t), y(t))$ for $a \leq t \leq b$, equal

$$\int_a^b \left((x'(t))^2 + (y'(t))^2\right)^{1/2} dt\ ?$$

   Answer requires notions of Lebesgue integration and of functions of bounded variation.

(4) *Integration and Differentiation as reverse notions:* When are integration and differentiation reverse operations? When are the two notions defined? How are they defined?
   Answer requires notions of Lebesgue integration and absolutely continuous functions.

(5) *Length, area, volume:* Can one assign a natural notion of length, area, volume, ... to arbitrary subsets of $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, dots?
   Answer requires notions of Lebesgue measure.

Summary: It is not possible to use or study Fourier series, partial differential equations, or essentially anything involving integration and limits of functions without the theory of Lebesgue measure and integration.

It is not possible to study probability theory without generalising these ideas to other measures.

Essentially any area of mathematics and its applications which involves ideas of limits (i.e. other than algebra) uses Lebesgue measure and integration, and its generalisations, as a basic tool

(1) First goal is to give a useful notion of size to *any* $E \subset \mathbb{R}^d$. This is called the *exterior measure* $m_*(E)$ of E (or *outer measure* or even just *measure*, sometimes with the qualifier *Lebesgue*).

(2) To obtain additivity

$$m_*(E_1 \cup \cdots \cup E_n) = m_*(E_1) + \cdots + m_*(E_n)$$

   on finite *disjoint* unions one needs to restrict to the "measurable" sets. (One usually uses the word *measure* when restricting outer measure to the measurable sets
   One also then obtains countable additivity on countable (infinite) disjoint unions.

Cannot expect additivity for uncountable disjoint unions. Consider

$$\mathbb{R}^d = \bigcup \{\{x\} : x \in \mathbb{R}^d\}.$$

Measure of left side is $+\infty$, but of any singleton is 0.

(3) Non measurable sets are highly pathological and essentially never arise in practice or in theory.

(4) There are many different approaches, but all lead to the same notion of outer measure.

Once one has the basic results and properties, one usually forgets the particular approach taken.

I follow the approach in the text, which is about as short as it gets.

1. Preliminaries

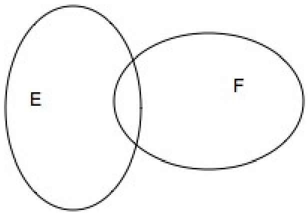

Figure 4. The distance between E and F is zero.

The *distance* between two sets, defined by

$$d(E,F) = \inf\{|x-y| : x \in E,\, y \in F\},$$

is quite different from the Hausdorff metric $d_{\mathcal{H}}$ or "dist", see MATH2320, [SS, p 345] and Figure 4. It is not even a metric.

A *closed rectangle*, usually just called a *rectangle*, in $\mathbb{R}^d$ is a set of form

$$R = [a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d \tag{9}$$

where $a_i \le b_i$ for each i.

The *volume* is the product (of real numbers)

$$|R| = (b_1 - a_1) \times \cdots \times (b_d - a_d).$$

It will equal the more generally defined notion of exterior measure when the latter is restricted to closed rectangles.

A *closed cube* or *cube* is a rectangle where all sides have equal length.

An *open rectangle* or *open cube* is defined similarly, except the intervals in (9) are open.

A union of rectangles is *almost disjoint* if the interiors are disjoint.

Lemma 1.1. *If $R = \bigcup_{k=1}^N R_k$ is an almost disjoint union of rectangles then $|R| = \sum_{k=1}^N |R_k|$.*

Proof. See text page 5, Fig 2.

(1) The result is direct if the union is obtained by partitioning each side $[a_i, b_i]$.
(2) By extending all edges of all R_k to the edges of R one obtains a partition $R = \bigcup_j \widehat{R}_j$ for which (1) is true.
(3) Moreover, each R_k is also of the form in (1) for a subset of the $\widehat{R}_j$ indexed by $j \in J_k$, say.
(4) This gives

$$\begin{aligned} |R| &= \sum_j |\widehat{R}_j| \text{ by (2)} \\ &= \sum_k \sum_{j \in J_k} |\widehat{R}_j| \text{ by rearranging} \\ &= \sum_{k=1}^N |R_k| \text{ by (3)} \end{aligned} \tag{10}$$

□

Lemma 1.2. *If $R \subset \bigcup_{k=1}^N R_k$ is a union of rectangles then $|R| \le \sum_{k=1}^N |R_k|$. (The union need not be almost disjoint.)*

Proof. Extend given rectangles as before, obtaining new rectangles $\widehat{R_j}$.

The new "maximal" rectangle contains R, and each $\widehat{R_j}$ may be a subset of more than one R_k. Thus "=" should be replaced by "≤" in the first and third lines in (10). □

Theorem 1.3. *Every open $U \subset \mathbb{R}^d$ is a countable union of almost disjoint (closed) cubes.*

Proof. Consider grids of side 2^{-n} where $n = 1, 2, 3, \ldots$. See Fig 3 p7. Consider the corresponding cubes of side 2^{-n}.

At stage 1 each cube of side 1 is accepted if it is a subset of U, rejected if a subset of U^c, and held in reserve if it meets both U and U^c.

At stage 2 all reserve cubes from stage 1 are further subdivided and then accepted, rejected or held in reserve according to the above criteria.

At stage 3 all reserve cubes from stage 2 are further subdivided and then accepted, rejected or held in reserve according to the above criteria. Etc.

Let $\mathcal{Q}_n$ be the collection of cubes accepted at stage n. Each such cube has side 2^{-n}.

Claim: $U = \bigcup_{n\geq 1} \mathcal{Q}_n$ is the required countable almost disjoint union.

First note it follows from the construction that the union is countable, almost disjoint, and $\bigcup_{n\geq 1} \mathcal{Q}_n \subset U$.

Next note that if $x \in U$ then since U is open, for some n there is one or more cubes of size 2^{-n} containing x which is a subset of U, *why?* Let n be the smallest such n for which this is true. The corresponding cube will be accepted at this stage and so $x \in \bigcup_{n\geq 1} \mathcal{Q}_n$. Hence $U \subset \bigcup_{n\geq 1} \mathcal{Q}_n$.

This proves the claim. □

2. Exterior measure

Definition 2.1. For $E \subset \mathbb{R}^d$ the *exterior measure* of E is

$$m_*(E) = \inf\left\{ \sum_{j\ge1} |Q_j| : E \subset \bigcup_{j\ge1} Q_j,\ Q_j \text{ are closed cubes} \right\}.$$

2.1. Remarks.

(1) $m_*(E) \in [0,\infty]$.
(2) $m_*(\emptyset) = 0$
(3) The cover can be finite or countably infinite, although the text has uses only countably infinite unions.
Alternatively, it is convenient to allow $Q_j = \emptyset$ and define $|\emptyset| = 0$.
(4) One can use closed rectangles instead of cubes, but the result is the same.

Proof Sketch. A rectangle can be covered by cubes such that the sum of their volumes is within ϵ of the rectangle's volume. □

(5) One cannot just use finite sums.

Proof sketch. Consider $E = \mathbb{Q} \cap [0,1]$. With finite covers one obtains the value 1, but with countable covers one obtains 0. □

(6) $m_*(\{x\}) = 0$. Proof in class.
(7) *If Q is a closed cube then $m_*(Q) = |Q|$.*

Proof Sketch. "$\le$" is immediate.

For "$\ge$" it is sufficient to show $|Q| \le \sum_j |Q_j|$ for any cover $(Q_j)_{j\ge1}$ of Q.
Enlarge the Q_j a little to open cubes S_j with $|S_j| \le (1+\epsilon)|Q_j|$.
Use compactness to get a finite subcover $S_1, \dots, S_N$ and then apply Lemma 1.2 to the closures to get

$$|Q| \le \sum_{j=i}^{N} |S_j| \le (1+\epsilon)\sum_{j=i}^{N} |Q_j| \le (1+\epsilon)\sum_{j\ge i} |Q_j|.$$

Since $\epsilon > 0$ is arbitrary, $|Q| \le \sum_j |Q_j|$. Result follows. □

(8) *If Q is an open cube then $m_*(Q) = |Q|$.*

Proof Sketch. Squeeze Q between a closed sub cube $\widehat{Q}$ with $|\widehat{Q}|$ near $|Q|$, and $\overline{Q}$. □

(9) *If R is a closed rectangle then $m_*(R) = |R|$.*

Proof Sketch. [34] "$\ge$" is similar to (7).

For "$\le$" use a grid of size ϵ to get almost disjoint cubes $\{Q_i : i = 1, \dots, N\}$ with

$$\bigcup_{i=1}^{M} Q_i \subset R \subset \bigcup_{i=1}^{N} Q_i,$$

where the left union is a rectangle, and $\sum_{i=M+1}^{N} |Q_i| \le c\epsilon$ where c depends on R but not on ϵ. Then

$$m_*(R) \le \sum_{i=1}^{N} |Q_i| \le \sum_{i=1}^{M} |Q_i| + c\epsilon \le |R| + c\epsilon.$$

Since $\epsilon > 0$ is arbitrary we are done. □

(10) $m_*(\mathbb{R}^d) = \infty$

Proof Sketch. $\mathbb{R}^d \supset C$ for arbitrarily large cubes C. □

[34] On p12 line 6 of text it should read: "... that intersect R and the complement of R ..."

(11) $m_*(C) = 0$ if C is the Cantor set.

Proof Sketch. C is covered by 2^n closed intervals of length 3^{-n}. □

2.2. **Properties of Exterior Measure.**

(1) Basically everything you might expect, *except* finite and countable additivity for disjoint unions, is true for outer measure.
We get finite and countable additivity for the *measurable* sets.

(2) *Monotonicity*: $E_1 \subset E_2 \Longrightarrow m_*(E_1) \leq m_*(E_2)$. Proof easy.

(3) *Countable Subadditivity*: $E \subset \bigcup_{i\geq 1} E_i \Longrightarrow m_*(E) \leq \sum_{i\geq 1} m_*(E_i)$.

Proof Sketch. Take an "$\epsilon/2^i$-efficient" cover of E_i by cubes. This gives an "ϵ-efficient" cover of E by cubes. □

(4) *Approximating by Open Sets from the Outside*:
$$m_*(E) = \inf\{m_*(U) : E \subset U,\ U \text{ open}\}.$$

Proof. $\leq$ is true by monotonicity.

Next take an $\epsilon/2$-efficient cover of E by closed cubes Q_j in Defn 2.1. Take an $\epsilon/2^{j+1}$ efficient cover of Q_j by an open cube Q_j^o.
Then $\mathcal{O} = \bigcup_{j\geq 1} Q_j^o$ is an open set containing E such that $m_*(\mathcal{O}) \leq m_*(E) + \epsilon$.
Result follows. □

(5) *Additivity for positively separated sets*:
$$d(E_1, E_2) > 0 \Longrightarrow m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2).$$

Proof Sketch. $\leq$ is true by subadditivity.

Suppose $d(E_1, E_2) = \delta > 0$. Take an ϵ-efficient cover of E by closed cubes Q_j in Defn 2.1. By subdividing cubes we can assume all have diameter $\leq \delta/3$, none meets both E_1 and E_2, and so some of the Q_j cover E_1 while the remainder cover E_2.
This gives $m_*(E_1) + m_*(E_2) \leq m_*(E) + \epsilon$ and we are done. □

(6) *Measure adds for countable almost disjoint unions of cubes*: $E = \bigcup_{i\geq 1} Q_i$ is a countable union of almost disjoint cubes implies $m_*(E) = \sum_{i\geq 1} |Q_i| = \sum_{i\geq 1} m_*(Q_i)$.
In particular, by Theorem 1.3 this applies to open sets E.

Proof Sketch. "$\leq$" is by definition of m_*.

Next fix N and $\epsilon > 0$.
Take cubes $\tilde{Q}_j$ strictly inside Q_j for $j = 1, \ldots, N$ such that
$$\begin{aligned}\sum_{j=1}^{N} |Q_j| &\leq \left(\sum_{j=1}^{N} |\tilde{Q}_j|\right) + \epsilon \\ &= m_*\left(\bigcup_{j=1}^{N} \tilde{Q}_j\right) + \epsilon, \text{ by } 5. \\ &\leq m_*(E) + \epsilon \text{ by monotonicity}\end{aligned}$$
This is true for all N and so $\sum_{i\geq 1} |Q_i| \leq m_*(E) + \epsilon$. But ϵ is arbitrary, so $\sum_{i\geq 1} |Q_i| \leq m_*(E)$.
Using 7. on page 33 we are done. □

3. Measurable Sets

(1) Suppose $E \subset \mathbb{R}^d$. By 4. on page 34, for each $\epsilon > 0$ there is an open set $U \supset E$ such that $m_*(U) \le m_*(E) + \epsilon$.

If we knew m_* was additive on disjoint sets we could deduce $m_*(U \setminus E) = m_*(U) - m_*(E) \le \epsilon$ (in case $m_*(E) < \infty$).

(2) *Definition*: $E \subset \mathbb{R}^d$ is *(Lebesgue) measurable* if for any $\epsilon > 0$ there is an open $U \supset E$ such that $m_*(U \setminus E) \le \epsilon$.

If E is measurable we define its *(Lebesgue) measure* $m(E) = m_*(E)$.

3.1. Properties.

(1) *Open sets are measurable.*

Follows from the definition.

(2) *Sets of measure 0, and hence subsets of sets of measure 0, are measurable.*

Proof Sketch: By 4. on page 34 and then monotonicity on page 34.

(3) *Countable unions of measurable sets are measurable.*

Proof Sketch: If $E = \bigcup_{j\ge1} E_j$ where each E_j is measurable, take an open $U_j \supset E_j$ where $m_*(U_j \setminus E_j) < \epsilon/2^j$.

Then $U = \bigcup_{j\ge1} U_j$ is the required open set.

(4) *Closed sets are measurable.*

Proof Sketch: By the previous result it is sufficient to show closed bounded sets F are measurable.

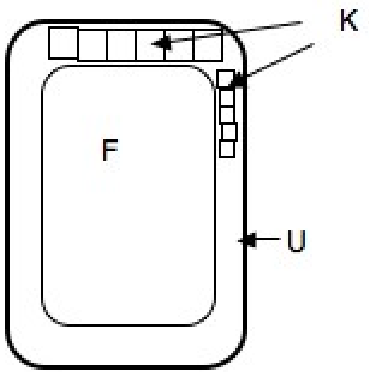

Figure 5. Diagram for proof that closed sets are measurable.

For such an F choose open $U \supset F$ such that $m_*(U) \le m_*(F) + \epsilon$. We W.T.S. $m_*(U \setminus F) \le \epsilon$.

Let (the open) $U \setminus F = \bigcup_j Q_j$ be a countable union of almost disjoint closed cubes. Then $m_*(U \setminus F) = \sum_{j\ge1} m_*(Q_j)$ by 6. on page 34.

For any N, $K := \bigcup_{j=1}^N Q_j$ is compact, is disjoint from the closed set F, and so $d(K, F) > 0$.[35]

Hence

$$m_*(U) \ge m_*(F \cup K) = m_*(F) + m_*(K) = m_*(F) + \sum_{j=1}^N m_*(Q_j),$$

Letting $N \to \infty$, and using the previous,

$$m_*(U \setminus F) = \sum_{j\ge1} m_*(Q_j) \le m_*(U) - m_*(F) \le \epsilon,$$

as required.

(5) *Complements of measurable sets are measurable.*[36]

Proof Sketch: Suppose E is measurable and choose open sets $U_n \supset E$ such that $m_*(U_n \setminus E) \le 1/n$.

---

[35] By arguments involving compact sets from the topology course. See text, Lemma 3.1, p18.

[36] Line 6-, p18 of text should be "so that ...", not "such that ...".

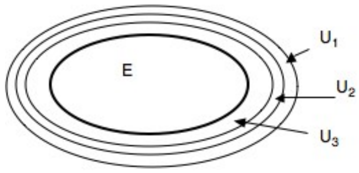

FIGURE 6. Diagram for proof that complements of measurable sets are measurable.

Then $E^c \supset S := \bigcup_n U_n^c$, which is measurable by 3. and 4.

Moreover, $E^c \setminus S \subset U_n \setminus E$ and so $m_*(E^c \setminus S) \leq 1/n \to 0$. Hence $m_*(E^c \setminus S) = 0$ and so $E^c \setminus S$ is measurable by 2.

Since $E^c = S \cup (E^c \setminus S)$ it is measurable by 3.

(6) *Countable intersections of measurable sets are measurable.*

Follows from previous results for complements and unions.

(7) *Countable additivity*: If $(E_j)_{j\geq 1}$ are disjoint measurable sets then $m(E) = \sum_{j\geq 1} m(E_j)$.

Proof. First assume all E_j are bounded.

Fix $\epsilon > 0$.

Choose closed $F_j \subset E_j$ s.t. $m(E_j \setminus F_j) < \epsilon/2^j$ (use measurability of E_j^c).[37]

For each N,

$$m(E) \geq m(\bigcup_{j=1}^{N} F_j) = \sum_{j=1}^{N} m(F_j) \geq \sum_{j=1}^{N} \left(m(E_j) - \frac{\epsilon}{2^j} \right),$$

using monotonicity for the first "$\geq$"; that the F_j are compact and disjoint and hence positively separated (see Footnote 35) and hence additivity holds, for the "$=$"; and subadditivity for the second "$\geq$".

Letting $N \to \infty$ and using the arbitrariness of ϵ, $m(E) \geq \sum_{j\geq 1} m(E_j)$. Using subadditivity completes the proof for the case the E_j are bounded.

Now for possibly unbounded E_j let $(S_k)_{k\geq 1}$ be a partition of $\mathbb{R}^d$ into bounded disjoint measurable sets. Then $E = \bigcup_{j,k} E_j \cap S_k$ and so

$$m(E) = \sum_{j,k} m(E_j \cap S_k) = \sum_j \sum_k m(E_j \cap S_k) = \sum_j m(E_j),$$

where the result for the bounded case is used for the first and third "$=$", and the fact the infinite series consists of non negative numbers is used for the second "$=$". □

3.2. Summary.

(1) The measurable sets include the open and closed sets.
(2) Finite and countably infinite set theoretic operations do not take one outside the class of measurable sets.
(3) Monotonicity, and countable additivity for disjoint unions, hold for the measure of measurable sets.
(4) The measure of a rectangle is the same as the standard volume.

3.3. Increasing Unions and Decreasing Intersections.

(1) If E_n are measurable and $E_n \nearrow E$ [38] then $m(E_n) \to m(E)$.

Proof Outline: Write $E = E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus E_2) \cup \ldots$ and use the fact this is a disjoint union.

(2) If $m(E_1) < \infty$ and $E_n \searrow E$ [39] then $m(E_n) \to m(E)$.

Proof Outline: $E_1 \setminus E_n \nearrow E_1 \setminus E$. Now use the previous result.

---

[37] This approximation by closed sets from the inside is important, and will be discussed later.

[38] $E_n \nearrow E$ means $E_1 \subset E_2 \subset E_3 \ldots$ and $E = \bigcup_n E_n$.

[39] $E_n \searrow E$ means $E_1 \supset E_2 \supset E_3 \ldots$ and $E = \bigcap_n E_n$.

3.4. **Approximating Measurable Sets.** Suppose E is measurable. Then for every $\epsilon > 0$:

(1) $\exists$ open $U \supset E$ such that $m(U \setminus E) < \epsilon$;
(2) $\exists$ closed $F \subset E$ such that $m(E \setminus F) < \epsilon$;
(3) $m(E) < \infty \Longrightarrow \exists$ compact $K \subset E$ such that $m(E \setminus K) < \epsilon$;
(4) $m(E) < \infty \Longrightarrow \exists$ a finite union F of closed cubes such that $m(E \triangle F) < \epsilon$.[40]

Proof Sketch. 1. is from the definition of measurability of E.

2. is from the definition of measurability of E^c and then take complements.

3. is from (2) by taking $K = F \cap \overline{B}_N(0)$ for sufficiently large N.

4. is from the definition of $m(E)$ to first obtain $E \subset \bigcup_{j\geq 1} Q_j$ with the Q_j closed cubes and $\sum_{j=1}^{\infty} m(Q_j) \leq m(E) + \epsilon/2$.

Choose N so that $\sum_{j=N+1}^{\infty} m(Q_j) \leq \epsilon/2$ and let $F = \bigcup_{j=1}^{N} Q_j$. □

3.5. **Invariance Properties of Lebesgue measure.**

(1) If E is translated then the exterior measure is unchanged. If E is dilated in all directions by δ then (for subsets of $\mathbb{R}^d$) its volume is multiplied by the factor δ^d.
(2) Translations and dilations of measurable sets are measurable.

Proof The first from the definition $m_*(E)$ and a similar result for volumes of cubes. The second from the definition of measurability.

3.6. **Construction of a Non Measurable Set.**

(1) $x, y \in [0,1]$ are *equivalent* if $x - y \in \mathbb{Q}$. We write $x \sim y$.
  (a) $x \sim x$, $x \sim y \Longrightarrow y \sim x$, $x \sim y$ & $y \sim z \Longrightarrow x \sim z$. (Because of these properties we say "$\sim$" is an *equivalence relation*.)
  (b) It follows two equivalence classes are either identical or disjoint. So we can write $[0,1] = \bigcup_{a\in A} E_a$ as a disjoint union of equivalence classes.
  (c) Each E_a is countable and so A is uncountable.

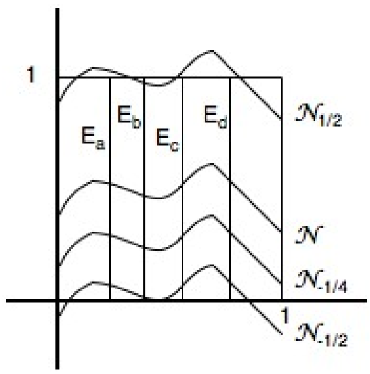

FIGURE 7. Schematic representation of proof of existence of a non measurable set.

(2) Let $\mathcal{N}$ contain exactly one element from each equivalence class. (This requires the uncountable axiom of choice.)
(3) For each $r \in \mathbb{Q}, -1 \leq r \leq 1$, let $\mathcal{N}_r = \{x + r : x \in \mathcal{N}\}$.
  (a) The $\mathcal{N}_r$ are disjoint and $[0,1] \subset \bigcup_r \mathcal{N}_r \subset [-1,2]$.
  (b) $m_*(\mathcal{N}_r) = m_*(\mathcal{N}) = m$ (say).
  (c) $\mathcal{N}$ is measurable $\Longrightarrow \mathcal{N}_r$ is measurable for all r.
(4) Now *assume* $\mathcal{N}$ is measurable. Then

$$1 \leq m + m + m + \dots \text{ (infinite series)} \leq 3.$$

This is a contradiction both if $m = 0$ and if $m > 0$.

Hence the assumption is wrong and so $\mathcal{N}$ is not measurable

(5) We have shown *countable* additivity fails for translated copies of $\mathcal{N}$.

