Real Analysis

Part I: MEASURE THEORY

1. Algebras of sets and σ-algebras

For a subset $A \subset X$, the complement of A in X is written $X - A$. If the ambient space X is understood, in these notes we will sometimes write A^c for $X - A$. In the literature, the notation A' is also used sometimes, and the textbook uses $\tilde{A}$ for the complement of A. The set of subsets of a set X is called the *power set of* X, written 2^X.

Definition. A collection $\mathcal{F}$ of subsets of a set X is called an *algebra of sets* in X if
1) $\emptyset, X \in \mathcal{F}$
2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

In mathematics, an "algebra" is often used to refer to an object which is simultaneously a ring and a vector space over some field, and thus has operations of addition, multiplication, and scalar multiplication satisfying certain compatibility conditions. However the use of the word "algebra" within the phrase "algebra of sets" is not related to the preceding use of the word, but instead comes from its appearance within the phrase "Boolean algebra". A Boolean algebra is not an algebra in the preceding sense of the word, but instead an object with operations `AND`, `OR`, and `NOT`. Thus, as with the phrase "Boolean algebra", it is best to treat the phrase "algebra of sets" as a single unit rather than as a sequence of words. For reference, the definition of a Boolean algebra follows.

Definition. A *Boolean algebra* consists of a set Y together with binary operations $\vee$, $\wedge$, and a unariy operation $\sim$ such that
1) $x \vee y = y \vee x; \quad x \wedge y = y \wedge x \qquad$ for all $x, y \in Y$
2) $(x \vee y) \vee z = x \vee (y \vee z); \quad (x \wedge y) \wedge z = x \wedge (y \wedge z) \qquad$ for all $x, y, z \in Y$
3) $x \vee x = x; \quad x \wedge x = x \qquad$ for all $x \in Y$
4) $(x \vee y) \wedge x = x; \quad (x \wedge y) \vee x = x \qquad$ for all $x \in Y$
5) $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \qquad$ for all $x, y, z \in Y$
6) $\exists$ an element $0 \in Y$ such that $0 \wedge x = 0$ for all $x \in Y$
7) $\exists$ an element $1 \in Y$ such that $1 \vee x = 1$ for all $x \in Y$
8) $x \wedge (\sim x) = 0; \quad x \vee (\sim x) = 1 \qquad$ for all $x \in Y$

Other properties, such as the other distributive law $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$, can be derived from the ones listed.

An equivalent formuation can be given in terms of partially ordered sets.

Definition. A *partially ordered set* consists of a set X together with a relation $\leq$ such that
1) $x \leq x \qquad \forall x \in X \qquad$ reflexive
2) $x \leq y, y \leq z \Rightarrow x \leq z \qquad$ transitive

3) $x \le y, y \le x \Rightarrow x = y$ (anti)symmetric

If A is a subset of a partially ordered set X, $g \in X$ is called the *greatest lower bound* for A if $g \le a$ for all $a \in A$, and if $g' \le a$ for all $a \in A$ then $g' \le g$. The *least upper bound* for A is defined in a similar fashion, reversing the inequality. In the case $X = \mathbb{R}$, the greatest lower bound and least upper bound for A go by the names *infimum* and *supremum* respectively, written $\inf A$ and $\sup A$.

Of course, in general a subset of a partially ordered set need not have a greatest lower bound nor a least upper bound. A partially ordered set in which every pair of elements has a greatest lower bound and a least upper bound is called a *lattice*. (Note: The word "lattice" has other meanings in other parts of mathematics.)

Given a Boolean algebra X, a partial order on X can be defined by setting $x \le y$ if and only if $x \wedge y = x$, which comes out to be equivalent to $x \vee y = y$. One can verify that with this definition, every pair of elements $\{x, y\}$ has a greatest lower bound and a least upper bound, given by $x \wedge y$ and $x \vee y$ respectively. Thus a Boolean algebra becomes a lattice which has the additional properties (5)–(8).

Conversely, given a lattice X, set $x \wedge y := \mathrm{glb}\{x, y\}$ and $x \vee y := \mathrm{lub}\{x, y\}$. Then properties (1)–(4) in the definition of Boolean algebra are satisfied. If property (5) is satisfied, the lattice is called *distributed* and if properties (6)–(8) are satisfied it is called *complemented.*

In summary,

Proposition. *A Boolean algebra is equivalent to a distributed complement lattice.* $\square$

If X is any set and $Y = 2^X$, then Y becomes a Boolean algebra with the assignments:
1) $\vee :=$ union
2) $\wedge :=$ intersection
3) $0 := \emptyset$
4) $1 := X$
5) $\sim :=$ complementation

In the special case where $X = \{\mathtt{T}, \mathtt{F}\}$ (standing for "True" and "False"), we obtain the standard Boolean algebra used in Logic (Propositional Calculus) in which the operations become `AND`, `OR` and `NOT`.

If $\mathcal{F}$ is algebra of sets in X, then $\mathcal{F}$ becomes a sub-Boolean-algebra of 2^X, since the definition of an algebra of sets, together with de Morgan's Law $(A \cup B)^c = A^c \cap B^c$ imply that the Boolean operations on 2^X restrict to operations on $\mathcal{F}$.

Definition. A collection $\mathcal{F}$ of subsets of a set X is called a *σ-algebra of sets* in X if
1) $\emptyset, X \in \mathcal{F}$
2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3) $\{A_i\}_{i=1}^{\infty} \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

Example. Let the *fundamental subsets* of $\mathbb{R}$ be those which are finite unions of intervals. The collection of fundamental subsets forms an algebra of sets which is not a σ-algebra.

Recall that a set is called *countable* if it is either finite or there exists a bijection between it and the natural numbers $\mathbb{N}$.

Proposition. *Let $\mathcal{F}$ be a σ-algebra of sets in X. Then any countable union of sets in $\mathcal{F}$ belongs to $\mathcal{F}$ and any countable intersection of sets in $\mathcal{F}$ belongs to $\mathcal{F}$.*

Proof. The statement for unions is part (3) of the definition, and the one for intersections then follows from part (2) of the definition and de Morgan's Law. □

Proposition. *Let X be a set and for each $j \in J$, let $\mathcal{F}_j \subset 2^X$ be a σ-algebra of sets in X. Then $\cap_{j \in J} \mathcal{F}_j$ is a σ-algebra of sets in X.*

Proof. Set $\mathcal{F} := \cap_{j \in J} \mathcal{F}_j$. Since $\emptyset \in \mathcal{F}_j$ and $X \in \mathcal{F}_j$ for all j, it follows that $\emptyset \in \mathcal{F}$ and $X \in \mathcal{F}$. If $A \in \mathcal{F}$, then $A \in \mathcal{F}_j$ for all j and so $A^c \in \mathcal{F}_j$ for all j and therefore $A^c \in \mathcal{F}$. If $\{A_i\}_{i=1}^\infty \in \mathcal{F}_j$ for all j then $\cup_{i=1}^\infty A_i \in \mathcal{F}_j$ for all j and so $\cup_{i=1}^\infty A_i \in \mathcal{F}$. □

A consequence of the preceding proposition is that if $\mathcal{U}$ is any collection of subsets of X there is always a "smallest σ-algebra of X containing $\mathcal{U}$", obtained by intersecting all σ-algebras of X containing $\mathcal{U}$. The smallest σ-algebra of X containing $\mathcal{U}$ is sometimes called the σ-algebra of X generated by $\mathcal{U}$.

Definition. The *Borel sets* are the σ-algebra in $\mathbb{R}$ generated by the fundamental sets.

Recall that a point $a \in A \subset \mathbb{R}^n$ is called an *interior point* of A if there exists $r > 0$ such that the open ball $B_r(a) := \{x \in \mathbb{R}^n \mid \|x - a\| < r\}$ is contained in A. The set A is called *open* if every point of A is an interior point.

Since it is easy to see that every fundamental set is a countable union of sets each of which is either or open it follows that

Proposition. *The fundamental sets are the algebra n $\mathbb{R}$ generated by the open sets.* □

Proposition. *Any uncountable subset of $\mathbb{R}$ has an accumulation point.*

Proof. Let A be a subset of $\mathbb{R}$ and suppose that A has no accumulation point. Recall from `MATB43` that the Bolzano-Weierstass theorem says that any infinite bounded set has an accumulation point. Therefore $A \cap [-n, n]$ is finite for all n, and so $A = \cup_n (A \cap [-n, n])$ is a countable union of finite sets and thus is countable. □

Proposition. *Any open subset of $\mathbb{R}$ is a countable union of open intervals.*

Proof. Let $A \subset \mathbb{R}$ be open. Then for each $a \in A$, there exists $r_a > 0$ such that $(a - r_a, a + r_a) \subset A$. Therefore A can be written as the union of (possibly uncountably many) intervals. By amalgamating intervals which intersect, we can write $A = \cup_{\alpha \in J} I_\alpha$ as a union of disjoint intervals I_α. Set $J' := \{\alpha \in J \mid \text{length}(I_\alpha) < \infty\}$. There can be at most two infinite length intervals in the disjoint union $\cup_{\alpha \in J} I_\alpha$, so it suffices to show that the index set J' is countable. Set $J_n := \{\alpha \in J' \mid \text{length}(I_\alpha) > 1/n\}$. The set $C_n := \{a_\alpha\}_{\alpha \in J_n}$ consisting of the centres of each of these intervals is a bounded subset of $\mathbb{R}$ with no accumulation point, since the distance between any pair of distinct points in C_n is at least $2/n$. Therefore the index set J_n is finite for each n. Thus $J' = \cup_n J_n$ is countable and so J is countable. □

Proposition. *The Borel sets are the σ-algebra generated by the open sets in $\mathbb{R}$.*

Proof. Since every open set is a countable union of intervals, every open set is a Borel set. Therefore the σ-algebra generated by the open sets is contained in the Borel sets. Conversely, since every every σ-algebra containing the open sets contains all the fundamental sets, every Borel set is contained in the σ-algebra generated by the open sets. □

In a topological space X, a set that can be written as a union of any collection (not necessarily finite) of open sets is always open and a set that can be written as any intersection of closed sets is always closed. A set that can be written as a countable union of closed sets is called an F_σ-set and a set which can be written as a countable intersection of open sets is called a G_δ-set. An F_σ-set need not be closed and a G_δ-set need not be open. For example $(-1/n, 1/n)$ is open in $\mathbb{R}$ for all n so $\{0\} = \cap_n(-1/n, 1/n)$ is a G_δ set in $\mathbb{R}$, but it is not open.

A set which can be written as a countable intersection of F_σ-sets is called an $F_{\sigma\delta}$-set and a set which can be written as a countable union of G_δ sets is called a $G_{\delta\sigma}$-set. Similarly we can define $F_{\sigma\delta\sigma}$-sets, $G_{\delta\sigma\delta}$-sets, etc.

Proposition. *Any $F_{\sigma\delta,\ldots,\sigma}$, $F_{\sigma\delta,\ldots,\sigma,\delta}$, $G_{\delta\sigma,\ldots,\sigma}$, or $G_{\delta\sigma,\ldots,\sigma,\delta}$ set is a Borel set.*

Proof. This follows from the earlier proposition stating that any countable union or intersection of sets in a σ-algebra lies in the σ-algebra. $\square$

2. Measures

Definition. Let $\mathcal{F}$ be an algebra of sets in X. A function $\mu : \mathcal{F} \to [0, \infty]$ is called a *content* on $\mathcal{F}$ if
1) $\mu(\emptyset) = 0$
2) $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$.

If there exists any set A such that $\mu(A) < \infty$, condition (1) can be derived from condition (2). Some books omit condition (1) in which case under their definition, but not under ours, the function $\mu(A) = \infty$ for all $A \in \mathcal{F}$ would be considered to be a content.

If μ is a content, it follows by induction that $\mu(A_1 \cup \ldots A_n) = \mu(A_1) + \ldots \mu(A_n)$ whenever $A_1 \ldots A_n$ are disjoint sets in $\mathcal{F}$.

The notation $A \amalg B$ is sometimes used for the union of A and B in the case where A and B are disjoint.

Proposition. Let μ be a content on an algebra of sets $\mathcal{F}$. If $A \subset B$ then $\mu(B - A) = \mu(B) - \mu(A)$ and in particular $\mu(A) \leq \mu(B)$.

Proof. $B = A \amalg (B - A)$ so $\mu(B) = \mu(A) + \mu(B - A)$. $\square$

In the case of sets which are not necessarily disjoint we have

Proposition. If $\{A_i\}_{i=1}^n \subset \mathcal{F}$, then $\mu(\cup A_i) \leq \sum_{i=1}^n \mu(A_i)$.

Proof.

$$\bigcup_{i=1}^n A_i = A_1 \amalg (A_2 - A_1) \amalg \big(A_3 - (A_1 \cup A_2)\big) \amalg \big(A_n - (A_1 \cup A_2 \ldots \cup A_{n-1})\big).$$

Since $A_j - \cup_{i=1}^{j-1} A_i \subset A_j$ we have $\mu\left(A_j - \cup_{i=1}^{j-1} A_i\right) \leq \mu(A_j)$. Therefore

$$\mu\left(\cup_{i=1}^n A_i\right) = \sum_{j=1}^n \mu(A_j - \cup_{i=1}^{j-1} A_i) \leq \sum_{i=1}^n \mu(A_i).$$

$\square$

Definition. Let $\mathcal{F}$ be a σ-algebra of sets in X. A function $\mu : \mathcal{F} \to [0, \infty]$ is called an *outer measure* on $\mathcal{F}$ if

1) $\mu(\emptyset) = 0$
2) $\mu(\cup A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$ for any countable collection $\{A_i\}_{i=1}^{\infty}$ of sets in $\mathcal{F}$.
   An outer measure satisfying the stronger condition

2') $\mu(\cup A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for any countable collection $\{A_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{F}$ is called a *measure* on $\mathcal{F}$.

We wish to construct a measure μ on some σ-algebra $\mathcal{M}$ in $\mathbb{R}$ such that

1) $\{\text{Fundamental Sets}\} \subset \mathcal{M}$
2) $\mu(\text{interval}) = \text{length of interval}$
3) μ is translation invariant. That is, $S \in \mathcal{M}$ implies $S + x \in \mathcal{M}$ with $\mu(S + x) = \mu(S)$ for all $x \in \mathbb{R}$.

Clearly there exists a content μ on $\{\text{Fundamental Set}\}$ satisfying these three conditions. The question is whether μ can be extended to some σ-algebra $\mathcal{M}$. Ideally one might hope to be able to choose $\mathcal{M} = 2^{\mathbb{R}}$, but we shall see later that this is impossible.

We can, however, define an outer measure on $2^{\mathbb{R}}$ satisfying the three conditions. Given a content $\mu : \mathcal{F} \to [0, \infty]$, define the associated outer measure $\mu^* : 2^X \to [0, \infty]$ by

$$\mu^*(A) := \inf \left\{ \sum_{n=1}^{\infty} \mu(I_n) \mid I_n \text{ is a collection of sets in } \mathcal{F} \text{ such that } A \subset \cup_n I_n \right\}.$$

In the case of the fundamental sets in $\mathbb{R}$ this becomes

$$\mu^*(A) := \inf \left\{ \sum_{n=1}^{\infty} \mu(I_n) \mid I_n \text{ is a collection of open intervals such that } A \subset \cup_n I_n \right\}.$$

Proposition. *μ^* is an outer measure on 2^X.*

Proof. We must check condition (2). Let $A = \cup_{n=1}^{\infty} A_n$. Given $\epsilon > 0$, by definition of inf, for each n there exists a cover $\{I_{n,k}\}_{k=1}^{\infty}$ of A_n by sets in $\mathcal{F}$ such that $\sum_{k=1}^{\infty} \mu(I_{n,k}) \leq \mu^*(A_n) + \frac{\epsilon}{2^{n+1}}$. Then $\{\cup_{n,k} I_{n,k}\}$ covers A and so

$$\mu^*(A) \leq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mu(I_{n,k}) \leq \sum_{n=1}^{\infty} \mu^*(A_n) + \sum_{n=1}^{\infty} \frac{\epsilon}{2^{n+1}} = \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon.$$

Since this is true for all $\epsilon > 0$, $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$, as required. $\square$

It is easy to see that $\mu^*(A) = \mu(A)$ if $A \in \mathcal{F}$. In the case $\mathcal{F} = \{\text{Fundamental Sets}\}$ with its standard content, it is also trivial that $\mu^*(A) = \mu^*(A + x)$ for any $A \in 2^{\mathbb{R}}$ and $x \in \mathbb{R}$.

Definition. A subset $E \subset X$ will be called *measurable* (with respect to the outer measure μ^*) if $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$ for all sets $A \subset X$.

The collection of measurable sets will be denoted $\mathcal{M}$.

Since in general $\mu^*(A) = \mu^*\big((A \cap E) \cup (A \cap E^c)\big) \leq \mu^*(A \cap E) + \mu^*(A \cap E^c)$, the issue is whether or not the reverse inequality $\mu^*(A \cap E) + \mu^*(A \cap E^c) \leq \mu^*(A)$ holds for every set $A \subset X$. In other words

Proposition. *A subset $E \subset X$ is measurable if and only if $\mu^*(A \cap E) + \mu^*(A \cap E^c) \leq \mu^*(A)$ for all sets $A \subset X$.* $\square$

Proposition. *$\emptyset \in \mathcal{M}$ and $X \in \mathcal{M}$.*

Proof. Trivial. $\square$

Proposition. *If $E \in \mathcal{M}$ then $E^c \in \mathcal{M}$.*

Proof. Trivial $\square$

Proposition. *If $\mu^*(E) = 0$ then $E \in \mathcal{M}$.*

Proof. Suppose $\mu^*(E) = 0$. Then $\mu^*(B) = 0$ for any $B \subset E$, and in particular $\mu^*(A \cap E) = 0$ for any $A \subset X$. Therefore

$$\mu^*(A) \leq \mu^*(A \cap E) + \mu^*(A \cap E^c) = 0 + \mu^*(A \cap E^c) \leq \mu^*(A)$$

and so each of the inequalities is an equality. Thus $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$. $\square$

Lemma. *If $D, E \in \mathcal{M}$ then $D \cup E \in \mathcal{M}$.*

Proof. Suppose $S \subset X$. Since E is measurable, the definition of measurable (applied with $A = S \cap D^c$) gives

$$\mu^*(S \cap D^c) = \mu^*(S \cap D^c \cap E) + \mu^*(S \cap D^c \cap E^c).$$

However since $D \cup E = D \cup (D^c \cap E)$ we get

$$S \cap (D \cup E) = S \cap \big(D \cup (D^c \cap E)\big) = (S \cap D) \cup (S \cap D^c \cap E)$$

and so

$$\mu^*\big(S \cap (D \cup E)\big) \leq \mu^*(S \cap D) + \mu^*(S \cap D^c \cap E).$$

Therefore, noting that $(D \cup E)^c = D^c \cup E^c$ we get

$$\begin{aligned} \mu^*\big(S \cap (D \cup E)\big) + \mu^*\big(S \cap (D \cup E)^c\big) &\leq \mu^*(S \cap D) + \mu^*(S \cap D^c \cap E) + \mu^*(S \cap D^c \cap E^c) \\ &= \mu^*(S \cap D) + \mu^*(S \cap D^c) = \mu^*(S) \end{aligned}$$

where the final two inequalities make use of the measurability of E and D respectively. Therefore $D \cup E$ is measurable. $\square$

Corollary. *The collection $\mathcal{M}$ of measurable sets forms an algebra of sets in X.* $\square$

Lemma. *If $E_1 \dots E_n$ are disjoint measurable sets then*

$$\mu^*\Big(A \cap (\cup_{i=1}^n E_i)\Big) = \sum_{i=1}^n \mu^*(A \cap E_i)$$

for any $A \subset X$.

Proof. The Lemma clearly holds when $n = 1$. Suppose by induction that the Lemma is known for any collection of $n-1$ disjoint measurable sets. Since the sets $\{E_i\}$ are disjoint, for any $A \subset X$ we have

$$\Big(A \cap (\cup_{i=1}^n E_i)\Big) \cap E_n^c = A \cap \big(\cup_{i=1}^{n-1} E_i\big) \qquad \text{and} \qquad \Big(A \cap (\cup_{i=1}^n E_i)\Big) \cap E_n = A \cap E_n.$$

Therefore the measurability of E_n and the induction hypothesis give

$$\begin{aligned} \mu^*\Big(A \cap (\cup_{i=1}^n E_i)\Big) &= \mu^*\Big(A \cap \big(\cup_{i=1}^{n-1} E_i\big)\Big) + \mu^*(A \cap E_n) \\ &= \sum_{i=1}^{n-1} \mu^*(A \cap E_i) + \mu^*(A \cap E_n) = \sum_{i=1}^{n} \mu^*(A \cap E_i). \end{aligned}$$

□

Corollary. *If $\{E_n\}_{i=1}^\infty$ are disjoint measurable sets then $\mu^*(\cup_{i=1}^\infty E_i) = \sum_{i=1}^\infty \mu^*(E_i)$.*

Proof. Set $B_n := \cup_{i=1}^n E_i$ for $n = 1, \ldots, \infty$. Since μ^* is an outer measure,

$$\mu^*(B_\infty) \leq \sum_{i=1}^\infty \mu^*(E_i).$$

To complete the proof, we must show the reverse inequality, $\sum_{i=1}^\infty \mu^*(E_i) \leq \mu^*(B_\infty)$. If $\mu^*(B_\infty) = \infty$ there is nothing to prove, so suppose $\mu^*(B_\infty) < \infty$. Applying the Lemma with $A := X$ gives $\mu^*(B_n) = \sum_{i=1}^n \mu^*(E_i)$ for each $n < \infty$. For all n we have $B_n \subset B_\infty$ and so $\mu^*(B_n) \leq \mu^*(B_\infty)$. Therefore taking the limit as $n \to \infty$ gives

$$\sum_{i=1}^\infty \mu^*(E_i) \leq \mu^*(B_\infty),$$

as desired. □

Theorem. *The collection $\mathcal{M}$ of measurable sets forms σ-algebra of sets in X.*

Proof. We must check that if E is a countable union of sets in $\mathcal{M}$ then E lies in $\mathcal{M}$. By repeated use of $B \cup C = B \amalg (C - B)$ we can write E as a countable union $E = \amalg_{i=1}^\infty E_i$ of disjoint measurable sets. Set $F_n := \amalg_{i=1}^n E_i$. Thus F_n is measurable for all n. Let $A \subset X$ be any set. Then for any n,

$$\mu^*(A) = \mu^*(A \cap F_n) + \mu^*(A \cap F_n^c) \geq \mu^*(A \cap F_n) + \mu^*(A \cap E^c).$$

Since the preceding Lemma implies $\mu^*(A \cap F_n) = \sum_{i=1}^n \mu^*(A \cap E_i)$ we have

$$\mu^*(A) \geq \sum_{i=1}^n \mu^*(A \cap E_i) + \mu^*(A \cap E^c)$$

for all n. Taking the limit as $n \to \infty$ gives

$$\mu^*(A) \geq \sum_{i=1}^\infty \mu^*(A \cap E_i) + \mu^*(A \cap E^c) \geq \mu^*(A \cap E) + \mu^*(A \cap E^c)$$

and so E is measurable. □

Summarizing the preceding corollary and theorem, we have shown

Theorem. *Let μ^* be an outer measure on 2^X. Then the collection $\mathcal{M}$ of measure sets (with respect to μ^*) forms a σ-algebra of sets in X and the restriction of μ^* to $\mathcal{M}$ forms a measure on $\mathcal{M}$.* □

Proposition. *Let $\mathcal{F}$ be an algebra of sets and let μ be a content on $\mathcal{F}$. Let μ^* be the outer measure associated to $\mathcal{F}$. Any $I \in \mathcal{F}$ is measurable with respect to μ^*. That is, $\mathcal{F} \subset \mathcal{M}$ (where $\mathcal{M}$ denotes the σ-algebra of μ^*-measurable sets).*

Proof. Let A be any set and write $A = A_1 \amalg A_2$ where $A_1 := A \cap I$ and $A_2 := A \cap I^c$. We must show $\mu^*(A_1) + \mu^*(A_2) \leq \mu^*(A)$. If $\mu^*(A) = \infty$ there is nothing to prove, so suppose $\mu^*(A) < \infty$. By definition of inf, given any $\epsilon > 0$ there exists a countable collection $\{I_n\}$ of sets in $\mathcal{F}$ which cover A and for which $\sum_n \mu(I_n) \leq \mu^*(A) + \epsilon$. Write $I_n = I'_n \amalg I''_n$ where $I'_n := I_n \cap I$ and $I''_n := I_n \cap I^c$. Since $A_1 \subset \cup_n I'_n$ we have

$$\mu^*(A_1) \leq \mu^*(\cup_n I'_n) \leq \sum_n \mu(I'_n)$$

and similarly $\mu^*(A_2) \leq \sum_n \mu(I''_n)$. Therefore

$$\mu^*(A_1) + \mu^*(A_2) \leq \sum_n \big(\mu(I'_n) + \mu(I''_n)\big) = \sum_n \mu(I_n) \leq \mu^*(A) + \epsilon.$$

Since this is true for all $\epsilon > 0$, we get $\mu^*(A_1) + \mu^*(A_2) \leq \mu^*(A)$, as desired. □

The measure $\mu : \mathcal{M} \to [0, \infty]$ associated to the content μ (alternatively called the measure extending the content μ) is defined as the restriction of the function μ^* to the collection $\mathcal{M}$ of measurable sets.

Let $\mathcal{F}_\sigma$ denote the collection of subsets of X which can be written as countable unions of sets in $\mathcal{F}$, and let $\mathcal{F}_{\sigma\delta}$ denote the collection of subsets of X which can be written as countable intersections of sets in $\mathcal{F}_\sigma$.

Proposition. *Let $\mathcal{F}$ be the algebra of Fundamental Sets in $\mathbb{R}$ with its standard content. For any measurable set A, there exist sets $B_1, B_2 \in \mathcal{F}_{\sigma\delta}$ such that $B_1 \subset A \subset B_2$ and $\mu(B_1) = \mu(A) = \mu(B_2)$. Furthermore, given any $\epsilon > 0$ there exists a closed set B'_1 and an open set B'_2 such that $B'_1 \subset A \subset B'_2$ and $\mu(B'_2) - \mu(A) < \epsilon$ and $\mu(A) - \mu(B'_1) < \epsilon$.*

Proof. By definition, given any k there exists an open set U_k such that $\mu(U_k) < \mu(A) + 1/k$. In particular, we can let B'_2 equal U_k for any $k > 1/\epsilon$. Set $B_2 := \cap_k U_k$. Then B_2 is a G_δ set and so is an $\mathcal{F}_{\sigma\delta}$ set. Note that $\mu(B_2) \leq \mu(U_k)$ for all k, since $B_2 \subset U_k$. Therefore $\mu(B_2) < \mu(A) + 1/k$ for all k and so $\mu(B_2) \leq \mu(A)$. However $A \subset B_2$ and so $\mu(A) \leq \mu(B)$. Therefore $\mu(A) = \mu(B_2)$. Applying the preceding argument to A^c gives an open set C' such that $\mu(C') - \mu(A^c) < \epsilon$ and a G_δ set C such that $\mu(C) = \mu(A^c)$, or equivalently $\mu(C - A^c) = 0$. Set $B_1 := C^c$ and $B'_1 := C'$. Then $B_1 \subset A$ and $\mu(B_1) = \mu(A)$ since $A - B_1 = C - A^c$ and similarly $\mu(A) - \mu(B'_1) < \epsilon$. Since B_1 is the complement of a G_δ set it is an $\mathcal{F}_{\sigma\delta}$ set and since B'_1 is the complement of an open set it is a closed set. □

The preceding proof in the more general setting yields

Proposition. *Let $\mathcal{F}$ be the algebra of sets and let μ be the measure obtained by extending a content on $\mathcal{F}$. For any measurable set A, there exist sets $B_1, B_2 \in \mathcal{F}_{\sigma\delta}$ such that $B_1 \subset A \subset B_2$ and $\mu(B_1) = \mu(A) = \mu(B_2)$. Furthermore, given any $\epsilon > 0$ there exist sets $B \in \mathcal{F}_\sigma$ such that $B \subset A$ and $\mu(A) - \mu(B) < \epsilon$.* □

For the rest of this section, let $\mathcal{F}$ be the algebra of Fundamental Sets in $\mathbb{R}$ with its standard content.