In fact the argument also shows *finite* additivity fails. *Why?*

[40] $E \triangle F := (E \setminus F) \cup (F \setminus E)$ is the *set theoretic difference* between E and F.

By way of background and putting the material of the course in a broader perspective, we discussed the following. Note however that it is not examinable.

All of mathematics can be formally done within set theory, the axioms for set theory, the fact that the axiom of choice [AC] is of a different nature from the other axioms, and the question of whether one can prove AC from the other axioms.

In any model of the axioms of set theory without AC, one can construct a model in which AC is true (Gödel) and another model in which AC is false (Cohen). These results show that AC can neither be proved false nor proved true from the other axioms of set theory.

In previous lectures we used the uncountable axiom of choice to establish the existence of a non-measurable set. The uncountable AC, while very plausible, is highly non constructive. It also implies that the set of real numbers can be well-ordered, which is a counterintuitive result.

In fact it is possible to have a model of set theory in which the countable AC is true along with the other axioms of set theory and such that all sets in $\mathbb{R}^n$ are Lebesgue measurable. This means that we will never be able to prove the existence of a set of a nonmeasurable set if we only use the standard axioms of set theory and the countable AC — that is we will need to also use uncountable AC!

We also discussed the idea of building up the Borel sets by proceeding through the "transfinite hierarchy" of countable ordinals, and pointed out that the ordinals are well-ordered.

I also pointed out that in probability theory, where one uses other measures besides Lebesgue measure, nonmeasurable sets are not necessarily pathological. In particular we saw a simple example where a coin is thrown twice. The set of outcomes is HH, HT, TH, TT, where the first letter is the result of the first throw and the second is the result of the second throw. In the standard setup, the measurable sets after the first throw but before the second correspond to "events" which can be known at that time and consists of the sets $\emptyset$, $\{HH, HT\}$, $\{TH, TT\}$, $\{HH, HT, TH, TT\}$, while the measurable sets after the second throw are all 8 subsets of $\{HH, HT, TH, TT\}$.

3.7. **σ-Algebras and Borel sets.**

(1) *Defn*: A *σ-algebra of sets* is a collection of sets closed under countable unions, countable intersections, and complements.

(2) The collection $\mathcal{M}$ of Lebesgue measurable sets is a σ-algebra (proved previously).

(3) For *any* collection $\mathcal{C}$ of subsets of $\mathbb{R}^d$ there is a smallest σ-algebra containing $\mathcal{C}$.

Proof: $\mathcal{P}(\mathbb{R}^d)$ is a σ-algebra containing $\mathcal{C}$. The intersection of any family of σ-algebras is itself a σ-algebra (*why?*)

It follows that the intersection $\mathcal{S}$ of the family of *all* σ-algebras containing $\mathcal{C}$ exists, is a σ-algebra containing $\mathcal{C}$, and moreover if $\mathcal{T}$ is any σ-algebra containing $\mathcal{C}$ then $\mathcal{S} \subset \mathcal{T}$.

(4) *Defn*: The *σ-algebra $\mathcal{B}$ of Borel sets* is the smallest σ-algebra containing the open (and closed) subsets of $\mathbb{R}^d$.

(5) Since $\mathcal{M}$ contains the open sets it follows $\mathcal{B} \subset \mathcal{M}$. In fact $\mathcal{B}$ is a proper subcollection (later assignment problem).

(6) Countable intersection of open sets are called G_δ sets. Countable unions of closed sets are called F_σ sets. These are the simplest Borel sets after the open and closed sets.

(One obtains all Borel sets by taking countable intersections, countable unions and complements in a manner indexed by the countable ordinals. Discussed in class.)

(7) *E is measurable iff $\exists$ a G_δ set $A \supset E$ such that $m(A \setminus E) = 0$ iff $\exists$ an F_σ set $B \subset E$ such that $m(E \setminus B) = 0$.*

Proof Sketch: Measurability of E follows because G_δ sets, F_σ sets and null sets are measurable. If E is measurable then the other direction follows from 1. and 2. on page 37.

(8) *E is measurable iff $E = B \cup N$ where B is Borel and N is null.*

Exercise.

4. Measurable Functions

Mon 2/4

Just as essentially any set we come across is measurable, so will essentially any function be measurable.

It is impossible to obtain a non measurable set or a non measurable function without using the uncountable axiom of choice. Since we cannot prove the uncountable axiom of choice from the other axioms of set theory, including the countable axiom of choice, we are very unlikely to need to deal with non measurable sets or functions.[41]

Defn: A *simple function* is a function of the form $f = \sum_{k=1}^{N} a_k \mathcal{X}_{E_k}$ where each E_k is measurable with *finite* measure.[42]

(These will be the basic functions used to define the Lebesgue integral.)

Defn: A *step function* is a function of the form $f = \sum_{k=1}^{N} a_k \mathcal{X}_{R_k}$ where each R_k is a rectangle. (These are the basic functions used to define the Riemann integral.)

4.1. Definitions and Basic Properties.

Unless otherwise clear from context, $f : E \to \mathbb{R}$ where $E \subset \mathbb{R}^d$ is measurable.

More generally, $f : E \to [-\infty, \infty]$. In such cases the set where $f = \pm\infty$ usually has measure 0.

4.1.1. *Equivalent Definitions.*

(1) *Defn*: Suppose $E \subset \mathbb{R}^d$ is measurable and $f : E \to [-\infty, \infty]$. Then f is *measurable* if $f^{-1}[-\infty, a)$ is measurable for all $a \in \mathbb{R}$.

(2) *Notation*: For a fixed domain E we often denote $\{x \in E : f(x) < a\}$, i.e. $f^{-1}[-\infty, a)$, by $\{f < a\}$. Similarly for other sets.

(3) *f is measurable iff $\{f \le a\}$ is measurable for every $a \in \mathbb{R}$ iff $\{f \ge a\}$ is measurable for every $a \in \mathbb{R}$ iff $\{f > a\}$ is measurable for every $a \in \mathbb{R}$.*

Proof:

$$\{f \le a\} = \bigcap_{k=1}^{\infty} \{f < a + 1/k\}, \ \{f < a\} = \bigcup_{k=1}^{\infty} \{f \le a - 1/k\},$$
$$\{f \ge a\} = \{f < a\}^c, \ \{f > a\} = \{f \le a\}^c.$$

(4) *f is measurable iff $-f$ is measurable.*

Proof: *Exercise.*

(5) *f is measurable iff $\{a < f < b\}$ is measurable for every $a, b \in \mathbb{R}$, $\{f = -\infty\}$ is measurable and $\{f = +\infty\}$ is measurable.*

Similarly for any combination of strict or weak inequalities.

Proof: *Exercise.*

(6) *If f is finite valued, then f is measurable iff $f^{-1}[U]$ is measurable for every open U iff $f^{-1}[C]$ is measurable for every closed C.*

(This is also true for extended valued functions if we additionally require that $\{f = -\infty\}$ and $\{f = +\infty\}$ are measurable.)

Proof: The point is that every open $U \subset \mathbb{R}$ is a *countable* union of open intervals, and every closed $C \subset \mathbb{R}$ is the complement of an open set.

(7) *If f is finite valued, then f is measurable iff $f^{-1}[B]$ is measurable for every Borel B.*

(This is also true for extended valued functions if we additionally require that $\{f = -\infty\}$ and $\{f = +\infty\}$ are measurable.)

Proof: Since open sets are Borel, one direction is immediate.

The main point is to show that if f is measurable and B is Borel then $f^{-1}(B)$ is measurable.

For this let $\mathcal{S} = \{E : f^{-1}(E) \text{ is measurable}\}$. Then $\mathcal{S}$ contains every open set (*why?*) and is a σ-algebra (*why?*). It follows $\mathcal{S} \supset \mathcal{B}$ (*why?*).

---

[41] I refer here to Lebesgue measure. In probability theory the situation is very different. In the theory of stochastic processes one typically uses a hierarchy of measures indexed by time, and an event E is measurable at time t iff membership of E can be determined by information knowable up to time t.

[42] The *characteristic function* $\mathcal{X}_E$ of E is defined by $\mathcal{X}_E(x) = 1$ if $x \in E$ and $\mathcal{X}_E(x) = 0$ if $x \notin E$.

Tues 3/4, Thurs 5/4

4.1.2. *Functions obtained from Measurable Functions are usually Measurable.*

(1) *Prop*: *f continuous (and hence finite valued) $\Longrightarrow$ f measurable. f measurable and finite-valued and ϕ continuous $\Longrightarrow$ $\phi \circ f$ is measurable.*

Proof: Exercise. Note $(\phi \circ f)^{-1}(-\infty, a) = f^{-1}[\phi^{-1}(-\infty, a)]$.

(a) For example, f measurable $\Longrightarrow$ $|f|$ measurable.

(b) It is not true that f measurable and ϕ continuous implies $f \circ \phi$ is continuous. See Exercise 35.

(2) *Prop*: *If $(f_n)_{n\geq 1}$ is a sequence of measurable functions with the same measurable domain then*

$$\sup_n f_n, \inf_n f_n, \ \limsup_n f_n, \ \liminf_n f_n,$$

are measurable. In particular, if $f_n(x) \to f(x)$ for all x then f is measurable.

Proof: $\sup_n f_n(x) > a$ iff $f_n(x) > a$ for some n. Hence $\{\sup f_n > a\} = \bigcup_n \{f_n > a\}$. Similarly for "inf".

Note that $\limsup_n f_n(x) = \inf_k \{\sup_{n\geq k} f_n(x)\}$ and similarly for "lim inf".

(3) *If f and g are measurable with common (measurable) domain then so are f^2, λf for any real number λ, $f + g$ and fg. (In the last two cases we require f and g are finite valued so that $f + g$ and fg are well defined.)*[43]

Proof: The first is from 1. with $\phi(x) = x^2$. The second since $\{\lambda f > a\} = \{f > a/\lambda\}$. The third since[44] $\{f + g > a\} = \bigcup_{r\in\mathbb{Q}} \{f > a - r\} \cap \{g > r\}$. (See Figure 8.) The fourth since $fg = \frac{1}{4}\big((f + g)^2 + (f - g)^2\big)$.

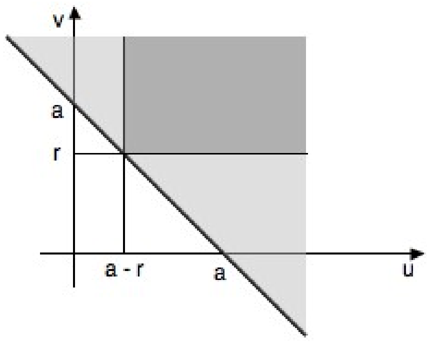

FIGURE 8. The shaded regions are $\{(u, v) : u + v > a\}$, the dark region is $\{(u, v) : u > a - r, \ v > r\}$. Note $u + v > a$ iff $u > a - r$ and $v > r$ for some $r \in \mathbb{Q}$. This is not true with $>$ replaced by $\geq$, why?

4.1.3. *Almost Everywhere.*

(1) In measure theory, we can generally neglect sets of measure 0. Recall that such sets are always measurable.

(2) *Defn*: A property of elements of $\mathbb{R}^d$ holds *almost everywhere*, written *a.e.*, if it holds for all but a set of elements of measure 0.

For example, $f = g$ a.e. means $m\{x : f(x) \neq g(x)\} = 0$.

(3) *If f is measurable and $f = g$ a.e. then g is measurable.*

Proof: Exercise.

(4) *If f_n are measurable and $f_n \to f$ a.e. then f is measurable.*

Proof: Exercise.

[43] For example, $\infty + a = \infty$, $\infty + \infty = \infty$, $\infty \times a$ are well defined if $a \neq 0$. But $\infty - \infty$ and $\infty \times 0$ are not well defined.

[44] It is clear that the right side is a subset of the left side. Conversely, suppose x is in the set given by the left side. Choose a rational r (sufficiently close to $g(x)$) such that $g(x) > r$ and $f(x) + g(x) - a > g(x) - r$. (*Why is this possible?*). It follows that $g(x) > r$ and $f(x) > a - r$.

4.2. **Approximating Measurable Functions.**

Theorem 4.1. *If $f : E\,(\subset \mathbb{R}^d) \to [0,\infty]$ is measurable then there exist simple functions ϕ_k s.t. $0 \le \phi_k(x) \uparrow f(x)$ for all $x \in E$.*

Proof. [45] First truncate f (draw diagram). For $x \in E$:

$$f_k(x) = \begin{cases} f(x) & |x| \le k,\ 0 \le f(x) \le k \\ k & |x| \le k,\ f(x) > k \\ 0 & \text{otherwise} \end{cases}$$

Then approximate f_k by a simple function to within error 2^{-k}. For $x \in E$:

$$\phi_k(x) = \frac{\ell}{2^k}, \quad \text{if } \frac{\ell}{2^k} \le f_k(x) < \frac{\ell+1}{2^k} \text{ for } \ell \ge 0 \text{ an integer.}$$

Check ϕ_k satisfies the required properties. □

Theorem 4.2. *If $f : E\,(\subset \mathbb{R}^d) \to [-\infty,\infty]$ is measurable then there exist simple functions ϕ_k s.t.*

$$0 \le \phi_k(x) \uparrow f(x) \text{ if } f(x) \ge 0, \quad 0 \ge \phi_k(x) \downarrow f(x) \text{ if } f(x) \le 0.$$

In particular, $|\phi_k(x)| \le |\phi_{k+1}(x)|$ and $\phi_k(x) \to f(x)$, for all x.

Proof. Write $f(x) = f^+(x) - f^-(x)$.[46] Construct $\phi_k^{(1)}$ and $\phi_k^{(2)}$ for f^+ and f^- as in the previous theorem. Let $\phi_k = \phi_k^{(1)} - \phi_k^{(2)}$. □

Theorem 4.3. *If $f : E\,(\subset \mathbb{R}^d) \to [-\infty,\infty]$ is measurable then there exist step functions ψ_k s.t. $\psi_k \to f$ a.e.*

4.3. **Remarks.**

(1) We only get a.e. convergence and we do not have the monotonicity properties of the previous two theorems.
(2) The first line of the proof in the text is misleading. The result does follow simply because we can approximate simple functions by step functions, it is necessary to consider the properties of the approximation. See proof below.

Proof of theorem.

(1) By Theorem 4.2 there is a sequence of simple functions $\phi_k \to f$ everywhere.
(2) Any simple function ϕ is of the form $\sum_j a_j \mathcal{X}_{A_j}$ for some *finite* collection of measurable sets A_j. We can also require these A_j be disjoint, by considering any intersections.
   (a) By "Approximating Measurable Sets 4." on page 37, for every measurable set A there is a finite union G of closed cubes such $m(A \triangle G) < \epsilon$.
   (b) G can be written as a sum of *almost disjoint* rectangles (consider the grid obtained by extending the sides of the cubes — draw a diagram).
   By taking the union $\tilde{G} \subset G$ of slightly smaller *disjoint* rectangles inside these rectangles we can ensure $m(G \setminus \tilde{G}) < \epsilon$ and so $m(A \triangle \tilde{G}) < 2\epsilon$.
   (c) By replacing A_j by A we obtain a step function ψ such that $\psi = \phi$ except on a set of measure 2ϵ.
(3) Hence *there exist step functions ψ_k such that if $E_k = \{\psi_k \ne \phi_k\}$ then $m(E_k) \le 2^{-k}$.*
(4) Let $F_n = \bigcup_{j \ge n+1} E_j$. Then $m(F_n) \le 2^{-n}$ and $\psi_k \to f$ except possibly on F_n.

---

[45] Proof in text not quite correct. To get $0 \le \phi_k(x) \uparrow f(x)$ rather than just $0 \le \phi_k(x) \to f(x)$ one needs to proceed as here with intervals of width 2^{-k}.

[46] $f^+(x) := f(x)$ if $f(x) \ge 0$, otherwise $f^+(x) := 0$. $f^-(x) := -f(x)$ if $f(x) \le 0$, otherwise $f^-(x) := 0$. (Note f^+ and f^- are both ≥ 0.) Then

$$f = f^+ - f^-, \ |f| = f^+ + f^-.$$

Hence

$$f^+ = \frac{1}{2}(|f| + f), \ f^- = \frac{1}{2}(|f| - f),$$

and so f^+ and f^- are measurable.

(5) Hence $\psi_k \to f$ except possibly on $F := \bigcap_{n\ge 1} F_n$. But $F_n \searrow F$ and so $m(F) = \lim_n m(F_n) = 0$.
Hence $\psi_k \to f$ a.e.

□

4.4. **Littlewood's Three Principles.** (Littlewood was a British mathematician (1885 – 1977) and a collaborator of G. H. Hardy.)

(1) Every set is "nearly" a finite union of cubes.[47]
See "Approximating Measurable Sets" 4., page 37.
(2) Every "function" is nearly continuous.
See Lusin's Theorem below.
(3) Every convergent sequence of functions is "nearly" uniformly convergent.
See Egorov's Theorem below.

Theorem 4.4 (Egorov)**.** *Suppose the measurable functions $f_k \to f$ a.e. on the measurable set E, where $m(E) < \infty$. Then for each $\epsilon > 0$ there is a closed $A \subset E$ with $m(E \setminus A) < \epsilon$ such that $f_k \to f$ uniformly on A.*

4.5. **Remarks on Egorov's Theorem.**

(1) Let $E = [-1, 1]$,

$$f_k(x) = \begin{cases} 0 & -1 \le x \le 0 \\ kx & 0 \le x \le \frac{1}{k} \\ 2 - kx & \frac{1}{k} \le x \le \frac{2}{k} \\ 0 & \frac{2}{k} \le x \le 1 \end{cases}, \quad f(x) = 0\ \forall x.$$

Then $f_k \to f$ pointwise on E, and uniformly on $[-1, 0] \cup [\epsilon, 0]$ for any $\epsilon > 0$. *Draw a diagram.*

(2) $m(E) < \infty$ is necessary. Consider $E = \mathbb{R}$ and $f_k(x) = x/k$.

Mon 24/4

Proof sketch of Egorov's Theorem. (Think of the above example 1.)

(1) By changing f_k and f on a set of measure 0, $f_k \to f$ everywhere on E.
(2) Let ("bad set") $B_k^n = \{x \in E : |f_j(x) - f(x)| \ge 1/n \text{ for some } j \ge k\}$.
Informally, "$x \in B_k^n$ if x is $1/n$ bad for *some* f_j with $j \ge k$".
The complement of B_k^n is the "good set" G_k^n,
i.e. $x \in G_k^n := (B_k^n)^c$ iff $|f_j(x) - f(x)| < 1/n$, $\forall j \ge k$.
Informally, "$x \in G_k^n$ if x is $1/n$ good for *all* f_j with $j \ge k$".
(3) Fix n. Then $B_k^n \downarrow \emptyset$ as $k \to \infty$ since $f_j(x) \to f(x)$ for all x.
(4) Fix $\epsilon > 0$. Then for each n, $\exists k = k(n)$ such that $m(B_k^n) \le \epsilon/2^n$.
Let $B = \bigcup_n B_{k(n)}^n$. Then
 (a) $m(B) \le \epsilon$,
 (b) $x \notin B \implies \forall n \Big(|f_j(x) - f(x)| < 1/n \text{ if } j \ge k(n) \Big)$.
That is, $f_j \to f$ uniformly on $E \setminus B$ where $m(B) \le \epsilon$.
(5) Since $E \setminus B$ is measurable, $\exists$ closed $A \subset E \setminus B$ such that $m\big((E \setminus B) \setminus A)\big) < \epsilon$. Then $f_j \to f$ uniformly on A and $m(E \setminus A) < 2\epsilon$.

□

Theorem 4.5 (Lusin)**.** *Suppose $f : E\, (\subset \mathbb{R}^d) \to \mathbb{R}$ where E is measurable. Then for each $\epsilon > 0$ there exists a closed $F \subset E$ such that $m(E \setminus F) < \epsilon$ and $f|_F$ is continuous.*

Moreover, there is a continuous function $\widetilde{f} : E \to \mathbb{R}^d$ such that $\widetilde{f} = f$ on F.

[47] Littlewood, thinking of $\mathbb{R}$, said "intervals".

4.6. **Remarks on Lusin's Theorem.**

(1) $f|_F$ is the restriction of f to F. It is not necessarily true that f is continuous on F.
For example, let $f = \mathbb{R} \to \mathbb{R}$ where $f = 0$ on the set $\mathbb{Q}$ of rationals and $f = 1$ on the set $\mathbb{I}$ of irrationals. Then f is nowhere continuous, but $f|_{\mathbb{I}} \equiv 0$ is continuous everywhere on its domain $\mathbb{I}$ ($\mathbb{I}$ is the set of irrationals).
Now choose a closed $F \subset \mathbb{I}$ (closed in $\mathbb{R}$) s.t. $m(\mathbb{I} \setminus F) < \epsilon$. Then $m(\mathbb{R} \setminus F) < \epsilon$ and $f|_F$ is continuous.

(2) What is a suitable $\widetilde{f}$ in this case?

(3) The statement in the text requires $m(E) < \infty$. This is not necessary, see the proof.

Proof of Lusin's Theorem. First assume $m(E) < \infty$.

(1) Let $f_n \to f$ a.e. on E where the f_n are *step functions*. (By a previous approximation result.)

(2) Since f_n is a step function, $f_n = \sum_{k=1}^N a_k \mathcal{X}_{R_k}$ where the R_k are rectangles.
By shrinking the rectangles a little, f_n restricted to $E \setminus B_n$ is continuous, where $m(B_n) < \epsilon 2^{-n}$.

(3) Let $B = \bigcup_{n \ge 1} B_n$.
Then $m(B) < \epsilon$, f_n restricted to $E \setminus B$ is continuous, $f_n \to f$ a.e. on $E \setminus B$.

(4) By Egorov, after removing another small set B' with $m(B') \le \epsilon$, $f_n \to f$ uniformly on $E \setminus (B \cup B')$, f_n restricted to $E \setminus (B \cup B')$ are continuous.

(5) Hence f restricted to $F' := E \setminus (B \cup B')$ is continuous, being a *uniform* limit of *continuous* functions.

(6) By taking a slightly smaller *closed* $F \subset F'$ we obtain $f|_F$ is continuous and $m(E \setminus F) < 3\epsilon$. This completes the proof (with ϵ replaced by 3ϵ) in case $m(E) < \infty$.

For $m(E) = \infty$

(1) Let $E = \bigcup_{k \ge 1} E_k$ where the E_k are bounded, and any bounded set B meets at most a *finite* number of the E_k. *Why is this possible?*

(2) For each k, take a closed $F_k \subset E_k$ where $m(E_k \setminus F_k) < \epsilon 2^{-k}$ and $f|_{F_k}$ is continuous. Note that the F_k are compact and disjoint.

(3) Let $F = \bigcup_{k \ge 1} F_k$. Then F is closed[48] and $m(E \setminus F) < \epsilon$.