Since in general, $\mathcal{F} \subset \mathcal{M}$, in our special case we get

Corollary. $\{\text{Borel sets}\} \subset \mathcal{M}$ □

Definition. The *Lebesgue measure* $\mu : \mathcal{M} \to [0, \infty]$ is defined as the measure extending the standard content on the fundamental sets.

If follows from the preceding results that the Lebesgue measure is indeed a measure on the σ-algebra $\mathcal{M}$ and furthermore it is translation invariant.

We next descibe an alternate approach to the construction of the σ-algebra $\mathcal{M}$.

Definition. Suppose $A, B \subset X$. The *symmetric difference* of A and B, denoted $A \mathbin{\Delta} B$, is defined by $A \mathbin{\Delta} B := (A - B) \cup (B - A)$.

Proposition. *Let $\mathcal{F}$ be an algebra of sets and suppose $A \in \mathcal{F}$. Then $B \in \mathcal{F}$ if and only if $A \mathbin{\Delta} B \in \mathcal{F}$.*

Proof. $A - B = A \cap B^c$ and similarly $B - A = B \cap A^c$ and so $A \mathbin{\Delta} B = (A \cap B^c) \amalg (B \cap A^c)$. Since $\mathcal{F}$ is an algebras of sets, this implies that if $B \in \mathcal{F}$ then $A \mathbin{\Delta} B$ in $\mathcal{F}$.

Conversely, we also have $B - A = (A \mathbin{\Delta} B) \cap A^c$ and $A \cup B = A \cup (B - A)$ and $A \cap B = (A \cup B) \cap (A \mathbin{\Delta} B)^c$ and $B = (A \mathbin{\Delta} B) \cup (B - A)$. Therefore if $A \mathbin{\Delta} B \in \mathcal{F}$ then $B - A$, $A \cup B$, $A \cap B$, and $B \in \mathcal{F}$. In particular, $B \in \mathcal{F}$. □

Definition. A *pseudo-metric* on a space X consists of a function $\rho : X \times X \to [0, \infty)$ such that

1) $\rho(x, x) = 0$
2) $\rho(x, y) = \rho(y, x)$ for all $x, y \in X$ — symmetry
3) $\rho(x, z) \le \rho(x, y) + \rho(y, z)$ for all $x, y, z \in X$ — triangle inequality

A pseudo-metric satisfying the stronger condtion

1') $\rho(x, y) = 0$ if and only if $x = y$

is called a *metric*.

Proposition. *Let μ^* be an outer-measure on a space Y. Define $\rho : 2^Y \times 2^Y \to [0, \infty]$ by $\rho(A, B) := \mu^*(A \mathbin{\Delta} B)$. Then ρ satisfies conditions (1)–(3) of the preceding definition (but may fail to be a pseudo-metric since the definition of pseudo-metric specifies that its values must lie in $[0, \infty)$ rather than $[0, \infty]$).*

Proof. Exercise. □

Proposition. *Let E be a subset of $\mathbb{R}$ such that there exists a sequence of measurable sets (E_n) for which $\rho(E, E_n) \to 0$. Then E is measurable.*

Proof. The proof is a variant on the proof that a set of outer measure 0 is measurable.

Given $\epsilon > 0$, there exists n such that $\rho(E, E_n) < \epsilon$. Since $E - E_n \subset E \,\Delta\, E_n$, we get $\mu^*(E - E_n) \leq \mu^*(E \,\Delta\, E_n) = \rho(E, E_n) < \epsilon$. Similarly $\mu^*(E_n - E) < \epsilon$.

Let A be any subset of $\mathbb{R}$. Since E_n is measurable, we have

$$\begin{aligned}
\mu^*(A \cap E) &= \mu^*(A \cap E \cap E_n) + \mu^*(A \cap E \cap E_n^c) \qquad \text{and} \\
\mu^*(A \cap E^c) &= \mu^*(A \cap E^c \cap E_n) + \mu^*(A \cap E^c \cap E_n^c) \qquad \text{and} \\
\mu^*(A) &= \mu^*(A \cap E_n) + \mu^*(A \cap E_n^c).
\end{aligned}$$

Thus using inequalities coming from containments gives

$$\begin{aligned}
\mu^*(A \cap E) + \mu^*(A \cap E^c) &= \mu^*(A \cap E \cap E_n) + \mu^*(A \cap E \cap E_n^c) + \mu^*(A \cap E^c \cap E_n) \\
&\quad + \mu^*(A \cap E^c \cap E_n^c) \\
&= \mu^*(A \cap E \cap E_n) + \mu^*\big(A \cap (E - E_n)\big) + \mu^*\big(A \cap (E_n - E)\big) \\
&\quad + \mu^*(A \cap E^c \cap E_n^c) \\
&= \mu^*(A \cap E_n) + \mu^*(E - E_n) + \mu^*(E_n - E) + \mu^*(A \cap E_n^c) \\
&\leq \mu^*(A \cap E_n) + \epsilon + \epsilon + \mu^*(A \cap E_n^c) = \mu^*(A) + 2\epsilon.
\end{aligned}$$

Since this is true for all $\epsilon > 0$, $\mu^*(A \cap E) + \mu^*(A \cap E^c) = \mu^*(A)$ and so E is measurable. $\square$

Notation: Set

$$\begin{aligned}
\mathcal{F}_{\text{fin}} &:= \{A \in \mathcal{F} \mid \mu(A) < \infty\} \\
\hat{\mathcal{F}} &:= \{A \in 2^{\mathbb{R}} \mid \exists \text{ sequence } (A_n) \textit{ in } \mathcal{F}_{\text{fin}} \text{ such that } \big(\rho(A_n, A)\big) \to 0\}
\end{aligned}$$

Proposition. *A subset E of $\mathbb{R}$ is measurable if and only if it can be written as a countable union of sets in $\hat{\mathcal{F}}$.*

Proof. All elements of $\hat{\mathcal{F}}$ are measurable by the preceding Proposition, and any countable union of measurable sets is measurable. Thus if E can be written as a countable union of sets in $\hat{\mathcal{F}}$ then E is measurable.

Conversely, suppose that E is measurable. Consider first the case where $\mu(E) < \infty$. For each n, by definition of inf there exists a collection of intervals $I_{k,n}$ such that $A \subset \cup_k I_{k,n}$ and $\sum_k \mu(I_{k,n}) \leq \mu(E) + 1/n$. Set $J_n := \cup_k I_{k,n}$. Then

$$\mu(E) \leq \mu(J_n) \leq \sum_k \mu(I_{k,n}) \leq \mu(E) + 1/n.$$

Since $E \subset J_n$, we have $E \,\Delta\, J_n = J_n - E$ and $J_n = E \amalg (E \,\Delta\, J_n)$. Therefore $\mu(J_n) = \mu(E) + \rho(E, J_n)$ and so $\rho(E, J_n) = \mu(J_k) - \mu(E) < 1/n$. Thus $\big(\rho(E, J_n)\big) \to 0$ and $\mu(J_n) < \mu(E) + 1/n < \infty$ so $J_n \in \mathcal{F}_{\text{fin}}$. Hence $E \in \hat{\mathcal{F}}$.

If E is an arbitrary measurable set, then $E = \cup_m E_m$ where $E_m = E \cap (-m.m)$. As the intersection of measurable sets, E_m is measurable and since it satisfies $\mu(E_m) \leq 2m < \infty$ the preceding case applies to E_m giving $E_m \in \hat{\mathcal{F}}$. Thus E is a countable union of sets in $\hat{\mathcal{F}}$. $\square$

Example. Since the rationals are countable with $\mathbb{Q} = \amalg_{r\in\mathbb{Q}}\{r\}$, we get that the rationals are measurable with $\mu(\mathbb{Q}) = \sum_{r\in\mathbb{Q}} \mu(\{r\}) = \sum_{r\in\mathbb{Q}} 0 = 0$. More generally, any countable subset of $\mathbb{R}$ is measurable and has measure 0.

From the definitions, we have

$$\{\text{Borel sets}\} \subset \{\text{Lebesgue measurable sets}\} \subset 2^{\mathbb{R}}.$$

We will show that both of these containments are strict.

Example. The Cantor set C is defined by as the subset of $[0,1]$ obtained as $C = \cap_n F_n$ where $F_1 := [0,1]$ and F_n is formed by deleting the middle third of every interval in F_{n-1}. Thus $F_2 = [0,1/3] \cup [2/3,1]$, $F_3 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$, etc. Equivalently,

$$C = \{x \in [0,1] \mid x \text{ has a ternary (base 3) expansion containing only 0's and 2's (no 1's)}\}.$$

(Note: Recall that expansions of real numbers in base n are not quite unique, since any positive number with a finite expansion also has an infinite expansion. For example, $0.999\ldots$ is an alternate decimal expansion of 1 since by definition $0.999\ldots = 9/10 + 9/10^2 + 9/10^3 + \ldots = (9/10)/(1-1/10) = 1$. Therefore the above description of C includes the endpoints of the intervals.)

Then $[0,1] = C \coprod \big(\amalg_{n=1}^{\infty}(F_{n+1} - F_n)\big)$ so

$$\mu(C) = \mu([0,1] - \sum_{n=1}^{\infty} \mu(F_{n+1} - F_n) = 1 - (1/3 + 2/9 + 4/27 + \ldots) = 1 - \frac{1/3}{1-(2/3)} = 1 - 1 = 0.$$

It follows that $\mu^*(A) \le \mu(C) = 0$ for any subset $A \subset C$ and in particular, and subset of C is measurable. $\square$

It follows from the example that the cardinality of the set of measurable sets is at least as large as the cardinality of the set of subsets of C. A surjection $\phi : C \to [0,1]$ is given by $\phi(x)$ equals the real number whose binary (base 2) expansion is the one obtained by replacing all 2's in the ternary expansion of x by 1's. The preimage under ϕ of a point in $[0,1]$ contains at most two points (the right endpoint of an interval gets mapped to the same place as the left endpoint of the next interval). Thus $\operatorname{Card}(C) = \operatorname{Card}([0,1]) = \operatorname{Card}(\mathbb{R})$ which is customarily denoted c, for "continuum". Thus we get Card(set of measurable sets) $\ge 2^{\operatorname{Card} C} = 2^c$. Since the cardinality of the set of measurable sets can be no more than the cardinality of all subsets of $[0,1]$ which is also 2^c we conclude that the Card(set of measurable sets) $= 2^c$.

Although, as we have just seen, there are lots of Lebesgues measurable sets, not every set is Lebesgue measurable. We now show that, as mentioned earlier, it is not possible to extend the content on the Borel sets to a translation-invariant measure on all of $2^{\mathbb{R}}$.

Theorem. *If μ is a translation-invariant measure defined on $2^{\mathbb{R}}$ such that $\mu([0,1]) < \infty$ then $\mu(A) = 0$ for all A. In particular, Lebesgue measure is not defined on all of $2^{\mathbb{R}}$: there exist sets which are not Lebesgue measurable.*

Proof. Define equivalence relations $\sim$ and $\sim'$ on $\mathbb{R}$ by $x \sim y$ if and only if $x - y \in \mathbb{Q}$ and $x \sim' y$ if and only if $x - y \in \mathbb{Z}$. (The equivalences classes under these relations form the quotient groups $\mathbb{R}/\mathbb{Q}$ and $\mathbb{R}/\mathbb{Z}$ respectively.) The set $[0,1)$ consists of exactly one element from each equivalence class of the relation $\sim'$. Using the axiom of choice, pick a subset S of $[0,1)$ which contains exactly one element from each equivalence class of the relation $\sim$. Suppose that S is measurable in the measure μ. For $s \in S$ and $x \in [0,1)$, define $s\hat{+}x \in [0,1)$ by

$$s\hat{+}x := \text{ the unique element of } [0,1) \text{ which is equivalent to } s+x \text{ under } \sim' .$$

(This operation is essentially the group operation in $\mathbb{R}/\mathbb{Z}$.) Explicitly,

$$s\hat{+}x = \begin{cases} s+x & \text{if } s+x<1; \\ s+x-1 & \text{if } s+x \geq 1. \end{cases}$$

For $x \in [0,1)$ set $S\hat{+}x := \{s\hat{+}x \mid s \in S\} \subset [0,1)$. Since μ is translation invariant, $\mu(S\hat{+}x) = \mu(S)$ for any x.

Lemma. *For any $x \in [0,1)$, there exists unique $q \in \mathbb{Q} \cap [0,1)$ such that $x \in S\hat{+}q$.*

Proof. Let s be the unique element of S such that $x \sim s$. Then by definition, $s - x \in \mathbb{Q}$. Set

$$q := \begin{cases} s-x & \text{if } x \leq s; \\ s-x+1 & \text{if } x > s. \end{cases}$$

Then $q \in \mathbb{Q} \cap [0,1)$ and $x \in S\hat{+}q$. If $q, q' \in \mathbb{Q} \cap [0,1)$ satisfy both $x \in S\hat{+}q$ and $x \in S\hat{+}q'$ then $s = x + q + \epsilon_1$ and $s = x + q + \epsilon_2$ where $\epsilon_1 = 0$ or 1 and $\epsilon_2 = 0$ or 1. Thus $q - q' \in \mathbb{Z}$ and since both q and q' lie in $[0,1)$, this implies $q = q'$.

Proof of Theorem (cont.) According to the Lemma, $[0,1) = \amalg_{q\in\left(\mathbb{Q}\cap[0,1)\right)} S\hat{+}q$. Since the rationals are countable, the definition of measure gives

$$\mu([0,1)) = \sum_{q\in\left(\mathbb{Q}\cap[0,1)\right)} \mu(S\hat{+}q) = \sum_{q\in\left(\mathbb{Q}\cap[0,1)\right)} \mu(S).$$

If $\mu(S) > 0$ this gives $\mu([0,1)) = \infty$ contradicting the hypothesis and therefore $\mu(S) = 0$ in which case it gives $\mu([0,1)) = 0$. It follows from translation invariance that $\mu(I) = 0$ for any half-open interval and since any set can be covered by countably many half-open intervals, we get $\mu(A) = 0$ for any set A. $\square$

Next we show that that there are measurable sets which are not Borel sets.

For each ordinal γ, inductively define a subset $\mathcal{B}_\gamma \subset 2^{\mathbb{R}}$ as follows. To begin, set $\mathcal{B}_0 := \mathcal{F}$, the fundamental sets. Having defined $\mathcal{B}_\beta$ for all ordinals $\beta < \gamma$, define

$$\mathcal{B}_\gamma := \{S \subset \mathbb{R} \mid S = \cup_{n=1}^{\infty} A_n \text{ where for each } n, \exists\, \beta < n \text{ such that either } A_n \in \mathcal{B}_\beta \text{ or } A_n^c \in \mathcal{B}_\beta\}$$

Thus $\mathcal{B}_1 = F_\sigma$, $\mathcal{B}_2 = F_{\sigma\delta\sigma}$, etc., but the process extends beyond the finite ordinals. Since the collection of Borel sets is closed under taking complements and countable unions, induction implies that if $A \in B_\gamma$ then A is a Borel set. Set $\mathcal{B} = \cup_{\gamma \in \{\text{countable ordinals}\}} B_\gamma$. Then $\mathcal{B}$ forms a σ-algebra all of whose elements are Borel sets, and so by definition $\mathcal{B} = \{\text{Borel sets}\}$.

An open interval in $\mathbb{R}$ is described by a pair of real numbers (giving its centre and length), so $\mathrm{Card}(\{\text{open invervals}\}) = c$. Fundamental sets can be formed from a choice of countably many open intervals, so $\mathrm{Card}(\mathcal{F}) = \aleph_0 \times c = c$, where $\aleph_0 = \mathrm{Card}(\mathbb{N})$. By induction we see that $\mathrm{Card}(\mathcal{B}_\gamma) = c$ for each γ, and thus $\mathrm{Card}(\mathcal{B}) = c$. However we noted earlier that $\mathrm{Card}(\mathcal{M}) = 2^c$, so it follows from the following Lemma that there are more Lebesgue measurable sets than just the Borel sets.

Lemma. *For any set X, $\mathrm{Card}(X) < \mathrm{Card}(2^X)$.*

Proof. Exercise. □

Applying our procedure to the algebra of sets in $\mathbb{R}^2$ based on rectangles rather then intervals allows us to define Lebesgue measure on $\mathbb{R}^2$, and even more generally we can define Lebesgue measure on $\mathbb{R}^n$ in a similar fashion using the volume of generalized rectangles, $\mu\big(\prod_{i=1}^n [a_i, b_i]\big) := \prod_{i=1}^n (b_i - a_i)$, as a starting point.

3. Measure Spaces

Definition. A *measure space* $(X, \mathcal{M}, \mu)$ consists of a set X, together with a σ-algebra of sets in X and a measure μ on $\mathcal{M}$. The elements of $\mathcal{M}$ are called *measurable sets* (in the measure space $(X, \mathcal{M}, \mu)$).

We sometimes refer to the "measure space X" where strictly speaking we ought to write "the measure space $(X, \mathcal{M}, \mu)$".

Examples.
1) X=any; $\mathcal{M}$ =any; $\mu(A) = 0$ for all $A \in \mathcal{M}$.
2) X=any; $\mathcal{M}$ =any; $\mu(A) = \{\, 0$ if A is finite; ∞ if A is infinite.
3) X=any; $\mathcal{M} = 2^X$; $\mu(A) = \begin{cases} \mathrm{Card}(A) & \text{if } A \text{ is finite;} \\ \infty & \text{if } A \text{ is infinite.} \end{cases}$
4) $X = \mathbb{R}^n$; $\mathcal{M} = \{\text{Lebesgue measurable sets}\}$; $\mu = \{\text{Lebesgue measure}\}$.

Notation. If $\{A_n\}_{n=1}^\infty$ is a collection of subsets of X, we write $\{A_n\} \nearrow$ to mean that $A_n \subset A_{n+1}$ for all n, and $\{A_n\} \searrow$ to mean that $A_n \supset A_{n+1}$ for all n. We write $\{A_n\} \nearrow A$ to mean $\{A_n\} \nearrow$ with $A = \cup_n A_n$ and $\{A_n\} \searrow A$ to mean $\{A_n\} \nearrow$ with $A = \cap_n A_n$. Similarly if $\{a_n\}_{n=1}^\infty$ is a collection of nonnegative real numbers we write $\{a_n\} \nearrow a$ to mean that that $a_n \le a_{n+1}$ for all n, and $a = \lim_{n\to\infty} a_n$ and if $\{f_n\}_{n=1}^\infty$ is a collection of nonnegative real-valued functions on X we write $\{f_n\} \nearrow f$ to mean that that $f_n(x) \nearrow f_{n+1}(x)$ for all $x \in X$.

Proposition.
1) *If $A = \cup_{n=1}^\infty A_n$ and A_n is measurable for all n then A is measurable and $\mu(A) \le \sum_{n=1}^\infty \mu(A_n)$.*
2) *If $A = \cup_{n=1}^\infty A_n$ and A_n is measurable for all n and $\{A_n\} \nearrow$, then A is measurable and $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*

3) If $A = \cap_{n=1}^{\infty} A_n$ and A_n is measurable for all n and $\{A_n\} \searrow$, then A is measurable and if $\mu(A_N)$ is finite for some N then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

Proof. In parts (1) and (2), the measurability of A comes from fact that according to the definition of a σ-algebra, the collection of measurable sets is closed under countable unions, For part (3) we must also use that σ-algebras are closed under complementation.

1) Set $B_n := A_n - \cup_{i=1}^{n-1} A_i$. Then $A = \amalg_{n=1}^{\infty} B_n$ and $B_n \subset A_n$ so $\mu(A) = \sum_{n=1}^{\infty} B_n \leq \sum_{n=1}^{\infty} \mu(A_n)$.
2) If $\mu(A_n) = \infty$ for some n then $\mu(A) = \infty = \lim_{n\to\infty} \mu(A_n)$. Therefore assume that $\mu(A_n) < \infty$ for all n. Write $A = \cup_{n=1}^{\infty} A_n = \amalg_{n=1}^{\infty} (A_n - A_{n-1})$, where by convention we set $A_0 := \emptyset$. Thus

$$\begin{aligned}\mu(A) &= \sum_{n=1}^{\infty} \mu(A_n - A_{n-1}) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k - A_{k-1}) = \lim_{n\to\infty} \sum_{k=1}^{n} \big(\mu(A_k) - \mu(A_{k-1})\big) \\ &= \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

3) When computing the limit, it suffices to restrict attention to those terms with $n \geq N$. For those terms we have

$$\begin{aligned}\mu(A_N) - \mu(A) &= \mu(A_N - \cap_n A_n) = \mu\big(A_N \cap (\cup_n A_n^c)\big) = \mu\big(\cup_n (A_N \cap A_n^c)\big) \\ &= \lim_{n\to\infty} \mu(A_N \cap A_n^c)\end{aligned}$$

by part (2). Since

$$\mu(A_N \cap A_n^c) = \mu(A_N - A_n) = \mu(A_N) - \mu(A_n)$$

we get

$$\mu(A_N) - \mu(A) = \lim_{n\to\infty} \big(\mu(A_N) - \mu(A_n)\big) = \mu(A_N) - \lim_{n\to\infty} \mu(A_n).$$

Therefore $\mu(A) = \lim_{n\to\infty} \mu(A_n)$. ☐

Notation. Let $f : S \to T$ be a function and suppose $B \subset T$. Then $f^{-1}(B) := \{s \in S \mid f(s) \in T\}$, called the inverse image of B under f.

Note: This notation does not require that the function f be invertible, although it agrees with the definition of $h(T)$ in the case where f is invertible and $h = f^{-1}$ is its inverse.

Observe that

$$\begin{aligned}f^{-1}(\cup_\alpha B_\alpha) &= \cup_\alpha B_\alpha f^{-1}(B_\alpha) \\ f^{-1}(\cap_\alpha B_\alpha) &= \cap_\alpha B_\alpha f^{-1}(B_\alpha) \\ f^{-1}(B^c) &= \big(f^{-1}(B)\big)^c \\ (g \circ f)^{-1}(B) &= f^{-1}\big(g^{-1}(B)\big)\end{aligned}$$

The definition of continuity for a function $f : X \to Y$ between topological spaces specifies that f is continuous if and only if $f^{-1}(U)$ is open in X for every open subset $U \subset Y$. As shown in `MATA37` and `MATB43` this agrees with the ϵ-δ definition in the case of Euclidean spaces.

Definition. Let X be a measure space and let Y be a topological space. A function $f : X \to Y$ is called *measurable* if $f^{-1}(U)$ is measurable for every open subset U of Y.

In the case $Y = \mathbb{R}$, since every open set is a countable union of open intervals it follows from the properties of f^{-1} that it suffices to check that the inverse image of intervals is measurable. In fact it is easy to see:

Proposition. *Let $f : X \to \mathbb{R}$. Then the following are equivalent:*
1) *f is measurable.*
2) *$f^{-1}(-\infty, a)$ is measurable for all $a \in \mathbb{R}$.*
3) *$f^{-1}(-\infty, a]$ is measurable for all $a \in \mathbb{R}$.*
4) *$f^{-1}(a, \infty)$ is measurable for all $a \in \mathbb{R}$.*
5) *$f^{-1}[a, \infty)$ is measurable for all $a \in \mathbb{R}$.* $\square$

In measure theory, generally speaking, sets of measure zero can be ignored in the sense that they tend not to affect the theorems or the calculations. It is therefore convenient to use the term that a property holds "almost everywhere" (sometimes abbreviated to a.e.) to mean that the set on which it fails to hold has measure zero. For example we might say that "$f = g$ almost everywhere", meaning that $\mu\big(\{x \mid f(x) \neq g(x)\}\big) = 0$.

Proposition. *Let X be a measure space and let $f, g : X \to Y$. Suppose f is measurable and $f = g$ almost everywhere. Then g is measurable.*

Proof. Set $E := \{x \in X \mid f(x) \neq g(x)\}$. Then for any subset $B \subset Y$, any deviation between $f^{-1}(B)$ and $g^{-1}(B)$ takes place within the set E. More precisely, the symmetric difference $\big(f^{-1}(B)\big) \,\Delta \big(g^{-1}(B)\big)$ is contained in E. Since E has measure 0, it follows that $\big(f^{-1}(B)\big) \,\Delta \big(g^{-1}(B)\big)$ is measurable for any $B \subset Y$. Therefore if $U \subset Y$ is open then both $f^{-1}(U)$ and $f^{-1}(U) \,\Delta\, g^{-1}(U)$ are measurable, and so it follows from an earlier proposition that $g^{-1}(U)$ is measurable. $\square$

Proposition. *Let X be a topological space and let $f : X \to \mathbb{R}$ be continuous. Suppose $\mathcal{M}$ is a σ-algebra on X which contains the open sets of X and let μ be a measure on $\mathcal{M}$. Then f is measurable. In particular, any continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue measurable.*

Proof. For any open $U \subset \mathbb{R}$, by definition of continuity, $f^{-1}(U)$ is open in X and thus by hypothesis it is measurable in the measure space $(X, \mathcal{M}, \mu)$. $\square$

Proposition. *Let $f : X \to \mathbb{R}$ be measurable and let $g : \mathbb{R} \to \mathbb{R}$ be continuous. Then $g \circ f : X \to \mathbb{R}$ is measurable.*

Proof. For any open $U \subset \mathbb{R}$, $f^{-1}(U)$ is measurable by the preceding Proposition and thus $(g \circ f)^{-1}(B) = f^{-1}\big(g^{-1}(B)\big)$ is measurable by measurability of f. $\square$

Proposition. *Let $f, g : X \to \mathbb{R}$ be measurable. Then*
1) *cf is measurable for any constant $c \in \mathbb{R}$.*
2) *$f + g$ is measurable.*
3) *$f - g$ is measurable.*
4) *fg is measurable.*
5) *f/g is measurable provided $\{g \mid g(x) = 0\}$ has measure 0.*

6) $|f|$ *is measurable.*
7) $\max\{f, g\}$ *is measurable.*
8) $\min\{f, g\}$ *is measurable.*

Proof.

1) If $c = 0$ then $cf \equiv 0$ which is clearly measurable. If $c > 0$ then $(cf)^{-1}(-\infty, a) = f^{-1}(-\infty, a/c)$ while if $c \leq 0$ then $(cf)^{-1}(-\infty, a) = f^{-1}(a/c, \infty)$.
2) If $x \in (f+g)^{-1}(-\infty, a)$ then $f(x)+g(x) < a$ so there exists a rational number r such that $f(x)+r < a-g(x)$. Thus $(f+g)^{-1}(-\infty, a) = \cup_{r\in\mathbb{Q}}\big(f^{-1}(-\infty, a)\cap g^{-1}(-\infty, a-r)\big)$ and so $f+g$ is measurable.
3) Since g is measurable, $-g$ is measurable by (1) and so $f - g = f + (-g)$ is measurable by (2).
4) Let $h(x) = x^2$. Since h is continuous and f is measurable, an earlier Proposition implies $f^2 = h\circ f$ is measurable. Similarly g^2 is measurable and $(f+g)^2$ is measurable. Therefore $fg = \big((f+g)^2 - f^2 - g^2\big)/2$ is measurable.
5) Define $\tilde{g} : X \to \mathbb{R}$ by

$$\tilde{g} := \begin{cases} g(x) & \text{if } g(x) \neq 0; \\ 1 & \text{if } g(x) = 0. \end{cases}$$

   Then $\tilde{g} = g$ almost everywhere so the measurability of g implies that $\tilde{g}$ is measurable, and similarly $f/g = f/\tilde{g}$ almost everywhere, so to show f/g measurable, it suffices to prove that $f/\tilde{g}$ is measurable. Let $h(x) = 1/x$. If U is any open subset of $\mathbb{R}$ then $h^{-1}(U)$ is an open subset of $\mathbb{R} - \{0\}$ by continuity of h. Since $\tilde{g}$ is never zero, $\left(\frac{1}{\tilde{g}}\right)^{-1}(U) = \tilde{g}^{-1}\big(h^{-1}(U)\big)$ which is therefore measurable since $\tilde{g}$ is measurable. Hence $1/\tilde{g}$ is measurable and so the product $f/\tilde{g} = f \times 1/\tilde{g}$ is measurable by part (4) and therefore f/g is measurable.
6) If $a < 0$ then $|f|^{-1}(a, \infty) = X$ which is measureable. If $a \geq 0$ then $|f|^{-1}(a, \infty) = f^{-1}(-\infty, -a) \cap f^{-1}(a, \infty)$, which is measurable. Therefore $|f|$ is measurable.
7) $\max\{f, g\} = (f + g + |f - g|)/2$.
8) $\min\{f, g\} = (f + g - |f - g|)/2$. □

Note: The converse of part (6) fails: it is possible for $|f|$ to be measurable even if f is not measurable. For example, let S be a non-measurable subset of $\mathbb{R}$ and define

$$f(x) = \begin{cases} 1 & \text{if } x \in S; \\ -1 & \text{if } x \notin S. \end{cases}$$

Then f is not measurable since $f^{-1}(0, \infty) = S$ is not measurable. However $|f| \equiv 1$ is measurable.

It is often convenient to work with functions which are never negative. For this reason we introduce the following notation which makes it convenient for us to write an arbitrary function as a difference of nonnegative functions.

Notation. For $f : X \to \mathbb{R}$, set $f^+ := \max\{f, 0\}$ and $f^- = -\min\{f, 0\}$, called the positive and negative parts of f respectively.

The definitions imply that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. It follows from the preceding Proposition that f is measurable if and only if f^+ and f^- are measurable.

Using induction, part (7) of the preceding Proposition implies that the maximum (and minimum) of any finite set of functions is measurable. This generalizes to infinite sets as follows.

Proposition. *Let $\{f_n\}_{n=1}^{\infty}$ be a set of measurable real-valued functions, $f_n : X \to \mathbb{R}$, and suppose that $\{f_n(x)\}_{n=1}^{\infty}$ is bounded above almost everywhere on X. (That is, the set of $x \in X$ for which it fails to be bounded above has measure 0.) Let $f : X \to \mathbb{R}$ by $f(x) := \sup\{f_n(x)\}$ except possibly on a set of measure 0. Then f is measurable. Similarly a function formed as the pointwise infimum of countably many measurable functions whose values are pointwise bounded below is measurable.*

Remark. Note that $f(x)$ makes sense since the axioms for $\mathbb{R}$ state that every bounded set has a (finite) supremum.

Proof of Prop. Ignoring a set of measure zero where the value of f has no effect on its measurability, $f^{-1}(a, \infty) = \cup_{n=1}^{\infty} f_n^{-1}(a, \infty)$ for any $a \in \mathbb{R}$, so f is measurable. The proof for $\inf\{f_n\}$ is similar. $\square$

Suppose (a_n) is a sequence in $\mathbb{R}$ with $\{a_n\}$. For each n, let $s_n = \sup\{a_k\}_{k=n}^{\infty}$. Then the sequence (s_n) of supremums is monotonically decreasing since each set is a subset of the preceding one. Any upper bound for $\{a_n\}$ is an upper bound for $\{s_n\}$ and any lower bound for $\{a_n\}$ is a lower bound for $\{s_n\}$. Therefore if the sequence $\{a_n\}$ is bounded both above and below the limit $\lim_{n\to\infty} s_n$ will always exist. We use the notation $\limsup(a_n) := \lim_{n\to\infty} \sup\{a_k\}_{k\geq n}$. Alternatively the notation $\overline{\lim}(a_n)$ is sometimes used for $\limsup(a_n)$. Similarly we define $\underline{\lim}(a_n) := \liminf(a_n)$ to be the limit of the sequence of infimums $\lim_{n\to\infty}\bigl(\inf\{a_k\}_{k\geq n}\bigr)$. It is easy to see that if $\lim_{n\to\infty}(a_n)$ exists then $\limsup(a_n)$ and $\liminf(a_n)$ both exist and equal $\lim_{n\to\infty}(a_n)$. Conversely, if $\limsup(a_n)$ and $\liminf(a_n)$ both exist and are equal then $\lim_{n\to\infty}(a_n)$ exists and equals the common value $\limsup(a_n) = \liminf(a_n)$.

Remark. Note that it is possible to have $\limsup(a_n) > -\infty$ even if $\{a_n\}$ is not bounded below. For example, the $\limsup$ of the sequence $0, -1, 0, -2, 0, -3, 0, -4, \ldots$ is 0.

Proposition. *Let $\{f_n\}_{n=1}^{\infty}$ be a set of measurable real-valued functions, $f_n : X \to \mathbb{R}$, and suppose that $\limsup\bigl(f_n(x)\bigr)$ exists for almost all $x \in X$, (i.e. except possibly for a set of measure 0). Define $f : X \to \mathbb{R}$ by $f(x) := \limsup\bigl(f_n(x)\bigr)$ (extended arbitrary to the set of measure 0 on which the $\limsup$ does not exist). Then f is measurable. Similarly $\liminf\bigl(f_n(x)\bigr)$ is measurable provided the pointwise limits exist.*

Proof. For each x, the sequence $\sup\bigl(\{f_k(x)\}_{k\geq n}\bigr)$ is monotonically decreasing, and so $\limsup\bigl(f_n(x)\bigr)$ can be rewritten as $\limsup\bigl(f_n(x)\bigr) = \inf\sup\bigl(\{f_k(x)\}_{k\geq n}\bigr)$. Therefore the measurability of f follows from two applications of the preceding Proposition. $\square$

Corollary. Let $\{f_n\}_{n=1}^{\infty}$ be a set of measurable real-valued functions, $f_n : X \to \mathbb{R}$, and suppose that $\lim_{n\to\infty}\bigl(f_n(x)\bigr)$ exists for almost all $x \in X$, (i.e. except possibly for a set of measure 0). Define $f : X \to \mathbb{R}$ by $f(x) := \lim_{n\to\infty}\bigl(f_n(x)\bigr)$ (extended arbitrary to the set of measure 0 on which the limit does not exist). Then f is measurable.

Proof. For any x for which the limit exists, it equals $\limsup\bigl(f_n(x)\bigr)$. $\square$

4. Integration

Definition. Suppose $A \subset X$. Define the *characteristic function* of A, denoted $\chi_A : X \to \mathbb{R}$ by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A; \\ 0 & \text{if } x \notin A. \end{cases}$$

.

It is clear from the definitions that the set A is measurable if and only if the function χ_A is measurable.

Any function of the form $f(x) = \sum_{i=1}^{n} c_i \, \chi_{A_i}(x)$ where $c_1, \ldots, c_n \in \mathbb{R}$ are constants and $A_1, \ldots, A_n$ are pairwise disjoint measurable sets is called a *simple function*. This concept generalizes the notion of "step function" which is the special case where all the sets A_i are intervals. Whereas Riemann integration is based on the area underneath step functions, Lebesgue integration is a generalization intuitively based on the area underneath simple functions. More precisely it is defined as follows.

If $f = \sum_{i=1}^{n} c_i \, \chi_{A_i} : X \to \mathbb{R}$ is a simple function and μ is a measure on X, define the *integral* of f with respect to the measure μ, denoted $\int f \, d\mu$, by $\int f \, d\mu := \sum_{i=1}^{n} c_i \mu(A_i)$.

Next if $f : X \to [0, \infty)$ is a measurable nonnegative real-valued function, define

$$\int f \, d\mu := \sup \left\{ \int s \, d\mu \mid s \text{ is a simple function with } s(x) \le f(x) \text{ for all } x \right\}.$$

We say that f is *integrable* with respect to μ if $\int f \, d\mu < \infty$.

Finally for arbitrary measurable $f : X \to \mathbb{R}$, write $f = f^+ - f^-$ where $f^+, f^- : X \to [0, \infty)$ and declare f to be integrable (w.r.t. μ) if and only if f^+ and f^- are integrable, in which case we define $\int f \, d\mu := \int f^+ \, du - \int f^- \, du$.

If $A \subset X$ is measurable, we set $\int_A f \, d\mu := \int \chi_A f \, d\mu$. Thus $\int_X f \, d\mu$ is the same as $\int f \, d\mu$.

For functions $f, g : X \to \mathbb{R}$, we write $f \le g$ to mean that $f(x) \le g(x)$ for all $x \in X$.

If g is integrable and f is measurable with $0 \le f \le g$ then it is immediate from the definitions that f is integrable with $\int f \, d\mu \le \int g \, d\mu$.

Proposition. *Let X be a compact subset of $\mathbb{R}^n$ (or more generally a compact Hausdorff space with $\mu(X) < \infty$) and let $f : X \to \mathbb{R}$ be continuous. Then f is integrable.*

Proof. We showed earlier that if f is continuous it is measurable. Since X is compact and f is continuous there exists a constant function M such that $f \le M$. We also know $\mu(X) < \infty$ since X is compact and thus bounded. Therefore M is integrable (with $\int M \, d\mu = M\mu(X) < \infty$) and so f is integrable. □

Theorem. *Let $f : X \to \mathbb{R}$ be measurable. Then there exists a sequence (s_n) of simple functions converging pointwise to f, i.e. $\lim_{n\to\infty} s_n(x) = f(x)$ for all $x \in X$.*

Proof. Given a positive integer n, partition $[-n, n]$ into $2n^2$ subintervals of length $1/n$ and for each $k = 0, \ldots, 2n^2 - 1$, write $I_k := \left[-n + \frac{k}{n}, -n + \frac{k+1}{n}\right]$ for the kth subinterval. Let $E_k := f^{-1}(I_k)$, which is a measurable set since f is measurable. Set $s_n := \sum_{k=0}^{2n^2-1} \left(-n + \frac{k}{n}\right) \chi_{E_k}$. Then by construction $f(x) - s_n(x) \le 1/n$ for all $x \in [-n, n]$ and so $\lim_{n\to\infty} s_n(x) = f(x)$ for all $x \in \mathbb{R}$. □

Notice that if $f(x) \geq 0$ on the interval $[-2n, 2n]$ then $s_n(x) \leq s_{2n}(x)$ for all $x \in [-2n, 2n]$ (where $s_q(x)$ is the simple function defined in the preceding proof). Thus we get

Corollary. *Let $f : X \to [0, \infty)$ be a nonnegative-valued measurable function. Then there exists a sequence (s_n) of simple functions such that $s_n \nearrow f$.* □

Lebesgue Monotone Convergence Theorem. *If (f_n) is a sequence of measurable non-negative real-valued functions such that $(f_n) \nearrow f$, then $(\int f_n\, d\mu) \nearrow \int f\, du$. In particular, if $\lim_{n\to\infty} f_n$ exists then f is integrable.*

Proof. The fact that the hypotheses imply that f is measurable is the statement of an earlier Proposition and since $f_n \leq f$, each f_n is integrable. We must show that $(\int f_n\, d\mu) \nearrow \int f\, du$.

Lemma. *Let $s : X \to \mathbb{R}$ be a simple function and suppose $\{E_n\} \nearrow X$. Then $(\int_{E_n} s\, du) \nearrow \int s\, d\mu$.*

Proof. Write $s = \sum_{i=1}^k c_i\, \chi_{F_i}$. Then $s\, \chi_{E_n} = \sum_{i=1}^k c_i\, \chi_{F_i \cap E_n}$. Therefore

$$\int_{E_n} s\, du = \sum_{i=1}^k c_i \int \chi_{F_i \cap E_n} = \sum_{i=1}^k c_i \mu(F_i \cap E_n).$$

However $\{F_i \cap E_n\} \nearrow F_i$ and so $\mu(F_i \cap E_n) \nearrow \mu(F_i)$. Thus

$$\left(\sum_{i=1}^k c_i \mu(F_i \cap E_n) \right) \nearrow \sum_{i=1}^k c_i \mu(F_i) = \int s\, du.$$

Proof of Theorem. (cont.)

Let $L := \lim_{n\to\infty} \int f_n\, d\mu$, with possibly $L = \infty$ in case the limit diverges. Since $f_n \leq f$, $\int f_n\, d\mu \leq \int f\, d\mu$ for all n and so $L \leq \int f\, d\mu$. By definition $\int f\, d\mu = \sup\{\int s\, d\mu \mid s \text{ simple}, 0 \leq s \leq f\}$. Suppose $0 \leq s \leq f$ with s simple. Pick some number $r \in (0, 1)$ and set $E_n := \{x \mid f_n(x) \geq rs(x)\}$. Then $\{E_n\} \nearrow X$ and

$$\int f_n\, d\mu \geq \int_{E_n} f_n\, d\mu \geq r \int_{E_n} s\, d\mu.$$

Taking the limit as $n \to \infty$ (and applying the Lemma) gives $L \geq r \int s\, d\mu$. This is true for all $r \in (0, 1)$, so taking the limit as $r \to 1$ gives $L \geq \int s\, d\mu$. Since this inequality holds for all simple functions s with $0 \leq s \leq f$, it follows that $L \geq \int f\, du$. □

Proposition. *Let f and g be integrable and let $c \in \mathbb{R}$ be a constant. Then*

1) $\int_{E_1 \amalg E_2} f\, d\mu = \int_{E_1} f\, d\mu + \int_{E_2} f\, d\mu$.

2) $f + g$ is integrable with $\int (f + g)\, d\mu = \int f\, d\mu + \int g\, d\mu$.

3) cf is integrable with $\int cf\, d\mu = c \int f\, d\mu$.

Proof. Part (3) is trivial.

Part (1) is trivial when f is a simple function and the general case follows from this by writing f as a limit of simple functions.

If $f, g \geq 0$ choose sequences of simple functions such that $(s_n) \nearrow f$ and $(t_n) \nearrow g$. Then $(s_n + t_n) \nearrow f + g$ so by the Lebesgue Monotone Convergence Theorem we get $\lim_{n\to\infty} \int s_n\, d\mu \nearrow \int f\, d\mu$, $\lim_{n\to\infty} \int t_n\, d\mu \nearrow \int g\, d\mu$, and $\lim_{n\to\infty} \int (s_n + t_n)\, d\mu \nearrow \int (f + g)\, d\mu$, so (2) follows in this case.

Similarly, since $\int -h\, d\mu = -\int h\, du$, applying the preceding case to $-f$, $-g$ shows that (2) holds when $f, g \leq 0$.

If $f \geq 0$ and $g \leq 0$, write $X = A \amalg B$ where $A = \{x \in X \mid |f(x) \geq g(x)|\}$ and $B = \{x \in X \mid |f(x) \leq g(x)|\}$. On A we have $f = (f + g) + (-g)$, expressing f as a sum of positive functions. Therefore our previous case shows that $\int_A f\, du = \int_A (f + g)\, du + \int_A (-g)\, du = \int_A (f + g)\, du - \int_A g\, du$ and so $\int_A (f + g)\, d\mu = \int_A f\, d\mu + \int_A g\, d\mu$. Similarly $\int_B (f + g)\, d\mu = \int_B f\, d\mu + \int_B g\, d\mu$. Since according to part (1), $\int h\, d\mu = \int_A h\, d\mu + \int_B h\, d\mu$ for any h, part (2) follows whenever $f \geq 0$ and $g \leq 0$. Similarly, part (2) holds in the case ≤ 0, $g \geq 0$.

Finally for arbitrary f and g, write X as the union of the four disjoint subintervals A_+^+, A_-^+, A_+^-, A_-^-, where

$$\begin{aligned}
A_+^+ &= \{x \in X \mid f(x) \geq 0, g(x) \geq 0\} \\
A_+^- &= \{x \in X \mid f(x) \geq 0, g(x) < 0\} \\
A_-^+ &= \{x \in X \mid f(x) < 0, g(x) \geq 0\} \\
A_-^- &= \{x \in X \mid f(x) < 0, g(x) < 0\},
\end{aligned}$$

and apply the preceding special cases, using that $\int h\, d\mu = \int_{A_+^+} h\, d\mu + \int_{A_+^-} h\, d\mu + \int_{A_-^+} h\, d\mu + \int_{A_-^-} h\, d\mu$. □

Proposition. *Let f be measurable. Then f is integrable if and only if $|f|$ is integrable. If they are integrable then $|\int f\, d\mu| \leq \int |f|\, d\mu|$.*

Proof. If f is integrable then f^+ and f^- are integrable so $|f| = f^+ + f^-$ is integrable. Conversely, if $|f|$ is integrable then $f^+ \leq |f|$ and $f^- \leq |f|$ imply that f^+ and f^- are integrable, and so f is integrable. Assume now that f and $|f|$ are integrable.

$$\left|\int f\, d\mu\right| = \left|\int (f^+ - f^-)\, d\mu\right| \leq \left|\int f^+\, d\mu\right| + \left|\int f^-\, d\mu\right| = \int f^+\, d\mu + \int f^-\, d\mu = \int |f|\, d\mu$$

□

Corollary. *If $\mu(X) < \infty$ and f is a bounded measurable function on X, then f is integrable on X.*

Note: The statement that a function is bounded is defined to mean that there is a bound on its absolute value, and thus means that the function is bounded both above and below.

Proof of Corollary. If M is a bound on $|f|$ then $\int_X |f|\, d\mu \leq M\mu(X)$ and so $|f|$ is integrable. Therefore f is integrable. □

Example. Define $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} n & \text{if } x \in (0, 1/n]; \\ 0 & \text{otherwise.} \end{cases}$$

Then $\int_{[0,1]} f_n \, d\mu = 1$ for all n, but $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in [0,1]$. Therefore

$$\lim_{n\to\infty} \int f_n \, d\mu = 1 \neq 0 = \int \lim_{n\to\infty} f_n \, d\mu.$$

$\square$

The example shows that in general $\lim_{n\to\infty} \int f_n \, d\mu \neq \int \lim_{n\to\infty} f_n \, d\mu$, although the Lebesgue Monotone Convergence Theorem gives some conditions under which equality holds. The Lebesgue Dominated Convergence Theorem gives other hypotheses under which equality holds.

Fatou's Lemma. *Let (f_n) be a sequence of nonnegative real-valued measurable functions. Then $\int \liminf(f_n) \, d\mu \le \liminf(\int f_n \, d\mu)$.*

Proof. Let $F_n(x) = \inf\{f_k(x) \mid k \ge n\}$. Then $(F_n) \nearrow \liminf(f_n)$ by definition of lim inf. Therefore the Lebesgue Monotone Convergence Theorem says that $\int \liminf(f_n) \, d\mu = \lim_{n\to\infty} \int F_n \, d\mu$. However $F_n(x) \le f_n(x)$ for all x and so $\int F_n \, d\mu \le \int f_n \, d\mu$ and therefore $\liminf \int F_n \, d\mu \le \liminf \int f_n \, d\mu$. Since $\int F_n \, d\mu \nearrow$, $\liminf \int F_n \, d\mu$ is the same as $\lim_{n\to\infty} \int F_n, d\mu$ so we have

$$\int \liminf(f_n) \, d\mu = \lim_{n\to\infty} \int F_n \, d\mu = \liminf \int F_n \, d\mu \le \liminf \int f_n \, d\mu.$$

$\square$

Lebesgue Dominated Convergence Theorem. *Let (f_n) be a sequence of measurable functions and suppose that $\big(f_n(x)\big) \to f(x)$ almost everywhere. Suppose there exists an integrable function g such that $|f_n(x)| \le g$ for all n and x. Then $(\int |f - f_n| \, d\mu) \to 0$. In particular, $(\int f_n \, d\mu) \to \int f \, d\mu$.*

Proof. $2g - |f - f_n| \ge 0$ for all n. Fatou's Lemma implies $\int \liminf(2g - |f - f_n|) \, d\mu \le \liminf \big(\int (2g - |f - f_n|) \, d\mu\big)$. However $\liminf(2g - |f - f_n|) = \lim_{n\to\infty}(2g - |f - f_n|) = 2g$. Therefore

$$\begin{aligned} \int 2g \, d\mu = \int \liminf(2g - |f - f_n|) \, d\mu &\le \liminf \left(\int (2g - |f - f_n|) \, d\mu \right) \\ &= \liminf \int 2g \, d\mu + \liminf \int -|f - f_n| \, d\mu \\ &= \int 2g \, d\mu - \limsup \int |f - f_n| \, d\mu \end{aligned}$$

and hence $\limsup \int |f - f_n| \, d\mu = 0$. In general, if (a_n) is a sequence of nonnegative numbers with $\limsup(a_n) = 0$ then $\lim_{n\to\infty} a_n$ exists and equals 0. Thus $(\int |f - f_n| \, d\mu) \to 0$. Since $|\int (f - f_n) \, d\mu| \le (\int |f - f_n| \, d\mu) \to 0$, it follows that $\lim_{n\to\infty} \int (f - f_n) \, d\mu = 0$, or equivalently $(\int f_n \, d\mu) \to \int f \, d\mu$. $\square$

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is Riemann-integrable on an interval $[a, b]$ if the supremum of the Riemann sums from below equals the infimum of the Riemann sums from above, which is equivalent to saying that the supremum of the integrals of the step functions which are pointwise less than or equal to $f(x)$ equals the infimum of the step functions which are pointwise greater than or equal to $f(x)$.

Proposition. *Let $f : \mathbb{R} \to \mathbb{R}$ be Riemann-integrable on $[a, b]$. Then f is Lebesgue integrable on $[a, b]$ and $\int_{[a,b]} f \, d\mu = \int_a^b f(x) \, dx$.*

Proof. Writing $f = f^+ - f^-$, reduces to the problem to the case where $f \geq 0$.

Since f is Riemann-integrable, for all n there exist step functions σ and τ such that $\sigma(x) \leq f(x) \leq \tau(x)$ for all $x \in [a, b]$ and $\int_a^b \big(f(x) - \sigma(x)\big) \, dx < 1/n$. Find a step function σ_1 such that $\sigma_1(x) \leq f(x)$ with $\int_a^b \big(f(x) - \sigma_1(x)\big) \, dx < 1$. Then $f - \sigma_1$ is nonnegative and integrable on $[a, b]$ so there exists a step function σ_2 such that $f - \sigma_1 - \sigma_2 \geq 0$ on $[a, b]$ with $\int_a^b \big(f(x) - \sigma_1(x) - \sigma_2(x)\big) \, dx < 1/2$. Continuing, for each positive integer n choose a step function $\sigma_n(x)$ such that $f - \sigma_1 - \ldots \sigma_n \geq 0$ on $[a, b]$ and

$$\int_a^b \big(f(x) - \sigma_1(x) - \ldots - \sigma_n(x)\big) \, dx < 1/n.$$

Write $\Sigma_n(x) := \sigma_1(x) + \ldots \sigma_n(x)$. Similarly there exists a step function $T_n(x)$ such that Then $f \leq T_n$ on $[a, b]$ and $\int_a^b \big(T_n(x) - f(x)\big) \, dx \leq 1/n$. Then $\int_a^b \big(T_n(x) - \Sigma_n(x)\big) \, dx \leq 2/n$. Therefore $\int_a^b \big(T(x) - \Sigma(x)\big) \, dx = 0$, where $\Sigma(x) := \lim_{n\to\infty} \Sigma_n(x)$ and $T(x) := \lim_{n\to\infty} T_n(x)$. For each $x \in [a, b]$, the sequence $\Sigma_n(x)$ is monotonically increasing and bounded above by $f(x)$ and $T_n(x)$ is monotonically decreasing and bounded below by $f(x)$. Thus each sequence converges with $\Sigma(x) \leq f(x) \leq T(x)$,

Let E be the collection of x's which are the endpoints of some interval in one of the step functions in either Σ_n or T_n for some n. Then E is countable. If $x \notin E$ then x is never an endpoint so there is an interval I_x about x for which $\Sigma_n(y) = \Sigma_n(x)$ for all $y \in I_x$ and thus $\Sigma(y) = \Sigma(x)$ for all $y \in I_x$ and similarly $T(y) = T(x)$ for all $y \in I_y$. Therefore $\int_a^b \big(T(t) - \Sigma(t)\big) \, dt \geq \big(T(x) - \Sigma(x)\big)\mu(I_x)$, which is a contradiction unless $T(x) = \Sigma(x)$, in which case $\Sigma(x) = f(x)$ by the Squeeze Principle. Thus aside from a set of measure zero, $\big(\Sigma_n(x)\big) \nearrow f(x)$. Therefore the Lebesgue monotone convergence theorem shows that $f(x)$ is measurable and $\int_{[a,b]} f d\mu = \lim_{n\to\infty} \Sigma_n(x) \, dx = \int f(x) \, dx$. $\square$

Example. Let $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & \text{if } x \in \mathbb{Q}; \\ 1 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

Then f is not Riemann-integrable since the supremum of the integral of the step functions below f is 0 while the infimum of the step functions above f is 1. However f **is** Lebesgue integrable since $f(x) = 1$ almost everywhere so $\int_{[0,1]} f \, d\mu = \int_{[0,1]} 1 \, d\mu = 1$. Therefore it is possible for a function to be Lebesgue-integrable even if it is not Riemann-integrable.

If we consider improper Riemann integrals, it is also possible to have functions which are Riemann-integrable in this sense which are not Lebesgue integrable. Let

$$f(x) = \begin{cases} (-1)^n 2^{n+1}/n & \text{if } x \in [1-\frac{1}{2^n}, 1-\frac{1}{2^{n+1}}]; \\ 0 & \text{if } x = 0. \end{cases}$$

Then $\int_{[1/2^{k+1},1]} |f|\, d\mu = \sum_{n=0}^{k} \frac{2^{n+1}}{2^{n+1}n} = \sum_{n=0}^{k} \frac{1}{n}$ which diverges as $k \to \infty$. Thus $|f|$ is not integrable on $[0,1]$ and so f is not integrable on $[0,1]$. However the improper integral $\lim_{\epsilon\to 0} \int_\epsilon^1 f(x)\, dx$ exists and equals $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n} = \ln(2)$. □

5. Product Measures

Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. Let $\mathcal{R}$ be the set of "measurable rectangles" in $X \times Y$. That is, $\mathcal{R} := \{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}$. Let $\mathcal{F}$ be the collection of subsets of $X \times Y$ which can be formed as finite disjoint unions of sets in $\mathcal{R}$.

Proposition. *The sets $\mathcal{F}$ form an algebra of sets in $X \times Y$.*

Proof. Clearly $\emptyset \in \mathcal{F}$ and $X \times Y \in \mathcal{F}$. If $E = A \times B$, $F = C \times D$ with $A, C \in \mathcal{A}$ and $B, D \in \mathcal{B}$ then $E \cup F$ can be carved up into a union of disjoint rectangles. Explicitly

$$E \cup F = \big((A - C) \times B\big) \amalg \big((A \cap C) \times (B \cup D)\big) \amalg \big((C - A) \times D\big).$$

It follows using induction that $\mathcal{F}$ is closed under finite unions. $(A \times B)^c = (A^c \times Y) \amalg (A \times B^c)$, so $\mathcal{F}$ is closed under complementation. Therefore $\mathcal{F}$ forms an algebra of sets in $X \times Y$. □

Let $\mu \times \nu$ denote the content on $\mathcal{F}$ determined by setting $(\mu \times \nu)(A \times B) := \mu(A)\nu(B)$ for $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then as in Section 2, we extend $\mu \times \nu$ to a measure on a σ-algebra $\mathcal{M}$ containing $\mathcal{F}$. The measure $\mu \times \nu$ is called the *product measure* of μ and ν. The σ-algebra $\mathcal{M}$ is denotes $\mathcal{A} \times \mathcal{B}$. The resulting measure space $(X \times Y, \mathcal{A} \times \mathcal{B}, \mu \times \nu)$ is called the product of the measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$. In the special case where $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are both $\mathbb{R}$ with its standard (Lebesgue) measure, the product measure is the Lebesgue measure on $\mathbb{R}^2$.

If $E \subset X \times Y$, and $x \in X$, we write E_x for the projection of E onto $\{x\} \times Y \cong Y$. That is $E_x := \{y \in Y \mid (x, y) \in E\}$. Similarly each $y \in Y$ gives a subset $E_y \subset X$. Let $\mathcal{R}_\sigma$ denote the collection of subsets of $X \times Y$ which can be written as countable unions of sets in $\mathcal{R}$, and let $\mathcal{R}_{\sigma\delta}$ denote the collection of subsets of $X \times Y$ which can be written as countable intersections of sets in $\mathcal{R}_\sigma$.

Lemma. *Suppose $E \subset \mathcal{R}_{\sigma\delta}$. Then $E_x \in \mathcal{B}$ for all $x \in X$ and similarly $E_y \in \mathcal{A}$ for all $y \in Y$.*

Proof. By symmetry it suffices to show that $E_x \in \mathcal{B}$. The lemma is trivial for $E \in \mathcal{R}$. We first show that it holds for $E \in \mathcal{R}_\sigma$. Suppose $E = \cup_{i=1}^{\infty} E_i$ with $E_i \in \mathcal{R}$ for all i. Then

$$\chi_{E_x}(y) = \chi_E(x, y) = \sup_i \chi_{E_i}(x, y) = \sup_i \chi_{(E_i)_x}(y).$$

Since $\chi_{(E_i)_x}$ is measurable for each i, it follows that χ_{E_x} is measurable, so E_x is measurable.