(4) Since each $f|_{F_k}$ is continuous so is $f|_F$. (Consider a sequence $(x_j)_{j \ge 1} \subset F$, $x_j \to x$. Then eventually $x_j \in F_k$, say, and now use the continuity of $f|_{F_k}$.)

For the last part of the theorem: By the Tietze extension theorem any continuous function defined on a closed subset of $\mathbb{R}$ can be extended to a continuous function defined on all of $\mathbb{R}$. □

[48] In general, a countable union of closed sets need not be closed. However, any convergent sequence $(x_j)_{j \ge 1} \subset F$ is bounded and so is a subset of the union of *finitely* many of the F_k, *why?* The union is closed and so it contains the limit x of the sequence, and so $x \in F$.

Part 4. INTEGRATION THEORY

1. Lebesgue Integral: basic properties

Tues 24/4

All sets and functions assumed measurable, unless otherwise stated.

The integral and its properties is developed for

(1) simple functions
(2) bounded functions supported on a set of finite measure
(3) positive functions
(4) arbitrary (real valued) functions

It is also defined for complex valued functions.

Motivation: For $f : \mathbb{R}^d \to \mathbb{R}$, $\int f$ will equal (the $d+1$ measure of the region above the $\mathbb{R}^d$ plane and below the graph) *minus* (the $d+1$ measure of the region below the $\mathbb{R}^d$ plane and above the graph).

1.1. Simple functions.

1.1.1. *Properties of simple functions.*

(1) Recall these are functions defined on $\mathbb{R}^d$ of the form $\phi = \sum_{k=1}^N a_k \mathcal{X}_{E_k}$ where $m(E_k) < \infty$.
(2) The *canonical form* is when a_k are all distinct and E_k are disjoint.
To put ϕ in canonical form let $\{c_1, \ldots, c_M\}$ be the set of values of ϕ and $F_k = \{\phi = c_k\}$. Then $\phi = \sum_{k=1}^M c_k \mathcal{X}_{F_k}$.

1.1.2. *Integral of simple functions. Definition*: If $\phi = \sum_{k=1}^M c_k \mathcal{X}_{F_k}$ is in canonical form then define

$$\int \phi = \int \phi(x)\,dx = \sum_{k=1}^M c_k m(F_k).$$

If E has finite measure then define

$$\int_E \phi = \int \phi \mathcal{X}_E.$$

1.1.3. *Properties.* Analogues of the following 6 properties will hold in the subsequent more general situations.

(1) *Independence of the representation*: Even if $\phi = \sum_{k=1}^N a_k \mathcal{X}_{E_k}$ is not in canonical form,
$$\int \phi = \sum_{k=1}^N a_k m(E_k).$$
Proof idea: The main ideas can be seen from doing the examples
  (a) $a\mathcal{X}_A + a\mathcal{X}_B$ where A and B are disjoint but the two coefficients are equal, *exercise.*
  (b) $a\mathcal{X}_A + b\mathcal{X}_B$ where $A \cap B \neq \emptyset$, *exercise.*
(2) *Linearity*: if ϕ and ψ are simple then
$$\int (a\phi + b\psi) = a\int \phi + b\int \psi.$$
Follows from 1, *exercise.*
(3) *Additivity*: if ϕ simple and E, F disjoint then
$$\int_E \phi + \int_F \phi = \int_{E\cup F} \phi.$$
Follows from 2, noting $\mathcal{X}_{E\cup F} = \mathcal{X}_E + \mathcal{X}_F$. *exercise.*
(4) *Monotonicity*: if $\phi \le \psi$ are simple then
$$\int \phi \le \int \psi.$$
Follows from 2 applied to $\psi - \phi$. *exercise.*

(5) *"Triangle inequality"*: if ϕ is simple then so is $|\phi|$ and
$$\left|\int \phi\right| \leq \int |\phi|.$$
Use 1.

(6) *Sets of measure 0 do not count*: If ϕ and ψ are simple and $\phi = \psi$ a.e. then
$$\int \phi = \int \psi.$$
Use 2 on $\phi - \psi$.

1.2. Bounded functions, finite measure support.

1.2.1. *Support of a function.*

(1) *Definition*: The *support* of f is $\{x : f(x) \neq 0\}$.[49]
f is *supported on* F if $\{x : f(x) \neq 0\} \subset F$.

(2) *Recall*: Suppose f is supported on E with $|f(x)| \leq M$ for all x. Then there is a sequence of simple functions $(\phi_n)_{n\geq 1}$ supported on E, with $|\phi_n(x)| \leq M$ for all x, such that $\phi_n(x) \to f(x)$ for all x.
(Follows immediately from Theorem 4.2.)

1.2.2. *Important Lemma.* Let f be a bounded function supported on a set E of finite measure. If $(\phi_n)_{n\geq 1}$ are simple functions supported on E, bounded in absolute value by M, and $\phi_n(x) \to f(x)$ for a.e. x, then:

(1) $\lim_{n\to\infty} \int \phi_n$ exists.
(2) If $f = 0$ a.e. then $\lim_{n\to\infty} \int \phi_n = 0$.
(3) If $(\psi_n)_{n\geq 1}$ are also simple functions supported on E, bounded in absolute value by M, and $\psi_n(x) \to f(x)$ for a.e. x, then
$\lim_{n\to\infty} \int \phi_n = \lim_{n\to\infty} \int \psi_n$.

Remarks:

(1) We will *define* $\int f = \lim_{n\to\infty} \int \phi_n$
(2) *Avoid blow up*! It is necessary that $|\phi_n(x)| \leq M$ for all x. Take $E = [0, 1]$, $f(x) = 0$ for all x, $\phi_n = n\mathcal{X}_{(0,1/n]}$.
(3) *Avoid escape to* ∞! It is necessary that the ϕ_n be supported on E. Take $E = [0, 1]$, $f(x) = 0$ for all x, $\phi_n = \frac{1}{n}\mathcal{X}_{[0,n]}$.

Proof. **1.** Suppose $\epsilon > 0$.
By Egorov, except on a set B (bad but small) of measure $< \epsilon$, $\phi_n \to f$ uniformly on $E \setminus B$.
So for all n sufficiently large:
$$\begin{aligned}\left|\int \phi_n - \int \phi_m\right| &\leq \int_{E\setminus B} |\phi_n - \phi_m| + \int_B |\phi_n - \phi_m| \\ &\leq \epsilon\, m(E \setminus B) + 2Mm(B) \quad \text{(using uniform cgce on } E \setminus B) \\ &\leq \epsilon\, m(E) + 2M\epsilon.\end{aligned}$$
Since $\epsilon > 0$ is arbitrary this means the sequence $(\int \phi_n)_{n\geq 1}$ is Cauchy and so has a limit.

Thurs 26/4

2. If $\phi_n(x) \to 0$ for a.e. x then a similar argument shows for all n sufficiently large
$$\left|\int \phi_n\right| \leq \epsilon\, m(E) + 2M\epsilon.$$
This implies $\lim_{n\to\infty} \int \phi_n = 0$.
3. Apply 2. to $\phi_n - \psi_n$. ☐

[49]Warning: Usually the support of a function is defined as the smallest *closed* set where the function is non-zero, i.e. the intersection of all closed sets on which the function is non zero.

1.2.3. *Definition.* If f is bounded and supported on a set E of finite measure then we define its *Lebesgue integral* by

$$\int f = \lim_{n\to\infty} \int \phi_n$$

for any sequence ϕ_n of simple functions supported on E, bounded in absolute value by some fixed M, with $\phi_n(x) \to f(x)$ for a.e. x.

For any such f and any $A \subset \mathbb{R}^d$ define

$$\int_A f = \int f\mathcal{X}_A.$$

Note: The first definition makes sense and is independent of the approximating sequence by the previous lemma. The second definition makes sense since if f is bounded and supported on a set of finite measure, then so is $f\mathcal{X}_A$.

1.2.4. *Properties.* Suppose f and g are bounded and supported on a set of finite measure. Then the following hold:

(1) *Linearity*:

$$\int (af + bg) = a\int f + b\int g.$$

Proof: Take sequences of simple functions (uniformly bounded and supported on a fixed set of finite measure) converging a.e. to f and g. *exercise.*

(2) *Additivity*: If E and F are disjoint then

$$\int_E f + \int_F f = \int_{E\cup F} f.$$

Proof: Follows from 1. *exercise.*

(3) *Monotonicity*: if $f \le g$ then

$$\int f \le \int g.$$

Proof: $\int f - \int g = \int (f-g)$ from 1. But $f - g \ge 0$ and so $\int(f-g) \ge 0$,*why*?

(4) *"Triangle inequality"*:

$$\left|\int f\right| \le \int |f|.$$

Proof: If $\phi_n \to f$ then $|\phi_n| \to |f|$. Now use the corresponding result for simple functions.

(5) *Sets of measure 0 do not count*: If $f = g$ a.e. then

$$\int f = \int g.$$

Proof: By 1. $\int f - \int g = \int (f-g) = 0$. For the second equality note that $h = 0$ a.e. implies $\int h = 0$, *why?*

1.2.5. *Bounded convergence theorem.* Assume $m(E) < \infty$. If $(f_n)_{n\ge1}$ is a sequence of functions supported on E, bounded in absolute value by M, and $f_n(x) \to f(x)$ for a.e. x, then

$$\lim_{n\to\infty} \int |f_n - f| = 0.$$

In particular,

$$\lim_{n\to\infty} \int f_n = \int f.$$

Proof. Suppose $\epsilon > 0$.

By Egorov, except on a set B (bad but small) of measure $< \epsilon$, $f_n \to f$ uniformly on $E \setminus B$.

So for all n sufficiently large:

$$\begin{aligned}\int |f_n - f| &= \int_{E\setminus B} |f_n - f| + \int_B |f_n - f| \\ &\le \epsilon\, m(E\setminus B) + 2Mm(B) \quad \text{(using uniform cgce on } E\setminus B) \\ &\le \epsilon\, m(E) + 2M\epsilon.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary this gives the first claim.

For the second, just note $\left|\int f_n - \int f\right| \le \int |f_n - f|$. □

Remarks:

(1) Proof is essentially the same as for the "Important Lemma" on page 45. The difference is that by this point we can use the basic properties of $\int f_n$ for f_n not just a simple function.

(2) The remaining convergence theorems are now relatively straightforward consequences.

(3) The bounded convergence theorem is in fact a special case of the slightly more general dominated convergence theorem, see later.

(4) *$f \ge 0$, f bounded, f supported on a set of finite measure, and $\int f = 0$, imply $f = 0$ a.e.*

(We later drop the requirements f bounded and f supported on a set of finite measure.)

Proof: Suppose $\{f > 0\}$ has positive measure. Then $F_n := \{f \ge 1/n\}$ has positive measure for some integer $n > 0$, *why?*. But then $\int f \ge \int \frac{1}{n}\mathcal{X}_{F_n} = \frac{1}{n} m(F_n) > 0$, contradiction.

Mon 30/4

1.2.6. *Riemann Integration.*

Theorem. *If f is Riemann integrable on $[a, b]$ then f is measurable and $\int_{[a,b]} f = R\int_a^b f$, where the second term is the Riemann integral.*

Proof. Assume $f : [a, b] \to \mathbb{R}$ is Riemann integrable. In particular, by definition, f is bounded.[50] In the following proof, "$\int$" denotes a Lebesgue integral and "$R\int$" denotes a Riemann integral.

1. By taking lower and upper sums in the definition of the Riemann integral, there are corresponding step[51] functions ϕ_k and ψ_k such that

$$\phi_k \uparrow \phi \le f, \quad \psi_k \downarrow \psi \ge f, \quad \int_{[a,b]} \phi_k \uparrow R\int_a^b f, \quad \int_a^b \psi_k \downarrow R\int_a^b f.$$

Why? Note that ϕ_k and ψ_k are measurable, and so are ϕ and ψ.

2. But

$$\int_{[a,b]} \phi_k \uparrow \int_{[a,b]} \phi, \quad \int_{[a,b]} \psi_k \downarrow \int_{[a,b]} \psi,$$

by the bounded convergence theorem. Note that by construction, all the ϕ_k and ψ_k are uniformly bounded by the upper and lower bounds for f. *Why?*

3. It follows from **1.** and **2.** that

$$\int_{[a,b]} \phi = \int_{[a,b]} \psi = R\int_a^b f.$$

4. Since $\phi \le \psi$ (*why?*) and $\int_{[a,b]} \phi = \int_{[a,b]} \psi$, it follows from the above Remark (4) applied to $\psi - \phi$ that $\psi = \phi$ a.e.

Since $\phi \le f \le \psi$ it now follows $\phi = f = \psi$ a.e. and in particular it follows f is measurable. Moreover, it then follows by monotonicity of the integral and the first equality in **3.** that

$$\int_{[a,b]} \phi = \int_{[a,b]} \psi = \int_{[a,b]} f,$$

why?

5. It follows from **3.** and **4.** that

$$\int_{[a,b]} f = R\int_a^b f.$$

□

[50] See Chapter 7 of my First Year Calculus/Analysis notes.

[51] Here step functions are allowed to use intervals $[a, b], (a, b], [a, b), (a, b)$.

Remarks

(1) A similar result and proof applies to functions of more than one variable.
(2) This allows us to compute integrals of standard functions in the usual manner.

1.3. Non negative functions.

Definition. Suppose $f : \mathbb{R}^d \to [0, \infty]$. Define

$$\int f = \sup\left\{\int g : 0 \le g \le f,\ g \text{ bdd}, g \text{ supported on a set of finite measure}\right\}.$$

If $E \subset \mathbb{R}^d$ then we define

$$\int_E f = \int f \mathcal{X}_E.$$

If $\int f < \infty$ then f is said to be *(Lebesgue) integrable.*

Example

(1) $|x|^{-\alpha}$ is integrable on the unit ball in $\mathbb{R}^d$ iff $\alpha < d$.
(2) $|x|^{-\alpha}$ is integrable on the complement of the unit ball in $\mathbb{R}^d$ iff $\alpha > d$.

The proof is by approximating by bounded functions and using results for Riemann integration.

1.3.1. *Properties.* Suppose $f \ge 0$ and $g \ge 0$. Then the following hold:

(1) *Linearity*: If $a, b \ge 0$ then

$$\int (af + bg) = a\int f + b\int g.$$

Proof: First check $\int af = a\int f$. *Why* is this true?
So it is sufficient to show $\int (f + g) = \int f + \int g$.

For "$\le$" consider any bounded simple ϕ supported on a set of finite measure such that $\phi \le f$, and similarly for $\psi \le g$. Since

$$\int \phi + \int \psi = \int \phi + \psi \le \int f + g$$

(*why?*), it follows $\int f + \int g \le \int f + g$, *why?*.

For "$\ge$" consider any bounded simple η supported on a set of finite measure with $\eta \le f + g$. Write

$$\eta = \eta_1 + \eta_2$$

where $\eta_1 = \min\{f, \eta\}$ and $\eta_2 = \eta - \eta_1$.

Then $\eta_1 \le f$, and η_1 is bounded with bounded support since $0 \le \eta_1 \le \eta$ and η is bounded with bounded support. Moreover, $\eta_2 \le g$ since

$$\eta_1(x) = f(x) \Longrightarrow \eta_2(x) = \eta(x) - \eta_1(x) \le f(x) + g(x) - f(x) = g(x),$$
$$\eta_1(x) = \eta(x) \Longrightarrow \eta_2(x) = \eta(x) - \eta_1(x) = 0.$$

Also, η_2 is bounded with bounded support since $\eta_2 = \eta - \eta_1$ and both η and η_1 are bounded with bounded support.

It follows from properties of the integral for simple functions that

$$\int \eta = \int \eta_1 + \eta_2 = \int \eta_1 + \eta_2 \le \int f + \int g.$$

Taking sups on the left side gives the result.

(2) *Additivity*: If E and F are disjoint then

$$\int_E f + \int_F f = \int_{E\cup F} f.$$

Proof: *exercise.*

(3) *Monotonicity*: if $f \le g$ then
$$\int f \le \int g.$$
Proof: $\int g - \int f = \int (g-f)$ from 1. But $g - f \ge 0$ and so $\int (g-f) \ge 0$, *why*?

(4) *If f is integrable then $m(\{f = +\infty\}) = 0$.*
Proof: Similar idea to Remark 4. page 47. See text.

(5) *$f \ge 0$ and $\int f = 0$, imply $f = 0$ a.e.*
Proof: Same as Remark 4. page 47.

Tues 1/5

1.3.2. *Fatou's Lemma.* Suppose $f_n \ge 0$ for all n, and $f_n \to f$ a.e. Then
$$\int f \le \liminf_{n\to\infty} \int f_n.$$

Remarks

(1) Escape to infinity and blowup, see remarks on page 45, both show we cannot expect equality without further conditions.
(2) We allow $\int f = \infty$ and $\liminf f_n = \infty$.

Proof of Fatou's Lemma. We reduce to the bounded convergence theorem by considering $0 \le g \le f$ were g is bounded and supported on a set of finite measure.

Let $g_n = \min\{f_n, g\}$. Then g_n are uniformly bounded and supported on a fixed set of finite measure. Moreover, $g_n \to g$ at every point where $f_n \to f$, *why?* So $\int g = \lim \int g_n$ by the bounded convergence theorem.

Since $g_n \le f_n$,
$$\int g = \lim \int g_n \le \liminf \int f_n.$$
Why? Taking sup over allowable g on the left side gives the result. □

The following is a special case of the Dominated Convergence Theorem on page 51, except in the case $\int f = \infty$.

Theorem. *Suppose $0 \le f_n \le f$ and $f_n \to f$ a.e. Then*
$$\lim \int f_n = \int f \quad (\textit{we allow} \int f = \infty).$$
In particular, the limit exists.

Proof. We have
$$\liminf \int f_n \le \limsup \int f_n \le \int f \le \liminf \int f_n.$$
The first inequality is trivial, the second since $\int f_n \le \int f$ for all n, the third by Fatou's lemma.

The result follows, *why?* □

An important case is:

Theorem (Monotone Convergence Theorem)**.** *Suppose $0 \le f_n \uparrow f$ a.e.*[52] *Then*
$$\int f = \lim \int f_n \quad (\textit{we allow} \int f = \infty).$$

The following is an immediate and useful corollary, and allows us to exchange summation and integration in many cases.

Theorem. *Suppose $g_k \ge 0$ for all k. Then*
$$\int \sum_{k=1}^{\infty} g_k = \sum_{k=1}^{\infty} \int g_k.$$

[52] $f_n \uparrow f$ means $f_n(x)$ is an increasing sequence for a.e. x, and $f_n(x) \to f(x)$ a.e.

Proof. Let $f_k = g_1 + \cdots + g_k$ and apply the previous theorem. *Exercise.* □

See text for two examples.

1.4. Arbitrary functions.

Definition. Write $f = f^+ - f^-$ as in Footnote 46. Then we define the *(Lebesgue) integral* by

$$\int f = \int f^+ - \int f^-.$$

Remarks

(1) This makes sense unless *both* $\int f^+ = \infty$ and $\int f^- = \infty$.
(2) If both terms are finite we say f is *integrable.*
(3) if g and h are integrable and $f = g - h$ where $g \geq 0$ and $h \geq 0$, then $\int f = \int g - \int h$. (It is not necessary that $g = f^+$ and $h = f^-$.)
To see this, write $g + f^- = h + f^+$ and integrate both sides.
(4) The integrability of f and the value of $\int f$ are unaffected by changing f on a set of measure 0. *Why?*

1.4.1. *Properties.* The integral of an integrable function is linear, additive, monotone and satisfies the triangle inequality.
Exercise.

Remark. Now that we can integrate functions that are not necessarily ≥ 0, it is easy to show that the condition $f_n \geq 0$ in Fatou's Lemma and the next two theorems in Section 1.3 can be replaced by $f_n \geq h$ for some h such that $\int h$ is finite. *Why?* The interesting new case is $h \leq 0$.

Thurs 3/5

Theorem (Only a little bit of f is far away)**.** *Suppose f is integrable. Then for each $\epsilon > 0$ there exists a ball $B_R(0)$ such that*

$$\int_{B_R(0)^c} |f| \leq \epsilon.$$

Proof. Consider $|f|\mathcal{X}_{B_n(0)} \uparrow |f|$ and apply the monotone convergence theorem. □

Theorem (Absolute continuity of the integral)**.** *Suppose f is integrable. Then for each $\epsilon > 0$ there exists $\delta > 0$ such that*

$$m(S) \leq \delta \Longrightarrow \int_S |f| \leq \epsilon.$$

Proof. (This is a bit tricky)

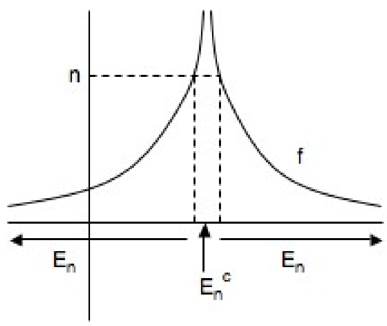

FIGURE 9. Diagram for proof of absolute continuity

Without loss of generality we may (and do) assume $f \geq 0$. *Why is this O.K.?* Define $E_n = \{f \leq n\}$. For any $S \subset \mathbb{R}^d$,

$$\int_S f = \int_{S\cap E_n} f + \int_{S\cap E_n^c} f. \tag{11}$$

For the second integral on the right in (11) we have

$$\int_{S\cap E_n^c} f \leq \int_{E_n^c} f = \int (f - f\mathcal{X}_{E_n}) \to 0$$

by the monotone convergence theorem, since $0 \leq f\mathcal{X}_{E_n} \uparrow f$ a.e. Hence there exists $N = N(\epsilon)$ such that

$$\int_{S\cap E_N^c} f < \epsilon/2. \tag{12}$$

Note that N depends on ϵ but not on the set S.

For the first integral on the right in (11), we have

$$\int_{S\cap E_N} f \leq N\, m(S) < \frac{\epsilon}{2}, \tag{13}$$

provided $m(S) < \epsilon/(2N(\epsilon))$ for the last inequality.

It follows from (12) and (13) that

$$\int_S f < \epsilon$$

provided $m(S) < \epsilon/(2N(\epsilon))$. □

Theorem (Dominated convergence theorem). *Suppose $f_n \to f$ a.e. Suppose $|f_n| \leq g$ a.e. where g is integrable. Then*

$$\int |f_n - f| \to 0 \quad \textit{as } n \to \infty,$$

and hence

$$\int f_n \to \int f \quad \textit{as } n \to \infty.$$

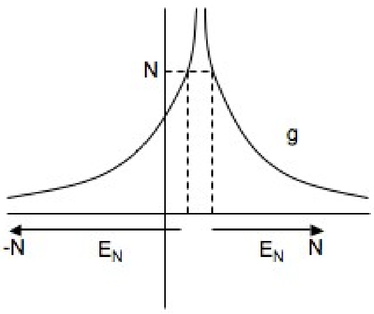

FIGURE 10. Diagram for proof of dominated convergence theorem

Proof. Fix $\epsilon > 0$.

1. First define

$$E_N = \{x : |x| \leq N,\ g(x) \leq N\}.$$

Then

$$g\mathcal{X}_{E_N} \uparrow g \text{ a.e. as } N \to \infty.$$

So we can[53] find and fix N such that

$$\int_{E_N^c} g < \epsilon.$$

[53]As in the proof of the Theorem on page 50.

2. Note

$$f_n \mathcal{X}_{E_N} \to f \mathcal{X}_{E_N} \text{ a.e. as } N \to \infty.$$

By the bounded convergence theorem

$$\int_{E_N} |f - f_n| \to 0, \quad \text{so} \int_{E_N} |f - f_n| < \epsilon$$

for all $n \geq n_0$, say.

Hence

$$\begin{aligned} \int |f - f_n| &\leq \int_{E_N} |f - f_n| + \int_{E_N^c} |f - f_n| \\ &\leq \epsilon + 2 \int_{E_N^c} g \quad (\textit{why?}) \\ &\leq \epsilon + 2\epsilon = 3\epsilon, \end{aligned}$$

for $n \geq n_0$. This proves the theorem. □

1.5. **Complex valued functions.**

(1) If $f(x) = u(x) = iv(x)$ then define $\int f = \int u + i \int v$.
(2) We say f is *integrable* iff u and v are integrable. This is equivalent to the real valued function $|f|$ being integrable. (Since $|u|, |v| \leq |f|$ and $|f| \leq |u| + |v|$.)
(3) The "triangle inequality" also holds for the integral of complex valued functions f. That is, $|\int f| \leq \int |f|$.

To see this we use the following "trick". Choose (the constant) θ so that $e^{i\theta} \int f$ is non-negative real.

Let $\widetilde{f} = e^{i\theta} f$, so $\int \widetilde{f}$ is non-negative real. Let $\widetilde{f} = \widetilde{u} + i\widetilde{v}$ where $\widetilde{u}$ and $\widetilde{v}$ are real valued functions. Note that $\int \widetilde{f} = \int \widetilde{u} \geq 0$, *why?*

It follows that

$$\left| \int f \right| = \left| \int \widetilde{f} \right| = \int \widetilde{f} = \int \widetilde{u} \leq \int |\widetilde{u}| \leq \int |\widetilde{f}| = \int |f|.$$

Why is each equality and inequality justified? □

2. The Normed Space L^1

Mon 7/5

We also discussed basic properties of normed spaces, inner product spaces, etc.

2.1. Basic Properties.

(1) The space $L^1 = L^1(\mathbb{R}^d)$ is the set of real-valued *integrable* functions defined on $\mathbb{R}^d$.
The norm $\|f\|$ is defined by
$$\|f\|_{L^1} = \|f\| = \int |f|.$$
(2) L^1 is a vector space essentially because it is closed under addition and scalar multiplication. (This implies it is a subspace of the vector space of *all* functions defined on $\mathbb{R}^d$.)
(3) $\|\cdot\|$ is a "norm" since
 (a) $\|f+g\| \le \|f\| + \|g\|$,
 (b) $\|af\| = |a|\|f\|$ for $a \in \mathbb{R}$,
 (c) $\|f\| = 0$ iff $f = 0$ a.e.
So we see it is not really a norm, unless we "identify" functions that are 0 a.e. More precisely, we could work with equivalence classes of functions, where two functions are equivalent if they are equal a.e.
(4) As in any normed space we can define a corresponding "metric". Here
$$d(f,g) = \|f-g\| = \int |f-g|.$$
This satisfies the triangle inequality, is symmetric, and $d(f,g) = 0$ iff $f = g$ a.e. So again it is not really a metric, unless we identify functions that are 0 a.e.
(5) We say $f_n \to f$ in L^1 sense if $f_n \to$ f in the sense of the metric, i.e. if $\|f_n - f\| \to 0$ as $n \to \infty$.
(6) We will see later that L^1 is complete.
(7) If we work with complex valued functions we similarly get a complex vector space.

2.2. Relationship between L^1 and a.e. convergence.

(1) We have seen (escape to infinity and blowup) that $f_n \to f$ a.e., or even everywhere, does not imply $f_n \to f$ in L^1 sense.
(2) However, the dominated convergence theorem gave an additional assumption under which a.e. convergence does imply L^1 convergence.
(3) L^1 convergence does not imply a.e. convergence.
For example, take the following sequence f_n of "moving blips" given by
$$\mathcal{X}_{[0,1]}, \mathcal{X}_{[0,1/2]}, \mathcal{X}_{[1/2,1]}, \mathcal{X}_{[0,1/3]}, \mathcal{X}_{[1/3,2/3]}, \mathcal{X}_{[2/3,1]}, \mathcal{X}_{[0,1/4]}, \mathcal{X}_{[1/4,2/4]}, \mathcal{X}_{[2/4,3/4]}, \mathcal{X}_{[3/4,1]}, \mathcal{X}_{[0,1/5]}, \dots, \mathcal{X}_{[4/5,1]}, \mathcal{X}_{[0,1/6]}, \dots, \mathcal{X}_{[5/6,1]}, \dots$$
The sequence $\int f_n$ is
$$1, 1/2, 1/2, 1/3, 1/3, 1/3, 1/4, 1/4, 1/4, 1/4, 1/5, \dots, 1/5, 1/6, \dots, 1/6, \dots$$
So $f_n \to \mathbf{0}$ in the L^1 sense, where $\mathbf{0}$ is the zero function.
But $f_n \not\to f$ anywhere on $[0,1]$.
(4) But if $f_n \to f$ sufficiently fast in the L^1 sense, then $f_n \to f$ a.e. See the next theorem.

Tues 8/5

Remark. The following theorem applies if $f_n \to f$ geometrically fast in the L^1 sense, that is $\|f_n - f\| \le Ar^n$ for some $A > 0$ and $0 \le r < 1$.

Theorem. *Suppose $f_n \to f$ fast, in the sense that $\sum_{n\ge1} \|f_n - f\| < \infty$.*[54] *Then $f_n \to f$ a.e.*

[54] Recall that for a sequence of positive real numbers, $a_n \to 0$ does not imply $\sum_n a_n < \infty$. For example, $1/n \to 0$ but $\sum_n 1/n = \infty$. But of course for positive a_n, $\sum_n a_n < \infty$ implies $a_n \to 0$.
If $0 \le r < 1$ and $0 \le a_n \le Ar^{-n}$ then $\sum_n a_n$. In this case we say that $a_n \to 0$ *geometrically fast.*

Remark. The proof of this result is important as it leads in to the proof of the following Riesz-Fischer theorem.

Proof. **1.** Write[55]

$$f_n = f_1 + (f_2 - f_1) + (f_3 - f_2) + \cdots + (f_n - f_{n-1}). \tag{14}$$

Let

$$g_n = |f_1| + |f_2 - f_1| + |f_3 - f_2| + \cdots + |f_n - f_{n-1}|. \tag{15}$$

2. Then $g_n \uparrow g$, say, where possibly $g(x) = \infty$ for some x. By the monotone convergence theorem $\int g = \lim_{n\to\infty} \int g_n$.

By assumption, $K := \sum_{n\ge 1} \int |f - f_n| < \infty$. Using the triangle inequality

$$\begin{aligned} \int g_n &= \int |f_1| + \sum_{k=2}^{n} \int |f_k - f_{k-1}| \\ &\le \int |f_1| + \sum_{k=2}^{n} \int \Big(|f - f_k| + |f - f_{k-1}| \Big) \le \int |f_1| + 2K, \end{aligned}$$

which is independent of n.

So $\int g < \infty$ by the monotone convergence theorem and so $g < \infty$ a.e. In particular, $\lim g_n(x)$ exists and is finite for a.e. x.

3. Because the series (15) converges a.e., the series (14) converges absolutely a.e., and in particular converges a.e. to h, say.

4. It remains to show $f = h$ a.e.

Since $|f_n| \le g_n \le g$, and since g is integrable, and since $f_n \to h$ a.e., by the dominated convergence theorem $f_n \to h$ in the L^1 sense.

We already know $f_n \to f$ in the L^1 sense. By uniqueness of limits in a metric space, $f = h$ a.e.

This completes the proof of the theorem. □

Theorem (Riesz Fischer)**.** *The space L^1 is complete in its metric.*

Proof.
1. Suppose $\|f_n - f_m\| \to 0$ as $m, n \to \infty$. Choose a subsequence $(f_{n_k})_{k\ge 1}$ so $\|f_{n_{k+1}} - f_{n_k}\| \le 2^{-k}$, and in particular $\sum_{k\ge 1} \|f_{n_{k+1}} - f_{n_k}\| \le \infty$.

2. Consider

$$f(x) := f_{n_1}(x) + \sum_{k\ge 1} \big(f_{n_{k+1}}(x) - f_{n_k}(x) \big) \tag{16}$$

$$g(x) := |f_{n_1}(x)| + \sum_{k\ge 1} \big| f_{n_{k+1}}(x) - f_{n_k}(x) \big|, \tag{17}$$

for each x such that the relevant series converges.

3. As in the proof of the previous theorem, $\int g < \infty$, hence g is finite a.e., hence the series in (17) converges to a finite limit a.e., and hence so does the series in (16). That is $f_{n_k} \to f$ a.e.

Also as in the previous proof, $f_{n_k} \to f$ in the L^1 sense by the dominated convergence theorem. *Why?*

4. To see that also $f_n \to f$ in the L^1 sense, write

$$\|f_n - f\| \le \|f_n - f_{n_k}\| + \|f_{n_k} - f\|.$$

Given $\epsilon > 0$ use the fact (f_n) is L^1-Cauchy to choose $N = N(\epsilon)$ so the first term on the right side is $< \epsilon/2$ for all $n, n_k > N$. Then choose n_k so the second term is $< \epsilon/2$, this being permissible since $f_{n_k} \to f$ in the L^1 sense.

Then $n > N$ implies $\|f_n - f\| \le \epsilon$, which gives the result. □

[55] We are writing the sequence $f_n(x)$ as a series, and we will prove the sequence converges a.e. by showing the series converges a.e., in fact converges absolutely a.e.

2.3. Dense subsets of $L^1(\mathbb{R}^d)$.

Theorem. *The following are dense in* $L^1(\mathbb{R}^d)$*:*

1. *Simple functions*
2. *Step functions*
3. *Continuous functions with compact support*

(That is, if $f \in L^1(\mathbb{R}^d)$ then for each $\epsilon > 0$ there is a simple function ϕ such that $\|f - \phi\| < \epsilon$. Equivalently, there is a sequence of simple functions $(\phi_n)_{n\geq 1}$ such that $\phi_n \to f$ in the L^1 sense.

Similarly for step and continuous compactly supported functions.)

Proof sketch. **1.** By setting $f = f^+ - f^-$, we see it is sufficient to approximate positive functions f.

If f is positive then there is a sequence of simple functions $0 \leq \phi_n \uparrow f$.

By the monotone convergence theorem, $\int \phi_n \uparrow \int f$.

It follows $\phi_n \to f$ in the L^1 sense, why?

2. It is now sufficient to show any simple function $\sum_{i=1}^N a_i \mathcal{X}_{E_i}$ can be approximated by a step function.

For this is is sufficient to show any characteristic function $\mathcal{X}_E$ can be approximated by a step function.

But given $\epsilon > 0$ any measurable E can be approximated by a finite union of almost disjoint cubes F_j such that $m\left(E \,\Delta\, \bigcup_j F_j\right) < \epsilon$. It follows $\|\mathcal{X}_E - \sum \mathcal{X}_{F_j}\| < \epsilon$.

3. It is sufficient to show that any step function $\sum_{i=1}^N a_i \mathcal{X}_{R_i}$ where R_i are closed rectangles can be approximated by a continuous compactly supported function.

So it is sufficient to show $\mathcal{X}_R$ where R is a closed rectangle can be approximated by a continuous compactly supported function h.

This is done by taking $h = 1$ on R and dying off rapidly to 0 in a narrow strip around R. □

2.4. Other L^1 spaces.

If $E \subset \mathbb{R}^d$ is measurable and $m(E) > 0$ then define

$$L^1(E) = \left\{ f : E \to \mathbb{R} \mid \int_E f \text{ exists and is finite} \right\}.$$

Equivalently, for any $f : E \to \mathbb{R}$ let $\widetilde{f}$ be the zero extension of f to $\mathbb{R}^d$. Then define

$$L^1(E) = \{ f : E \to \mathbb{R} \mid \widetilde{f} \in L^1(\mathbb{R}^d)\}, \quad \|f\|_{L^1(E)} = \|\widetilde{f}\|_{L^1(\mathbb{R}^d)}.$$

It follows almost immediately that $L^1(E)$ is a complete metric space.

Thurs 10/5

2.5. Invariance Properties.

Translates, dilates and reflections of measurable functions are measurable. Moreover:

1. $\int f(x-h)\,dx = \int f(x)\,dx$. (Note that the graph of the function on the left is the translate of the graph of f by the vector h.)
   Proof straightforward, think of $\mathcal{X}_E$ case.
2. $r^d \int f(rx)\,dx = \int f(x)\,dx$. (Note that the graph of the function on the left is the dilation of the graph of f by the factor r^{-1}.)
   To see this think of the case $\mathcal{X}_E$.
3. $\int f(-x)\,dx = \int f(x)\,dx$.

2.6. **Consequences of invariance properties.**

(1) Suppose $h \in \mathbb{R}^d$ and $f(h-x)\,g(x)$ is integrable (for example, if both f and g are bounded and have bounded support). Then the integral on the right below also exists and

$$\int f(h-x)\,g(x)\,dx = \int f(x)\,g(h-x)\,dx.$$

Proof: The right integral is obtained from the left by replacing x by $h-x$. This is a translation by h followed by a reflection and so does not change the value of the integral.

Note: It is helpful, and often useful, to take f as an approximate delta function — in other words a non-negative function supported on a small interval containing the origin and having integral equal to one. Then either of the above integrals is an approximation to $g(h)$ obtained by using f to "average" g near h.

The right side is the *convolution* $f * g$ evaluated at h. The result says that $f * g = g * f$, i.e. convolution is commutative.

(2) For any a and $\epsilon > 0$

$$\int_{|x|\le\epsilon} \frac{dx}{|x|^a} = \epsilon^{-a+d} \int_{|x|\le 1} \frac{dx}{|x|^a}, \qquad \int_{|x|\ge\epsilon} \frac{dx}{|x|^a} = \epsilon^{-a+d} \int_{|x|\ge 1} \frac{dx}{|x|^a}.$$

Proof: For the first result

$$\begin{aligned}\text{L.H.S.} &= \int \frac{1}{|x|^a}\mathcal{X}_{B_\epsilon(0)}(x)\,dx = \int \frac{1}{|x|^a}\mathcal{X}_{B_1(0)}\left(\frac{x}{\epsilon}\right)\,dx \\ &= \epsilon^d \int \frac{1}{\epsilon^a|x|^a}\mathcal{X}_{B_1(0)}(x)\,dx = \text{R.H.S.}\end{aligned}$$

The first and second equalities are straightforward and the third is from the dilation result.

The second result is similar.

2.7. L^1 **continuity of translations.**

Theorem. *Suppose $f \in L^1(\mathbb{R}^d)$. Then $\|f_h - f\| \to 0$ as $h \to 0$, where $f_h(x) = f(x-h)$.*

Proof. Suppose $\epsilon > 0$ and choose a step function ψ such that $\|f - \psi\| < \epsilon$.

Then

$$\|f_h - f\| \le \|f_h - \psi_h\| + \|\psi_h - \psi\| + \|\psi - f\|.$$

The first and third terms on the right are equal by translation invariance, and for suitable choice of ψ the third is $< \epsilon$ by the density of step functions.

The second term is $< \epsilon$ for suitably small h since for any rectangle R, $\int \mathcal{X}_R - (\mathcal{X}_R)_h = \int \mathcal{X}_R - \mathcal{X}_{R_h}$ can clearly be made arbitrarily small by taking h sufficiently small.

(Alternatively, use compactly supported functions ψ. The fact they are uniformly continuous shows the middle term on the right can be made $< \epsilon$ for suitable ψ. The first and third terms are $< \epsilon$ as before by the density of continuous compactly supported functions. □

3. Interchanging Integration with Differentiation and Limits

Mon 14/5 This is very standard. See, for example, "Real Analysis" by Folland, p56.

In the following we have a function $f(x,t)$ of two variables $x \in E \subset \mathbb{R}^d$ and $t \in [a,b]$. You may want to think of the case $d = 1$. Note that all integrals are with respect to the x variable.

Let

$$F(t) = \int_E f(x,t)\,dx.$$

In particular we assume that $f(x,t)$ is integrable in x for each $t \in [a,b]$. We want to know if:

(1) Assuming $f(x,t)$ is continuous in t for each $x \in E$, does it follow that $F(t)$ is continuous?
(2) Assuming $f(x,t)$ is differentiable in t for each $x \in E$ does it follow that $F'(t)$ exists and $F'(t) = \int_E \frac{\partial}{\partial t} f(x,t)\,dx$?

The answer is "yes", provided that the functions $f(\cdot,t)$, respectively $\frac{\partial}{\partial t} f(\cdot,t)$, are uniformly integrable over E. That is, provided they are bounded by some positive integrable function independent of t.

The proof is straightforward using

(1) the dominated convergence theorem, and
(2) the observation that the relevant limit exists at t_0 if the limit both exists for any sequence $t_n \to t_0$ and is independent of the sequence.

Theorem. *Suppose $E \subset \mathbb{R}^d$, $f : E \times [a,b] \to \mathbb{R}$, and $f(x,t)$ is integrable in x for each $t \in [a,b]$. Let*

$$F(t) = \int_E f(x,t)\,dx.$$

Suppose $f(x,t)$ is continuous in t for each $x \in E$. Suppose for all $x \in E$ and $t \in [a,b]$ that $|f(x,t)| \le g(x)$ where $g : E \to \mathbb{R}$ is integrable. Then $F(t)$ is continuous in t, i.e.

$$\lim_{t \to t_0} \int_E f(x,t)\,dx = \int_E f(x,t_0)\,dx \quad \forall t_0 \in [a,b]. \tag{18}$$

Suppose $\frac{\partial}{\partial t} f(x,t)$ exists for all t and all x. Suppose for all $x \in E$ and $t \in [a,b]$ that $\left|\frac{\partial}{\partial t} f(x,t)\right| \le k(x)$ where $k : E \to \mathbb{R}$ is integrable. Then F is differentiable and

$$\frac{\partial}{\partial t}\bigg|_{t=t_0} \int_E f(x,t)\,dx = \int_E \frac{\partial}{\partial t}\bigg|_{t=t_0} f(x,t)\,dx \quad \forall t_0 \in [a,b]. \tag{19}$$

Proof. Let $(t_n)_{n\ge1}$ be any sequence in $[a,b]$ such that $t_n \to t_0$. Let $f_n(x) = f(x,t_n)$ and $f(x) = f(x,t_0)$. Since $f_n(x) \to f(x)$ and $|f_n(x)| \le g(x)$ for all x, where $\int_E g < \infty$, it follows from the dominated convergence theorem that

$$\lim_{t_n \to t_0} \int_E f(x,t_n)\,dx = \int_E f(x,t_0)\,dx \quad \forall t_0 \in [a,b].$$

Since the right side is independent of the sequence (t_n), (18) follows.

Again let $(t_n)_{n\ge1}$ be any sequence in $[a,b]$ such that $t_n \to t_0$. Then

$$h_n(x) := \frac{f(x,t_n) - f(x,t_0)}{t_n - t_0} \to \frac{\partial f}{\partial t}(x,t_0) \quad \text{as } t_n \to t_0.$$

By the mean value theorem, for some ξ_n between t_n and t_0,

$$|h_n(x)| = \left|\frac{\partial f}{\partial t}(x,\xi_n)\right| \le k(x).$$

It follows by the dominated convergence theorem that

$$\int h_n(x)\,dx \to \int_E \frac{\partial f}{\partial t}(x,t_0)\,dx \quad \text{as } t_n \to t_0,$$

$$\text{i.e. } \frac{\int_E f(x,t_n) - \int_E f(x,t_0)}{t_n - t_0} \to \int_E \frac{\partial f}{\partial t}(x,t_0)\,dx \quad \text{as } t_n \to t_0.$$

Since this is true for any sequence $(t_n)_{n\ge1}$ from $[a,b]$, (19) follows. □

4. Fubini's Theorem

Tues 15/5

4.1. **Set and function slices.** *See p75 of text for diagram.*

(1) We work in $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.
To fix ideas, think of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.
(2) If $E \subset \mathbb{R}^d$ we write E^y for the "horizontal" y-slice of E where $y \in \mathbb{R}^{d_2}$. Similarly we write E_x for the "vertical" x-slice where $x \in \mathbb{R}^{d_1}$.
(3) We also write $f^y(x) = f(x, y)$ for the function on $\mathbb{R}^{d_1}$ but thought of as a function on the "horizontal slice" through y.
Similarly for $f_x(y) = f(x, y)$.

Theorem. *Suppose $f(x, y)$ is integrable on $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then*

(1) *For a.e. y the function $x \mapsto f(x, y)$ is integrable.*
(2) *The function $y \mapsto \int_{R^{d_1}} f(x, y)\, dx$ is integrable.*
(3) *Moreover, $\int_{R^{d_2}} \left(\int_{R^{d_1}} f(x, y)\, dx\right) dy = \int_{\mathbb{R}^d} f$.*

(By switching the roles of x and y it follows that

(1) For a.e. x the function $y \mapsto f(x, y)$ is integrable.
(2) The function $x \mapsto \int_{R^{d_2}} f(x, y)\, dy$ is integrable.
(3) $\int_{R^{d_1}} \left(\int_{R^{d_2}} f(x, y)\, dy\right) dx = \int_{\mathbb{R}^d} f$.

In particular, changing the order of integration in the double integral does not affect the result.)

Proof Sketch. There are no major new techniques. But the proof in the text is about as clean as it gets.

1. First show that the class of functions satisfying 1–3 is closed under finite linear combinations (easy).

2. Then show the same class is closed under a.e. increasing and a.e. decreasing limits. The main point is to apply the monotone convergence theorem (twice)

3. Then show that $\mathcal{X}_E$ satisfies 1-3 if E is a $\mathcal{G}_\delta$ set, i.e. E is a countable intersection of open sets.

This is done first for open cubes, then for the boundary of open cubes (the double integral is zero), then for finite unions of almost disjoint closed cubes, then for open sets (by approximating by finite unions of almost disjoint closed cubes and using also Step 2), then for $\mathcal{G}_\delta$ sets (by using again Step 2).

4. Then show that $\mathcal{X}_N$ satisfies 1-3 if $m(N) = 0$ (cover N by a $\mathcal{G}_\delta$ set also of measure 0, and use Step 3).