Now suppose that $E = \cap_{i=1}^{\infty} E_i \in \mathcal{R}_{\sigma\delta}$ where $E_i \in \mathcal{R}_\sigma$ for all i. Then

$$\chi_{E_x}(y) = \chi_E(x, y) = \inf \chi_{E_i}(x, y) = \inf \chi_{(E_i)_x}(y).$$

Since $\chi_{(E_i)_x}$ is measurable for each i, it follows that χ_{E_x} is measurable, so E_x is measurable. $\square$

Lemma. *Suppose $E \in \mathcal{R}_{\sigma\delta}$ with $(\mu \times \nu)(E) < \infty$. Define $g : X \to \mathbb{R}$ by $g(x) := \nu(E_x)$. Then g is integrable with $\int g\, d\mu = (\mu \times \nu)(E)$.*

Proof. The lemma is trivial for $E \in \mathcal{R}$. Suppose next that $E = \cup_{i=1}^{\infty} E_i \in \mathcal{R}_\sigma$, where $E_i \in \mathcal{R}$ for all i. By writing $E_i - (\cup_{j<i} E_j)$ as a disjoint union of sets in $\mathbb{R}$, we may reduce to the case where the sets E_i are disjoint. Set $g_i(x) := \nu\big((E_i)_x\big)$. Then g_i is a nonnegative measurable function and since $E_x = \amalg_{i=1}^{\infty} (E_i)_x$ we get $g(x) = \sum_{i=1}^{\infty} g_i(x)$. By the Lebesgue Monotone Convergence Theorem

$$\int g\, d\mu = \sum_{i=1}^{\infty} \int g_i\, d\mu = \sum_{i=1}^{\infty} (\mu \times \nu)(E_i) = (\mu \times \nu)(E).$$

Thus the lemma holds for $E \in \mathcal{R}_\sigma$.

Finally, suppose that $E \in \mathcal{R}_{\sigma\delta}$. Then there exist sets $E_i \in \mathcal{R}_\sigma$ with $E_{i+1} \subset E_i$ such that $E = \cap_{i=1}^{\infty} E_i$. Given $\epsilon > 0$, by a proposition from Section 2, there exists $C \in \mathcal{R}_\sigma$ with $E \subset C$ such that $(\mu \times \nu)(C) - (\mu \times \nu)(E) < \epsilon$. In particular, $(\mu \times \nu)(C) < \infty$. Set $C_i := E_i \cap C$. Notice that $C_i \in \mathcal{R}_\sigma$ and $E = \cup_{i=1}^{\infty} C_i$ since $E \subset C$. Thus by replacing E_i with C_i, we may assume that $(\mu \times \nu)(E_i) < \infty$ for all i. Set $g_i(x) := \nu\big((E_i)_x\big)$. Then $\int g_i(x)\, d\mu < \infty$ for almost all x. Thus

$$g(x) = \nu(E_x) = \lim_{i\to\infty} \nu\big((E_i)_x\big) = \lim_{i\to\infty} g_i(x)$$

for almost all x. The previous case applies to E_i for each i and since $g_i \le g_1$ the Lebesgue Dominated Convergence Theorem applies to give

$$\int g\, du = \lim_{i\to\infty} \int g_i\, d\mu = \lim_{i\to\infty} (\mu \times \nu)(E_i) = (\mu \times \nu)(E).$$

$\square$

A measure space $(Z, \mathcal{C}\, \lambda)$ is called a "complete measure space" of $\mathcal{C}$ contains all sets of measure 0. For example, Lebesgue measure is complete as is any measure produced by the procedure described in section 2. Furthermore, any incomplete measure can be completed by applying the procedure of section 2 to it.

Lemma. *Suppose that $(Y, \mathcal{B}, \nu)$ is complete. If $(\mu \times \nu)(E) = 0$ then $\mu(E_x) = 0$ for almost all x.*

Proof. Suppose $(\mu \times \nu)(E) = 0$. By a proposition from Section 2, there exists $C \in \mathcal{R}_{\sigma\delta}$ such that $E \subset C$ and $(\mu \times \nu)(C) = 0$. Let $g(x) = \nu(C_n)$. The previous lemma gives $\int g\, d\mu = (\mu \times \nu)(C) = 0$. But $g \ge 0$ and so $g(x) = 0$ for almost all x. I.e. $\nu(C_x) = 0$ for almost all x. However $E_x \subset C_x$ so by completeness of μ for almost all x we conclude that E_x is measurable with $\nu(E_x) = 0$. $\square$

Lemma. *Suppose that* $(Y, \mathcal{B}, \nu)$ *is complete. Let* $E \subset X \times Y$ *be measurable with* $(\mu \times \nu)(E) < \infty$*. Then for almost all* x*, the set* $E_x \subset Y$ *is measurable, the function* $g(x) := \nu(E_x)$ *is measurable, and* $\int g \, d\mu = (\mu \times \nu)(E)$*.*

Proof. By a proposition from section 2, there exists $C \in \mathcal{R}_{\sigma\delta}$ such that $E \subset C$ and $(\mu \times \nu)(E) = (\mu \times \nu)(C)$. Set $G := C - E$. Since C and E are measurable, so is G and we have $(\mu \times \nu)(G) = (\mu \times \nu)(C) - (\mu \times \nu)(E) = 0$. Therefore the previous lemma gives $\nu(G_x) = 0$ for almost all x. Hence $g(x) = \nu(E_x) = \nu(C_x)$ for almost all x. Thus the lemmas above imply that g is measurable with $\int g \, d\mu = (\mu \times \nu)(C) = (\mu \times \nu)(E)$. □

Theorem (Fubini). *Let* $(X, \mathcal{A}, \mu)$ *and* $(Y, \mathcal{B}, \nu)$ *be complete measure spaces. Let* $f : X \times Y \to \mathbb{R}$ *be integrable in the product measure* $\mu \times \nu$*. Then*

1) *The function* $f_x(y) := f(x, y)$ *is integrable on* Y*.*
2) *The function* $f_y(x) := f(x, y)$ *is integrable on* X*.*
3) *The function* $I(x) := \int_Y f(x, y) \, d\nu$ *is integrable on* X*.*
4) *The function* $J(y) := \int_X f(x, y) \, d\mu$ *is integrable on* Y*.*
5)
$$\int_X \left(\int_Y f \, d\nu \right) d\mu = \int_{X \times Y} f \, d(\mu \times \nu) = \int_Y \left(\int_X f \, d\mu \right) d\nu$$

Note: In part (3), the notation $\int_Y f(x, y) \, d\nu$ is used to denote the integral $\int_Y f_x \, d\nu$ of the function in part (1) and similar notation is used in part (4). In part (5), the notation $\int_X \left(\int_Y f \, d\nu \right) d\mu$ is used to denote $\int_X I \, d\mu$ and $\int_Y \left(\int_X f \, d\mu \right) d\nu$ is used to denote $\int_Y J \, d\nu$.

Proof. By symmetry, it suffices to prove (1), (3), and the first equality of (5), If the theorem holds for two functions it also holds for their sum and difference. Therefore it suffices to consider the case where f is nonnegative. The previous lemma asserts that the theorem holds when f is the characteristic function of a set of finite measure, so it holds for any simple function which is zero outside of some set of finite measure. By a theorem from section 4, there exists a sequence of simple functions s_n such that $(s_n) \nearrow f$. Then $\left((s_n)_x\right) \nearrow f_x$ and so f_x is measurable. Since $s_n \leq f$ and f is integrable, it follows that s_n is zero outside of a set of finite measure, so the theorem applies to s_n for each n. The Lebesgue Monotone Convergence Theorem gives

$$\int_Y f(x, y) \, d\nu = \lim_{n \to \infty} \int_Y s_n(x, y) \, d\nu$$

and

$$\int_X \left(\int_Y f \, d\nu \right) d\mu = \lim_{n \to \infty} \int_X \left(\int_Y s_n \, d\nu \right) d\mu = \lim_{n \to \infty} \int_{X \times Y} s_n \, d(\mu \times \nu) = \int_{X \times Y} f \, d(\mu \times \nu)$$

□

Corollary (of the proof of Fubini). *Let* $(X, \mathcal{A}, \mu)$ *and* $(Y, \mathcal{B}, \nu)$ *be complete measure spaces such that* $\mu(X) < \infty$ *and* $\nu(Y) < \infty$*. Let* $f : X \times Y \to \mathbb{R}$ *be a nonnegative function* *which is measurable in the product measure $\mu \times \nu$. Then f is integable so the conclusions of Fubini's theorem apply.*

Proof. The only place in the proof of Fubini where the integrability of f is used is to conclude that the step functions s_n described in the proof are zero outside of a set of finite measure. However the hypotheses that $\mu(X) < \infty$ and $\nu(Y) < \infty$ make that trivial. $\square$

6. $\mathcal{L}_p$ spaces

Definition. Let V be a vector space over $\mathbb{F}$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A *norm* on V is a function $V \to \mathbb{R}^+$ which, using the notation $\|x\|$ for the norm of x, satisfies
1) $\|x\| = 0$ if and only if $x = 0$
2) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.
3) $\|\alpha x\| = |\alpha| \, \|x\|$ for all $\alpha \in F$, $x \in V$.

Notation. $\mathcal{C}(X;\mathbb{F}) := \{f : X \to \mathbb{F} \mid f \text{ is continuous and } |f| \text{ is bounded}\}$.

Recall from `MATB43` that if X is a closed bound subset of $\mathbb{R}^n$ then $|f|$ is automatically bounded.

Examples of normed vector spaces.
1) $V = \mathbb{F}^n$ with $\|x\| = \sqrt{|x_1|^2 + \ldots |x_n|^2}$
2) $V = \mathcal{C}(X;\mathbb{F})$ with $\|f\| = \sup_{x \in X} |f(x)|$. This is called the "sup norm" on $\mathcal{C}(X;\mathbb{F})$. When discussing $\mathcal{C}(X;\mathbb{F})$ as a normed vector space, $\| \; \|$ shall refer to the sup norm unless stated otherwise.
3) $V = \mathcal{C}([0,1];\mathbb{F})$ with $\|f\| = \int_0^1 |f(x)| \, dx$

Given a normed vector space V, we can define a metric on V by $d(x, y) := \|x - y\|$. In example (1) above, this gives the standard distance function on $\mathbb{R}^n$.

Notation. Let (X, σ, μ) be a measure space. Suppose $p \in [1, \infty)$. Let

$$\mathcal{L}^p(X;\mathbb{F}) := \left\{ f : X \to \mathbb{F} \;\Big|\; \int |f|^p \, d\mu < \infty \right\} \Big/ \sim,$$

where $f \sim g$ if and only if $f = g$ almost everywhere. For $f \in \mathcal{L}^p(X;\mathbb{F})$ set $\|f\|_p := \int |f|^p \, d\mu$. If $f : X \to \mathbb{F}$, we will sometimes say "f lies in $\mathcal{L}^p(X;\mathbb{F})$" when strictly speaking we ought to say "the equivalence class of f lies in $\mathcal{L}^p(X;\mathbb{F})$." We will sometimes write simply $\mathcal{L}^p(X)$ when either $\mathbb{F}$ is understood or it is a statement which is true for both $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$.

We will show that $\| \; \|_p$ forms a norm on $\mathcal{L}^p(X)$.

Definition. If $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$, then p and q are called *conjugate exponents*.

Hölder's Inequality. *Suppose $f \in \mathcal{L}^p(X)$ and $g \in \mathcal{L}^q(X)$ where $\frac{1}{p} + \frac{1}{q} = 1$. Then $fg \in \mathcal{L}^1(X)$ and $\|fg\|_1 \leq \|f\|_p \|g\|_q$.*

Proof. The result is trivial if either $f = 0$ almost everywhere or $g = 0$ almost everywhere, so suppose neither of these holds.

Set $F := \frac{f}{\|f\|_p}$ and $G := \frac{g}{\|g\|_q}$. Then $\|F\|_p = 1$, $\|G\|_q = 1$, and the inequality holds for f and g if and only if it holds for F and G.

Note that the function $\exp(t) = e^t$ is convex (portion of the graph joining any two points is never above its secant line) because the second derivative is positive. In other words $\exp\big(\lambda x + (1-\lambda)y\big) \le \lambda \exp(x) + (1-\lambda)\exp(y)$ for all $x\, y \in \mathbb{R}$ and $\lambda \in [0,1]$. If $F(t) = 0$ or $G(t) = 0$ it is trivial that

$$|F(t)|\,|G(t)| \le \frac{1}{p}|F(t)|^p + \frac{1}{q}|G(t)|^q.$$

If $F(t) \ne 0$ and $G(t) \ne 0$, applying the convexity with $x := \ln\big(|F(t)|^p\big)$ and $y := \ln\big(|G(t)|^q\big)$ again gives

$$|F(t)|\,|G(t)| = \exp\left(\frac{1}{p}\ln\big(|F(t)|^p\big) + \frac{1}{q}\ln\big(|G(t)|^q\big)\right) \le \frac{1}{p}|F(t)|^p + \frac{1}{q}|G(t)|^q.$$

Thus the preceding inequality holds for all $t \in X$ and therefore

$$\int |FG|\,d\mu \le \frac{1}{p}\int |F|^p\,d\mu + \frac{1}{q}\int |G|^q\,d\mu = \frac{1}{p}(\|F\|_p)^p + \frac{1}{q}(\|G\|_q)^q = \frac{1}{p}1^p + \frac{1}{q}1^q = \frac{1}{p} + \frac{1}{q} = 1$$

Thus the inequality holds for F and G, as desired. □

Minkowski's Inequality. *If $f, g \in \mathcal{L}^p(X)$, then $f + g \in \mathcal{L}^p$ with $\|f+g\|_p \le \|f\|_p + \|g\|_p$.*

Proof. For $p = 1$ the result follows immediately from $|f(x) + g(x)| \le |f(x)| + |g(x)|$, so assume $p > 1$.

The function $t \mapsto t^p$ is convex and therefore $\left(\frac{1}{2}|f| + \frac{1}{2}|g|\right)^p \le \frac{1}{2}|f|^p + \frac{1}{2}|g|^p$. Thus $f + g \in \mathcal{L}^p(X)$.

Let q be the conjugate exponent of p. Then $q + p = pq$ or equivalently $p = q(p-1)$.

Since $f + g \in \mathcal{L}^p(X)$ and $p = q(p-1)$ we deduce that $(f+g)^{p-1} \in \mathcal{L}^q(X)$. Therefore Hölder implies $\int |f|\,|f+g|^{p-1}\,d\mu \le \|f\|_p\,\|(f+g)^{p-1}\|_q$ and similarly $\int |g|\,|f+g|^{p-1}\,d\mu \le \|g\|_p\,\|(f+g)^{p-1}\|_q$.

$$|f+g|^p = |f+g|\,|f+g|^{p-1} \le (|f| + |g|)|f+g|^{p-1} = |f|\,|f+g|^{p-1} + |g|\,|f+g|^{p-1}.$$

Therefore

$$\begin{aligned}
(\|f+g\|_p)^p &\le \|f\|_p\,\|(f+g)^{p-1}\|_q + \|g\|_p\,\|(f+g)^{p-1}\|_q \\
&= (\|f\|_p + \|g\|_p)\left(\int |f+g|^{(p-1)q}\right)^{1/q} = (\|f\|_p + \|g\|_p)\left(\int |f+g|^p\right)^{1/q} \\
&= (\|f\|_p + \|g\|_p)\,(\|f+g\|_p)^{p/q} = (\|f\|_p + \|g\|_p)\,(\|f+g\|_p)^{(p-1)} \\
&= (\|f+g\|_p)^p
\end{aligned}$$

□

Corollary. *$\mathcal{L}^p(X)$ is a vector space over $\mathbb{F}$ and $\| \ \|_p$ is a norm on $\mathcal{L}^p(X;\mathbb{F})$.*

Proof. If $f \sim g$ then $|f|^p = |g|^p$ almost everywhere and so $\|f\|_p = \|g\|_p$. Therefore $\| \ \|_p$ is well defined. Minkowski's inequality guarantees that $\mathcal{L}^p(X)$ forms a vector space. It is trivial that $\|0\|_p = 0$ and since $\|f|^p \geq 0$ it is also clear that if $\|f\|_p = 0$ then $|f|^p = 0$ almost everywhere and so $f = 0$ almost everywhere. Property (2) in the definition of norm is also trivial and property (3) is Minkowski's inequality. □

Intuitively, we wish to regard the *sup* norm on $\mathcal{C}(X)$ as a limit of the $\mathcal{L}^p$ norms as $p \to \infty$. For motivation consider positive real numbers $a_1 > a_2 > \ldots > a_n$ and let $L = \lim_{p\to\infty} = ({a_1}^p + \ldots + {a_n}^p)^{1/p}$. Then using L'Hôpital's rule,

$$\begin{aligned}\ln L &= \lim_{p\to\infty} \frac{\ln({a_1}^p + \ldots + {a_n}^p)}{p} = \lim_{p\to\infty} \frac{{a_1}^p \ln(a_1) + \ldots + {a_n}^p \ln(a_n)}{{a_1}^p + \ldots + {a_n}^p} \\ &= \lim_{p\to\infty} \frac{\ln(a_1) + (a_2/a_1)^p \ln(a_2) \ldots + (a_n/a_1)^p \ln(a_n)}{1 + (a_2/a_1)^p \ldots + (a_n/a_1)^p} = \ln(a_1)\end{aligned}$$

and thus $L = a_1$. In other words, the limit converges to the maximum of $\{a_1, \ldots, a_n\}$. If σ is a nonnegative real-valued step function then $\|\sigma\|_p$ is a sum of the above form and so $\lim_{p\to\infty} \|\sigma\|_p = \max\{\sigma(x)\}$ where the maximum is taken over the (finite) set of values of σ, which is equivalent to $\sup_{x\in X} \sigma(x)$. More generally, if $f \in \mathcal{C}([0,1])$, by approximating one can show that the Riemann integral (and thus also the Lebesgue integral since they are equal) satisfies $\lim_{p\to\infty} \|f\|_p = \sup_{x\in[0,1]} f(x)$. With this as motivation, we would like to define an norm, known as the $\mathcal{L}_\infty$ norm, by $\|f\|_\infty := \sup_{x\in X} |f(x)|$, but first we must decide upon a suitable domain for $\| \ \|_\infty$, and make the definition more precise by ignoring function values on sets of measure zero in the appropriate fashion.

Definition. Let (X, σ, μ) be a measure space. We say that a function a function $f : X \to \mathbb{F}$ is *essentially bounded* if there exists N such that $\mu(\{x \mid |f(x)| > N\}) = 0$.

Set $\mathcal{L}^\infty(X) := \{f : X \to \mathbb{F} \mid f \text{ is essentially bounded}\}/\sim$, where $f \sim g$ if and only if $f = g$ almost everywhere. For $f \in \mathcal{L}^\infty(X)$, define

$$\|f\|_\infty := \inf\{N \mid \mu(\{x \mid |f(x)| > N\} = 0)\},$$

called the *essential suprement* of f.

Notice that if $X \subset \mathbb{R}^n$ and $f \in \mathcal{C}(X;\mathbb{R})$, the set of bounded continuous functions on X, then the essential supremum of f equals the supremum $\sup\{|f(x)|\}_{x\in X}$ since for a continuous function, $|f(x)| > N$ implies that $|f(y)| > N$ on some open set containing x, and in particular this inequality holds on a set of nonzero measure. Thus in this case $\mathcal{C}(X;\mathbb{R}) \subset \mathcal{L}^\infty(X;\mathbb{R})$ with $\|f\| = \|f\|_\infty$.

Proposition. *If f is essentially bounded then $\mu(\{x \mid |f(x)| > \|f\|_\infty\}) = 0$.*

Proof. By definition of the essential supremum, $\mu(\{x \mid |f(x)| > \|f\|_\infty + 1/n\}) = 0$ for every positive n. Since $\{x \mid |f(x)| > \|f\|_\infty\} = \cup_{n=1}^\infty \{x \mid |f(x)| > \|f\|_\infty + 1/n\}$ the result follows. □

Proposition. $\big(\mathcal{L}^\infty(X), \| \, \|_\infty\big)$ *forms a normed vector space.*

Proof. We must check that $\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty$ for $f, g \in \mathcal{L}^\infty(X)$.

$$\{x \mid |f(x)+g(x)| > \|f\|_\infty + \|g\|_\infty\} \subset \{x \mid |f(x)| > \|f\|_\infty\} \cup \{x \mid |g(x)| > \|g\|_\infty\}$$

which has measure 0 since it is the union of two sets with measure 0. $\square$

We now consider the question of how, if at all, the sets $\mathcal{L}^p(X)$ and $\mathcal{L}^{p'}(X)$ are related as p and p' vary.

Example. Let $X = [0,1]$ and let

$$f(x) = \begin{cases} x^{-1/2} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

We can see that f is integrable and compute $\int_{[0,1]} f \, d\mu$ by considering the functions

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

Each f_n is integrable and $f_n \nearrow f$, so by the Lebesgue Monotone Convergence Theorem, f is integrable with $\int_{[0,1]} f \, du = \lim_{n\to\infty} \int_{[0,1]} f_n \, d\mu = 2$. Thus $f \in \mathcal{L}^1([0,1])$. However using the same method we can see that $\int_{[0,1]} f_n^2 \, d\mu = \infty$, so $f \in \mathcal{L}^1([0,1]) - \mathcal{L}^2([0,1])$. If $\mu(X) = \infty$, it is also possible to have functions which are in $\mathcal{L}^2(X) - \mathcal{L}^1(X)$. For example, if $X = [1, \infty)$ and $f(x) = 1/x$ then $f \in \mathcal{L}^2(X) - \mathcal{L}^1(X)$. As shown in the next two propositions, this is not possible if $\mu(X) < \infty$.

Proposition. *Suppose $\mu(X) < \infty$. Then $\mathcal{L}^\infty(X) \subset \mathcal{L}^p(X)$ for all p and the inequality $\|f\|_p \le \big(\mu(X)\big)^{1/p} \|f\|_\infty$ is satisfied for all $f \in \mathcal{L}^\infty(X)$.*

Proof. Suppose $f \in \mathcal{L}_\infty(X)$ and let $M := \|f\|_\infty$. Let $E := \{x \mid |f(x)| \le M\}$. Since $\mu(E^c) = 0$, we have

$$\int |f|^p \, d\mu = \int_E |f|^p \, d\mu \le \int_E M^p \, d\mu = M^p \mu(E) = M^p \mu(X) < \infty.$$

Thus $f \in \mathcal{L}^p(X)$ and $\|f\|_p \le M\big(\mu(X)\big)^{1/p} = \big(\mu(X)\big)^{1/p} \|f\|_\infty$ $\square$

Proposition. *Suppose $\mu(X) < \infty$ and $0 < p < p' \le \infty$. Then $\mathcal{L}^{p'}(X) \subset \mathcal{L}^p(X)$. Furthermore, if (f_n) is a sequence in $\mathcal{L}^{p'}(X)$ with $(f_n) \to f$ in the metric coming from the $\| \, \|_{p'}$ norm then $(f_n) \to f$ in the metric coming from the $\| \, \|_p$ norm.*

Proof. The containment $\mathcal{L}^\infty(X) \subset \mathcal{L}^p(X)$ was shown in the previous Proposition, and the convergence statement follows immediately from the inequality $\|f - f_n\|_p \le \big(\mu(X)\big)^{1/p} \|f - f_n\|_\infty$.

Now consider the case $p' < \infty$. Suppose $f \in \mathcal{L}^{p'}(X)$. Let $E = \{x \mid |f(x)| \leq 1\}$. Then since $|f^p(x)| \leq |f^{p'}(x)|$ when $x \in E^c$,

$$\begin{aligned}\int |f|^p \, d\mu &= \int_E |f|^p \, d\mu + \int_{E^c} |f|^p \, d\mu \leq \int_E 1 \, d\mu + \int_{E^c} |f|^{p'} \, d\mu \leq \int_X 1 \, d\mu + \int_X |f|^{p'} \, d\mu \\ &= \mu(X) + (\|f\|_{p'})^{p'} < \infty.\end{aligned}$$

Thus $f \in \mathcal{L}^p(X)$ and so we have shown $\mathcal{L}^{p'}(X) \subset \mathcal{L}^p(X)$.

Suppose now that $(f_n) \to f$ in the metric coming from the $\| \ \|_{p'}$ norm. We wish show that $(f_n) \to f$ in the metric coming from the $\| \ \|_p$ norm. By replacing f_n by $f_n - f$ we see that it suffices to prove the statement in the case $f = 0$. So suppose $(f_n) \to 0$ in the metric coming from the $\| \ \|_{p'}$ norm. Equivalently $(\|f_n\|_{p'}) \to 0$ which is equivalent to $\big((\|f_n\|_{p'})^\alpha\big) \to 0$ whenever $\alpha > 0$. In particular, $\big((\|f_n\|_{p'})^p\big) \to 0$.

Let $r = p'/p > 1$ and let s be the conjugate exponent to r. Then Hölder's inequality gives

$$\begin{aligned}0 \leq (\|f_n\|_p)^p &= \int |f_n|^p \, d\mu = \|(f_n)^p\|_1 \leq \|(f_n)^p\|_r \, \|1\|_s \\ &= \left(\int (|f_n|^p)^r \, d\mu\right)^{1/r} \left(\int 1^s \, d\mu\right)^{1/s} = \left(\int |f_n|^{p'} \, d\mu\right)^{1/r} (\mu(X))^{1/s} \\ &= (\|f_n\|_{p'})^p \, (\mu(X))^{1/s}.\end{aligned}$$

Since the last term tends to 0 as $n \to \infty$, the squeezing principle gives $(\|f_n\|_p)^p \to 0$ and therefore $(\|f_n\|_p) \to 0$. □

Remark. The statement in the Proposition that convergence of sequences is preserved is equivalent to saying the inclusion map $\mathcal{L}^{p'}(X) \to \mathcal{L}^p(X)$ is continuous whenever $\mu(X) < \infty$ and $p \leq p' \leq \infty$. For $p' < \infty$ there is no direct analogue of the formula $\|f\|_p \leq \|f\|_\infty \big(\mu(X)\big)^{1/p}$ because another term enters into the right hand side coming from the $\mu(E)$ which need not equal 0.

Recall that a sequence (a_n) in a metric space is called a *Cauchy sequence* if there for every $\epsilon > 0$ there exists N such that $d(a_n, a_k) < \epsilon$ whenever $n, k \geq N$. A metric space X is called *complete* if every Cauchy sequence in X converges.

Example. As discussed in `MATA37` and `MATB43`, $\mathbb{R}$ is complete as a consequence of the least upper bound axiom. This can be used to show that $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete metric spaces (with their standard metric). $(0,1]$ is not a complete metric space because $(1/n)$ is a Cauchy sequence in $(0,1]$ which does not converge to any point in $(0,1]$. In fact, it follows from the definitions that a subset of a complete metric space is complete if and only if it is a closed subset.

Definition. A *Banach space* is a normed vector space which is complete in the metric coming from the norm.

Lemma. *Let X be a measure space. Then $\mathcal{L}^\infty(X)$ is a Banach space.*

Proof. Let (f_n) be a Cauchy sequence in $\mathcal{L}^\infty$. Then for any positive integer k there exists N_k such that $\|f_n - f_m\|_\infty < 1/k$ for all $n, m \geq N_k$. Given m, n, k set $I_{m,n,k} := \{x \mid |f_n(x) - f_m(x)| \geq 1/k\}$ and set $I := \cup_{k=1}^\infty \cup_{n,m \geq N_k} I_{m,n,k}$. By construction, $\big(f_n(x)\big)$ is a Cauchy sequence in $\mathbb{F}$ whenever $x \notin I$. By definition of N_k, $\mu(I_{m,n,k}) = 0$ whenever $n, m \geq N_k$ and therefore $\mu(I) = 0$. Thus $\big(f_n(x)\big)$ is a Cauchy sequence in $\mathbb{F}$ for almost all x.