5. Then show that $\mathcal{X}_E$ satisfies 1-3 if E is measurable. (Let $G = E \cup N$ where G is a $\mathcal{G}_\delta$ set and N is a null set and the union is disjoint. Then note $\mathcal{X}_E = \mathcal{X}_G - \mathcal{X}_N$ and use Step 1.)

6. $f = f^+ - f^-$ so by Step 1 it is sufficient to assume $f \geq 0$.

Then there exist simple functions $\phi_n \uparrow f$. But simple functions satisfy 1–3 by Steps 5 and 1, and so f satisfies 1–3 by Step 2.

□

4.2. **Applications of Fubini's Theorem.**

4.2.1. *Tonelli's Theorem.*

(1) This differs from Fubinis theorem in that it applies to any *non negative* function f, but without the integrability restriction that $\int_{\mathbb{R}^d} f < \infty$.
(2) In applications one often wants to apply Fubini to $f : \mathbb{R}^d \to \mathbb{R}$ but does not know f is integrable. In this case one can often first apply Tonelli's theorem to $|f|$ to show $\int_{\mathbb{R}^d} |f| < \infty$.
This implies $\int f^+$ and $\int f^-$ are finite (*why?*) and so $\int f$ exists and is finite, i.e. f is integrable. Then one is justified in applying Fubini's theorem to f.

Theorem. *Suppose $f \geq 0$ and measurable on $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then*

(1) *For a.e. y the function $x \mapsto f(x, y)$ is measurable.*

(2) *The function* $y \mapsto \int_{R^{d_1}} f(x,y)\,dx$ *is measurable (may take value* $+\infty$*).*
(3) *Moreover,* $\int_{R^{d_2}} \left(\int_{R^{d_1}} f(x,y)\,dx\right) dy = \int_{\mathbb{R}^d} f$ *(may take value* $+\infty$*).*

(By switching the roles of x and y it follows that

(1) For a.e. x the function $y \mapsto f(x,y)$ is measurable.
(2) The function $x \mapsto \int_{R^{d_2}} f(x,y)\,dy$ is measurable (may take value $+\infty$)..
(3) $\int_{R^{d_1}} \left(\int_{R^{d_2}} f(x,y)\,dy\right) dx = \int_{\mathbb{R}^d} f$ (may take value $+\infty$).

In particular, changing the order of integration in the double integral does not affect the result.)

Proof. **1.** Define the truncations of f for $k = 1, 2, \ldots$:

$$f_k(x,y) = \begin{cases} f(x,y) & \text{if } |(x,y)| \le k \text{ and } f(x,y) \le k, \\ 0 & \text{otherwise.} \end{cases}$$

Then $0 \le f_x(x,y) \uparrow f(x,y)$ for all (x,y).

2. For a.e. y the slice f_k^y is measurable for *every* k (uses Fubini 1).
Moreover for such y, $f_k^y \uparrow f^y$, and so by monotone convergence theorem

$$\int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx \uparrow \int_{\mathbb{R}^{d_1}} f(x,y)\,dx.$$

3. By Fubini, each integral on the left is measurable as a function of y and hence so is the integral on the right as a function of y .
By the monotone convergence theorem,

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx\right) dy \uparrow \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)\,dx\right) dy.$$

4. By Fubini, the integral on the left equals $\int_{\mathbb{R}^d} f_k$. Hence by the monotone convergence theorem the left side converges to $\int_{\mathbb{R}^d} f$. By the uniqueness of limits the final part of Tonelli follows. The other parts have already been proved. □

4.2.2. *Sets and Slices.*

(1) If follows from Tonelli that *if* $E \subset \mathbb{R}^d$ *is measurable then a.e. y-slice and a.e. x-slice is measurable. Moreover:*

$$m(E) = \int_{\mathbb{R}^{d_2}} m(E^y)dy = \int_{\mathbb{R}^{d_1}} m(E_x)dx. \tag{20}$$

Thurs 17/5

(2) In particular, *if* $E = E_1 \times E_2$ *is measurable then so are* E_1 *and* E_2 *and*

$$m(E) = m(E_1) \times m(E_2).$$

(3) It is not true that if every y-slice of E is measurable then so is E.
In fact there are (weird) examples where all y-slices and all x-slices are measurable, but E is not measurable. See text pp 81, 82.
(4) A result we want is:
if $E_1 \subset \mathbb{R}^{d_1}$ *and* $E_2 \subset \mathbb{R}^{d_2}$ *are measurable then so is* $E_1 \times E_2$.
(See Proposition 3.6. The main points is that we can approximate E_1 and E_2 by $\mathcal{G}_\delta$ sets to within measure 0. But the product of two $\mathcal{G}_\delta$ sets is Borel (*why*) and hence measurable. One also has to deal with the product of a set of measure zero and another set. See text, pp 83, 84. There are no essentially new ideas involved, but one does need to be careful.)

4.2.3. *Integrals and Areas.* The following ties integration, and area under the graph, together. *Draw a diagram.*

Theorem. *Suppose* $f \geq 0$ *and* $f : \mathbb{R}^d \to \mathbb{R}$. *Let*

$$\mathcal{A} = \{(x, y) : 0 \leq y \leq f(x)\} \subset \mathbb{R}^d \times \mathbb{R}.$$

Then

(1) f *is measurable on* $\mathbb{R}^d$ *iff* $\mathcal{A}$ *is a measurable subset of* $\mathbb{R}^{d+1}$.
(2) *If this condition holds then* $\int_{\mathbb{R}^d} f = m(\mathcal{A})$.

Proof. Again, we will not prove the measurability. See text p85.

Assuming both f and $\mathcal{A}$ are measurable, then by (**20**)

$$m(\mathcal{A}) = \int_{\mathbb{R}^d} m(\mathcal{A}_x)\, dx = \int_{\mathbb{R}^d} f(x) dx.$$

Why the second equality? □

Part 5. HILBERT SPACES

0. Motivation

I begin with some motivation. This includes a summary of the basic information for finite dimensional inner product spaces in both the real and complex settings.

0.1. Standard Inner Product on $\mathbb{R}^n$ and $\mathbb{C}^n$. The *standard inner product* in $\mathbb{R}^n$ and $\mathbb{C}^n$ respectively, is defined for $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ with $u_i, v_i \in \mathbb{R}$ and $\mathbb{C}$ respectively, by

$$\langle u, v\rangle = u_1 v_1 + \cdots + u_n v_n \text{ in } \mathbb{R}^n,$$
$$\langle u, v\rangle = u_1 \overline{v_1} + \cdots + u_n \overline{v_n} \text{ in } \mathbb{C}^n,$$

where the bar denotes complex conjugate. The corresponding *(Euclidean) norms* are defined by

$$\|u\| = \langle u, u\rangle^{1/2}, \text{ i.e. } \left(|u_1|^2 + \cdots + |u_n|^2\right)^{1/2},$$

in both the real and complex case. We think of the norm of u as the length of u, i.e. the distance from the origin.

(Notice that there is a natural correspondence between $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ given by

$$(u_1 + iv_1, \ldots, u_n + iv_n) \leftrightarrow (u_1, v_1, \ldots, u_n, v_n),$$

and that the norms are preserved.)

There are a number of inequalities and equalities which follow for the standard inner product. These are easiest to prove in the setting of a general inner product space, which we now introduce.

0.2. Inner Product Spaces. Motivated by $\mathbb{R}^n$ and $\mathbb{C}^n$, an *inner product space* is defined to be a vector space V over either $\mathbb{R}$ or $\mathbb{C}$ (not necessarily finite dimensional), together with an *inner product* $\langle \cdot, \cdot\rangle : V \times V \to \mathbb{R}$ $(\mathbb{C})$ which is required to be:

(1) *Conjugate Symmetric*: for all $u, v \in V$,

$$\langle u, v\rangle = \overline{\langle v, u\rangle},$$

In the real case, this is just usual the usual symmetry condition $\langle u, v\rangle = \langle u, v\rangle$.

(2) *Linear* in the first argument: for all scalars α and all $u, u_1, u_2, v \in V$,

$$\langle \alpha u, v\rangle = \alpha\langle u, v\rangle, \quad \langle u_1 + u_2, v\rangle = \langle u_1, v\rangle + \langle u_2, v\rangle.$$

By conjugate symmetry, it follows that for all scalars α and all $u, v, v_1, v_2 \in V$,

$$\langle u, \alpha v\rangle = \overline{\alpha}\langle u, v\rangle, \quad \langle u, v_1 + v_2\rangle = \langle u, v_1\rangle + \langle u, v_2\rangle.$$

This is called *antilinearity* in the second argument. (In the physics literature, the linearity is in the second argument and the antilinearity is in the first.)

(3) *Positive Definite*:

$$\langle u, u\rangle \geq 0, \quad \langle u, u\rangle = 0 \text{ iff } u = 0 \text{ (the zero vector)}.$$

In an inner product space the *norm* is defined by

$$\|u\| = \langle u, u\rangle^{1/2}.$$

We see it is a norm after proving the Cauchy Schwartz inequality.

The *metric* corresponding to this norm is (as for any norm) given by

$$d(u, v) = \|u - v\|.$$

0.3. **Examples.**

(1) $\mathbb{R}^n$ and $\mathbb{C}^n$.

(2) $\mathcal{C}[a,b]$ is the set of continuous real, or complex, valued functions on $[a,b]$. It is an inner product space under the usual definition of the sum of two functions, multiplication by a constant, and L^2 *inner product and norm* given by

$$\langle f,g\rangle = \int_a^b fg, \quad \langle f,g\rangle = \int_a^b f\overline{g}, \quad \|f\| = \left(\int_a^b |f|^2\right)^{1/2}.$$

in the real and complex case respectively. *Exercise.* Note that we only need Riemann integration here. This space is infinite dimensional.

(3) $L^2(E)$, where E is any measurable subset of $\mathbb{R}^d$, is by definition the set of square integrable functions, i.e. measurable functions $f : E \to \mathbb{R}$ (or $\mathbb{C}$) such that $\int_E |f|^2 < \infty$. It is an inner product space with the same L^2 *inner product and norm* given by

$$\langle f,g\rangle = \int_E fg, \quad \langle f,g\rangle = \int_E f\overline{g}, \quad \|f\| = \left(\int_E |f|^2\right)^{1/2}.$$

More precisely, just as for $L^1(E)$, we should identify functions that are equal a.e. Also, we need to use Lebesgue integration.

The main point is to show that $f, g \in L^2(E)$ implies that $f+g \in L^2(E)$ and that $f\overline{g}$ is integrable. The rest of the proof that $L^2(E)$ is an inner product space is straightforward. But

$$|f+g|^2 \le (|f|+|g|)^2 \le 2|f|^2 + 2|g|^2, \quad |f\overline{g}| = |f||g| \le \frac{1}{2}|f|^2 + \frac{1}{2}|g|^2.$$

This shows $f+g \in L^2(E)$ and $f\overline{g}$ is integrable.

Note: An important fact about the previous two examples is that $\mathcal{C}[a,b]$ is *not* complete in the L^2 norm. For example, consider the functions

$$f_n(x) = \begin{cases} 0 & -1 \le x \le 0 \\ nx & x \le 0 \le \frac{1}{n} \\ 1 & \frac{1}{n} \le x \le 1 \end{cases} \quad \text{if } n \ge 1, \quad f(x) = \begin{cases} 0 & -1 \le x \le 0 \\ 1 & 0 < x \le 1 \end{cases}.$$

Then $f_n \in \mathcal{C}[-1,1]$ and $f \in L^2[-1,1] \setminus \mathcal{C}[-1,1]$. It is easy to see (*why?*) that $f_n \to f$ in the L^2 sense and in particular (f_n) is Cauchy. But (f_n) has no limit $g \in \mathcal{C}[-1,1]$, since if it did then by uniqueness of limits in $L^2[-1,1]$ it follows $g = f$ a.e. But no continuous function g has this property.[56]

On the other hand, we will soon see that $L^2(E)$ is always complete. *This is an important reason for developing the theory of Lebesgue integration.*

0.4. **Properties.** The *Cauchy Schwartz inequality* states:

$$|\langle u,v\rangle| \le \|u\|\,\|v\|,$$

with equality iff u and v are linearly dependent. A slick proof is to assume $v \ne 0$ (otherwise the result is trivial) and let $\lambda = \langle u,v\rangle / \langle v,v\rangle$. Then a calculation gives (*exercise*)

$$0 \le \langle u-\lambda v, u-\lambda v\rangle = \langle u,u\rangle - \frac{|\langle u,v\rangle|^2}{\langle v,v\rangle}.$$

The result follows, *exercise.*

It follows that $\|u\|$ defines a norm, that is

(1) *Homogeneity*

$$\|\alpha u\| = |\alpha| \cdot \|u\|.$$

(2) *Positivity*

$$\|u\| \ge 0, \quad \|u\| = 0 \text{ iff } u = 0 \text{ (the zero vector)}.$$

[56]Suppose g is continuous and $g = f$ a.e. Consider any open interval $(-1/n, 1/n) \ni 0$. Since $f(x) = 1$ for every $x \in (0,1/n)$ it follows $g(x) = 1$ for a.e. $x \in (0,1/n)$. Similarly, $g(x) = 0$ for a.e. $x \in (-1/n, 0)$. It follows that g cannot be continuous at 0 since there exists $0 < x_n \to 0$ with $g(x_n) = 1$ and $0 > x_n \to 0$ with $g(x_n) = 0$.

(3) *Triangle Inequality*

$$\|u+v\| \le \|u\| + \|v\|.$$

The proof of homogeneity and positivity is straightforward, *exercise*. For the triangle inequality we use Cauchy Schwartz to get

$$\begin{aligned}\|u+v\|^2 = \langle u+v, u+v\rangle &= \|u\|^2 + \langle u,v\rangle + \langle v,u\rangle + \|v\|^2 = \|u\|^2 + 2\mathfrak{Re}\langle u,v\rangle + \|v\|^2 \\ &\le \|u\|^2 + 2|\langle u,v\rangle| + \|v\|^2 \le \|u\|^2 + 2\|u\|\,\|v\| + \|v\|^2 = \big(\|u\|+\|v\|\big)^2,\end{aligned}$$

and take square roots.

Motivated by what happens in $\mathbb{R}^2$ and $\mathbb{R}^3$, we define u and v to be *orthogonal* and write $u \perp v$ if $\langle u,v\rangle = 0$. For a real vector space we also define the *angle* $\theta \in [0,\pi]$ between u and v to be given by

$$\cos\theta = \frac{\langle u,v\rangle}{\|u\|\,\|v\|}.$$

For a complex vector space we can define the *angle* $\theta \in [0,\pi/2]$ between u and v to be given by

$$\cos\theta = \frac{|\langle u,v\rangle|}{\|u\|\,\|v\|}.$$

We also have

$$\langle u,v\rangle = 0 \Longrightarrow \|u+v\|^2 = \|u\|^2 + \|v\|^2 \quad \text{(Pythagoras's Theorem)}$$

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2 \quad \text{(Parallelogram law)}$$

Both are easy *exercises* from the definition of the norm and the properties of an inner product. Note what the second says about the sides and diagonals of a parallelogram in $\mathbb{R}^2$.

An *orthonormal basis* for a *finite* dimensional inner product space V is a basis $\{e_1,\dots,e_n\}$ such that $\langle e_i, e_j\rangle = 0$ if $i \ne j$ and $= 1$ if $i = j$. That is, the vectors have unit length and are mutually orthogonal.

The standard example in either $\mathbb{R}^n$ or $\mathbb{C}^n$ is $e_i = (0,\dots,0,1,0,\dots,0)$ where the 1 is in the ith place.

It is easy to check that any orthonormal *set* in a finite dimensional space V is linearly independent (*why?*), and so is a basis iff the cardinality is the same as the dimension of V.

If $\{e_1,\dots,e_k\}$ is an orthonormal set and $a_i := \langle u, e_i\rangle$, then

$$\|u\|^2 \ge \sum_{i=1}^{k} |a_i|^2. \quad \text{(\textit{Bessel's inequality} in the finite case).}$$

Equality holds iff $\{e_1,\dots,e_k\}$ is an orthonormal *basis.* (*Parsevals's identity* in the finite case.) They follow from Pythagoras's theorem, *why?*

We say that a_i is the *component* of u in the direction e_i and the vector $a_i e_i$ is the *projection* of u on the space spanned by e_i.

The *Gram Schmidt* process converts a basis of a finite dimensional inner product space into an orthonormal basis. (We will later see the extension in the infinite dimensional case for Hilbert spaces).

Namely, suppose $w_1,\dots,w_n$ is a basis for $\mathbb{R}^n$ or $\mathbb{C}^n$. Then we construct $e_1,\dots,e_n$ as follows:

(1) $e_1 = w_1/\|w_1\|$. (Unit length vector in same direction as w_1.)
(2) $e_2 = (w_2 - \langle w_2, e_1\rangle e_1)/\|w_2 - \langle w_2, e_1\rangle e_1\|$. (Subtract from w_2 its projection on the space spanned by e_1, and then normalise to unit length. This gives a unit length vector orthogonal to e_1 and in the space spanned by w_1, w_2.)
(3) $e_3 = (w_3 - \langle w_3, e_1\rangle e_1 - \langle w_3, e_2\rangle e_2)/\|w_3 - \langle w_3, e_1\rangle e_1 - \langle w_3, e_2\rangle e_2\|$. (Subtract from w_3 its projection on the space spanned by e_1, e_2, and then normalise to unit length. This gives a unit length vector orthogonal to e_1, e_2 and in the space spanned by w_1, w_2, w_3.)
(4) Etc.

Draw a diagram.

Any two finite dimensional spaces with the same dimension are *isomorphic as vector spaces* in that there is a one-to-one correspondence which preserves vector addition and scalar multiplication. If the bases are $\{e_1, \dots, e_n\}$ and $\{f_1, \dots, f_n\}$ then such a correspondence is given by $\sum_{i=1}^n a_i e_i \leftrightarrow \sum_{i=1}^n a_i f_i$.

Any two finite dimensional *inner product* spaces with the same dimension are *isomorphic as inner product spaces* in that the inner product, and hence the norm and metric, are also preserved under suitable one-to-one correspondences. In this case one takes orthonormal bases $\{e_1, \dots, e_n\}$ and $\{f_1, \dots, f_n\}$ and the same correspondence as before. *Check* that this preserves the inner products.

1. The Hilbert space $L^2(E)$

Mon 21/5

1.1. The Notion of a Hilbert Space.

We will usually be dealing with Hilbert spaces which are spaces of functions, and usually $L^2(E)$ for some $E \subset \mathbb{R}^d$. For this reason, elements of a general Hilbert space are usually denoted by letters f, g, h etc.

Definition 1.1. A *Hilbert space* $\mathcal{H}$ is an inner product space which is complete and separable.

"Complete" means that any Cauchy sequence (f_n) in the induced norm (equivalently, the induced metric) converges in the norm to some $f \in \mathcal{H}$.

"Separable" means there exists a countable dense subset of $\mathcal{H}$ with respect o the induced norm (that is, with respect to the induced metric). In many texts, the separability requirement is omitted.

1.2. Completeness of $L^2(E)$.

We first consider $L^2(\mathbb{R}^d)$. The proof of the following theorem is very similar to that for $L^1(\mathbb{R}^d)$.

Theorem 1.2. *The inner product space $L^2(\mathbb{R}^d)$ is complete.*

The following is very similar to that for the space $L^1(\mathbb{R}^d)$.

Proof. Suppose $(f_n)_{n\ge 1}$ is Cauchy in $L^2(\mathbb{R}^d)$. By passing to a subsequence, for which we abuse notation and also write as $(f_n)_{n\ge1}$, we can assume

$$\|f_n - f_{n-1}\| \le 2^{-n} \text{ if } n \ge 2. \tag{21}$$

Why? (The norm $\|\cdot\|$ is the L^2 norm given by $\|h\| = \left(\int |h|^2\right)^{1/2}$.)

Until Step 5 we will work with this subsequence, not the original sequence.

Step 2: Write the (sub)sequence f_n as a sequence of partial sums for a telescoping series:

$$f_n = f_1 + \sum_{k=2}^{n} \big(f_k - f_{k-1}\big). \tag{22}$$

The sequence $\big(f_n(x)\big)$ converges iff the corresponding infinite *series*

$$f(x) := f_1(x) + \sum_{n=2}^{\infty} \big(f_n(x) - f_{n-1}(x)\big) \tag{23}$$

converges, and the limit is the same. (Because the limit of a series is, by definition, just the limit of the sequence of partial suns.)

We will show this series converges a.e. by showing that it converges absolutely a.e.

Step 3: We consider the series of absolute values corresponding to (23):

$$g(x) := \big|f_1(x)\big| + \sum_{n=2}^{\infty} \big|f_n(x) - f_{n-1}(x)\big|, \tag{24}$$

with partial sums

$$g_n(x) := \big|f_1(x)\big| + \sum_{k=2}^{n} \big|f_k(x) - f_{k-1}(x)\big|, \tag{25}$$

In order to show the series in (24) converges a.e. we need to show $g < \infty$ a.e. We will do this by showing $\int g^2 < \infty$, which implies $g^2 < \infty$ a.e. and so $g < \infty$ a.e.

First note

$$\begin{aligned}\|g_n\| &\le \|f_1\| + \sum_{k=2}^{n} \|f_k - f_{k-1}\| \quad \text{(by the triangle inequality)}\\ &\le \|f_1\| + \frac{1}{2} \quad \text{(from (21))}\end{aligned}$$

Hence $\int g_n^2 \le \left(\|f_1\|+\frac{1}{2}\right)^2$. Because $g_n^2 \uparrow g^2$ it follows from the monotone convergence theorem that $\int g^2 \le \left(\|f_1\|+\frac{1}{2}\right)^2 < \infty$. Hence $g < \infty$ a.e. Hence (**24**) converges a.e. and hence (**23**) converges (absolutely) a.e. That is $f_n \to f$ a.e.

Step 4: We now know that f in (**23**) is defined a.e., and it is the limit of (the subsequence) (f_n) a.e. We claim $f \in L^2(\mathbb{R}^d)$ and $f_n \to f$ in the L^2 sense.

First note that $|f| \le g$ a.e. from (**23**) and (**24**). Similarly, $|f_n| \le |g_n| \le g$. It follows that $|f_n - f|^2 \le 4g^2$. Since $|f_n - f|^2 \to 0$ a.e., it follows by the dominated convergence theorem that $\int |f_n - f|^2 \to 0$, and so $\|f_n - f\| \to 0$.

Step 5: We now have an *original* sequence (f_n) which is Cauchy in the L^2 metric, and a *subsequence* which converges to some $f \in L^2(\mathbb{R}^d)$ in the L^2 metric. By standard metric space theory (*exercise*) it follows the original sequence converges to f in the L^2 metric.[57]

Hence $L^2(\mathbb{R}^d)$ is complete. □

Remark: In the theorem we proved that a certain subsequence of the original sequence converged a.e. It is not necessarily true that the original sequence converges a.e. A counterexample is given by the "moving blip" example in the measure and integration lecture notes.

Corollary 1.3. *If $E \subset \mathbb{R}^d$ is measurable then $L^2(E)$ is complete.*

(The following proof is not usually written out as it is pretty "obvious" the result is true. But there is something to show.)

Proof. Consider

$$\widetilde{L}^2(E) = \{f \in L^2(\mathbb{R}^d) \mid f(x) = 0 \text{ if } x \notin E\} \subset L^2(\mathbb{R}^d),$$

with the induced inner product. There is a natural one-one correspondence between $L^2(E)$ and $\widetilde{L}^2(E)$ which preserves inner products, and hence norms, metrics and the corresponding notions of convergence. *Why?*

Since $L^2(\mathbb{R}^d)$ is complete, completeness of $L^2(E)$ is equivalent to $\widetilde{L}^2(E)$ being closed in $L^2(\mathbb{R}^d)$, *why?*

To see $\widetilde{L}^2(E)$ is closed in $L^2(\mathbb{R}^d)$ suppose $f_n \in \widetilde{L}^2(E)$ and $f_n \to f$ in the $L^2(\mathbb{R}^d)$ sense. We W.T.S. that $f = 0$ a.e. on E^c.

But

$$\int |\mathcal{X}_E f_n - \mathcal{X}_E f|^2 \le \int |f_n - f|^2 \to 0.$$

Hence $\mathcal{X}_E f_n \to \mathcal{X}_E f$ in the $L^2(\mathbb{R}^d)$ sense. Since $\mathcal{X}_E f_n = f_n$ and $f_n \to f$ in the $L^2(\mathbb{R}^d)$ sense, by uniqueness of limits in a metric space, $\mathcal{X}_E f = f$ a.e., and so $f = 0$ a.e. on E^c. □

1.3. **Separability of $L^2(E)$.** We first consider $L^2(\mathbb{R}^d)$.

Theorem 1.4. *The inner product space $L^2(\mathbb{R}^d)$ is separable.*

Proof. See [SS] p 160. The idea is to show that step functions with rational values and supported on rectangles with rational vertices, are dense in the L^2 norm. □

Corollary 1.5. *The inner product space $L^2(E)$ where $E \subset \mathbb{R}^d$, is separable.*

Proof 1. Any non empty subset of a separable metric space is separable.[58] Since the Hilbert space $L^2(E)$ is isomorphic to $\widetilde{L}^2(E) \subset L^2(\mathbb{R}^d)$, we are done. □

---

[57] In the text this is written out without referring to general metric spaces. See the last paragraph on p 160 of the proof of Theorem 1.2.

[58] (This is not that obvious.) Suppose (X, d) is a metric space and suppose $Y \subset X$ with the induced metric. Let $S = \{s_1, s_2, \dots\}$ be a countable dense subset of X.

For each $s_i \in S$ and each $n > 0$ choose $y_{in} \in B_{1/n}(s_i) \cap Y$ provided this set is non empty. Then the set S^* of y_{in} chosen in this manner is a countable subset of Y.

We claim S^* is dense in Y. For suppose $y \in Y$ and n is a positive integer. By density of S in X there exists $s_i \in S$ such that $d(y, s_i) < 1/n$. By the construction of S^* since $B_{1/n}(s_i) \cap Y \neq \emptyset$, $y_{in} \in B_{1/n}(s_i) \cap Y$ and in particular $d(y_{in}, s_i) < 1/n$. Then by the triangle inequality $d(y, y_{in}) \le d(y, s_i) + d(s_i, y_{in}) < 1/n + 1/n = 2/n$. It follows S^* is dense in Y.

Proof 2. Let $\{f_n : n \geq 1\}$ be any countable dense subset of $L^2(\mathbb{R}^d)$. Then $\{f_n \mathcal{X}_E : n \geq 1\}$ is a countable dense subset of $\widetilde{L}^2(E) \subset L^2(\mathbb{R}^d)$. *Why?* □

1.4. **Density Results.** We have already used a slightly stronger form of (2) in the following Theorem, see Theorem 1.4. But we note it again for completeness and comparison with the L^1 case.

Theorem 1.6. *The following families of functions are dense in $L^2(\mathbb{R}^d)$ (in the L^2 sense).*

(1) *The simple functions.*
(2) *The step functions.*
(3) *The continuous functions of compact support.*

Proof. Compare with Theorem 2.4 p 71, and Exercise 6 p194. The proof is similar to the L^1 case, but the result does not follow from this since L^1 convergence does not imply L^2 convergence.
1. This is as on page 72 for (i), except that now we apply the dominated convergence theorem to $|f - \phi_k|^2 \leq 4|f|^2$.
2. This is as on page 72 for (ii), except that now $\|\mathcal{X}_E - \psi\|_{L^2} \leq \sqrt{2\epsilon}$.
3. This is as on page 72 for (iii), except that now $\|f - g\|_{L^2} \leq \sqrt{2\epsilon}$. □

Remark 1.7. Similar results apply to $L^2(E)$ for measurable $E \subset \mathbb{R}^d$, by multiplying the approximating functions by $\mathcal{X}_E$. In the case of (2) we require E to be a closed rectangle and in the case of (3) we require E to be a closed set to make this method work. □

Tues 22/5

1.5. **Relationships between Different Types of Convergence.** Suppose $E \subset \mathbb{R}^d$. Consider functions f_n and f defined on E. We write $f_n \to f$ *uniformly* if $\|f_n - f\|_{\sup} \to 0$, where $\|g - f\|_{\sup} := \sup_{x\in E} |g(x) - f(x)|$ defines the *sup norm* (or *uniform norm*).

Theorem 1.8. *Suppose $m(E) < \infty$. Then for functions with domain E, the set of uniformly bounded functions (measurable) functions defined on E is a subset of $L^2(E)$, which in turn is a subset of $L^1(E)$.*

Moreover,

$$f_n \to f \text{ uniformly} \quad \Longrightarrow \quad f_n \to f \text{ in the } L^2\text{-sense} \quad \Longrightarrow \quad f_n \to f \text{ in the } L^1\text{-sense}\,.$$

The inverse implications are not true.

Proof. For the set inclusion relations note that

$$\int_E |f| \leq m(E)^{1/2} \left(\int_E |f|^2\right)^{1/2}, \qquad \left(\int_E |f|^2\right)^{1/2} \leq m(E)^{1/2}\|f\|_{\sup}.$$

The first inequality follows from the Cauchy Schwarz inequality, *why?* The second is straightforward, *why?*

Similarly,

$$\int_E |f_n - f| \leq m(E)^{1/2} \left(\int_E |f_n - f|^2\right)^{1/2},$$
$$\left(\int_E |f_n - f|^2\right)^{1/2} \leq m(E)^{1/2}\|f_n - f\|_{\sup}.$$

This gives the convergence implications.

Examples showing the inverse implications do not hold are the following.

Let $f_n = \sqrt{n}\mathcal{X}_{[0,1/n]}$. Then $f_n \to 0$ in the L^1 sense but not in the L^2 sense, *why?*[59]

Let $f_n = \sqrt[3]{n}\mathcal{X}_{[0,1/n]}$. Then $f_n \to 0$ in the L^2 sense but not in the uniform sense, *why?*[60] □

[59]Note also that if f_n converged to *any* g in the L^2 sense, it must also converge to g in the L^1 sense, by the first part of the theorem.

[60]Note also that if f_n converged to *any* g in the uniform sense, it must also converge to g in the L^2 sense, by the first part of the theorem.

1.6. L^∞ **norm.** Since we work with equivalence classes of functions modulo equality a.e., it is more natural to replace the sup norm by the *essential sup norm*, usually called the L^∞ *norm*. The definition is

$$\operatorname{ess\,sup} f = \|f\|_\infty = \inf\{M : |f| \le M \text{ a.e.}\}.$$

This agrees with the supremum for continuous functions, and if two functions agree a.e. then they have the same L^∞ norm.

One then defines

$$L^\infty(E) = \{f : E \to \mathbb{R}(\mathbb{C}) \mid \|f\|_\infty < \infty\}.$$

The previous theorem is then true with the sup norm replaced by the L^∞ norm, and uniform convergence replaced by convergence in the L^∞ norm.

It is fairly straightforward to check that $L^\infty(E)$ with the L^∞ norm is, like $L^1(E)$, a complete normed vector space, i.e. a Banach space. *Exercise.*

2. Hilbert spaces

In this section we discuss general Hilbert spaces.

2.1. Examples.

Example 2.1.

(1) Both $\mathbb{R}^n$ and $\mathbb{C}^n$ are Hilbert spaces. We already know they are inner product spaces.
They are complete essentially because $\mathbb{R}$ and $\mathbb{C}$ are complete. *Explain*,
They are separable because n-tuples with rational entries form a countable dense subset.
(2) We have seen that $\mathcal{C}[a, b]$ is an inner product space with the L^2 norm, but is not complete.
(3) We have seen that $L^2(E)$, for measurable $E \subset \mathbb{R}^d$, is a Hilbert space.
(4) The natural generalisation of the inner product spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ to the infinite dimensional case is the following. Let

$$\ell_2^{\mathbb{R}} = \left\{ u = (u_1, u_2, \dots) \;\middle|\; u_i \in \mathbb{R},\ \sum_{i \geq 1} |u_i|^2 < \infty \right\},$$
$$\ell_2^{\mathbb{C}} = \left\{ u = (u_1, u_2, \dots) \;\middle|\; u_i \in \mathbb{C},\ \sum_{i \geq 1} |u_i|^2 < \infty \right\},$$
$$\langle u, v \rangle = \sum_{i \geq 1} u_i v_i, \quad \langle u, v \rangle = \sum_{i \geq 1} u_i \overline{v_i}.$$

Note that

$$|u_i \overline{v_i}| = |u_i|\,|v_i| \leq \frac{1}{2}\left(|u_i|^2 + |v_i|^2\right),$$

and so the series for $\langle u, v \rangle$ converge (absolutely). By the next proposition, $\ell_2^{\mathbb{R}}$ and $\ell_2^{\mathbb{C}}$ are Hilbert spaces.

Notation: If we wish to refer to either $\ell_2^{\mathbb{R}}$ or $\ell_2^{\mathbb{C}}$ we sometimes write ℓ_2.

It is sometimes notationally convenient to consider sequences $u = (\dots, u_{-2}, u_{-1}, u_0, u_1, u_2, \dots)$ with $\sum_{i=-\infty}^{+\infty} |u_i|^2 < \infty$. There is no new theory, just identify $u = (\dots, u_{-2}, u_{-1}, u_0, u_1, u_2, \dots)$ with $(u_0, u_{-1}, u_1, u_{-2}, u_2, \dots)$.

If we want to make a distinction we write $\ell_2(\mathbb{N})$ and $\ell_2(\mathbb{Z})$, or $\ell_2^{\mathbb{R}}(\mathbb{N})$ and $\ell_2^{\mathbb{R}}(\mathbb{Z})$, or $\ell_2^{\mathbb{C}}(\mathbb{N})$ and $\ell_2^{\mathbb{C}}(\mathbb{Z})$.

Proposition 2.2. *$\ell_2^{\mathbb{R}}$ and $\ell_2^{\mathbb{C}}$ are Hilbert spaces.*

Proof. It is straightforward to check that $\ell_2^{\mathbb{R}}$ and $\ell_2^{\mathbb{C}}$ are inner product spaces.

A countable dense subset is obtained in each case by taking those sequences all of whose entries are rational *and* which are ultimately 0. (*Exercise*)

To show completeness in the $\ell_2^{\mathbb{R}}$ case consider a Cauchy sequence $(a^n)_{n \geq 1}$ where $a^n = (a_1^n, a_2^n, \dots)$.

For each fixed i, the sequence of real numbers $(a_i^n)_{n \geq 1}$ is Cauchy (*why?*) and so $\lim_{n \to \infty} a_i^n = a_i$ for some $a_i \in \mathbb{R}$.

We claim that if $a := (a_1, a_2, \dots)$ then $\lim_{n\to\infty} a^n = a$ in the $\ell_2^{\mathbb{R}}$ sense. This needs justification.[61] So suppose $\epsilon > 0$. Then for some $N(\epsilon)$,

$$\begin{aligned} \sum_{i=1}^{\infty} |a_i^m - a_i^n|^2 &\le \epsilon \quad \text{if } m,n \ge N(\epsilon), \\ \therefore\ \forall k\ \sum_{i=1}^{k} |a_i^m - a_i^n|^2 &\le \epsilon \quad \text{if } m,n \ge N(\epsilon), \\ \therefore\ \forall k\ \sum_{i=1}^{k} |a_i^m - a_i|^2 &\le \epsilon \quad \text{if } m \ge N(\epsilon) \text{ since } \lim_{n\to\infty} a_i^n = a_i, \\ \therefore\ \sum_{i=1}^{\infty} |a_i^m - a_i|^2 &\le \epsilon \quad \text{if } m \ge N(\epsilon). \end{aligned}$$

The third displayed line is justified since we are taking the limit inside a *finite* sum.

Assuming $a \in \ell_2^{\mathbb{R}}$, it follows that $\lim_{n\to\infty} a^n = a$ in the $\ell_2^{\mathbb{R}}$ sense. One easy way to check that $a \in \ell_2^{\mathbb{R}}$ is to note that $a = (a - a^m) + a^m$. But $a - a^m \in \ell_2^{\mathbb{R}}$ since $\|a - a^m\| \le 1$ for $m \ge N(1)$, and $a^n \in \ell_2^{\mathbb{R}}$ by definition. Since $\ell_2^{\mathbb{R}}$ is a vector space and hence closed under addition, it follows that $a \in \ell_2^{\mathbb{R}}$.

The proof in the $\ell_2^{\mathbb{C}}$ case is almost exactly the same. □

Thurs 24/5

2.2. **Convergence Properties.** A useful fact is that convergence preserves sums, scalar products, norms and inner products.

Convergence in a Hilbert space always means, unless specified otherwise, convergence in the Hilbert space norm, i.e. convergence in the Hilbert space metric. Convergence of a series means convergence of the sequence of partial sums.

Proposition 2.3. *Suppose in a Hilbert space that $u_n \to u$ and $v_n \to v$, and that for a sequence of scalars $\alpha_n \to \alpha$. Then*

$$\alpha_n u_n \to \alpha u,\ \ u_n + v_n \to u + v,\ \ \|u_n\| \to \|u\|,\ \ \langle u_n, v_n\rangle \to \langle u, v\rangle.$$

Proof. One has

$$\begin{aligned} \|\alpha_n u_n - \alpha u\| &= \|\alpha_n(u_n - u) + (\alpha_n - \alpha)u\| \\ &\le |\alpha_n|\,\|u_n - u\| + |\alpha_n - \alpha|\,\|u\| \to 0, \end{aligned}$$

using the fact the sequence $|\alpha_n|$ is bounded since $\alpha_n \to \alpha$.

Similarly,

$$\begin{aligned} \|(u_n + v_n) - (u + v)\| &\le \|u_n - u\| + \|v_n - v\| \to 0, \\ \big|\|u_n\| - \|u\|\big| &\le \|u_n - u\| \to 0, \\ \big|\langle u_n, v_n\rangle - \langle u, v\rangle\big| &= \big|\langle u_n, v_n - v\rangle + \langle u_n - u, v\rangle\big| \\ &\le \big|\langle u_n, v_n - v\rangle\big| + \big|\langle u_n - u, v\rangle\big| \\ &\le \|u_n\|\,\|v_n - v\| + \|u_n - u\|\,\|v\| \to 0. \end{aligned}$$

The second inequality is by the triangle inequality. In the last line we use the fact that the $\|u_n\|$ are uniformly bounded since $\|u_n\| \le \|u\| + \|u_n - u\|$ by the triangle inequality. □

---

[61] Interchange of limits is not always justified. Consider the infinite matrix defined for $m, n \ge 1$ by

$$a_{mn} = \begin{cases} 1 & \text{if } n \ge m \\ 0 & \text{if } n < m \end{cases}.$$

Then

$$\forall m\ \lim_{n\to\infty} a_{mn} = 1, \quad \forall n\ \lim_{m\to\infty} a_{mn} = 0.$$

2.3. **Orthonormal Basis.** The notion of an orthonormal basis is fundamental in the theory of Hilbert spaces. We will see this when we have a look at Fourier series.

Definition 2.4. [62] Suppose $\mathcal{H}$ is a Hilbert space with a countable orthonormal subset $B = \{e_1, e_2, e_3, \dots\}$, either finite or infinite.[63] Then B is an *orthonormal basis* for $\mathcal{H}$ if every $f \in \mathcal{H}$ is of the form

$$f = \sum_{n\geq 1} a_n e_n, \quad \text{that is } \Big\| f - \sum_{n\geq 1} a_n e_n \Big\| \to 0 \text{ as } k \to \infty \text{ in the infinite case,} \tag{26}$$

for some $a_n \in \mathbb{R}$ or $\mathbb{C}$.

Remark 2.5. Taking inner products with e_i in (26) we expect that

$$a_i = \langle f, e_i \rangle. \tag{27}$$

To justify this we use Proposition 2.3 to see that

$$\sum_{n=1}^{k} a_n e_n \to f \quad \text{implies} \quad \left\langle \sum_{n=1}^{k} a_n e_n, e_i \right\rangle \to \langle f, e_i \rangle,$$

which gives the result, *why?* □

Example 2.6.

(1) For $\mathbb{R}^n$ and $\mathbb{C}^n$, the standard orthonormal basis is given by $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ where the 1 is in the ith place, with $i = 1, \dots, n$.
(2) For $L^2[a, b]$ a nice orthonormal basis is given by Fourier series, discussed later.
(3) For $\ell_2^{\mathbb{R}}$ and $\ell_2^{\mathbb{C}}$, the standard orthonormal basis is given by $e_i = (0, \dots, 0, 1, 0, \dots)$ where the 1 is in the ith place, with $i \geq 1$.

Remark 2.7. Note that B is *not* a basis in the usual algebraic sense, unlike the case when we used Zorn's lemma to prove every vector space has an (algebraic) basis. The difference is that here f is only required to be a *limit* of finite linear combinations of the e_i, whereas in the algebraic case every f *equals* a finite linear combination. The algebraic situation is not very useful since it turns out that the basis is uncountable, the axiom of choice is needed to show existence of a basis, and the basis is not "natural" in any useful sense.

On the other hand, orthonormal bases as defined here can be extremely useful and natural. We will see this when we discuss Fourier series. □

Mon 28/5

2.4. **Properties of an Orthonormal Basis.** The following essentially combines Theorems 2.3 and 2.4 from [SS, pp165–168].

Theorem 2.8. *Every Hilbert space $\mathcal{H}$ has an orthonormal subset $B = \{e_1, e_2, \dots\}$ such that finite linear combinations of the e_i are dense in $\mathcal{H}$.*

If $B = \{e_1, e_2, \dots\}$ is any orthonormal subset of $\mathcal{H}$ such that finite linear combinations of the e_i are dense in $\mathcal{H}$, then B is an orthonormal basis. More precisely, if $f \in \mathcal{H}$ and $a_n = \langle f, e_n \rangle$, then

$$f = \sum_{n\geq 1} a_n e_n \text{ (orthonormal expansion, norm convergence).} \tag{28}$$

Moreover,

$$\|f\|^2 = \sum_{n\geq 1} |a_n|^2 \text{ (Parseval's identity).} \tag{29}$$

If $\{e_1, e_2, \dots\}$ is an orthonormal set (but not necessarily a basis) then $\sum_{n\geq 1} |a_n|^2 \leq \|f\|^2$ (Bessel's inequality). If equality holds then $\{e_1, e_2, \dots\}$ is an orthonormal basis.

[62] This looks like a stronger requirement than that given in the last few lines of [SS, p164]. But it is equivalent because of Theorem 2.8.

[63] Any subset of B must be linearly independent. To see this suppose that $\alpha_1 e_1 + \dots + \alpha_k e_k = 0$. Taking inner products of both sides with e_n, it follows that with $\alpha_n = 0$ for all n.

Proof.
Step 1: Because $\mathcal{H}$ is separable there is a countable dense subset $\{g_1, g_2, g_3, \dots\}$.

By successively removing any g_i which is a linear combination of previous g_j's, we can assume that each finite subset $\{g_1, \dots, g_n\}$ is linearly independent. For this new sequence which we also denote by $\{g_1, g_2, g_3, \dots\}$, finite linear combinations of the g_i will be dense.

Applying the Gram Schmidt process from page 63 we obtain an orthonormal set $\{e_1, e_2, e_3, \dots\}$, such that for each n the sets $\{e_1, \dots, e_n\}$ and $\{g_1, \dots, g_n\}$ span the same n-dimensional subspace of $\mathcal{H}$. Since finite linear combinations of the g_i are dense it follows that finite linear combinations of the e_i are also dense.

From now until the end of Step 5 we assume that B is an orthonormal subset of $\mathcal{H}$ and that finite linear combinations of the e_i are dense in $\mathcal{H}$.

Our main goal is the Claim in Step 4.

Step 2: *Claim*: If $\langle h, e_i\rangle = 0$ for all e_i then $h = 0$.

To prove this, use the result in Step 1 to obtain a sequence $(h_n)_{n\geq 1}$ of linear combinations of the e_i such that $h_n \to h$. Since $\langle h, e_i\rangle = 0$ for all e_i it follows that $\langle h, h_n\rangle = 0$ for all h_n. From this we have the following nice argument:

$$\|h\|^2 = |\langle h, h\rangle| = |\langle h, h - h_n\rangle| \leq \|h\|\,\|h - h_n\| \to 0 \text{ as } n \to \infty.$$

Hence $\|h\| = 0$ and so $h = 0$.

Step 3: *Claim*: Let $a_n := \langle f, e_n\rangle$. Then

$$e_1, \dots, e_N \perp f - \sum_{n=1}^{N} a_n e_n, \tag{30}$$

$$\|f\|^2 = \sum_{n=1}^{N} |a_n|^2 + \left\| f - \sum_{n=1}^{N} a_n e_n \right\|^2. \tag{31}$$

For (30) note $\langle f, e_n\rangle = a_n$. For (31) note $f = a_1e_1 + \cdots + a_ne_n + \left(f - \sum_{n=1}^{N} a_n e_n\right)$, and use Pythagoras's theorem.

Step 4: *Claim*: $f = \sum_{n=1}^{\infty} a_n e_n$.

(We prove this by using (31) to show that the sequence of partial sums $\sum_{n=1}^{N} a_n e_n$ is Cauchy and so converges by completeness of $\mathcal{H}$ to some $g \in \mathcal{H}$. Then we use (31) again to show that $g = f$.)

From (31) it follows $\sum_{n=1}^{\infty} |a_n|^2 \leq \|f\|^2 < \infty$ and so the series $\sum_{n=1}^{\infty} |a_n|^2$ converges. Next compute for $M > N$ that

$$\left\| \sum_{n=1}^{M} a_n e_n - \sum_{n=1}^{N} a_n e_n \right\|^2 = \sum_{n=N+1}^{M} |a_n|^2 \to 0 \text{ as } M, N \to \infty,$$

since $\sum_{n=1}^{\infty} |a_n|^2 < \infty$. Hence the limit $\sum_{n=1}^{\infty} a_n e_n$ exists (in the L^2 sense) and equals g, say, by completeness of $\mathcal{H}$.