Define $f : X \to \mathbb{F}$ by

$$f(x) = \begin{cases} \lim_{n\to\infty} f_n(x) & \text{if } x \notin I; \\ 0 & \text{if } x \in I. \end{cases}$$

Suppose $x \notin I$. Given k there exists M_x such that $|f(x) - f_m(x)| < 1/k$ for all $m \geq M_x$. (Note: Although it does not appear in the notation, M_x depends on k as well as on x.) By replacing M_x by a larger number if necessary, we may assume that $M_x \geq N_k$. Therefore if $n \geq N_k$ then $|f_{M_x}(x) - f_n(x)| < 1/k$ by definition of N_k, using $x \notin I_{n,M_x,k} \subset I$. Thus for $n \geq N_k$ we have

$$|f(x) - f_n(x)| \leq |f(x) - f_{M_x}(x)| + |f_{M_x}(x) - f_n(x)| < 1/k + 1/k = 2/k.$$

Since this holds for all x outside the set I, which has measure 0, it follows that $f - f_n \in \mathcal{L}^\infty(X)$ for all n and $\|f - f_n\| \leq 2/k$ whenever $n \geq N_k$. The latter statement implies that $\|f - f_n\|_\infty \to 0$ in $\mathbb{R}$. Therefore $f = (f - f_n) + f_n$ lies in $\mathcal{L}^\infty(X)$ since it is the sum of two elements of $\mathcal{L}^\infty(X)$, and $f_n \to f$ in $\mathcal{L}^\infty(X)$ since it is equivalent to $\|f - f_n\|_\infty \to 0$ in $\mathbb{R}$. □

To show that $\mathcal{L}^p(X)$ is complete for $p < \infty$ we will first need to establish some other properties.

The closure of a subset A of topological space X is the smallest closed subset of X containing A. It is denoted $\bar{A}$. A subset A is called *dense* if $\bar{A} = X$.

Although intuitively one tends to visualize dense subsets of X as being "most" of X, there is no direct connection between denseness and measure theoretic concepts. For example, $\mathbb{Q}$ is dense in $\mathbb{R}$ even though it has measure 0.

Lemma. *$\{$simple functions$\}$ is dense in $\mathcal{L}^p(X)$.*

Proof. Given $f \in \mathcal{L}^p(X)$ we must find a sequence (s_n) of simple functions such that (s_n) converges to f in the metric coming from the $\mathcal{L}^p$-norm. Equivalently, we must find a sequence (s_n) of simple functions such that the sequence $(\|f - s_n\|_p)$ of real numbers converges to 0 in $\mathbb{R}$.

Consider first the case where $f \geq 0$. Since f is measurable, as shown earlier, there exists a sequence (s_n) of simple functions such that $(s_n) \nearrow f$. Thus $\lim_{n\to\infty}(f - s_n)(x) = 0$ for all x and so $\lim_{n\to\infty}(f - s_n)^p(x) = 0$ for all x. Since there exists integrable g such that $(f - s_n)^p \leq g$ (namely $g = |f^p|$, for example) the Lebesgue Dominated Convergence Theorem applies and gives

$$\lim_{n\to\infty} \|f - s_n\|_p := \lim_{n\to\infty} \left(\int (f - s_n)^p \, d\mu\right)^{1/p} = \left(\int \lim_{n\to\infty} (f - s_n)^p \, d\mu\right)^{1/p} = \left(\int 0 \, d\mu\right)^{1/p} = 0$$

and so (s_n) converges to f in the $\| \ \|_p$-norm.

For arbitrary $f \in \mathcal{L}^p(X)$, write $f = f^+ - f^-$ where $f^+, f^- \geq 0$. Choose sequences of simple functions (s_n), (t_n) such that $(s_n) \to f^+$ and $(t_n) \to f^-$ in $\| \ \|_p$. Then $\|f - (s_n + t_n)\|_p \leq \|f^+ - s_n\|_p + \|f^- - t_n\|_p$ and so the sequence $(s_n + t_n)$ of simple functions converges to f in the $\| \ \|_p$-norm. ☐

Lemma. *A normed vector space V is complete if and only if it has the property that whenever (x_n) is a sequence in V such that the series $\sum_{n=1}^\infty \|x_n\|$ converges in $\mathbb{R}$, the series $\sum_{n=1}^\infty x_n$ converges in V.*

Proof. To simplify the discussion, we will refer to the property mentioned in the Lemma by the words "absolute convergence implies convergence". Recall that a series is defined as convergent if its sequence of partial sums converges.

Suppose first that V is complete. We wish to show absolute convergence implies convergence. Let (x_n) be a sequence in V such that the series $\sum_{n=1}^\infty \|x_n\|$ converges in $\mathbb{R}$. Then the sequence of partial sums $S_n := \sum_{k=1}^n \|x_k\|$ forms a Cauchy sequence in $\mathbb{R}$. Since for $n \leq m$ the sequence of partial sums $P_n := \sum_{k=1}^n x_k$ satisfies

$$\|P_m - P_n\| = \|x_{n+1} + \ldots + x_m\| \leq \|x_{n+1}\| + \ldots + \|x_m\| = S_m - S_n,$$

the sequence (P_n) forms a Cauchy sequence in V and thus converges because V is complete. Thus the series $\sum_{n=1}^\infty x_n$ converges in V.

Conversely suppose that V has the property that absolute convergence implies convergence. Let (x_n) be a Cauchy sequence in V. If a Cauchy sequence has a convergent subsequence then it converges, so we will show that (x_n) has a convergent subsequence. Set $N_1 := 1$ and having chosen $N_1, \ldots, N_{k-1}$, choose N_k such that $N_k > N_{k-1}$ and $\|x_n - x_m\| < 1/2^k$ whenever $n, m \geq N_k$. Consider the series $x_{N_1} + \sum_{k=2}^\infty (x_{N_k} - x_{N_{k-1}})$ in V. The corresponding series of norms satisfies

$$\|x_{N_1}\| + \sum_{k=2}^\infty \|x_{N_k} - x_{N_{k-1}}\| \leq \|x_{N_1}\| + \sum_{k=2}^\infty 1/2^{k-1} = x_{N_1} + 1 < \infty$$

so it converges in $\mathbb{R}$. Therefore V's "absolute convergence implies convergence property" says that the series $x_{N_1} + \sum_{k=2}^\infty (x_{N_k} - x_{N_{k-1}})$ converges in V. However the kth partial sum of this series is x_{N_k}. Thus $(x_{N_k})_{k=1}^\infty$ forms a convergent subsequence of (x_n) and so the Cauchy sequence (x_n) converges. ☐

Remark. Note the use of the vector space operations in V in the preceding proof. Consider, for example, the space $X = (0, 1]$ where all the elements are positive and so $\|x\| = x$ and thus there is no distinction between convergence and absolute convergence. It is certainly not valid to conclude that X is complete based on the tautology that within X absolute convergence implies convergence!

Theorem (Riesz-Fischer). *Let X be a measure space. Then $\mathcal{L}^p(X)$ is a Banach space for all $p \in [1, \infty]$.*

Proof. The case $p = \infty$ was done earlier, so suppose $p < \infty$. By the Lemma above, it suffices to show that absolute convergence implies convegence within $\mathcal{L}^p(X)$. Let (f_n) be a sequence in $\mathcal{L}^p(X)$ such that $\sum_{n=1}^{\infty} \|f_n\|_p$ converges in $\mathbb{R}$. We must show that $\sum_{n=1}^{\infty} f_n$ converges in $\mathcal{L}^p(X)$.

Let $M = \sum_{n=1}^{\infty} \|f_n\|_p$. Then $M < \infty$ by hypothesis. Let $h_k(x) = \sum_{n=1}^{k} |f_n(x)|$. Then $\|h_k\|_p \leq \sum_{n=1}^{k} \|f_n\|_p \leq M$ for all k. It follows that $E := \{x \mid \lim_{k\to\infty} h_k(x) = \infty\}$ has measure 0.

Define $h : X \to \mathbb{R}$ by

$$h(x) := \begin{cases} \lim_{k\to\infty} h_k(x) = \sum_{n=1}^{\infty} |f_n(x)| & \text{if } x \notin E; \\ 0 & \text{if } x \in E. \end{cases}$$

Then $(h_k^p) \nearrow h^p$ almost everywhere, so by the Lebesgue Monotone Convergence Theorem $\int h^p \, d\mu = \lim_{k\to\infty} \int h_k^p \, d\mu \leq M^p$.

If $x \notin E$ then by definition $\sum_{n=1}^{\infty} |f_n(x)|$ converges and since $\mathbb{R}$ is complete "absolute convergence implies convergence" gives $\sum_{n=1}^{\infty} f_n(x)$ converges. Define $g : X \to \mathbb{F}$ by

$$g(x) := \begin{cases} \sum_{n=1}^{\infty} f_n(x) & \text{if } x \notin E; \\ 0 & \text{if } x \in E. \end{cases}$$

We will show that $\sum_{n=1}^{k} f_n$ converges in $\mathcal{L}^p(X)$ by showing that g belongs to $\mathcal{L}^p(X)$ and that the series converges to g in the $\| \ \|_p$ norm. Since $|g| \leq |h|$, $\int |g|^p \, d\mu \leq M^p$, so $g \in \mathcal{L}^p(X)$. Set $g_k = \sum_{n=1}^{k} f_n$. Then $g_k \in \mathcal{L}^p(X)$ with $\|g_k\|_p \leq M$ for all k. Since $2^p|g|^p$ is integrable with $|g_k - g|^p < 2^p|g|^p$ for all p, the Lebesgue Dominated Convergence Theorem applies to $(g_k - g)$ and yields $\left(\int |g_k - g|^p \, d\mu\right) \to \int 0 \, d\mu = 0$. Therefore $(g_k) \to g$ in $\mathcal{L}^p(X)$. That is, the series $\sum_{n=1}^{k} f_n$ converges in $\mathcal{L}^p(X)$ as desired. □

Let X be a compact subset of $\mathbb{R}^n$. Recall that continuous functions on compact subsets of $\mathbb{R}^n$ are integrable and so $\mathcal{C}(X) \subset \mathcal{L}^p(X)$ for all p.

Proposition. *Let X be a compact subset of $\mathbb{R}^n$ (or more generally a compact Hausdorff space with $\mu(X) < \infty$). Then the continuous functions are dense in $\mathcal{L}^p(X)$ for every $p \in [1, \infty)$.*

Proof. Since the simple functions are dense in $\mathcal{L}^p(X)$ it suffices to show that every simple function is a limit of continuous functions in the metric coming from the $\| \ \|_p$-norm. Since limits commute with sums and Lebesgue Dominated implies that $\lim_{n\to\infty} \|f_n\|_p = \|\lim_{n\to\infty} f_n\|_p$, by writing the simple function as a finite sum of characteristic functions we see that it suffices to show that every characteristic function is a limit of continuous functions.

Consider the function χ_S where $S \subset X$ is measurable. We showed earlier that since S is measurable, for every $\epsilon > 0$ there exists a closed set B and an open set A with $B \subset S \subset A$ and $\mu(B - A) < \epsilon$. According to Urysohn's Lemma (`MATC27`) given any inclusion $B \subset A$ of subsets of X in which B is closed and A is open, there exists a continuous function $f : [a, b] \to [0, 1]$ such that $f(A) = 0$ and $f(B) = 1$. Hence for every positive integer n, there exists sets $A_n \subset B_n$ with $\mu(B_n - A_n) < 1/n$ and a continuous function $f_n : X \to [0,1]$ such that $f_n(A_n) = 0$ and $f_n(B_n) = 1$. Thus $\chi_S(x) = f_n(x)$ if $x \notin B_n - A_n$ and $|\chi_S(x) - f_n(x)| \leq 1$ if $x \in B_n - A_n$. Therefore

$$(\|\chi_S - f_n\|_p) = \left(\int_{B_n - A_n} |\chi_S(x) - f_n(x)|^p \, d\mu\right)^{1/p} \leq \big(\mu(B_n - A_n)\big)^{1/p} \leq \frac{1}{n^{1/p}}$$

and so $f_n \to \chi_S$ in the metric coming from the $\| \ \|_p$ norm. $\square$

Example. Let $A \subset [0,1]$ be the (generalized) Cantor set defined by

$$A = \{x \in [0,1] \mid x \text{ has a decimal expansion containing no 5's}\}.$$

and let $B = [0,1] - A$. Since B is a union of open sets it is open so it is a Borel set and thus measurable. If $I \subset [0,1]$ is any nontrivial subinterval,

$$\mu(B \cap I) = \mu(I)\left(\frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^3} + \ldots\right) = \mu(I)\frac{1}{10}\frac{1}{1 - (1/10)} = \frac{\mu(I)}{9} > 0$$

and $\mu(A) = \frac{8\mu(I)}{9} > 0$. The function χ_A belongs to $\mathcal{L}^\infty([0,1])$ with $\|\chi_A\|_\infty = 1$. Let $f : [0,1] \to \mathbb{F}$ be continuous. Let $S = \{x \in [0,1] \mid \|f(x) - \chi_A(x)\| < 1/2$. If $S = \emptyset$ then

$$\mu(\{x \in [0,1] \mid \|f(x) - \chi_A(x)\| \geq 1/2\}) = 1 > 0$$

and so $\|f - \chi_A\|_\infty \geq 1/2$. If $S \neq \emptyset$ then either there exists $x \in S \cap B$ or there exists $x \in S \cap A$. If there exists $x \in S \cap B$ then $|f(x)| < 1/2$ so by continuity there exists an interval I about x in which $|f(x)| < 1/2$. Then

$$\{x \in [0,1] \mid \|f(x) - \chi_A(x)\| \geq 1/2\} \supset A \cap I$$

so

$$\mu(\{x \in [0,1] \mid \|f(x) - \chi_A(x)\| \geq 1/2\}) \geq \mu(A \cap I) = 9\mu(I)/9 > 0$$

and so $\|f - \chi_A\|_\infty \geq 1/2$. Similarly, if there exists $x \in S \cap A$ then $|f(x)| > 1/2$ so by continuity there exists an interval I about x in which $|f(x)| > 1/2$. Then

$$\{x \in [0,1] \mid \|f(x) - \chi_B(x)\| \geq 1/2\} \supset B \cap I$$

so

$$\mu(\{x \in [0,1] \mid \|f(x) - \chi_A(x)\| \geq 1/2\}) \geq \mu(B \cap I) = \mu(I)/9 > 0$$

and so $\|f - \chi_A\|_\infty \geq 1/2$. Thus in all cases $\|f - \chi_A\|_\infty \geq 1/2$. But since $\|f - \chi_A\|_\infty \geq 1/2$ for any continuous function f, it is not possible to find a sequence of continuous functions which converges to χ_A in the $\| \ \|_\infty$ norm. Thus the set of continuous functions is not dense in $\mathcal{L}^\infty([0,1])$.

Part II: FUNCTIONAL ANALYSIS

7. Approximation

Recall the notation $\mathcal{C}(X;\mathbb{F}) := \{f : X \to \mathbb{F} \mid f \text{ is continuous and } |f| \text{ is bounded}\}$ and the "sup norm" on $\mathcal{C}(X;\mathbb{F})$ is defined by $\|f\| := \sup_{x\in X}\{|f(x)|\}$.

Certain subsets of $\mathcal{C}(X;\mathbb{F})$ may have particularly nice properties and it is often useful be able to approximate (in $\| \ \|$ norm) arbitrary functions by properties from this subset. The question of whether the subspace has this property is equivalent to asking if it is dense in the metric coming from the $\| \ \|$ norm. For example, if $X = \mathbb{R}$, we might be interesting in knowing whether or not an arbitrary continuous function can be approximated by a polynomial. In this section will we discuss a sufficient condition under which we can show that a subset A of $\mathcal{C}(X;\mathbb{F})$ has the property that $\bar{A} = \mathcal{C}(X;\mathbb{F})$, or equivalently that given arbitrary $f \in \mathcal{C}(X;\mathbb{F})$ for every $\epsilon > 0$ there exists $p \in A$ such that $\|f - p\| < \epsilon$.

Throughout this section we will assume X is a compact (i.e. closed and bounded) subset of $\mathbb{R}^n$ although more generally the results are valid whenever X is a "compact Hausdorff space" (defined in `MATC27`). Since X is compact, the condition "$|f|$ is bounded" in the definition of $\mathcal{C}(X;\mathbb{F})$ becomes redundant because continuous functions on compact sets are always bounded (`MATB43`).

We first consider the case $\mathbb{F} = \mathbb{R}$.

Lemma. *Let $f(t) = |t|$. Given $\epsilon > 0$, for any $c \geq 0$ there exists a polynomial $p(t)$ (depending on ϵ) such that $|f(t) - p(t)| < \epsilon$ for all $t \in [-c, c]$.*

Proof. First consider the special case $c = 1$. Let $T_N(t) := \sum_{k=0}^{N} a_k t^k$ be the Nth Taylor polynomial about 0 for $h(t) := (1-t)^{1/2}$. By examination of the remainder term one can check that this Taylor series converges uniformly to $h(t)$ for all $t \in [0, 1]$. (Recall from `MATA37` that this is not automatic: in general, a Taylor series need not converge to the function which gave rise to it.) Thus there exists N such that $|h(t) - T_N(t)| < \epsilon$ for all $t \in [0, 1]$. Set $p(t) := T_N(1 - t^2)$. Then $1 - t^2 \in [0, 1]$ for any $t \in [-1, 1]$ and it follows that $|f(t) - p(t)| < \epsilon$ for all $t \in [-1, 1]$.

Now let c be arbitrary. Given $\epsilon > 0$, by the special case there exists a polynomial $q(t)$ having the property that $|f(t) - q(t)| < \epsilon/c$ for all $t \in [-1, 1]$. Multiplying by c and using the fact that $f(Nz) = Nf(z)$ gives $|f(ct) < cq(t)| < \epsilon$ for all $t \in [-1, 1]$. Making the change of variable $x := ct$, shows that this is equivalent to saying $|f(x) < cq(x/c)| < \epsilon$ for all $x \in [-c, c]$. Therefore setting $p(x) := cq(x/c)$ gives a polynomial $p(x)$ having the property that $|f(x) - p(x)| < \epsilon$ for all $x \in [-c, c]$. $\square$

Define a partial order on $\mathcal{C}(X;\mathbb{R})$ by $f \leq g$ if and only if $f(x) \leq g(x)$ for all $x \in X$. (Note: The notation '$\leq$' is being used differently in this section than it was in the section on $\mathcal{L}^p$-spaces where it was based on $|f(x)|$ rather than on $f(x)$.)

Using the notation of Section 1, every pair f, g of elements in $\mathcal{C}(X;\mathbb{R})$ has a greatest lower bound, given by $(f \wedge g)(x) := \min\{f(x), g(x)\}$, and a least upper bound, given by $(f \vee g)(x) := \max\{f(x), g(x)\}$. That is, $\mathcal{C}(X;\mathbb{R})$ becomes a lattice under the partial order $\leq$.

Lemma. *Suppose X contains at least two points. Let L be a closed sublattice of $\mathcal{C}(X;\mathbb{R})$ having the property that for any $x \neq y \in X$ and any $a, b \in \mathbb{R}$ there exists $h \in L$ such that $h(x) = a$ and $h(y) = b$. Then $L = \mathcal{C}(X;\mathbb{R})$.*

Proof. Suppose $f \in \mathcal{C}(X;\mathbb{R})$. Given $\epsilon > 0$ we must show that there exists $g \in L$ such that $\|f - g\| < \epsilon$, or equivalently that $f(z) - \epsilon < g(z) < f(z) + \epsilon$ for all $z \in X$.

Pick $x \in X$. By hypothesis, for any $y \neq x$ there exists $f_y \in L$ such that $f_y(x) = f(x)$ and $f_y(y) = f(y)$.

Let $U_y = \{z \in X \mid f_y(z) < f(z) + \epsilon\}$. Then U_y is open with both $x \in U_y$ and $y \in U_y$. Since $\{U_y\}_{y \neq x}$ covers X which is compact, by Heine-Borel (`MATB43`) there exists a finite subcollection $\{U_{y_1}, \ldots, U_{y_k}\}$ such that $X = \{U_{y_1} \cup \ldots \cup U_{y_k}\}$. We write simply f_j for the function $f_{y_j} \in L$. Set $g_x := f_1 \wedge f_2 \wedge \ldots \wedge f_k \in L$. Observe that g_x has the property that $g_x(x) = x$ and $g_x(z) < f(z) + \epsilon$ for all $z \in X$.

Next set $V_x := \{z \in X \mid g_x(z) > f(z) - \epsilon\}$. Then V_x is open with $x \in V_x$. Since $\{V_x\}$ covers X which is compact there exists a finite subcollection $\{V_{x_1}, \ldots, U_{x_m}\}$ such that $X = \{V_{x_1} \cup \ldots \cup V_{x_m}\}$. We write simply g_j for the function $g_{x_j} \in L$. Set $g := g_1 \vee g_2 \vee \ldots \vee g_m \in L$. Then $g \in L$ and has the property that $f(z) - \epsilon < g(z) < f(z) + \epsilon$ for all $z \in X$. □

A subset $A \subset \mathcal{C}(X;\mathbb{F})$ is said to form an *algebra* over $\mathbb{F}$ if any scalar multiple, sum, difference, or product of elements of A lies in A.

Remark. This is the more common use of the term "algebra" mentioned in Section 1.

Lemma. *Let A be a subalgebra of $\mathcal{C}(X;\mathbb{R})$ which is closed (in the metric coming from the $\| \ \|$ norm). Then $|f| \in A$ for all $f \in A$.*

Proof. Suppose $f \in A$. Let $h(t) = |t|$. By an earlier Lemma, given $\epsilon > 0$ there exists a polynomial $q(t)$ such that $|h(t) - q(t)| < \epsilon/2$ for all $t \in \big[-\|f\|, \|f\|\big]$. Let $p(t) := q(t) - q(0)$ be the polynomial obtained by replacing the constant term of $q(t)$ by 0. For all $t \in \big[-\|f\|, \|f\|\,\big]$ we have

$$|h(t) - p(t)| = |h(t) - q(t) - q(0)| = |h(t) - q(t) + 0 - q(0)| \leq |h(t) - q(t)| + |0 - q(0)| = \epsilon. \quad (*)$$

Since $f \in A$ and A is an algebra, any polynomial in f lies in A, and in particular the function $g(x) := p(f)(x)$ lies in A. (By the notation $p(f)$ we mean the function $\sum_{j=1}^k c_j f^j$ where p is the polynomial $p(t) = \sum_{j=1}^k c_j t^j$.) Since $f(x) \in [-\|f\|, \|f\|\,]$ for all $x \in X$ it follows from $(*)$ that $|h\big(f(x)\big) - p\big(f(x)\big)| < \epsilon$ for all $x \in X$.

In other words, for all $\epsilon > 0$ we have found a function $p(f) \in A$ such that $\|\,|f| - p(f)\| < \epsilon$. Since A is closed, this implies $|f| \in A$. □

Using identities

$$f \vee g = \frac{f + g + |f - g|}{2} \qquad f \wedge g = \frac{f + g - |f - g|}{2}$$

we get

Corollary. *If A is a closed subalgebra of $\mathcal{C}(X;\mathbb{R})$ then A forms a sublattice of $\mathcal{C}(X;\mathbb{R})$.*

Definition. A set $S \subset \mathcal{C}(X;\mathbb{F})$ is said to *separate points* if for any $x \neq y \in X$ there exists $f \in S$ such that $f(x) \neq f(y)$.

Real Stone-Weierstrass Theorem. *Let P be a subalgebra of $\mathcal{C}(X;\mathbb{R})$ such that P separates points and contains the constant function $f(x) = 1$. Then P is dense in $\mathcal{C}(X;\mathbb{R})$.*

Proof. Set $A := \bar{P}$. We must show that $A = \mathcal{C}(X;\mathbb{R})$. It follows from the fact that limits commute with all algebra operations (addition, multiplication, etc.) that A is an algebra. Since the algebra A contains one constant function it contains them all. If X has only one point, then the only elements of $\mathcal{C}(X;\mathbb{R})$ are constant functions, so $A = \mathcal{C}(X;\mathbb{R})$ in this case. Therefore assume X has at least two points. By an earlier Lemma, it suffices to show that A has the property that given $x \neq y \in X$ and $a, b \in \mathbb{R}$ there exists $f \in A$ such that $f(x) = a$ and $f(y) = b$. Given $x \neq y$, since A separates points, there exists $g \in A$ such that $g(x) \neq g(y)$. Then $f(z) := a\frac{g(z)-g(y)}{g(x)-g(y)} + b\frac{g(z)-g(x)}{g(y)-g(x)}$ has the desired property. □

Remark. Some books include the condition $1 \in P$ as part of the definition of an "algebra". Using that convention, the hypothesis that P contains the function 1 is redundant since it is included in the statement that P forms an algebra.

Complex Stone-Weierstrass Theorem. *Let P be a subalgebra of $\mathcal{C}(X;\mathbb{C})$ such that P separates points, contains the constant function $f(x) = 1$, and contains the complex conjugate of each of its elements. Then P is dense in $\mathcal{C}(X;\mathbb{C})$.*

Proof. Set $A := \bar{P}$. We must show that $A = \mathcal{C}(X;\mathbb{C})$. If X has only one point the theorem is trivial, so assume X has at least two points. Let $B = \{f \in A \mid f(X) \subset \mathbb{R}\}$. Since $1 \in A$ it is clear that $1 \in B$. Suppose $x \neq y \in X$. Since A separates points, there exists $f \in A$ such that $f(x) \neq g(y)$. Therefore either $(\operatorname{Re} f)(x) \neq (\operatorname{Re} f)(y)$ or $(\operatorname{Im} f)(x) \neq (\operatorname{Im} f)(y)$. Observe that $\operatorname{Re}(f) = (f + \bar{f})/2$ and $\operatorname{Im}(f) = (f - \bar{f})/(2i)$ and therefore the hypothesis that $\bar{f} \in A$ implies that $\operatorname{Re} f \in B$ and $\operatorname{Im} f \in B$. Hence B separates points and so by the Real Stone-Weierstass Theorem, $B = \mathcal{C}(X;\mathbb{R})$.

For arbitrary $g \in \mathcal{C}(X;\mathbb{C})$ write $g = \operatorname{Re} g + i \operatorname{Im} g$. Since $\operatorname{Re} g, \operatorname{Im} g \in \mathcal{C}(X;\mathbb{R}) = B \subset A$, and A is an algebra, the linear combination $g = \operatorname{Re} g + i \operatorname{Im} g$ belongs to A. Thus $A = \mathcal{C}(X;\mathbb{C})$. □

Since we showed earlier that the continuous functions are dense in $\mathcal{L}^p([a,b])$ in both real and complex cases we get immediately

Corollary. *Let P be a subalgebra of $\mathcal{C}([a,b];\mathbb{F})$ such that P separates points, contains the constant function $f(x) = 1$, and contains the complex conjugate of each of its elements. Then P is dense in $\mathcal{L}^p([a,b];\mathbb{F})$.* □

The preceding theorems show existence of uniform approximations and are useful for theoretical purposes. However they do not tell the full story since sometimes actually finding an appropriate approximating function is needed.

Examples.

1) The polynomials are dense in $\mathcal{C}(X;\mathbb{R})$. In other words, any continuous function on a compact subset of $\mathbb{R}^n$ can be uniformly approximated by a polynomial. Notice that

this uniform approximation is significantly different from the convergence of the Taylor series of an analytic function to the function. For one thing, not every continuous function is analytic; in fact, a function has to be infinitely differentiable to even **have** a Taylor series and not all continuous functions are differentiable. However even if f is analytic, its Taylor series does not usually converge uniformly. A Taylor series forms a good approximation only in the vicinity of the point at which the expansion takes place.

2) The "trigonometric polynomials", $\mathbb{C}[\{e^{2\pi nt}\}_{n\in\mathbb{Z}}]$, are dense in $\mathcal{C}([0,1];\mathbb{C})$. This suggests that Fourier series might be useful in providing a uniform approximation. This is only partly true however. For a differentiable function f, Dini's Theorem says that its Fourier series does indeed converge uniformly to f. In fact more generally, Dini's Theorem states that is suffices to know that there exist positive constants M and α such that f satisfies Hölder's inequality:

$$\|f(x) - f(y)\| \leq M\|x - y\|^{\alpha} \text{ for all } x, y \in X.$$

(If f is differentiable, the inequality is satisfied with $\alpha = 1$.) However although the Stone-Weierstrass guarantees that every continuous function can be uniformly approximated by trigonometric polynomials, there is no guarantee that its Fourier series will be the appropriate functions in general.

3) According to (1) above, every continuous function f has an approximation by polynomials, but its Taylor polynomials are not appropriate. Instead, (2) suggests that it might be better to look for polynomials whose relation with f is more analogous to a Fourier expansion than a Taylor expansion. As discussed in `MATC46`, a Sturm-Liouville problem produces a collection of eigenfunctions into which functions can be expanded in a generalized Fourier series. Some of the differential equations have eigenfunctions which are polynomials so under some conditions one might get polynomials which converge uniformly by considering expansions into, for example, Chebyshev polynomials.