Fix i. Since $\left\langle f - \sum_{n=1}^{N} a_n e_n, e_i \right\rangle = 0$ for $N \geq i$, it follows on letting $N \to \infty$ that $\langle f - g, e_i\rangle = 0$ for all i. Hence $f = g$ from Step 2.

Step 5: *Claim*: $\|f\|^2 = \sum_{n\geq 1} |a_n|^2$.

This now follows from (31), since $\left\| f - \sum_{n=1}^{N} a_n e_n \right\|^2 \to 0$ as $N \to \infty$ by Step 4.

Step 6: If $\{e_1, e_2, \dots\}$ is an orthonormal set, but not necessarily a basis, then Step 3 is still valid. It follows from (31) that $\sum_{n\geq 1} |a_n|^2 \leq \|f\|^2$. If equality holds then, from (31), $\left\| f - \sum_{n=1}^{N} a_n e_n \right\|^2 \to 0$ as $N \to \infty$ and so $f = \sum_{n\geq 1} a_n e_n$. In particular, $\{e_1, e_2, \dots\}$ is an orthonormal basis. □

Remark 2.9. It follows from Theorem 2.8 that any Hilbert space is isomorphic to $\mathbb{R}^n$ (or $\mathbb{C}^n$) for some n, or is isomorphic to $\ell_2^{\mathbb{R}}$ (or $\ell_2^{\mathbb{C}}$). *Why?* □

3. Fourier series

Tues 29/5

For technical convenience we work in $L^2[-\pi,\pi]$. Any interval $[a,b]$ is OK but the constants are messier.

I take a different approach to proving the trigonometric functions form an orthonormal basis, which avoids the material on kernels in Chapter 3.2. and instead uses the Stone Wierstrass theorem from MATH2320.

3.1. Orthonormal Trigonometric Functions in $L^2[-\pi,\pi]$.

It is convenient to use the inner product

$$(f,g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\overline{g} = ⨍_{-\pi}^{\pi} f\overline{g}.$$

Trig functions on $[-\pi,\pi]$.

(1) In the real case we check by integration that the set of functions

$$S_{\mathbb{R}} := S := \{1, \sqrt{2}\cos nx, \sqrt{2}\sin nx : n \in \mathbb{N}\}$$

forms an orthonormal set in $L^2([-\pi,\pi])$. For example,

$$⨍_{-\pi}^{\pi} 2\cos^2 nx\, dx = ⨍_{-\pi}^{\pi} 2\sin^2 nx\, dx,$$

from a diagram sketch, and their sum is $⨍_{-\pi}^{\pi} 2 = 2$. So both integrals equal 1.

To prove $(\sin nx, \cos mx) = 0$ if $n \neq m$, note

(32) $$\sin nx \cos mx = \frac{1}{2}(sin(n+m)x - \sin(n-m)x),$$

so

$$⨍_{-\pi}^{\pi} \sin nx \cos mx\, dx = 0.$$

Similarly for other cases.

In the complex case we check that

$$S_{\mathbb{C}} := S := \{e^{inx} : n \in \mathbb{Z}\}$$

forms an orthonormal set by using $e^{inx} = \cos nx + i\sin nx$. *Exercise.*

(2) We have seen that in both the real and complex case, S is an orthonormal subset of $L^2[-\pi,\pi]$. We will show in the next theorem that S is an orthonormal *basis.*

(3) It is natural to consider the set S of functions as they arise physically in harmonic motion, travelling waves, harmonics in music; and mathematically in solving the harmonic (Laplace), heat and wave equations.

3.2. The Fourier Basis.

Theorem. *S is an orthonormal basis, both in the real and complex cases.*

Proof. We know S is an orthonormal set.

The remaining point is to show that $\operatorname{span}(S)$, the set of finite liner combinations of functions from S, is dense in $L^2[-\pi,\pi]$. Then by Theorem 2.8, S is an orthonormal basis.

We first do the real case.

1. Let

$$C^*[-\pi,\pi] = \big\{ f \in C[-\pi,\pi] : f(-\pi) = f(\pi) \big\}.$$

We *claim* $\operatorname{Span}(S_{\mathbb{R}})$ is dense in $C^*[-\pi,\pi]$ in the sup norm, and hence in the L^2 norm.

Let S^1 denote the unit circle. $\mathcal{C}(S^1)$, the space of continuous functions on S^1, is essentially the same as the space $C^*[-\pi,\pi]$, *why?* We will not distinguish between these two spaces

We use the Stone Wierstrass theorem (MATH2320) to show that *Span($S_{\mathbb{R}}$)* is dense in $\mathcal{C}(S^1)$ in the uniform norm (i.e. max norm, i.e. sup norm), and hence in the L^2 norm by the earlier remarks on different types of convergence.

For this we note that S^1 is compact. We need to check that $\mathrm{Span}(S_{\mathbb{R}})$ is an algebra (i.e. closed under addition, scalar multiplication, and products), contains the constant functions, and separates points (i.e. $x \neq y \Longrightarrow f(x) \neq f(y)$ for some $f \in \mathrm{Span}(S_{\mathbb{R}})$).

The fact that $\mathrm{Span}(S_{\mathbb{R}})$ is closed under multiplication follows from (**32**) and similar identities. The fact $S_{\mathbb{R}}$ separates points is fairly clear. (Note this is not true if we work on $\mathcal{C}[-\pi,\pi]$ rather than $\mathcal{C}(S^1)$, since all functions in $S_{\mathbb{R}}$ take the same values at $-\pi$ and π.)

This completes the proof of the *claim*.

2. We *claim* $C^*[-\pi,\pi]$ is dense in $\mathcal{C}[-\pi,\pi]$ in the L^2 norm.

To see this suppose $f \in C[-\pi,\pi]$. By changing f near π, but by an arbitrarily small amount in the L^2 norm, we get a function in $C^*[-\pi,\pi]$. This proves the *claim*.

3. We *claim* $\mathcal{C}([-\pi,\pi])$ is dense in $L^2([-\pi,\pi])$.

This follows from Remark 1.7 with $E = [-\pi,\pi]$.

4. From **1,2,3** it follows that $\mathrm{Span}(S_{\mathbb{R}})$ is dense in $L^2([-\pi,\pi])$.

This completes the proof in the real case. We now do the complex case.

5. Suppose $f = u + iv \in L^2[-\pi,\pi]$. Suppose $\epsilon > 0$.

From the real case, $\exists \phi, \psi \in \mathrm{Span}(S_{\mathbb{R}})$ such that $\|u-\phi\|_{L^2} < \epsilon$ and $\|u-\psi\|_{L^2} < \epsilon$. It follows $\|f-g\| < 2\epsilon$ where $g = \phi + i\psi$.

We claim $g \in \mathrm{Span}(S_{\mathbb{C}})$. This follows (*why?*) from the facts

$$\cos nx = \frac{1}{2}(e^{inx} + e^{-inx}), \quad \sin nx = \frac{1}{2i}(e^{inx} - e^{-inx}), \quad 1 = e^{i0x}.$$

This completes the proof in the complex case. □

3.3. Properties of Fourier Series.

We can now apply Theorem 2.3 of the text or Theorem 2.8 here, to immediately deduce the following important and far from obvious facts:

Theorem. *Suppose $f \in L^2_{\mathbb{C}}[-\pi,\pi]$ and let $c_n = \fint_{-\pi}^{\pi} f(x)e^{-inx}\,dx$ for $n \in \mathbb{Z}$. Then*

1. *$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}$ with convergence in the L^2 sense.*
2. *$\fint_{-\pi}^{\pi} |f|^2 = \sum_{n=-\infty}^{\infty} |c_n|^2$.*
3. *$f = 0$ a.e. $\Longleftrightarrow c_n = 0\ \forall n$.*
4. *$f \leftrightarrow (c_n)_{-\infty}^{\infty}$ is an isomorphism between $L^2[-\pi,\pi]$ and $\ell_2(\mathbb{Z})$.*

In the real case, let

$$a_0 = \fint_{-\pi}^{\pi} f(x)\,dx, \quad a_n = \sqrt{2}\fint_{-\pi}^{\pi} f(x)\cos nx\,dx, \ n > 0,$$
$$b_n = \sqrt{2}\fint_{-\pi}^{\pi} f(x)\sin nx\,dx, \ n > 0.$$

Then

1. *$f(x) = a_0 + \sqrt{2}\sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)$ with convergence in the L^2 sense.*
2. *$\fint_{-\pi}^{\pi} |f|^2 = a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)$.*
3. *$f = 0$ a.e. $\Longleftrightarrow a_n, b_n = 0\ \forall n$.*
4. *$f \leftrightarrow \dots b_3 b_2 b_1 a_0 a_1 a_2 a_3 \dots$ is an isomorphism between $L^2[-\pi,\pi]$ and $\ell_2(\mathbb{Z})$.*

Note that if $f \in L^2_{\mathbb{R}}[-\pi,\pi]$ then $f \in L^2_{\mathbb{C}}[-\pi,\pi]$, *why?* We can express the c_n $(n \in \mathbb{Z})$ in terms of the a_m $(m \geq 0)$ and b_m $(m > 0)$, and conversely. Namely, for $n > 0$,

$$c_0 = a_0, \quad c_n = \frac{a_n}{\sqrt{2}} - i\frac{b_n}{\sqrt{2}}, \quad c_{-n} = \frac{a_n}{\sqrt{2}} + i\frac{b_n}{\sqrt{2}}.$$

Conversely (for $n > 0$)

$$a_0 = c_0, \quad a_n = \frac{1}{\sqrt{2}}(c_n + c_{-n}) = \sqrt{2}\,\Re(c_n), \quad b_n = \frac{i}{\sqrt{2}}(c_n - c_{-n}) = -\sqrt{2}\,\Im(c_n).$$

Check the above.

End of course for 2012!

4. Closed subspaces and orthogonal projection

4.1. Subspaces.

(1) A *subspace* of a Hilbert space $\mathcal{H}$ is a subspace in the vector space sense, i.e. is closed under addition and scalar multiplication.

Examples are subspaces of $\mathbb{R}^n$ and $\mathbb{C}^n$ in the usual sense. If $S = \{e_1, e_2, \dots\}$ is an orthonormal subset of $\mathcal{H}$, or indeed for any $S \subset \mathcal{H}$, then Span(S) is a subspace.

(2) A *closed* subspace is a subspace closed in the norm of $\mathcal{H}$.

(3) Any *finite* dimensional subspace is closed (shown in class).

If $S = \{e_1, e_2, \dots\}$ is an infinite orthonormal subset of $\mathcal{H}$ then Span(S) is not closed. For example, $\sum_{n\geq 1} \frac{1}{n} e_n \in \mathcal{H}$ (*why?*), but $\notin$ Span(S), *why?*

(4) A closed subspace of a Hilbert space is itself a Hilbert space with the induced inner product. (Any subspace is an inner product space, completeness follows from the fact we have a *closed* subspace. Separability is not surprising but requires proof. One argument is in Exercise 11.

A non closed subspace is an inner product space, but not a Hilbert space since it is not complete. *Why?*

4.2. Closest point on a closed subspace.

Theorem. *Suppose $\mathcal{S}$ is a closed subspace of a Hilbert space $\mathcal{H}$ and $f \in \mathcal{H}$. then*

(1) *There is a unique closest $g_0 \in \mathcal{S}$ to f, i.e.*

$$\|f - g_0\| = \inf_{g \in \mathcal{S}} \|f - g\|.$$

(2) *$f - g_0 \perp g$ for all $g \in \mathcal{S}$.*

Proof. See text pp175–177 for details. Here are the key points.

(1) Choose a minimising sequence $(g_n)_{n\geq 1}$, i.e.

$$g_n \in \mathcal{S}, \quad \|f - g_n\| \to d := \inf_{g \in \mathcal{S}} \|f - g\|.$$

(2) Use the *parallelogram law*[64]

$$2(\|f - g_n\|^2 + \|f - g_m\|^2) = \|2f - (g_n + g_m)\|^2 + \|g_n - g_m\|^2,$$

to show $(g_n)_{n\geq 1}$ is Cauchy. It follows $g_n \to g_0(\text{say}) \in \mathcal{S}$.

(3) Prove $(f - g_0, g) = 0$ for $g \in \mathcal{S}$ by showing that otherwise

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \|f - (g_0 + \epsilon g)\|^2 \neq 0.$$

(4) Prove g_0 is unique by noting that if $\widetilde{g}_0$ also minimises, then by what we have proved,

$$(f - g_0, g_0 - \widetilde{g}_0) = 0, \quad (f - \widetilde{g}_0, g_0 - \widetilde{g}_0) = 0.$$

Subtracting, $(g_0 - \widetilde{g}_0, g_0 - \widetilde{g}_0)$, so $g_0 = \widetilde{g}_0$.

□

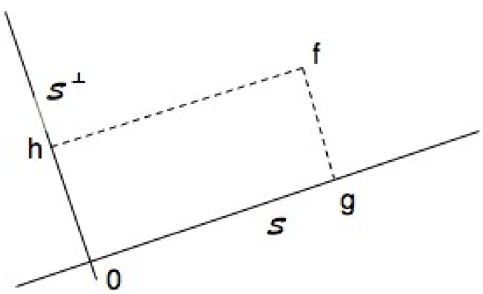

Figure 11. Orthogonal complement

[64] $2(\|u\|^2 + \|v\|^2) = \|u + v\|^2 + \|u - v\|^2$ for any $u, v \in \mathcal{H}$. What does this say in $\mathbb{R}^2$ about the sum of the square of the 4 edges, and the sum of the squares of the 2 diagonals, of a parallelogram?

4.3. **Orthogonal Complement.**

(1) If $\mathcal{S}$ is a subspace of $\mathcal{H}$ then the *orthogonal complement* is

$$\mathcal{S}^{\perp} = \{f \in \mathcal{H} : (f, g) = 0 \ \forall g \in \mathcal{S}\}.$$

(2) Then $\mathcal{S}^{\perp}$ is a subspace, *why?*
(3) Moreover, $\mathcal{S} \cap \mathcal{S}^{\perp} = \{0\|$, since $f \in \mathcal{S} \cap \mathcal{S}^{\perp} \Longrightarrow (f, f) = 0 \Longrightarrow f = 0$.
(4) $\mathcal{S}^{\perp}$ is *closed*, since $f_n \to f$ and $(f_n, g) = 0$ for all n implies $(f, g) = 0$ by Cauchy Schwarz.
(5) *Proposition*: *If* $\mathcal{S}$ *is a* closed *subspace, then* $\mathcal{H} = \mathcal{S} \oplus \mathcal{S}^{\perp}$.

We say $\mathcal{H}$ is the *direct sum* of $\mathcal{S}$ and $\mathcal{S}^{\perp}$. This means that every $f \in \mathcal{H}$ can be written uniquely in the form $f = g + h$ with $g \in \mathcal{S}$ and $h \in \mathcal{S}^{\perp}$.

Proof. The existence is immediate from the previous theorem with $g = g_0$ and $h = f - g_0$.

The uniqueness is because if we have two such sums $f = g_1 + h_1 = g_2 + h_2$ then $g_1 - g_2 = h_2 - h_1$. But then both equal zero by point 3., *why?* □

4.4. **Orthogonal projection.**

(1) If $\mathcal{H} = \mathcal{S} \oplus \mathcal{S}^{\perp}$ as in 5 (Proposition) above, then the *orthogonal projection* $P_{\mathcal{S}} : \mathcal{H} \to \mathcal{S}$ is defined by

$$P_{\mathcal{S}}(f) = g, \text{ where } f = g + h \text{ and } g \in \mathcal{S},\ h \in \mathcal{S}^{\perp}.$$

(2) $P_{\mathcal{S}}$ is linear, is the identity on $\mathcal{S}$, is 0 on $\mathcal{S}^{\perp}$, and $\|P_{\mathcal{S}}(f)\| \le \|f\|$.
(3) Suppose $\{e_1, e_2, \dots\}$ is a finite or countable set of orthonormal set of vectors. Suppose $\mathcal{S}$ is the closure of its span (which is the same as the span for a finite collection). Then

$$P_{\mathcal{S}}(f) = \sum_k (f, e_k) e_k.$$

This follows since
(a) the right side belongs to $\mathcal{S}$, and
(b) $(f - \sum_k (f, e_k) e_k, e_j) = (f, e_j) - (f, e_j) = 0$, and so $f - \sum_k (f, e_k) e_k \in \mathcal{S}^{\perp}$, *why?*

4.5. **Example of orthogonal projection.** Suppose $f \in L^2[-\pi, \pi]$. Then[65]

$$f(\theta) = \sum_{n=-\infty}^{\infty} a_n e^{in\theta}, \quad \text{where } a_n = ⨍_{-\pi}^{\pi} f(\phi) e^{-in\phi} d\phi.$$

By preceding comments 3., the orthogonal projection onto the space spanned by $\{e^{-iN\theta}, \dots, e^{iN\theta}\}$ is

$$\begin{aligned} S_N f(\theta) &:= \sum_{n=-N}^{N} a_n e^{in\theta} \\ &= \sum_{n=-N}^{N} \left(⨍_{-\pi}^{\pi} f(\phi) e^{-in\phi} d\phi \right) e^{in\theta} \\ &= \sum_{n=-N}^{N} ⨍_{-\pi}^{\pi} f(\phi) e^{in(\theta-\phi)} d\phi \\ &= ⨍_{-\pi}^{\pi} f(\phi) \left(\sum_{n=-N}^{N} e^{in(\theta-\phi)} \right) d\phi \\ &= ⨍_{-\pi}^{\pi} f(\phi) \frac{\sin(N+\frac{1}{2})(\theta-\phi)}{\sin\frac{1}{2}(\theta-\phi)} d\phi. \end{aligned}$$

The last equality comes from summing the geometric series

$$w^{-N} + \cdots + w^{N} = \dots = \frac{w^{N+\frac{1}{2}} - w^{-(N+\frac{1}{2})}}{w^{\frac{1}{2}} - w^{-\frac{1}{2}}}.$$

[65]The reason for using θ and ϕ is that we are often thinking of functions defined on the unit circle S^1.

Let

$$w = e^{i(\theta-\phi)},$$

and use the fact

$$e^{it} - e^{-it} = 2i \sin t.$$

Then

$$S_N f(\theta) = \int_{-\pi}^{\pi} f(\phi) D_N(\theta - \phi)\, d\phi = \frac{1}{2\pi}(f * D_N)(\theta),$$

(see top p74), where D_N is the *Dirichlet kernel*

$$D_N(\theta) := \frac{\sin(N + \frac{1}{2})\theta}{\sin \frac{1}{2}\theta}.$$

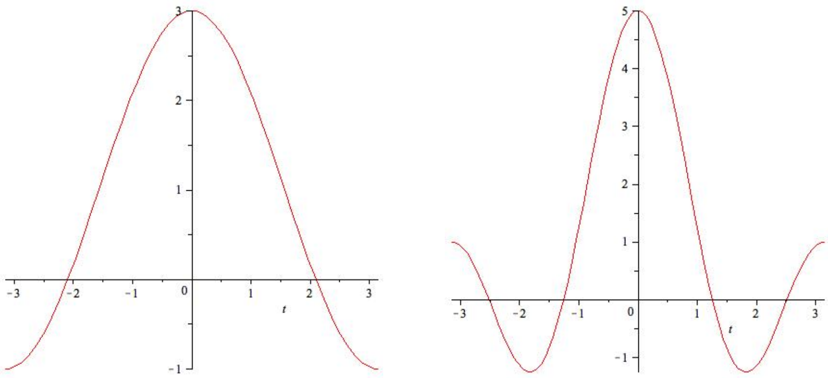

FIGURE 12. Graphs of D_1, D_2.

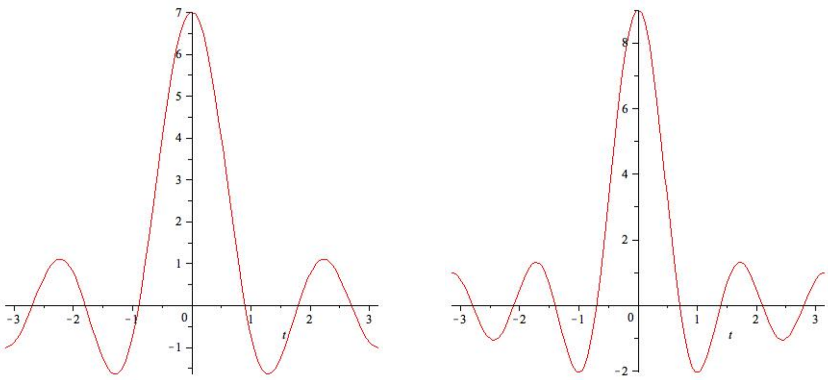

FIGURE 13. Graph of D_3, D_4.

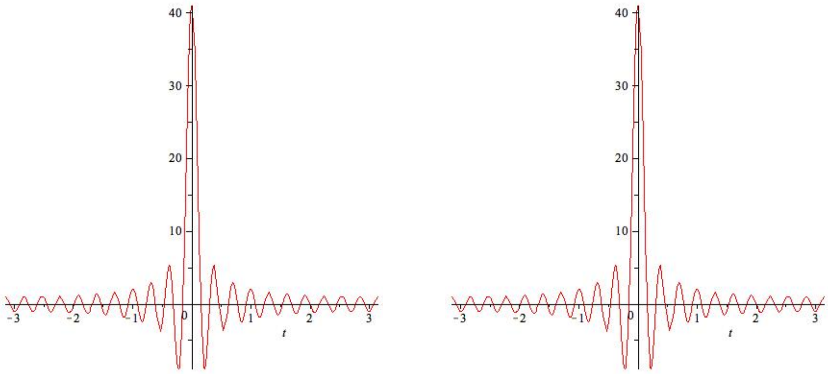

FIGURE 14. Graph of D_{10}, D_{20}.

5. Linear transformations

5.1. Basic definitions and properties. Suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces.

(1) $T : \mathcal{H}_1 \to \mathcal{H}_2$ is a *linear operator* or *linear transformation* if T is linear in the usual sense, i.e.
$$T(\alpha f + \beta g) = \alpha T(f) + \beta T(g).$$
(2) A linear operator T is *bounded* if
$$\|T\| := \sup\left\{\frac{\|T(f)\|}{\|f\|} : f \neq 0\right\} = \sup\{\|T(f)\| : \|f\| = 1\} < \infty.$$
(The second and third terms are equal by linearity.)

(3) *Proposition*: $\|T\| = \sup\{(Tf, g) : \|f\| \leq 1, \|g\| \leq 1$.

(Straightforward, see text p181, Lemma 5.1.)