8. Hilbert Space

In this section, let $\mathbb{F}$ equal $\mathbb{C}$ unless stated otherwise. The modifications for the case $\mathbb{F} = \mathbb{R}$ are straightforward.

Definition. An *inner product* on a (complex) vector space V consists of a function $(\ ,\ ) : V \times V \to \mathbb{C}$ such that

1) $(ax + by, z) = a(x, z) + b(y, z)$ for all $x, y, z \in V,\ a, b \in \mathbb{C}$
2) $(x, y) = \overline{(y, x)}$ for all $x, y \in V$

An inner product is called *positive definite* if it also satisfies

3) $(x, x) \geq 0$ for all $x \in V$ and $(x, x) = 0$ if and only if $x = 0$.

A (complex) *inner product space* consists of a (complex) vector space V together with a positive definite inner product on V.

Remark 1. In many books, the inner product of x and y is written as $\langle x, y\rangle$ rather than (x, y).

Remark 2. It follows from the given properties that $(x, y + z) = (x, y) + (x, z)$ and $(x, ay) = \bar{a}(x, y)$. Physicists tend to use the opposite convention and require instead $(x, ay) = a(x, y)$ so that $(ax, y) = \bar{a}(x, y)$ under their convention.

Remark 3. For any inner product, $(x, x) \in \mathbb{R}$ as a consequence of (2) but a positive definite inner product further restricts its values to the nonnegative reals.

Notation. Set $\|x\| := (x, x)^{1/2}$.

Many of the familiar properties of Euclidean space carry over to inner product spaces.

Proposition (Cauchy-Schwartz). *In an inner product space,* $|(x, y)| \leq \|x\|\,\|y\|$

Proof. If $x = 0$ or $y = 0$ the inequality is trivial so suppose $x \neq 0$ and $y \neq 0$. Dividing both sides by $\|x\|\,\|y\|$ reduces the Proposition to showing that $|(x, y)| \leq 1$ whenever $\|x\| = \|y\| = 1$ so suppose now that $\|x\| = \|y\| = 1$. Since the inner product is positive definite $\|(x, y)y - x\|^2 \geq 0$. However

$$\begin{aligned}
\|(x, y)y - x\|^2 &= \big((x, y)y - x, (x, y)y - x\big) \\
&= \big((x, y)y, (x, y)y\big) - \big((x, y)y, x\big) - \big(x, (x, y)y\big) + (x, x) \\
&= (x, y)\overline{(x, y)}(y, y) - (x, y)(y, x) - \overline{(x, y)}(x, y) + (x, x) \\
&= |(x, y)|^2\|y\|^2 - (x, y)\overline{(x, y)} - \overline{(x, y)}(x, y) + \|x\|^2 \\
&= |(x, y)|^2 1^2 - |(x, y)|^2 - |(x, y)|^2 + 1^2 = -|(x, y)| + 1.
\end{aligned}$$

Thus $|(x, y)| \leq 1$ as desired. □

Corollary (Triangle Inequality). *In an inner product space,* $\|x + y\| \leq \|x\| + \|y\|$.

Proof.

$$\begin{aligned}
\|x + y\|^2 &= (x + y, x + y) = (x, x) + (x, y) + (y, x) + (y, y) \\
&\leq \|x\|^2 + \|x\|\|y\| + \|x\|\|y\| + \|y\|^2 \\
&= \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.
\end{aligned}$$

Taking the square root of both sides yields the desired inequality. □

Since the other properties required of a norm are trivial from the definition, it follows that the assignment $\|v\| := (v,v)^{1/2}$ yields a norm on a positive definite inner product on V.

Definition. A *Hilbert space* is an inner product space which forms a Banach space under the norm coming from the inner product.

Examples.

1) $V = \mathbb{C}^n$ with $(x,y) := x \cdot y := \sum_{i=1}^n x_i \overline{y_i}$. The norm coming from the inner product becomes the standard norm on $\mathbb{C}^n$ yielding the standard distance in $\mathbb{C}^n$, and $\mathbb{C}^n$ is complete in this metric so forms a Hilbert space.
2) $V = \mathcal{L}^2(X;\mathbb{C})$ with $(f,g) = \int f\bar{g}\,d\mu$. Since $(f,f) = \int f\bar{f}\,d\mu = \int |f|^2\,d\mu = (\|f\|_2)^2$, the norm coming from this inner product is the $\|\ \|_2$ which we previously showed forms a Banach space. Therefore $\mathcal{L}^2(X;\mathbb{C})$ forms a Hilbert space.
3) Consider the special case of (2) in which $X = \{0,1,2,\dots,\}$ and $\mu(S) = \mathrm{Card}(S)$. In this case, the Hilbert space $\mathcal{L}^2(X;\mathbb{C})$ is denoted l^2. Observe that a function on X is equivalent to a sequence $(a_n)_{n=0}^{\infty}$ and the condition that the correspond function lies in $\mathcal{L}^2(X;\mathbb{C})$ is equivalent to requiring that $\sum_{n=1}^{\infty}|a_n|^2 < \infty$.

Notation. We write $x \perp y$ to mean $(x,y) = 0$. If A is a subset of an inner product space X, we set $A^{\perp} := \{x \in X \mid x \perp a \text{ for all } a \in A\}$.

Theorem (Pythagoras). *If $x \perp y$ then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.*

Proof. Expand $\|x+y\|^2 = (x+y,x+y)$ and set $(x,y) = (y,x) = 0$. □

Theorem.

1) *(Parallelogram Law).*
Let V be an inner product space. Then $\|x-y\|^2 + \|x+y\|^2 = 2(\|x\|^2 + \|y\|^2)$ for all $x, y \in V$.
2) *(Jordan - von Neumann Theorem)*
Let V be a normed vector space such that $\|x-y\|^2 + \|x+y\|^2 = 2(\|x\|^2 + \|y\|^2)$ for all $x, y \in V$. Then on V there exists a positive definite inner product whose associated norm is the given norm on V.

Proof.

1)

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

2) Observe that expanding shows that in any inner product space

$$\|x+y\|^2 - \|x\|^2 - \|y\|^2 = (x,y) + (y,x) = (x,y) + \overline{(x,y)} = 2\,\mathrm{Re}(x,y).$$

Either by expanding or by substitution using $\|iy\|^2 = \|y\|^2$ we get

$$\|x+iy\|^2 - \|x\|^2 - \|y\|^2 = 2\,\mathrm{Re}(x,iy) = 2\,\mathrm{Re}\,\bar{i}(x,y) = -2\,\mathrm{Re}\,i(x,y) = 2\,\mathrm{Im}(x,y).$$

It follows that in any inner product space

$$2(x,y) = \|x+y\|^2 - \|x\|^2 - \|y\|^2 + i(\|x+iy\|^2 - \|x\|^2 - \|y\|^2).$$

With this as motivation, given an norm on V which satisfies the Parallelogram Law we define $(\ ,\ ) : V \to \mathbb{C}$ by

$$(x,y) := \big(\|x+y\|^2 - \|x\|^2 - \|y\|^2 + i(\|x+iy\|^2 - \|x\|^2 - \|y\|^2)\big)/2.$$

Then

$$\begin{aligned}(x,x) &= \big(\|2x\|^2 - \|x\|^2 - \|x\|^2 + i(\|1+i)x\|^2 - \|x\|^2 - \|x\|^2\big)/2 \\ &= \big(4\|x\|^2 - 2\|x|^2 + i(2\|x\|^2 - 2\|x\|^2)\big)/2 = (2\|x\|^2 - 0)/2 = \|x\|^2.\end{aligned}$$

Calculation shows that if the Parallelogram Law is satisfied then $(\ ,\ )$ satisfies the properties required to be an inner product. (Exercise.) □

A set $\{e_j\}_{j\in J}$ in an inner product space V is called *orthonormal* if $e_i \perp e_j$ for all $i \neq j \in J$ and $\|e_j\| = 1$ for all $j \in J$.

Theorem (Bessel's Inequality). *Let $\{e_1, \ldots, e_n\}$ be an orthonormal set in an inner product space V. Then $\sum_{j=1}^{n} |(x,e_j)|^2 \leq \|x\|^2$ for any $x \in V$.*

Proof.

$$\begin{aligned}0 \leq \left\| x - \sum_{j=1}^{n}(x,e_j)e_j \right\|^2 &= \left(x - \sum_{j=1}^{n}(x,e_j)e_j, x - \sum_{j=1}^{n}(x,e_j)e_j \right) \\ &= \|x\|^2 - \sum_{j=1}^{n}\big(x,(x,e_j)e_j\big) - \sum_{j=1}^{n}(x,e_j)(e_j,x) + \sum_{j=1}^{n}\big((x,e_j)e_j,(x,e_j)e_j\big) \\ &= \|x\|^2 - 2\sum_{j=1}^{n}|(x,e_j)|^2 + \sum_{j=1}^{n}|(x,e_j)|^2 = \|x\|^2 - \sum_{j=1}^{n}|(x,e_j)|^2\end{aligned}$$

□

Let $\mathcal{H}$ be a Hilbert space. A maximal orthonormal subset of $\mathcal{H}$ is called a *Hilbert space basis* for $\mathcal{H}$. Zorn's Lemma (equivalent to the Axiom of Choice) shows that every Hilbert space has a basis.

Proposition. *Let $\{e_j\}_{j\in J}$ be a Hilbert space basis for $\mathcal{H}$. If $x \perp e_j$ for all $j \in J$ then $x = 0$.*

Proof. If $x \neq 0$ then $\{e_j\}_{j\in J} \cup \left\{\frac{x}{\|x\|}\right\}$ is an othornormal subset of $\mathcal{H}$ which properly contains $\{e_j\}_{j\in J}$ contradicting the maximality of $\{e_j\}_{j\in J}$. □

If $\{e_j\}_{j\in J}$ is a Hilbert space basis for a Hilbert space $\mathcal{H}$, the complex numbers $(x,e_j)_{j\in J}$ are called the *Fourier coefficients* of x with respect to the basis $\{e_j\}_{j\in J}$.

Proposition. *Let $\{e_j\}_{j \in J}$ be a Hilbert space basis for $\mathcal{H}$ and suppose $x \in \mathcal{H}$. Then at most countably many of the Fourier coefficients (x, e_j) are nonzero.*

Proof. $\{j \in J \mid (x, e_j) \neq 0\} = \cup_n A_n$ where $A_n := \{j \in J \mid |(x, e_j)|^2 > 1/2^n\}$. According to Bessel's Inequality, it is not possible to find a subset of A_n containing more than $2^n \|x\|^2$ elements, so in particular A_n is finite for all n. The union of a countable collection of finite sets is countable so x as at most countably many nonzero Fourier coefficients. $\square$

Proposition. *Let $\{e_j\}_{j \in J}$ be a Hilbert space basis for $\mathcal{H}$ and suppose $x \in \mathcal{H}$. Let $\{e_{j_1}, e_{j_2}, \ldots, e_{j_n}, \ldots\}$ be the (countable) set of all e_j such that $(x, e_j) \neq 0$. Then $x = \sum_{n=1}^{\infty} (x, e_{j_n}) e_{j_n}$.*

Remark. Recall that a series is defined as convergent if its sequence of partial sums converges. Changing the order of the terms in the summation changes the sequence of partial sums and can, in general, change the limit of the sequence. Since x determines only the set $\{e_{j_n}\}$ and not its order, the Proposition says that in this case the order is irrelevant: the series always converges to x regardless of the order of the terms in the summation.

Proof of Proposition. After picking an ordering of the set of nonzero Fourier coefficients of x, let $s_n = \sum_{i=1}^{n} (x, e_{j_i}) e_{j_i}$. We wish to show that $(s_n) \to x$. Pick an integer m. If $n > m$ then

$$\|s_n - s_m\|^2 = \left\| \sum_{i=m+1}^{n} (x, e_{j_i}) e_{j_i} \right\|^2 = \sum_{i=m+1}^{n} |(x, e_{j_i})|^2 \leq \|x\|^2$$

by Bessel's inequality. Therefore the monotonically increase sequence of real numbers $(\|s_n - s_m\|^2)_{n=m}^{\infty}$ is bounded so it is a Cauchy sequence. That is, given $\epsilon > 0$, there exists N such that $\|s_{n'} - s_m\|^2 - \|s_n - s_m\|^2 < \epsilon$ whenever $n, n' > N$. However $\|s_{n'} - s_m\|^2 - \|s_n - s_m\|^2 = \|s_{n'} - s_n\|^2$ and so this says that (s_n) forms a Cauchy sequence in $\mathcal{H}$. Since $\mathcal{H}$ is complete, there exists $y \in \mathcal{H}$ such that $(s_n) \to y$.

It remains to show that $y = x$. For any $j \in J$,

$$(y - x, e_j) = (y, e_j) - (x, e_j) = \lim_{n \to \infty} (s_n, e_j) - (x, e_j).$$

If j not in the set of nonzero Fourier coefficients of x, then $(x, e_j) = 0$ and $(s_n, e_j) = 0$ for all n and so $(y - x, e_j) = 0$. If $j = j_k$ for some k in the set of nonzero Fourier coefficients of x, then since $(s_n, e_{j_k}) = (x, e_{j_k})$ for all $n \geq k$, we again get $(y - x, e_j) = 0$. Thus $(y - x) \perp e_j$ for all $j \in J$ and so $y - x = 0$ by an earlier Proposition. $\square$

Notation. We write $x = \sum_{j \in J} (x, e_j) e_j$ with the understanding that at most countably many terms in the sum are nonzero and that the sum of the series is independent of the order in which that countable collection of nonzero terms is summed.

Corollary (Parseval's Identity). *Let $\{e_j\}_{j \in J}$ be a Hilbert space basis for $\mathcal{H}$ and suppose $x, y \in \mathcal{H}$. Let $\{e_{j_1}, e_{j_2}, \ldots, e_{j_n}, \ldots\}$ be the (countable) set of all e_j such that $(x, e_j) \neq 0$. Then $(x, y) = \sum_{j \in J} (x, e_j) \overline{(y, e_j)}$. In particular, $\|x\|^2 = \sum_{j \in J} |(x, e_j)|^2$.* $\square$

Remark. A Hilbert space is also a vector space and as such it has a vector space basis as well as a Hilbert space basis. However a Hilbert space basis is not (in general) a vector space basis. The difference is that every element of a vector space can be written uniquely as a **finite** linear combination of basis elements, whereas in a Hilbert space every element is a **countable infinite** linear combination of basis elements where convergence is used to define the meaning of such a sum. Thus a Hilbert space basis will (in general) be much smaller than a vector space basis for the same space. Of course, for an arbitrary vector space, there is no concept of a Hilbert space basis since there is no metric in which to discuss the issue of convergence. From now on, when discussing a Hilbert space, the term *basis* will refer to a Hilbert space basis (rather than a vector space basis) unless specified otherwise.

Theorem. $\{e^{2\pi nx}\}_{n\in\mathbb{Z}}$ *forms an orthonormal basis for* $\mathcal{L}^2([0,1])$.

Proof. Write e_n for $e^{2\pi nx}$. First be show that $\{e_n\}$ forms an orthonormal set. If $n \neq m$,

$$\begin{aligned}(e_n, e_m) &= \int_0^1 e^{-2\pi nx}e^{2\pi mx}\,dx \int_0^1 e^{2\pi(n-m)x}\,dx \\ &= \left[\frac{e^{2\pi i(n-m)x}}{2\pi i(n-m)}\right]_0^1 = \frac{e^{2\pi i(n-m)}}{2\pi i(n-m)} - \frac{1}{2\pi i(n-m)} = 0.\end{aligned}$$

and

$$(e_n, e_n) = \int_0^1 e^{-2\pi nx}e^{2\pi nx}\,dx = \int_0^1 1\,dx = 1.$$

Therefore $\{e_n\}$ forms an orthonormal set.

To show that $\{e_n\}$ is a basis, we must show that if $f \in \mathcal{L}^2([0,1])$ such that $(f, e_n) = 0$ for all n then $f = 0$. Therefore suppose $f \in \mathcal{L}^2([0,1])$ such that $(f, e_n) = 0$ for all n. Let W be the linear span of $\{e_n\}_{n\in\mathbb{Z}}$ and let $\bar{W}$ denote the closure of W in the metric coming from the $\|\ \|_2$ norm. For $g \in \bar{W}$ there exists $g_n \in W$ such that $(g_n) \to g$ in the $\|\ \|_2$ norm. Since $(f, e_n) = 0$ for all n, we get $(f, h) = 0$ for all $h \in W$. In particular, $(f, g_n) = 0$ for all n. Using Cauchy-Schwartz, we have

$$|(f, g) - (f, g_n)| = |(f, g - g_n)| \leq \||f\||_2 \|g - g_n\|_2.$$

Since $(\|g - g_n\|_2) \to 0$, this implies that $(f, g) = \lim_{n\to\infty}(f, g_n) = 0$.

Note that $e_n e_k = e_{n+k}$ and so the set $\{e_n\}$ is closed under multiplication. Therefore W is closed under multiplication and so forms a subalgebra of $\mathcal{C}([0,1];\mathbb{C})$. However W does not separate the points 0 and 1. The unit circle S^1 can be regarded as a line segment with endpoints identified. (In the the terminology of `MATC27` there is a homeomorphism $\big([0,1]/(\{0\} \amalg \{1\})\big) \cong S^1$.) Under this correspondence we have $\mathcal{C}(S^1;\mathbb{C}) = \{k \in \mathcal{C}([0,1];\mathbb{C}) \mid k(0) = k(1)\}$. Applying the Stone-Weierstrass Theorem with $X := S^1$ gives that every continuous function $k(x)$ such that $k(0) = k(1)$ can be uniformly approximated by elements of W. That is, there exist a sequence (k_n) in W such that $(k_n) \to k$ in the metric coming from the sup norm $\|\ \|$. As we showed earlier, for spaces of finite measure, convergence in sup norm is stronger than convergence in $\|\ \|_p$ norm, so

$(k_n) \to k$ in the metric coming from the $\| \ \|_2$ norm. Since $k_n \in W$, we have $(f, k_n) = 0$ for all n, as as above it follows that $(f, k) = 0$. Thus we have shown that $(f, k) = 0$ for any continuous function k such that $k(0) = k(1)$.

We showed earlier that for any p, the continuous functions on a compact subset of $\mathbb{R}^N$ are dense in the metric coming from the $\| \ \|_p$ norm. Since $\mathcal{L}^2([0,1]) = \mathcal{L}^2(S^1)$ (changing the value of a function at the point 1 so that $k(1) = k(0)$ does not change its equivalence class in $\mathcal{L}^2([0,1])$), applying this to S^1 shows that the subset of continous functions $k(x)$ for which $k(0) = k(1)$ are dense. Therefore for any $\phi \in \mathcal{L}^2([0,1])$ there exists a sequence there exists a sequence of continuous functions converging to ϕ in the $\| \ \|_2$ norm and thus we conclude that $(f, \phi) = 0$ for all $\phi \in \mathcal{L}^2([0,1])$. In particular $(f, f) = 0$. But if $\|f\|^2 = (f, f) = 0$ then $f = 0$. $\square$

If A, B are subsets of a metric space, we define the *distance* from A to B, denoted $d(A, B)$ by $d(A, B) := \inf\{d(a, b) \mid a \in A, b \in B\}$.

Obviously if $A \cap B \neq \emptyset$ then $d(A, B) = 0$, but observe that even if A and B are disjoint closed subsets, it is possible that $d(A, B) = 0$. For example if $A = x$-axis $\subset \mathbb{R}^2$ and B is the graph of $y = 1/x$ in $\mathbb{R}^2$ then $d(A, B) = 0$ even though A, B are closed and disjoint.

Remark. If A and B are disjoint closed subsets and either A or B is compact, then $d(A, B) > 0$.

Definition. A subset A of a real or complex vector space is called *convex* if for every pair of elements $x, y \in A$ the line segment joining x to y lies in A. That is, if $x, y, \in A$ then $tx + (1 - t)y$ for all $t \in [0, 1]$.

Closest Point Lemma. *If C is a closed convex closed convex subset of a Hilbert space and x is a vector not in C then $d(\{x\}, C) > 0$ and there is a unique vector in C which is closest to x.*

Proof. Let $\delta = d(\{x\}, C) = \inf_{c \in C} \|x - c\|$. Then there exists a sequence (c_n) in C such that $(\|x - c_n\|) \to \delta$ in $\mathbb{R}$. Using the Parallelogram Law we get

$$\begin{aligned}
\|c_m - c_n\|^2 &= \|(x - c_n) - (x - c_m)\|^2 \\
&= 2\|x - c_n\|^2 + 2\|x - c_m\|^2 - \|(x - c_n) + (x - c_m)\|^2 \\
&= 2\|x - c_n\|^2 + 2\|x - c_m\|^2 - 4\|x - (c_n + c_m)/2\|^2.
\end{aligned}$$

Since C is convex, $(c_n + c_m)/2 \in C$ and therefore $\delta \leq \|x - (c_n + c_m)/2\|^2$ by definition of δ. Thus the previous equation implies

$$\|c_m - c_n\|^2 = \|(x - c_n) - (x - c_m)\|^2 \leq 2\|x - c_n\|^2 + 2\|x - c_m\|^2 - 4\delta^2$$

and taking the limit as $m, n \to \infty$ gives $0 \leq \lim_{m,n\to\infty}(\|c_m - c_n\|^2) \leq 2\delta^2 + 2\delta^2 - 4\delta = 0$ resulting in $\lim_{m,n\to\infty}(\|c_m - c_n\|^2) = 0$ by the Squeezing Principle. Therefore c_n is a Cauchy sequence in C and since C is closed there exists $c \in C$ such that $c_n \to c$. Then $\|x - c\| = \lim_{n\to\infty} \|x - c_n\| = \delta$. Thus there is no point in C which is closer to x than c is. If also follows that $\delta > 0$ since if $\delta = 0$ then $x = c$ contradicting the hypothesis. (The fact that $d(\{x\}, C) > 0$ could also have been deduced from the fact that $\{x\}$ is compact.)

Suppose c' in C also has the property that $\|x-c'\| = \delta$. Consider the sequence $(b_n) = (c_1, c', c_2, c', c_3, c', \ldots)$ in C formed by alternating terms of the sequence (c_n) with the constant sequence (c'). Then $\lim_{n\to\infty} \|x - b_n\| = \delta$, so, as above, the sequence (b_n) converges. However (b_n) has a subsequence converging to c and also a subsequence converging to c'. Therefore $c = c'$. □

Recall that $A^\perp$ denotes $\{x \in X \mid x \perp a$ for all $a \in A$. We write $A^{\perp\perp}$ for $(A^\perp)^\perp$.

Proposition. *Let A be a subset of a Hilbert space $\mathcal{H}$ (not necessarily a subspace, i.e. A might not be a vector space). Then*
1) $A^\perp$ is a closed subspace of $\mathcal{H}$
2) $A \subset A^{\perp\perp}$
3) $A \subset B$ implies $B^\perp \subset A^\perp$
4) $A^\perp = A^{\perp\perp\perp}$
5) If M is a closed subspace then $M + M^\perp = \mathcal{H}$ and $M \cap M^\perp = 0$. Thus $\mathcal{H} \cong M \oplus M^\perp$ as vector spaces.
6) $A^{\perp\perp}$ is the smallest closed subspace that contains A. In particular, if M is a closed subspace then $M^{\perp\perp} = M$

Proof. (1)–(4) are easy and (6) follows from the earlier properties. We will prove (5).

The equation $M \cap M^\perp = 0$ is trivial. Let x belong to $\mathcal{H}$. Let $m_0 \in M$ be the unique vector in M which is closest to x. Write $x = m_0 + (x - m_0)$. We will show that $x - m_0 \in M^\perp$. This suffices to prove (5) since it demonstrates that $x \in M + M^\perp$ for all $x \in \mathcal{H}$, so that $M + M^\perp = \mathcal{H}$.

For all $m \in M$, $\|x - m_0\| \le \|x - (m_0 + m)\|$. Squaring and expanding gives

$$\begin{aligned}&(x,x) - (x,m_0) - (m_0,x) + (m_0,m_0)\\ &\le (x,x) - (x,m_0) - (x,m) - (m_0,x) - (m,x) + (m_0,m_0) + (m_0,m) + (m,m_0) + (m,m).\end{aligned}$$

Therefore

$$\begin{aligned}0 &\le -(x,m) - (m,x) + (m_0,m) + (m,m_0) + (m,m)\\ &= (m_0 - x, m) + (m, m_0 - x) + (m,m)\\ &= 2\,\mathrm{Re}(m_0 - x, m) + \|m\|^2. \qquad (*)\end{aligned}$$

Since M is a subspace, the inequality $(*)$ holds for tm for all $t \in \mathbb{C}$. Consider the special case where $t \in \mathbb{R}$. Define a quadratic polynomial $q(t) : \mathbb{R} \to \mathbb{R}$ by

$$q(t) = 2\,\mathrm{Re}(m_0 - x, tm) + \|tm\|^2 = 2\,\mathrm{Re}(m_0 - x, m)t + \|m\|^2 t^2.$$

According to $(*)$, $q(t) \ge 0$ for all $t \in \mathbb{R}$. However if $at^2 + bt \ge 0$ for all t its graph must meet the t-axis only once, so $b = 0$. Thus $\mathrm{Re}(m_0 - x, m) = 0$ for all $m \in M$. Write the complex number $(m_0 - x, m) = 0$ in the polar form $(m_0 - x, m) = re^{i\theta}$ where $r = |(m_0 - x, m)| \in \mathbb{R}$. Since $e^{i\theta}m$ also lies in M, the preceding conclusion applies to it giving

$$0 = \mathrm{Re}(m_0 - x, e^{i\theta}m) = \mathrm{Re}\, e^{-i\theta}(m_0 - x, m) = \mathrm{Re}\, r = r = |(m_0 - x, m)|$$

for all $m \in M$. In other words $m_0 - x \in M^\perp$ as desired.

Thus every $x \in \mathcal{H}$ can be written in the form $x = m_0 + (x - m_0)$ □

If M is a closed subspace then the *projection* $P_M : \mathcal{H} \to M$ of $\mathcal{H}$ onto M is defined as the composition $\mathcal{H} \cong M \oplus M^\perp \to M$. Thus $P_M(x)$ is the unique element of M such that $x - P_M(x) \in M^\perp$.

9. Continuous Linear Transformations of Banach Spaces

In studying Banach spaces, subvector spaces of the Banach space which are not closed are rarely of much interest. For this reason, some books use the term "subspace" to refer only to a closed subspace and use the term "linear manifold" to refer to a subspace which is not necessarily closed. This can be doubly confusing since not only because the reader might misinterpret their use of the term subspace but also because this use of the word "manifold" does not agree with its use in other parts of mathematics. In these notes we will instead write "closed linear subspace" (sometimes abbreviated to "closed subspace") and use "vector subspace" or "linear subspace" for one which is not necessarily closed, and we will avoid the use of the term "linear manifold".

In this section we will again assume $\mathbb{F} = \mathbb{C}$ unless stated otherwise.

Proposition. *Let $\mathcal{N}$ and $\mathcal{N}'$ be normed vector spaces (not necessarily Banach spaces). A linear transformation $\phi : \mathcal{N} \to \mathcal{N}'$ is continuous if and only if there exists (real) M such that $|\phi(x)| \le M$ for all $x \in B$ with $\|x\| = 1$.*

Proof. Suppose there exists M such that $|\phi(x)| \le M$ whenever $\|x\| = 1$. Let $(x_n) \to x$ be a convergent sequence in $\mathcal{N}$. We wish to show $\bigl(\phi(x_n)\bigr) \to \phi(x)$ in $\mathcal{N}'$. Since ϕ is linear, by subtracting x it suffices to consider the case where $(x_n) \to 0$. Again using ϕ linear, whenever $x_n \ne 0$, since $\bigl\|y/\|y\|\bigr\| = 1$ we get $\|\phi(x_n)\| = \|x_n\| \, \|\phi(x/\|x_n\|)\| \le \|x_n\|M$. Furthermore, if $x_n = 0$ then $\phi(x_n) = 0$ so $\phi(x_n) \le \|x_n\|M$ also holds in this case. Thus $\|\phi(x_n)\| \le \|x_n\|M$ for all n. Hence taking the limit as $n \to \infty$ gives $\|\phi(x_n)\| \to 0$ and so $\phi(x_n) \to 0$, as required.

Conversely, suppose ϕ is continuous. If there does not exists M such that $|\phi(x)| \le M$ whenever $\|x\| = 1$ then for each integer $n > 0$ there exists $y_n \in \mathcal{N}$ such that $\|y_n\| = 1$ and $|\phi(y_n)| > n$. Let $x_n = y_n/n$. Then $\|x_n\| = \|\phi(y_n)\|/n = 1/n$ and $\|\phi(x_n)\| = \|\phi(y_n)\|/n > 1$. However this contradicts the continuity of ϕ since the first condition says that $(x_n) \to 0$ in $\mathcal{N}$ while the second implies that $\bigl(\phi(y_n)\bigr) \not\to 0$ in $\mathcal{N}'$. $\square$

Definition. Let $\mathcal{N}$ and $\mathcal{N}'$ be normed vector spaces (not necessarily Banach spaces). A linear transformation $\phi : \mathcal{N} \to \mathcal{N}'$ is called *bounded* if there exists M such that $|\phi(x)| \le M$ whenever $\|x\| = 1$, or equivalently if ϕ is continuous. If ϕ is a bounded linear transformation, we set $\|\phi\| := \sup\{\|\phi(x)\| \mid x \in \mathcal{N} \text{ with } \|x\| = 1\}$.

Remark. Although this is similar to the definition of the sup norm and we are using the same notation for it and we even call it the sup norm, strictly speaking the sup norm would be defined by $\inf\{\|\phi(x)\| \mid x \in \mathcal{N}\}$. However since $\|\phi(Nx)\| = N\|\phi(x)\|$ the result would always be ∞ unless $\mathcal{N} = 0$. Since $\sup\{\|\phi(x)\| \mid x \in \mathcal{N}\}$ is therefore a useless concept in this context, we have stolen its name and notation for the related concept $\sup\{\|\phi(x)\| \mid x \in \mathcal{N} \text{ with } \|x\| = 1\}$.

Definition and Notation. We write $B(\mathcal{N}, \mathcal{N}')$ for the set of bounded linear transformations from $\mathcal{N}$ to $\mathcal{N}'$. A linear transformation from $\mathcal{N}$ to itself is called a *linear operator*.

We write simply $B(\mathcal{N})$ for the set of bounded linear operators on $\mathcal{N}$. A linear transformation from $\mathcal{N}$ to $\mathbb{F}$ is called a *linear functional.* We write $\mathcal{N}^*$ for the set of bounded linear functionals on $\mathcal{N}$, and refer to it as the *(Banach) dual space* of $\mathcal{N}$.

Remark. Although are using the same terminology and notation, the dual space of $\mathcal{N}$ in the sense of vector spaces is larger than its dual space as a normed vector space, since the linear functionals in the vector space dual are not required to be bounded.

Example. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\mathcal{H} = \mathcal{L}^2(X)$. Given $f \in \mathcal{H}$, there is a corresponding "multiplication operator", $M_f \in B(\mathcal{H})$ defined by $M_f(g) = fg$ where $(fg)(x) := f(x)g(x)$.

Lemma. *For x, y in a Banach space, the inequality $\big|\|x\| - \|y\|\big| \leq \|x - y\|$ holds in $\mathbb{R}$.*

Proof. $\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|$ so subtracting yields $\|x\| - \|y\| \leq \|x - y\|$. Similarly $\|y\| - \|x\| \leq \|y - x\| = \|x - y\|$. Therefore $\big|\|x\| - \|y\|\big| \leq \|x - y\|$. $\square$

Corollary. *If $(y_n) \to y$ in a Banach space, then $(\|y_n\|) \to \|y\|$ in $\mathbb{R}$.*

Proof. $0 \leq \big|\|y\| - \|y_n\|\big| \leq \|y - y_n\|$ so $\lim_{n\to\infty}\big|\|y\| - \|y_n\|\big| = 0$ by the Squeezing Principle. $\square$

Proposition. *Let $\mathcal{N}$ be a normed vector space and let $\mathcal{B}$ be a Banach space. Then $B(\mathcal{N}, \mathcal{B})$ forms a Banach space under $\|\ \|$.*

Proof. If ϕ is a bounded linear transformation and $\alpha \in \mathbb{F}$, it is clear that $\alpha\phi$ is also a bounded linear transformation with $\|\alpha\phi\| = |\alpha|\|\phi\|$. If ϕ, ϕ' are bounded linear transformations then whenever $x \in \mathcal{B}$ with $\|x\| = 1$ we have $\|(\phi + \phi')(x)\| \leq \|\phi(x)\| + \|\phi'(x)\| \leq \|\phi + \|\phi'\|$ and so $\phi + \phi'$ is a bounded linear transformation with $\|\phi + \phi'\| \leq \|\phi| + \|\phi'\|$. Therefore $B(\mathcal{N}, \mathcal{B})$ becomes a normed vector space under $\|\ \|$.

Suppose now that (ϕ_n) is a Cauchy sequence in $B(\mathcal{N}, \mathcal{B})$. Then for each $x \in \mathcal{N}$ we have $\|\phi_m(x) - \phi_n(x)\| \leq \|x\|\,\|\phi_m - \phi_n\|$ and so $\big(\phi_n(x)\big)$ is a Cauchy sequence in $\mathcal{B}$. Therefore, since $\mathcal{B}$ is a Banach space, $\big(\phi_n(x)\big)$ converges in $\mathcal{B}$. Define $\phi : \mathcal{N} \to \mathcal{B}$ by $\phi(x) := \lim_{n\to\infty} \phi_n(x)$. It is clear that ϕ is a linear transformation. We wish to show that ϕ is bounded (so that $\phi \in B(\mathcal{N}, \mathcal{B})$) and that $(\phi_n) \to \phi$ in $B(\mathcal{N}, \mathcal{B})$.

For any n and m, the Lemma above yields $\big|\|\phi_m\| - \|\phi_n\|\big| \leq \|\phi_m - \phi_n\|$. Therefore the fact that (ϕ_n) is a Cauchy sequence in $B(\mathcal{N}, \mathcal{B})$ imples that $(\|\phi_n)\|)$ is a Cauchy sequence in $\mathbb{R}$ so it converges and in particular it is bounded by some $M \in \mathbb{R}$. Then for all $x \in \mathcal{N}$ with $\|x\| = 1$ we have $\|\phi_n(x)\| \leq M$. According to the Corollary above, $(\|\phi_n(x)\|) \to \|\phi(x)\|$ in $\mathbb{R}$, so taking the limit as $n \to \infty$ gives $\phi(x) \leq M$. Since this is true for all x with $\|x| = 1$, we see that ϕ is bounded so $\phi \in B(\mathcal{N}, \mathcal{B})$.

Given $\epsilon > 0$, there exists N such that $\|\phi_m - \phi_n\| < \epsilon/2$ for all $n, m \geq N$. Suppose $x \in \mathcal{N}$ with $\|x\| = 1$. There exists N_x such that $\|\phi(x) - \phi_k(x)\| < \epsilon/2$ for all $k > N_x$. Choose $m \geq \max\{N, N_x\}$. Then for all $n \geq N$ we have

$$\|\phi(x) - \phi_n(x)\| = \|\phi(x) - \phi_m(x) + \phi_m(x) - \phi_n(x)\| \leq \|\phi(x) - \phi_m(x)\| + \|\phi_m(x) - \phi_n(x)\| < \epsilon.$$

Therefore $\|\phi - \phi_n\| \leq \epsilon$ for all $n \geq N$. Thus $(\phi_n) \to \phi$ in $B(\mathcal{N}, \mathcal{B})$. $\square$

Observe that if $\mathcal{H}$ is a Hilbert space then any $a \in \mathcal{H}$ yields a bounded linear functional $\phi_a \in \mathcal{H}^*$ given by $\phi_a(x) = (x, a)$. This linear functional is indeed bounded since by Cauchy-Schwartz, $|\phi_a(x) = |(x, a)| \leq \|x\| \, \|a\|$ so $\|\phi_a\|$ is bounded with $\phi_a \leq \|a\|$. In fact, we have equality $\|\phi_a\| = \|a\|$ since $\|\phi_a\| \geq \phi_a(a/\|a\|) = (a, a)/\|a\| = \|a\|^2/\|a\| = \|a\|$ gives the reverse inequality.

Proposition. *Let $\mathcal{H}$ be a Hilbert space. Let A and B be bounded linear operators on $\mathcal{H}$ such that $(Ax, y) = (Bx, y)$ for all $x, y \in \mathcal{H}$. Then $A = B$.*

Proof. By subtraction, we may reduce to the case where $B = 0$. Therefore suppose $(Ax, y) = 0$ for all $x, y \in \mathcal{H}$. Setting $y := Ax$ gives $||Ax||^2 = 0$ for all x. Therefore $Ax = 0$ for all x and so A is the zero operator. $\square$

Riesz Representation Theorem. *Let ϕ be a bounded linear functional on a Hilbert space $\mathcal{H}$. Then there exists unique $a \in H$ such that $\phi = \phi_a$.*

Proof. If $\phi_a = \phi_{a'}$ then $(x, a) = (x, a')$ for all $x \in X$. Therefore $(x, a - a') = 0$ for all $x \in X$ and in particular $\|a - a'\|^2 = (a - a', a - a') = 0$ giving $a = a'$. Therefore a, if it exists, is unique.

Let $N = \ker \phi \subset \mathcal{H}$. Then N is a closed linear subspace of $\mathcal{H}$. If $N = \mathcal{H}$ then $\phi = 0$ so we can choose $a = 0$. If $N \neq \mathcal{H}$ then $N^\perp \neq 0$ so choose $b \in N^\perp$ and normalize it so that $\phi(b) = 1$. Then for any $x \in \mathcal{H}$,

$$\phi\big(x - \phi(x)b\big) = \phi(x) - \phi(x)\phi(b) = \phi(x) - \phi(x) = 0$$

so $x - \phi(x)b \in N$. Therefore $(x - \phi(x)b) \perp b$. In other words

$$0 = \big(x - \phi(x)b, b\big) = (x, b) - \phi(x)(b, b) = (x, b) - \phi(x)\|b\|^2.$$

It follows that $\phi(x) = (x, b)/\|b|^2 = \left(x, \frac{b}{\|b\|^2}\right)$. Thus $a := b/\|b\|^2$ has the desired property. $\square$

Corollary. *Let $\mathcal{H}$ be a Hilbert space. Then the map $a \to \phi_a$ is a Banach space isomorphism $\mathcal{H} \cong \mathcal{H}^*$.* $\square$

Theorem (Hahn-Banach). *Let M be a linear subspace of a normed vector space X and let f be a bounded linear functional on M. Then there exists an extension $\hat{f} \in X^*$ of $f \in M^*$ such that $\|\hat{f}\| = \|f\|$.*

Remark. We say that a function g "extends" a function f if Domain $f \subset$ Domain g and $g|_{\text{Domain } f} = f$.

Proof of Theorem.
Case I: $\mathbb{F} = \mathbb{R}$.

By definition of the norm $\| \ \|$ on X^*, the inequality $\|f\| \leq \|g\|$ will be satisfied for any extension. The issue is whether an extension can be chosen in such a way that the reverse inequality holds.

Consider the set $\mathcal{S}$ of all linear functionals g which extend f to some subspace of X containing M and which satisfy $\|g\| = \|f\|$. Partially order this set by defining $g_1 \preceq g_2$

if and only if g_2 extends g_1. Zorn's Lemma implies that the partially ordered set $(\mathcal{S}, \preceq)$ has maximal elements so choose a linear functional $\tilde{f} \in \mathcal{S}$ which is maximal in this partial order. Set $N :=$ Domain f. It suffices to show that $N = X$. Suppose not. Then there exists $e \in X - N$. We will be done if we can show that $\tilde{f}$ can be extended to a linear functional $\hat{f}$ on the linear span of N and e in such a way that $\|\hat{f}\| = \|f\|$.

Since $\hat{f}(x)$ is to be given by $\tilde{f}(x)$ for $x \in N$, we must define $\hat{f}(e)$ appropriately. Set $\hat{f}(e) = \lambda$ where we must show that λ can be chosen so that $\|\hat{f}\| = \|f\|$. Need $|\hat{f}(x + \alpha e)| \leq \|f\| \, \|x + \alpha e\|$ for all $x \in N$ and $\alpha \in \mathbb{R}$. It suffices to prove that $\hat{f}(x + \alpha e) \leq \|f\| \, \|x+\alpha e\|$ for all $x \in N$ and $\alpha \in \mathbb{R}$ because the fact that it holds for both x, α and $-x, -\alpha$ implies that the previous inequality holds for all x, α. Thus we want to choose λ so that $\tilde{f}(x) + \alpha\lambda \leq \|f\| \, \|x + \alpha e\|$ for all x, α. If $\alpha > 0$, the required condition is

$$\lambda \leq \|f\| \left\| \frac{x}{\alpha} + e \right\| - \tilde{f}\left(\frac{x}{\alpha}\right)$$

so we need

$$\lambda \leq \|f\| \, \|z + e\| - \tilde{f}(z) \qquad \text{for all } z \in N \tag{1}$$

If $\alpha < 0$, the required condition is

$$\lambda \geq -\|f\| \left\| \frac{x}{-\alpha} - e \right\| + \tilde{f}\left(\frac{x}{-\alpha}\right)$$

so we need

$$\lambda \geq -\|f\| \, \|w - e\| + \tilde{f}(w) \qquad \text{for all } w \in N \tag{2}$$

Therefore it suffices to show that there exists $\lambda \in \mathbb{R}$ such that both (1) and (2) are satisfied. In other words, we need to know that

$$-\|f\| \, \|w - e\| + \tilde{f}(w) \leq \|f\| \, \|z + e\| - \tilde{f}(z) \tag{3}$$

for all $w, z \in N$. Equation (3) is equivalent to

$$\tilde{f}(w + z) \leq \|f\|(\|z + e\| + \|w - e\|)$$

for all $w, z \in N$. However $\|z + w\| = \|z + e + w - e\| \leq \|z + e\| + \|w - e\|$ and so using $\|\tilde{f}\| = \|f\|$ gives

$$|\tilde{f}(w + z)| \leq \|\tilde{f}\| \, \|w + z\| \leq \|\tilde{f}\|(\|z + e\| + \|w - e\|) = \|f\|(\|z + e\| + \|w - e\|)$$

as desired.

Case II: $\mathbb{F} = \mathbb{C}$.

Given a complex-valued linear functional on M, set $g(x) := \big(f(x) + \overline{f(x)}\big)/2$ and $h(x) := \big(f(x) - \overline{f(x)}\big)/(2i)$. Then g and h are real-valued linear functionals on M which satisfy $\|g\| = \|f\| = \|h\|$ and $f(x) = g(x) + ih(x)$ and

$$g(ix) = f(ix) + \overline{f}(ix)/2 = if(x) - i\overline{f}(x)/2 = -f(x) - \overline{f}(x)/(2i) = -h(x).$$

By Case I, extend g to a real-valued linear functional $\hat{g}$ on X such that $\|\hat{g}\| = \|g\|$. Set $\hat{h}(x) := -\hat{g}(ix)$ and define $\hat{f} : X \to \mathbb{C}$ by $\hat{f}(x) := \hat{g}(x) + i\hat{h}(x)$. It is straightforward to check that $\hat{f}(a+ib)(x) = (a+ib)\hat{f}(x)$ for all $a, b \in \mathbb{R}$, $x \in X$ so $\hat{f}$ is a linear functional. Given $x \in X$, write $\hat{f}(x) = re^{i\theta} \in \mathbb{C}$. Then $\hat{f}(e^{-i\theta}x) = r \in \mathbb{R}$ so $\hat{g}(e^{-i\theta}x) = r$ so and $\hat{h}(e^{-i\theta}x) = 0$. Therefore

$$\|\hat{f}(x)\| = |re^{i\theta}| = r = |\hat{f}(e^{-i\theta}x)| = |\hat{g}(e^{-i\theta}x)| \leq \|\hat{g}\| \, \|e^{-i\theta}x\| = \|f\| \, \|x\|.$$

so $\hat{f}$ is the desired extension. □

10. Banach Algebras

Definition. A *Banach Algebra* consists of a Banach space X together with a mulitplication on X making X into an algebra (with identity) over $\mathbb{F}$ in such a way that $\|xy\| \leq \|x\| \, \|y\|$. A *Banach $*$-Algebra* consists of a Banach algebra X over $\mathbb{C}$ together with an operation $* : X \to X$ such that

1) $1^* = 1$
2) $(x+y)^* = x^* + y^*$
3) $(\alpha x)^* = \bar{\alpha} x^*$ for all $\alpha \in \mathbb{C}$
4) $(xy)^* = y^* x^*$
5) $x^{**} = x$

Example. If $\mathcal{B}$ is a Banach space, then $B(\mathcal{B})$, the collection of bounded linear operators on $\mathcal{B}$ forms a Banach algebra. In this case, the identity operator is sometimes written as I in place of our generic notation 1. If $\mathcal{H}$ is a complex Hilbert space then $B(\mathcal{H})$ forms a Banach $*$-algebra, where A^* is the adjoint operator to A (defined, using the Riesz Representation Theorem, by $(x, A^*y) = (Ax, y)$, as in the next section. In particular, if X is a compact Hausdorff space, then $\mathcal{C}(X; \mathbb{C})$ forms a Banach $*$-algebra.

Definition. Let X be a Banach algebra and suppose $A \in X$. The *spectrum* A, denoted $\sigma(A)$ is defined by $\sigma(A) := \{\lambda \in \mathbb{C} \mid A - \lambda I \text{ is not invertible } \}$. The complement of $\sigma(A)$ is called the *resolvent set*, denoted $\rho(A)$. Thus $\rho(A) := \{\lambda \in \mathbb{C} \mid A - \lambda I \text{ is invertible}\}$.

In the special case $X = B(\mathcal{B})$, we define the *point spectrum* of A, denoted $\pi_0(A)$ is defined by $\pi_0(A) := \{\lambda \in \mathbb{C} \mid Ax = \lambda x \text{ for some nonzero } x \in \mathcal{B}\}$.

Clearly $\pi_0(A) \subset \sigma(A)$. If $\mathcal{B} \cong \mathbb{C}^n$ is finite dimensional, then $X := B(\mathcal{B})$ is isomorphic to the matrix algebra $M_{n\times n}(\mathbb{C})$. In this case, for $A \in M_{n\times n}(\mathbb{C})$, $\sigma(A) = \pi_0(A) = \{\text{eigenvalues of } \ \text{A}\}$. That is, for a matrix, $A - \lambda I$ fails to be invertible if and only if it has a nonzero kernel, in which case it also fails to be surjective. However for an arbitrary operator, $A - \lambda I$ can fail to be invertible either because it fails to injective or because it fails to be injective (or both), and these conditions are not equivalent and thus the spectrum need not equal the point spectrum in general.

Example. Let $\mathcal{B} = l^2$. Define the "right shift operator", $R : l^2 \to l^2$ by

$$R(a_0, a_1, a_2, \ldots, a_n, \ldots) = (0, a_0, a_1, a_2, \ldots, a_n, \ldots)$$

and the “left shift operator”, $L : l^2 \to l^2$ by

$$L(a_0, a_1, a_2, \ldots, a_n, \ldots) = (a_1, a_2, \ldots, a_n, \ldots).$$

Then $\|R\| = \|L\| = 1$ so both R and L lie in $B(l^2)$. Although $L \circ R = I$, neither L nor R is invertible since the composition $R \circ L \neq I$. For matrices this type of behavious cannot happen: for matrices, if $AB = I$ then it is always true that $BA = I$. R is injective but not surjective, and L is surjective but not injective.

Proposition. *Let X be a Banach algebra and let A belong to X. Then $\sigma(A^n) = \sigma(A)^n$.*

Proof. Let $\lambda \in \mathbb{C}$. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ be the solutions to the equation $t^n = \lambda$. Then $t^n - 1 = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n) \in \mathbb{C}[t]$ and so the equation $A^n - \lambda I = (A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_n I)$ holds in X. If $A^n - \lambda I$ is invertible then, since they commute, each of the factors is invertible or equivalently if $A - \lambda_i I$ is noninvertible for some i then $A^n - \lambda I$ is noninvertible. Thus $\sigma(A)^n \subset \sigma(A^n)$. Conversely, if $(A - \lambda_i I)$ is invertible for all i, then their product A is invertible. Hence if $\lambda \in \sigma(A^n)$ then $\lambda_i \in \sigma(A)$ for some i. Therefore $\sigma(A^n) \subset \sigma(A)^n$. □

We will show that $\sigma(A)$ is always a non-empty compact subset of $\mathbb{C}$.

Proposition. *Let X be a Banach algebra and let A belong to X. If $\|1 - A\| < 1$ then A is invertible.*

Proof. Suppose $\|1 - A\| < 1$. Then

$$\left\| \sum_{n=0}^{\infty} (1 - A)^n \right\| \leq \sum_{n=0}^{\infty} \|(1 - A)^n\| \leq \sum_{n=0}^{\infty} \|1 - A\|^n$$

which is a geometric progression and thus converges. Thus there exists $B \in B(\mathcal{B})$ such that $\sum_{n=0}^{\infty} (1 - A)^n = B$. Then

$$A\big(1 + (1 - A) + (1 - A)^2 + \ldots (1 - A)^n\big) = 1 - (1 - A)^{n+1}.$$

so taking the limit as $n \to \infty$ gives $AB = 1$ since $\lim_{n \to \infty} (1 - A)^{n+1} = 0$. Similarly $BA = 1$. □

Corollary. *Let X be a Banach algebra and let A belong to X. If $\|A\| < 1$ then $1 - A$ is invertible with $\|(1 - A)^{-1}\| \leq \frac{1}{1 - \|A\|}$.*

Proof. Apply the preceding proposition to $1 - A$. As above, $(1 - A)^{-1} = \sum_{i=0}^{\infty} A^n$ and so $\|(1 - A)^{-1}\| \leq \sum_{i=0}^{\infty} \|A^n\| \leq \sum_{i=0}^{\infty} \|A\|^n = \frac{1}{1 - \|A\|}$. □

Proposition. *If $|\lambda| > \|A\|$ then $\lambda \in \rho(A)$ with $\|(\lambda - A)^{-1}\| \leq \frac{1}{\lambda - \|A\|}$.*

Proof. $\|A/\lambda\| = \|A\|/|\lambda| < 1$ and so $1 - A/\lambda$ is invertible, or equivalently $\lambda - A$ is invertible. $\|(\lambda - A)^{-1}\| = \frac{1}{|\lambda|} \|\big(1 - (A/\lambda)^{-1}\big)\| \leq \frac{1}{|\lambda|} \frac{1}{1 - \|(A/\lambda)\|} = \frac{1}{\lambda - \|A\|}$. □

The preceding proposition says that $\sigma(A)$ is a bounded subset of $\mathbb{C}$.

Lemma. *Let X be a Banach algebra, let A belong to X, and let $\lambda \in \rho(A)$. Suppose $\mu \in \mathbb{C}$ satifisfies $|\mu - \lambda| < \frac{1}{\|(\lambda - A)^{-1}\|}$. Then $\mu \in \rho(A)$.*

Proof. Since $|\mu - \lambda| < \frac{1}{\|(\lambda - A)^{-1}\|}$, the corollary above says that $(\mu - \lambda)(\lambda - A)^{-1} + 1$ is invertible. Therefore, since $\lambda - A$ is invertible, the product

$$(\lambda - A)\big((\mu - \lambda)(\lambda - A)^{-1} + 1\big) = \mu - \lambda + \lambda - A = \mu - A$$

is invertible. Thus $\mu \in \rho(A)$. □

Corollary. *Let X be a Banach algebra and let A belong to X. Then the resolvent set $\rho(A)$ is an open subset of $\mathbb{C}$.*

Proof. According to the Lemma, if $\lambda \in \rho(A)$ then every μ within a sufficiently small neighbourhood of A lies in $\rho(A)$. Therefore $\rho(A)$ is an open subset of $\mathbb{C}$.

□

Corollary. *Let X be a Banach algebra and let A belong to X. Then $\sigma(A)$ is a compact subset of $\mathbb{C}$.*

Proof. Since $\rho(A)$ is open in $\mathbb{C}$, its complement, $\sigma(A)$, is closed. Also, $\sigma(A)$ is bounded, as shown above. □

The preceding Corollary generalizes the fact that for a matrix A the set of eigenvalues of A is finite, and thus in particular is compact.

Given $A \in X$, define the *resolvent function* of A, $R : \rho(A) \to X$ by $R(z) := (z - A)^{-1}$.

Theorem (Resolvent Equation). *Let X be a Banach algebra, let A belong to X, and let z, z_0 belong to $R(z)$. Then*

$$R(z) - R(z_0) = (z_0 - z)R(z)R(z_0).$$

Proof. The motivation comes from the identity

$$\frac{1}{z - a} - \frac{1}{z_0 - a} = \frac{z_0 - a - z + a}{(z - a)(z_0 - a)} = -\frac{z - z_0}{(z - a)(z_0 - a)}$$

for $a \neq z, z_0 \in \mathbb{C}$. A formal proof of the identity in X is obtained by multiplying out $\big(R(z) - R(z_0)\big)(z - A)(z_0 - A)$ and then simplifying the answer to $z_0 - z$. □

Note: For any $\lambda, \lambda' \in \mathbb{C}$, the elements $A - \lambda$, $A - \lambda'$ commute so the multiplication formulas involving only terms of this form can be manipulated using the same rules that apply to multiplication of elements of $\mathbb{C}$.

Corollary. *$R : \rho(A) \to X$ is continuous.*

Proof. Suppose $z_0 \in \rho(A)$. Then by definition, $A - z_0$ is invertible. Set $B := (A - z_0)^{-1}$ and set $w := z - z_0$. For any $z \in \rho(A)$ we have

$$R(z) = (z - A)^{-1} = \big(w - (A - z_0)\big)^{-1} = \big(w - B^{-1}\big)^{-1} = B(wB - 1)^{-1}$$

To show $\lim_{z \to z_0} R(z) = R(z_0)$, it suffices to consider only z such that $|z - z_0| < 1/(2\|B\|)$, in which case $\|wB\| < 1/2$ and so the lemma above gives $\|wB - 1\|^{-1} \leq \frac{1}{1-\|wB\|} \leq 2$. Therefore for such z we have

$$\|R(z)\| \leq \|B\| \, \|(wB-1)^{-1}\| \leq 2\|B\|.$$

Thus the resolvent equation gives

$$\|R(z) - R(z_0)\| \leq |z_0 - z| \, 2\|B\| \, \|R(z_0)\|$$

and so $\lim_{z \to z_0} \|R(z) - R(z_0)| = 0$. $\square$

Recall, (`MATC34`), that a function $\mathbb{C} \to \mathbb{C}$ is called *holomorphic* if it is complex differentiable.

Proposition. *Let X be a Banach algebra and let A belong to X. Let $\phi : X \to \mathbb{C}$ be a linear functional. Then $\phi \circ R : \mathbb{C} \to \mathbb{C}$ is a holomorphic on the (open) domain $\rho(A)$.*

Proof. Let $f = \phi \circ R$. For $z, z_0 \in \rho(A)$,

$$\frac{f(z) - f(z_0)}{z - z_0} = \frac{1}{z - z_0} \phi\big(R(z) - R(z_0)\big) = \frac{1}{z - z_0} \phi\big((z_0 - z) R(z) R(z_0)\big) = -\phi\big(R(z) R(z_0)\big)$$

Therefore $f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} = -\phi\big(R(z_0)^2\big)$. Thus f is (complex) differentiable at every point in its domain. $\square$

Corollary. *Let X be a Banach algebra and let A belong to X. Then $\sigma(A) \neq \emptyset$.*

Proof. If $A = \lambda 1$ is a multiple of the identity then the Corollary holds since $\sigma(\lambda 1) = \{\lambda\} \neq \emptyset$, so we may assume A is not a multiple of the identity. Suppose $\sigma(A) = \emptyset$. Then $R(z) = \mathbb{C}$ so $\phi \circ R$ is an entire complex function for every linear functional $\phi \in X^*$.

On the compact set $K := \{z \mid |z| \leq 2\|A\|\}$, there exists M such that $\phi \circ R(z) \leq M$ for all $z \in K$ since any continuous function on a compact set is bounded. For z in the complement of K, we have $z > 2\|A\|$ and so

$$\|(z - A)^{-1}\| = \left\| \frac{1}{z\|A\|} \left(1 + \frac{A}{z} + \frac{A^2}{z^2} + \dots\right) \right\| \leq \frac{1}{2\|A\|} \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) = \frac{1}{\|A\|}.$$

Thus $|\phi \circ R(z)| \leq \max\{M, 1/\|A\|\}$ for all $z \in \mathbb{C}$. Hence $\phi \circ R$ is a bounded entire function and so by Liouville's Theorem, (`MATC34`), $\phi \circ R$ is a constant.