(4) A linear operator is *continuous* if $f_n \to f \implies T(f_n) \to T(f)$.

(5) Proposition: *A linear operator is continuous iff it is bounded.*

($\Longleftarrow$): $\|T(f_n) - T(f)\| \leq \|T\|\,\|f_n - f\|$.

($\Longrightarrow$): Suppose T is not bounded. Then $\exists f_n$ for $n \geq 1$ such that $\|T(f_n)\| \geq n\|f_n\|$. Consider $g_n = f_n/n\|f_n\|$. Show $g_n \to 0$ but $\|T(g_n)\| \geq 1$.

5.2. Linear Functionals. Suppose $\mathcal{H}$ is a Hilbert space.

(1) A *linear functional* is a linear transformation $\ell : \mathcal{H} \to$ (the field of scalars), i.e. $\mathbb{R}$ or $\mathbb{C}$.

(2) If $g \in \mathcal{H}$ then
$$\ell(f) = (f, g)$$
defines a linear functional with $\|\ell\| = \|g\|$.

(Proof is straightforward.)

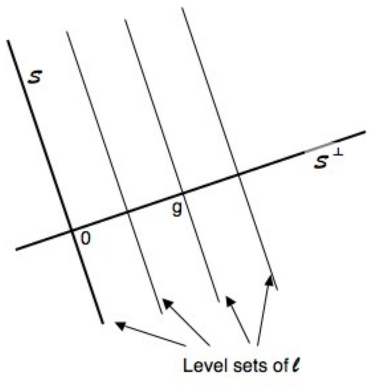

Figure 15. level sets of the linear functional $\ell(f) = (f, g)$.

Theorem 5.1 (Riesz representation theorem)**.** *Every linear functional on $\mathcal{H}$ is of the form $\ell(f) = (f, g)$ for a unique $g \in \mathcal{H}$.*

Proof Outline. **1.** Let $\mathcal{S} = \{f : \ell(f) = 0\}$ (this is the *null space* or *kernel* of ℓ). Then $\mathcal{S}$ is a subspace (since $\mathcal{S}$ is linear) and is closed (since $\mathcal{S}$ is continuous).

2. If $\mathcal{S} = \mathcal{H}$ then ℓ is the zero operator and we take $g = 0$. Otherwise, choose $h \in \mathcal{S}^\perp$ s.t. $\|h\| = 1$.

3. Check that $\ell(f)h - \ell(h)f \in \mathcal{S}$, and so $\big(\ell(f)h - \ell(h)f, h\big) = 0$.

4. Use this to show $\ell(f) = (f, \overline{\ell(h)}h)$. Let $g = \overline{\ell(h)}h$.

5. *Uniqueness*: Suppose $\ell(f) = (f, g_1) = (f, g_2)$ for all f. Then $(f, g_1 - g_2) = 0$, and taking $f = g_1 - g_2$ shows $g_1 = g_2$. □

Note: It follows $\mathcal{S}^\perp$ is one dimensional (in the real or complex sense).

To see this, suppose h_1 and h_2 are two unit vectors in $\mathcal{S}^\perp$. Then from the uniqueness result, $\overline{\ell(h_1)}h_1 = \overline{\ell(h_2)}h_2$, and so h_1 and h_2 are linearly dependent.

Concluding Remark. I recommend you look at the remainder of the chapter, pp183–193, to see a little history and important applications.

Part 6. DIFFERENTIATION and INTEGRATION

1. Questions and Examples

1.1. Integral average. If $m(E) < \infty$ we define the integral average of f over E by

$$\fint_E f = \frac{1}{m(E)} \int_E f.$$

1.2. Questions. Consider measurable $f : \mathbb{R} \to \mathbb{R}$.

(1) If f is integrable, does the derivative of the integral exist a.e. and does it equal f a.e.?

That is, does

$$\lim_{h \to 0} \frac{\int_a^{x+h} f(y)\, dy - \int_a^x f(y)\, dy}{h} = f(x) \text{ a.e.?}$$

Equivalently, does (for the $h > 0$ case)?

$$\lim_{h\to 0} \left(\fint_{[x,x+h]} f \right) \to f(x) \text{ as } h \to 0, \text{ for a.e. } x?$$

We will see the answer is YES.

We will also generalise this to $\mathbb{R}^d$ for any d.

(2) Conversely, when is f differentiable a.e.? And when is the derivative of the integral of f equal to f a.e.?

That is, when does f' exist a.e.? And when does

$$\int_a^x f'(y)\, dy = f(x) - f(a) \text{ a.e.?} \tag{33}$$

We will see that f' exist a.e. if f has *bounded variation.*

We will see (**33**) is true if and only if there is an *absolutely continuous* function equal to f a.e. If f is replaced by this absolutely continuous representative, then (**33**) is true *everywhere.*

There are generalisations to higher dimensions, but these are quite subtle. See "Functions of Bounded Variation and Free Discontinuity Problems" by Ambrosio, Fusco, Pallara (QA316 .A52 2000).

1.3. Examples and Remarks for Question 1.

(1) If f is continuous then we know from Riemann integration theory that the answer to Question 1. is YES everywhere.

(2) If f is not continuous we cannot expect the answer to Question 1. to be YES everywhere. For example, let $f(x) = 1$ if $x > 0$, and $f(x) = 0$ if $x \le 0$. What happens at 0?

A measurable function may be nowhere continuous, take the characteristic function of the irrationals. But the answer to Question 1. is still YES *almost everywhere.*

1.4. Examples and Remarks for Question 2.

(1) If f is differentiable and the derivative is continuous then we know from Riemann integration theory that the answer to Question 2. is YES everywhere.

(2) Consider $f(x) = |x|$. The derivative exists except at the origin. But (**33**) is still true (in fact everywhere).

(3) Consider $f(x) = 1$ if $x > 0$, and $f(x) = 0$ if $x \le 0$. Then $f'(x) = 0$ a.e. (everywhere except at $x = 0$) but (**33**) is not true. It is for this reason that we will require f to be continuous (but not f'!).

(4) Even if f is continuous and the derivative exists a.e., then (**33**) may not be true. In fact, we can also have f continuous, increasing, $f' = 0$ a.e, and (**33**) false.

See text pp126, 127. Or see the diagram and discussion of the Cantor-Lebesgue function (this is a clickable link, or go directly to http://en.wikipedia.org/wiki).

2. Differentiation of the integral

(1) Suppose f integrable on $[a,b]$ and let

$$F(x) = \int_a^x f(y)\,dy \quad a \le x \le b.$$

Then does $F'(x)$ exist a.e. and does

$$F'(x) \left(\text{i.e. } \lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(y)\,dy, \text{ i.e. } \lim_{\substack{|I|\to 0\\ I=[x,x+h]}} ⨍_I f \right) = f(x) \text{ a.e.?}$$

(2) Reformulating, we ask if for closed intervals I the following limit exists and if

$$\lim_{\substack{|I|\to 0\\ x\in I}} ⨍_I f = f(x) \quad \text{a.e.}$$

This is a stronger requirement since x need not be an endpoint of I.

We will show the answer is YES.

(3) For technical reasons we will instead take *open* intervals I.

But the limit exists and equals $f(x)$ for open intervals I iff it exists and equals $f(x)$ for closed intervals I iff it exists and equals $f(x)$ for open intervals I but with the condition $x \in \overline{I}$, *why?*

(4) More generally, we will prove if f is integrable on $\mathbb{R}^d$ then

$$\lim_{\substack{m(B)\to 0\\ x\in B}} ⨍_B f = f(x) \quad \text{a.e.} \tag{34}$$

Here B ranges over open balls containing x. (We could require x to be the centre of B but this is a weaker requirement.)

One can also take closed balls, or open balls that have x in their closure, without changing any limits, *why?*

(5) If f is continuous at x then (34) is true, *why?* (see text p100)

2.1. Maximal function.

Definition. If f integrable over $\mathbb{R}^d$ (more generally, if $\int_E |f| < \infty$ for every bounded $E \subset \mathbb{R}^d$) then the *maximal function* is

$$f^*(x) = \sup_{x\in B} ⨍_B |f|.$$

The sup is over *all open* balls $B \ni x$.

Note that f^* depends only on $|f|$, rather than f.

2.2. Examples.

(1) $f = \mathcal{X}_{[-1,1]}$.

If $0 \le x < 1$ then $f^*(x) = ⨍_I f = 1$, for any interval I with $x \in I \subset [-1,1]$.

If $x > 1$ then $f^*(x) = ⨍_{[-1,x]} f = 2/(x+1)$, *why?* Note that it is OK to take a closed interval, *why?*

Similarly for $x \le 0$, by symmetry of f. So

$$f^*(x) = \begin{cases} 1 & |x| \le 1 \\ \frac{2}{|x|+1} & |x| \ge 1 \end{cases}.$$

(2) $f(x) = |x|^\alpha$ where $-1 < \alpha < 0$.

If $x > 0$ then (*why?*)

$$\begin{aligned} f^*(x) &= \sup_{y<0} ⨍_{[y,x]} |x|^\alpha \, dx \\ &= \sup_{y<0} \frac{1}{x-y} \frac{x^{\alpha+1} + |y|^{\alpha+1}}{\alpha+1} \\ &= \sup_{t>0} \frac{1}{x+t} \frac{x^{\alpha+1} + t^{\alpha+1}}{\alpha+1} \\ &= x^\alpha \sup_{t>0} \frac{1 + (t/x)^{\alpha+1}}{(\alpha+1)(1+t/x)} = cx^\alpha, \end{aligned}$$

where $c := \sup_{s>0} \dfrac{1+s^{\alpha+1}}{(\alpha+1)(1+s)}$ and does not depend on x. The sup is achieved for some $s \in (0,\infty)$, *why?* If $\alpha = -1/2$ then calculation shows $c = 1+\sqrt{2}$.

So $f^*(x) = c|x|^\alpha$ with c as above.

Theorem. *Suppose f is integrable on $\mathbb{R}^d$. Then*
(1) *f^* is measurable*
(2) *$f^* < \infty$ a.e.*
(3) *$m\{f^* > \alpha\} \le \dfrac{3^d}{\alpha}\|f\|$, where $\|f\| = \int |f|$.*

2.3. **Remarks.**
(1) The actual value of the constant 3^d is not particularly important.
(2) The conclusions of the theorem are still true, but with a different constant, if we use balls *centred* at x in the definition of f^*, *why?* This defines the *centred* maximal function.
We can also use closed balls B and require $x \in B^\circ$ (the interior of B), which does not change $⨍_B |f|$, *why?* We can also allow $x \in B$ in this case without changing $f^*(x)$, *why?* Or use cubes instead of balls, which changes $f^*(x)$ by at most a fixed constant factor, *why?*.
(3) We will see in Corollary 2.2 that $|f| \le f^*$ a.e. (*why* do we expect this for continuous f, and why do we not expect "="?)
(4) *Tchebychev's inequality* (Ex 9, p91) gives a similar, but much easier, estimate for $|f|$ as in (3) above, namely
$m\{|f| > \alpha\} \le \dfrac{1}{\alpha}\|f\|$. So we see f^* is not "much" bigger than $|f|$ in terms of the size of sets where the two functions f^* and $|f|$ are big.
(5) However, it is not necessarily true that $\int |f^*| < \infty$, or even that f^* is locally integrable! See Exercise 4 and 5 p146 for counter examples.

We first need to prove the following important result.

Theorem (Vitali covering lemma). *Suppose $\{B_1,\dots,B_N\}$ is a finite collection of open balls in $\mathbb{R}^d$. Then there exists a disjoint subcollection $\{B_{i_1},\dots,B_{i_k}\}$ such that*

$$\bigcup_{\ell=1}^N B_\ell \subset \bigcup_{j=1}^k m(\widetilde{B_{i_j}}),$$

where $\widetilde{B}$ is the ball with the same centre as B and 3 times the radius.
It follows

$$m\left(\bigcup_{\ell=1}^N B_\ell\right) \le 3^d \sum_{j=1}^k m(B_{i_j}).$$

Proof. Pick a ball B_{i_1} from $\{B_1,\dots,B_N\}$ having maximal size.
Discard all remaining balls in $\{B_1,\dots,B_N\}$ that meet B_{i_1}.
Pick a ball B_{i_2} from the remaining balls in $\{B_1,\dots,B_N\}$ having maximal size.
Discard all remaining balls in $\{B_1,\dots,B_N\}$ that meet B_{i_2}.
Etc.

Any ball B in $\{B_1, \dots, B_N\}$ must meet some B_{i_j}, have equal or smaller radius, and hence be covered by $\widetilde{B_{i_j}}$. □

We now prove the theorem.

Proof Sketch. **1.** Suppose $\overline{x} \in E = \{f^* > \alpha\}$. Then $⨍_B |f| > \alpha$ for some $B \ni \overline{x}$. Select $\epsilon > 0$ $|x - \overline{x}| < \epsilon \Longrightarrow x \in B$. Then $x \in E$ for all such x and so E is open. Hence E is measurable.

3. Suppose $K \subset E$ and K is compact. We will show the required estimate for K. Since $m(E) = \sup\{m(K) : K \subset E,\ K\text{compact}\}$ see Thm 3.4(iii) p21 (and even if $m(E) = \infty$, *why?*), the result will follow.

For each $x \in K$ there exists a ball $B_x \ni x$ such that $⨍_{B_x} |f| > \alpha$.
By compactness $\exists$ a finite subcover of K which we denote by $\{B_1, \dots, B_N\}$.
Even though this subcover is not disjoint, by Vitali $\exists$ a disjoint subcollection $\{B_{i_1}, \dots, B_{i_k}\}$ such that the 3 times dilations $\{\widetilde{B_{i_1}}, \dots, \widetilde{B_{i_k}}\}$ covers $\bigcup_{\ell=1}^N B_\ell$.
Hence (*why is each inequality true?*),

$$m(K) \le m\left(\bigcup_{\ell=1}^N B_\ell\right) \le m\left(\bigcup_{j=1}^k \widetilde{B_{i_j}}\right) \le \sum_{j=1}^k m(\widetilde{B_{i_j}})$$
$$\le 3^d \sum_{j=1}^k m(B_{i_j}) \le \frac{3^d}{\alpha} \sum_{j=1}^k \int_{B_{i_j}} |f| \le \frac{3^d}{\alpha} \|f\|.$$

This establishes the third (and main) part of the theorem.

2. Note

$$m(\{f^* = \infty\}) \le m(\{f^* > \alpha\}) \le \frac{3^d}{\alpha}\|f\| \to 0 \text{ as } \alpha \to \infty.$$

This establishes the second part of the theorem. □

2.4. **Lebesgue Differentiation Theorem.** In the following theorem we consider $f \in L^1(\mathbb{R}^d)$. However, it is clear that the existence and value of the limit in (35) is independent of the value of f outside any fixed neighbourhood of x. In fact, all that is required in the theorem is that $f \in L^1_{loc}(\mathbb{R}^d)$, which means that $\int_E |f| < \infty$ for any E with $m(E) < \infty$. We then say that f is *locally integrable*. The result in this case is then immediate, just apply the theorem to $\mathcal{X}_{B_N} f$, where B_N is a the ball centred at the origin and of radius N. It follows the result is true for a.e. $x \in B_N$, and hence for a.e. x since N is arbitrary.

Theorem 2.1. *Suppose $f \in L^1(\mathbb{R}^d)$ (or $L^1_{loc}(\mathbb{R}^d)$). Then*

$$\lim_{\substack{|B| \to 0 \\ x \in B}} ⨍_B f(y)\, dy = f(x) \text{ for a.e. } x. \tag{35}$$

where B ranges over open balls containing x.

Proof and Discussion.
First Note:

(1) The result is easily seen to be true if f is replaced by a continuous function, *why?*
(2) Recall for $f \in L^1(\mathbb{R}^d)$ there is a sequence of continuous functions (g_n) such that $\|f - g_n\|_{L^1} \to 0$ as $n \to \infty$.
(3) From Tchebychev's theorem (easy) and the maximal function theorem (hard) we have

$$m\big\{|f - g_n| > \alpha\big\} \le \frac{1}{\alpha}\|f - g\|_{L^1},$$
$$m\big\{|(f - g_n)^*| > \alpha\big\} \le \frac{A}{\alpha}\|f - g\|_{L^1}.$$

for some constant A (in fact, 3^d). Recall $(f - g_n)^* = \sup_{x \in B} ⨍_B \big(f(y) - g_n(y)\big)\, dy$.

We will use these facts in the following proof.

Step 1. We do not initially know that the limit in (35) even exists. But (35) is equivalent to showing

$$\limsup_{\substack{|B|\to 0\\ x\in B}} \left| ⨍_B \big(f(y)-f(x)\big)\,dy\right| = 0 \text{ for a.e. } x. \tag{36}$$

Note that the lim sup always exists[66] although it possibly is $+\infty$.

This is equivalent to showing

$$m\left\{ x : \limsup_{\substack{|B|\to 0\\ x\in B}} \left| ⨍_B \big(f(y)-f(x)\big)\,dy\right| > 0 \right\} = 0, \tag{37}$$

which in turn is equivalent to showing that for *every* $\alpha > 0$,

$$m\left\{ x : \limsup_{\substack{|B|\to 0\\ x\in B}} \left| ⨍_B \big(f(y)-f(x)\big)\,dy\right| > \alpha \right\} = 0, \tag{38}$$

why?

Step 2. In order to show that the measure on the left side of (38) is indeed zero, suppose (g_n) is a sequence of continuous functions with $\|f-g_n\|\to 0$ and write

$$\begin{aligned}\left| ⨍_B \big(f(y)-f(x)\big)\,dy\right| &\le \left| ⨍_B \big(f(y)-g_n(y)\big)\,dy\right| + \left| ⨍_B \big(g_n(y)-g_n(x)\big)\,dy\right| + \left| ⨍_B \big(g_n(x)-f(x)\big)\,dy\right| \\ &=: I_1 + I_2 + I_3,\end{aligned}$$

(where each I_j depends on n, x and B). It follows that

$$\begin{aligned}\limsup_{\substack{|B|\to 0\\ x\in B}} \left| ⨍_B \big(f(y)-f(x)\big)\,dy\right| &\le \limsup_{\substack{|B|\to 0\\ x\in B}} I_1 + \limsup_{\substack{|B|\to 0\\ x\in B}} I_2 + \limsup_{\substack{|B|\to 0\\ x\in B}} I_3 \\ &= \limsup_{\substack{|B|\to 0\\ x\in B}} I_1 + 0 + |g_n(x)-f(x)|.\end{aligned}$$

(For the second term we use the fact $⨍_B g_n \to g_n(x)$ as $|B|\to 0$ for $x\in B$, by the continuity of g_n at x. For the third term we just note that $g_n(x)-f(x)$ is a constant in the relevant integral. Note that we could not have obtained the 0 if we had estimated the second integral by $⨍_B \big|g_n(y)-g_n(x)\big|\,dy$, *why?*)

It follows that

$$\begin{aligned}\limsup_{\substack{|B|\to 0\\ x\in B}} \left| ⨍_B \big(f(y)-f(x)\big)\,dy\right| &\le \sup_{x\in B} \left| ⨍_B \big(f(y)-g_n(y)\big)\,dy\right| + |g_n(x)-f(x)| \\ &\le \sup_{x\in B} ⨍_B \big|f(y)-g_n(y)\big|\,dy + |g_n(x)-f(x)| \\ &= (f-g_n)^*(x) + |g_n(x)-f(x)| \end{aligned} \tag{39}$$

(The first inequality is clear. But it may seem like too coarse an estimate and that we are "throwing away too much", since we are really only interested in *very small* balls containing x. But if $\sup_{x\in B} \left|⨍_B \big(f(y)-g_n(y)\big)\,dy\right|$ is big it will "usually" happen for $|B|$ very small.)

---

[66] There are a number of equivalent definitions of the limsup. For example,

$$\limsup_{\substack{|B|\to 0\\ x\in B}} G(x,B) = b$$

means

$$\begin{gathered}\exists (B_n)_{n\ge 1} \text{ such that } |B_n|\to 0,\ x\in B_n \ \&\ \lim G(x,B_n) = b; \\ \forall (B_n)_{n\ge 1},\ \text{if } |B_n|\to 0,\ x\in B_n \ \&\ \lim G(x,B_n) = a,\ \text{then } a\le b.\end{gathered}$$

Step 3. From (39), using the maximal function estimate and Tchebychev's theorem,

$$m\left\{x : \limsup_{\substack{|B|\to 0\\ x\in B}} \left| \fint_B \big(f(y)-f(x)\big)\,dy\right| > \alpha\right\} \le m\left\{(f-g_n)^* > \frac{\alpha}{2}\right\} + m\left\{|f-g_n| > \frac{\alpha}{2}\right\}$$

$$\le \frac{c}{\alpha}\|f-g_n\| + \frac{2}{\alpha}\|f-g_n\|,$$

for some constant c, in fact $c = 2\times 3^d$. Since $\|f-g_n\|\to 0$ it follows (38) is true, hence (37) is true, hence (36) is true, and hence (35). □

Corollary 2.2. *Suppose* $f \in L^1(\mathbb{R}^d)$ *(or* $L^1_{loc}(\mathbb{R}^d)$*). Then* $|f(x)| \le f^*(x)$ *for a.e.* x.

Proof. Applying the previous theorem to $|f|$ gives

$$|f(x)| = \lim_{\substack{|B|\to 0\\ x\in B}} \fint_B |f(y)|\,dy \le \sup_{x\in B} \fint_B |f(y)|\,dy = f^*(x),$$

for a.e. x. □

Definition 2.3. If $E \subset \mathbb{R}^d$ and $x \in \mathbb{R}^d$ then the *density* of E at x is

$$\theta(E,x) = \lim_{\substack{|B|\to 0\\ x\in B}} \frac{m(B\cap E)}{m(B)},$$

provided the limit exists. (We will see in the next Corollary that the limit exists a.e.)

Often the density is defined to be

$$\theta(E,x) = \lim_{r\to 0} \frac{m(B_r(x)\cap E)}{m(B_r(x))},$$

provided the limit exists.

Remark 2.4.

(1) If the limit exists in the first sense then it certainly exists in the second. *Why?*

(2) If E is a closed rectangle in $\mathbb{R}^2$ then the limit in the *second* sense exists everywhere. It takes the values 1 in the interior, either 1/2 or 1/4 on the boundary, and 0 in the complement of E. *Why?*

The limit in the *first* sense is 1 in the interior and 0 in the complement of E, but does note exist on the boundary of E. *Why?*

(3) It might help to think of applying a microscope at x and dialling up the magnification. In the second case the microscope is centred at x and in the first case it is a wobbly microscope.

□

Corollary 2.5. *Suppose* $E \subset \mathbb{R}^d$ *is measurable. Then*

$$\theta(E,x) = \begin{cases} 1 & a.e.\ x\in E \\ 0 & a.e.\ x\notin E \end{cases}.$$

Proof. This is just the Lebesgue Differentiation Theorem with $f = \mathcal{X}_E$. □