Pick $z_1 \neq z_2$ belong to $\mathbb{C}$. Suppose $R(z_1), R(z_2)$ are linearly independent elements of X. Define a linear function ϕ on the 2-dimensional subspace generated by $R(z_1), R(z_2)$ by setting $\phi\big(R(z_1)\big) = 1$ and $\phi\big(R(z_2)\big) = 0$. Use the Hahn-Banach theorem to extend ϕ to a linear functional on X. This gives a contradiction because $\phi \circ R$ is not constant. Therefore $R(z_1), R(z_2)$ are linearly dependent for all $z_1, z_2 \in \mathbb{C}$. That is, for all $z_1, z_2 \in \mathbb{C}$ there exists λ such that $R(z_1) = \lambda R(z_2)$. Equivalently $z_2 - A = \lambda z_1 - \lambda A$. Since $z_2 \neq z_1$, we see that $\lambda \neq 1$ and so we can solve to get $A = \frac{z_2 - \lambda z_1}{1 - \lambda}$ which is a multiply of the identity, contradicting our assumption. Therefore $\sigma(A) = \emptyset$ is not possible. $\square$

11. Operators on a Hilbert Space

Let $A : B \to B'$ be a bounded linear operator between Banach spaces. Define its adjoint $A^* : B'^* \to B^*$ by $A^*(f)(x) := f\big(A(x)\big)$. Notice that if $\|x\| = 1$ then $\|A(x)\| \le \|A\|$ and so $\|A^*(f)(x)\| \le \|f\|\|A\|$. Since this holds for all x having norm 1, it follows that $\|A^*(f)\| \le \|f\|\|A\|$ and thus $\|A^*\| \le \|A\|$ and in particular A^* is bounded.

For the rest of this section we consider the special case where $B = B(\mathcal{H})$ for a Hilbert space $\mathcal{H}$.

For $A \in B(\mathcal{H})$, using the ismorphism $\mathcal{H}^* \cong \mathcal{H}$, we can regard A^* as an element of $B(\mathcal{H})$. Explicitly, for $y \in \mathcal{H}$, A^*y is determined by the equation $(x, A^*y) = (Ax, y)$ for all $x \in X$. It is immediate that $A^{**} = A$ and it follows that $\|A^*\| = \|A\|$.

Example. Let $\mathcal{H} = l^2$. Then the left and right shift operators are adjoints of each other.

In the case where $\mathcal{H} = \mathbb{C}^n$, $B(\mathcal{H}) = M_{n\times n}(\mathbb{C})$ and A^* becomes the conjugate transpose of A.

The following proposition is easily checked.

Proposition. *$B(\mathcal{H})$ forms a Banach $*$-algebra with adjoint as the $*$-operation.*

Generalizing the corresponding familiar notions for matrices.

Definition. An operator $A \in B(\mathcal{H})$ is called

a) *self-adjoint* or *Hermitian* if $A = A^*$. In the finite dimensional case, a real self-adjoint matrix in $A \in M_{n\times n}(\mathbb{R})$ is called *symmetric.*

b) *normal* if $AA^* = A^*A$

c) *unitary* if $A^* = A^{-1}$

Proposition. $\sigma(A^*) = \overline{\sigma(A)}$

Proof. Suppose $A - \alpha I$ is invertible. Thus there exists B such that $B(A - \alpha I) = I = (A - \alpha I)B$. Then taking adjoints gives $B^*(A^* - \bar{\alpha} I) = I = (A^* - \bar{\alpha} I)B^*$ and so $A^* - \bar{\alpha} I$ is invertible. Therefore α lies outside $\sigma(A)$ implies that $\bar{\alpha}$ lies outside $\sigma(A^*)$ and by symmetry the converse is also true. Hence $\sigma(A^*) = \overline{\sigma(A)}$ ☐

Corollary. *If A is self-adjoint then $\sigma(A) \subset \mathbb{R}$.* ☐

Next we examine the spectral theorem for normal operators in the finite dimensional case. In this special case we are dealing with matrices and the results can be proved using the methods of `MATB24`.

Proposition. *Let V be a finite dimensional (complex) vector space. Let $A_1, \ldots, A_k$ be pairwise commuting operators in* $\operatorname{Hom}(V, V)$. *(That is, $A_iA_j = A_jA_i$ for all i and j.) Then there exists an eigenvector common to all the A_i.*

Proof. The proposition is trivial for $k = 1$. Suppose by induction that it is true for $k - 1$. Let x be a common eigenvector for $A_1, \ldots A_{k-1}$. Then for each $j = 1, \ldots k-1$ there exists λ_j such $A_j(x) = \lambda_j x$. Let $W \subset V$ be the linear span of $\{x, A_k x, A_k^2 x, \ldots A_k^s, \ldots\}$. By construction, $A_k(W) \subset W$ and thus the restriction of A_k to W is a linear transformation on W. Let r be the least integer such that $\{x, A_k x, A_k^2 x, \dots A_k^{r-1}\}$ is linearly independent and thus a basis for W. If $w = \sum_{i=1}^{r-1} c_i A_k^i x$ is any element of W then

$$A_j w = \sum_{i=1}^{r-1} c_i A_j A_k^i x = \sum_{i=1}^{r-1} c_i A_k^i A_j x = \sum_{i=1}^{r-1} c_i A_k^i \lambda_j x = \lambda_j w.$$

Thus every element of W is an eigenvector for $A_1, \dots, A_{k-1}$. Regarding A_k as a linear transformation on W it has some eigenvector $y \in W$. Then y is a common eigenvector for $A_1, \dots, A_k$. □

Theorem (Spectral Theorem for finite dimensional normal operators). *Let V be a finite dimensional (complex) inner product space and let $A \in \operatorname{Hom}(V, V)$ be a normal operator. Then V has an orthonormal basis of consisting eigenvectors of A.*

Proof. According to the previous proposition, A and A^* have a common eigenvector x_1. Normalize x_1 so that $\|x_1\| = 1$. Set $W_1 := \langle x_1 \rangle$. Then $A(W_1) \subset W_1$ and $A^*(W_1) \subset W_1$. It follows that if $y \in W_1^\perp$, then for all $w \in W$ we have $(Ay, w) = (y, A^* w) = 0$ since $A^* w \in W_1$. Therefore $A(W_1^\perp) \subset W_1^\perp$ and similarly $A^*(W_1^\perp) \subset W_1^\perp$. Thus we have a decomposition $V \cong W_1 \oplus W_1^\perp$ of the inner product space V which is compatible with the operators A and A^*. By induction, the restriction $A|_{W_1^\perp}$ has an orthonormal basis of eigenvalues of A, and by definition of $W_1^\perp$, they are all orthogonal to the eigenvector x_1. □

Corollary. *Any normal matrix is diagonizable* □

We present without proof the generalization of this spectral theorem to compact normal operators.

A topological space K is called *compact* if every cover of K by open sets has a finite subcover. In the case of subsets of $\mathbb{R}^n$, the Heine-Borel Theorem says that this condition is equivalent to saying that K is closed and bounded, but that does not hold for subsets of an arbitrary Banach space.

We say that an operator $A \in B(\mathcal{B})$ is compact if the closure $\overline{A(S)}$ of the image of any bounded set S is compact.

Compact operators satisfy the following properties.

Proposition.
1) *If the image of A is finite dimensional then A is compact.*
2) *If the image under A of some bounded subset of $\mathcal{B}$ contains the unit ball of some infinite dimensional subset of $\mathcal{B}$, then A is not compact. In particular, the identity operator on $\mathcal{B}$ is not compact unless $\mathcal{B}$ is finite dimensional.*
3) *If A is compact then given any bounded sequence $(x_n) \in \mathcal{B}$, the sequence $\big(A(x_n)\big)$ has a convergent subsequence.*
4) *If A and B are compact then $A + B$ is compact.*
5) *If A is compact then αA is compact for all $\alpha \in \mathbb{C}$.*
6) *If $A \in B(\mathcal{B})$ is compact then AB and BA are compact for all $B \in B(\mathcal{B})$.*

7) The collection of compact operators forms a closed subset of $B(\mathcal{B})$.

In the terminology of `MATC02`, (4)–(6) say the compact operators from a two-sided ideal in the ring $B(\mathcal{B})$. A consequence of (6) and (2) is that a compact operator cannot be invertible unless $\mathcal{B}$ is finite dimensional. Thus when $\mathcal{B}$ is infinite dimensional, $0 \in \sigma(A)$ for every compact operator $A \in \mathcal{B}$.

Proposition. *Let $A \in B(\mathcal{B})$ be compact. Then*
1) The point spectrum (eigenvalues) $\pi_0(A)$ of A is a countable subset of $\mathbb{C}$.
2) Either $\pi_0(A)$ is either finite of consists of a sequence λ_n converging to 0.
3)

$$\sigma(A) = \begin{cases} \pi_0(A) \cup \{0\} & \textit{if } \mathcal{B} \textit{ is infinite dimensional;} \\ \pi_0(A) & \textit{if } \mathcal{B} \textit{ is finite dimensional.} \end{cases}$$

Property (2) is known as the "Fredholm alternative". A variant of the Fredholm alternative appears in the properties of the eigenvalues of self-adjoint differential operators as described in `MATC46`.

Theorem (Spectral Theorem for compact normal operators). *Let $\mathcal{H}$ be a Hilbert space and let $A \in B(\mathcal{H})$ be a compact normal operator. Then $\mathcal{H}$ has an orthonormal (Hilbert space) basis consisting of eigenvectors of A.*

Using measure theory and the notion of a compact Hausdorff space (`MATC27`), there is a generalization of the spectral theorem to normal operators which are not necessarily compact.

Theorem (Spectral Theorem for normal operators). *Let $\mathcal{H}$ be a Hilbert space and let $A \in B(\mathcal{H})$ be a normal operator. Let $\mathcal{A}$ be the Banach subalgebra of $B(\mathcal{H})$ generated by A. Then there exists a compact Hausdorff space X and a measure μ on the Borel sets of X and a norm-preserving Banach $*$-algebra homomorphism $\Phi : \mathcal{A} \to B\big(\mathcal{L}^2(X,\mu)\big)$ such that $\Phi(A)$ is a multiplication operator M_f for some $f \in \mathcal{L}^2(X,\mu)$.*

To see that this generalizes our previous spectral theorem, suppose that A is compact. Set $X := \sigma(A) = \pi_0(A) \cup \{0\}$, $\mathcal{M} := 2^X$; $\mu(S) := \operatorname{Card}(S)$. Define $f \in \mathcal{L}^2(X)$ by

$$f(x) = \begin{cases} \lambda & \text{if } x = \lambda \text{ is an eigenvalue;} \\ 0 & \text{if } x = 0, \end{cases}$$

observing that if 0 is an eigenvalue then the two definitions agree.

Another way to look at the preceding theorem is as follows.

Theorem. *Let $\mathcal{H}$ be a Hilbert space, let $A \in B(\mathcal{H})$ be a normal operator, and let $\mathcal{A}$ be the Banach subalgebra of $B(\mathcal{H})$ generated by A. Let $f : \sigma(A) \to \mathbb{C}$ be given by*

$$f(x) = \begin{cases} \lambda & \textit{if } x = \lambda \textit{ is an eigenvalue;} \\ 0 & \textit{if } x = 0. \end{cases}$$

Then the map $A \to f$ determines a (norm-preserving) Banach $*$-isomorphism from $\mathcal{A}$ to $\mathcal{C}\big(\sigma(A);\mathbb{C}\big)$.

More generally,

Theorem (Gelfand). *Let $\mathcal{A}$ be a commutative Banach $*$-algebra. Then there is a compact Hausdorff space X such that there is a (norm-preserving) Banach $*$-isomorphism from $\mathcal{A}$ to $\mathcal{C}(X;\mathbb{C})$.*

The space X is the spectrum of the ring $\mathcal{A}$ as defined in the subject of algebraic geometry. It is a topological space formed from the collection of prime ideals in $\mathcal{A}$.

12. Additional properties of Banach spaces

Notation. In a metric space (X,d), for any $x \in X$ and $r \in [0,\infty)$ set $B_r(x) := \{y \in X \mid d(y,x) < r\}$, the "open ball of radius r about x", and set $B_r[x] := \{y \in X \mid d(y,x) \leq r\}$, the "closed ball of radius r about x".

Lemma (Baire Category Theorem). *Let (X,d) be a complete metric space and let (S_n) be a sequence of subspaces of X such that none of the sets $\overline{S_n}$ contain a nonempty open set. Then $\cup_n S_n \neq X$.*

Proof. Since $\overline{(S_1)}$ is closed, its complement $\overline{(S_1)}^c$ is open. By hypothesis $\overline{(S_1)}$ does not contain a nonempty open set, so in particular $\overline{(S_1)}$ does not contain all of X. Thus $\overline{(S_1)}^c$ is an open set containing x_1 for some $x_1 \in X$ and so contains the open ball $B_1 = B_{r_1}(x_1)$ for some $r_1 > 0$. Choosing a smaller ball if necessary, we may assume that the radius of r_1 is less than 1. Let $F_1 = B_{r_1/2}[x_1]$ and let $\mathring{F}_1 = B_{r_1/2}(x_1)$. By hypothesis $\overline{(S_2)}$ does not contain a nonempty open set, so in particular $\overline{(S_2)}$ does not contain the open set $\mathring{F}_1$. Thus $\overline{(S_2)}^c$ is an open set containing x_2 for some $x_2 \in \mathring{F}_1$ and so contains the open ball $B_2 = B_{r_2}(x_2)$ for some r_2 with $0 < r_2 < 1/2$. Let $F_2 = B_{r_1/2}[x_2]$ and let $\mathring{F}_2 = B_{r_1/2}(x_2)$. Notice that the construction implies that $F_2 \subset F_1$. Continuing, for each n construct a closed ball F_n such that $F_n \subset F_{n-1}$ and diameter $F_n < 1/2^n$. By construction no point of $\cap_n F_n$ can lie in S_n for any n. According to the Cantor intersection theorem (`MATB43`), $\cap_n F_n$ is a single point and in particular $\cap_n F_n \neq \emptyset$. Thus there exists $x \in X$ which does not lie in $\cup_n S_n$. □

Principal of Uniform Boundedness. *Let X be a Banach space and let Y be a normed vector space. Let $\{A_\alpha\}_{\alpha\in I}$ be a collection of bounded linear operators from X to Y. Suppose that for all $x \in X$ there exists K_x such that $\|A_\alpha x\| \leq K_x$ for all $\alpha \in I$. Then there exists K such that $\|A_\alpha\| \leq K$ for all α.*

Proof. Let $S_n = \{x \mid \|A_\alpha x\| < n \ \forall \alpha\}$. Then $\cup_n S_n = X$ and each S_n is closed by continuity of A_α. By the Baire Category Theorem, there exists n_0 such that $\overline{S_{n_0}}$ contains a nonempty open set. Since S_{n_0} is closed, $\overline{S_{n_0}} = S_{n_0}$. Thus S_{n_0} contains an open set U. Pick $u_0 \in S_{n_0}$. By definition, $\|A_\alpha u\| \leq n_0$ for all $u \in U$ and $\alpha \in I$. Let $T = U - u_0 := \{u - u_0 \mid u \in U\}$. Then T is an open set containing 0 and so T contains the open ball $B_{2r}(0)$ for some $r > 0$, and this in turn contains the closed ball $B_r[0]$. If $t = u - u_0 \in T$, with $u \in U$, then $\|A_\alpha x\| = \|A_\alpha u - A_\alpha u_0\| \leq \|A_\alpha u\| + \|A_\alpha u_0\| \leq n_0 + n_0 = 2n_0$ for all $\alpha \in I$. Any $x \in X$ can be written as $x = \|x\|t/r$ where $t = rx/\|x\| \in B_r[0] \subset T$. Hence for any $x \in X$ and $\alpha \in I$ we have $\|A_\alpha x\| = \frac{\|x\|}{r}\|A_\alpha t\| \leq 2n_0\|x\|/r$. Therefore the theorem holds with $K = 2n_0/r$. □

Corollary 1. *Let X be a Banach space and let Y be a normed vector space. Let (A_n) be a sequence of bounded linear operators from X to Y such that $(A_n x)$ converges in Y for each $x \in X$. Define $A : X \to Y$ by $Ax := \lim_{n\to\infty} A_n x$. Then A is a bounded linear operator.*

Proof. For each $x \in X$, since $(A_n x)$ converges the sequence $\|A_n x\|$ is bounded. That is, there exists K_x such that $\|A_n x\| \leq K_x$ for all n. Thus there exists K such that $\|A_n\| \leq K$ for all n. In other words for each x we have, $\|A_n x\| \leq K\|x\|$ for all n and so taking the limit as $n \to \infty$ gives $\|Ax\| \leq K\|x\|$. Therefore $\|A\|$ is bounded with $\|A\| \leq K$. □

Corollary 2 (Banach-Steinhaus Theorem). *Let X be a Banach space. Let $\{\phi_\alpha\} \subset X^*$ be a collection of bounded linear functionals on X such that for each x there exists K_x such that $|\phi_\alpha| \leq K_x$ for all α. Then there exists K such that $\|\phi_\alpha\| \leq K$ for all α.* □

Definition. Let X be a Banach algebra and suppose $A \in X$. The *spectral radius of A*, denoted $r(A)$ is defined by $r(A) := \sup\{|\lambda| \mid \lambda \in \sigma(A)\}$.

For a matrix A, $r(A)$ is the maximum of the absolute values of its eigenvalues.

It follows from the an earlier proposition that $r(A) \leq \|A\|$. It is possible to have $r(A) < \|A\|$. For example, let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then $r(A) = 0$ but $\|A\| = 1$. Notice however that $A^2 = 0$. There is actually a way of computing $r(A)$ from norms, but we must use not just $\|A\|$ but also the norms of its powers.

Theorem (Spectral Radius Formula). *Let X be a Banach algebra and let A belong to X. Then $\lim_{n\to\infty}(\|A^n\|)^{1/n}$ converges and $r(A) = \lim_{n\to\infty}(\|A^n\|)^{1/n}$*

Proof. If $\lambda \in \sigma(A)$ then for all n we have $\lambda^n \in \sigma(A^n)$ and so $|\lambda^n| \leq \|A^n\|$, or equivalently, $|\lambda| \leq \|A^n\|^{1/n}$. It follows that $r(A) \leq \|A\|^n$ for all n and so

$$r(A) \leq \liminf(\|A^n\|)^{1/n} \leq \limsup(\|A^n\|)^{1/n} \leq \|A\|.$$

To finish the proof, we show that $\limsup(\|A^n\|)^{1/n} \leq r(A)$.

Let $D := \{z \in \mathbb{C} \mid |z| < 1/r(A)\}$ and $D' := \{z \in \mathbb{C} \mid |z| < 1/\|A\|\} \subset D$. If $|w| > r(A)$ then w lies in $\rho(A)$ so the resolvent function $R(w)$ is defined. Thus $R(1/z)$ is defined for all $z \in D$. If z lies in the subset D' then $1 + Az + A^2z^2 + \ldots$ converges and for such z we have

$$R(1/z) = \big((1/z) - A\big)^{-1} = z(1 - zA)^{-1} = z(1 + Az + A^2z^2 + \ldots).$$

Let $\phi \in X^*$ be a linear functional. Define $f : D \to \mathbb{C}$ by $f_\phi(z) := \phi\big(R(1/z)\big)$. The function $f_\phi(z)$ is complex analytic and its power series expansion on the subset D' is given by $z(1 + \phi(A)z + \phi(A^2)z^2 + \ldots)$. The radius of convergence of the power series for $f_\phi(z)$ (about 0) equals the distance from 0 to the closest singularity of $f_\phi(z)$ in the complex plane. Thus the radius of convergence of $z(1 + \phi(A)z + \phi(A^2)z^2 + \ldots)$ is at least $1/r(A)$. Hence the series converges for all z with $|z| < 1/r(A)$ and in particular $\{\phi(z^nA^n)\}$ is bounded for any $|z| < 1/r(A)$ and any linear functional $\phi \in X^*$.

Suppose $|z| < 1/r(A)$. Regard z^nA^n as a linear functional on X^*. Applying the Banach-Steinhaus Theorem to the collection $\{z^nA^n\} \subset X^{**}$ gives a constant K such that

$\|z^n A^n\| \le K$ for all n. Therefore $|z| \le \frac{K^{1/n}}{\|A^n\|^{1/n}}$ for all n. Taking the limit as $n \to \infty$ gives $|z| \le \frac{1}{\limsup \|A^n\|^{1/n}}$. Since this is true for all z with $|z| < 1/r(A)$ we conclude that $\frac{1}{r(A)} \le \frac{1}{\limsup \|A^n\|^{1/n}}$. Thus $\limsup \|A^n\|^{1/n} \le r(A)$, as desired. $\square$

Open Mapping Theorem. *Let $A : X \to Y$ be a surjective bounded linear transformation between Banach spaces and let U be an open subset of X. Then $A(U)$ is an open subset of Y.*

Proof.
Step 1: For any $r > 0$ there exists $r' > 0$ such that $B_{r'}(0) \subset \overline{A\big(B_r(0)\big)}$.

Let $r > 0$ be given. Since $\cup_n \overline{A\big(B_{nr}(0)\big)} = Y$, by the Baire Category Theorem there exists n such that $\overline{A\big(B_{nr}(0)\big)}$ contains a nonempty open set. Since for all $s > 0$ the selfmap of Y given $y \mapsto sy$ takes open sets to open sets, it follows that $\overline{A\big(B_{nr}(0)\big)}$ contains an a nonempty open set for all n and in particular $\overline{A\big(B_{r/2}(0)\big)}$ contains a nonempty open set U. In general, if C, V are any subsets with V open and $\bar{C} \cap V \ne \emptyset$, it is always true that $C \cap V \ne \emptyset$. Therefore $A\big(B_{r/2}(0)\big) \cap U \ne \emptyset$ so there exists $x_0 \in B_{r/2}(0)$ such that that $A(x_0) \in U$. Let

$$U' := U - A(x_0) := \{y \in Y \mid y = u - A(x_0) \text{ for some } u \in U\}$$

be an open set containing 0 obtained by translating U by $A(x_0)$. Then

$$U' \subset \overline{A\big(B_{r/2}(0)\big)} - A(x_0) = \overline{A\big(B_{r/2}(-x_0)\big)}.$$

However for any x, $B_{r/2}(x) \subset B_r(x)$ and so we have $U' \subset \overline{A\big(B_r(0)\big)}$. Since U' is an open set containing 0, there exists $r' > 0$ such that $B_{r'}(0) \subset U'$ and so we get $B_{r'}(0) \subset \overline{A\big(B_r(0)\big)}$.
Step 2: For any $r > 0$ there exists $r' > 0$ such that $B_{r'}(0) \subset A\big(B_r(0)\big)$.

By Step (1), for all $\delta > 0$ there exists $\epsilon_\delta > 0$ such that $B_{\epsilon_\delta}(0) \subset \overline{A\big(B_\delta(0)\big)}$. It suffices to show that $\overline{A\big(B_{\delta/2}(0)\big)} \subset A\big(B_\delta(0)\big)$. Suppose $y \in \overline{A\big(B_{\delta/2}(0)\big)}$. Choose (δ_i) such that $\delta_i > 0$ and $\sum_{i=1}^\infty \delta_i < \delta/2$. Since $y \in \overline{A\big(B_{\delta/2}(0)\big)}$ there exists $x_1 \in B_{\delta/2}$ such that $\|y - Ax_1\| < \epsilon_{\delta_1}$. Equivalently $y - Ax_1 \in B_{\epsilon_{\delta_1}}(0)$ which is contained in $\overline{A\big(B_{\delta_1}(0)\big)}$. Let $\epsilon_2 = \min\{\epsilon_{\delta_1}, \epsilon_{\delta_2}\}$. Then there exists $x_2 \in B_{\delta_1/2}$ such that $\|y - Ax_1 - Ax_2\| < \epsilon_2$. Continuing, get sequence (x_n) such that $x_n \in B_{\delta_{n-1}}$ and

$$\|x_1\| + \|x_2\| + \ldots + \|x_n\| < \delta/2 + \sum_{i=1}^{n-1} \delta_i.$$

Therefore $\sum_{n=1}^\infty \|x_n\| < \delta$. Since $\|\sum_{n=1}^\infty x_n\| \le \sum_{n=1}^\infty \|x_n\|$, completeness of X gives that the series $\sum_{n=1}^\infty x_n$ converges to some $x \in X$ satisfying $\|x\| < \delta$. Since A is continuous,

$$A(x) = A\left(\sum_{i=1}^\infty x_i\right) = \lim_{n\to\infty} \sum_{i=1}^n A(x_i) = y.$$

Thus $\overline{A\big(B_{\delta/2}(0)\big)} \subset A\big(B_\delta(0)\big)$.

Step 3: Conclusion of the proof.

Let $V \subset X$ be open. To show that $A(V)$ is open, we show that every point in $A(V)$ is an interior point within $A(V)$. Suppose $y \in A(V)$. Choose $x \in V$ such that $A(x) = y$. The translated set $V - x$ contains 0 so it contains $B_r(0)$ for some $r > 0$. Then $A\big(B_r(0)\big) \subset A(V) - y$. By Step (2), there exists $r' > 0$ such that $B_{r'}(0) \subset A\big(B_r(0)\big)$. Translating back gives

$$B_{r'}(y) = B_{r'}(0) + y \subset A(v) - y + y = A(V).$$

Thus $A(V)$ contains an open ball about y and so y is an interior point of $A(V)$. □

Corollary (Bounded Inverse Theorem). *Let $A : X \to Y$ be a bijective bounded linear transformation. Then A^{-1} is a bounded linear transformation.*

Proof. Let $B = A^{-1}$. The fact that B is a linear transformation is a standard property of vector spaces. That is, given $y, y' \in Y$ find $x, x' \in X$ such that $A(x) = y$ and $A(x') = y'$. Then $A(x + x') = y + y'$ and so $B(y + y) = x + x' = B(y) + B(y')$ and similarly $B(\alpha y) = \alpha B(y)$. The issue is whether B is continuous. A function $Y \to X$ is continuous if and only the pre-image of every open set is open. Given open $U \subset X$, $B^{-1}(U) = A(U)$ is open by the Open Mapping Theorem. Thus B is continuous. □

If X, Y are Banach spaces it is easy to see that setting $\|(x, y)\| := \|x\| + \|y\|$ defines a norm on $X \oplus Y$ under which $X \oplus Y$ becomes a Banach space. We will use this as our standard norm on $X \oplus Y$ although there are other possible norms we might have chosen. For example $(\|x\|^p + \|y\|^p)^{1/p}$ defines a norm on $X \oplus Y$ for all p and the norm $(\|x\|^2 + \|y\|^2)^{1/2}$ is often used.

Definition. Let X, Y be vector spaces. For $f : X \to Y$, the *graph* $\Gamma(f)$ of f is defined by $\Gamma(f) := \{\big(x, f(x)\big) \mid x \in X\} \subset X \oplus Y$.

If X, Y are normed vector spaces, it is easy to see that if $A : X \to Y$ is a bounded linear transformation then its graph is a closed subset of $X \oplus Y$. If X and Y are Banach spaces then the converse is also true.

Closed Graph Theorem. *Let X, Y be Banach spaces and let $A : X \to Y$ be a linear transformation whose graph is a closed subset of $X \oplus Y$. Then A is bounded.*

Proof. Since $\Gamma(A)$ is a closed linear subspace of the Banach space $X \oplus Y$, it forms a Banach space. Let $P : \Gamma(A) \to X$ be the projection $P(x, y) := x$. Then P is linear. For any $(x, Ax) \in \Gamma(A)$,

$$\|P(x, Ax)\| = \|x\| \leq \|(x, Ax)\|$$

and so P is bounded with $\|P\| \leq 1$. Since P is a bijection, by the bounded inverse theorem its inverse $Q := P^{-1} : X \to \Gamma(A)$ is bounded. Therefore there exists K such that $\|Q(x\| \leq K\|x\|$ for all $x \in X$. Then

$$\|x\| + \|Ax\| := \|x + Ax\| = \|Q(x)\| \leq K\|x\|$$

for all $x \in X$. Thus $\|Ax\| \leq |K - 1|\|x\|$ for all $x \in X$ and so A is bounded. □